A web-based application for eliciting narrative discourse from Greek-speaking people with and without language impairments

Spyridoula Stamouli[1]*, Michaela Nerantzini[2],
Ioannis Papakyritsis[2], Athanasios Katsamanis[1],
Gerasimos Chatzoudis[1], Athanasia-Lida Dimou[1], Manos Plitsis[1,3],
Vassilis Katsouros[1], Spyridoula Varlokosta[1,4] and Arhonto Terzi[2]

[1]Institute for Language and Speech Processing, Athena Research Center, Athens, Greece, [2]Department of Speech and Language Therapy, University of Patras, Patras, Greece, [3]Department of Informatics and Telecommunications, University of Athens, Athens, Greece, [4]Faculty of Philology, Linguistics Department, National and Kapodistrian University of Athens, Athens, Greece

In this paper we present a web-based data collection method designed to elicit narrative discourse from adults with and without language impairments, both in an in-person set up and remotely. We describe the design, methodological considerations and technical requirements regarding the application development, the elicitation tasks, materials and guidelines, as well as the implementation of the assessment procedure. To investigate the efficacy of remote elicitation of narrative discourse with the use of the technology-enhanced method presented here, a pilot study was conducted, aiming to compare narratives elicited remotely to narratives collected in an in-person elicitation mode from ten unimpaired adults, using a within-participants research design. In the remote elicitation setting, each participant performed the tasks of a narrative elicitation protocol *via* the web application in their own environment, with the assistance of an investigator in the context of a virtual meeting (video conferencing). In the in-person elicitation setting, the participant was in the same environment with the investigator, who administered the tasks using the web application. Data were manually transcribed, and transcripts were processed with Natural Language Processing (NLP) tools. Linguistic features representing key measures of spoken narrative discourse were automatically calculated: linguistic productivity, content richness, fluency, syntactic complexity at clausal and inter-clausal level, lexical diversity, and verbal output. The results show that spoken narratives produced by the same individuals in the two different experimental settings do not present significant differences regarding the linguistic variables analyzed, in sixty six out of seventy statistical tests. These results indicate that the presented web-based application is a feasible method for the remote collection of spoken narrative discourse from adults without language impairments in the context of online assessment.

KEYWORDS

narrative discourse, elicitation, web application, remote assessment, aphasia

1. Introduction

Acquired speech and language disorders are increasingly relevant for a significant percentage of the adult population, given the current aging rate, and they have a direct and severe effect on their quality of life, since they limit daily communication. The effective support of adults with language impairments requires individualized, systematic, and regular intervention by speech and language therapists (SLTs). Timely assessment is an essential step for the identification of their communication abilities and deficits, for prognosis of functional recovery, as well as for the design of individualized intervention plans.

Direct, face-to-face (FTF) clinical services have been considered the gold standard of behavioral appraisal and intervention in the field of speech and language pathology. Effective service delivery requires clinicians to be able to offer real-time directions and feedback cues that are directly responsive to the patient's actions, utterances, or other type of behaviors during the clinical session. However, it could be argued that for some patients with communication disorders, FTF service delivery is not an ideal or a viable option. Especially for patients with significant physical and communication disabilities, FTF sessions may often require disproportionately high levels of physical, cognitive, and emotional effort on the part of the patient, as well as need for caregiver assistance, transport, and increased financial cost.

Teleassessment and telerehabilitation practices have been considered an effective alternative to in-person clinical services, long before the COVID-19 pandemic made them an urgent necessity. A survey of the American Speech-Language-Hearing Association, completed by 476 SLTs, indicated that 64% of the clinicians endorsed providing services *via* telepractice, 37.6% used telepractice for screenings, 60.7% used telepractice for assessment, and 96.4% used it for intervention (American Speech-Language-Hearing Association, 2016). Remote clinical services are particularly relevant in the context of stroke-induced speech and language impairments, such as aphasia, given the high levels of unmet needs and the increased service demands. Additionally, effectiveness of treatment in aphasia is linked to early appraisal and thus it is beneficial for clinicians to have a means to carry out comprehensive assessments of different aspects of the patient's communication abilities, including narrative discourse, with minimal effort from the part of the patient and the clinician, and without the need for transport away from the patient's residence.

To address these needs, a research project was set up, aiming to develop a technology-enhanced platform offering adults with acquired speech and language disorders the opportunity for remote long-term speech and language therapy in their own environment, without the physical presence of an SLT. With the aim to assist STLs in the process of patients' assessment and monitoring of the intervention outcome, the platform integrates a web-based application for the collection of spoken narratives from individuals with speech and language impairments and unimpaired controls. Data collected with this application serve as an evaluation corpus, against which a machine learning system for the automatic assessment of the severity of impairment will be tested for its accuracy and robustness. The system focuses on aphasia, as one of the most complex types of chronic acquired language disorders,

affecting the communicative abilities of a significant percentage of the adult population in multiple language modules and modalities.

The purpose of this paper is to present the design, methodological considerations and requirements of the web-based application for the elicitation of spoken narrative discourse from Greek-speaking people with aphasia (PWA) and unimpaired adults. The application allows the online administration of a comprehensive protocol of seven narrative tasks and is designed for remote as well as for in-person administration. Subsequently, we present the first phase of the evaluation of the presented method regarding its efficacy and feasibility in collecting data remotely, as compared to the traditional in-person setting. More specifically, a pilot study which involves only neurotypical adults will be presented, aiming to examine whether the linguistic properties of spoken narratives collected remotely are comparable to the ones collected in a FTF set up using the presented web-based application. The second phase of the method's evaluation will include language impaired participants, namely PWA.

1.1. The study of discourse in aphasia—Narrative discourse

The term *discourse* is commonly used to describe the way in which language is used and structured beyond the level of the sentence to convey an understandable message (Armstrong, 2000; Wright, 2011). The need to collect and study extensive discourse samples from PWA was identified in the late 1970s, when a discrepancy between language performance on standardized aphasia tests and real-life language use for social interactions, termed as *functional communication*, was identified (Holland, 1979).

Aphasia is an acquired language disorder, as a result of focal damage to the left cerebral hemisphere, caused by a cerebral vascular accident (CVA) or a traumatic brain injury (TBI) (Obler and Gjerlow, 1999).[1] Aphasia can affect the production and comprehension of both spoken and written language, at all language levels (phonological, morphological, syntactic and semantic) and to varying degrees, depending on the area and severity of the brain injury (Harley, 2001), causing mild, moderate or severe language impairment.

The aim of aphasia rehabilitation is to improve PWA's functional communication, i.e., individuals' language skills to achieve communication goals in the context of everyday social interactions. Thus, since the late 1970s there is an increasing focus on the study of contextualized language use, since there is an agreement in the aphasia literature that the controlled administration conditions of standardized aphasia assessment protocols mainly focus on isolated components of language - phonology, morphology, syntax and semantics- at word and sentence level, that do not simulate the cognitive requirements and conditions of real-life communication (Holland, 1982; Armstrong,

---

1   Moreover, another type of aphasia, Primary Progressive Aphasia (PPA), has been identified (Mesulam, 2001). PPA is a neurodegenerative clinical condition associated with frontotemporal dementia (FTD), which primarily affects language functions and is characterized by their gradual loss.

2000; Beeke et al., 2011; Olness and Ulatowska, 2011; Doedens and Meteyard, 2022). In this context, the study of discourse provides a naturalistic and ecologically valid framework for language assessment, which can reveal different aspects of language abilities, as well as weaknesses, of PWA in more natural communicative contexts, unravel interactions between individual language components and assess intervention effects in connected speech (Dietz and Boyle, 2018).

According to the systematic literature review of Bryant et al. (2016), which covers 40 years of the study of discourse in aphasia (1976–2015), studies using discourse analysis methods doubled in the second half of the 1990s, with the most significant increase observed from the late 2000s to 2015. The growing interest in the study of aphasia at the level of discourse over the last two decades has also been influenced by international frameworks and procedures in terms of assessment and rehabilitation, such as the International Classification of Functioning, Disability and Health (ICF) of the World Health Organization (World Health Organization, 2001). In this context, new approaches to the assessment and rehabilitation of communication skills of PWA have begun to develop, focusing on functional communication of individuals and their active participation in daily life, examining not only linguistic factors, such as the nature of the language impairment, but also social and psychological ones, such as social participation, social identity, self-esteem, and mental resilience. Using tools such as systematic observation and assessments of PWA's ability to respond effectively to specific communication situations, like maintaining a dialogue with another interlocutor or recounting a story, the functional approach highlights the level of discourse as a field of study in aphasia, emphasizing the role of the communicative setting and the context in discourse production (Armstrong et al., 2011).

For the study of discourse of PWA and unimpaired controls, different discourse types have been analyzed: exposition, procedural, narrative and conversational discourse. Exposition refers to discourse produced to describe or explain a topic, procedural discourse refers to the description of a process (e.g., how to make a peanut butter and jelly sandwich), narrative represents recounting of a fictional or factual story and conversational discourse is the interactive communication between two or more people. Among these discourse types, narrative is the one that has been more extensively investigated (Bryant et al., 2016). The study of narrative discourse is compatible with the functional approach to the rehabilitation of language disorders in PWA, since narrative is favored as an integral part of human communication, representative of everyday language use. Moreover, narrative offers a controlled framework for the analysis of coherence and discourse organization, provided by global story macrostructure.

1.2. Methods for eliciting narrative discourse from PWA and neurotypical controls

In recent surveys on the use of discourse data for the assessment of aphasia in clinical settings (Bryant et al., 2017; Cruice et al., 2020), as well as in research and clinical settings (Stark et al., 2021b),

it has been reported that most clinicians and researchers collect spoken discourse data from PWA and unimpaired controls using a variety of discourse elicitation methods. For instance, free narrative production tasks, such as personal narratives, or structured and semi-structured tasks, such as picture-elicited story production or story retelling, have been frequently used.

