NEURAL CORRELATES OF CONNECTED SPEECH DEFICITS AFTER LEFT HEMISPHERE RESECTIVE SURGERY

By

ANGELICA RAE MCCARRON

A Thesis Submitted to The Honors College

In Partial Fulfillment of the Bachelors degree
With Honors in

Speech, Language, and Hearing Sciences

THE UNIVERSITY OF ARIZONA

M A Y 2 0 1 6

Approved by:

Stephen M. Wilson, Ph. D.
Department of Speech, Language, and Hearing Sciences

ABSTRACT

Transient aphasias are very common in patients during the acute days following neurosurgical resection in the language-dominant hemisphere. This study used quantitative analysis of connected speech to characterize the incidence and nature of these aphasias, to assess their recovery over time, and to determine whether there are systematic relationships between the location of surgical resection and specific deficits in various language domains. A cohort of 114 patients underwent neurosurgical resection in the language-dominant hemisphere. Language was evaluated by means of standardized assessment and connected speech elicitation prior to surgery, 2-3 days post-surgery and 1 month post-surgery. Voxel-based lesion-symptom mapping was used to elucidate any relationships between site of lesion and the specific language deficits identified by the connected speech analysis. High proportions of closed class words were associated with resections of the left middle and inferior temporal gyrus. Low proportions of closed class words were associated with resections of the left inferior frontal gyrus. Most language deficits observed 2-3 days post-surgery resolved by 1 month post-surgery. However, still present at one month were disruptions to fluency (including increased incidence of retracings and false starts),lexical access impairments, and speech sound errors (including increased incidence of phonological errors and distortions). These findings demonstrate that transient aphasias generally resolve within one month post-surgery. However, subtle language deficits left undetected by standardized language assessments sometimes persist beyond 1 month post-surgery.

INTRODUCTION

Aphasia is an acquired communication disorder caused by damage to the brain's language network that can adversely affect verbal expression, auditory comprehension, reading, writing, or any combination thereof. Furthermore, aphasia can be transient or chronic depending on lesion location and etiology.

Surgery in the language-dominant hemisphere (Chang et al., 2015) may be associated with damage to areas involved in language function and processing (Penfield & Roberts, 1959). Cortical stimulation mapping is a technique often used to curb this damage by identifying and preserving these language regions during surgery. Consequently, long-term language outcomes after resective surgery in the dominant hemisphere are exceptional overall. However, many patients still present with transient language deficits in the days following resective surgery for gliomas or intractable epilepsy in the language-dominant hemisphere.

Some previous studies have investigated these transient aphasias and which language domains they affect. For example, left temporal lobectomy has been shown to result in acute deficits in word finding (including visual and descriptive naming.) Acute comprehension, repetition, reading and writing deficits were also demonstrated in these patients. Deficits resolved by one-year post-surgery (Loring et al., 1994). Furthermore, resection of infiltrative gliomas in and around Broca's area has proven to cause more significant language deficits than resection of well-circumscribed lesions both pre- and post-operatively (Lubrano et al., 2009). According to that same study, the acute deficits can occur in any number of language domains and generally resolve within 3 to 6 months. However, some studies have shown that deficits following glioma resection in and around left hemisphere language regions in the frontal and temporal lobes persist beyond 6 months in a small percentage of patients (Sanai et al., 2008). This study found that motor speech function is mediated by cortical regions that can extend well beyond the classic boundaries of Broca's area. Furthermore, lesion sites that resulted in anomia were distributed throughout the frontal lobe. Another set of naming sites were found to be located within the superior and middle temporal gyri (Sanai et al., 2008). Additionally, neurosurgery in and around the opercular region has been shown to result in apraxia of speech, a motor speech disturbance, and aphasia, with the latter being much more transient than the former (Peraud et al., 2003). Acutely, the language deficits included word-finding, naming, and repetition, and reading difficulties, along with dysarthria and apraxia of speech; almost all deficits resolve within the 6- and 18-month postoperative period with some characteristics of apraxia of speech (motor speech deficits) persisting in a small percentage of patients.

Though these and similar studies have provided a wealth of information as to the functional neuroanatomy of language and the nature of deficits following neurosurgical resection in the language-dominant hemisphere, only a few findings make assertions about actual speech production and fluency of language. Furthermore, these assertions only address a few specific aspects of production and fluency deficits. Therefore, our knowledge of how transient aphasias impact production of connected speech—spoken language that is produced in a continuous sequence—is very limited. Furthermore, production of connected speech draws on every major language domain including phonology, morphology, syntax, and the lexicon. Each of these domains depends on the functionality of distinct brain regions within the language network. The implication of this is that there are many aspects of language that can be encapsulated by a sample of connected speech. A single fluency rating as measured in Wilson et al. (2015) is insufficiently sensitive to the complex aspects of connected speech and therefore, conflates a lot of different linguistic functions that may be compromised.

Our study attempted to address these limitations and sought to assess specifically which domains of language are impacted by transient aphasias as well as how they are impacted.
Our study had three primary goals. The first was to determine the incidence and nature of transient aphasias following surgical resection in the left (language-dominant) hemisphere in a large cohort of patients with a variety of lesion locations using quantified analysis of connected speech samples collected 2-3 days post-surgery. The second goal was to use voxel-based lesion symptom mapping (VLSM.) in order to determine whether the pattern of resulting transient language deficits was related to the location of surgical resection.

Finally, our third goal was to determine the nature of recovery by examining the longitudinal changes in language deficits in a subset of our patient cohort using connected speech samples collected prior to surgery, acutely after surgery, and one month after surgery, as a short-term follow-up time point.

Because the surgical resections are discrete and lesion locations are not limited by vascular distribution or neurodegeneration patterns as in stroke and primary progressive aphasia, respectively, this cohort provides a unique insight into the neural organization of language. Furthermore, the use of quantified analysis of connected speech allows fluency to be measured in ways not possible when using aphasia batteries, regardless of comprehensiveness, which is a limitation of previous studies. Additionally, examination of the longitudinal effects of neurosurgical resection provides insight into recovery patterns of specific language deficits.

METHODS

The cohort of patients who participated in this study is the same cohort presented in Wilson et al. (2015). As such, the description of the methodologies used in this study is based heavily on the description provided in the aforementioned paper.

Patient Population

A total of 114 patients who underwent neurosurgical resection in the left hemisphere were included in this study. Inclusion criteria are as follows: 1) a left (language-dominant) hemisphere resection in perisylvian neural regions, including the anterior temporal lobe; 2) left hemisphere dominance for language confirmed either by performance on a Wada test, presence of presurgical language deficits, or magnetoencephalography (MEG) lateralization; 3) fluency in English; 4) availability of postsurgical FLAIR and diffusion-weighted imaging (DWI) suitable for delineation of resections and any associated infarction; 5) Western Aphasia Battery (WAB) administered 2-3 days post-surgery; and 6) connected speech sample elicited through a picture description of a picnic scene from the WAB.

The study was approved by the UCSF Human Research Protection Program, and all participants gave written informed consent. Analysis of de-identified data took place at the University of Arizona.

Surgical Methods

As described by Wilson et al. (2015): “All patients underwent craniotomy using monitored anesthesia care without intubation and general anesthesia. Generous local anesthetic infiltration was applied to create a scalp block. Surgical exposure was tailored for each case, depending on the target lesion and/or seizure focus. Patients were sedated with either propofol or dexmedetomidine at the start of the procedure. Intraoperative language mapping with electrical stimulation was performed in the majority of patients. The patients were fully awake for the mapping, and intraoperative electrocorticography was used to monitor for stimulation-induced after-discharges. The intraoperative language tasks included counting, confrontation naming, and occasionally reading. After mapping, patients were re-sedated with either propofol or dexmedetomidine for the remainder of the procedure. Essential language sites were identified with stimulation mapping and defined as those resulting in a loss of function in at least 2 of 3 stimulations. The majority of sites were at least 1 cm from the resection margin. The resection was usually performed with an ultrasonic aspirator guided by intraoperative neuronavigation. Subpial resection was used where possible.”

Language Assessments

Language assessment included three components: Western Aphasia Battery (WAB), Boston Naming Test (BNT) and a connected speech sample. A speech-language pathologist, neuropsychologist, or trained research assistant administered the WAB and BNT 2-3 days post-surgery for the entire cohort of patients. A subset of 27 patients had these assessments administered 2-3days pre-surgery and 1 month post-operation. In addition to the WAB and BNT, a connected speech sample was elicited using a picture prompt (specifically, the picnic scene) from the WAB, shown below. Again, this was done for the entire cohort of patients 2-3 days post-surgery, with the same subset of 27 patients having samples elicited at the pre- and post-surgery time points as well.

