Content Word Production during Discourse in Aphasia: Deficits in Word Quantity, Not Lexical–Semantic Complexity

Reem S. W. Alyahya¹,²,³, Ajay D. Halai¹, Paul Conroy⁴, and Matthew A. Lambon Ralph¹

Abstract

Although limited and reduced connected speech production is one, if not the most, prominent feature of aphasia, few studies have examined the properties of content words produced during discourse in aphasia, in comparison to the many investigations of single-word production. In this study, we used a distributional analysis approach to investigate the properties of content word production during discourse by 46 participants spanning a wide range of chronic poststroke aphasia and 20 neurotypical adults, using different stimuli that elicited three discourse genres (descriptive, narrative, and procedural). Initially, we inspected the discourse data with respect to the quantity of production, lexical–semantic diversity, and psycholinguistic features (frequency and imageability) of content words. Subsequently, we created a “lexical–semantic landscape,” which is sensitive to subtle changes and allowed us to evaluate the pattern of changes in discourse production across groups. Relative to neurotypical adults, all persons with aphasia (both fluent and nonfluent) showed significant reduction in the quantity and diversity of production, but the lexical–semantic complexity of word production directly mirrored neurotypical performance. Specifically, persons with aphasia produced the same rate of nouns/verbs, and their discourse samples covered the full range of word frequency and imageability, albeit with reduced word quantity. These findings provide novel evidence that, unlike in other disorders (e.g., semantic dementia), discourse production in poststroke aphasia has relatively preserved lexical–semantic complexity but demonstrates significantly compromised quantity of content word production. Voxel-wise lesion-symptom mapping using both univariate and multivariate approaches revealed left frontal regions particularly the pars opercularis, IC, and central and frontal opercular cortices supporting word retrieval during connected speech, irrespective of word class or their lexical–semantic complexity.

INTRODUCTION

Word retrieval deficits, which are common after brain damage, undermine single-word, sentence, and discourse production, which impact the engagement in conversations. In comparison to the many investigations of single-word production, there have been fewer explorations of the properties of content words produced during discourse, how these might vary across subgroups of aphasia (i.e., an acquired language impairment after brain damage), and whether the results depend on the elicitation stimuli. These enquiries were addressed in the current study.

Discourse production provides a rich data source to assess linguistic content/structure and language use in a more naturalistic context (Dipper & Pritchard, 2017). Different elicitation stimuli tend to elicit different discourse genres, and studies vary in the employed elicitation stimuli. Composite picture description is commonly used in clinical practice and research (e.g., Stark, 2019; Goodglass & Kaplan, 1972), because of advantages in time and consistency of responses across neurotypical adults, resulting in a reliable baseline to compare the language profiles of patient groups. This task, however, might elicit more concrete words compared to other more natural modes of connected speech, such as storytelling narratives (Bird, Howard, & Franklin, 2000). Storytelling based on personal past experiences, based on familiar stories, or using sequential pictures has been employed in research more than clinical settings (Alyahya, Conroy, Halai, & Lambon Ralph, 2021; Stark, 2019; Bird & Franklin, 1996). Personal narratives, however, can result in inconsistent responses at the lexical and semantic levels across and within individuals. Another natural mode of connected speech is to use procedural discourse (description of a familiar task, e.g., changing a car tire; e.g., Alyahya, Halai, Conroy, & Lambon Ralph, 2020; Stark, 2019; Basilakos et al., 2014), although this has been utilized less frequently in the aphasiology literature and clinical practice compared to picture description and personal storytelling, as indicated by a systematic review on studies that utilized discourse samples from people with aphasia (Bryant, 2021 Massachusetts Institute of Technology

© 2021 Massachusetts Institute of Technology

https://doi.org/10.1162/jocn_a_01772
Ferguson, & Spencer, 2016). It has been shown that the elicitation stimuli might result in different language patterns in persons with aphasia. Specifically, there is evidence to indicate that storytelling probes greater word quantity, lexical diversity, and propositional density than picture description in people with aphasia and neurotypical adults (Alyahya et al., 2020b; Stark, 2019). However, no studies explored whether the elicitation stimuli are important to the varying psycholinguistic properties of the words produced during connected speech. This was addressed in the current study.

Discourse production relies heavily on the production of content words, especially of the two main word classes, nouns and verbs. Although persons with aphasia generally present with greater verb deficits compared to nouns, there are strong indications that these noun–verb differences at a single-word level might be attributable to the semantic and psycholinguistic distinctions between them, with verbs tending to be semantically more complex and more abstract than nouns (Alyahya, Halai, Conroy, & Lambon Ralph, 2018a; Vigliocco, Vinson, Drucks, Barber, & Cappa, 2011). Word class effects have also been mapped onto the neural substrates to expand our knowledge of the language system. There is evidence to indicate that nouns and verbs are underpinned by shared neural correlates (e.g., Alyahya et al., 2018a; Crepaldi et al., 2013; Siri et al., 2008). Most studies that compared the neural correlates of nouns and verbs employ single-word tasks (naming or lexical generation). To date, one study has explored the neural correlates of nouns and verbs during picture description in poststroke aphasia using network lesion-symptom mapping. This study suggested that object word production was supported by posterior networks across the occipital, parietal, and posterior inferior temporal regions, whereas action word production was supported by frontal networks (Gleichgerrcht et al., 2016). The study, however, did not account for variations in the lexical–semantic properties of the object and action words.