The elicitation of personal narratives in the context of the study of aphasia typically involves the narration the person's "stroke story", whereby the participants narrate the events of their stroke. Personal narratives have been employed in aphasia research because of their *multi-functionality*; as personal stories of an individual's experience, they actively involve both functions of narratives, i.e., the referential function, mainly related to the temporal arrangement of events, and the evaluative function, which conveys the narrator's attitudes, emotions, and opinions toward the narrated events (Labov, 1972). The evaluative function of personal narratives is critical for the study of PWA's communicative competence or functionality, since it reflects the intrapersonal and interpersonal function of narration (Olness and Ulatowska, 2011); PWA have the opportunity to talk about themselves and their chronic condition, and, therefore, to establish their sense of identity, self-image and self-expression and, at the same time, to share their experiences with others in the context of a social interaction (Fromm et al., 2011). For these reasons, personal narratives and, especially, illness narratives, are considered a natural context to investigate the degree in which PWA are able to accomplish the intrapersonal and interpersonal functions of storytelling, depending on their language impairments and severity level (Armstrong and Ulatowska, 2007; Olness et al., 2010; Olness and Englebretson, 2011), and, thus, they represent a discourse elicitation task, which is compatible to functional and social models in aphasia assessment and therapy. Moreover, in terms of specific measures of language ability, it has been reported that personal narratives involve more *correct information units* (*CIUs*, Nicholas and Brookshire, 1993) (Doyle et al., 1995), which are considered an objective measure of functional, real-life communication abilities in aphasia (Doedens and Meteyard, 2020), and are characterized by greater complexity, as evaluated by criteria such as vocabulary range, utterance length, subordination, etc., in relation to picture-elicited narratives (Glosser et al., 1988). However, the speech samples include personal events, and are therefore less comparable with each other in terms of informational content, than picture-elicited narratives.

Structured or semi-structured tasks involve two sub-types, which have been widely used in aphasia research: (a) the production of narrative discourse based on visual stimuli and (b) the retelling of stories that have been presented orally. Picture-elicited narratives, termed also as "expositional narratives" (Stark, 2019), or "picture descriptions" (MacWhinney et al., 2011), involve narration based on a single picture or a picture sequence. Tasks of this type do not burden memory, provide a controlled context for narrative production and guide participants to produce comparable stories in terms of narrative macrostructure, lexical elements, main events, and information units. The picture prompts that have been most widely used to elicit narrative discourse include the *Cookie Theft*, which is part of the Boston Diagnostic Aphasia Examination (BDAE) (Goodglass and Kaplan, 1972), the *Picnic*, which is part of the Western Aphasia Battery (WAB-R) (Kertesz, 2006), and

the *Cat Rescue*, the *Birthday Cake*, the *Fight*, and the *Farmer and His Directions* single picture and picture sequences stimuli by Nicholas and Brookshire (1993). It is worth noting that single-picture stimuli, such as the *Cookie theft*, the *Picnic* and the *Cat Rescue*, are not transparently associated with the narrative discourse type. Many studies and aphasia assessment procedures require participants to describe either single pictures or picture sequences as isolated scenes, resulting in the production of a static description of situations and characters, rather than narratives, which typically involve the identification of underlying temporal and causal relations between events (Armstrong, 2000; Wright and Capilouto, 2009). It has been demonstrated that the discourse type elicited (description vs. narrative) as well as the quality of narrative production heavily depend on task instructions given to participants to produce discourse (Olness, 2006; Wright and Capilouto, 2009). Moreover, single pictures have been reported to elicit lower narrative levels, in terms of narrative structure complexity (Lemme et al., 1984; Bottenberg et al., 1985), lower cohesive harmony (Bottenberg et al., 1985) and lower lexical diversity (Fergadiotis and Wright, 2011) than picture sequences.

The second type involves the retelling of a story that participants have previously heard. This type of task does not require speakers to construct the narrative content themselves but requires them to retain the events of the story and their temporal succession, to recall them from memory and reproduce them. The most well-known protocol of this type is the *Story Retell Procedure* (Doyle et al., 2000), which is based on the visual stimuli of Nicholas and Brookshire (1993) and has been evaluated as a reliable and valid method of narrative elicitation (Doyle et al., 2000; McNeil et al., 2001, 2002, 2007). This method, despite the processing demands on working memory posed to participants, produces comparable speech samples, in terms of measures of linguistic performance, to the ones obtained from picture-elicited narratives and procedural discourse tasks (McNeil et al., 2007).

Given the diverse clinical patterns in aphasia and task effects that have been observed in discourse production, regarding parameters such as verbal productivity, fluency, information content, grammatical accuracy, and complexity (Doyle et al., 1998; Stark, 2019), it is recommended (Brookshire and Nicholas, 1994; Armstrong, 2000; Olness, 2006; Stark, 2019; Stark and Fukuyama, 2021) that a combination of elicitation methods should be used to obtain a comprehensive language sample most resemblant of actual language use. To this end, AphasiaBank, the largest repository of multi-lingual and multi-modal data for the study of communication in aphasia, implements a standard protocol for the elicitation of spoken discourse from PWA and unimpaired controls (MacWhinney et al., 2011), which includes personal narratives, picture-elicited narratives, familiar story telling and procedural discourse.[2] Personal narratives include the narration of the PWA's "stroke story" and an "illness story" from neurotypical controls, as well as the story of an important event from both groups. The *Cat Rescue* picture prompt (Nicholas and Brookshire, 1993), as well as the *Refused Umbrella* and the *Broken window* picture sequences are used for the collection of picture-elicited narratives. The

AphasiaBank protocol also includes the narration of the traditional fairytale of *Cinderella*, after the revision of a wordless picture book (Grimes, 2005), which is removed prior to narration (Saffran et al., 1989). Finally, PWA are prompted to produce procedural discourse, in which they describe the procedure of making a *Peanut Butter and Jelly Sandwich*.

Moreover, it is evidenced that the amount of data collected per participant is an important issue related to sufficient discourse sampling, so speech samples are representative of participants' language abilities (Brookshire and Nicholas, 1994; Boles and Bombard, 1998; Armstrong, 2000). Brookshire and Nicholas (1994) found that test-retest stability of two measures of spoken discourse, words per minute and percentage of correct information units, increased as sample size increased. They suggested that a speech sample obtained from 4 to 5 different tasks containing a total of 300–400 words per participant represents a sufficient sample size to achieve acceptable high test-retest stability. Boles and Bombard (1998) investigated adequacy of sample size in terms of time duration of the conversational discourse sampled per participant. They found that 10-min samples represented an adequate conversation length to reliably measure the variables of conversational repair, speaking rate and utterance length.

Despite the recommendations for collecting spoken discourse samples with a variety of discourse elicitation methods, it is reported that the number of samples collected for the assessment and analysis of spoken discourse usually ranges from one to four, with most investigators collecting one or two samples per person (Stark et al., 2021a). Several studies highlight significant barriers in implementing discourse data collection methods in clinical practice, the most typical of which is the lack of tools and resources, such as computer software or hardware and audio equipment, as well as inadequate training, knowledge, and skills in discourse collection (Bryant et al., 2017; Cruice et al., 2020; Stark et al., 2021a).

1.3. Remote assessment of language abilities

The elicitation of narratives has been primarily obtained *via* direct FTF set up. However, technological applications for language or communicative skills assessment in educational or clinical settings have some significant advantages over FTF services: (a) they are more practical, since all materials and prompts are integrated in a comprehensive computer environment, and all equipment needed, such as microphone, audio player, and speakers, is built-in and easily accessible *via* the application interface; (b) uniformity of task presentation and administration is facilitated regarding several parameters, such as order of tasks presentation, stimuli presentation order and timing, and instructions format; (c) data storage and management is significantly simplified, since data are stored and logged in the application's backend, allowing the investigators easy data access, tracking and filtering. Moreover, keeping in mind that the ultimate goal of the data collection method reported here is to assess individuals with aphasia, it should be noted that the available research for the delivery of remote clinical services, including teleassessment and telerehabilitation, has produced promising results.

---

2   The materials and guidelines for the administration of the AphasiaBank protocol can be accessed at https://aphasia.talkbank.org/protocol/english/.

Regarding teleassessment, remote appraisal of language and other communication functions involves either the adaptation of conventional, already available assessment tools for online use, or the development of novel instruments specifically developed for remote administration. In fact, there is a significant body of research on the validity of computerized testing of neurotypical populations (Newton et al., 2013). Although studies conducted in the early 1990s have reported considerably poorer performance on computerized compared to pen and paper tests, the disadvantage found for scores of computerized assessments is now getting smaller and the two procedures are considered comparable (Noyes and Garland, 2008). The discrepancies often reported between the two methods have been attributed to a number of factors, including computer experience, anxiety and participant perceptions toward computerized testing (Newton et al., 2013).

Although there are numerous assessments of speech, language and communication available for aphasia, the literature on the validation of teleassessment versions of these tools is still lacking. However, several studies have demonstrated the potential validity of web-based versions of widely used aphasia assessments. More specifically, Theodoros et al. (2008), Hill et al. (2009), and Palsbo (2007) have compared online vs. FTF administration of the short versions of the BDAE-3 (Goodglass et al., 2001) and the Boston Naming Test (BNT, 2nd edition). Newton et al. (2013) compared computer-delivered and paper-based language tests, including parts of the Comprehensive Aphasia Test (CAT) and the Test of Reception of Grammar (TROG), and Dekhtyar et al. (2020) validated the Western Aphasia Battery–Revised (WAB-R) for videoconference administration.

In the study of Theodoros et al. (2008) and Hill et al. (2009), 32 patients with aphasia due to stroke or traumatic brain injury were grouped in terms of severity and were assessed both FTF and remotely on the BNT and the BDAE-3. This assessment battery targets both oral and written language, expression and comprehension (spontaneous speech, picture description, naming, repetition, auditory comprehension, reading and writing), thus the computerized version involved the online presentation of a variety of visual and oral stimuli and the recording of both verbal (words, phrases, sentences etc.) and non-verbal responses (while using a touch screen). Overall, test scores obtained from the two delivery modes were comparable within each aphasia severity level. Additionally, the majority of participants were comfortable with the online administration process, confident with the results obtained, and equally satisfied with either FTF or web-based delivery. Researchers reported that severity of aphasia might have influenced the ability to assess two of the eight subtest clusters in the online condition, namely the naming cluster and the paraphasia tally cluster; the latter is based on the quantity and type of paraphasic errors. However, these clusters also displayed a high level of agreement between the FTF and the remote method for all severity levels. Clinical comments also indicated that the remote administration of some subtests was more laborious when assessing patients with severe aphasia. The authors concluded that although aphasia severity may increase the challenges of remote assessment, it does not have an impact on assessment accuracy.

In Palsbo's (2007) randomized agreement study, 24 poststroke patients were assigned to either a remote or a FTF assessment

of functional communication based on a subset of the BDAE i.e., the subsection of *Conversational and Expository Speech* (description of the Cookie Theft picture) and the sections of *Auditory Comprehension* (commands and complex ideational material, respectively). These tasks involved the use of visual and oral stimuli, and the recording of both verbal and non-verbal responses of patients. Overall, it was found that remote assessment of functional communication was equivalent to FTF administration; percentage agreement within the 95% limits of agreement ranged from 92 to 100% for each measure of functional communication. However, it should be pointed out that percentage of exact agreement between clinicians was much lower when the BDAE was administered remotely, than when it was administered by the FTF examiner. Given that the authors did not randomize the clinicians between remote and FTF assessments, it is not clear whether this discrepancy was related to the remote administration.

