Behavioral constructs and neural correlates of descriptive and responsive speech in post-stroke aphasia

Mara Barberis[1,2,6], Ella Eycken[1,2,6], Pieter De Clercq[1,2], Robin Lemmens[2,4,5], Hugo Van hamme[3], Maaike Vandermosten[1,2,6]

[1]Experimental Otorhinolaryngology (ExpORL), Department of Neurosciences, KU Leuven, Belgium

[2]Leuven Brain Institute, KU Leuven, Belgium

[3]Processing Speech and Images (PSI), Department of Electrical Engineering (ESAT), KU Leuven, Belgium

[4]Department of Neurology, University Hospitals Leuven, Belgium

[5]Experimental Neurology, Department of Neurosciences, KU Leuven, Belgium

[6]Leuven Interdisciplinary Language Institute, KU Leuven, Belgium

Corresponding author

Mara Barberis

Herestraat 49 (ON2), PO721

3000 Leuven

Belgium

+3216194566

mara.barberis@kuleuven.be

Abstract

Discourse refers to language that extends beyond a single clause and is used to express ideas, feelings or to interact with others. Persons with post-stroke aphasia often encounter difficulties in discourse, which can severely impede everyday communication. In recent years, discourse has received more attention in both research and clinical practice due to its high ecological validity. However, little is known about the underlying behavioral constructs and neural correlates of discourse in post-stroke aphasia. Therefore, this study explores the behavioral and neural constructs of descriptive and responsive speech in post-stroke aphasia based on micro- and macrostructural linguistic variables. Discourse was assessed in 49 persons with chronic aphasia following left-hemispheric (n=39) or bilateral (n=10) stroke using picture description from the CAT (descriptive speech) and verbal responses to everyday scenarios from the ANELT (responsive speech). Factor analysis and voxel-wise lesion symptom mapping were performed to investigate the underlying behavioral and neural constructs. Factor analysis revealed four latent constructs that explained 64% of the total variance in the discourse variables. These latent constructs mainly had high loadings on grammatical, phonological, lexical-semantic and macrostructural variables, with fluency variables spread across the different factors. After correction for multiple comparisons, only the lexical-semantic factor was significantly associated with a left cluster covering inferior parietal, superior and middle temporal and lateral occipital regions. In conclusion, this study informs us about the underlying behavioral constructs and neural correlates of descriptive and responsive speech, and highlights the need for further research on macrostructural deficits in persons with post-stroke aphasia.

Keywords

Stroke; Aphasia; Discourse; Connected speech; Voxel-wise Lesion Symptom Mapping

Highlights

- Discourse relied on 4 constructs: grammatical, phonological, lexical-semantic, macrostructure
- Fluency variables were distributed across the 4 constructs
- Lexical-semantic deficits were associated with a temporoparietal-to-occipital cluster

1. Introduction

Discourse refers to language that extends beyond a single clause and is used to express ideas, feelings or to interact with others (Armstrong, 2000; Kong, 2023) and is thereby the most ecologically valid level of expressive language (Doedens & Meteyard, 2020; Fromm et al., 2022; Gordon, 2020). In persons with post-stroke language difficulties (aphasia), discourse can be impaired in both microstructure, i.e., phonological, lexical and syntactic processes, and macrostructure, i.e., organization and regulation of meaningful, coherent speech (Armstrong, 2000). Discourse impairments often lead to frustration, reduced quality of life, psychosocial stress or social isolation (Stark, 2023). As a consequence, improving discourse is paramount for persons with post-stroke aphasia and their relatives (Wallace et al., 2017; Worrall et al., 2011). Therefore, clinicians and researchers have been advocating to systematically include discourse in aphasia assessment (Azios et al., 2021; Dietz & Boyle, 2018).

Recently, there has been a paradigm shift from isolated language tasks (e.g., confrontation naming, word repetition) to more naturalistic discourse tasks (e.g., picture description, personal narratives). This shift has led researchers to investigate the underlying behavioral constructs of discourse in post-stroke aphasia (e.g., Casilio et al., 2019; Gordon, 2020; Vermeulen et al., 1989) and its neural correlates (e.g., Alyahya et al., 2020, 2021, 2022; Angelopoulou et al., 2024; Borovsky et al., 2007; Casilio et al., 2024; Ding et al., 2020, 2023; Gleichgerrcht et al., 2021; Stark et al., 2019; Wilmskoetter et al., 2019; Zevgolatakou et al., 2023), since discourse requires interaction and context integration of isolated linguistic stimuli (Garrett, 1984; Hamilton & Huth, 2020). Earlier studies may not have captured the neural underpinnings of macrostructural aspects of discourse, such as discourse planning or coherence. There is extensive evidence that the neural language network (typically assessed with isolated language tasks) shows involvement of left-hemispheric frontal and temporoparietal areas (Fedorenko et

al., 2024), but less is known on the neural correlates of discourse. The few studies conducted in this area, suggest that discourse production relies on a bilateral, but left-dominant neural network beyond the traditional perisylvian areas, including frontal, anterior and posterior temporal, and parietal regions, and their underlying white matter tracts (Alyahya, 2023; Schnur et al., 2023). The left lateralization observed in discourse tasks is however less pronounced than that observed in isolated language tasks (Hamilton & Huth, 2020).

Certain studies specifically investigated the latent constructs of discourse, and subsequently linked these constructs to stroke-induced brain lesions using lesion and/or disconnection mapping analysis. Except for the study of Casilio et al. (2024), all studies were restricted to left-hemispheric stroke-induced lesions, hence mainly revealing information about the left-hemispheric language network. Alyahya et al. (2020) found that the number of (distinct) content words, informativeness and speech rate from narrative, descriptive and procedural discourse tasks administered in persons with chronic post-stroke aphasia could be captured in three components: verbal quantity (number of (distinct) content words), verbal quality (informativeness) and motor speech (words per minute). Alyahya and colleagues concluded that these components were mainly associated with left frontal regions and some deep temporal structures, as well as their underlying white matter tracts. Ding et al. (2020) assessed the content and structure of a narrative (storytelling) task in the acute stage post-stroke. Principal component analysis of thirteen discourse variables derived from quantitative production analysis (Saffran et al., 1989) resulted in four components: structural complexity, lexical selection, syntactic accuracy and production fluency. Lesions in the left temporal-parietal region were associated with lexical selection and structural complexity (i.e., word production within increasingly complex combinations). Lesions in the inferior frontal cortex on the other hand hampered syntactical accuracy. Production fluency (words per minute) was associated

with lesions in the left precentral and postcentral gyrus, posterior superior temporal sulcus, inferior parietal lobe, insula, thalamus and basal ganglia. Gleichgerrcht et al. (2021) automatically extracted twenty syntactic variables from three picture description tasks administered in persons with chronic aphasia. Discourse variables were then reduced to two components: sentence complexity and sentence production. Sentence production was linked to the left pars opercularis of the left inferior frontal gyrus, the precentral gyrus, the postcentral gyrus and the supramarginal gyrus. Critical regions identified for sentence complexity were the pars opercularis and pars triangularis as well as the left insula. This finding differs from that of Ding et al. (2020), who associated sentence complexity (cf. 'structural complexity') with temporo-parietal regions. Zevgolatakou et al. (2023) included 21 speech production variables, thirteen of which were obtained from quantitative production analysis (Saffran et al., 1989) that measured morphological, structural and lexical properties based on a narrative (storytelling) task in persons with chronic aphasia. Principal component analysis with bifactor rotation resulted in four components. The first component 'complexity' was linked to a large temporo-parietal region and reflects overall aphasia severity and the length and complexity of discourse. Next, a lexical syntax factor (consisting of lexical items with a syntactic function) spanned a relatively small set of fronto-parietal regions and finally, semantics and phonology were identified as independent components but no significant neural correlates were found. Recently, Casilio et al. (2024) investigated the neural correlates of the four spontaneous speech-based APROCSA dimensions (Casilio et al., 2019) in persons with acute post-stroke aphasia. Paraphasia (misselection of words and sounds) and logopenia (paucity of words) were primarily associated with posterior regions spanning both the dorsal and ventral streams. Agrammatism and motor speech were primarily associated with anterior regions of the dorsal stream, which was also found in the studies of Alyahya et al. (2020), Ding et al. (2020) and Gleichgerrcht et al. (2021).

The present study explores the underlying behavioral and neural constructs of discourse in post-stroke aphasia and thereby contributes to the relatively understudied field of discourse production. Our study adds several novel aspects to the studies discussed above. Previous studies on the underlying constructs of discourse and their neural correlates have been focusing on a single discourse task, eliciting descriptive, procedural or narrative speech. Other studies have merged the output of multiple discourse tasks prior to extracting discourse variables. In the current study, discourse variables are extracted per task, acknowledging that different tasks may challenge the language system in different ways (Fergadiotis & Wright, 2011; Schnur & Wang, 2024; Stark, 2019), allowing us to investigate whether two discourse tasks (descriptive and responsive speech) yield similar underlying behavioral constructs. Furthermore, this study is – to the best of our knowledge – the first study to include responsive speech. Responsive speech was elicited using the ANELT (Blomert et al., 1995; Ruiter et al., 2016), in which individuals with aphasia have to verbally respond to everyday life situations (e.g., rescheduling an appointment), presented by the test administrator in a script-like manner. Second, the majority of studies are limited to microlinguistic discourse variables, while discourse is preeminent for macrolinguistic investigations. Therefore, this study includes discourse variables spanning different subdomains of language, including both micro- and macrolinguistic variables. We hypothesize that this will result in behavioral constructs that reflect both micro- and macrostructural aspects of discourse production. On the neural level, we hypothesize that microstructural discourse production will mainly be associated with the left core language network (Fedorenko et al., 2024), whereas macrostructural discourse production is likely to be associated with a more bilateral neural network beyond the traditional perisylvian areas (Alyahya, 2023; Schnur et al., 2023).

