University of Groningen

Language in individuals with left hemisphere tumors

Rofes, Adrià; Talacchi, Andrea; Santini, Barbara; Pinna, Giampietro; Nickels, Lyndsey; Bastiaanse, Roelien; Miceli, Gabriele

Published in:
Journal of Clinical and Experimental Neuropsychology

DOI:
[10.1080/13803395.2018.1426734](https://doi.org/10.1080/13803395.2018.1426734)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2018

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Rofes, A., Talacchi, A., Santini, B., Pinna, G., Nickels, L., Bastiaanse, R., & Miceli, G. (2018). Language in individuals with left hemisphere tumors: Is spontaneous speech analysis comparable to formal testing? *Journal of Clinical and Experimental Neuropsychology*, 40(7), 722-732.
<https://doi.org/10.1080/13803395.2018.1426734>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Language in individuals with left hemisphere tumors: Is spontaneous speech analysis comparable to formal testing?

Adrià Rofes, Andrea Talacchi, Barbara Santini, Giampietro Pinna, Lyndsey Nickels, Roelien Bastiaanse & Gabriele Miceli

To cite this article: Adrià Rofes, Andrea Talacchi, Barbara Santini, Giampietro Pinna, Lyndsey Nickels, Roelien Bastiaanse & Gabriele Miceli (2018) Language in individuals with left hemisphere tumors: Is spontaneous speech analysis comparable to formal testing?, Journal of Clinical and Experimental Neuropsychology, 40:7, 722-732, DOI: [10.1080/13803395.2018.1426734](https://doi.org/10.1080/13803395.2018.1426734)

To link to this article: <https://doi.org/10.1080/13803395.2018.1426734>

View supplementary material [↗](#)

Published online: 31 Jan 2018.

Submit your article to this journal [↗](#)

Article views: 184

View Crossmark data [↗](#)

Language in individuals with left hemisphere tumors: Is spontaneous speech analysis comparable to formal testing?

Adrià Rofes^{a,b}, Andrea Talacchi^c, Barbara Santini^c, Giampietro Pinna^d, Lyndsey Nickels^e, Roelien Bastiaanse^f and Gabriele Miceli^g

^aGlobal Brain Health Institute, Trinity College Dublin, Dublin, Ireland; ^bDepartment of Cognitive Science, Johns Hopkins University, Baltimore, MD, USA; ^cSection of Neurosurgery, Department of Neurosciences, University of Verona, Verona, Italy; ^dDepartment of Neurosurgery, University Hospital, Verona, Verona, Italy; ^eARC Center of Excellence in Cognition and its Disorders, Department of Cognitive Science, Macquarie University, Sydney, Australia; ^fCenter for Language and Cognition (CLCG), University of Groningen, Groningen, The Netherlands; ^gCenter for Mind/Brain Sciences (CIMEC), University of Trento, Trento, Italy

ABSTRACT

Background: The relationship between spontaneous speech and formal language testing in people with brain tumors (gliomas) has been rarely studied. In clinical practice, formal testing is typically used, while spontaneous speech is less often evaluated quantitatively. However, spontaneous speech is quicker to sample and may be less prone to test/retest effects, making it a potential candidate for assessing language impairments when there is restricted time or when the patient is unable to undertake prolonged testing.

Aim: To assess whether quantitative spontaneous speech analysis and formal testing detect comparable language impairments in people with gliomas. Specifically, we addressed (a) whether both measures detected comparable language impairments in our patient sample; and (b) which language levels, assessment times, and spontaneous speech variables were more often impaired in this subject group.

Method: Five people with left perisylvian gliomas performed a spontaneous speech task and a formal language assessment. Tests were administered before surgery, within a week after surgery, and seven months after surgery. Performance on spontaneous speech was compared with that of 15 healthy speakers.

Results: Language impairments were detected more often with both measures than with either measure independently. Lexical-semantic impairments were more common than phonological and grammatical impairments, and performance was equally impaired across assessment time points. Incomplete sentences and phonological paraphasias were the most common error types.

Conclusions: In our sample both spontaneous speech analysis and formal testing detected comparable language impairments. Currently, we suggest that formal testing remains overall the better option, except for cases in which there are restrictions on testing time or the patient is too tired to undergo formal testing. In these cases, spontaneous speech may provide a viable alternative, particularly if automated analysis of spontaneous speech becomes more readily available in the future. These results await replication in a bigger sample and/or other populations.

ARTICLE HISTORY

Received 3 June 2017

Accepted 18 December 2017

KEYWORDS

Aphasia; brain tumor; glioma; language assessment; spontaneous speech

Quantitative analyses of spontaneous speech have been used to assess and to characterize language abilities in daily life in people with acquired neurological deficits (e.g., Bastiaanse & Jonkers, 1998; Miceli, Silveri, Romani, & Caramazza, 1989; Saffran, Berndt, & Schwartz, 1989). Here, we studied whether spontaneous speech analysis and formal testing are equally good measures to detect language impairments in people with left hemisphere tumors (gliomas). This is relevant because spontaneous speech has rarely been studied in this population (see Rofes et al., 2017; cf. Satoer, Vincent, Smits, Dirven, & Visch-Brink, 2013),

and also because these tasks could be an ecologically valid complement to formal testing and/or be used as a rapid assessment when there exist time restrictions or the patient is too tired to complete a full formal battery.

In a typical spontaneous speech task, patients are asked to respond to open-ended questions, to describe a picture, or to tell a story (for a review see Prins & Bastiaanse, 2004). There is a long tradition of spontaneous speech studies in people with aphasia due to stroke, including a thorough description of language impairments and recommendations for data acquisition and analysis (e.g., Gordon, 2006; Herbert,

Hickin, Howard, Osborne, & Best, 2008; Kasl & Mahl, 1965; Rochon, Saffran, Berndt, & Schwartz, 2000; Rossi & Bastiaanse, 2008; Van Lancker Sidtis & Postman, 2006; Vermeulen, Bastiaanse, & Van Wageningen, 1989; Wagenaar, Snow, & Prins, 1975; Wingfield, 1996; Yorkston & Beukelman, 1980). Some of the variables discussed in these studies have been revisited in descriptions of the spontaneous speech of people with different neurodegenerative disorders, including Alzheimer's, Parkinson's, and different types of primary progressive aphasia (e.g., Colman et al., 2009; Fraser, Rudzicz, Graham, & Rochon, 2013; Garrard, Rentoumi, Gesierich, Miller, & Gorno-Tempini, 2014; Kansal, Abraham, & Onyike, 2015; Wilson et al., 2010).