In the study by Newton et al. (2013), 15 patients with aphasia were assessed in three conditions, FTF or remotely, with and without the presence of a clinician, on two language comprehension tasks, i.e., a sentence-to-picture matching task and a grammaticality judgment task, that required oral and/or visual stimuli but non-verbal responses. PWA also expressed their perceptions of each condition *via* questionnaire rating scales. High correlation of the test scores across the three conditions was attested, which suggests the remote test format was sensitive to the same factors and measured the same constructs as the FTF test version. However, it was also found that computerized administration could increase test difficulty, given that participants performed significantly lower on the remote test condition. Overall, PWA preferred the FTF assessment method, although some participants felt comfortable with the remote administration. The authors conclude that remote testing can be used for the assessment of PWA, but comparison between scores obtained by remote and FTF methods should be exercised with caution.

Dekhtyar et al. (2020) compared in-person vs. remote administrations of the WAB-R, a comprehensive test that is often considered a core outcome measure for language impairment in aphasia (Wallace et al., 2019), in 20 PWA with a variety of aphasia severities. Despite the presence of some performance inconsistencies attributed to individual variability (five of the 20 participants showed changes in aphasia classification; however, this was a result of minimal changes of the actual scores), there were no significant differences between the FTF and online conditions for the WAB-R scores, and high participant satisfaction was reported for the videoconference administration. The authors concluded that the two methods of administration of the WAB-R test can be used interchangeably.

Apart from the above attempts to validate teleassessment in aphasia, Choi et al. (2015) and Guo et al. (2017) developed tablet-based aphasia assessment applications based on conventional evaluation protocols. Choi et al. (2015) developed a mobile aphasia screening test (MAST) designed as an iPad application, based on a conventional, widely used screening, K-FAST (Ha et al., 2009), the Korean version of the Frenchay Aphasia Screening Test. Sixty stroke patients, 30 with and 30 without aphasia were assessed FTF using K-FAST and the Korean version of WAB, and remotely using MAST. MAST uses a word-to-picture

matching task to assess auditory comprehension, as well as a picture description and a phonemic verbal fluency task to assess verbal expression. Patient responses are stored in a central web portal accessible to the service providers. The system scores the comprehension task automatically, whereas the verbal expression section is scored manually offline. The authors found that MAST had high diagnostic accuracy and correlated significantly with the conventional test and screening.

Going beyond aphasia screening, Guo et al. (2017) developed and validated "Access2Aphasia", a tablet videoconferencing application for the remote comprehensive assessment of aphasia at the impairment, activity and participation levels, based on the ICF framework (World Health Organization, 2001). Thirty PWA were randomized into either FTF or remote administration of the spoken word to picture matching and the naming tasks of the Psycholinguistic Assessment of Language Processing in Aphasia (PALPA), and the Assessment of Living with Aphasia (ALA) questionnaire. The study found moderate to almost perfect agreement of the online and the conventional assessment, and comparable intra- and inter-rater reliability for the two conditions.

Regarding specifically the remote elicitation and analysis of narrative discourse, Brennan et al. (2004) and Georgeadis et al. (2004) assessed 40 patients with a recent onset of either a stroke or TBI on the production and comprehension of spoken narratives, both in-person and by videoconference. The authors used a standardized discourse elicitation protocol, the Story Retell Procedure (Doyle et al., 2000). In each condition, patients listened to three pre-recorded stories accompanied by a series of black-and-white line drawings, that, in the remote condition, were scanned and displayed on a computer monitor. After the completion of the story, all pictures were displayed together, and the clinician asked each participant to retell the story using her/his own words. In both conditions, the patient's narrative was recorded and analyzed offline, in terms of the percent of information units, i.e., percent of intelligible utterances that convey accurate information relevant to the story (McNeil et al., 2001). The researchers did not report any significant differences in the patients' performance between the two assessment conditions. Additionally, a high level of acceptance of the remote version of the narrative elicitation procedure was reported. However, it is worth mentioning that patients with TBI were less likely, compared to stroke patients, to use videoconferencing for communication with the clinician. Very recently, AphasiaBank has also released an electronic version of the AphasiaBank standardized discourse elicitation protocol, which can be used for assessing PWA remotely in the context of a videoconference.[3]

The above literature review underscores the continuous and increasing need for developing and testing novel, remote assessment methodologies across all modalities and domains of communication, addressed to PWA of different types and severity levels. The available body of research has investigated web-based versions of conventional Aphasia tests and tools that specifically elicit and analyze narrative discourse. All past research

_____

3   The scenarios for the remote administration of the AphasiaBank protocol to PWA and unimpaired controls are available at the AphasiaBank website (Sections 1–4), https://aphasia.talkbank.org/protocol/english/.

studies support the validity, feasibility and reliability of web-based assessment for PWA and indicate that conventional assessment procedures can be modified to accommodate computer delivery. Although most discrepancies found between the two modes of administration were minimal and non-systematic, there is some evidence, from both neurotypical populations (Noyes and Garland, 2008) and PWA (Newton et al., 2013), to indicate a small systematic disadvantage of scores obtained *via* computer-based and/or remote assessment. Although this discrepancy seems to concern more performance on standardized tests and test tasks rather than production of narrative discourse, it does imply that caution should be exercised when comparing scores collected *via* different methods of elicitation, i.e., FTF vs. remote assessment.

2. A method for online narrative discourse data collection

The presented method was designed as a result of the COVID-19 pandemic restrictions, with the aim to enable spoken narrative discourse data collection in both in-person and remote settings from PWA and neurotypical adults, using a web-based application. The speech samples collected with the presented method are intended to be used as evaluation data for a machine learning algorithm aiming to predict aphasia severity level on the basis of several linguistic features.

2.1. Protocol for narrative discourse elicitation

Since it is generally accepted that different discourse elicitation methods may impose different cognitive and linguistic demands, the literature suggests variety in discourse elicitation methods to address the diversity of clinical characteristics in aphasia (see Section 1.3). To address the issues of task variety and sufficient sample size per participant, we implemented a protocol for eliciting narrative discourse which comprises four discourse elicitation methods: (i) free narrative production (personal narrative), (ii) story production based on a single picture or a picture sequence (picture-elicited narrative), (iii) familiar story telling, and (iv) retelling of a previously heard story. These discourse elicitation methods are represented in seven narrative tasks. Four of them are adopted from the Kakavoulia et al. (2014) narrative elicitation protocol developed for the assessment of Greek-speaking PWA (Varlokosta et al., 2016), which includes the following tasks:

A. "*Stroke story*": PWA narrate the personal story of their stroke, while unimpaired participants narrate a health-related incident about themselves ("health or accident story").
B. "*the Party*": Story production based on a six-picture sequence. The picture stimuli are original and depict an adult every-day life incident: a young man, disturbed by the noise caused by a party in the adjoining apartment, gets upset and visits his neighbors to complain about it.
C. "*the Ring*": Retelling of an unknown recorded story, supported by a five-picture series. The story is original, with the structure of a traditional fairy tale. It is about the love story of a prince and a

young woman, that is hindered by the prince's evil stepmother who steals a ring, the evidence of the prince's love for the young woman. The participants are offered visual support by a sequence of five pictures depicting main events of the story while narrating the story.

D. "*Hare and Tortoise*": Retelling of a familiar recorded story, which is an adaptation of the Aesop's fable, without visual support.

Additionally, three tasks, shown below, are adopted from the AphasiaBank standard discourse protocol, to achieve the collection of larger language samples from each participant and to be compatible with the elicitation methodology of the AphasiaBank database and comparable to the studies conducted with this methodology. The two protocols share the task of the free production of a personal story (*Stroke story*).

E. "*Refused umbrella*": Story production based on a six-picture sequence.
F. "*Cat rescue*": Story production based on a single picture.
G. "*Cinderella*": Narration of the traditional Cinderella fairytale. Participants first go through the wordless storybook to refresh their memory and afterwards they narrate the story without looking at the pictures.

The AphasiaBank protocol has been widely used for the collection of spoken discourse from PWA and healthy controls. The repository hosts data from nearly 300 PWA and 200 age-matched unimpaired controls (MacWhinney and Fromm, 2016). The AphasiaBank data are being analyzed with CLAN programs and have been used for the investigation of several research topics addressing PWA and unimpaired controls, such as discourse, grammar, gesture, lexicon, fluency, automatic classification, social factors, and treatment effects. Thus, it comprises tasks that have been multiply validated across different types of aphasia at different levels of analysis. The Greek protocol for oral narrative discourse collection has been used to collect data from 22 PWA and 22 age and education-matched controls, the spoken discourse samples of whom comprise the Greek Corpus of Aphasic Discourse (GREECAD) (Varlokosta et al., 2016). Table 1 presents the correspondence between discourse elicitation methods and narrative tasks of the implemented protocol.

The *Party*, the *Cat rescue* and the *Refused umbrella* tasks involve narration on the basis of picture stimuli. However, *Cat rescue* involves a single picture stimulus, and, therefore, differs from *the Party* and the *Refused umbrella* tasks, which involve a six-picture sequence stimulus. The *Party* picture sequence depicts an incident of adult everyday life, compared to the *Refused umbrella* which represents a child life incident. The *Ring* and *Hare and tortoise* represent two retelling tasks, since they require the reproduction of a recorded story, but differ significantly in the linguistic and cognitive demands they impose to participants. The *Ring* is an unknown lengthy story, with many episodes and a complex plot, characteristics which increase the linguistic and cognitive demands of the task. The visual support offered by five pictures, which depict the main events of the story, is expected to compensate for the increased task demands. Hare and tortoise is a familiar story, especially in the Greek culture, thus no visual support is offered. Regarding the *Cinderella* narration task, several terms have

been used to describe it. It has been termed as "story narrative" (MacWhinney et al., 2011), "story retelling" (Stark and Fukuyama, 2021) and "storytelling" (Fromm et al., 2022). We chose not to refer to this task as "retelling", since it does not require reproduction of an already heard story, but as "familiar storytelling", to indicate the activity of the narration of a well-known fairytale.

2.2. Web application for data collection

The narrative discourse elicitation protocol presented in Section 2.1., including task instructions, as well as visual and audio stimuli for each task, is integrated in a custom web application for use on computer. The web application development had to meet specific requirements regarding different aspects of use: (a) consistent administration, (b) secure audio recording and good sound quality, (c) friendly and easy-to-use interface.