2. Methods

2.1. Participants

Demographic and clinical characteristics of the participant group are presented in Table 1. This study comprises 49 persons (25 males, mean age 67 years) who were diagnosed with aphasia by a speech-language pathologist in the acute stage following ischemic (n = 38) or hemorrhagic (n = 11) stroke. Participants were selected from a larger, ongoing aphasia study based on the availability of acute neuroimaging data. Both left-hemispheric (n = 39) and bilateral (n = 10) strokes were included. Hand preference data were available for 39 participants (30 right-handed, 9 left-handed). Persons with a psychiatric or neurodegenerative disorder were excluded. Brain imaging data were acquired in the (sub)acute stage post-stroke, while discourse data were collected in the chronic stage post-stroke. In the chronic stage, aphasia severity, naming abilities and communicative effectiveness were additionally assessed to clinically characterize the participant group (Table 1). Participants did not have to meet a particular aphasia type or severity level to be included, as we want to investigate a representative sample of the population with aphasia. Participants had to score below the cut-off on the ScreeLing (Visch-Brink et al., 2010), the Dutch Naming Test (Van Ewijk et al., 2018) or report communication difficulties according to the Communicative Effectiveness Index (Lomas et al., 1989) to ensure that they still had aphasia in the chronic phase post-stroke. All participants gave written consent to participate in the study. The study was conducted in accordance with the Declaration of Helsinki and approved by the Medical Ethical Committee of the University Hospitals Leuven and the University of Leuven (B322201731747, approved on 08 March 2017).

Table 1: Demographic and clinical characteristics of the participant group.

	Mean (standard deviation)	Range
Age (years)	67.04 (13.83)	38 – 88
Time post onset for behavioral assessment (months)	31.31 (23.34)	4 – 115
Time post onset for neuroimaging (days)	4.47 (6.61)	0 – 32
Lesion volume (ml)	78.99 (94.01)	0.33 – 426.94
Aphasia severity[1]	60.62 (12.60)	17 – 72
Semantics	21.22 (3.77)	10 – 24
Phonology	19.50 (5.07)	5 – 24
Syntax	19.90 (4.73)	2 – 24
Naming abilities[2]	215.69 (68.68)	7 – 274
Communicative effectiveness[3]	6.75 (2.17)	2.38 – 10.00

[1]Based on the ScreeLing (Visch-Brink et al., 2010). Total scores < 68 are indicative of aphasia, maximum score = 72. The ScreeLing contains three subtests (semantics, phonology and syntax). Subtest scores < 22 are indicative of a deficit in the respective language domain (18-21.5 mild deficit; 14-17.5 moderate deficit; 10-13.5 severe deficit; <10 very severe deficit), maximum subtest score = 24. [2]Based on the Dutch Naming Test (Van Ewijk et al., 2018). Total scores lower than 255 are indicative of aphasia, maximum score = 276. [3]Based on the sixteen-item Communicative Effectiveness Index (CETI) self-report questionnaire (Lomas et al., 1989). Scores on the sixteen items are averaged to a score on ten, where lower scores reflect lower communicative effectiveness.

2.2. Discourse tasks

Discourse was elicited using two monologic discourse tasks, comprising descriptive and responsive speech. Descriptive speech was prompted using the picture description task from the Dutch adaptation of the Comprehensive Aphasia Test (CAT) (Visch-Brink et al., 2015). Participants were given five minutes to describe a black and white picture. Task instruction was as follows: 'Look carefully at this picture, and then describe what is happening using complete sentences'. If participants were known to have hemianopia or neglect and essential parts of the picture were not described, the test administrator was allowed to ask an extra question ('what is happening here?' while pointing at the omitted information in the picture). This extra question was thus not asked to elicit additional language, but to minimize the influence of comorbid visual deficits. Responsive speech was elicited using the revised version of the Amsterdam Nijmegen Everyday Language Test (ANELT) (Ruiter et al., 2016). The ANELT consists of ten items, preceded by two test items in which the test administrator is allowed to give feedback about the correctness of the response. In the ANELT, persons are asked to verbally respond to everyday situations in a script-like manner (Blomert et al., 1995). For example, the test administrator says: 'You call the municipality. You want to speak to someone from the municipality but apparently the number is incorrect. You get me on the line by mistake. What do you say?' In addition, three ANELT items include a picture of an object. For example, the task instruction of ANELT item 4 is as follows: 'This [show picture 4: broken glasses] is yours. You are now in the eyewear store. I am the salesperson. What do you say to me?' In contrast to the picture description task, there are no time restrictions implemented in the ANELT. There was no minimum speech output requirement to be included in the analysis, as we did not want to exclude severe cases of aphasia. Two individuals with very severe aphasia could however not complete the ANELT task. Individuals with aphasia produced on average 57.57 words in the CAT (SD = 51.91 words) and 424.35 words in the ANELT (SD = 305.56 words). Audio

was recorded in a quiet space at the participant's home using a BOYA BY-M1 Universal Lavalier Microphone with a sample rate of 16 kHz and a bit-depth of 16.

2.3. Transcription

Transcripts were obtained using a Dutch automatic speech recognition model (Poncelet & Van hamme, 2022), which was trained on 270 hours of Belgian Dutch speech from the Spoken Dutch Corpus (Oostdijk, 2000) and finetuned on nine hours of discourse (personal narratives) from eighteen cognitively healthy older persons (9 male, mean age 73 years). To ensure conformity of the transcriptions and subsequent labeling, all annotators (two second-year master's students speech and language pathology and one research assistant) received a 1.5 hour training from author M.B., followed by three exercise samples after which feedback was provided on adherence to the orthographic transcription protocol, utterance splitting and correctly assigning aphasia-specific labels. Human annotators manually corrected the automatically generated transcripts and simultaneously labeled unintelligible speech, non-speech sounds and disfluencies (e.g., word repetition, false start) according to the orthographic transcription protocol of the Spoken Dutch Corpus (Taalunie, 2000), which is the most standardized transcription system available for Dutch. This protocol was further adapted for aphasic speech by adding labels for semantic errors, phonemic errors, neologisms and filled pauses (Barberis et al., 2023). The guidelines of ASTA (Boxum et al., 2013) were used for utterance splitting and mean length of utterance (MLU) calculation. ASTA is a well-established approach for investigating spontaneous speech in Dutch-speaking individuals with aphasia. According to ASTA, an utterance is defined as a grammatical unit. If no grammatical unit can be identified, an utterance can be determined by a decreasing intonation pattern. The interrater reliability of utterance splitting and MLU calculation according to the ASTA approach is reported to be excellent (ICC = 0.98) by van der Scheer et al. (2011).

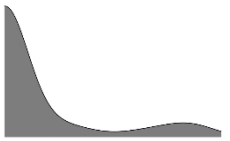
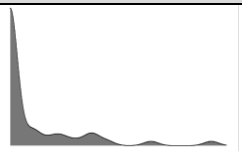
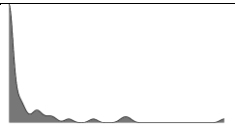
2.4. Discourse variables

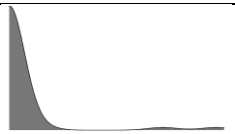
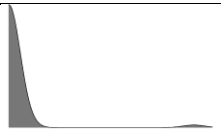
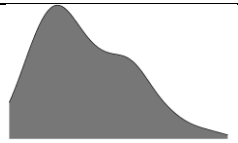
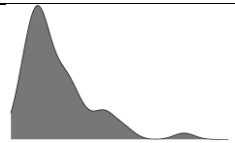
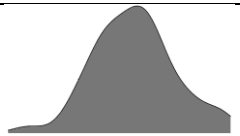
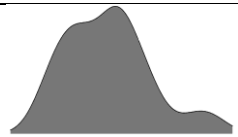
Sixteen discourse variables that span different subdomains of language were included in this study (Table 2). Semantic and phonemic paraphasias as well as neologisms were manually labeled and included as a proportion of the total number of words in the sample. On the lexical level, number of words and the Brunet Index (BI) were calculated. Brunet Index (a text-length independent measure) was chosen over Type Token Ratio and Honoré's Statistic (taking into account words only occurring once) as a measure of lexical diversity, since previous research, using a sample that largely overlapped with the current sample, has shown that the Brunet Index is more relevant to differentiate between individuals with aphasia and neurotypical controls compared to Type Token Ratio and Honoré's Statistic (Barberis et al., 2024). Grammatical discourse variables included mean length of utterance (MLU), noun-verb rate and the proportion of subordinate clauses. Although mean length of utterance can also be affected by lexical retrieval deficits, MLU was – in accordance to the review of Bryant et al. (2016) – used as a proxy of grammatical complexity in the current study. To investigate macrostructure, we included the proportion of utterances that do not contain key information (digression). Finally, fluency was included by means of the speech rate (words per minute), filled pauses, long (> 400 ms) and short (150-400 ms) pauses (Le et al., 2018; Pakhomov et al., 2010), false starts, abandoned words and repetitions within a word.

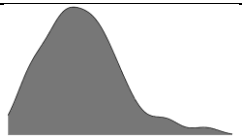
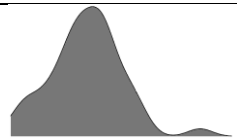
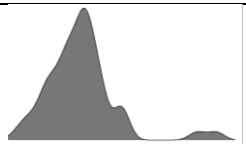
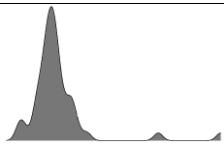
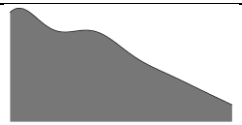
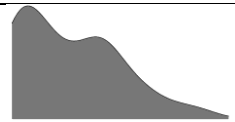
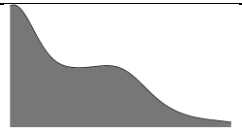
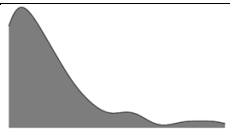
Discourse variables were extracted from the samples in both an automatic and manual manner. The first part of the automatic feature extraction pipeline involved chunking of the audio recording. Robust Voice Activity Detection (Tan et al., 2020) was used to find segments of voice activity. Next, transcripts were obtained for each segment using the automatic speech recognition model described in the previous section. Montreal Forced Aligner (McAuliffe et al., 2017) was used to calculate word-level and phone-level timings from the generated

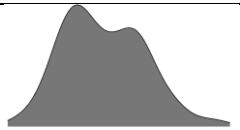
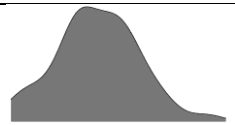
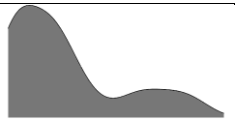
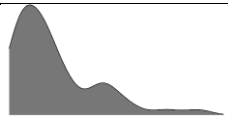
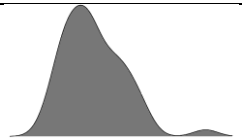
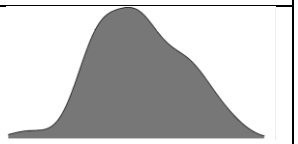
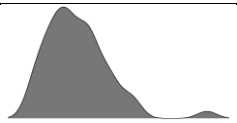
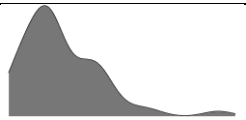
transcripts. From this output, automatic features such as word count, speech rate and pauses were derived. The manually extracted features included counting the occurrences of each of the aphasic errors, labeled while correcting the automatically generated transcripts. The sixteen variables were extracted for each discourse task separately, given that different discourse elicitation methods can challenge the language system in different ways (Fergadiotis & Wright, 2011; Schnur & Wang, 2024; Stark, 2019).