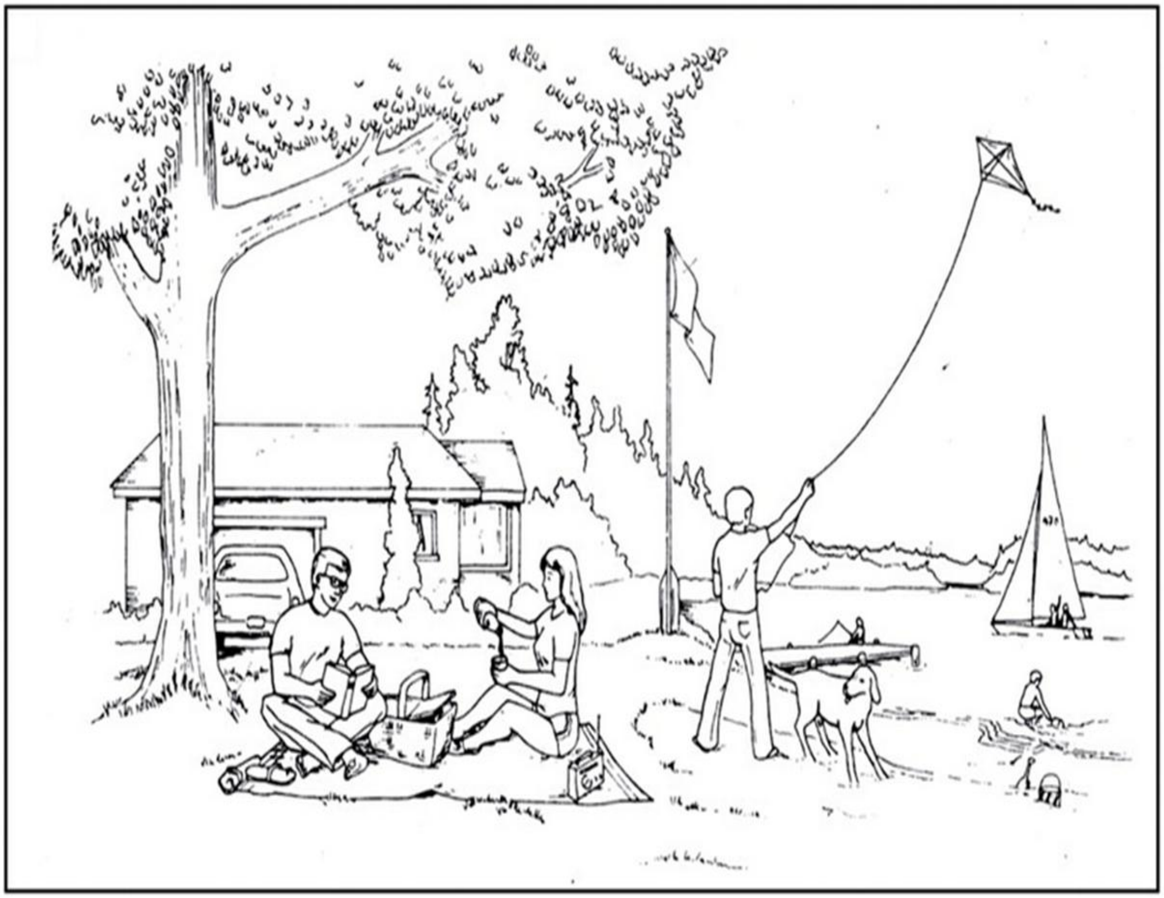

Figure 1. Picnic scene for description

The WAB is a comprehensive and validated aphasia battery. It is used to formally assess patients with aphasia-like symptoms and yields a score in the form of an Aphasia Quotient (AQ) that quantifies severity of aphasia and yields subscores in 5 different language domains, including fluency and naming. Our study used the AQ as a standard measure and did not use any of the subscores as measurements of language abilities because of their limitations in sensitively measuring subtle deficits. The naming subscore of the WAB, for instance, was not used in this study because the object-naming task was judged to be too simplistic and, therefore, subject to ceiling effects. In place of the WAB naming subtest, we administered a 15-item version of the BNT. The BNT was used because it includes challenging, low-frequency items, which makes it less subject to ceiling effects and more effective at quantifying naming.

Fluency, as measured by the WAB, is subjectively rated by the clinician and is based on the spontaneous speech of the patient. To better assess subtle expressive language deficits, our study uses quantified analysis of connected speech rather than using a subjective score of fluency from the WAB. The connected speech sample was elicited by asking patients to describe a picture using complete sentences. 114 patients provided a connected speech sample 2-3 days post-surgery, a subset of this cohort provided connected speech samples at the pre-surgery time point as well. Unsuccessful acquisition of language data prior to surgery was due solely to circumstantial factors and, therefore, there is not a bias in the sample. However, follow-up testing at 1 month post-surgery was only attempted in patients whose AQ was at least 93.8 and whose BNT score was at least 12 when tested 2-3 days post-surgery. Many patients of the total 114 met the criteria for follow-up, but due to the large amounts of time required to transcribe and code connected speech, the pre-surgery and follow-up samples of 27 patients were successfully processed, with all 114 patients having their respective 2-3 days post-surgery samples fully processed.

Since follow-up data was only obtained in patients who presented as aphasic 2-3 days post-surgery, the patients tested 1 month post-surgery do not necessarily represent an unbiased sample of the whole cohort. In order to avoid this potential bias and construct an unbiased sample of the whole cohort at 1 month post-surgery, scores for this time point were imputed for an additional 11 patients who had pre- and post-surgery language data. 12 of these patients presented as non-aphasic during post-surgery assessment. The post-surgery scores of 11 of the 12 non-aphasic patients were imputed as the follow-up scores to create a cohort of 38 patients with longitudinal language data; that is, language data collected at all three described time points. This is the best possible match to the actual breakdown of the full cohort. 78 of 114 patients presented as aphasic, comprising 68.4% of the full cohort. The follow-up sample of data is comprised of 26 aphasic patients and 1 non-aphasic patient. With scores imputed from 11 non-aphasic patients, the subset cohort with follow-up data is comprised of a total of 38 patients, 26 of whom were aphasic, comprising 68.4% of this subset.

This process of imputation of scores tends to underestimate the 1 month post-surgery scores of patients since subtle deficits present 2-3 days after surgery that were left undetected by the language assessment process would likely have resolved over the month following surgery.

Paired t-tests were implemented to compare scores on all language measures (which will be described in detail below) between presurgical and (2-3 days) postsurgical assessments; between

(2-3 days) postsurgical assessments and (1 month postsurgical) follow-up assessments; and between presurgical and (1 month postsurgical) follow-up assessments. Furthermore, patients were classified as phasic according to the WAB when their AQ was less than 93.8. Of note is the fact that a patient can be impaired in one or more language domains and still be classified as non-aphasic so long as their AQ is greater than or equal to 93.8. BNT scores were considered abnormal if below 2 or more SDs from the means determined by non-brain-damaged controls.

Transcription and coding of the connected speech samples in this study followed the CHAT system for transcription and included word- and utterance-level error codes designed specifically for quantifying aphasic connected speech. The use of CHAT in this study provided a standardized means of representing naturalistic samples of aphasic connected speech while comprehensively capturing phonological, lexical, morphosyntactic, and fluency deficits. In order to capture precise timing information of the patients' productions, the samples were transcribed using EUDICO Linguistic Annotator or ELAN. The speech samples where transcribed using standard English orthography excepting the use of the International Phonetic Alphabet (IPA) to transcribe distortions, phonemic paraphasias, and neologisms. All connected speech samples were either transcribed and coded by the author of this paper, or were checked and edited by her in order to maximize consistency in the transcription and coding process, since certain aspects of the process are subjective. For example, determining utterance boundaries can be highly subjective. In this study, utterance boundaries were identified based on principles outlined by Saffran et al. (1989). As aforementioned, errors were coded and the word and utterance level where appropriate. Word level errors included phonological errors, neologisms, semantic errors, formal lexical device errors, morphological errors, or omissions. Retracings and corrections of errors were also recorded. A total of five-utterance level error codes are outlined in the CHAT system. These include grammatical errors (both of agrammatism and paragrammatism), empty speech, circumlocutory speech, jargon, and perseverative utterances. In this study we implemented three additional post-codes for the utterance level: semantically anomalous utterances, non-sentence utterance (for example, isolated noun or verb phrases), and embeddings. Utterances could meet the criteria for classification by multiple utterance-level error codes. These codes were only used when word-level codes were insufficient for capturing the errors in question. This avoided double-counting of errors which could have artificially inflated our results.