The parameter of consistency of administration is associated with the specifications regarding the order of task presentation, timing and modality of task stimuli, as well as content and modality of task instructions, which should follow the same principles in both the pen-and-paper and online administration, ensuring that the elicitation of spoken discourse samples is representative of the participants' language abilities, and that the elicitation is carried out in a consistent way across investigators. Presentation of tasks follows a linear sequence, beginning with the AphasiaBank tasks (A, E, F, G), followed by the Greek elicitation protocol tasks (B, C, D). Task administration follows a consistent flow, starting with task instructions, which are presented in both written and audio format to ensure that the goal of the task is clear to the participant. Subsequently, the task stimuli are presented, according to each elicitation method. In narrative tasks prompted by a picture or a picture sequence (tasks B, E, F), the picture(s) is/are presented to the participant on screen. In the case of picture sequences, the pictures appear all at once, with a number indicating their order (Figure 1). The order of pictures is pointed out by the investigator, who allows the participant enough time to go through the picture sequence and form a mental representation of the story. The pictures appear in their actual size or can be enlarged (e.g., *the Party*), so the participant can examine them in detail, serially, using the "next" navigation arrow (Figure 2).

In the case of the *Cinderella* task, the picture book has been integrated in its digital format, following the exact pagination of the printed book version. The participant can examine the book linearly, flipping through its pages back and forth, using the navigation arrows.

In tasks requiring the retelling of a recorded story (tasks C and D), story playback is initiated and controlled using an integrated audio player. In case retelling is supported by pictures (task C), the pictures appear on screen while the participants are listening to the recorded story; picture enlargement option is also available (Figure 3).

Once stimuli presentation is completed, voice recording begins using the microphone icon, which appears on each task screen at the same position. Recording ends by pressing the microphone icon again and it is submitted by pressing the "next task" icon.

TABLE 1 Correspondence between narrative tasks and narrative discourse elicitation methods.

Narrative tasks	Discourse types			
	Personal narrative	Picture-elicited narrative	Familiar storytelling	Story retelling
Stroke/health story	X			
the Party		X		
the Ring				X
Hare and tortoise				X
Cat rescue		X		
Refused umbrella		X		
Cinderella			X	

FIGURE 1
Home screen of the *Refused umbrella* task (picture adopted from the AphasiaBank protocol, MacWhinney et al., 2011).

2.3. Administration specifications

An elaborated Elicitation and Administration Guide[4] is provided to the investigators who administer the protocol. The guidelines serve the following main goals: (a) to ensure elicitation of the narrative discourse type, and not descriptive or conversational discourse, (b) to minimize verbal interventions from the part of the investigator, so the audio files acquired will be as "free" of non-participants' voice as possible, and (c) to facilitate uniformity of administration across settings and investigators and, therefore, acquisition of comparable speech samples across participants. The guidelines follow the AphasiaBank instructions[5] and include

---

4   The Elicitation and Administration Guide is available in Greek at https://www.planv-project.gr/files/applications/Elicitation_Administration_Guide_PLan-V.pdf.

5   Instructions for the remote and local administration of the AphasiaBank protocol are available at https://aphasia.talkbank.org/protocol/english/.
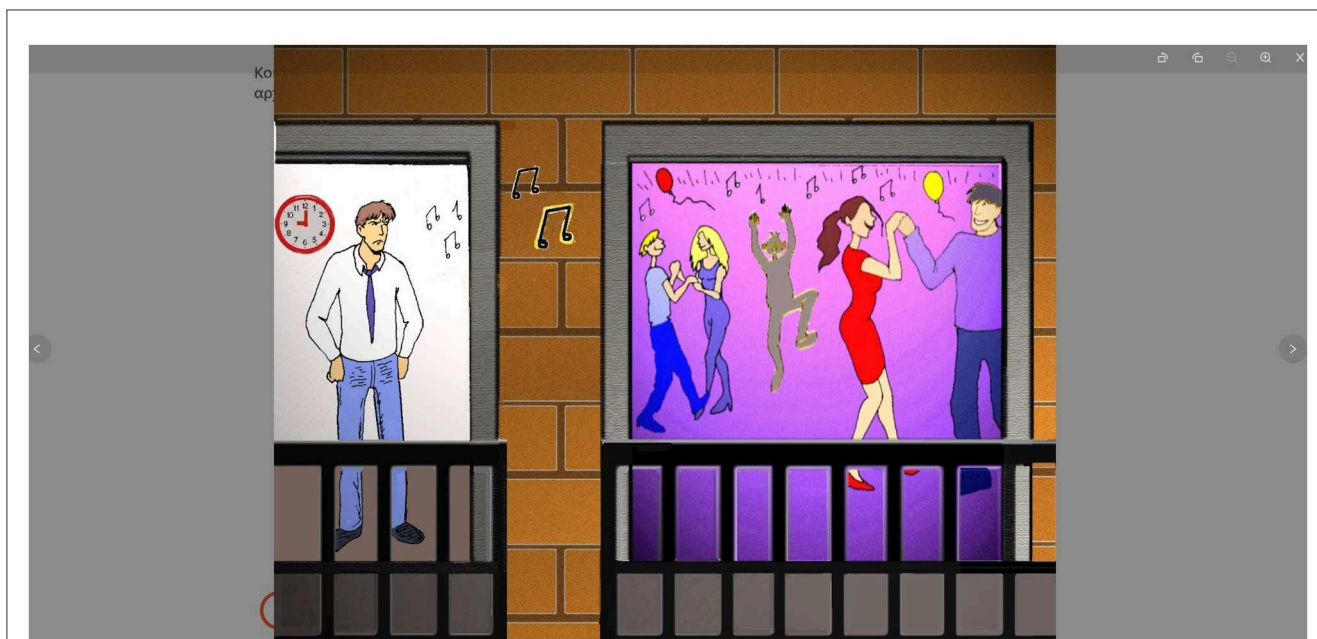
FIGURE 2
Picture enlarging functionality of the *Party* task (picture adopted from the Kakavoulia et al., 2014 protocol).

specific prompts that should be used in case there are difficulties in story production, as well as troubleshooting questions for each task.

The instructions of the seven discourse tasks emphasize the production of the narrative mode and prompt participants toward the production of stories with a beginning, middle and end (Wright and Capilouto, 2009). At the same time, specific instructions are given regarding verbal encouragements and interruptions. Investigators are instructed to avoid verbal encouragements and use non-verbal cues instead, such as facial expressions, eye contact and gestures. Moreover, type and degree of verbal encouragements and facilitating questions are controlled, allowing for gradually more specific prompts, which range from general encouragements to story-specific aids, in case of no response or serious speech halting.

The administration procedure is preceded by a brief introduction about the data collection process and its purpose. Participants are requested to narrate the stories at their own pace, and it is clarified that the investigator will not intervene, unless it is needed. They are also informed about issues of personal data protection and data access rights.

Participants sit in front of a computer screen or a laptop, which has a built-in microphone. In case there is an external microphone, it is placed close to the participants, between them and the screen. Investigators are advised not to stand next to the participants while they are examining the picture stimuli and producing the stories, so no shared knowledge of the story is assumed by the participant, a factor which might affect the linguistic characteristics of the narrative (Holler and Wilkin, 2009). For this reason, investigators are facing the participant and are not sitting next to her/him. Since the elicitation protocol is lengthy, the web application design allows its administration in several sessions. When a session of a number of tasks is completed, the investigator can exit the application, while all data collected are saved in the database. Login to the

application with the same account initiates a new session, where the investigator can select the next task from a dropdown menu, skipping the tasks that have been already completed.

The same procedure, as well as elicitation and administration guidelines, are being followed in case a participant is assessed remotely in her/his own environment. Remote administration is carried out *via* the web-based application in the context of a teleconference, while the investigator is offering guidance to the participant for navigating through the application (see more in Section 3.2). Verbal encouragements and interruptions are still recommended to be avoided, as in face-to-face sessions. However, they are less likely to occur in remote elicitation settings, since the investigators are advised to keep their microphone muted during participants' recording. Moreover, the factor of spatial proximity between the participant and the investigator while narrating a story, which might favor the assumption of shared knowledge of the story, is not present in the remote elicitation setting. It should also be noted that full implementation of remote administration is feasible with the use of any remote desktop software, if desired, which allows the investigator greater control of the application operation and minimizes participants' interactions with the computer.

2.4. Technical specifications

The application is written in Javascript, using the ReactJS environment for the development of the interactive interface with browser access. Audio files and images are uploaded and stored on a remote computer (server), which uses Flask server software, while user accounts are stored in databases. Development issues were primarily related to the quality of sound recording, which is essential for collecting samples for speech processing purposes,

Θα ακούσετε μια ιστορία για το τι συμβαίνει σε αυτές τις εικόνες. Μόλις τελειώσει θα πρέπει να την επαναλάβετε όσο το δυνατόν πιο ολοκληρωμένα. Κοιτάξτε πρώτα τις εικόνες.

FIGURE 3
Screen of the *Ring* task, with audio and visual stimuli (picture adopted from the Kakavoulia et al., 2014 protocol).

especially from people whose speech impairments might have affected their voice quality in terms of intensity. More specifically, the AudioWorklet technology was used, so that the recording and any other processing can be done in a separate calculation thread, leaving the basic functions of the user's computer unaffected. This ensures there are no distortions or interruptions in the audio files. The files are recorded in high quality, with a sample rate of 44,100 Hz, 16 bit (cd quality), and sent *via* streaming during the recording to ensure a smooth user experience. Recordings are saved in a password-protected database, which provides data tracking information for user, task, session, date and time.

3. Evaluation of the web-based data collection method in neurotypical adults

The web application for narrative discourse data collection has so far been implemented for collecting narratives from PWA in an in-person setting, where SLTs have been administering the protocol of narrative tasks in the context of patients' in-house treatment. Moreover, spoken narratives are being collected from unimpaired

individuals, both in in-person and remote settings. Data are being collected, manually transcribed and processed with Natural Language Processing (NLP) tools, with the aim of quantifying text properties of spoken discourse production at several types and degrees of impairment.

To evaluate and investigate the efficacy of the presented web-based data collection method in remote elicitation settings, a pilot study was conducted, aiming to compare narratives elicited remotely to narratives collected in an in-person elicitation mode from unimpaired adults. The main research question of the study is: *Are spoken narratives produced in remote elicitation settings comparable to the ones produced in in-person settings, in terms of specific characteristics of language production at the discourse level*?

3.1. Participants

The study involved ten unimpaired, Greek-speaking adult participants (five male and five female) and applies a random sampling method. However, since the objective of the study is to investigate the efficacy of an online discourse elicitation method, which addresses a specific adult population, PWA, certain

TABLE 2 Participants' demographics and cognitive tests scores.