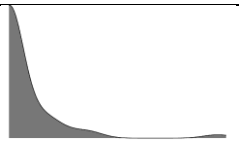
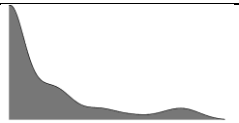
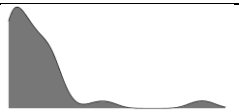
Table 2: Overview of the included discourse variables, their operationalization and extraction method, grouped by subdomain of language. The discourse variables are extracted for both descriptive (picture description from CAT) and responsive (ANELT) speech tasks. Their density plot, mean, standard deviation (SD) and range are included in the final two columns.

Subdomain of language	**Discourse variable**	**Operationalization**	**Extraction method**	**CAT** **Density plot** **Mean (SD)** **Range**	**ANELT** **Density plot** **Mean (SD)** **Range**
Semantic	Semantic paraphasia	Substitution of content words for (un)related content words (Casilio et al., 2019), calculated as a proportion of the total number of words.	Manual	 0.00298 (0.00666) 0 – 0.02247	 0.00080 (0.00157) 0 – 0.00744
Phonology	Phonemic paraphasia	Phoneme deletion, insertion, substitution or transposition (Casilio et al., 2019; Vermeulen et al., 1989), calculated as a proportion of the total number of words.	Manual	 0.01245 (0.02785) 0 – 0.15000	 0.00289 (0.01075) 0 – 0.07273

	Neologism	Nonwords, i.e., word forms that are not real words (Casilio et al., 2019), calculated as a proportion of the total number of words.	Manual	 0.00273 (0.01192) 0 – 0.06742	 0.00133 (0.00796) 0 – 0.05455
Lexical	Number of words	Total number of words produced, including non-words, phonemic language errors, repetitions, minimal responses, comments and stereotypes, in accordance with Boxum et al. (2013) and Vandenborre et al. (2018).	Automatic	 87.57143 (51.91379) 11 – 227	 424.34043 (305.56443) 48 – 1594
	Brunet Index	Text-length independent measure of lexical diversity. Calculated as $W^{(u^{-0.165})}$ where w = total number of word tokens and u = total number of unique word types (Parsapoor et al., 2021). Lower values indicate more diverse word use (Sanborn et al., 2020).	Automatic	 10.53571 (1.45511) 6.72 – 13.72	 14.19 (0.95027) 12.66 – 16.66

Grammatical	Mean length of utterance	Average number of words per utterance, based on the guidelines provided by (Boxum et al., 2013).	Manual	 8.23494 (4.39346) 1 – 22	 6.11511 (1.94342) 2 – 12
	Noun-verb rate	Total number of nouns divided by the total number of verbs, excluding auxiliaries and modals.	Manual	 1.28973 (0.62987) 0.18182 – 3.66667	 0.71705 (0.44769) 0.17500 – 3.00
	Subordinate clause	Clause that forms part of and is dependent on a main clause, often introduced by a conjunction. Calculated as a proportion of the total number of utterances.	Manual	 0.25046 (0.23487) 0 – 0.80	 0.13387 (0.11617) 0 – 0.43636
Macrostructure	Digression	Utterance that does not contain key information, but is straying from the main topic. Calculated as a proportion of the total number of utterances.	Manual	 0.15948 (0.17957) 0 – 0.66667	 0.18508 (0.21408) 0 – 0.95122

Fluency	Speech rate	Total number of words divided by the total duration in minutes.	Automatic	 86.44653 (36.30977) 12.49 – 185.05	 96.22128 (26.86249) 40.07 – 168.12
	Filled pause	Filler words (e.g., 'uh') calculated as a proportion of the total number of words.	Automatic	 0.05249 (0.05232) 0 – 0.18056	 0.02941 (0.02763) 0 – 0.12348
	Short pause	Number of silent regions between words that are 150-400 ms (Le et al., 2018; Pakhomov et al., 2010) divided by the total number of words.	Automatic	 0.18876 (0.06873) 0.08 – 0.44	 0.17851 (0.04278) 0.07 – 0.27
	Long pause	Number of silent regions between words that are > 400 ms (Le et al., 2018; Pakhomov et al., 2010) divided by the total number of words.	Automatic	 0.22008 (0.10279) 0.076 – 0.63	 0.19383 (0.07097) 0.1 – 0.47

	False start	A word is temporarily interrupted, but the target word is eventually produced (e.g., ca-camera). Calculated as a proportion of the total number of words.	Manual	 0.00489 (0.01286) 0 – 0.07692	 0.00275 (0.00729) 0 – 0.04545
	Word abandoned	Words are abandoned after one or two phonemes; the target word is not produced. Calculated as a proportion of the total number of words.	Manual	 0.01054 (0.01680) 0 – 0.06044	 0.00726 (0.01533) 0 – 0.09091
	Word repetition	Immediate repetition of a previously used word, calculated as a proportion of the total number of words.	Automatic	 0.03428 (0.04694) 0 – 0.23077	 0.01358 (0.01835) 0 – 11.220

265

2.5. Factor analysis

Factor analysis was conducted using the *psych* package in R (version 4.4.1) on the initial dataset containing 32 variables (sixteen variables for each of the two discourse tasks) to reveal the underlying behavioral constructs of descriptive and responsive speech. Orthogonal varimax rotation was applied, as this minimizes the number of variables with high loadings on each factor, resulting in a more interpretable factor solution. Factors were extracted according to the minimal residuals method, given the small sample size and non-normally distributed data. The number of factors was determined by parallel analysis. Variables with loadings below 0.4 or a high uniqueness (> 0.75) were removed, as this indicates that they do not contribute meaningfully to the factor structure. The resulting factor solution is interpreted based on the loading of the remaining variables to identify the underlying behavioral constructs.

2.6. Voxel-wise Lesion Symptom Mapping

Lesion masks were segmented on T2-weighted FLAIR images (n = 45) and on non-enhanced contrast computed tomography images (n = 4). The lesions were manually segmented using MRIcron (v. 02092019), guided by the radiological report. After segmentation, all brain scans and lesion masks were normalized to the template for older individuals in MNI space with the Clinical Toolbox implemented in SPM (Rorden et al., 2012). Finally, lesion masks were smoothed using a Gaussian kernel with a size of 4 mm using SPM12 (Welcome Trust Centre for Neuroimaging, London, UK), implemented in MATLAB 2021b (MathWorks, Massachusetts, USA). The Harvard-Oxford atlas (Kennedy et al., 1998) was used to define brain regions.

Based on the lesion masks, Voxel-wise Lesion Symptom Mapping (VLSM) was used to derive neural correlates of the discourse variables extracted from descriptive and responsive speech

(cf. Section 2.4). This technique allows for a whole-brain (voxel per voxel) analysis of the relationship between brain lesions and behavioral variables, by setting up a linear regression model where behavioral outcomes are predicted by the (binary) lesion status of brain voxels (Bates et al., 2003). In this study, univariate VLSM was performed for each of the four factors derived from the behavioral discourse variables (cf. Section 2.5) to allow for an analysis of the neural correlates of the underlying latent constructs of descriptive and responsive speech. Further, voxels damaged in less than 10% of the subjects (corresponding to five subjects in our analysis) were discarded to minimize statistical power loss (Baldo et al., 2022). As a result, only left-hemispheric lesions were analyzed (Figure 1). The lesion overlap plot in Figure 1 reveals a broad distribution of damage across left-hemispheric regions. The most prominently lesioned areas are spread across the frontal lobe (inferior and middle frontal gyri, precentral gyrus, frontal orbital cortex), the parietal lobe (postcentral gyrus, supramarginal gyrus, angular gyrus), temporal lobe (middle temporal gyrus), occipital lobe (lateral occipital cortex) and insular and opercular regions. The linear regression model was further adjusted to correct for multiple comparisons across voxels and to adjust for confounding factors, as suggested by Baldo et al. (2022) and as implemented in previous studies (Stark et al., 2019; Wilson et al., 2010). Age, months post stroke onset and lesion volume were implemented as confound variables and were thus regressed out of the data. To correct for multiple comparisons, cluster-based permutation testing (Winkler et al., 2014) was performed based on the method implemented in Stark et al. (2019). As described in their study, voxel-wise 'uncorrected' statistics (one-tailed, $p < 0.05$) were computed by comparing the behavioral (i.e., factor) score in each voxel in subjects with and without a lesion in that voxel. Note that in our study positive z-statistics ($z > 1.645$) were analyzed, as lesioned voxels are assumed to have higher accordance to the factor compared to non-lesioned voxels. Only voxels that survived threshold (i.e., voxels that were significant) were retained and further corrected for multiple comparisons by setting

clusters of surviving voxels against a cluster threshold. The cluster threshold was defined by the following procedure. First, voxel-wise statistics (one-tailed, $p < 0.05$) were computed again, although now for 5000 permutations of the behavioral data (i.e., where the behavioral score is randomly shuffled across subjects), and again only significant voxels were retained. Second, for each permutation the size of the largest cluster of surviving (significant) voxels was estimated, after which all largest cluster sizes across all 5000 permutations were ranked. In the last step, the 250th largest of the 5000 ranked cluster sizes was set as the 'corrected' cluster threshold (corresponding to a corrected threshold of $p < 0.05$). As such, for each latent factor a map of 'corrected' z-statistics was computed, representing cluster(s) of (lesioned) voxels that survived the cluster threshold and that thus correlated significantly with that factor. Lastly, cluster plots and cluster distribution maps were generated to further analyze the significant cluster(s). All analyses were conducted in NiiStat implemented in MATLAB (http://www.nitrc.org/projects/niistat) and Python (Python Software Foundation. Python Language Reference, version 3.12. Available at http://www.python.org).