Little research in this area has involved people with brain tumors. Satoer et al. (2013) found that people with brain tumors produced shorter sentences than healthy speakers, in a spontaneous speech task at 2 weeks, 7 weeks, and 3 months after surgery. Interestingly, these authors found that the incidence of shorter than normal sentences was positively correlated with low scores on an object-naming task, and suggested that in this population shorter sentences may relate to syntactic or to lexical impairments. In a single-case study of a patient with a left hemisphere glioma, Satoer, Kloet, Vincent, Dirven, and Visch-Brink (2014) did not detect major impairments on a comprehensive language battery across assessment times. However, the patient was slower at responding, was less fluent, and produced more incomplete sentences and repetitions than a group of healthy speakers. The authors suggested that the slower rate of response corresponded to difficulties selecting a correct speech plan, whereas the high number of repetitions and incomplete sentences was the consequence of a lexical problem that was not measurable with the language battery.

Aims and hypotheses

Our overall aim was to assess whether spontaneous speech analysis and formal testing detect comparable language impairments in our sample of people with left hemisphere gliomas before surgery, and at two separate time points after surgery. Specifically, we asked two main questions: (a) Do spontaneous speech and formal neuropsychological testing detect comparable language impairments in our patient sample? and (b) which language levels (lexical-semantic, grammatical, phonological), assessment times (T1, T2, T3), and spontaneous speech variables (diversity of nouns, mean length of utterance, etc.) are more commonly impaired in our patient sample,

as opposed to in healthy speakers? To answer these questions, we asked individuals with a brain tumor to complete a formal language battery and a spontaneous speech interview. We used this information to evaluate whether spontaneous speech should be preferred to formal testing for routine clinical evaluation of language disorders in this population. This is a relevant question for the clinical setting because spontaneous speech is less time consuming to administer and may be less prone to test/retest effects than formal testing. Hence, it may be preferable when there are time restrictions, when the patient is too tired to respond, or when the formal test stimuli are administered in assessment sessions that are relatively close to one another.

We focused on features of speech production that require computational demands in spontaneous speech comparable to those that are needed for formal testing. As these features should pose similar difficulties in both conditions, their analysis should allow comparisons between the efficacy of spontaneous speech and formal testing in detecting language deficits. For example, lexical retrieval difficulties in spontaneous speech (e.g., low percentage/diversity of nouns and/or verbs) may co-occur with poor naming abilities in formal testing. Herbert et al. (2008) studied the conversation of people with poststroke aphasia and their partners, and found correlations between performance on a picture-naming task and various aspects of lexical retrieval in conversation, such as the proportion of nouns. They argued that the proportion of nouns and performance on a picture-naming task correlate, as the processes necessary for the spoken production of nouns are the same regardless of whether they are recruited for picture naming or in a narrative speech task (Caramazza, 1997; Levelt, 1989).

Herbert et al. (2008) also indicated that the proportion of nouns may not always correlate with fluency measures, as there may be patients with normal fluency scores and low proportion of nouns and vice versa. Along the same lines, Bastiaanse and Jonkers (1998) indicated that people with poststroke aphasia typically produce fewer and less diverse verbs than healthy controls in connected speech. They argued that this may be due to the fact that some of their patients had problems with grammatical processing, which may have led them to produce verbs with less complex grammatical structures (for example, intransitive verbs such as "to walk" that do not take more than one argument, and transitive verbs such as "to kiss," which take more than one argument). These authors also argued that "retrieving verbs in spontaneous speech is more complex than in

action naming test, because in spontaneous speech linguistic processing takes place at more than one level" (Bastiaanse & Jonkers, 1998, p. 965).

By the same token, syntactic deficits in spontaneous speech (e.g., reduced mean length of utterance or MLU, low rate of subordination, etc.) may co-occur with difficulties in formal tasks that require grammatical processing (e.g., sentence completion).

Additionally, features of spontaneous speech may change at different assessment points with relation to the time of surgery. Before surgery (T1), patients may present with no or mild impairments, as they are typically selected for awake surgery based on normal or near-normal performance on formal tasks (e.g., Bello et al., 2007). Within a week after surgery (T2), the typical time selected for post-operative evaluations, they may perform more poorly due to surgery-related issues (tissue resection is a traumatic event for the brain; adaptive or maladaptive brain reorganization may take place, which may modify mechanisms premorbidly involved in language processing). Seven months after surgery (T3), at follow-up assessments, patients may perform normally or show only mild disorders. Hypothesizing which specific language impairments may arise in our sample is out of the scope of this paper. These impairments have been shown to depend on the subjects' premorbid abilities, and lesion site and type, among other factors (e.g., Benzagmout, Gatignol, & Duffau, 2007; Ojemann, 1979; Rofes, Spena, Miozzo, Fontanella, & Miceli, 2015; Santini et al., 2012).

Method

Participants

Five Italian-speaking patients (3 female, 2 male) with a perisylvian tumor in the left hemisphere participated in this study. They ranged between 39 and 65 years in age and had received formal education for 8–17 years. Patient S.O. had already undergone

awake surgery at another center in 2005, for the same tumor. Individual characteristics and tumor variables are reported in Table 1. Four patients had a tumor in the prefrontal cortex, and one patient in the parietal lobe. Three patients had a low-grade glioma (i.e., Tumor Grade II), which is characterized by a slow growth and typically infiltrates subcortical pathways. The two other patients had a high-grade glioma (i.e., Tumor Grade III or IV), which has a fast growth rate and typically does not infiltrate subcortical pathways. All patients presented with no or only mild language impairments before surgery, as this is an inclusion criterion for patients to undergo awake surgery (e.g., Bello et al., 2007; Rofes et al., 2017).

The healthy speakers included 15 Italian-speaking volunteers (11 female) without history of neurological disorder, with a mean age of 55.7 years (range 33–67) and mean education of 13.8 years (range 8–17). The healthy and patient groups did not significantly differ in age ($U = 2.235$, $p = .135$, two-tailed) or education ($U = 0.14$, $p = .017$, two-tailed).

Materials and procedure

All patients were assessed before surgery (T1), within a week after surgery (T2), and seven months after surgery (T3). Healthy speakers were assessed only once. We examined variables in spontaneous speech and performed formal testing focused on lexical-semantic, grammatical, and phonological levels of processing. Additionally, measurements of fluency were analyzed, since these are known to be discriminative in aphasic spontaneous speech (Vermeulen et al., 1989). All the materials and variables used in the assessment protocol are listed in Table 2. A description of formal tasks and spontaneous speech is provided below. For a more detailed description, please see Supplemental Material.