Sex (N)	Age (years)	Education (ISCED level)	MMSE score	5-objects score
	Mean (SD) Min−Max		Mean (SD) Min−Max	Mean (SD) Min−Max
F (5)	65.8 (4.08)	4–5	29.2 (0.74)	25 (0.00)
	61–70		28–30	−25
M (5)	61.8 (5.06)	6–8	30 (0.00)	25 (0.00)
	57–70		−30	−25

inclusion criteria have been applied, so the sample is comparable to the characteristics of the population that the method intends to assess. Therefore, since aphasia is more common in older than younger adults[6] (Ellis and Urban, 2016), the participants' age ranges from 57 to 70 years (mean age = 63.8). Moreover, since it is evidenced that educational level affects language abilities of PWA (González-Fernández et al., 2011), as well as of unimpaired individuals (Radanovic et al., 2004), the independent variable of educational level should be controlled. Therefore, the sample involves participants with a minimum ISCED level 4, i.e., individuals who have completed upper secondary education (N = 3), participants having completed short-cycle tertiary education (ISCED level 5, N = 2), as well as participants who hold a Bachelor's, Master's, or Doctoral degree (ISCED levels 6, 7, 8, N = 2, 1, and 2, respectively), according to the International Standard Classification of Education (UNESCO-UIS, 2012). All participants were screened using neuropsychological tools, namely the Mini Mental State Examination (MMSE) (Folstein et al., 1975) and the 5-Object cognitive screening test (Papageorgiou et al., 2014), to ensure that their cognitive abilities were within the norms. Participants' demographic details are presented in Table 2.

The study was approved by the Bioethics Committee of the coordinating organization, University of Patras, Greece. All participants have signed a Participation Consent Form, after being informed on the study's aims and objectives via a Participant Information Sheet.

3.2. Data collection procedure

The study uses a within-subjects experimental design, since all participants took part in both experimental conditions: in-person and remote elicitation set up. The same investigator administered the tasks to the same participant in both conditions, to eliminate the effect of the investigators' communication style on participants' language production. Investigators were two linguists and one SLT, members of the research team, experienced in administering assessment protocols and familiar with the specific narrative elicitation tasks. To eliminate the effect of prior knowledge of the stories and task familiarity on language production, as a result of the administration order between the two conditions, the sample was split into two groups, with the first group being first investigated in

---

6   National Institute of Deafness and Other Communication Disorders. Available online at: https://www.nidcd.nih.gov/health/aphasia (last accessed April 6, 2022).

the remote condition and the second in the in-person condition. The two sessions had a minimum time distance of 1 week from each other.

In the remote elicitation setting, participants performed the narrative elicitation tasks via the web application in their own environment, with the assistance of an investigator in the context of a virtual meeting (video conferencing). The participants shared their screen, so the investigator could help them enter their account credentials and navigate through the application. The investigators followed the same guidelines regarding order of tasks, instructions, and interventions. Recording was directly done through the participants' laptop built-in microphone, and not through the investigators' laptop speakers, to avoid sound distortion. Investigators had their microphone muted, to avoid voice interference and overlaps during the participants' narration.

In the in-person elicitation setting, the participant was in the same room with the investigator, who administered the tasks via the web application following the same guidelines.

3.3. Data transcription and processing

A total of 139 spoken narrative samples was collected, 70 elicited in the in-person condition and 69 in the remote condition. Ten participants performed seven narrative tasks each, in two conditions, with one missing data-point for one narrative task in the remote condition. The recorded narratives were manually transcribed in an orthographic format by three researchers using the ELAN software. Transcriptions were manually time-aligned, to allow for the automatic calculation of duration, and manually segmented into utterances, following the AphasiaBank guidelines for utterance segmentation; each utterance includes only one main clause along with its depended subordinate clauses. Repetitions, reformulations, and false starts were not included in utterances transcription for uniform word count and MLU calculation. According to the AphasiaBank guidelines, the period and the question mark were used as utterance terminators. Also, wide use of comma was applied, to indicate boundaries of clauses and phrases, that would facilitate NLP tools to perform accurate syntactic parsing. All transcripts were evaluated and normalized by a single researcher to ensure uniformity in utterance segmentation, as well as consistent application of orthographic and punctuation criteria (e.g., use of comma before a subordinate clause, use of full stop only at utterance final position, word contractions etc.). Table 3 presents an overview of the study dataset.

TABLE 3 Overview of the study dataset.

	In-person	Remote
Narrative samples (*N*)	70	69
Total *N* of tokens	14,137	14,101
Mean *N* of tokens per participant (SD)	1,411.9 (424.57)	1,452.58 (423.006)
Min–Max *N* of tokens per participant	918–2,441	936–2,492
Total duration (s)	7,827.86	7,817.004
Mean duration per participant(s) (SD)	782.24 (206.38)	783.63 (214.86)
Min–Max duration per participant(s)	595.99–1,218.34	561.54–1,292.38

Transcripts were extracted by participant and task in plain text format and processed with the Neural NLP Toolkit for Greek[7] (Prokopidis and Piperidis, 2020). The Neural NLP Toolkit for Greek is a state-of-the-art suite of NLP tools for the automated processing of Greek texts, developed at the Institute for Language and Speech Processing/Athena Research Center (ILSP/ATHENA RC). It currently integrates modules for part of speech (POS tagging), lemmatization, dependency parsing and text classification. The toolkit is based on code, models and language resources developed at the NLP group of ILSP.

3.4. Measures of narrative discourse production

According to recent literature reviews (Bryant et al., 2016; Pritchard et al., 2018) on the analysis of discourse in aphasia, more than 500 linguistic variables are being used to measure spoken language abilities and intervention outcomes of PWA. To address the variety and heterogeneity in methods, measures and analyses of spoken discourse samples, recent research initiatives are being undertaken toward the standardization of measures and methods (Stark et al., 2021b), as well as the identification and evaluation of primary linguistic variables for the reliable assessment of language abilities in aphasia across discourse types and elicitation methods (Stark, 2019). Moreover, given the growing availability of shared databases, such as AphasiaBank, as well as of tools for automated language analysis, statistical and machine learning methods are being increasingly applied for the automatic analysis, assessment and classification of PWA's speech samples, quantifying their linguistic properties and translating them into features used for the computational modeling of aphasia (Stark and Fukuyama, 2021; Fromm et al., 2022).

Following Stark (2019), who extracted from AphasiaBank data a set of eight primary linguistic variables which serve as proxies for various language abilities at spoken discourse level, the same set of features was selected to measure spoken language production

---

of participants at both experimental conditions. These features correspond to the language abilities of linguistic productivity (MLU), content richness (propositional density), fluency (words per minute), syntactic complexity (verbs per utterance, open/closed class words, noun/verb ratio), lexical diversity (lemma/token ratio) and gross linguistic output (number of words). These linguistic variables have been evaluated by Stark (2019) in a large sample of PWA and unimpaired controls, drawn from the AphasiaBank database, across three discourse types, expositional, narrative, and procedural discourse, corresponding to four discourse elicitation tasks of the AphasiaBank protocol: *Broken window, Cat rescue, Cinderella* and *Peanut Butter and Jelly* (procedural discourse). Her analysis showed significant effects of discourse type on the linguistic properties of spoken discourse in both groups, with similar findings across groups regarding discourse type sensitivity to primary linguistic variables.

In the present study, the measure representing lexical diversity was modified, since lemma/token ratio measure, which considers inflectional variants of the same lemma as the same type, was favored over the most commonly used type/token ratio measure, which treats inflected forms of the same lemma as different types. This decision was based on studies of lexical diversity in narrative discourse of PWA and unimpaired controls (Fergadiotis and Wright, 2011; Fergadiotis et al., 2013) which performed a lemma-based analysis of lexical diversity. In these studies, different inflected forms of the same word, for example eat, eats, ate, were counted as one and the same type. The reason for counting only unique lexical representations as separate types was to avoid conflating the measure of lexical diversity with the one of grammaticality, as reflected on the use of different inflected forms of the same lemma. Lemma/token ratio has also been applied in measuring lexical diversity in EFL learner corpora (Granger and Wynne, 1999), as the use of different lemmas (such as go, come, leave, enter, return) indicates greater lexical richness than the use of different inflected forms of the same lemma (such as go, goes, going, gone, went). For these reasons, as well as given that Greek is a highly inflected language, the present study adopts the lemma-based analysis of lexical diversity as a more representative measure of speakers' vocabulary range.

Moreover, two additional measures of syntactic complexity at the inter-clausal level were implemented, subordinate/all clauses ratio and mean dependency tree height. These measures were selected as relevant to NLP-based linguistic features extraction for the automatic processing of language data. They have been widely employed as highly effective measures of linguistic complexity in various fields of the automatic processing of texts, such as automatic text readability assessment (Vajjala and Meurers, 2012), Second Language Acquisition (SLA) research (Chen et al., 2021), automatic analysis of language production in aphasia (Gleichgerrcht et al., 2021), automatic Primary Progressive Aphasia (PPA) subtyping (Fraser et al., 2014) and automatic Alzheimer's Disease identification (Fraser et al., 2016). Subordinate/all clauses ratio represents the ratio of all subordinate clauses (complement, adverbial, relative clauses) to all clauses produced in the same narrative, including subordinate and main clauses. Mean dependency tree height measures the height of the dependency tree

TABLE 4 Linguistic measures for analyzing discourse production and their correspondence to language abilities (*a.–h.* adapted from Stark, 2019).

Measure	Definition	Language ability
a. Mean length of utterance (MLU)	Average number of words per utterance (excluding repetitions and reformulations)	Linguistic productivity
b. Propositional density	Number of verbs, adjectives, adverbs, prepositions, and conjunctions divided by the total number of words	Content richness
c. Words per minute	Total number of tokens divided by total story duration	Fluency
d. Verbs per utterance	Average number of verbs per utterance	Syntactic complexity
e. Lemma-token ratio	Number of lemmas divided by the total number of words (tokens)	Lexical diversity
f. Open-closed class words ratio	Ratio of open class words (nouns, verbs, adjectives, adverbs) divided by closed class words (all other classes)	Syntactic complexity
g. Noun-verb ratio	Ratio of nouns to verbs	Syntactic complexity
h. Number of words (tokens)	Total number of words produced	Gross output
i. Subordinate/all clauses ratio	Number of subordinate clauses (complement, adverbial, relative clauses) divided by the number of all clauses produced	Syntactic complexity
j. Mean dependency tree height	Height of the dependency tree (syntax tree)	Syntactic complexity

(corresponding to the syntax tree). The higher the dependency tree, the more complex the syntactic structure.

Table 4 presents the linguistic measures applied in the present study for the analysis of narrative discourse production and their correspondence to language abilities.