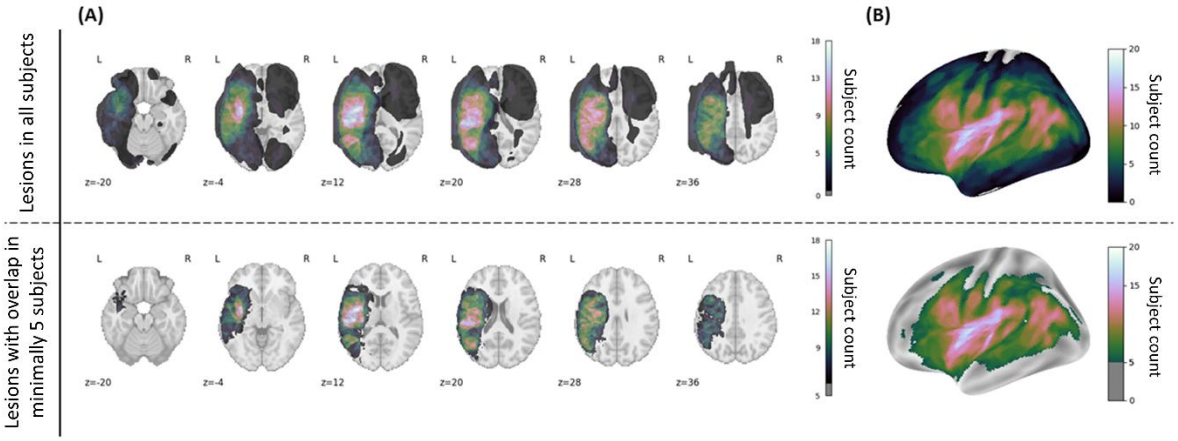


Figure 1: Lesion overlap plots. Overlap plots containing lesions in all subjects are visualized against overlap plots containing lesions in minimally five subjects. Panel (A) shows axial slices as located in the MNI-space by means of their respective MNI-z-coordinates. In panel (B),

surface plots of the left hemisphere depict the lesion overlap. Color grading is used to visualize the number of participants sharing a lesion in the depicted voxels.

3. Results

3.1. Underlying constructs of descriptive and responsive speech

A factor analysis with varimax rotation was conducted to investigate the underlying behavioral constructs of descriptive and responsive speech in post-stroke aphasia. From the initial set of 32 variables, six variables (CAT phonemic paraphasia, CAT neologism, CAT noun-verb rate, CAT word abandoned, CAT short pause and ANELT short pause) were removed as they did not meaningfully contribute to the factor structure (loadings < 0.4 or uniqueness > 0.75).

Based on the remaining 26 variables (fifteen from the ANELT and eleven from the CAT), four factors were retained, explaining 64% of the total variance. Factor loadings are presented in Figure 2. Factor 1 explained 22% of the total variance and had high loadings on variables of grammatical complexity (subordinate clauses and mean length of utterance) from both the ANELT and CAT, hence this factor was interpreted as 'grammatical'. In addition, two fluency variables (speech rate and long pauses from both ANELT and CAT) and one macrostructural variable (digression rate from ANELT) loaded on the grammatical factor, though note that digression rate loaded only slightly above the absolute cut-off of 0.4. Factor 2 accounted for 21% of the total variance and was characterized by high loadings on both phonological variables (phonemic paraphasia and neologism from ANELT), one grammatical variable (ANELT noun-verb rate) and three fluency variables (ANELT word abandoned, ANELT and CAT false start and CAT word repetition). Given the dominant loadings for phonemic paraphasia (0.99), neologism (0.95) and false start (0.94 for ANELT and 0.83 for CAT), this factor was named 'phonological'. Factor 3 (lexical-semantic) explained 13% of the total variance and showed

high loadings on all lexical-semantic variables (number of words, Brunet Index and semantic paraphasia) from both the ANELT and CAT. In addition, two ANELT variables (digression and word repetition) had loadings > 0.4 on this third factor. Factor 4 (macrostructure) accounted for 8% of the total variance and had the highest loadings for filled pauses, followed by word repetition of the ANELT and digression of both the ANELT and CAT. Digression loaded negatively on macrostructure, while filled pauses and ANELT word repetition loaded positively on macrostructure. In conclusion, the four factors were largely consistent with the theoretical categorization of the variables according to the subdomains of language in Table 2, with the difference being that fluency variables were spread across all four factors, rather than clustering in a factor of its own.

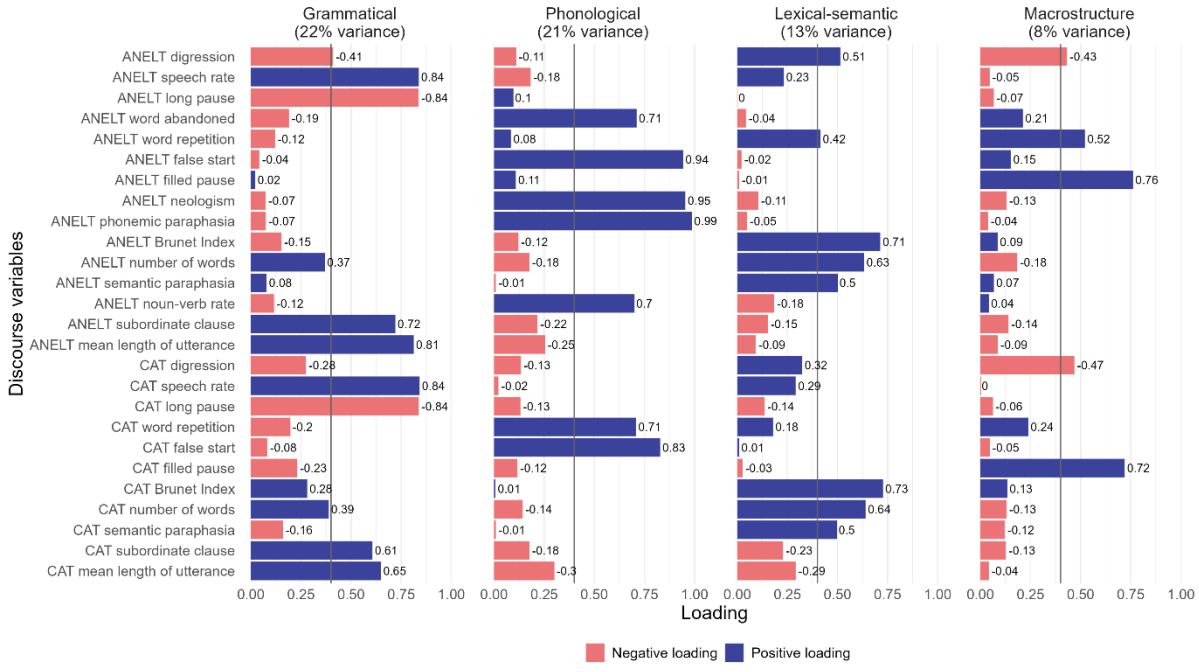


Figure 2: Factor loadings for each of the 26 variables, grouped per factor. Positive loadings are indicated in blue, negative loadings are indicated in red.

3.2. Neural correlates of descriptive and responsive speech

For each of the four factors, univariate VLSM (one-tailed, uncorrected p-value = 0.05) was conducted to analyze the relationship between the respective factor and lesioned brain regions. After controlling for age, time post stroke onset, and lesion volume, and after correcting for multiple comparisons (by means of cluster-based permutation testing with 5000 permutations, corrected p-value = 0.05), a significant neural correlate was found for one of the four factors. More specifically, one cluster of lesioned voxels (N = 10 328 voxels) could be connected to the lexical-semantic factor (corrected p-value = 0.05, corresponding to a cluster threshold of N^t = 7035 voxels). This cluster is visualized Figure 3.

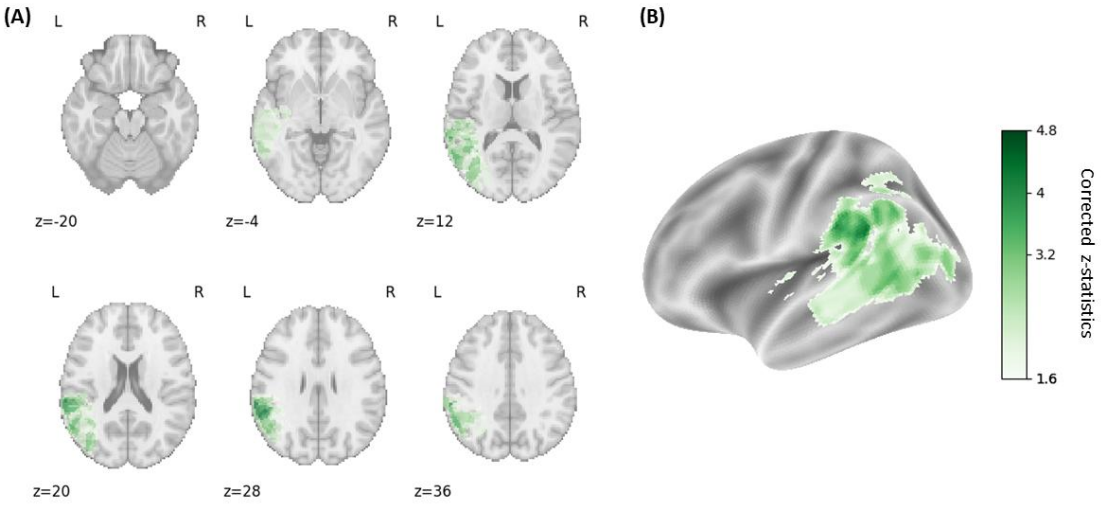


Figure 3: Cluster image for the lexical-semantic cluster. (A) Axial slices (accompanied by their MNI Z-coordinates) and (B) left-hemispheric surface plot of the cluster significantly correlating with the lexical-semantic factor. The color gradient represents the corrected z-statistics of the voxels within the cluster, where higher z-statistics have higher color grading.