There is a large body of evidence on the effect of lexical–semantic properties and, particularly, word imageability on single-word production (Alyahya, Halai, Conroy, & Lambon Ralph, 2018b, 2020a; Bird, Howard, & Franklin, 2003) and comprehension in poststroke aphasia (e.g., Alyahya et al., 2018a; Sandberg & Kiran, 2014) with increased accuracy and efficiency in processing concrete over abstract words, including nouns and verbs. Word frequency strongly influences single-word and connected speech production in patients with semantic dementia (Hoffman, Meteyard, & Patterson, 2014; Bird et al., 2000). However, the effect of imageability and frequency during discourse production in aphasia remains relatively unexplored. Studies have reported that people with fluent and nonfluent aphasia mainly use light (high-frequency and semantically empty) verbs (e.g., “go, do”) during sentence production (Berndt, Haendiges, Mitchum, & Sandson, 1997), and they tend to produce more high-frequency and high-imageability verbs compared to neurotypical adults during storytelling (Bird & Franklin, 1996).

To date, the differences in lexical–semantic properties (word imageability and frequency) of nouns and verbs during discourse production have not been examined in poststroke aphasia and across different discourse genres. This might be because of a methodological challenge with discourse data, which is how to quantify and analyze the pattern of vocabulary elicited. Perhaps most commonly, studies have measured deficits in fluency and diversity by coding the number and type of tokens (e.g., Alyahya et al., 2020b; Crepaldi et al., 2011; Bastiaanse & Jonkers, 1998). Other investigations in semantic dementia have adopted a more distributional analysis approach in which the production rate of words at different points along a psycholinguistic continuum is evaluated in controls and patients. This is a method that can be much more sensitive to subtle changes (Hoffman et al., 2014; Bird et al., 2000).

In this study, we adopted this distributional analysis approach to create a “lexical–semantic landscape” of the vocabulary produced during discourse (with respect to word imageability and frequency). We examined how this landscape differs between a large cohort of people with a wide range of poststroke aphasia severity/classifications and a group of neurotypical adults. We also explored differences between subgroups of aphasia: fluent versus nonfluent aphasia as well as high versus low performers on semantic and phonology domains. In addition, we assessed the influence of different discourse stimuli (composite picture description, storytelling narrative, and procedural discourse) on the lexical–semantic properties of the words produced during connected speech. By comparing the lexical–semantic landscape (i.e., production rate across different levels of word frequency and imageability), it is possible to evaluate the pattern of changes in discourse production in people with aphasia and neurotypical controls, specifically, whether there is a reduction in the overall number of words produced and/or change in the shape of the lexical–semantic landscape. This approach has a second useful feature in that it is possible to see how different types of word (e.g., nouns and verbs) are located with respect to different levels of frequency and imageability within the lexical–semantic landscape. Finally, we mapped the neural correlates of word retrieval during connected speech within the lexical–semantic landscape using univariate and multivariate lesion-symptom mapping approaches. Although there are strong advocates for both univariate and multivariate lesion-symptom mapping approaches (Sperber, Wiesen, & Karnath, 2019; DeMarco & Turkeltaub, 2018; Mah, Husain, Rees, & Nachev, 2014; Tyler, Marslen-Wilson, & Stamatakis, 2005; Bates et al., 2003), these alternative analyses tackle different fundamental questions (localization vs. prediction modeling) and have opposite strengths and weaknesses (Hebart & Baker, 2018; Haufe et al., 2014). Previous studies have shown that (a) multivariate
approaches do not always perform better than univariate approaches, (b) different multivariate approaches are not equal, and (c) using both univariate and multivariate approaches is more likely to be complementary as they can be used to answer different questions (Zhao, Halai, & Lambon Ralph, 2020; Schumacher, Halai, & Lambon Ralph, 2019; Sperber et al., 2019). Therefore, we present results from a univariate approach and two multivariate approaches.