All these measures were automatically calculated using the tokenization, lemmatization, POS tagging and dependency parsing modules of the Neural NLP tool for Greek. More specifically, the measures of mean dependency tree height and subordinate clauses ratio were calculated using the dependency parser module. Since the dependency parser analyzes sentences, which consist of multiple main clauses -together with their subordinate clauses- connected with coordinating conjunctions, manual post-processing of transcripts was carried out, in order to convert utterances into sentences. As part of this process, several utterances were merged into a single sentence, using mainly intonational criteria, as shown in the following example:

> and then the witch found a pumpkin from the garden.
> she gave it a hit.
> and she transformed it into a carriage.
> and the two mice that accompanied Cinderella, she transformed them into two very nice horses.

Each one of the above lines corresponds to a separate utterance of the transcript. At post-processing, following the speaker's intonation contour, the four utterances were merged into a single sentence, beginning with a capital letter, and ending in a full stop or a question mark:

> And then the witch found a pumpkin from the garden, she gave it a hit, and she transformed it into a carriage, and the two mice that accompanied Cinderella, she transformed them into two very nice horses.

Post-processing was carried out by a single researcher, to ensure uniformity of sentence segmentation. Extensive evaluation of the automatically computed feature values was performed by two researchers, which led to script modifications, until minimal calculation errors were identified. For example, there were cases of passive participles tagged as verbs by the POS tagger, on the basis of their morphological characteristics, but actually had the syntactic role of an adjectival modifier (example 1) or a nominal subject (example 2) or object (example 3). The script written to calculate POS from the POS tagger output was modified to assign the POS tag "adjective" (example 1) and "noun" (examples 2–3) to the respective words.

(1) *Η καημένη* (POS: VERB | VerbForm: Part | syntactic role: amod) *η Σταχτοπούτα τα έβλεπε όλα αυτά.* (*Poor Cinderella was watching all this.*)

(2) *Στη γιορτή αυτή υπάρχουν διάφοροι καλεσμένοι.* (POS: VERB | VerbForm=Part | syntactic role: nsubj) (*In this party there are several guests.*)

(3) *Κι έτσι λοιπόν ο πρίγκιπας βρίσκει την αγαπημένη* (POS: VERB | VerbForm=Part | syntactic role: obj) *του.* (*And so, the prince finds his beloved.*)

4. Results

As a first step of analysis, correlations between the linguistic measures presented in Section 3.4 were calculated, to explore whether the selected measures contribute significantly in describing the language abilities of the participants at the discourse level. Linguistic variables were averaged across all narrative tasks and across the two elicitation conditions, in-person and remote. Results are presented in Table 5.

The results indicate some strong and expected correlations between certain linguistic variables. MLU correlates with three variables of syntactic complexity, i.e., verbs per utterance, ratio of dependent to all clauses, and mean tree height, indicating the interdependence of the utterance length with the structural complexity at utterance and sentence level. Accordingly, a strong correlation is found between the aforementioned variables of syntactic complexity, indicating that there is a close two-way relationship between the number of verbs in an utterance, the density of subordination at inter-clausal level, and the overall structural complexity at sentence level, as represented by the syntax tree. These results reveal some predictable but meaningful interactions between variables of linguistic productivity (MLU) and syntactic complexity (VPU, MTH, and Dep/Cl).

TABLE 5 Correlations between linguistic variables of narrative discourse production across elicitation condition.

	MLU	WPM	VPU	TTR	NVR	NoW	OCCW	Dep/Cl	MTH	PrD
MLU	1									
WPM	−0.470	1								
VPU	**0.981**	−0.412	1							
LTR	0.487	−0.477	0.410	1						
NVR	0.656	−0.642	0.538	0.762	1					
NoW	−0.195	0.382	−0.110	−0.710	−0.461	1				
OCCW	−0.343	0.078	−0.267	−0.165	−0.269	0.312	1			
Dep/Cl	**0.952**	−0.486	**0.958**	0.541	0.676	−0.167	−0.192	1		
MTH	**0.987**	−0.496	**0.968**	0.496	0.692	−0.169	−0.335	**0.952**	1	
PrD	−0.581	0.545	−0.519	−0.829	−0.890	0.551	0.243	−0.662	−0.638	1

MLU, mean length per utterance; WPM, words per minute; VPU, verbs per utterance; LTR, lemma-token ratio; NVR, noun-verb ratio; NoW, number of words; OCCWR, open-closed class words ratio; Dep/Cl, dependent clauses divided by total number of clauses; MTH, mean tree height; PrD, propositional density. Values in bold indicate strong correlations (> ±0.95).

We subsequently analyzed the linguistic variables under investigation across the two experimental settings, and across the seven narrative tasks. As described in Section 3.2, the two experimental groups consisted of the same participants. Therefore, these two samples are considered as dependent, which entails paired measurements of the same participant. Subjects within each group are independent. Given the small sample size (<30), it is hard to test the sample data for normality. Therefore, the non-parametric equivalent of the paired *t*-test, the Wilcoxon signed-rank test, was conducted. Supplementary Table 1 presents the results of the dependent-sample Wilcoxon test conducted for each one of the ten variables under investigation per narrative, making a total of 70 tests. Formally, the test hypothesis is formulated as follows:

(1) Null hypothesis (H0): The difference between the pairs follows a symmetric distribution around zero.
(2) Alternative hypothesis (HA): The difference between the pairs does not follow a symmetric distribution around zero.

The results did not reveal any statistically significant differences (*p*-value > 0.05) for 66 out of the 70 tests. The fact that no statistically significant differences were demonstrated for the vast majority of the linguistic variables analyzed per narrative task suggests that the administration condition did not have a significant effect on the linguistic properties of the spoken narratives produced by the study participants. In four cases, statistically significant mean differences emerged, i.e., in MLU, Verbs per Utterance and Mean Tree Height for *Cinderella,* and in Noun-Verb ratio for *the Ring* task. In the case of the *Cinderella* task, the means of MLU, VPU and MTH were significantly higher in the remote elicitation condition, as was the mean of NVR in the case of the *Ring*. All these measures correspond to features of syntactic complexity, while three of them, MLU, Verbs per Utterance and Mean Tree Height, were strongly correlated variables (Table 5). The *Cinderella* task is found to be one of the most sensitive discourse elicitation tasks to measure propositional density and syntactic complexity in spoken narratives of both PWA and unimpaired controls (Stark, 2019). Although this may not directly serve as an explanation of the rejection of the null hypothesis for three syntactic complexity

variables, further investigation in a larger sample of participants is needed to explore whether the remote condition elicits more syntactically complex language, at least in the case of *Cinderella*, which has been demonstrated as a good elicitation task to measure syntactic complexity.

5. Discussion

This paper presents the design, methodological considerations and requirements of a web-based method designed to facilitate spoken narrative discourse data collection from Greek-speaking PWA and unimpaired adults in remote as well as in in-person administration conditions for online assessment purposes. The application comprises a 7-task protocol for narrative discourse elicitation, guidelines to ensure reliable elicitation of narrative discourse type, appropriate sampling of participants' spoken discourse, and consistent administration across investigators and settings. Transcription procedures of speech samples, as well as procedures for the automated analysis of transcripts with available NLP tools for Greek texts, enabling the calculation of linguistic features that can serve as indicators of language abilities are also described.

As a first step for validating the presented method, a pilot study was conducted including only unimpaired adults and comparative results of narrative discourse produced in two different conditions of data collection, i.e., in-person and remote setting, are presented. The aim of the study was to investigate the feasibility of the presented method to elicit spoken narratives in remote collection settings that are comparable to the ones produced in in-person settings, in terms of specific characteristics of language production at the discourse level. Spoken narratives were collected with the use of the presented application in both conditions, using a within-subjects experimental design.

A set of ten linguistic variables representing various language abilities at the spoken discourse level, i.e., linguistic productivity, content richness, fluency, syntactic complexity at utterance and sentence level, lexical diversity, and verbal output, was selected to quantify spoken language production of participants in both

experimental conditions. Statistical analysis for each variable per narrative task indicated non-significant differences for most of the paired samples mean difference measurements, a finding which indicates the efficacy of the presented method to collect spoken narrative discourse samples from neurotypical adults with similar linguistic properties in both conditions.

A major limitation of the presented study is related to the small number of participants. Even though the statistical hypotheses were tested on a sufficient dataset of speech samples per participant, in terms of task variety, time duration and number of tokens (Brookshire and Nicholas, 1994; Boles and Bombard, 1998), the small sample size compromises the generalizability of results to a broader population of neurotypical adults.

Moreover, the fact that the method's validation included only unimpaired controls, and not PWA, does not allow considering the present study results generalizable to the population of PWA. PWA represent a vulnerable population that often suffers from coexisting chronic physical dysfunctions, cognitive impairments as well as negative social and emotional outcomes, such as depression and low social participation (Kauhanen et al., 2000; Mayo et al., 2002; Hilari et al., 2015). These conditions might affect PWA's ability to use technology-enhanced assessment environments remotely, without in-person supervision by an investigator. Despite the fact than none of the study participants needed help in navigating through the application or in setting up the teleconference, replicating the study with PWA, or even with unimpaired participants of lower educational level, might reveal issues related to the independent use of technology that did not occur in the present study.

Therefore, the planned next step for validating the presented method is to conduct the same study on participants with aphasia, to demonstrate its feasibility in collecting comparable data from adults with language impairments in remote and in-person assessment settings. This study can include participants of different aphasia severity levels, with the aim to identify possible differences in linguistic variables of spoken discourse across elicitation conditions and to investigate the effect of aphasia severity on these differences. Moreover, evaluation of the selected linguistic variables can identify effects of narrative tasks on language production properties in each population, PWA and neurotypical adults, and in comparison with each other, which can be further explored across elicitation condition, i.e., remote and in-person. A complementary area of future research is to investigate the contribution of additional linguistic variables, such as features related to informational content and narrative macrostructure, in the description of language abilities of both PWA and unimpaired adult populations, as well as the effect of elicitation condition on these variables.

Future work will also involve replication of the study in a larger sample of unimpaired adults of different age groups and educational levels, with the aim to investigate the impact of these demographic variables on different elicitation conditions. Age and educational level might have an effect on participants' performance in the remote elicitation condition, which is heavily related on technology skills, so future research could contribute to testing this hypothesis.

Given the above limitations, our findings are aligned with prior studies (Palsbo, 2007; Theodoros et al., 2008; Hill et al., 2009; Dekhtyar et al., 2020) which compare online assessment methods addressed to individuals with language and communication disorders in remote and in-person conditions, suggesting that both settings produce comparable results in terms of language production. Concerns raised regarding participants' technology skills need to be considered for the method's effective implementation, by adding special instructions for participants and troubleshooting guidelines for investigators in case of remote elicitation of spoken discourse from either impaired or unimpaired individuals. Instructions can include issues such as setting up a teleconference, screen sharing, microphone muting and unmuting, as well as navigating through the application. However, it is worth noting that the use of the presented web-based application is still feasible even in case of participants with limited or no technology skills, with the use of a remote desktop software by the investigator, which will allow full control of the participant's computer.