The cluster that significantly correlated to the lexical-semantic factor was mainly located in the left inferior parietal, superior and middle temporal as well as in the lateral occipital cortex (Figure 3). Within the cluster, the voxel with the highest z-statistic (z = 4.80) was located in the

posterior division of the supramarginal gyrus (SMG). Most voxels of the cluster resided in the SMG (17.95% of the cluster's voxels) and angular gyrus (9.86%), and in the posterior part of the superior temporal gyrus (STG) (7.13%) and the posterior (9.55%) and temporo-occipital (9.87%) parts of the middle temporal gyrus (MTG), and in the lateral occipital cortex (superior (14.98%) and inferior (7.50%) division). Raw voxel amounts can be found in the Supplementary Material (Table 1). Note that this approach to representing cluster distribution may overlook the involvement of smaller brain areas. Therefore, the Supplementary Material provides an overview of the degree to which each brain area overlaps with the lexical-semantic cluster.

4. Discussion

The current study investigated the underlying behavioral constructs and neural correlates of descriptive and responsive speech in 49 persons with post-stroke aphasia. Factor analysis revealed four latent factors that mainly had high loadings on grammatical, phonological, lexical-semantic and macrostructural variables, with fluency variables spread across the different factors. After multiple comparison correction, only the lexical-semantic factor could be significantly associated with a left cluster covering inferior parietal, superior and middle temporal and lateral occipital regions.

4.1. Underlying constructs of descriptive and responsive speech

To investigate the underlying constructs of descriptive and responsive speech in post-stroke aphasia, sixteen variables spanning different subdomains of language were extracted from both a picture description task (CAT) and a responsive speech task (ANELT). Several variables did not meaningfully contribute to the factor solution and were thus removed from the final analysis. In contrast to 'long pauses', the variable 'short pauses' from both the ANELT and

CAT was omitted, which is understandable given that typical deficits associated with aphasia (e.g., word finding difficulties) result in long rather than short pauses. In addition, several variables from the CAT (phonemic paraphasia, neologism, noun-verb rate and word abandoned) were omitted, presumably because the CAT elicits a relatively shorter language sample compared to the ANELT (mean number of words in ANELT = 424.35 [SD = 305.56], mean number of words in CAT = 57.57 [SD = 51.91]), which might not allow for certain behaviors to become apparent. This raises the important matter of selecting a discourse task that is sensitive to the variables of interest. For example, Boschi et al. (2017) state that picture description tasks tend to limitedly elicit syntactic structures and primarily prompt declarative present tense. The two tasks in the current study – descriptive and responsive speech – may place different cognitive and linguistic demands on the speaker. As mentioned previously, the CAT elicits shorter, more constrained language samples, resulting in the omission of certain variables from the factor analysis. This may have subtly influenced the resulting factor structure. On the other hand, the more open-ended and question-response nature of the ANELT might be more sensitive to discourse deficits in the broad micro- and macrostructural sense. In addition, one could argue that variables from two different discourse tasks should not be analyzed together, but the fact that variables are pairwise clustered over the factors (for example subordinate clause and MLU of both the CAT and ANELT load on grammatical complexity; Brunet Index, number of words and semantic paraphasia of both the CAT and ANELT load on lexical-semantic), justifies our approach combining all variables in one factor analysis, rather than conducting a factor analysis per discourse task. This finding suggests that the descriptive and responsive speech task yield a similar underlying behavioral structure in the current study.

Factor analysis with the remaining 26 variables revealed four underlying behavioral constructs: grammatical, phonological, lexical-semantic and macrostructure. The four factors found in our

study are in line with the theoretical categorization of the variables according to the subdomains in Table 2, with the difference being that fluency variables are spread across all four factors, rather than clustering in a factor of its own. This suggests that grammatical, phonological, lexical-semantic and macrostructural deficits have an impact on fluency in differential ways. This was also found for the APROCSA dimensions (Casilio et al., 2019), where for example speech rate, false starts, abandoned utterances and pauses were distributed across the four dimensions. In contrast to prior work (e.g., Alyahya et al., 2020; Casilio et al., 2019; Vermeulen et al., 1989), we did not find evidence for a separate motor speech component. Motor speech components in previous work often relied on speech rate and pauses, which loaded on the grammatical factor in the current study. One specific fluency variable, speech rate, has also been found to be a factor or component on its own (Alyahya et al., 2020; Ding et al., 2020; Gordon, 2020). We should however consider that speech rate is perhaps a multi-faceted measure rather than merely a proxy of fluency in the strict sense, given that speech rate can be affected by different aspects, such as lexical retrieval, syntactic formulation or motor speech (Gordon & Clough, 2020).

Mean length of utterance and the proportion of subordinate clauses, both measures of grammatical complexity, loaded highly on the grammatical factor. In contrast to our expectations, noun-verb rate did not load highly on the grammatical factor, but instead it loaded highly on the second factor, phonological. One explanation for this could be that selection of nouns and verbs is influenced not only by grammatical processing, but also by phonological demands. Moreover, verbs often lack inflection when produced in spontaneous speech by individuals with aphasia (Bastiaanse et al., 2002), resulting in comparing phone-rich with phone-poor words, which is more phonological in nature. Next, the fluency variables speech rate and long pauses loaded highly on the grammatical factor. It is evident that speech rate and

long pauses have opposite loadings, as long pauses decrease the speech rate, which is frequently observed in persons with grammatical deficits. Finally, digression had a weak, negative loading on the grammatical factor (-0.41), which means grammatical skills are associated with less digression. A key finding in our and earlier (Ding et al., 2020; Gordon, 2020; Vermeulen et al., 1989) work, is that the factor or component reflecting grammatical complexity has the highest proportion of explained variance. This is somewhat evident given that we are investigating constructs underlying discourse, which refers to language that extends beyond a single clause.

The second factor exhibits very high loadings on phonemic paraphasia, neologism and false start. Prior work (Gordon, 2020; Vermeulen et al., 1989; Zevgolatakou et al., 2023) also found a separate factor mainly reflective of phonological difficulties. In addition, this factor shows that phonological difficulties are associated with abandoning (especially in the ANELT) and repeating (especially in the CAT) words. The latter fluency features could theoretically also be associated with for example lexical-semantic difficulties, but were in this sample associated with phonological difficulties.

Lexical-semantic difficulties become apparent through a higher proportion of semantic paraphasias, higher Brunet Index (indicative of less diverse word use) and a higher number of words. The latter finding may seem atypical, but can be explained as follows: in case of lexical-semantic deficits, individuals with aphasia may use more words to self-cue, i.e., circumlocution (Tilton-Bolowsky et al., 2023), and retrieve the correct word. This elaboration does not only explain the increased number of words, but it also explains the positive loading of digression and word repetition on the lexical-semantic factor. In previous studies with a similar approach to studying discourse production, the lexical and semantic variables were often clustered in separate constructs (Alyahya et al., 2020; Ding et al., 2020; Zevgolatakou et al., 2023).

However, the study of Vermeulen et al. (1989) found a similar factor 'vocabulary', which reflected both lexical diversity and naming.

The fourth factor is one that we believe is discourse-related, and would not arise when merely including isolated language tasks. This factor seems to capture higher-order processes that are beyond the microstructural level, and was therefore named 'macrostructure'. Difficulties with for example conceptualizing an idea at the macrolinguistic level, can be positively associated with filled pauses (from both discourse tasks) and word repetition (from the CAT). The negative loadings on digression can be explained in a similar way to the grammatical factor: difficulties in conceptualizing and organizing ideas tend to hinder, rather than facilitate, digression. We also note that this is the only factor with loadings > 0.4 for the CAT digression variable, which strengthens our interpretation of this factor as macrostructural in nature. On the other hand, we should be careful with interpreting this factor as it accounted for the least variance and included only one macrostructural variable (digression). The other variables that loaded highly on the fourth factor (e.g., filled pauses, repetitions) could also be associated with lower-level lexical access difficulty or motor planning issues. Despite these caveats, this is – to the best of our knowledge – the first study where one of the factors underlying discourse reflects a macrostructural component.

It is challenging to compare our four-factor solution to prior work given the heterogeneity of the discourse variables included across studies. Establishing a core outcome measure set specific to discourse could help address this issue. A second aspect hampering comparison to prior work, is the differences between analytical approaches. Both factor analysis (FA) and principal component analysis (PCA) have been frequently used to address similar research questions (e.g., Alyahya et al., 2020; Casilio et al., 2019; Ding et al., 2020; Gleichgerrcht et al.,

2021; Gordon, 2020; Vermeulen et al., 1989; Zevgolatakou et al., 2023). However, FA and PCA differ in goals and estimation methods, which impacts replication of previous findings (Wilson & Hula, 2019). The current study used FA because the goal was to identify the latent constructs underlying a set of discourse variables while retaining the explainability of the factor solution. In contrast, the primary goal of PCA is data reduction.