Measures of nonverbal IQ (Raven Color Matrices; Carlesimo et al., 1996), attention (Trail Making Test, TMT A, B, B-A; Giovagnoli et al., 1996), inhibition (Stroop test; Caffarra, Vezzadini, Dieci, Zonato, &

Table 1. Patient characteristics and tumor characteristics.

	Age (years)	Gender	Edu.	Employment	Hand.	Tumor characteristics		
						Grade	Type	Site
A.A.	52	F	>17	Accounting	Right	II	Oligodendroglioma	MTG
C.R.	39	F	8	Housewife	Right	II	Oligodendroglioma	IPL
C.E.	59	M	>17	Geologist	Right ^a	IV	Glioblastoma	MTG, ITG
R.A.	65	F	>17	Sales (retired)	Right	III	Anaplastic astrocytoma	STG
S.O.	45	M	8	Carpenter	Right	II	Oligoastrocytoma	IFG, MFG

Note. F = female; M = male; Hand. = handedness; Edu. = education in years; IFG = inferior frontal gyrus; IPL = inferior parietal lobe; ITG = inferior temporal gyrus; MFG = middle frontal gyrus; MTG = middle temporal gyrus; STG = superior temporal gyrus; Grade = tumor grade (i.e., I–II are low-grade gliomas; III–IV are high-grade gliomas).

^aLeft-handed taught to write with the right hand.

Table 2. Spontaneous speech and formal testing measures administered.

	Spontaneous speech	Formal testing
Fluency/ speech rate	Number of words Total number of sentences Words per minute ^a	NA
Lexical- semantic	Retrieval of adjectives ^a Retrieval of verbs Retrieval of nouns Diversity for all lemmas Diversity for nouns ^a Diversity for verbs	Auditory lexical decision ^a Visual lexical decision Auditory noun comprehension ^a Visual noun comprehension Auditory verb comprehension Oral action naming: finite verbs ^a Oral object naming ^a
Grammatical	Mean length of utterance by word ^a Inflected verbs (percentage) Embedded sentences (percentage)	Auditory sentence comprehension Visual sentence comprehension Constrained oral picture description Written picture description
Phonological	Phonological paraphasias ^a Fragments <i>Conduite d'approche</i> ^a	Nonword repetition ^a Phonological paraphasias Fragments <i>Conduite d'approche</i> ^a
Error analyses	Determiner errors Morphological errors: nouns Morphological errors: verbs Neologisms Semantic paraphasias Circumlocutions ^a Interruptions Perseverations ^a Unfinished sentences ^a Fillers Anomic pauses (hesitations) ^a Other	Determiner errors Morphological errors: nouns Morphological errors: verbs Neologisms Semantic paraphasias Circumlocutions lexical

Note. The spontaneous speech measures and formal testing tasks are ordered per language domain. NA = language domains that were not assessed in this study. Further details available in Supplemental Material.

^aFrequently impaired tasks in our sample.

Venneri, 2002), and verbal memory (Digit Span Forward/Backward, Orsini et al., 1987; Spinnler & Tognoni, 1987) were considered to rule out non-language-specific factors that may affect the results. These measures were not counted in as part of the comparisons between spontaneous speech and formal tasks.

Spontaneous speech

We asked four questions to elicit connected speech—Question 1 was phrased differently for patients (1a) and for healthy speakers (1b).

- (1a) Please tell me what happened to you. How come you are at the hospital today?
- (1b) Please tell me about any time you hurt yourself and had to go to the hospital.

- (2) Please tell me what you will do today. What will you do tomorrow? And in the holidays?
- (3) Please tell me the story of the Little Red Riding Hood.
- (4) Please tell everything you see happening in this picture (Cookie Theft, Goodglass & Kaplan, 1972).

To keep the setting as similar as possible across patients, examiners asked each probe question and let the patient talk, adding as few comments or related questions as possible. Responses were recorded, and the audio files were transcribed by the first author, following descriptive conventions (Dardano & Trifone, 1997). Measures of cohesion, coherence, and prosody and story grammar were not analyzed, as the focus of this study was on phonological, lexical-semantic, and syntactic variables. To ensure transcription reliability, two experienced neuropsychologists, both native speakers of Italian, listened to each recording and marked disagreements between the transcriptions made by the first author and the audio files. The interrater agreement (i.e., percentage of nondiscordant scores over the total number of words transcribed by the first author and the errors marked by the two independent raters) was above 90% for all transcriptions (mean agreement = 96; $SD = 2.97$). Instances of disagreement were discussed until a final consensus was reached.

We analyzed a maximum of 300 words from the answer to each question, and excluded from the word count repairs, false starts, repetitions, fillers, and comments on the narrative. The data collected from each individual consisted of a maximum of 1200 words (300 words $\times$ 4 questions). Analyses of spontaneous speech were performed for each participant on the data across the four questions, as this increases test-retest stability and reliability (Boyle, 2015; Brookshire & Nicholas, 1994a, 1994b). We performed calculations on combined data (e.g., all questions combined for subject S. O. at T1) rather than calculating individual scores for each question and then averaging them. We considered a reduced number of nouns and verbs and/or a reduced diversity of these word classes, and the occurrence of word-level errors (semantic paraphasias, circumlocutions) as indices of lexical-semantic impairment, resulting from loss of vocabulary and/or of word meaning (Vermeulen et al., 1989). Reduced MLU and reduced percentages of inflected verbs and of embedded sentences were considered as indicators of syntactic deficits. Segmental errors (phonemic paraphasias, fragments, *conduites d'approche*, neologisms) were considered as an index of phonological deficits. Reduced speech rate, unfinished sentences, and fillers/

repairs were considered as “mixed” deficits, as they may result from both lexical–semantic and grammatical impairments. Lexical counts (i.e., “bike” in “I have a new bike” is a noun) were aided with a part-of-speech tagger, TreeTagger, trained for Italian (Schmid, Baroni, Zanchetta, & Stein, 2007). A detailed description of each linguistic measure is provided in the Supplemental Material. The scores of each patient were compared to those of the healthy group by means of modified *t* tests (Crawford, Garthwaite, & Porter, 2010; Crawford & Howell, 1998), based on cutoff scores (i.e., the scores below/above which the patient’s performance differed significantly from that of healthy speakers, $p < .05$, two-tailed). Individual scores for each patient in each separate assessment time (e.g., T2) were compared with the mean scores of the healthy speakers, who were only assessed once.