METHODS

Participants

Forty-six participants (32 men) who had developed aphasia after a single left hemorrhagic or ischemic stroke were tested in the chronic stage (>12 months poststroke, mean = 69.43 months, SD = 48.86, range = 16–280 months), with a mean age of 63.21 years (SD = 11.93, range = 44–87 years) and mean years of education of 12.65 (SD = 2.59, range = 9–19). The Boston Diagnostic Aphasia Examination (BDAE; Goodglass & Kaplan, 1972) was administered to each participant, and they were diagnosed and classified using the BDAE standard aphasia classification criteria. No restrictions were placed according to aphasia type or severity (spanning from mild anoma to global aphasia), with a mean severity of 2.85 (SD = 1.2, range = 1–5). The exclusion criteria included having suffered more than one stroke or any other neurological conditions, severe motor–speech disorders as described in the participant’s clinical workup, participants who did not produce any response in any of the discourse samples, any contraindications for MRI scanning, and being premorbidly left-handed. All participants were native English speakers with normal or corrected-to-normal vision and/or hearing. Detailed demographic information is presented in Supplementary Table 1. Discourse samples were also collected from 20 age/education-matched healthy/neurotypical right-handed native English-speaking adults (nine men, mean age = 68.85 years, SD = 8.47, range = 57–84 years; years of education: mean = 14, SD = 2.8, range = 9–19). All participants reported no abnormal neurological conditions or history of brain injury. Informed consent was obtained before participation under approval from a local ethics committee.

Discourse Samples: Elicitation, Transcription, and Coding

Three elicitation stimuli were employed with no time limit. First, a relatively simple and commonly used composite picture description was utilized to elicit a descriptive discourse using the “Cookie Theft” picture from the BDAE (Goodglass & Kaplan, 1972). This was included because it is widely used in clinical assessments and research. Second, a storytelling narrative using the “Dinner Party” pictorial script (Fletcher & Birt, 1983) was employed to elicit a more natural mode of connected speech with reduced working memory load compared (because of the presence of picture stimuli) compared to other forms of storytelling narratives. Participants were presented with this series of eight black-and-white picture sequences and were asked to look through them and then describe in detail what was going on in these pictures. Finally, participants were asked to provide a procedural discourse “how they prepare a cup of tea,” which is another natural form of connected speech but without the use of picture stimuli. No prompts or questions were provided throughout the testing by the examiner, except for nonverbal encouragement.

Discourse samples were digitally recorded, then transcribed verbatim (orthographically), and checked against the recording to correct for any discrepancies. This was followed by content analyses carried out on each transcript. Transcription, coding, and analyses were conducted by the first author (R. S. W. A.), a qualified and experienced speech and language pathologist. The following measures were computed from each transcript: (1) content word counts, which included all words that were intelligible, informative, and relevant to the discourse (adapted from Nicholas & Brookshire, 1993). Contractions (e.g., it’s or haven’t) were counted as two separate words; (2) number of nouns and lexical verbs (excluding auxiliary such as “is” in “is going” and modal verbs such as “should” in “should go”) were extracted from the content word count and used as a measure of quantity. The present tense form of the verb “to be” was accepted as a lexical verb (i.e., “is” in “she is happy”); (3) type counts (i.e., the number of distinct words) for nouns and lexical verbs, which were used as a measure of lexical–semantic diversity. Type counts were selected over ratios, because unlike ratios, they do not inflate estimates of lexical diversity (particularly in nonfluent aphasia), as shorter samples produced by persons with severe aphasia can often appear to be richer in diversity because of their higher ratios compared to longer samples. The use of ratios would restrict comparisons between participants and across discourses of different lengths.

The second step was to obtain the lexical–semantic properties (word frequency and imageability) for each noun and verb presented in each transcript (details are provided in Table 2). Frequency values represent lemma frequency per million words (combined written and spoken counts) obtained from the British National Corpus Consortium (2007). Imageability ratings were drawn from the MRC Psycholinguistic Database (Coltheart, 1981) and a corpus of published norms (Bird, Franklin, & Howard, 2001), as these databases include ratings for words within specified word class to disambiguate any noun–verb ambiguous words (e.g., “brush”). Frequency and imageability values were obtained from the respective database using an automated approach (LOOKUP function) implemented in MATLAB (2018a). Frequency values were obtained for all words, and imageability values were...
successfully obtained for 96.3% of the words in all transcripts (see details in Supplementary Table 2). Only words with a defined imageability value were entered into the imageability analyses.

**Statistical Analyses**

Preliminary analyses were conducted to inspect the nature of the discourse data in which the properties of the words produced during discourse stimuli were compared across the three discourse stimuli and between neurotypical adults and persons with aphasia. The comparisons were also conducted within the aphasia group by splitting the group into participants with fluent aphasia (n = 25: anomia, conduction, and transcortical sensory aphasia) versus nonfluent aphasia (n = 21: global, mixed nonfluent, Broca’s, and transcortical mixed aphasia). We expected participants with aphasia to produce fewer words than neurotypical controls and participants with nonfluent aphasia to produce fewer words than those with fluent aphasia. Specifically, we examined differences in the quantity, diversity, and lexical–semantic properties of nouns and verbs, using mixed 2 × 3 × 2 ANOVAs on word count (to measure quantity of production), type count (to measure diversity of production), and imageability and frequency values (measures of lexical–semantic properties). In all ANOVAs, Group (neurotypical vs. persons with aphasia) was entered in the models as the between-participant factor, whereas Discourse (storytelling narrative vs. procedural) and Word class (nouns vs. verbs) were entered as the within-participant factors. All significant interactions were explored using t tests and post hoc analyses and Bonferroni corrected for multiple comparisons. Similar analyses were then conducted, but the between-participant factor was set to persons with fluent versus nonfluent aphasia.