4.2. Neural correlates of descriptive and responsive speech

The current study informs us about how acute stroke-induced brain lesions influence long-term discourse production. From the four latent constructs, only the lexical-semantic construct could be significantly associated with lesions, showing a left cluster covering the inferior parietal (AG, SMG, parietal operculum), the posterior temporal (STG, MTG, HG, PT) and the lateral occipital cortex. The location of a temporoparietal cluster for lexical-semantic processing aligns with the findings of Ding et al. (2020) who concluded that damage to temporo-parietal regions affected word retrieval and the ability to produce these words within increasingly complex combinations. Similarly, Casilio and colleagues (2024) found that paraphasia (misselection of words and sounds) and logopenia (paucity of words) were associated with damage to primarily left posterior temporal and parietal regions, and also Borovsky and colleagues (2007) observed a cluster in the posterior temporal regions for the Type Token Ratio, a measure of lexical diversity. However, other lesion symptom mapping studies on discourse have indicated more anteriorly located regions involved in lexical-semantic processing. More specifically, Alyahya et al. (2020) provided evidence that 'verbal quantity' (mainly reflecting content word count and type count) is associated with the pre/post-central gyri and the inferior frontal gyrus, and Wilmskoetter et al. (2019) concluded that low lexical diversity and lexical sophistication (i.e., the use of sophisticated or 'advanced' words) are respectively linked to lesions in the left insula and the left anterior temporal gyrus. Finally, Stark et al. (2019) found both anterior (temporal)

as well as posterior (temporoparietal) involvement for semantically related and unrelated paraphasias respectively. Although it has long been assumed, based on the first case studies of Broca and Wernicke, that frontal regions are involved in production whereas posterior regions are involved in comprehension (Tremblay & Dick, 2016), more recent fMRI studies have shown overlap in frontal and temporal areas regarding language production and comprehension (Menenti et al., 2011; Silbert et al., 2014). This is in line with our and other lesion symptom mapping studies using productive discourse tasks (Borovsky et al., 2007; Casilio et al., 2025; Ding et al., 2020), suggesting that posterior temporoparietal regions are not exclusive for comprehension, but are also vital for producing lexical-semantic content in a discourse task. Another functional specialization that is often assumed within the language network is between semantic (recruiting ventral temporal regions) versus phonological (recruiting dorsal temporoparietal-to-frontal regions) (Hickok & Poeppel, 2007). Our temporoparietal cluster for lexical-semantics contained the expected ventral regions (MTG) but also contained dorsal regions (PT, Heschl, SMG, STGp), hence did not fully align with the specialized regions indicated by the dual route model. Regardless of its exact language function, the temporoparietal cluster seems critical for language processing, given that lesions in this region lead to long-lasting and more severe language problems (Wilson et al., 2023).

It is noteworthy that significant neural correlates were only found for the lexical-semantic construct. This may be due to methodological issues related to discourse measures. For example, lexical-semantic measures have been used and reported more frequently than macrostructural measures in studies. This may result in more robust and sensitive measures of lexical-semantic abilities compared to macrostructural abilities. On the other hand, our factor analysis, which contained 32 variables and 49 subjects, may have been methodologically suboptimal for elucidating precise lesion-behavior relationships. Although our sample

contained sufficient behavioral variance, the lack of variance in lesion locations may have resulted in the absence of significant neural correlates for the other three factors. Some factors, such as macrostructure, are hypothesized to involve right-hemispheric areas as well, which were not represented in the current study.

4.3. Limitations and future research

Although post-stroke language deficits are often associated with left-hemispheric lesions (Carey & Johnstone, 2014), language is organized in both hemispheres (Bozic et al., 2010; Bradshaw et al., 2017; Hartwigsen et al., 2010; Hope et al., 2017; Poeppel, 2014; Price, 2012; Schevenels et al., 2020; Vigneau et al., 2011), which suggests the right-hemispheric areas involved could be supportive for optimal language performance. Especially for macrostructural components, such as pragmatics, the right hemisphere could play a role (Fedorenko et al., 2024). However, we had insufficient lesion overlap in the right hemisphere to reveal a potential influence of right-hemispheric lesions on discourse macrostructure. Second, we did not investigate disruption of functional connectivity, although we recognize the importance of white matter tracts such as the anterior and long segments of the arcuate fasciculus, the frontal aslant tract, the left uncinate fasciculus and the inferior fronto-occipital fasciculus for spoken discourse production, as summarized by Alyahya (2023). It must be noted that 1.64% of the voxels in the cluster relating to the lexical-semantic factor, could not be attributed to any of the brain regions based on the Harvard-Oxford Atlas. These voxels were possibly part of deeper subcortical structures or white matter tracts. Future research could implement disconnectomics (Thiebaut de Schotten et al., 2020) to obtain a more comprehensive view. Disconnectomics are maps that allow to visualize which (proportions of) white matter connective pathways have been disrupted due to the lesion, as such allowing to unravel the neural connective correlates of discourse production.

Another important limitation of the current study is that we did not formally assess the reliability of the transcription and manually extracted discourse variables, as is recommended by Stark et al. (2023). However, all transcriptions were based on the most standardized orthographic annotation protocol available for Dutch, and the accuracy of the transcription and scoring was checked for the initial samples of each annotator. Although we did not calculate interrater reliability for our manual measures, previous research in Dutch individuals with aphasia has shown that measures such as mean length of utterance and noun-verb rate typically demonstrate high interrater reliability (van der Scheer et al., 2011). We acknowledge that certain variables, such as digression, are likely more sensitive to differences in scoring and that reliability may be lower for these measures. In addition, we did not include cognitive variables, which would have been relevant to further specify the underlying behavioral constructs of discourse, especially at the macrostructural level. Furthermore, we want to raise the concern that digression based on the ANELT was a suboptimal measure, as this variable loaded on all but one factor. Digression refers to utterances that do not contain key information, but are straying from the main topic instead. This might have been more straightforward to determine for a constrained task like picture description from the CAT, compared to responsive speech from the ANELT where participants have more speech degrees of freedom and it can thus be debated when an utterance should be labeled as 'digression'. Next, the inclusion of more macrolinguistic variables would have been relevant to further establish the preliminary, macrostructural factor. Only few studies have addressed macrolinguistic features (Alyahya et al., 2022; Barker et al., 2017). Future studies investigating discourse should therefore systematically include macrolinguistic variables as well. We recommend to include measures of coherence, cohesion, informativeness and story grammar to capture a broader set of macrostructural abilities. Despite the limited inclusion of macrostructural variables in the current study, it is the first one to find a macrostructural factor when investigating the behavioral

constructs underlying discourse production. Another limitation is that the four factor solution was not further validated using confirmatory factor analysis in an independent sample of persons with post-stroke aphasia. Finally, it is recognized that discourse assessment enables a comprehensive characterization of language in context (Bryant et al., 2016; Stark et al., 2021). However, its implementation in both research and clinical practice is currently limited, mainly due to the high workload of discourse analysis (Bryant et al., 2019; Stark et al., 2021). To overcome this barrier, we used automatic speech recognition and semi-automatic speech assessment. Automated discourse analysis has the potential to enable time-efficient and thus more feasible discourse analysis, and allows to include a variety of discourse variables spanning all subdomains of language.

5. Conclusion

The current study suggests that the underlying constructs of discourse production reflect both micro- and macrolinguistic components. In addition, the neural correlates of lexical-semantics in discourse were found to be consistent with those found in more controlled, isolated language tasks. This finding further establishes the involvement of the left inferior parietal, superior and middle temporal and lateral occipital lobe in lexical-semantic abilities. Future research should focus on establishing macrostructural components of discourse and exploring potential neural correlates outside of the traditional, left-hemispheric perisylvian language network. With further studies and consistent findings, this line of research has the potential to provide prognostic insights, such as identifying which acutely damaged brain areas are predictive of long-term discourse deficits.

Data availability statement

The behavioral dataset and the factor analysis script used in this study are available on Open Science Framework. For the voxel-wise lesion symptom mapping, we used the NiiStat software package implemented in MATLAB. The Python script is available on GitHub. The MRI data cannot be shared, as lesioned MRI data cannot be considered anonymous.

Author contributions

Mara Barberis conceptualization, data curation, methodology, formal analysis, resources, writing - original draft, visualization, funding acquisition; Ella Eycken methodology, formal analysis, visualization, writing – original draft; Pieter De Clercq methodology, software; Robin Lemmens methodology, writing - review & editing; Hugo Van hamme methodology, software, writing - review & editing, supervision; Maaike Vandermosten conceptualization, resources, writing - review & editing, supervision, project administration, funding acquisition.

Funding sources

This work was supported by the Research Foundation Flanders (Fonds Wetenschappelijk Onderzoek), Award ID 1SH1Q24N to Mara Barberis and Award ID G0D8520N to Maaike Vandermosten.

Acknowledgements

We would like to thank all individuals with aphasia who participated in our study. In addition, we are thankful to the nursing and (para)medical staff at the stroke unit to make this collaboration possible. We are also truly grateful for the help during behavioral data collection from Jana Sevenants, Klara Spooren, Fien Poesen, Hanneke Lantin, Ilka Bocken, Sara Dupont, Eline Biermans, Valerie Deckers and Nele Faes. In addition, a heartfelt thank you to Sara

663 Dupont, Eline Biermans and Andi Smet for their help with lesion delineation, to Lotte Van den
664 Eynde for her help with data visualization and to Bastiaan Tamm for the automated feature
665 extraction.

References

Alyahya, R. S. W. (2023). The Structural Neural Correlates of Spoken Discourse. In A. P.-H. Kong (Ed.), *Spoken Discourse Impairments in the Neurogenic Populations* (pp. 111–119). Springer International Publishing. https://doi.org/10.1007/978-3-031-45190-4_8

Alyahya, R. S. W., Halai, A. D., Conroy, P., & Lambon, M. A. (2020). A unified model of post-stroke language deficits including discourse production and their neural correlates. *Brain*, *143*(5), 1541–1554. https://doi.org/10.1093/brain/awaa074

Alyahya, R. S. W., Halai, A. D., Conroy, P., & Ralph, M. A. L. (2021). Content word production during discourse in aphasia: Deficits in word quantity, not lexical–semantic complexity. *Journal of Cognitive Neuroscience*, *33*(12), 2494–2511. https://doi.org/10.1162/jocn_a_01772

Alyahya, R. S. W., Lambon Ralph, M. A., Halai, A., & Hoffman, P. (2022). The cognitive and neural underpinnings of discourse coherence in post-stroke aphasia. *Brain Communications*, *4*(3). https://doi.org/10.1093/braincomms/fcac147

Angelopoulou, G., Kasselimis, D., Varkanitsa, M., Tsolakopoulos, D., Papageorgiou, G., Velonakis, G., Meier, E., Karavassilis, E., Pantoleon, V., Laskaris, N., Kelekis, N., Tountopoulou, A., Vassilopoulou, S., Goutsos, D., Kiran, S., Weiller, C., Rijntjes, M., & Potagas, C. (2024). Investigating silent pauses in connected speech: integrating linguistic, neuropsychological, and neuroanatomical perspectives across narrative tasks in post-stroke aphasia. *Frontiers in Neurology* , *15*. https://doi.org/10.3389/fneur.2024.1347514

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

Azios, J. H., Archer, B., & Lee, J. B. (2021). Detecting behavioural change in conversation: procedures and preliminary data. *Aphasiology*, *35*(7), 961–983. https://doi.org/10.1080/02687038.2020.1812031