The whole dataset included two hours of spontaneous speech from patients and 1 hour and 42 min of speech from non-brain-damaged speakers. Administering the spontaneous speech task took on average 7.24 min ($SD = 1.62$) in healthy speakers. For people with brain tumors, the average time was 6.93 min ($SD = 3.75$) at T1, 9.69 min ($SD = 3.84$) at T2, and 6.24 min ($SD = 2.34$) at T3. The time to transcribe each sample was approximately 1 hour and 30 min (this includes the time needed for the experienced neuropsychologist to amend the first-pass transcripts). The time to analyze each sample was approximately 2 hours.

Formal tasks

Language was formally assessed with subtests from an Italian battery for aphasia (*Batteria per l’Analisi dei Deficit Afasici*, BADA; Miceli, Laudanna, & Capasso, 2006) and two picture naming tasks: object naming and production of finite verbs (Rofes, de Aguiar, & Miceli, 2015). As shown in Table 2, we included subtests of the BADA that tap lexical–semantic and grammatical abilities, both in production and in comprehension. We included a test for sublexical language processing (non-word repetition), to gain a better understanding of the phonological versus lexical–semantic nature of language deficits. Formal language impairments were diagnosed based on cutoff scores indicated in the corresponding articles/manuals. More specifically, the BADA indicates that a task is below the norm when the participant makes one or more errors (Miceli et al., 2006). The battery was normed with the help of 102 healthy volunteers ranging in age between 40 and 70 years and in education between 6 and 17 years. The naming tasks contain cutoff scores for different

age and education ranges (Rofes, de Aguiar, et al., 2015). The two tasks were normed for name agreement with the help of 65 people without brain damage ranging in age between 21 and 85 years and in education between 5 and 18 years. The cutoff scores for both tasks can be seen in Table S1 (Supplemental Material). Specific scores below the norm at a specific point in time (i.e., T1, T2, T3) can be checked in the Supplemental Material (Excel sheet). To avoid repetition effects, we used two different short versions of the BADA (i.e., Version 1 in T1 and T3, and Version 2 in T2).

Differences between the different versions of the BADA tests at the different assessment points (i.e., T1, T2, T3) were evaluated with Fisher’s exact test. Differences between the object naming task and the finite verb production task across assessment points were analyzed with McNemar’s tests. The time taken to administer the battery of formal language tasks ranged between 45 and 60 min. The time to analyze the battery of tests was very short, as the computerized version of the BADA provides a summary sheet with the final scores for each task seconds after finalizing the battery of tests.

Results

Individual results are summarized in Table 3. A detailed discussion of individual results is reported in the Supplemental Material.

Do spontaneous speech and formal neuropsychological testing detect comparable language impairments in our patient sample?

A spontaneous speech score was impaired when it fell significantly below/above that of the 15 healthy speakers (using modified *t* tests, Crawford et al., 2010; Crawford & Howell, 1998). A formal test score was considered to be impaired when it fell below the range of other cognitively unimpaired normative samples, as defined by the standardized tests (Miceli et al., 2006; Rofes, de Aguiar, et al., 2015). When we collapsed the results for the five participants across the three testing sessions ($N = 45$) there was no significant difference in the number of impaired scores for spontaneous speech (32/45, 72%) and formal testing (28/45, 63%), Fisher exact test, $z = 0.67$, $p = .505$, two-tailed).

A language level (lexical–semantic, grammatical, phonological) was impaired when at least one score tapping that level was impaired. For example, in spontaneous speech, the lexical–semantic level was impaired when patients produced fewer nouns than healthy speakers

Table 3. Individual results on the impairments for each language domain with spontaneous speech and formal testing across assessment times.

	A.A.			C.R.			C.E.			R.A.			S.O.		
	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3
Fluency/speech rate	—	S	—	S	—	S	—	S	—	NS	NS	NS	—	—	—
Lexical–semantic	—	S+FT	S+FT	S+FT	S+FT	FT	S+FT	S+FT	S+FT	S+FT	S+FT	S+FT	S+FT	S+FT	S
Grammatical	—	S+FT	S	FT	FT	S+FT	S+FT	S	S	S	S+FT	S+FT	FT	—	S
Phonological	—	S+FT	S	FT	S+FT	FT	S	S	S	—	S+FT	S+FT	FT	S	—

Note. Impairments detected with spontaneous speech (S), formal testing (FT), both measures (S+FT). Impairments not detected at this level (—). Fluency deficits can only be detected with spontaneous speech. Impairments in spontaneous speech were defined as scores significantly below those of 15 healthy speakers (using modified *t* tests, Crawford et al., 2010; Crawford & Howell, 1998). Impairments on formal testing were defined as those failing below the range of other healthy populations, as defined by the standardized tests. The comprehensive language battery typically indicates that a task is below the norm when the participant makes one or more errors (*Batteria per l'Analisi dei Deficit Afasici*, BADA, Miceli et al., 2006); the cutoff scores we used for object naming task and production of finite verbs (Rofes, de Aguiar, et al., 2015) were based on different standardized samples that were matched to our participants in terms of age and education.

and, in formal testing, when oral object naming was below the norm. To understand whether one measure was intrinsically more sensitive than the other, we compared the number of instances in which both measures converged towards indicating damage to the same language level (e.g., when both spontaneous speech and formal testing detected an impairment at the lexical–semantic level), as opposed to the number of times when language impairments were picked only by spontaneous speech or only by formal testing (e.g., when formal testing detected an impairment at the lexical–semantic level but spontaneous speech did not, or vice versa). We did this for the five participants across the three testing sessions ($N = 45$). Results showed that both measures concurrently pointed to the same level of language impairment significantly more often (21/39 or 54%, where 39 corresponds to the total number of impairments—i.e., 45 time points where an impairment could have occurred minus 6 time points where an impairment was not detected) than a level was found to be impaired only on formal testing (7/39 times, 18%; Fisher exact, $z = 3.31$, $p = .002$, two-tailed), or only on spontaneous speech analysis alone (11/39 times, 28%; Fisher exact, $z = 2.31$, $p = .038$, two-tailed). There was no significant difference in the number of times that formal testing alone and spontaneous speech alone detected a particular level of language impairment (Fisher exact, $z = 1.07$, $p = .421$, two-tailed).

Which language levels (lexical–semantic, grammatical, phonological), assessment times (T1, T2, T3), and spontaneous speech variables are more commonly impaired in our patient sample, as opposed to in healthy speakers?