Following transcription and coding of each sample, morphemes were parsed for part of speech using the program *mor*. In many cases, morphemes can be ambiguous for part of speech. The ambiguous morphemes in each sample were manually disambiguated by the transcriber.

The coded transcriptions were analyzed to obtain quantified measures in four broad language domains: 1) speech rate and speech sound errors; 2) lexical content; 3) morphosyntax and complexity; and 4) disruptions to fluency. This and all statistical analyses of the data were done using MATLAB.

<u>Language Assessment Measures</u>

(1) Speech rate and speech sound errors

There are three separate measures included within this language domain: words per minute, phonological errors, and distortions. Speech rate, given in words per minute, is calculated as the total number of countable words produced divided by the total time spent speaking. Countable words included real words and neologisms, but excluded false starts, fillers, unintelligible speech, words subsequently retraced and direct responses to questions posed by the examiner. The time spent speaking was the total duration of the speech sample, excluding examiner speech, excluded utterances, interruptions, and pauses between examiner and patient speech.

Phonological errors were defined as neologisms or phonemic paraphasias. Neologisms were defined as word with unknown targets or words with known targets that included more than one deletion, omission, or substitution. This definition of neologisms comes from the CHAT system; notably, this definition is broader than most typical uses of the term. Phonemic paraphasias were defined as a deletion, omission, or substitution of the onset, nucleus, or coda within one syllable of the target word. In calculating the measure of phonological errors, neologisms were weighted as two, since they contained at least two errors in comparison to phonemic paraphasias. These measures were calculated per hundred words in order to normalize for the length of sample. Furthermore, the measures are presented as the log of number of errors per hundred words in order to account for large variability (broad range) across samples in the number of phonological errors per hundred words

(2) Lexical content

A composite measure of impaired lexical access (log per hundred words) was calculated as the sum of word-level semantic errors, neologisms with unknown targets, incomplete utterances (due to a trailing off, which is generally the result of word-finding problems), utterances characterized by empty speech, circumlocutory utterances, and semantically anomalous utterances. Utterances were classified as semantically anomalous when they were not appropriate for the context of the picnic scene and when the anomaly could not be attributed to a specific word-level problem.

The proportion of closed class words (a subset of the lexicon comprised of functors and the like) was calculated by dividing the total number of closed class words by the sum of open (a subset of the lexicon comprised of content words) and closed class words. The proportion of verbs was calculated in a similar fashion: by dividing the total number of verbs by the sum of verbs and nouns. Retraced words were not included in either proportion calculation. These proportions indirectly measure agrammatism, which is characterized by decreased proportions of closed class words and verbs, as well as empty speech, which is characterized by increased proportions of closed class words and verbs.

(3) Morphosyntax and complexity

In our study, as in many others, mean length of utterance was used as a metric of syntactic complexity. It was calculated by dividing the total number of countable words in each sample (excepting false starts, fillers, unintelligible speech, and retraced words or phrases) by the total number of utterances in the given sample.

Bound morphemes (per hundred words) were counted and measured morphological complexity. This measure included both inflectional and derivational morphemes. Incorrectly used bound morphemes are still included in the count because they demonstrated an attempt to use morphology.

Utterances containing embeddings (per hundred words) were coded using the utterance-level post-code and counted as a measure of syntactic complexity. Embeddings were defined to contain a subject or finite verb form. Multiple embeddings in a single utterance were counted. Embeddings were counted even when agrammatical so long as the defining criteria were clearly present.

A composite measure of morphosyntactic errors (log per hundred words) was calculated to capture deficits in this domain. This composite measure was the sum of word-level morphological errors, formal lexical device errors, omitted words (these were only coded when they could be identified with a reasonable level of confidence, as identifying these is somewhat subjective), utterances that were incomplete sentences (for example, isolated noun or verb phrases), and agrammatical or paragrammatical utterances. The utterance-level was postcode for agrammatism and paragrammatism was applied when utterances contained more than one word-level morphosyntactic error or when the specific word-level errors could not be confidently identified (in cases where syntax was severely garbled.) As such, this postcode was given double weight in the sum for the composite measure.

(4) Disruptions to fluency

To quantitatively assess fluency in this study, we used three additional measures which together capture disruptions to fluency and provide an overall impression of non-fluent connected speech. These measures include false starts, retracings, and filled pauses (for example, “uh” and “um.”) False starts were defined as partial words that were started, but left unfinished. In other words they were abandoned and, in most cases, after one or a few phonemes had been produced. Retracings were defined as a sequence of one or more words that were redundant due to the presence of subsequent repetitions, elaborations, amendments, or alternative expressions. Again, to normalize for sample length, these measures were calculated per hundred words.

A sample transcription with accompanying error codes is provided below.

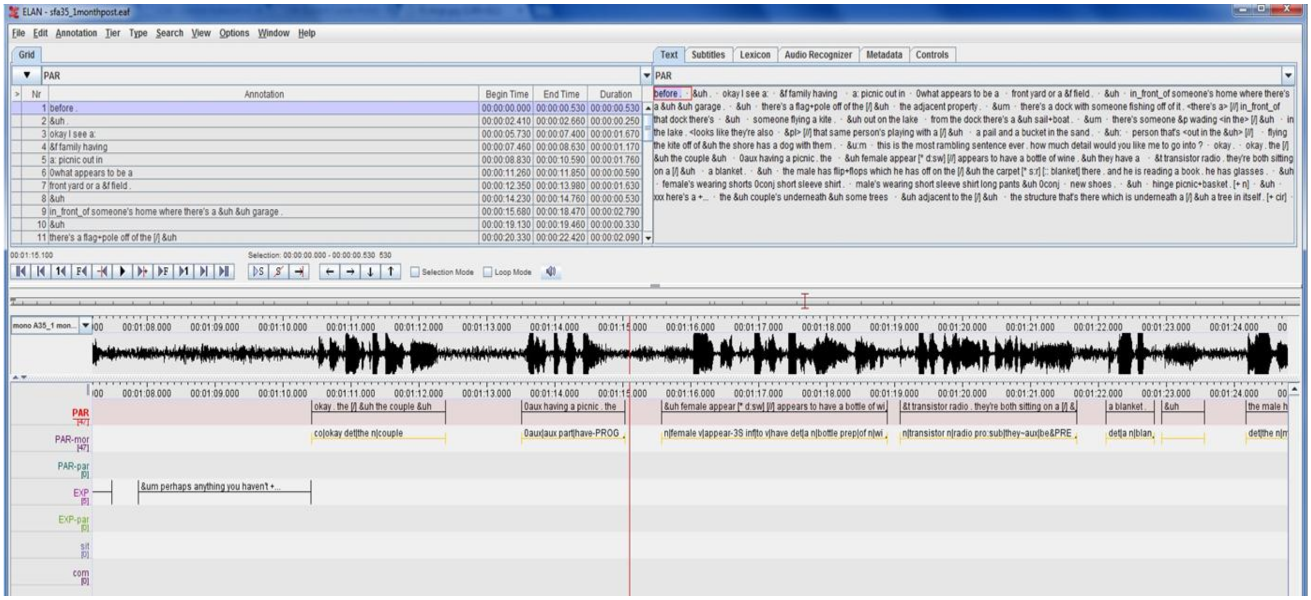

Figure 2. Sample transcription of aphasic connected speech

<u>Neuroimaging & Lesion Mapping</u>

The voxel-based lesion symptom mapping technique was used to identify lesion locations predictive of the language deficits revealed by the quantified connected speech analysis. The imaging analysis was carried out using images that were acquired 2-3 days post-surgery. 87 patients of the cohort of 114 underwent the neuroimaging procedures. A trained research assistant, with input from the neurosurgeon and a neuroradiologist, delineated the surgical site of lesion for each of the 87 patients. In this way, clinical MR images were used to demarcate the location and extent of each resection on each slice using MRIcron. These images were normalized to MNI space using SPM5, and smoothed.

VLSM was carried out and statistical maps were thresholded at voxelwise $p < 0.05$, then corrected for multiple comparisons based on cluster size using a permutation procedure.

For specific details on the imaging procedures utilized in this study refer to Wilson et al. (2015).

RESULTS

Language Assessment

Cohort of **38** patients assessed 2-3 days **pre**-surgery
Cohort of **38** patients assessed 2-3 days **post**-surgery
Cohort of **114** patients assessed 2-3 days **post**-surgery
Cohort of **27** patients assessed 2-3 days **pre**-surgery
Cohort of **27** patients assessed 2-3 days **post**-surgery
Cohort of **27** patients assessed 1 month post-surgery for **follow-up**
Cohort of **11** patients assessed 2-3 days **pre**-surgery
Cohort of **11** patients assessed 2-3 days **post**-surgery

Table 1. Legend for language assessment results, Figures 3-8

The table (provided above) applies to all graphs presented in the results section. The blue samples represent a patient cohort of 38, 26 of whom were aphasic 2-3 days post-surgery and 12 of whom were not. This cohort represents an unbiased sample of the overall cohort of 114 patients. The red samples represent a patient cohort of 27. All but one of these individuals presented as aphasic 2-3 days post-surgery. As such, follow-up language assessment and neuroimaging data were obtained 1 month following surgery. This cohort represents a biased sample of the overall cohort. To avoid this bias, language assessment scores were imputed from 11 non-aphasic patients; this cohort is represented in green.