To allow for a more sensitive measure in capturing systematic variations between the groups and across the discourse genres, imageability and frequency distribution patterns were created (Figures 3 and 4). Specifically, the range of word imageability (and frequency) was divided into bands (presented on the x axis of Figures 3 and 4), and the number of words produced within each of these bands was computed for each participant followed by computing the group mean. This distributional analysis approach indicates how often the participants produced words of different imageability and frequency ranges. Subsequently, the imageability and frequency bands were used to generate contour maps (Figures 5 and 6), that is, a “2-D lexical–semantic landscape” of the words produced during discourse. The number of words produced within each of these Frequency × Imageability cell of the contour maps was computed for each participant, and then we calculated the group mean (for both groups of persons with aphasia and neurotypical adults). This was done for the discourse that best captured the imageability and frequency distributions as per the findings of the distributional analysis described above. Further analyses were conducted to examine whether the 2-D lexical–semantic landscapes are consistent across different aphasia subgroups defined in accordance to three orthogonal aphasia-related domains: fluency, semantic ability, and phonological ability. Specifically, we employed a multivariate decomposition algorithm (principal component analysis) on a detailed neuropsychological/aphasiological battery that assessed different aspects of language and cognitive skills, including repetition, naming, word retrieval, fluency, phonemic discrimination, semantic processing, single-word and sentence comprehension, working memory, and executive function. For more details on the neuropsychological battery and our previous work on conceptualizing aphasia as a graded multidimensional space, see Alyahya et al. (2018a). This data-driven approach maximizes the amount of shared variance in a heterogeneous sample and accounts for systematic variations across tests. The principal component analysis generated components related to semantics, phonology, and fluency. We used participants’ component scores along these three components, and the whole aphasia group was then split into high versus low performer subgroups based on the median value of the scores on each of these components.

**Acquisition, Processing, and Analyses of Neuroimaging Data**

High-resolution structural T1-weighted MRI scans were acquired for each patient on a 3.0-T Philips Achieva scanner (Philips Healthcare) using an eight-element SENSE head coil. A T1-weighted inversion recovery sequence with 3-D acquisition was utilized with the following parameters: repetition time = 9.0 msec, echo time = 3.95 msec, acquired voxel size = 1.0 × 1.0 × 1.0 mm³, slice thickness = 1 mm, matrix size = 256 × 256, 150 contiguous slices, flip angle = 8°, field of view = 256 mm, inversion time = 1150 msec, SENSE acceleration factor = 2.5, and total scan acquisition time = 575 sec.

Participants’ structural T1-weighted MRI scans were preprocessed and analyzed with SPM software (SPM12: Wellcome Trust Centre for Neuroimaging, www.fil.ion.ucl.ac.uk/spm/) running under MATLAB (2018a). The images were normalized into standard Montreal Neurological Institute (MNI) space using a modified unified segmentation-normalization tool optimized for focal brain lesions (Seghier, Ramlackhansingh, Crinion, Leff, & Price, 2008). Structural scans from an age- and education-matched control group (18 men and four women; mean age = 69.13 years, SD = 5.85; mean years of education = 13, SD = 2.66) were used as reference to identify abnormal tissue in the group with stroke using a fuzzy clustering with fixed prototypes (FCP) approach. This produces a whole-brain map where each voxel is a probability of abnormality compared to the control group. We applied a binary threshold to this image to obtain a binary lesion...
image (i.e., $U$-threshold = 0.5). The images generated for each patient were visually inspected with respect to the original scan and manually corrected if necessary and were used to generate a lesion overlap map (Figure 1). Images were then smoothed with an 8-mm FWHM Gaussian kernel, to account for the global intraindividual shape differences.

To identify the neural correlates associated with word retrieval during connected speech within the lexical–semantic landscape, we conducted voxel-based correlational methodology (VBCM; Tyler et al., 2005), a variant of voxel-lesion symptom mapping (Bates et al., 2003), which identifies statistical relationships between brain and behavior by correlating the value per voxel (as a continuous variable) with the behavioral performance. We created multiple regression models using the FCP whole-brain images (percent abnormality) with the behavioral measure of interest (noun and verb retrieval within the lexical–semantic landscape) entered as regressors. Lesion volume (estimated using the automated lesion identification tool; Seghier et al., 2008) and demographic variables (age, months after stroke onset, and years of formal education) were entered in the models as covariates to control for participants’ variabilities on these variables and because of the importance of these demographic variables for the population with stroke. Specifically, age and education are related to brain health and general executive/language performance, and time poststroke can predict language performance in people with poststroke aphasia (Hope et al., 2017). We used the statistical nonparametric mapping toolbox (Version 13.1.08; warwick.ac.uk/tenichols/snpm) to create each multiparticipant model and 5000 permutations where implemented to determine statistical inference. The results were thresholded at $p < .005$ voxel level and cluster corrected using FWE at $p < .05$.