Specific to language levels, we counted the number of impaired language levels across all time points, regardless of whether difficulties emerged in spontaneous speech or in formal testing. Lexical–semantic [14/15 times (5 participants $\times$ 3 time points); 93%],

phonological (12/15; 80%), and grammatical impairments (13/15; 87%) were diagnosed equally often, $\chi^2(2) = 1.15$, $p = .858$. Additionally, we counted the number of times in which spontaneous speech analysis and formal testing concurrently indicated an impairment of the same language level. The two measures converged more often in the presence of lexical–semantic impairment (12/14 times; 86%), than in the presence of phonological (4/12; 33%) or grammatical (5/13; 38%) impairments, $\chi^2(2) = 8.99$, $p = .013$. Note that in the latter analyses the denominators correspond to the number of impaired language levels across all time points (e.g., 14 for lexical–semantic impairments). The difference in the occurrence of grammatical and phonological impairments was not significant (Fisher exact, $p = 1$, two-tailed). Fluency measures (total number of words, total number of sentences, and number of words per minute) were reduced on 7/15 occasions (47%). Fluency measures only applied to spontaneous speech. They were not evaluated during formal testing.

Regarding differences between the time points in the extent of impairment, we analyzed the number of times in which spontaneous speech analysis and formal testing converged in indicating damage to the same language level. Language difficulties were numerically, but not significantly, more frequent within a week from surgery (T2: 11/15, or 73%) than before (T1: 5/15, or 34%) or seven months after surgery (T3: 6/15, 44%), $\chi^2(2) = 5.51$, $p = .076$. Additionally, we examined the number of times in which only one measure detected impairments. Spontaneous speech analysis but not formal testing detected impairments at T1 (2/15, 13.4%), at T2 (3/15, 20%), and at T3 (6/15, 40%). Formal testing but not spontaneous speech detected impairments at T1 (4/15, 27%), at T2 (1/15, 6.7%), and at T3 (2/15, 13%). In both cases, no differences across assessment times were observed [spontaneous speech: $\chi^2(2) = 3.13$, $p = .217$; formal testing: $\chi^2(2) = 2.37$, $p = .463$]. Specific to T3, spontaneous speech did not detect more additional language impairments than

formal testing at T3 (an additional 6 impairments for spontaneous speech vs. an additional 2 impairments for formal testing), Fisher exact, $p = .215$, two-tailed.

Finally, the analysis of spontaneous speech difficulties took many dimensions of performance into account. Some of these were quantitative (e.g., the number of times in which subjects scored below the norm in measures of fluency, lexical retrieval/diversity, and syntactic complexity), others qualitative (occurrence of various error types in patients' speech, as compared to the same error types in healthy volunteers). In our sample, pathological numbers of incomplete sentences (10/15, 67%) of assessment sessions) and phonological paraphasias (in 9/15, 60% of sessions) were the most common errors. Perseverations and circumlocutions were observed less often (6/15, 40%), followed by *conduite d'approche*, hesitations, and reduced number of words per minute (5/15, 33.3% of sessions). Number of adjectives, noun diversity, mean length of utterance, determiner errors, morphological errors on nouns, semantic paraphasias, and comments on the narrative were all affected in 3/15 assessments (20%); number of nouns, verb diversity, number of embedded sentences, amendments, morphological errors in verbs, and neologisms were below norm in 2/15 assessments (13.4%).

Discussion

We studied the perioperative performance of five individuals who underwent awake surgery for the removal of a left perisylvian tumor using spontaneous speech and formal language tasks. Patients were evaluated before surgery (T1), within a week after surgery (T2), and seven months after surgery (T3). Spontaneous speech scores were compared with those of a group of age-matched healthy speakers ($N = 15$). The main goal of this study was (a) to assess whether spontaneous speech and formal testing detected comparable language impairments in our sample. We also examined (b) which language levels, assessment times, and spontaneous speech variables were most often affected in our patient sample. Ultimately, we used this information to discuss whether spontaneous speech analysis is preferable to formal testing in this population.

Did spontaneous speech analyses and formal testing detect comparable language impairments in our patient sample?

Both measures were equally good at detecting language impairments and identified language deficits in a comparable number of instances at the phonological,

lexical–semantic, and grammatical levels. In addition, the number of instances in which the two measures converged in detecting language problems at a given level was significantly higher than the number of times when only one measure indicated damage to one level, and the other measure did not. The fact that both measures are equally good does not imply that they assess the same language capacities, nor that we should use either without the other. For example, spontaneous speech and formal testing may tackle the same abilities, but in different computational contexts, which may influence results. Word retrieval is a good case in point. Object naming, probably the most widely used task for language evaluation, provides information on single-word retrieval, largely unencumbered by additional cognitive/linguistic requirements. Spontaneous speech also requires word retrieval, but in this case grammatical and memory processes pose additional computational demands that may affect performance accuracy. Consequently, word retrieval difficulties may co-occur (or dissociate) in object naming and in spontaneous speech for different reasons (Herbert et al., 2008; cf. Bastiaanse & Jonkers, 1998).

In this regard, additional analyses on possible non-linguistic factors showed that the Digit Span Backward—a measure of verbal working memory—was impaired at almost all time points (12/15 times; 80%) and more often so than Digit Span Forward (5/15 times; 33%), Fisher exact test, $z = 2.58$, $p = .025$, two-tailed. To the best of our knowledge, no studies in the stroke aphasia literature exist that examine the relationship between working memory and spontaneous speech production. Investigating the relationship between nonlinguistic cognition and spontaneous speech is a potential direction for future work, as little is known about cognitive effects of tumor and surgical treatment in this population (Talachchi, Santini, Savazzi, & Gerosa, 2011). No correlations were found between nonlinguistic factors and language measures (but not spontaneous speech) in a previous sample tested by our group (Santini et al., 2012), but we are unaware of any studies in this population examining how nonlinguistic factors relate to spontaneous speech. In the second language acquisition literature, it has been shown that working memory capacity (measured by word span measures) correlates with speech rate (the total number of words divided by the total time) but not with other lexical measures, when learners of a foreign language are asked to describe a picture (Finardi & Prebianca, 2006). If this were to be true also in our population, impairments in working memory may be related to low rate (words per minute), but not to deficits in lexical retrieval, or lexical diversity. The performance of subject S.O. militates

against this possibility: Across assessment times, he had below-normal scores in Digit Span Backward, but produced a normal number of words per minute, and showed lexical-semantic impairment.

Which language levels are most frequently affected in glioma cases?

Lexical-semantic impairments were more frequent than grammatical or phonological disorders. These results are only indicative because of our admittedly small sample size and because we used more measures of lexical-semantic impairments than of phonological impairment, which were obtained by counting the number of words with phonological errors produced during spontaneous speech and formal testing, and a nonword repetition task in formal testing. That said, our results are in agreement with other data from tumor patients (i.e., Satoer et al., 2013) and suggest that people with brain gliomas may present with lexical-semantic impairments. Fluency impairments were also common, observed in four of five subjects and in 7/15 sessions. In this study, we looked at fluency measures in spontaneous speech by looking at the number of words per minute. However, we did not use a measure of fluency in formal testing, as the formal fluency measures (e.g., category or letter fluency) cannot be considered to measure the same processes as fluency in spontaneous speech.