The presented web-based data collection method is currently being employed for collecting spoken language data from PWA in-person and from unimpaired individuals, either remotely or in-person. This dataset (speech samples, transcripts, features measured and labels for aphasia class) serves as a golden corpus, which provides the ground truth, on the basis of which a machine-learning system for the automatic classification of aphasia in Greek will be assessed and evaluated. A substantial amount of manual work is carried out for compiling this corpus; manual transcription and time-alignment, utterance segmentation and sentence splitting. Since a large amount of data is required to train accurate linguistic models for automatic classification purposes, ongoing research activities are being carried out that aim to automate manual work involved in the data processing and analysis pipeline (Chatzoudis et al., 2022) for aphasia classification purposes, such as Automatic Speech Recognition in Greek for transcription and time alignment.

In sum, the presented method, as evidenced from the present study findings, offers an applicable, feasible and valid framework for both in-person and remote online elicitation of spoken narrative discourse samples for the assessment of language abilities of adult populations without language disorders. The next step will be to investigate its feasibility to collect comparable spoken discourse data from adults with language disorders in remote and FTF settings. In line with the current literature on language and communication disorders assessment and intervention, which highlights the need for the modification of conventional pen-and-paper methods to accommodate technology-enhanced tools and applications, the present paper provides some initial evidence toward the reliable implementation of technology applications for remote language data collection and assessment of language skills.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The study involving human participants was reviewed and approved by the Bioethics Committee, University of Patras, Patras, Greece. The participants provided their written informed consent to participate in this study.

Author contributions

SS, AK, and VK: conceptualization. SS, MN, IP, A-LD, AK, SV, and AT: methodology. GC, MP, AK, and VK: software. SS, MN, and A-LD: data collection, transcription, and annotation. SS, GC, and MP: data processing and analysis. SS, MN, and IP: writing. SS, MN, IP, A-LD, SV, and AT: review and editing. AT: scientific coordination. IP: project administration. All authors have read and agreed to the published version of the manuscript.

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomm.2023.919617/full#supplementary-material

SUPPLEMENTARY TABLE 1
Measures of linguistic productivity, content richness, fluency, syntactic complexity, lexical diversity and verbal output, statistically compared between in-person and remotely elicited speech samples using the non-parametric Wilcoxon signed-rank test for comparison of means. *P*-values are listed, with an asterisk (*) indicating statistical significance.

References

American Speech-Language-Hearing Association (2016). *2016 SIG 18 Telepractice Survey Results*. Available online at: https://www.asha.org/siteassets/practice-portal/telepractice/2016-telepractice-survey.pdf (accessed April 13, 2022).

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

Armstrong, E., Ferguson, A., and Simmons-Mackie, N. (2011). "Discourse and functional approaches to aphasia," in *Aphasia and Related Neurogenic Communication Disorders*, eds I. Papathanasiou, P. Coppens, and C. Potagas (Sudbury, MA: Jones and Bartlett Pub), 217–231.

Armstrong, E., and Ulatowska, H. K. (2007). "Stroke stories: conveying emotive experiences in aphasia," in *Clinical aphasiology: Future Directions*, eds. M. J. Ball, and J. S. Damico (New York, NY: Psychology Press).

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

Boles, L., and Bombard, T. (1998). Conversational discourse analysis: appropriate and useful sample sizes. *Aphasiology* 12, 547–560. doi: 10.1080/02687039808249557

Bottenberg, D., Lemme, M., and Hedberg, N. (1985). "Analysis of oral narratives of normal and aphasic adults," in *Clinical Aphasiology*, ed R. H. Brookshire (Minneapolis, MN: BRK), 241–247.

Brennan, D. M., Georgeadis, A. C., Baron, C. R., and Barker, L. M. (2004). The effect of videoconference based telerehabilitation on story retelling performance by brain-injured subjects and its implications for remote speech-language therapy. *Telemed. J. e- Health* 10, 147–154. doi: 10.1089/tmj.2004.10.147

Brookshire, R., and Nicholas, L. (1994). Speech sample-size and test-retest stability of connected speech measures for adults with aphasia. *J. Speech Lang. Hear. Res.* 37, 399–407. doi: 10.1044/jshr.3702.399

Bryant, L., Ferguson, A., and Spencer, E. (2016). Linguistic analysis of discourse in aphasia: a review of the literature. *Clin. Linguist. Phonet.* 30, 489–518. doi: 10.3109/02699206.2016.1145740

Bryant, L., Spencer, E., and Ferguson, A. (2017). Clinical use of linguistic discourse analysis for the assessment of language in aphasia. *Aphasiology* 31, 1105–1126. doi: 10.1080/02687038.2016.1239013

Chatzoudis, G., Plitsis, M., Stamouli, S., Dimou, A.-L., Katsamanis, A., and Katsouros, V. (2022). "Zero-shot cross-lingual aphasia detection using automatic speech recognition," in *INTERSPEECH 2022* (Incheon).

Chen, X., Alexopoulou, T., and Tsimpli, I. (2021). Automatic extraction of subordinate clauses and its application in second language acquisition research. *Behav. Res. Methods* 53, 803–817. doi: 10.3758/s13428-020-01456-7

Choi, Y. H., Park, H. K., Ahn, K. H., Son, Y. J., and Paik, N. J. (2015). A Telescreening tool to detect aphasia in patients with stroke. *Telemed. J. e- Health* 21, 729–734. doi: 10.1089/tmj.2014.0207

Cruice, M., Botting, N., Marshall, J., Boyle, M., Hersh, D., Pritchard, M., et al. (2020). UK speech and language therapists' views and reported practices of discourse analysis in aphasia rehabilitation. *Int. J. Lang. Commun. Disord.* 55, 417–442. doi: 10.1111/1460-6984.12528

Dekhtyar, M., Braun, E. J., Billot, A., Foo, L., and Kiran, S. (2020). Videoconference administration of the western aphasia battery-revised: feasibility and validity. *Am. J. Speech Lang. Pathol.* 29, 673–687. doi: 10.1044/2019_AJSLP-19-00023

Dietz, A., and Boyle, M. (2018). Discourse measurement in aphasia research: have we reached the tipping point? *Aphasiology* 32, 459–464. doi: 10.1080/02687038.2017.1398803

Doedens, W. J., and Meteyard, L. (2020). Measures of functional, real-world communication for aphasia: a critical review. *Aphasiology* 34, 492–514. doi: 10.1080/02687038.2019.1702848

Doedens, W. J., and Meteyard, L. (2022). What is functional communication? A theoretical framework for real-world communication applied to aphasia rehabilitation. *Neuropsychol. Rev.* 32, 937–973. doi: 10.1007/s11065-021-09531-2

Doyle, P. J., Goda, A. J., and Spencer, K. A. (1995). The communicative informativeness and efficiency of connected discourse by adults with aphasia under structured and conversational sampling conditions. *Am. J. Speech Lang. Pathol.* 4, 130–134. doi: 10.1044/1058-0360.0404.130

Doyle, P. J., McNeil, M. R., Park, G. H., Goda, A. J., Spencer, K., Lustig, A., et al. (2000). Linguistic validation of four parallel forms of a story retelling procedure. *Aphasiology* 14, 537–549. doi: 10.1080/026870300401306

Doyle, P. J., McNeil, M. R., Spencer, K. A., Goda, A. J., Cottrell, K., and Lustig, A. P. (1998). The effects of concurrent picture presentations on retelling of orally presented stories by adults with aphasia. *Aphasiology* 12, 561–574. doi: 10.1080/02687039808249558

Ellis, C., and Urban, S. (2016). Age and aphasia: a review of presence, type, recovery and clinical outcomes. *Top. Stroke Rehabil.* 23, 430–439. doi: 10.1080/10749357.2016.1150412

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

Fergadiotis, G., Wright, H. H., and West, T. M. (2013). Measuring lexical diversity in narrative discourse of people with aphasia. *Am. J. Speech Lang. Pathol.* 22, S397–S408. doi: 10.1044/1058-0360(2013/12-0083)

Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). Mini-mental state: a practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* 12,189–198. doi: 10.1016/0022-3956(75)90026-6

Fraser, K. C., Meltzer, J. A., Graham, N. L., Leonard, C., Hirst, G., Black, S. E., et al. (2014). Automated classification of primary progressive aphasia subtypes from narrative speech transcripts. *Cortex* 55, 43–60. doi: 10.1016/j.cortex.2012.12.006

Fraser, K. C., Meltzer, J. A., and Rudzicz, F. (2016). Linguistic features identify Alzheimer's disease in narrative speech. *J. Alzheimers Dis.* 49, 407–422. doi: 10.3233/JAD-150520

Fromm, D., Greenhouse, J., Pudil, M., Shi, Y., and MacWhinney, B. (2022). Enhancing the classification of aphasia: a statistical analysis using connected speech. *Aphasiology.* 36, 1492–1519. doi: 10.1080/02687038.2021.1975636

Fromm, D., Holland, A., Armstrong, E., Forbes, M., MacWhinney, B., Risko, A., et al. (2011). "Better but no cigar": persons with aphasia speak about their speech. *Aphasiology* 25, 1431–1447. doi: 10.1080/02687038.2011.608839

Georgeadis, A., Brennan, D., Barker, L., and Baron, C. (2004). Telerehabilitation and its effect on story retelling by adults with neurogenic communication disorders. *Aphasiology* 18, 639–652. doi: 10.1080/02687030444000075

Gleichgerrcht, E., Roth, R., Fridriksson, J., den Ouden, D., Delgaizo, J., Stark, B., et al. (2021). Neural bases of elements of syntax during speech production in patients with aphasia. *Brain Lang.* 222, 105025. doi: 10.1016/j.bandl.2021.105025

Glosser, G., Wiener, M., and Kaplan, E. (1988). Variations in aphasic language behaviors. *J. Speech Hear. Disord.* 53, 115–124. doi: 10.1044/jshd.5302.115

González-Fernández, M., Davis, C., Molitoris, J. J., Newhart, M., Leigh, R., and Hillis, A. E. (2011). Formal education, socioeconomic status, and the severity of aphasia after stroke. *Arch. Phys. Med. Rehabil.* 92, 1809–1813. doi: 10.1016/j.apmr.2011.05.026

Goodglass, H., and Kaplan, E. (1972). *Boston Diagnostic Aphasia Examination (BDAE)*. Philadelphia, PA: Lea and Febiger.

Goodglass, H., Kaplan, E., and Barresi, B. (2001). *Boston Diagnostic Aphasia Examination, 3rd Edn*. Philadelphia, PA: Lippincott Williams and Wilkins.