To supplement the univariate analysis, we conducted multivariate analyses using two approaches. First, we used support-vector regression lesion symptom mapping (SVR-LSM) toolbox (DeMarco & Turkeltuab, 2018). In this approach, we loaded the binary lesion images as the features, created a separate model for noun and verb retrieval within the lexical–semantic landscape, and included demographic variables (age, education, time post-onset) in all models as covariates. The following settings were used: MATLAB SVM implementation, hyperparameter optimization (Bayes optimization with default settings) and lesion threshold = 3. The resulting beta weights were evaluated by permutation testing ($n = 10,000$) and thresholded at voxel-wise $p < .005$ and cluster-wise $p < .05$. We ran each behavioral model twice, with and without correction for lesion volume using the “regress on both” option as recommended by DeMarco and Turkeltuab (2018). Second, we used the pattern recognition of neuroimaging toolbox (PRoNTo V2.1, www.mini.cs.ucl.ac.uk/pronto; Schrouff et al., 2013) as an alternative multivariate approach because it formally evaluates model predictions and does not truncate beta weights post hoc. For this approach, we entered the FCP (percent abnormality) images as continuous values, which quantify the amount of abnormality at each voxel across the whole brain including both the lesioned and intact hemispheres. We created separate models for noun versus verb retrieval within the lexical–semantic landscape, with demographic variables (age, education, time postonset) entered in all models as covariates. We followed the pipeline through in two pathways: (i) restricted to lesion territory (similar to SVR-LSM) and (ii) using the whole brain as input (similar to the VBCM). Data from the whole brain were included in case there were any meaningful variations that might otherwise be missed when using a lesion mask. The models were estimated using the relevance vector regression (Tipping, 2001) implementation as this method is quick and does not require hyperparameter optimization. PRoNTo relies on kernel methods to overcome the high dimensionality problem in neuroimaging (using $n \times n$ pairwise similarity matrix), and features were mean centered. A 10-fold cross-validation scheme was used to determine model performance (i.e., trained on 90% of the data and tested on 10% the held-out set). We also report $r$ values that represent the cross-validated correlation value of the models. The correlation is between the observed (true) scores and the model estimated (predicted) scores in the left-out cases. Model inference, specifically, is evaluated using permutation testing whereby the observed values are shuffled randomly and the prediction model is recalculated to obtain a distribution of model performance by chance ($n = 5$ iterations). The real cross-validated correlation is then compared to the
null distribution to determine significance with a $p < .05$ alpha threshold. As with the SVR-LSM, we ran each model with and without lesion volume as a covariate. The anatomical labels were obtained using Harvard-Oxford atlas (Desikan et al., 2006).

**RESULTS**

Table 1 provides information on the discourse data produced by the neurotypical and aphasia groups.

**The Properties of the Words Produced during Discourse**

Descriptive statistics on the quantity and diversity of nouns and verbs produced during discourse are reported in Table 1 and illustrated in Figure 2. Descriptive statistics on the frequency and imageability values are reported in Table 2 for the groups of neurotypical adults, all persons with aphasia, persons with fluent aphasia, and persons with nonfluent aphasia across the three discourse genres. Results from the $2 \times 3 \times 2$ ANOVAs are reported in Table 3. In summary, the findings indicated, as expected, that (1) persons with aphasia produced less quantity and diversity of nouns and verbs compared to neurotypical adults across all three discourse genres (storytelling narrative, descriptive, and procedural). Furthermore, there were more words (especially nouns) in terms of quantity and diversity produced during the storytelling narrative compared to the other two discourse genres. The group of persons with aphasia produced more verbs than nouns during the procedural discourse; (2) the average imageability values for nouns were significantly higher than those for verbs, and this was consistent between the neurotypical and aphasia groups and across the three discourse genres; and (3) the average frequency values did not differ for the two word classes and between the neurotypical and aphasia groups across the three discourse genres; and (4) the group of persons with fluent aphasia produced greater quantity and diversity of nouns and verbs compared to the group of persons with nonfluent aphasia across the three discourse genres (storytelling narrative, descriptive, and procedural). Furthermore, there was greater production of words (especially nouns) in terms of quantity and diversity during storytelling narratives compared to the other two discourse genres. The persons with fluent aphasia produced more verbs than nouns during the procedural discourse. Details are provided in Supplementary Results.