Which assessment times are most frequently affected in glioma cases?

Instances of impaired performance increased, but not significantly, from before surgery to one week after surgery, probably due to surgery-related factors, such as physical and psychological adaptation following the surgical procedure (Sanaï, Mirzadeh, & Berger, 2008; Santini et al., 2012). It is possible that across assessments, the incidence of some error types, such as hesitations and unfinished sentences, may be influenced by psychological factors including anxiety (Kasl & Mahl, 1965; cf. Talacchi et al., 2011). Emotional stress is a factor that may have influenced linguistic performance throughout the disease course, and at all assessment times. Indeed, there is evidence in stroke aphasia that anxiety can affect discourse measures (Cahana-Amitay et al., 2015). At T1, patients were waiting to undergo the surgical procedure; at T2, they had undergone surgery only a few days earlier and were in early stages of recovery; at T3 they had started additional adjuvant treatments and were anxious about later outcome. Studies comparing neurological and

non-neurological populations about to undergo surgery, or including patients in particularly stressful medical situations, could provide relevant information on the extent to which this may influence performance.

Which spontaneous speech variables are most frequently affected in our sample?

Like Satoer et al. (2013, 2014), incomplete sentences were amongst the most common errors in the spontaneous speech of the five patients tested here. Satoer et al. argued that such errors were due to grammatical impairments. In our sample, incomplete sentences occurred with the same frequency as phonological paraphasias; hence, our participants may have not only grammatical impairments but also lexical and/or phonological impairments. We also found high numbers of perseverations and circumlocutions, problems with numbers of words per minute, *conduites d'approche*, and hesitations. Less frequently, we found low number of words, reduced noun/verb diversity, low number of embedded sentences, and increased morphological errors. All of these error types may be used to support the presence of grammatical, lexical-semantic and/or phonological impairments.

Is spontaneous speech preferable to formal testing in clinical practice?

Overall, our data license spontaneous speech analysis in the assessment of language deficits in awake surgery patients. Spontaneous speech and formal testing provide comparably sensitive measures of language impairments and yield converging (albeit not entirely parallel) results for most language parameters. The issue then becomes which approach should be adopted in clinical practice. The following section includes practical arguments from the perspective of the patient and that of the surgical team.

From the patient's perspective, spontaneous speech seems preferable. Formal testing requires detailed evaluations, which may take several hours (hence, several sessions, possibly on different days). In addition, patients must often travel to reach a neurosurgical or neuro-oncological center, and prolonged stays outside of their environment may be distressing. During formal testing, some tasks may prove particularly frustrating (e.g., because they rely heavily on language processes that are disproportionately impaired in the patient). By contrast, an evaluation by spontaneous speech is faster and possibly less stressful, as the relevant data are collected during a short conversation, which more

closely resembles communicative interaction in daily life. In this study, our battery of formal tasks required 45–60 min to complete, while spontaneous speech took approximately 10 min, even a week after surgery (9.69 min, $SD = 3.84$), which is the stage at which we would expect task administration to take longest.

Some aspects of spontaneous speech-based assessments are preferable also for the awake surgery team. In individuals undergoing awake surgery for brain gliomas, language must be evaluated repeatedly and, at least initially, at close intervals (T1 and T2 are frequently spaced by less than a week; T2 and T3 by a few weeks). Since most currently used batteries do not have multiple parallel versions, repeated assessments may be affected by test/retest effects, and yield inaccurate measures (i.e., when assessing whether a test produces the same results on separate repeated administrations). Spontaneous speech corpora are less prone to (even though, not entirely exempt from) such a drawback. For example, Brookshire and Nicholas (1994a) suggested that the test/retest reliability of parameters such as words per minute, correct information units per minute (words in context that convey information relevant to the stimulus), and percentage of correct information units (number of correct information units in a sample divided by the total number of words in that sample) is unreliable when working with short spontaneous speech samples containing an answer to only one question. For example, relative to words per minute in two separate sessions they indicate that “Aphasic subject’s average change scores for sets of stimuli ranged from about six to eight words per minute. When the scores were based on individual stimuli, change scores often doubled or tripled” (Brookshire & Nicholas, 1994a, p. 124). Additionally, Boyle (2015) studied test/retest reliability of phonological paraphasias, semantic paraphasias, fragments, and repetitions in corpora collected during two separate sessions spaced by 2–7 days, in nine people with poststroke aphasia, and in a case of posttraumatic aphasia. The analysis of phonological and semantic paraphasias over time yielded particularly stable measurements. Interestingly, in both Brookshire and Nicholas (1994a) and Boyle (2015) stable measurements were obtained when collapsing scores from different questions (i.e., picture description and story retell), but not when considering individual questions. In our study, we also collapsed scores from different questions. Having said that, and particularly at T2 (i.e., two weeks from T1 and within a week after surgery), some patients commented that they had answered the same questions before surgery. When this happened, we encouraged them to retell the narrative with the same level of detail as before surgery.

From the perspective of the surgical team, however, other practical considerations do not allow the recommendation of spontaneous speech analyses over formal testing for routine clinical use. Corpora must be transcribed and analyzed, and both steps are time consuming. In our case, transcription required, on average, 1 hour and 30 min (this includes the time needed for the experienced neuropsychologist to amend the first-pass transcripts). Analysis is also a lengthy process, despite computer-based tools allowing some automated analyses. While these reduce the time needed for a thorough evaluation, current tools only do so to a very limited extent. In our study, analyzing the speech corpora took on average 2 hours per session. Lexical retrieval, lexical diversity, and MLU variables can be obtained in a matter of seconds using Computerized Language Analysis (CLAN) (MacWhinney, 2000). This procedure, however, still requires previous transcription work, which has to be done by hand.

Promisingly, Jarrold et al. (2014) and Fraser et al. (2013) have proposed automated methods to transcribe and analyze spontaneous speech in people with dementia. Such automatic speech analysis uses a variety of engineering tools to transcribe the audio file of individual subjects (e.g., Automated Speech Recognition), to measure the number and frequency of word types (e.g., mechanisms of lexical feature extraction such as automatic part-of-speech tagging and linguistic inquiry and word count software), and to discriminate different dementia types with machine learning algorithms. These tools are still under development but provide the hope that in the near future, faster spontaneous speech analyses will be available than currently. Further work also seems necessary to decide on a manageable number of spontaneous speech variables that are meaningful for identification of impairment in people with brain tumors. Reasonable candidates in a computerized approach may be: (a) the number of words per minute as a fluency measure; (b) noun and verb percentages and type/token ratios or another measure of diversity for lexical retrieval and diversity; (c) sentence length or mean length of utterance for grammatical performance; and (d) detection of (correct and incorrect) word fragments or of phonological paraphasias, as a phonological measure. Further work may give firmer grounds to decide which variables should be best included in such an analysis.