Granger, S., and Wynne, M. (1999). "Optimising measures of lexical variation in EFL learner corpora," in *Corpora Galore*, ed J. Kirk (Amsterdam; Atlanta: Rodopi), 249–257

Grimes, N. (2005). *Walt Disney's Cinderella*. New York, NY: Random House.

Guo, Y. E., Togher, L., Power, E., Hutomo, E., Yang, Y. F., Tay, A., et al. (2017). Assessment of aphasia across the international classification of functioning, disability and health using an iPad-based application. *Telemed. J. e- Health* 23, 313–326. doi: 10.1089/tmj.2016.0072

Ha, J. W., Pyun, S.-B., Lee, H. Y., Hwang, Y.-M., and Nam, K. (2009). Reliability and validity analyses of the Korean version of Frenchay Aphasia Screening Test in brain-damaged patients. *Kor. J. Commun. Disord.* 14, 46–57. Available online at: http://www.e-csd.org/journal/view.php?number=382

Harley, T. (2001). *The Psychology of Language: From Data to Theory*, *2nd Edn*. New York, NY: Psychology Press.

Hilari, K., Cruice, M., Sorin-Peters, R., and Worrall, L. (2015). Quality of life in aphasia: state of the art. *Folia Phoniatr. Logopaed.* 67, 114–118. doi: 10.1159/000440997

Hill, A. J., Theodoros, D. G., Russell, T. G., Ward, E. C., and Wootton, R. (2009). The effects of aphasia severity on the ability to assess language disorders via telerehabilitation. *Aphasiology* 23, 627–642. doi: 10.1080/02687030801909659

Holland, A. (1979). "Some practical considerations in aphasia rehabilitation," in *Rationale for Adult Aphasia Therapy*, eds. M. Sullivan, and M. Kommers (Omaha, NE: University of Nebraska Medical Center), 167–180.

Holland, A. (1982). Observing functional communication of aphasic adults. *J. Speech Hear. Disord.* 47, 50–56. doi: 10.1044/jshd.4701.50

Holler, J., and Wilkin, K. (2009). Communicating common ground: How mutually shared knowledge influences speech and gesture in a narrative task. *Lang. Cogn. Process.* 24, 267–289. doi: 10.1080/01690960802095545

Kakavoulia, M., Stamouli, S., Foka-Kavalieraki, P., Economou, A., Protopapas, A., and Varlokosta, S. (2014). A battery for eliciting narrative discourse by Greek speakers with aphasia: Principles, methodological issues, and preliminary results. *Glossologia* 22, 41–60 (in Greek). Available online at: http://glossologia.phil.uoa.gr/?q=node/107

Kauhanen, M. L., Korpelainen, J. T., Hiltunen, P., Määtt,ä, R., Mononen, H., Brusin, E., et al. (2000). Aphasia, depression, and non-verbal cognitive impairment in ischaemic stroke. *Cerebrovasc. Dis.* 10, 455–461. doi: 10.1159/000016107

Kertesz, A. (2006). *Western Aphasia Battery, Revised Edn*. San Antonio, TX: Pearson Assessment.

Labov, W. (1972). *Language in the Inner City: Studies in the Black English Vernacular*. Philadelphi, PA: University of Pennsylvania Press.

Lemme, M. L., Hedberg, N. L., and Bottenberg, D. F. (1984). "Cohesion in narratives of aphasic adults," in *Clinical Aphasiology*, ed. R. H. Brookshire (Minneapolis, MN: BRK), 215–222.

MacWhinney, B., and Fromm, D. (2016). AphasiaBank as BigData. *Semin. Speech Lang.* 37, 10–22. doi: 10.1055/s-0036-1571357

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

Mayo, N. E., Wood-Dauphinee, S., Côté, R., Durcan, L., and Carlton, J. (2002). Activity, participation, and quality of life 6 months poststroke. *Arch. Phys. Med. Rehabil.* 83, 1035–1042. doi: 10.1053/apmr.2002.33984

McNeil, M., Doyle, P., Fossett, T., Park, G., and Goda, A. (2001). Reliability and concurrent validity of the information unit scoring metric for the story retelling procedure. *Aphasiology* 15, 991–1006. doi: 10.1080/0268704014300 0348

McNeil, M. R., Doyle, P. J., Park, G. H., Fossett, T. R. D., and Brodsky, M. B. (2002). Increasing the sensitivity of the Story Retell Procedure for the discrimination of normal elderly subjects from persons with aphasia. *Aphasiology.*16, 815–822. doi: 10.1080/02687030244000284

McNeil, M. R., Sung, J. E., Yang, D., Pratt, S. R., Fossett, T. R. D., Doyle, P. J., et al. (2007). Comparing connected language elicitation procedures in persons with aphasia: concurrent validation of the Story Retell Procedure. *Aphasiology* 21, 775–790. doi: 10.1080/02687030701189980

Mesulam, M. M. (2001). Primary progressive aphasia. *Ann. Neurol.* 49, 425–432. doi: 10.1002/ana.91

Newton, C., Acres, K., and Bruce, C. (2013). A comparison of computerized and paper-based language tests with adults with aphasia. *Am. J. Speech Lang. Pathol.* 22, 185–197. doi: 10.1044/1058-0360(2012/12-0027)

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

Noyes, J. M., and Garland, K. J. (2008). Computer- vs. paper-based tasks: are they equivalent? *Ergonomics.* 51, 1352–1375. doi: 10.1080/00140130802170387

Obler, L. K., and Gjerlow, K. (1999). *Language and the Brain*. Cambridge: Cambridge University Press.

Olness, G. S. (2006). Genre, verb, and coherence in picture-elicited discourse of adults with aphasia. *Aphasiology* 20, 175–187. doi: 10.1080/02687030500472710

Olness, G. S., and Englebretson, E. F. (2011). On the coherence of information highlighted by narrators with aphasia. *Aphasiology* 25, 713–726. doi: 10.1080/02687038.2010.537346

Olness, G. S., Matteson, S. E., and Stewart, C. T. (2010). "Let me tell you the point": how speakers with aphasia assign prominence to information in narratives. *Aphasiology* 24, 697–708. doi: 10.1080/02687030903438524

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

Palsbo, S. E. (2007). Equivalence of functional communication assessment in speech pathology using videoconferencing. *J. Telemed. Telecare* 13, 40–43. doi: 10.1258/135763307779701121

Papageorgiou, S. G., Economou, A., and Routsis, C. (2014). The 5 objects test: a novel, minimal-language, memory screening test. *J. Neurol.* 261, 422–431. doi: 10.1007/s00415-013-7219-1

Pritchard, M., Hilari, K., Cocks, N., and Dipper, L. (2018). Psychometric properties of discourse measures in aphasia: acceptability, reliability, and validity. *Int. J. Lang. Commun. Disord.* 53, 1078–1093. doi: 10.1111/1460-6984.12420

Prokopidis, P., and Piperidis, S. (2020). "A neural NLP toolkit for Greek," in *SETN 2020: 11th Hellenic Conference on Artificial Intelligence. September 2020* (Athens), 125–128.

Radanovic, M., Mansur, L. L., and Scaff, M. (2004). Normative data for the Brazilian population in the Boston diagnostic aphasia examination: Influence of schooling. *Braz. J. Med. Biol. Res.* 37, 1731–1738. doi: 10.1590/S0100-879X2004001100019

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

Stark, B. C. (2019). A comparison of three discourse elicitation methods in aphasia and age-matched adults: implications for language assessment and outcome. *Am. J. Speech Lang. Pathol.* 28, 1067–1083. doi: 10.1044/2019_AJSLP-18-0265

Stark, B. C., Dutta, M., Murray, L. L., Bryant, L., Fromm, D., MacWhinney, B., et al. (2021b). Standardizing assessment of spoken discourse in aphasia: a working group with deliverables. *Am. J. Speech Lang. Pathol.* 11, 491–502. doi: 10.1044/2020_AJSLP-19-00093

Stark, B. C., Dutta, M., Murray, L. L., Fromm, D., Bryant, L., Harmon, T. G., et al. (2021a). Spoken discourse assessment and analysis in aphasia: An international survey of current practices. *J. Speech Lang. Hear. Res.* 64, 4366–4389. doi: 10.1044/2021_JSLHR-20-00708

Stark, B. C., and Fukuyama, J. (2021). Leveraging big data to understand the interaction of task and language during monologic spoken discourse in speakers with and without aphasia. *Lang. Cognit. Neurosci.* 36, 562–585. doi: 10.1080/23273798.2020.1862258

Theodoros, D., Hill, A., Russell, T., Ward, E., and Wootton, R. (2008). Assessing acquired language disorders in adults via the Internet. *Telemed. J. e- Health* 14, 552–559. doi: 10.1089/tmj.2007.0091

UNESCO-UIS (2012). *International Standard Classification of Education-ISCED 2011.* UNESCO-UIS Institute for Statistics. Available online at: http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf (accessed April 6, 2022).

Vajjala, S., and Meurers, D. (2012). "On improving the accuracy of readability classification using insights from second language acquisition," in *Proceedings of the Seventh Workshop on Building Educational Applications Using NLP, 163-173.* Montreal, QC: Association of Computational Linguistics. Available online at: https://aclanthology.org/W12-2019.pdf (accessed June 22, 2022).

Varlokosta, S., Stamouli, S., Karasimos, A., Markopoulos, G., Kakavoulia, M., Nerantzini, M., et al. (2016). "A greek corpus of aphasic discourse: collection, transcription, and annotation specifications," in *Proceedings of LREC 2016 Workshop. Resources and Processing of Linguistic and Extra-Linguistic Data from People with Various Forms of Cognitive/Psychiatric Impairments (RaPID-2016), Monday 23rd of May 2016,* 14–21. Available online at: https://ep.liu.se/ecp/128/003/ecp16128003.pdf (accessed April 13, 2022).

Wallace, S. J., Worrall, L., Rose, T., Le Dorze, G., Breitenstein, C., Hilari, K., et al. (2019). A core outcome set for aphasia treatment research: The ROMA consensus statement. *Int. J. Stroke* 14, 180–185. doi: 10.1177/1747493018806200

World Health Organization (2001). *International Classification of Functioning, Stability and Health.* ICF. Available online at: https://apps.who.int/iris/handle/10665/42407 (accessed April 6, 2022).

Wright, H. H. (2011). Discourse in aphasia: an introduction to current research and future directions. *Aphasiology* 25, 1283–1285. doi: 10.1080/02687038.2011.613452

Wright, H. H., and Capilouto, G. J. (2009). Manipulating task instructions to change narrative discourse performance. *Aphasiology* 23, 1295–1308. doi: 10.1080/02687030902826844