**Impact of Lexical–Semantic Properties on Content Words Produced during Discourse**

*Imageability Effects*

*Neurotypical adults versus aphasia groups.* The imageability distribution pattern used by the neurotypical

---

**Table 1. Descriptive Statistics of the Measures Extracted from the Three Discourses Produced by the Neurotypical and Aphasia Groups**

<table>
<thead>
<tr>
<th>Discourse Genre</th>
<th>Storytelling Narrative</th>
<th>Descriptive</th>
<th>Procedural</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measure</td>
<td>Range</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Neurotypical Group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of tokens</td>
<td>101–706</td>
<td>205.3</td>
<td>140.4</td>
</tr>
<tr>
<td>Duration (sec)</td>
<td>43–322</td>
<td>128</td>
<td>61.4</td>
</tr>
<tr>
<td>Quantity of verb</td>
<td>16–152</td>
<td>42.4</td>
<td>28.4</td>
</tr>
<tr>
<td>Diversity of verbs</td>
<td>15–47</td>
<td>25.1</td>
<td>8.6</td>
</tr>
<tr>
<td>Quantity of nouns</td>
<td>20–144</td>
<td>56</td>
<td>28.9</td>
</tr>
<tr>
<td>Diversity of nouns</td>
<td>13–74</td>
<td>32.4</td>
<td>15.2</td>
</tr>
</tbody>
</table>

Aphasia Group

| Number of tokens | 8–454 | 156.7 | 114.53 | 6–315 | 66.57 | 59.70 | 5–262 | 59.02 | 57.74 |
| Duration (sec)   | 42–620 | 190.04 | 109.78 | 11–310 | 95.09 | 66.22 | 10–219 | 61.95 | 42.66 |
| Quantity of verb | 0–114 | 18.80 | 17.58 | 0–52 | 07.86 | 07.41 | 0–38 | 08.00 | 07.82 |
| Diversity of verbs | 0–49 | 12.52 | 09.99 | 0–23 | 05.76 | 04.49 | 0–21 | 06.09 | 05.42 |
| Quantity of nouns | 0–66 | 23.89 | 17.85 | 0–40 | 08.5 | 07.02 | 0–30 | 08.24 | 06.60 |
| Diversity of nouns | 0–43 | 15.58 | 10.97 | 0–24 | 07.23 | 05.17 | 0–17 | 05.41 | 03.39 |
Figure 2. The quantity and diversity of nouns and verbs produced during discourse. Bar graphs showing the group mean and standard errors (error bars) of the quantity of nouns and verbs (turquoise bars) and the diversity of nouns and verbs (purple bars) produced during three different discourse genres among the (A) neurotypical adults and all persons with aphasia and (B) the fluent and nonfluent aphasia subgroups.

Table 2. Mean (SD) of Frequency and Imageability of the Words Produced by the Neurotypical Adults and Persons with Aphasia across the Three Discourse Genres

<table>
<thead>
<tr>
<th>Word Type</th>
<th>Discourse Group</th>
<th>Frequencya</th>
<th>Imageabilityb</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Storytelling</td>
<td>Descriptive</td>
<td>Procedural</td>
</tr>
<tr>
<td>All</td>
<td>Neurotypical</td>
<td>2.34 (0.13)</td>
<td>2.34 (0.19)</td>
</tr>
<tr>
<td>All</td>
<td>All persons with aphasia</td>
<td>2.48 (0.49)</td>
<td>2.17 (0.70)</td>
</tr>
<tr>
<td>Fluent aphasia</td>
<td>2.44 (0.22)</td>
<td>2.14 (0.07)</td>
<td>2.04 (0.04)</td>
</tr>
<tr>
<td>Nonfluent aphasia</td>
<td>2.56 (0.65)</td>
<td>2.26 (0.84)</td>
<td>2.00 (0.60)</td>
</tr>
<tr>
<td>Nouns</td>
<td>Neurotypical</td>
<td>1.89 (0.08)</td>
<td>1.64 (0.11)</td>
</tr>
<tr>
<td>All</td>
<td>All persons with aphasia</td>
<td>1.92 (0.27)</td>
<td>1.69 (0.49)</td>
</tr>
<tr>
<td>Fluent aphasia</td>
<td>1.96 (0.19)</td>
<td>1.69 (0.36)</td>
<td>1.60 (0.3)</td>
</tr>
<tr>
<td>Nonfluent aphasia</td>
<td>1.86 (0.35)</td>
<td>1.69 (0.65)</td>
<td>1.70 (0.28)</td>
</tr>
<tr>
<td>Verbs</td>
<td>Neurotypical</td>
<td>2.79 (0.25)</td>
<td>3.04 (0.32)</td>
</tr>
<tr>
<td>All</td>
<td>All persons with aphasia</td>
<td>2.92 (0.59)</td>
<td>2.71 (0.8)</td>
</tr>
<tr>
<td>Fluent aphasia</td>
<td>2.91 (0.38)</td>
<td>2.71 (0.69)</td>
<td>2.49 (0.43)</td>
</tr>
<tr>
<td>Nonfluent aphasia</td>
<td>2.93 (0.65)</td>
<td>2.81 (0.8)</td>
<td>2.48 (0.87)</td>
</tr>
</tbody>
</table>

a Lemma frequency that were log-transformed.

b Imageability ratings on a scale from 100 to 700.
Table 3. Findings from ANOVAs on the Effects of Group, Discourse Genre, Word Class, and Their Interactions on the Quantity, Diversity, and Psycholinguistic Properties (Imageability and Frequency) of the Words Produced during Discourse