Conclusion

In our sample of five participants with glioma, spontaneous speech analyses and formal neuropsychological testing detected comparable language impairments and may therefore be thought to be equally good for

detecting language impairment in people with brain tumors. Nevertheless, parallels between the two measures must be drawn cautiously as there are many practical and theoretical differences between them. Furthermore, our results await replication in bigger samples. At this stage, before future work brings about a rapid spontaneous speech assessment, formal testing remains the better option except for cases where testing time is limited or the patient is too tired to undergo formal testing.

Acknowledgments

We express our gratitude to Monica Ricci, Giovanna Cappelletti, Vânia de Aguiar, German Kruszewski, Marco Baroni, and Claudio Bendazzoli for help and comments at different stages of this project.

Disclosure statement

The authors report no conflict of interest concerning the materials or methods used in this study or the findings specified in this paper.

Funding

This work was supported by PAT (Provincia Autonoma di Trento) to G.M.; Fondazione CaRiTRO (Cassa di Risparmio di Trento e Rovereto) to G.M.; the Erasmus Mundus PhD Program IDEALAB (Macquarie University, Newcastle University, University of Groningen, University of Trento and University of Potsdam [agreement number 2014-0025]; the Global Brain Health Institute (Trinity College Dublin and University of California, San Francisco); a Cognitive Science Postgraduate Grant from Macquarie University to A.R.; and an Australian Research Council Future Fellowship [grant number FT120100102] to L.N.

ORCID

Lyndsey Nickels <http://orcid.org/0000-0002-0311-3524>

References

- Bastiaanse, R., & Jonkers, R. (1998). Verb retrieval in action naming and spontaneous speech in agrammatic and anomic aphasia. *Aphasiology*, 12(11), 951–969.
- Bello, L., Galluci, M., Fava, M., Carraba, G., Giussani, C., Acerbi, F., ... Gaini, S. M. (2007). Intraoperative subcortical language tract mapping guides surgical removal of gliomas involving speech areas. *Neurosurgery*, 60(1), 67–82.
- Benzagmout, M., Gatignol, P., & Duffau, H. (2007). Resection of world health organization grade II gliomas involving Broca's area: Methodological and functional considerations. *Neurosurgery*, 61(4), 741–753.
- Boyle, M. (2015). Stability of word-retrieval-errors with the AphasiaBank stimuli. *American Journal of Speech-Language Pathology*, 24(4), 953–960.
- Brookshire, R. H., & Nicholas, L. E. (1994a). Test-retest stability of measures of connected speech in aphasia. *Clinical Aphasiology*, 22(1), 19–133. Retrieved from <http://aphasiology.pitt.edu/archive/00000163/01/22-09>
- Brookshire, R. H., & Nicholas, L. E. (1994b). Speech sample size and test-retest stability of connected speech measures for adults with aphasia. *Journal of Speech, Language, and Hearing Research*, 37(2), 399–407.
- Caffarra, P., Vezzadini, G., Dieci, F., Zonato, F., & Venneri, A. (2002). Una versione abbreviata del test di Stroop: dati normativi nella popolazione italiana [A short version of the Stroop test: normative data from an Italian population]. *Nuova Rivista di Neurologia*, 12(4), 111–115. Retrieved from <http://cat.inist.fr/?aModele=afficheN&cpsidt=14013019>
- Cahana-Amitay, D., Oveis, A. C., Sayers, J. T., Pineles, S. L., Spiro, A., III, & Albert, M. L. (2015). Biomarkers of “Linguistic Anxiety” in aphasia: A proof-of-concept case study. *Clinical Linguistics & Phonetics*, 29(5), 401–413.
- Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, 14(1), 177–208.
- Carlesimo, G. A., Caltagirone, C., Gainotti, G. U. I. D., Fadda, L., Gallassi, R., Lorusso, S., ... Parnetti, L. (1996). The mental deterioration battery: Normative data, diagnostic reliability and qualitative analyses of cognitive impairment. *European Neurology*, 36(6), 378–384.
- Colman, K. S., Koerts, J., van Beilen, M., Leenders, K. L., Post, W. J., & Bastiaanse, R. (2009). The impact of executive functions on verb production in patients with Parkinson's disease. *Cortex*, 45(8), 930–942.
- Crawford, J. R., Garthwaite, P. H., & Porter, S. (2010). Point and interval estimates of effect sizes for the case-controls design in neuropsychology: Rationale, methods, implementations, and proposed reporting standards. *Cognitive Neuropsychology*, 27(3), 245–260.
- Crawford, J. R., & Howell, D. C. (1998). Comparing an individual's test score against norms derived from small samples. *The Clinical Neuropsychologist*, 12(4), 482–486.
- Dardano, M., & Trifone, P. (1997). *La nuova grammatica della lingua italiana* [The new grammar of Italian language]. Bologna: Zanichelli Editore.
- Finardi, K., & Prebianca, G. V. V. (2006). Working memory capacity and speech production in L2: Evidences from a picture description task. *Revista de estudos da linguagem*, 14, 231–260.
- Fraser, K., Rudzicz, F., Graham, N., & Rochon, E. (2013). Automatic speech recognition in the diagnosis of primary progressive aphasia. In *Proceedings of the fourth workshop on speech and language processing for assistive technologies* (pp. 47–54). Retrieved from ftp://ftp.db.toronto.edu/public_html/public_html/cs/ftp/pub/gh/Fraser-et-al-SLPAT-2013.pdf
- Garrard, P., Rentoumi, V., Gesierich, B., Miller, B., & Gorno-Tempini, M. L. (2014). Machine learning approaches to diagnosis and laterality effects in semantic dementia discourse. *Cortex*, 55(1), 122–129.
- Giovagnoli, A. R., Del Pesce, M., Mascheroni, S., Simoncelli, M., Laiacina, M., & Capitani, E. (1996). Trail making test: Normative values from 287 normal adult controls. *The Italian Journal of Neurological Sciences*, 17(4), 305–309.
- Goodglass, H., & Kaplan, E. (1972). *The assessment of aphasia and related disorders*. Philadelphia: Lea & Febiger.