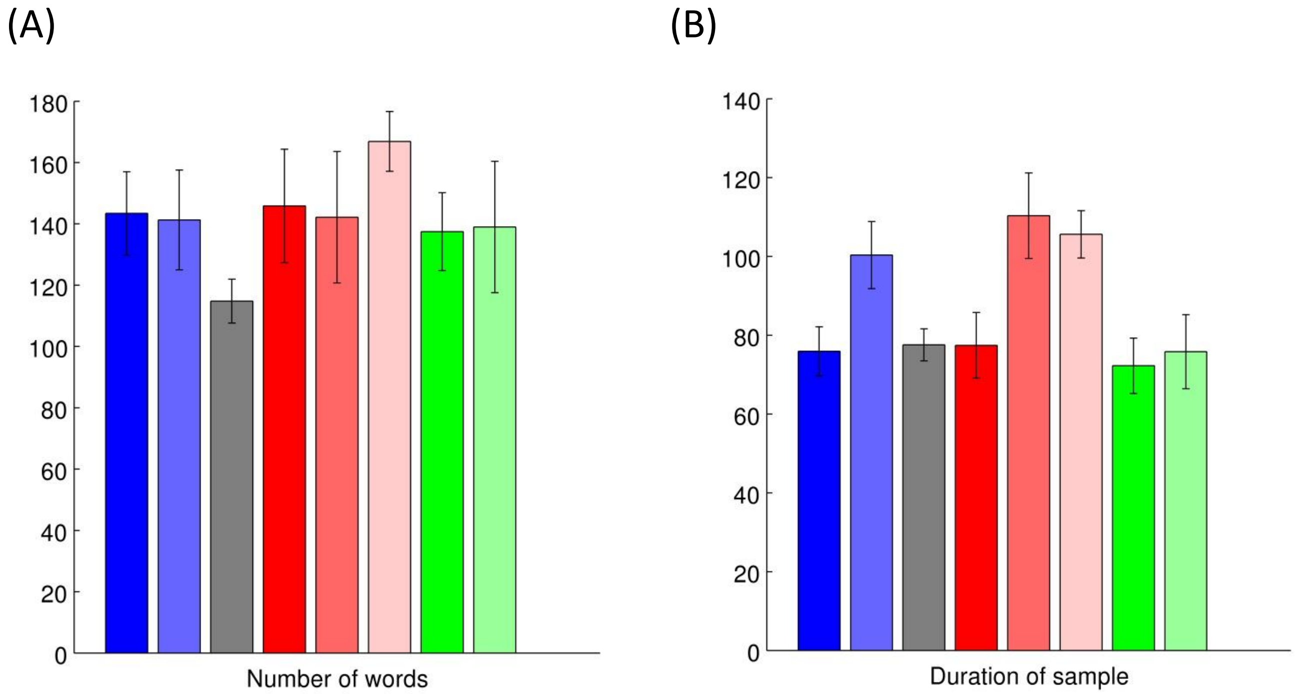

Figure 3. Sample length (words) and duration (seconds)

Graphs depicting average connected speech sample length (in words) and average sample duration (in seconds) for each cohort at various time points are provided in Figure 3 to demonstrate the nature of the connected speech samples collected in this study. Duration of sample excludes pauses between examiner and patient speech, examiner speech, as well as utterances excluded because they were answers to questions posed by the examiner during sample collection.

(1) Speech Rate and Speech Sound Errors

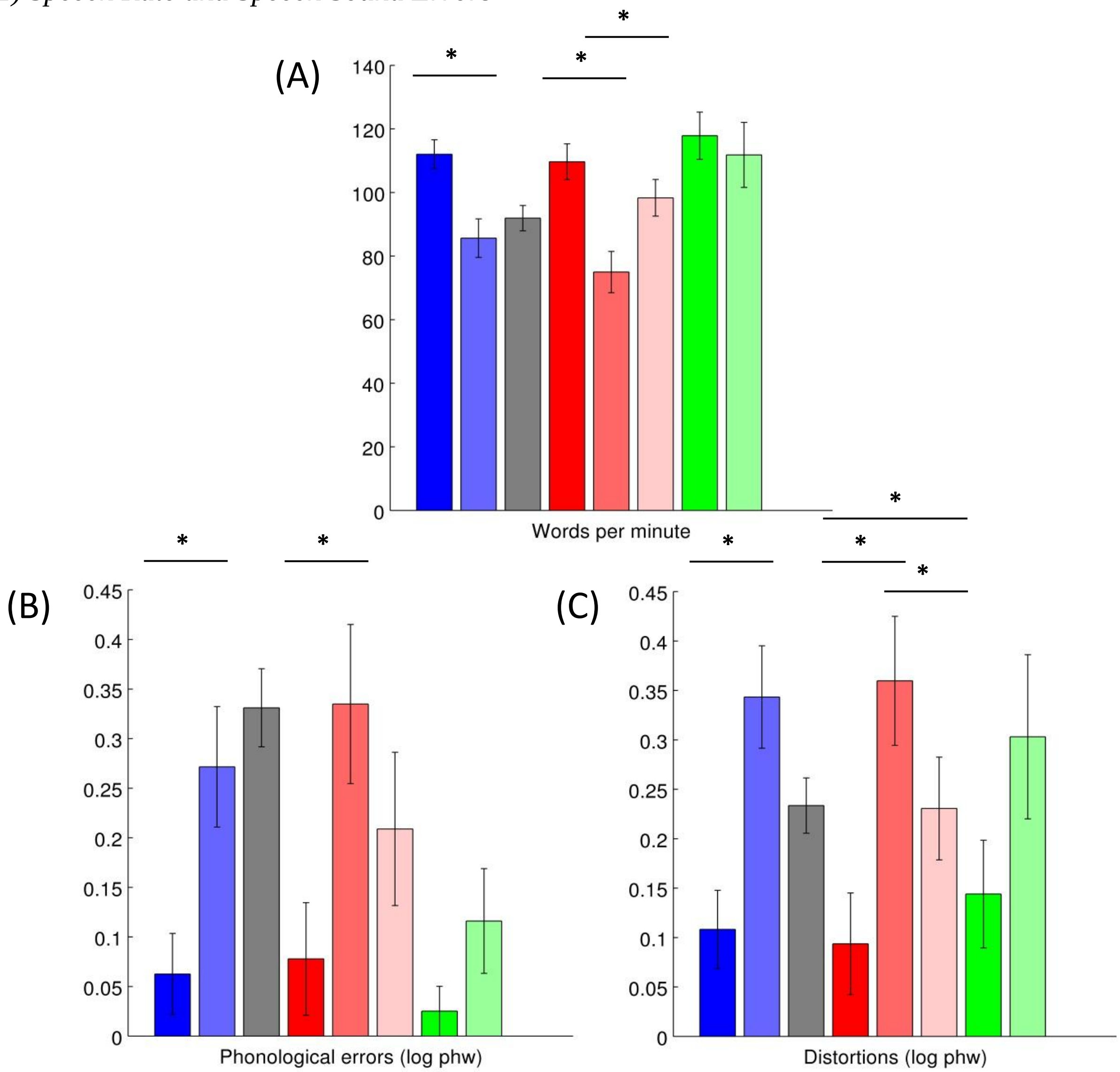

Figure 4. Measures of fluency

Phonological errors (log per hundred words) and distortions (log per hundred words): The prevalence of each increased significantly (more than doubles) following surgery in the unbiased cohort of 38 patients that is comprised of both aphasic and non-aphasic individuals ($p=0.0002$ and $p<0.0001$, respectively.) As expected, in the cohort of 27 aphasic patients with longitudinal language data across all three time points, the prevalence of phonological errors and distortions increased significantly from pre-operation to 2-3 days post-operation ($p=0.0007$ and $p=0.0003$, respectively.) In regards to recovery within the 27-patient cohort, the prevalence of phonological errors decreased, but not significantly (p=0.1355). Phonological errors did not return to normal pre-operative levels (as determined from a pre-operative to 1 month post-operative comparison in which p=0.01.) That is to say, phonological errors 1 month post-surgery were more prevalent than they were pre-surgery, indicating subtle phonemic encoding deficits. Distortions in this same cohort decreased significantly (p=0.0365) from 2-3 days post-surgery to 1 month post-surgery. In keeping with the pattern observed in phonological errors, and despite this decrease being more marginal, distortions were still more prevalent post-operatively (1 month) than pre-operatively (p=0.0029), indicating the presence of subtle motor speech deficits 1 month after surgery.