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Quantity (Word Count)</th>
<th>Diversity (Type Count)</th>
<th>Mean Imageability Values</th>
<th>Mean Frequency Values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F(1, 64) = 36.46, p &lt; .0001, (\eta^2 = .36): neurotypical &gt; persons with aphasia</td>
<td>F(1, 64) = 53.69, p &lt; .0001, (\eta^2 = .46): neurotypical &gt; persons with aphasia</td>
<td>ns</td>
<td>ns</td>
<td></td>
</tr>
<tr>
<td><strong>Discourse</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F(2, 128) = 80.32, p &lt; .0001, (\eta^2 = .56): narrative &gt; picture description and procedural</td>
<td>F(2, 128) = 59.43, p &lt; .0001, (\eta^2 = .48): narrative &gt; picture description and procedural</td>
<td>ns</td>
<td>ns</td>
<td></td>
</tr>
<tr>
<td><strong>Word class</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F(1, 64) = 26.3, p &lt; .0001, (\eta^2 = .29): nouns &gt; verbs</td>
<td>F(1, 64) = 9.82, p = .003, (\eta^2 = .13): nouns &gt; verbs</td>
<td>F(1, 53) = 1083.63, p &lt; .0001, (\eta^2 = .95): nouns &gt; verbs</td>
<td>ns</td>
<td></td>
</tr>
<tr>
<td><strong>Group × Discourse</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F(2, 128) = 11.4, p &lt; .0001, (\eta^2 = .15): narrative by neurotypical group &gt; aphasia group</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
<td></td>
</tr>
<tr>
<td><strong>Group × Word Class</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ns</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Discourse × Word Class</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F(2, 128) = 20.06, p &lt; .0001, (\eta^2 = .24): nouns &gt; verbs during narrative</td>
<td>F(2, 128) = 13.99, p &lt; .0001, (\eta^2 = .18): nouns &gt; verbs during narrative</td>
<td>ns</td>
<td>ns</td>
<td></td>
</tr>
<tr>
<td><strong>Group × Discourse × Word Class</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ns</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Significant effects \((p < .01)\) of 2 (group: neurotypical vs. aphasia) × 3 (discourse: narrative vs. picture description vs. procedural) × 2 (word class: nouns vs. verbs) mixed ANOVA. All significant main and interaction effects were further explored using post hoc \(t\) tests and Bonferroni corrected for multiple comparisons \((p < .01)\). ns = not significant.
group (Figure 3A) is fairly similar across the three discourse genres, with more words produced with high imageability (501–600) and followed by low imageability (<300 and 300–400). It is apparent that this distribution was affected by the word class, in which there was (i) increased production of concrete frequently used nouns (501–600 and 601–700, e.g., "door, plate, milk"), but no production of nouns with low imageability in all three discourse genres, and (ii) increased production of semantically "light" and abstract verbs with lower imageability (<300 and 301–400, e.g., "think, decide, go, get, put"), followed by mid-imageability verbs (401–500 and 501–600, e.g., "lay, ask, fill") and very few high-imageability verbs that were produced in the storytelling narrative only (601–700, e.g., "ring"). Results from a 3 discourse genre × 2 word class × 5 imageability band repeated-measures ANOVA are reported in Supplementary Results.

The imageability distribution used by persons with aphasia (Figure 3B) was similar in pattern to the distribution by the neurotypical group for the three discourse genres and with the same word class distinction. The only difference being that the aphasia group produced fewer words than the neurotypical group in all bands across the three discourse genres (note the scale in Figure 3). Results from a 3 discourse genre × 2 word class × 5 imageability band repeated-measures ANOVA are reported in Supplementary Results. These results are similar to those of the neurotypical adults. A direct comparison between the neurotypical and aphasia groups was conducted using a mixed ANOVA and reported in the Supplementary Results.

**Fluent versus nonfluent aphasia subgroups.** Figure 3C and D illustrates the imageability distribution related to the content words produced during the three discourse genres by the fluent versus nonfluent aphasia subgroups. Details of the analyses comparing the two groups are reported in Supplementary Results and indicated that the nonfluent aphasia group produced fewer words in all imageability bands across the three discourse genres than the fluent group while maintaining a similar pattern of distribution.

**Frequency Effects**

**Neurotypical adults versus aphasia groups.** Figure 4A illustrates the frequency distribution of the neurotypical group in all three discourse genres, and it indicated that more words were produced within mid-frequency values (1–2.5) for the storytelling narrative and picture description, whereas more words were produced within a lower frequency band (<1) for procedural discourse. Specifically, there were more nouns produced within mid-frequency (2–2.5, e.g., "fish, floor, minutes") in storytelling narrative and picture description as well as more nouns within lower frequency (<1) in procedural discourse, which represent prompt-specific compound vocabulary (e.g., "teabag,

![Figure 3](http://direct.mit.edu/jocn/article-pdf/doi/10.1162/jocn_a_01772/1957807/jocn_a_01772.pdf)
teapot”), and thus this might not be generalizable to all procedural discourses. There were very few nouns produced in the high-frequency bands (2.5–3 and 3–3.5) and no nouns at the highest frequency band (>3.5) across all discourse genres. On the other hand, there were more verbs produced within higher frequency bands (>3, e.g., “see, can, do”) and a limited number of verbs produced within low- and mid-frequency bands. Results from a 3 discourse genre × 2 word class × 7 frequency band repeated-measures ANOVA are reported in Supplementary Results.