- Gordon, J. K. (2006). A quantitative production analysis of picture description. *Aphasiology*, 20(02–04), 188–204.
- Herbert, R., Hickin, J., Howard, D., Osborne, F., & Best, W. (2008). Do picture-naming tasks provide a valid assessment of lexical retrieval in conversation in aphasia? *Aphasiology*, 22(2), 184–203.
- Jarrold, W., Peintner, B., Wilkins, D., Vergry, D., Richey, C., Gorno-Tempini, M. L., & Ogar, J. (2014, June). Aided diagnosis of dementia type through computer-based analysis of spontaneous speech. In *Proceedings of the ACL workshop on computational linguistics and clinical psychology* (pp. 27–36). Retrieved from <http://www.aclweb.org/anthology/W14-32#page=39>
- Kansal, K., Abraham, E., & Onyike, C. U. (2015). Longitudinal analysis of electronic messages to characterize language decline in ftd-als. *Alzheimer's & Dementia: the Journal of the Alzheimer's Association*, 11(7), 875–876.
- Kasl, S. V., & Mahl, G. F. (1965). Relationship of disturbances and hesitations in spontaneous speech to anxiety. *Journal of Personality and Social Psychology*, 1(5), 425–433.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Miceli, G., Laudanna, A., & Capasso, R. (2006). *Batteria per l'Analisi dei Deficit Afasici* [Battery for the analysis of aphasic deficits]. Bologna: EMS.
- Miceli, G., Silveri, M. C., Romani, C., & Caramazza, A. (1989). Variation in the pattern of omissions and substitutions of grammatical morphemes in the spontaneous speech of so-called agrammatic patients. *Brain and Language*, 36(3), 447–492.
- Ojemann, G. A. (1979). Individual variability in cortical localization of language. *Journal of Neurosurgery*, 50(2), 164–169.
- Orsini, A., Grossi, D., Capitani, E., Laiacona, M., Papagno, C., & Vallar, G. (1987). Verbal and spatial immediate memory span. Normative data from 1355 adults and 1112 children. *The Italian Journal of Neurological Sciences*, 8(6), 539–548.
- Prins, R., & Bastiaanse, R. (2004). Review analysing the spontaneous speech of aphasic speakers. *Aphasiology*, 18(12), 1075–1091.
- Rochon, E., Saffran, E. M., Berndt, R. S., & Schwartz, M. F. (2000). Quantitative analysis of aphasic sentence production: Further development and new data. *Brain and Language*, 72(3), 193–218.
- Rofes, A., de Aguiar, V., & Miceli, G. (2015). A minimal standardization setting for language mapping tests: An Italian example. *Neurological Sciences*, 36(7), 1113–1119.
- Rofes, A., Mandonnet, E., Godden, J., Baron, M. H., Colle, H., Darlix, A., ... Wager, M. (2017). Survey on current cognitive practices within the European low-grade glioma network: Towards a European assessment protocol. *Acta Neurochirurgica*, 159(7), 1167–1178.
- Rofes, A., Spena, G., Miozzo, A., Fontanella, M., & Miceli, G. (2015). Advantages and disadvantages of intraoperative language tasks in awake surgery: A three-task approach for prefrontal tumors. *Journal of Neurosurgical Sciences*, 59(4), 337–349. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/26159550>
- Rossi, E., & Bastiaanse, R. (2008). Spontaneous speech in Italian agrammatic aphasia: A focus on verb production. *Aphasiology*, 22(4), 347–362.
- Saffran, E. M., Berndt, R. S., & Schwartz, M. F. (1989). The quantitative analysis of agrammatic production: Procedure and data. *Brain and Language*, 37(3), 440–479. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/2804622>
- Sanai, N., Mirzadeh, Z., & Berger, M. (2008). Functional outcome after language mapping for glioma resection. *The New England Journal of Medicine*, 358(1), 18–27.
- Santini, B., Talacchi, A., Squintani, G., Casagrande, F., Capasso, R., & Miceli, G. (2012). Cognitive outcome after awake surgery for tumors in language areas. *Journal of Neuro-Oncology*, 108(2), 319–326.
- Satoer, D., Kloet, A., Vincent, A., Dirven, C., & Visch-Brink, E. (2014). Dynamic aphasia following low-grade glioma surgery near the supplementary motor area: A selective spontaneous speech deficit. *Neurocase*, 20(6), 704–716.
- Satoer, D., Vincent, A., Smits, M., Dirven, C., & Visch-Brink, E. (2013). Spontaneous speech of patients with gliomas in eloquent areas before and early after surgery. *Acta Neurochirurgica*, 155(4), 685–692.
- Schmid, H., Baroni, M., Zanchetta, E., & Stein, A. (2007). The enriched treetagger system. In *Proceedings of the EVALITA 2007 workshop*. Retrieved from https://www.evalita.it/sites/evalita.fbk.eu/files/presentations2007/04_EVALITA_UniTn_Baroni_POS_Zanchetta.pdf
- Spinnler, H., & Tognoni, G. (1987). *Standardizzazione e taratura italiana di test neuropsicologici* [Italian standardization and calibration of neuropsychological tests]. Milan: Masson Italia Periodici.
- Talacchi, A., Santini, B., Savazzi, S., & Gerosa, M. (2011). Cognitive effects of tumour and surgical treatment in glioma patients. *Journal of Neuro-Oncology*, 103(3), 541–549.
- Van Lancker Sidtis, D., & Postman, W. A. (2006). Formulaic expressions in spontaneous speech of left- and right-hemisphere-damaged subjects. *Aphasiology*, 20(5), 411–426.
- Vermeulen, J., Bastiaanse, R., & Van Wagoningen, B. (1989). Spontaneous speech in aphasia: A correlational study. *Brain and Language*, 36(2), 252–274.
- Wagenaar, E., Snow, C., & Prins, R. (1975). Spontaneous speech of aphasic patients: A psycholinguistic analysis. *Brain and Language*, 2(3), 281–303.
- Wilson, S. M., Henry, M. L., Besbris, M., Ogar, J. M., Dronkers, N. F., Jarrold, W., ... Gorno-Tempini, M. L. (2010). Connected speech production in three variants of primary progressive aphasia. *Brain*, 133(7), 2069–2088.
- Wingfield, A. (1996). Cognitive factors in auditory performance: Context, speed of processing, and constraints of memory. *Journal of the American Academy of Audiology*, 7(3), 175–182. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8780990>
- Yorkston, K. M., & Beukelman, D. R. (1980). An analysis of connected speech samples of aphasic and normal speakers. *Journal of Speech and Hearing Disorders*, 45(1), 27–36.