Figure 4A shows the other measure used to capture motor speech function, effective words per minute. With regard to the 38-patient cohort, rate of speech decreased significantly (p=0.0003) pre-surgery to 2-3 days post-surgery. This aligns with expectations that the rate of output should decrease acutely after surgery, when articulatory control is likely to be diminished depending on lesion location and when neural reorganization has not yet occurred. As expected, this significant decrease in rate of speech output was demonstrated in the 27-patient aphasic cohort (p=0.0001). In regards to recovery, there was a significant increase in words per minute from 2-3 days post-surgery to 1 month post-surgery (p=0.0012). Furthermore, effective wpm returned to normal, pre-operative levels (p=0.0947), as there was no significant difference between 2-3 days pre-surgery and 1 month post-surgery.

(2) Lexical Content

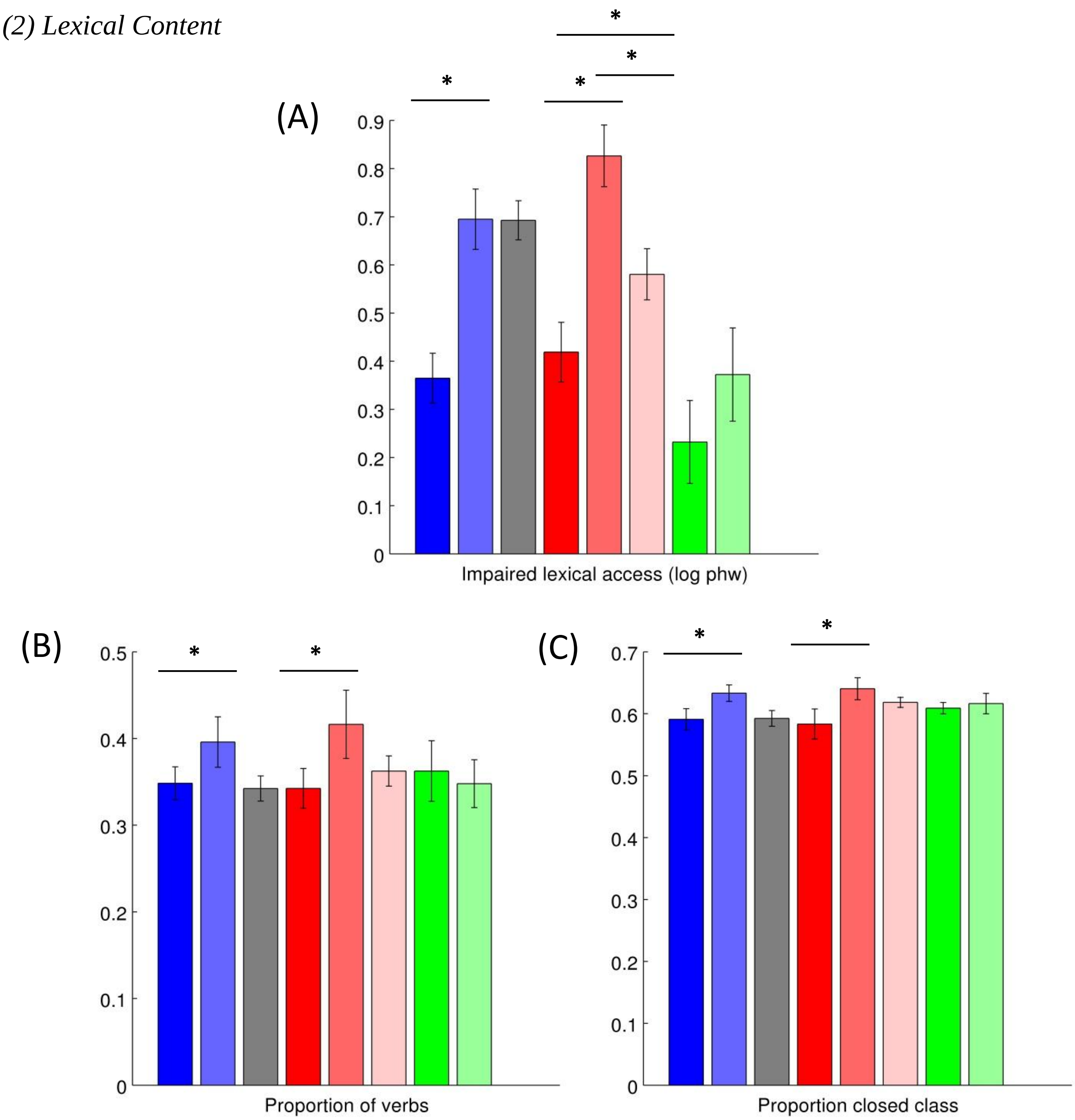

Figure 5. Measures of lexical access

Figure 5A depicts the composite measure of impaired lexical access. The degree of impairment increased significantly pre-surgery to 2-3 days post-surgery in both the 38-patient cohort and the 27-patient aphasic cohort. ($p=<0.0001$ and $p=0.0001$, respectively.) In regards to recovery, we turn our attention to the 27-patient cohort which is marked by longitudinal data. There was a significant decrease in the degree of impairment to lexical access between the 2-3 days and 1 month post-surgery period (p=0.0024). Despite this, lexical access did not return to normal, pre-operative levels, as there was a significant increase between the pre-operative and 1 month post-operative period (p=0.0153).

Figure 5B depicts the measure of proportion of verbs, which when low is indicative of agrammatic speech that is lacking in functors, and indicative of impaired lexical access resulting in empty speech when the proportion is high. The increases in the proportion of verbs used by individuals in both the 27-patient and 38-patient cohort from pre-surgery to 2-3 days post-surgery were marginally significant (p=0.0378 and 0.0679, respectively). In regards to recovery, there was a non-significant decrease in the proportion of verbs from the 2-3 days to 1 month post-surgery period in the 27-patient cohort (p=0.1599). Additionally, there was no significant difference between the pre-surgery and 1 month post-surgery levels in this same cohort (p=0.3723).

Figure 5C depicts the measure of proportion of closed class words, which is indicative of empty speech (and thus, impaired lexical access) when the proportion is high. Marginally significant increases in this proportion were observed in the 27-patient cohort and the 38-patient cohort from pre-surgery to 2-3 days post-surgery (p=0.0154 and p=0.0161, respectively.) In regards to recovery, there was a non-significant decrease in the proportion of closed class from 2-3 days to 1 month post-surgery in the 27-patient cohort (p=0.248). Furthermore, there was no significant difference between the pre-surgery and 1 month post-surgery levels in this same cohort (p=0.1567).