Baldo, J. V., Ivanova, M. V., Herron, T. J., Wilson, S. M., & Dronkers, N. F. (2022). *Voxel-Based Lesion Symptom Mapping* (pp. 95–118). https://doi.org/10.1007/978-1-0716-2225-4_5

Barberis, M., De Clercq, P., Tamm, B., Van hamme, H., & Vandermosten, M. (2024). Automatic recognition and detection of aphasic natural speech. *Interspeech 2024*, 1990–1994. https://doi.org/10.21437/Interspeech.2024-517

Barberis, M., Van hamme, H., & Vandermosten, M. (2023). *Protocol transcriptie natuurlijke spraak bij afasie*. https://osf.io/pmgwc/?view_only=77dddda4789d4343ab0b1e8b4cf4e0a7

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

Bastiaanse, R., Hugen, J., Kos, M., & van Zonneveld, R. (2002). Lexical, morphological, and syntactic aspects of verb production in agrammatic aphasics. *Brain and Language*, *80*(2), 142–159. https://doi.org/10.1006/brln.2001.2585

Bates, E., Wilson, S. M., Saygin, A. P., Dick, F., Sereno, M. I., Knight, R. T., & Dronkers, N. F. (2003). Voxel-based lesion–symptom mapping. *Nature Neuroscience*, *6*(5), 448–450. https://doi.org/10.1038/nn1050

Blomert, L., Koster, C. H., & Kean, M.-L. (1995). *Amsterdam-Nijmegen Test voor Alledaagse Taalvaardigheid (ANTAT)*. Hogrefe.

Borovsky, A., Saygin, A. P., Bates, E., & Dronkers, N. (2007). Lesion correlates of conversational speech production deficits. *Neuropsychologia*, *45*(11), 2525–2533. https://doi.org/10.1016/j.neuropsychologia.2007.03.023

Boschi, V., Catricalà, E., Consonni, M., Chesi, C., Moro, A., & Cappa, S. F. (2017). Connected Speech in Neurodegenerative Language Disorders: A Review. *Frontiers in Psychology*, *8*, 269. https://doi.org/10.3389/fpsyg.2017.00269

Boxum, E., van der Scheer, F., & Zwaga, M. (2013). *Analyse voor Spontane Taal bij Afasie*. Vereniging voor Klinische Linguïstiek.

Bozic, M., Tyler, L. K., Ives, D. T., Randall, B., & Marslen-Wilson, W. D. (2010). Bihemispheric foundations for human speech comprehension. *Proceedings of the National Academy of Sciences*, *107*(40), 17439–17444. https://doi.org/10.1073/pnas.1000531107

Bradshaw, A. R., Thompson, P. A., Wilson, A. C., Bishop, D. V. M., & Woodhead, Z. V. J. (2017). Measuring language lateralisation with different language tasks: a systematic review. *PeerJ*, *5*, e3929. https://doi.org/10.7717/peerj.3929

Bryant, L., Ferguson, A., & Spencer, E. (2016). Linguistic analysis of discourse in aphasia: A review of the literature. *Clinical Linguistics and Phonetics*, *30*(7), 489–518. https://doi.org/10.3109/02699206.2016.1145740

Bryant, L., Ferguson, A., Valentine, M., & Spencer, E. (2019). Implementation of discourse analysis in aphasia: investigating the feasibility of a Knowledge-to-Action intervention. *Aphasiology*, *33*(1), 31–57. https://doi.org/10.1080/02687038.2018.1454886

Carey, D. P., & Johnstone, L. T. (2014). Quantifying cerebral asymmetries for language in dextrals and adextrals with random-effects meta analysis. *Frontiers in Psychology*, *5*, 1128. https://doi.org/10.3389/fpsyg.2014.01128

Casilio, M., Kasdan, A. V, Bryan, K., Shibata, K., Schneck, S. M., Levy, D. F., Entrup, J. L., Onuscheck, C., de Riesthal, M., & Wilson, S. M. (2025). Four dimensions of naturalistic language production in aphasia after stroke. *Brain*, *148*(1), 291–312. https://doi.org/10.1093/brain/awae195

738 Casilio, M., Rising, K., Beeson, P. M., Bunton, K., & Wilson, S. M. (2019). Auditory-
739 perceptual rating of connected speech in Aphasia. *American Journal of Speech-Language*
740 *Pathology*, *28*(2), 550–568. https://doi.org/10.1044/2018_AJSLP-18-0192
741 Dietz, A., & Boyle, M. (2018). Discourse measurement in aphasia research: have we reached
742 the tipping point? In *Aphasiology* (Vol. 32, Issue 4, pp. 459–464). Routledge.
743 https://doi.org/10.1080/02687038.2017.1398803
744 Ding, J., Martin, R. C., Cris Hamilton, A., & Schnur, T. T. (2020). Dissociation between frontal
745 and temporal-parietal contributions to connected speech in acute stroke. *Brain*, *143*(3),
746 862–876. https://doi.org/10.1093/brain/awaa027
747 Ding, J., Middleton, E. L., & Mirman, D. (2023). Impaired discourse content in aphasia is
748 associated with frontal white matter damage. *Brain Communications*, *5*(6).
749 https://doi.org/10.1093/braincomms/fcad310
750 Doedens, W. J., & Meteyard, L. (2020). Measures of functional, real-world communication for
751 aphasia: a critical review. In *Aphasiology* (Vol. 34, Issue 4, pp. 492–514). Routledge.
752 https://doi.org/10.1080/02687038.2019.1702848
753 Fedorenko, E., Ivanova, A. A., & Regev, T. I. (2024). The language network as a natural kind
754 within the broader landscape of the human brain. *Nature Reviews Neuroscience*, *25*(5),
755 289–312. https://doi.org/10.1038/s41583-024-00802-4
756 Fergadiotis, G., & Wright, H. H. (2011). Lexical diversity for adults with and without aphasia
757 across discourse elicitation tasks. *Aphasiology*, *25*(11), 1414–1430.
758 https://doi.org/10.1080/02687038.2011.603898
759 Fromm, D., Greenhouse, J., Pudil, M., Shi, Y., & MacWhinney, B. (2022). Enhancing the
760 classification of aphasia: a statistical analysis using connected speech. *Aphasiology*,
761 *36*(12), 1492–1519. https://doi.org/10.1080/02687038.2021.1975636

Garrett, M. F. (1984). *The organization of processing structure for language production: Applications to aphasic speech* (D. Caplan, A. Lecours, & A. Smith, Eds.). MIT Press. https://api.semanticscholar.org/CorpusID:56603168

Gleichgerrcht, E., Roth, R., Fridriksson, J., den Ouden, D., Delgaizo, J., Stark, B., Hickok, G., Rorden, C., Wilmskoetter, J., Hillis, A., & Bonilha, L. (2021). Neural bases of elements of syntax during speech production in patients with aphasia. *Brain and Language*, *222*. https://doi.org/10.1016/j.bandl.2021.105025

Gordon, J. K. (2020). Factor Analysis of Spontaneous Speech in Aphasia. *Journal of Speech, Language, and Hearing Research*, *63*, 4127–4147. https://doi.org/10.23641/asha

Gordon, J. K., & Clough, S. (2020). How fluent? Part B. Underlying contributors to continuous measures of fluency in aphasia. *Aphasiology*, *34*(5), 643–663. https://doi.org/10.1080/02687038.2020.1712586

Hamilton, L. S., & Huth, A. G. (2020). The revolution will not be controlled: natural stimuli in speech neuroscience. *Language, Cognition and Neuroscience*, *35*(5), 573–582. https://doi.org/10.1080/23273798.2018.1499946

Hartwigsen, G., Price, C. J., Baumgaertner, A., Geiss, G., Koehnke, M., Ulmer, S., & Siebner, H. R. (2010). The right posterior inferior frontal gyrus contributes to phonological word decisions in the healthy brain: Evidence from dual-site TMS. *Neuropsychologia*, *48*(10), 3155–3163. https://doi.org/10.1016/j.neuropsychologia.2010.06.032

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393–402. https://doi.org/10.1038/nrn2113

Hope, T. M. H., Leff, A. P., Prejawa, S., Bruce, R., Haigh, Z., Lim, L., Ramsden, S., Oberhuber, M., Ludersdorfer, P., Crinion, J., Seghier, M. L., & Price, C. J. (2017). Right hemisphere structural adaptation and changing language skills years after left hemisphere stroke. *Brain*, *140*(6), 1718–1728. https://doi.org/10.1093/brain/awx086

Kennedy, D., Lange, N., Makris, N., Bates, J., Meyer, J., & Caviness Jr, V. (1998). Gyri of the human neocortex: an MRI-based analysis of volume and variance. *Cerebral Cortex*, *8*(4), 372–384. https://doi.org/10.1093/cercor/8.4.372

Kong, A. P.-H. (2023). *Spoken Discourse Impairments in the Neurogenic Populations* (A. P.-H. Kong, Ed.). Springer International Publishing. https://doi.org/10.1007/978-3-031-45190-4

Le, D., Licata, K., & Mower Provost, E. (2018). Automatic quantitative analysis of spontaneous aphasic speech. *Speech Communication*, *100*, 1–12. https://doi.org/10.1016/j.specom.2018.04.001

Lomas, J., Pickard, L., Bester, S., Elbard, H., Finlayson, A., & Zoghaib, C. (1989). The communicative effectiveness index: Developmental and psychometric evaluation of a functional communication measure for adult aphasia. *Journal of Speech and Hearing Disorders*, *54*(1), 113–124. https://doi.org/10.1044/jshd.5401.113

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi. *Interspeech 2017*, 498–502. https://doi.org/10.21437/Interspeech.2017-1386

Menenti, L., Gierhan, S. M. E., Segaert, K., & Hagoort, P. (2011). Shared language: Overlap and segregation of the neuronal infrastructure for speaking and listening revealed by functional MRI. *Psychological Science*, *22*(9), 1173–1182. https://doi.org/10.1177/0956797611418347

Oostdijk, N. (2000). The Spoken Dutch Corpus: Overview and first evaluation. *Proceedings of LREC-2000, Athens*, *2*.

Pakhomov, S. V. S., Smith, G. E., Chacon, D., Feliciano, Y., Graff-Radford, N., Caselli, R., & Knopman, D. S. (2010). Computerized analysis of speech and language to identify

psycholinguistic correlates of frontotemporal lobar degeneration. *Cognitive and Behavioral Neurology*, *23*(3), 165–177. https://doi.org/10.1097/WNN.0b013e3181c5dde3

Parsapoor, M., Alam, M. R., & Mihailidis, A. (2021). *Towards AI-powered Language Assessment Tools*. https://doi.org/10.48550/arXiv.2209.12652

Poeppel, D. (2014). The neuroanatomic and neurophysiological infrastructure for speech and language. *Current Opinion in Neurobiology*, *28*, 142–149. https://doi.org/10.1016/j.conb.2014.07.005

Poncelet, J., & Van hamme, H. (2022). Learning to jointly transcribe and subtitle for end-to-end spontaneous speech recognition. *Proceedings SLT 2022*. https://doi.org/10.48550/arXiv.2210.07771

Price, C. J. (2012). A review and synthesis of the first 20years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage*, *62*(2), 816–847. https://doi.org/10.1016/j.neuroimage.2012.04.062

Rorden, C., Bonilha, L., Fridriksson, J., Bender, B., & Karnath, H.-O. (2012). Age-specific CT and MRI templates for spatial normalization. *NeuroImage*, *61*(4), 957–965. https://doi.org/10.1016/j.neuroimage.2012.03.020

Ruiter, M. B., Lotgering, E., & Rietveld, A. C. (2016). Adapted ANELT scenarios for ANELT-CU scoring method. In *[Unpublished test stimuli]*.

Saffran, E. M., Berndt, R. S., & Schwartz, M. F. (1989). The quantitative analysis of agrammatic production: Procedure and data. *Brain and Language*, *37*(3), 440–479. https://doi.org/10.1016/0093-934X(89)90030-8

Sanborn, V., Ostrand, R., Ciesla, J., & Gunstad, J. (2020). Automated assessment of speech production and prediction of MCI in older adults. *Applied Neuropsychology:Adult*, *29*(5), 1250–1257. https://doi.org/10.1080/23279095.2020.1864733

Schevenels, K., Price, C. J., Zink, I., De Smedt, B., & Vandermosten, M. (2020). A Review on Treatment-Related Brain Changes in Aphasia. *Neurobiology of Language*, *1*(4), 402–433. https://doi.org/10.1162/nol_a_00019

Schnur, T., Brown, D., & Guess, K. (2023). The Functional Neural Correlates of Spoken Discourse. In A. P.-H. Kong (Ed.), *Spoken Discourse Impairments in the Neurogenic Populations* (pp. 121–132). Springer.

Schnur, T. T., & Wang, S. (2024). Differences in Connected Speech Outcomes Across Elicitation Methods. *Aphasiology*, *38*(5), 816–837. https://doi.org/10.1080/02687038.2023.2239509

Silbert, L. J., Honey, C. J., Simony, E., Poeppel, D., & Hasson, U. (2014). Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proceedings of the National Academy of Sciences*, *111*(43). https://doi.org/10.1073/pnas.1323812111

Stark, B. C. (2019). A comparison of three discourse elicitation methods in aphasia and age-matched adults: Implications for language assessment and outcome. *American Journal of Speech-Language Pathology*, *28*(3), 1067–1083. https://doi.org/10.1044/2019_AJSLP-18-0265

Stark, B. C. (2023). Historical Review of Research in Discourse Deficits and Its Recent Advancement. In *Spoken Discourse Impairments in the Neurogenic Populations* (pp. 3–22). Springer International Publishing. https://doi.org/10.1007/978-3-031-45190-4_1

Stark, B. C., Basilakos, A., Hickok, G., Rorden, C., Bonilha, L., & Fridriksson, J. (2019). Neural organization of speech production: A lesion-based study of error patterns in connected speech. *Cortex*, *117*, 228–246. https://doi.org/10.1016/j.cortex.2019.02.029

Stark, B. C., Bryant, L., Themistocleous, C., den Ouden, D.-B., & Roberts, A. C. (2023). Best practice guidelines for reporting spoken discourse in aphasia and neurogenic

communication disorders. *Aphasiology*, *37*(5), 761–784. https://doi.org/10.1080/02687038.2022.2039372

Stark, B. C., Dutta, M., Murray, L. L., Fromm, D., Bryant, L., Harmon, T. G., Ramage, A. E., & Roberts, A. C. (2021). Spoken discourse assessment and analysis in aphasia: an international survey of current practices. *Journal of Speech, Language and Hearing Research (JSLHR)*, *64*(11), 4366–4389. https://doi.org/10.1044/2021_JSLHR-20-00708

Taalunie. (2000). *Protocol voor Orthografische Transcriptie*.

Tan, Z.-H., Sarkar, A. kr., & Dehak, N. (2020). rVAD: An unsupervised segment-based robust voice activity detection method. *Computer Speech & Language*, *59*, 1–21. https://doi.org/10.1016/j.csl.2019.06.005

Thiebaut de Schotten, M., Foulon, C., & Nachev, P. (2020). Brain disconnections link structural connectivity with function and behaviour. *Nature Communications*, *11*(1), 5094. https://doi.org/10.1038/s41467-020-18920-9

Tilton-Bolowsky, V. E., van der Stelt, C. M., DeMarco, A., Laks, A., Dobbins, C., Snider, S. F., Turkeltaub, P. E., & Vallila-Rohter, S. (2023). Evaluating circumlocution in naming as a predictor of communicative informativeness and efficiency in discourse. *Aphasiology*, *38*(2), 237–260. https://doi.org/10.1080/02687038.2023.2183076

Tremblay, P., & Dick, A. S. (2016). Broca and Wernicke are dead, or moving past the classic model of language neurobiology. *Brain and Language*, *162*, 60–71. https://doi.org/10.1016/j.bandl.2016.08.004

Van Der Scheer, F., Zwaga, M., & Jonkers, R. (2011). Normering van de ASTA, analyse voor spontane taal bij afasie. *Stem-, Spraak- En Taalpathologie*, *17*(2), 19–30.

Van Ewijk, L., Dijkhuis, L., Hofs-van Kats, M., Hendrickx, M., Wijngaarden, M., & De Hilster, C. (2018). *Nederlandse benoemtest (NBT)*. Bohn Stafleu van Loghum.

884 Vandenborre, D., Visch-Brink, E., van Dun, K., Verhoeven, J., & Mariën, P. (2018). Oral and
885 written picture description in individuals with aphasia. *International Journal of Language*
886 *and Communication Disorders*, *53*(2), 294–307. https://doi.org/10.1111/1460-6984.12348
887 Vermeulen, J., Bastiaanse, R., & Van Wageningen, B. (1989). Spontaneous Speech in Aphasia:
888 A Correlational Study. *Brain and Language*, *36*(2), 252–274.
889 Vigneau, M., Beaucousin, V., Hervé, P.-Y., Jobard, G., Petit, L., Crivello, F., Mellet, E., Zago,
890 L., Mazoyer, B., & Tzourio-Mazoyer, N. (2011). What is right-hemisphere contribution to
891 phonological, lexico-semantic, and sentence processing? *NeuroImage*, *54*(1), 577–593.
892 https://doi.org/10.1016/j.neuroimage.2010.07.036
893 Visch-Brink, E. G., van de Sandt-Koenderman, M., & El Hachioui, H. (2010). *ScreeLing*. Bohn
894 Stafleu Van Loghum.
895 Visch-Brink, E., Vandenborre, D., de Smet, H. J., Mariën, P., Swinburn, K., Porter, G., &
896 Howard, D. (2015). *Comprehensive Aphasia Test - Nederlandstalige bewerking (CAT-*
897 *NL)*. Pearson.
898 Wallace, S. J., Worrall, L., Rose, T., & Le Dorze, G. (2017). Which treatment outcomes are
899 most important to aphasia clinicians and managers? An international e-Delphi consensus
900 study. *Aphasiology*, *31*(6), 643–673. https://doi.org/10.1080/02687038.2016.1186265
901 Wilmskoetter, J., Fridriksson, J., Gleichgerrcht, E., Stark, B. C., Delgaizo, J., Hickok, G.,
902 Vaden, K. I., Hillis, A. E., Rorden, C., & Bonilha, L. (2019). Neuroanatomical structures
903 supporting lexical diversity, sophistication, and phonological word features during
904 discourse. *NeuroImage: Clinical*, *24*, 101961. https://doi.org/10.1016/j.nicl.2019.101961
905 Wilson, S. M., Entrup, J. L., Schneck, S. M., Onuscheck, C. F., Levy, D. F., Rahman, M.,
906 Willey, E., Casilio, M., Yen, M., Brito, A. C., Kam, W., Davis, L. T., de Riesthal, M., &
907 Kirshner, H. S. (2023). Recovery from aphasia in the first year after stroke. *Brain*, *146*(3),
908 1021–1039. https://doi.org/10.1093/brain/awac129

Wilson, S. M., Henry, M. L., Besbris, M., Ogar, J. M., Dronkers, N. F., Jarrold, W., Miller, B. L., & Gorno-Tempini, M. L. (2010). Connected speech production in three variants of primary progressive aphasia. *Brain*, *133*(7), 2069–2088. https://doi.org/10.1093/brain/awq129

Wilson, S. M., & Hula, W. D. (2019). Multivariate Approaches to Understanding Aphasia and its Neural Substrates. *Current Neurology and Neuroscience Reports*, *19*(8), 53. https://doi.org/10.1007/s11910-019-0971-6

Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., & Nichols, T. E. (2014). Permutation inference for the general linear model. *NeuroImage*, *92*, 381–397. https://doi.org/10.1016/j.neuroimage.2014.01.060

Worrall, L., Sherratt, S., Rogers, P., Howe, T., Hersh, D., Ferguson, A., & Davidson, B. (2011). What people with aphasia want: Their goals according to the ICF. *Aphasiology*, *25*(3), 309–322. https://doi.org/10.1080/02687038.2010.508530

Zevgolatakou, E., Thye, M., & Mirman, D. (2023). Behavioural and neural structure of fluent speech production deficits in aphasia. *Brain Communications*, *5*(1). https://doi.org/10.1093/braincomms/fcac327