(3) Morphosyntax and Complexity

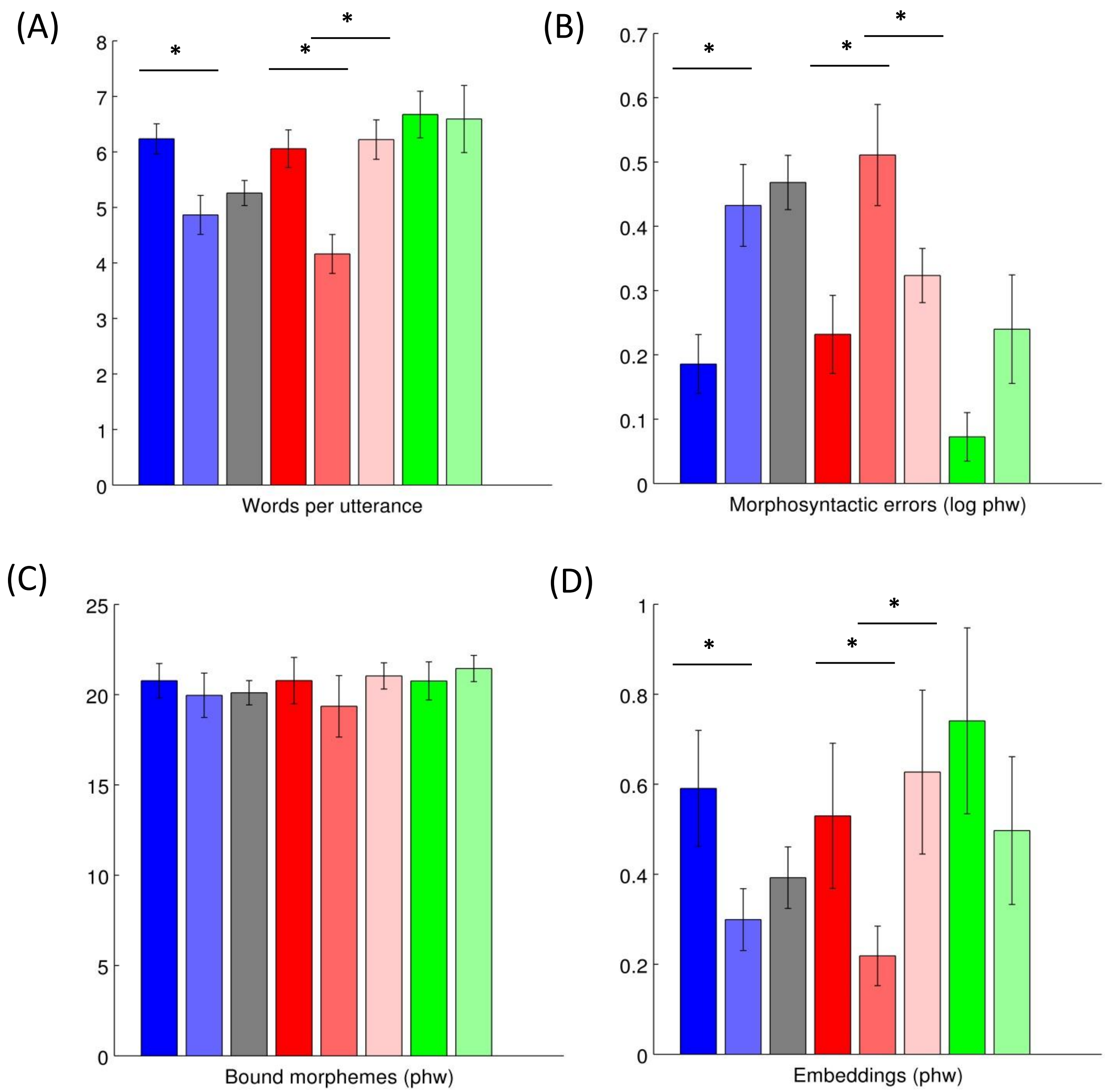

Figure 6. Measures of morphosyntax and syntactic complexity

Figure 6A depicts words per utternace, an indirect measure of syntactic complexity. A significant decrease in utterance length was observed in the 38-patient mixed aphasic and non-aphasic cohort (p=0.009). This same significant decrease was observed in the 27-patient aphasic cohort (p=0.003). In regards to recovery, following surgery, a significant increase in utterance length was observed in the 27-patient cohort between the 2-3 days and 1 month post-surgery levels. As such, there was no significant difference between the pre-surgery and 1 month post-surgery levels among this cohort (p=0.6188).

Figure 6B depicts the composite quantified measure of total morphosyntactic errors, which directly measures morphosyntactic impairments. As expected, a significant increase in total morphosyntactic errors was observed in both the 38-patient and 27-patient cohorts (p=0.0002 and p=0.001, respectively). In regards to recovery, following surgery, a significant decrease in morphosytnactic errors was observed in the 27-patient aphasic cohort (p=0.0145) from 2-3 days to 1 month post-surgery. As such, there was no signficant difference between the pre-surgery and 1 month post-surgery levels amont this cohort (p=0.1122).

Figure 6C depicts the number of bound morphemes per hundred words. Increases in this measure indicate an intact morphosyntax language domain. Decreases in this measure indicate the reverse. There were no significant changes observed in this measure among any of the patient cohorts. The indicates that the number of bound morphemes was not significantly affected by resective surgery.

Figure 6D depicts the number of embeddings per hundred words, which is a measure of syntactic complexity. Marginally significant decreases in the number of embeddings were observed in both the 27-patient and 38-patient cohort from pre-surgery to 2-3 days post-surgery (p=0.0823 and p=0.0426, respectively). Furthermore, there was a singnificant increase in number of embeddings observed in the 27-patient cohort from 2-3 days to 1 month post-surgery (p=0.0439). There was no significant difference observed between the pre-surgery and 1 month post-surgery level in this same cohort (p=0.5778).

(4) Disruptions to Fluency

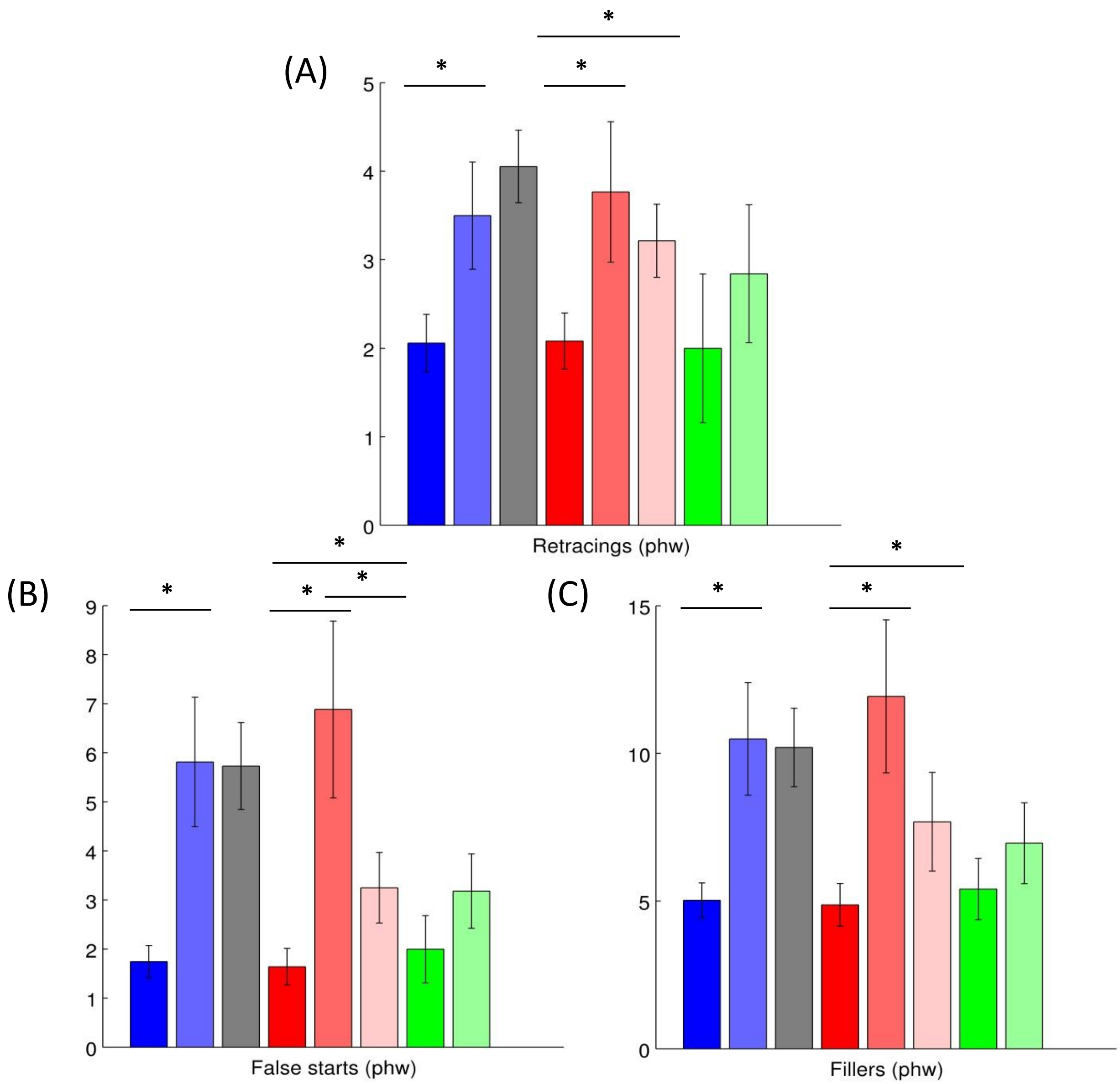

Figure 7. Measures of fluency

Figure 7A depicts the average number of retracings that occurred per hundred words amongst the patient cohorts. Though retracings are common colloquially, a high degree of retracings can indicate disfluency of speech, anda significant increase in the prevalence of retracings certainly indicates a disruption to fluency. A significant increase in the number of retracings was observed in both the 27- and 38-patient cohorts between pre-surgery and 2-3 days post-surgery (p=0.0331 and p=0.014, respectively). In regards to recovery over time, there was no significant decrease in retracings observed between 2-3 days and 1 month post-surgery in the 27-patient cohort. As such, there was a significant difference between the prevalence of retracings between pre-surgery and 1 month post-surgery levels within this same cohort (p=0.02).

Figure 7B depicts the number of false starts per hundred words. Similarly, this is a measure of disruptions to fluency. A significant increase in prevalence of false starts was observed in both the 27- and 38-patient cohorts between pre-surgery and 2-3 days post-surgery (p=0.0056 and p=0.0033, respectively). In regards to recovery, a significant decrease in false start prevalence was observed in the 27-patient cohort between 2-3 days and 1 month post-surgery (p=0.0619). Despite this, there was still a significant difference between the prevalence of retracings pre-surgery and 1 month post-surgery (p=0.0448.)

Figure 7C indicates the number of filled pauses per hundred words which also contributes to an overall measure of fluency of speech. A significant increase in prevalence of filled pauses was observed in both the 27- and 38-patient cohorts between pre-surgery and 2-3 days post-surgery (p=0.0057 and p=0.0034, respectively). In regards to recovery, a non-significant decrease in filled pause prevalence was observed in the 27-patient cohort between 2-3 days and 1 month post-surgery (p=0.115). As such, there was a marginally significant difference between the prevalence of retracings pre-surgery and 1 month post-surgery (p=0.0502). Fillers can be considered to be very marginally impaired, but relatively recovered to normal, pre-surgical levels by 1 month post-surgery.

(5) Standard Measures

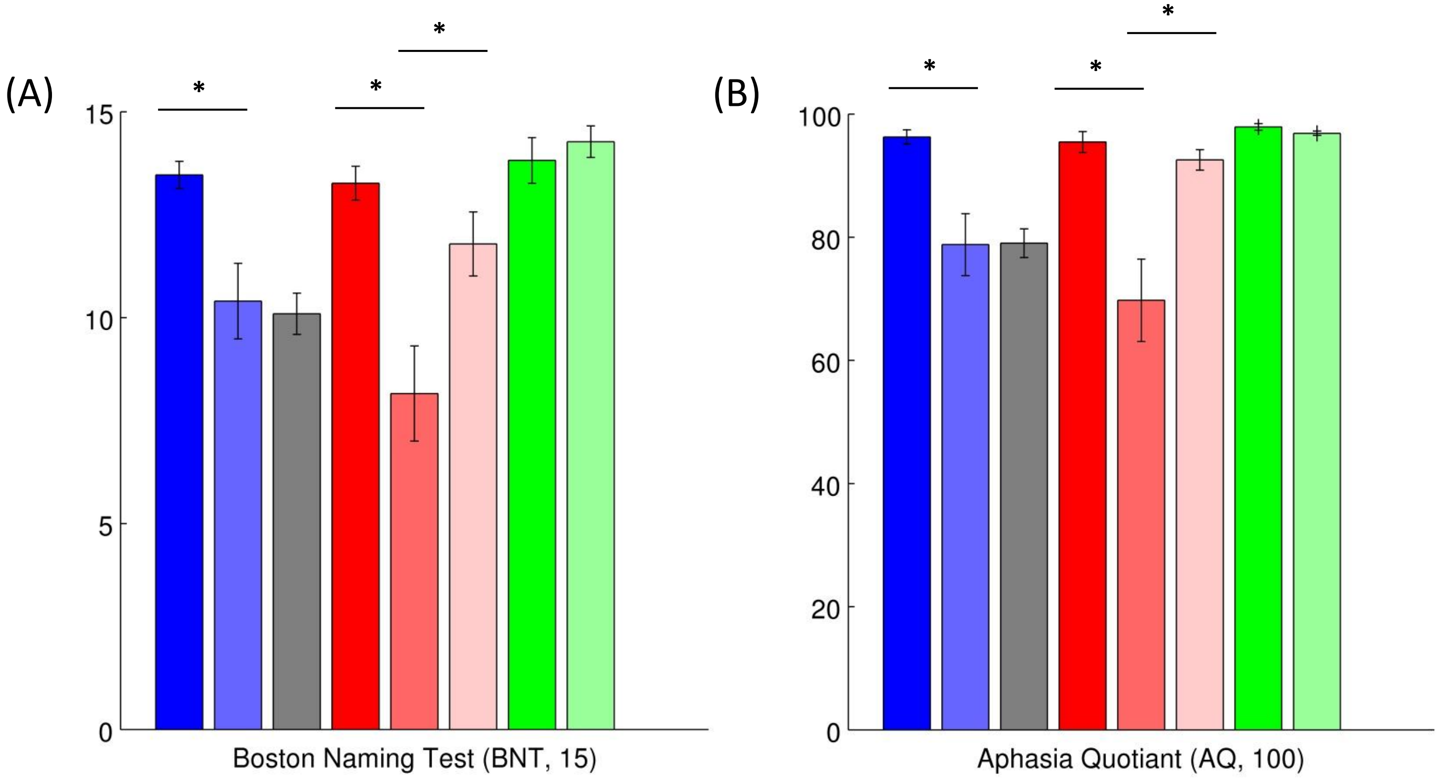

Figure 8. Measures of standardized language assessment

Figure 8A demonstrates the results from standardized language assessment using a 15-item BNT. There was a significant decrease in score between pre-surgery and 2-3 days post-surgery in both the 27- and 38-patient cohort (p=0.0003 and p=0.0016, respectively). In regards to recovery, there was a significant increase in score between 2-3 days and 1 month post-surgery in the 27-patient cohort (p=0.0026). Comparisons between pre-surgery and 1 month post-surgery scores yielded no significant differences in this same cohort (p=0.0527).

Figure 8B demonstrates the results from standardized language assessment using the WAB, scored out of 100. There was a significant decrease in score between pre-surgery and 2-3 days post-surgery in both the 27- and 38-patient cohort (p=0.0014 and p=0.0017, respectively). In regards to recovery, there was a significant increase in score between 2-3 days and 1 month post-surgery in the 27-patient cohort (p=0.0044). Comparisons between pre-surgery and 1 month post-surgery scores yielded no significant differences in this same cohort (p=0.1605).

Voxel-based Lesion Symptom Mapping

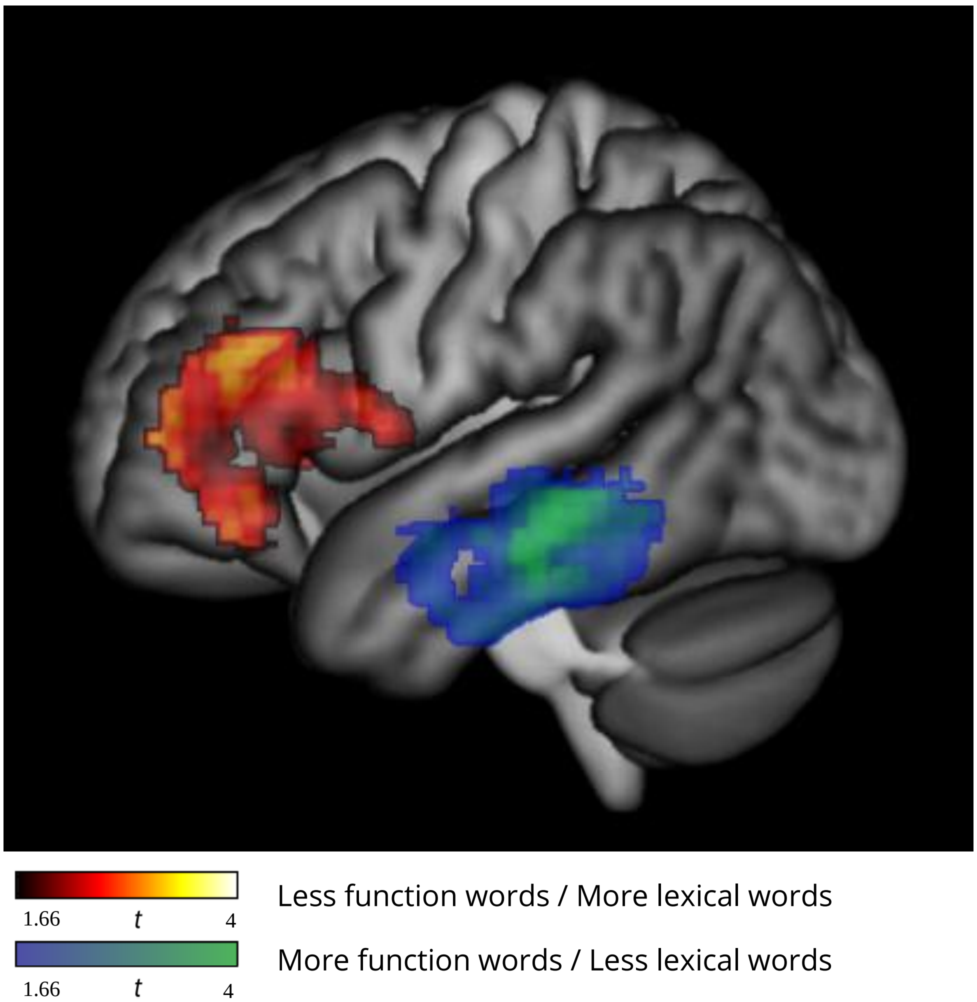

Figure 9. Voxel-based lesion symptom mapping (VLSM)

VLSM analyses of 87 of the 114-patient cohort yielded systemic relationships between lesion site or location of resection and particular language deficits. Increased proportions of function words were correlated with resections of the left middle and inferior temporal gyri (p = 0.04).

Decreased proportions of function words were correlated with resections of the left inferior frontal gyrus (p = 0.03).

Additionally, there were non-significant trends for increased phonological errors (e.g. /kaɪp/ for 'kite') associated with resection of the left ventral precentral gyrus, and increased lexical access impairments (e.g. 'parking lot' for 'driveway') associated with resections of the left middle and inferior temporal gyri.

DISCUSSION

We found that transient aphasias were very common following resective neurosurgery in the language-dominant left hemisphere and that all language domains can be affected by these transient aphasias. Longitudinal assessment and analysis of language pre- and post-operatively suggests that, though recovery is rapid and most deficits resolve within 1 month, there are some subtle deficits not captured by standardized language assessments that do persist at the 1 month time point following surgery.

Nearly language measures used in this study (rate of speech/speech sound errors, lexical content, morphosyntax/syntactic complexity, fluency, BNT performance, and AQ performance) were adversely and significantly affected acutely following surgery. Many deficits demonstrated 2-3 days post-surgery resolved by 1 month post-surgery. However, some persisted.

Distortions and phonological errors are still subtly impaired 1 month-post surgery, though words per minute had recovered back to pre-surgical levels in the 27-patient aphasic cohort. This indicates that there are some deficits in the phonological encoding and/or motor speech domains of language. The composite measure of lexical assess impairment yielded indicated deficits 1-month post-surgery in this domain. Proportions of closed class words and verbs recovered to normal, pre-surgical levels in this cohort, indirectly indicating relatively preserved lexical content and reduced prevalence of empty speech. Morphosyntax was the only language domain for which all deficits observed 2-3 days post-surgery resolved by 1 month post-surgery. In regards to fluency and fluency disruptions, retracings and false starts were still significantly more prevalent 1 month post-surgery and filled pauses were marginally still prevalent.

BNT performance and AQ performance returned to normal, pre-surgical levels by 1 month post-operation. Interestingly, quantified connected speech analysis indicates that these standardized assessments of language are insensitive to subtle linguistic deficits still present 1 month post-operation, especially the fluency, motor speech, and lexical access language domains.

A limitation of this data set is that follow-up language assessment was only pursued with patients that qualified as aphasic 2-3 days post-operatively. This is problematic because standardized language assessment has been shown by this study and others to be insensitive to subtle language deficits. Thus, imputed follow-up scores for the non-aphasic patients can be assumed to be underestimation of actual follow-up scores would have been had follow-up been pursued with these patients. This is because the subtle deficits not captured, but likely present would have improved over the course of 1 month following surgery.

We observed some systematic relationships between the location or site of surgical resection and deficits in particular language domains. Our findings suggest that distinct brain regions mediate grammatical and lexical processes, as revealed by selective reduction in function words or lexical words respectively. Resections within the left inferior frontal gyrus were associated with impaired grammar and thus, use of a low proportion of function words. This is to say that patients with grammatical impairments often omit function words such as determiners (e.g. ‘the’, ‘a’) or auxiliary verbs (e.g. ‘is’, ‘was’). Conversely, patients with lexical deficits often use pronouns in the place of nouns (e.g. ‘he’s flying it’ instead of ‘the boy is flying a kite’). Impairments of lexical access were associated with resections of the left middle and inferior temporal gyri in our study.

Our findings are broadly consistent with lesion-symptom mapping in other patient groups, such as primary progressive aphasia (Wilson et al., 2010.) The findings are also consistent with functional imaging studies in normal controls, which have shown frontal activation for function words and temporal activation for lexical words (Friederici et al., 2000) as well as modulation of frontal regions by sentence complexity regardless of lexicality, with modulation of temporal regions exclusively when there is lexical content (Pallier et al., 2011.)

CONCLUSIONS

Transient aphasias are common following resective surgery in the language-dominant left hemisphere. There are systematic relationships between the location of surgical sites and deficits in specific language domains. Almost all language domains are affected adversely in the acute period following resective surgery. Many language deficits resolve within 1 month, but aspects of fluency, motor speech control, and lexical access can remain impaired at 1 month post-operation. Quantified, in-depth connected speech analysis is a valuable means of language assessment because it is more sensitive to language deficits across domains than standardized language assessment. The patient cohort in this study provides a unique window into the neural basis of language. Resections are discrete and their locations are not limited by vascular distribution or patterns of neurodegeneration, reducing the limitations of interpreting the lesion-symptom mapping executed in this study. Furthermore, evaluation of patient language 2–3 days after surgery, allowed us to observe language function before substantial neural reorganization had taken place.

ACKNOWLEDGEMENTS

This research was supported by the National Institutes of Health (R01 DC013270), the Undergraduate Biology Research Program and WAESO. We thank the patients who participated in the research.

REFERENCES

Bates E, Wilson SM, Saygin AP, Dick F, Sereno MI, Knight RT, Dronkers NF. *Voxel-based lesion-symptom mapping*. Nat Neurosci 2003; 6: 448–50.

Chang EF, Raygor KP, Berger MS. *Contemporary model of language organization: an overview for neurosurgeons.* J Neurosurg 2015; 122:250–261.

Benzagmout M, Gatignol P, Duffau H. *Resection of World Health Organization grade II gliomas involving Broca's Area: methodological and functional considerations.* Neurosurg 2007; 61:741–753.

Duffau H, Capelle L, Denvil D, Sichez N, Gatignol P, Lopes M, Mitchell MC, Sichez JP, Van Effenterre R. *Functional recovery after surgical resection of low grade gliomas in eloquent brain: hypothesis of brain compensation.* J Neurol Neurosurg Psychiatry 2003;74:901–907.

Falconer MA, Serafetinides EA. *A follow-up study of surgery in temporal epilepsy.* J Neuro Neurosurg Psychiat 1963; 26: 154-167.

Helmstaedter C, Gleibner U, Zentner J, Elger CE. *Neuropsychological consequences of epilepsy surgery in frontal lobe epilepsy*. Neuropsychol 1998; 36(7): 681-689.

Katz A, Awad IA, Kong AK, Chelune GJ, Naugle RI, Wyllie E, Beauchamp G, Luders H. *Extent of resection in temporal lobectomy for epilepsy: Memory changes and neurologic complications.* Epilepsia 1989; 30(6):763-771.

Loring DW, Meador KJ, Lee GP. *Effects of temporal lobectomy on generative fluency and other language functions.* Arch Clin Neuropsychol 1994; 9: 229–278.

Lubrano V, Draper L, Roux FE. *What makes surgical tumor resection feasible in Broca's Area? Insights into intraoperative brain mapping.* Neurosurgery 2010; 66:868-875.

Friederici AD, Optiz B, von Cramon DY. *Segregating semantic and syntactic aspects of processing in the human brain: an fMRI investigation of different word types.* Cereb Cortex 2000; 10: 698-705.

Pallier C,Devauchelle AD, Dehaene S. *Cortical representation of the constituent structure of sentences*. Proc Natl Acad Sci 2011; 108: 2522-7.

Penfield W, Roberts L. *Speech and Brain Mechanisms*. Princeton, NJ: Princeton University Press, 1959.

Peraud A, Ilmberger J, Reulen HJ. *Surgical resection of gliomas WHO grade II and III located in the opercular region.* Acta Neurochir 2004; 146: 9–18.

Roberts L. *Handedness and Cerebral Dominance.* Transactions of Amer Neurological Assoc 1955; 80: 143-147.

Sanai N, Mirzadeh Z, Berger MS. *Functional outcome after language mapping for glioma resection.* N Engl J Med 2008; 358:18-27.

Wilson SM, Henry ML, Besbris M, Ogar JM, Dronkers NF, Jarrold W, Miller BL, Gorno Tempini ML. *Connected speech production in three variants of primary progressive aphasia*. Brain 2010; 133: 2069–88.

Wilson SM, Lam D, Babiak MC, Perry DW, Shih T, Hess CP, Berger MS, Chang EF. *Transient aphasias after left hemisphere resective surgery*. JNeurosurg 2005; 127:581-93.