The frequency distribution for the persons with aphasia group (Figure 4B) was similar in shape to the frequency distribution used by the neurotypical group across the three discourse genres, with the same word class distinction, while producing fewer words in all bands. Results from a 3 discourse genre × 2 word class × 7 frequency band repeated-measures ANOVA are reported in Supplementary Results. The data between neurotypical and aphasia groups were directly compared using a mixed ANOVA and reported in the Supplementary Results.

Fluent versus nonfluent aphasia subgroups. Figure 4C and D illustrated the frequency distribution related to the content words produced during the three discourse genres by the fluent versus nonfluent aphasia subgroups. Details of the analyses between the two groups are reported in Supplementary Results and indicated that the nonfluent aphasia group produced fewer words in all bands across the three discourse genres than the fluent group while maintaining a relatively similar pattern of distribution.

Lexical–Semantic Landscape

Neurotypical adults versus aphasia groups. The constructed contour maps (Figure 5) demonstrate that the lexical–semantic landscape of the words produced during discourse by the aphasia group was broadly similar to the one of the neurotypical adult group. Both maps contain two peaks: (i) at the top of the map and represents words with high imageability and mid-frequency and (ii) at the bottom right side corner of the map and represents words with low imageability and high frequency. Interestingly, by constructing contour maps for each word class, it becomes clear that each peak predominantly reflects different word classes for both groups: The space for nouns consists of one peak within the high-imageability and mid-frequency range, which contains commonly used everyday nouns (e.g., “chair, cat, bedroom, restaurant”) with a lack of abstract nouns. Conversely, the contour maps for the verbs, in both groups, consist of one main peak for low-imageability and high-frequency words, which contain regularly used light verbs (e.g., “go, have, get, make, do, want, be”) and abstract emotional and cognitive verbs (e.g., “think, assume, like”). The verb contour map also includes another smaller peak for words with high imageability and mid-frequency (i.e., overlapping with the peak for nouns). This part of the space contains concrete, frequently used verbs (e.g., “ring, eat”) and might be influenced by task demands. Overall, there was very low use of low-frequency abstract words. The low-frequency abstract

![Figure 4](http://example.com/fig4.png)

**Figure 4.** Frequency distribution of the content words produced during different discourse genres. The group mean of the number of words produced within each frequency band by (A) neurotypical adult group, (B) persons with aphasia group, (C) persons with fluent aphasia subgroup, and (D) persons with nonfluent aphasia subgroup. Top row: nouns and verbs combined; center row: nouns; bottom row: verbs. Error bars represent SEM.
verb peak becomes broader in the neurotypical group, extending toward concrete verbs with mid-frequency (e.g., "concentrate, surprise, admire"), although the number of verbs produced in these ranges was very small. To visualize the similarities/differences in the lexical–semantic landscapes and to determine statistical differences, we generated subtraction maps between the neurotypical and aphasia groups. This subtraction map was constructed by performing t tests between the lexical–semantic landscapes of the two groups on each Imageability × Frequency cell (Figure 5) with a Bonferroni correction ($p \leq .001$). The subtraction maps indicated areas of the lexical–semantic landscape that were used significantly more often by the neurotypical group compared to the aphasia group (colored green on Figure 5). The results indicated that the difference between groups reflected a global reduction in the number of words produced, rather than a reshaping of the space.

**Aphasia subgroups.** The resultant contour maps from the aphasia subgroups (Figure 6) demonstrated two main characteristics that are consistent with the other results. First, irrespective of how the persons with aphasia were subdivided (by fluency, semantic, or phonological domains), each subgroup still utilized the same parts of the lexical–semantic landscape and differences between the maps reflected a generalized reduction in the number of words produced by the low-performer groups rather than a change in the shape of the lexical–semantic landscape. Second, the largest differences between the maps resulted from splitting the groups of persons with aphasia according to fluency, whereas there were only modest differences between the groups split on the two other orthogonal aphasia domains (semantics and phonology).

**Neuroimaging Results**

To identify the neural correlates associated with word retrieval during connected speech accounting for the lexical–semantic property of the word, we identified the two peaks within the neurotypical adults' lexical–semantic landscape (representing a noun peak and a verb peak). Then, we computed the number of words produced by each person with aphasia at these peaks and used these values as the measure of interest in the univariate and multivariate lesion-symptom mapping analyses. Correlation analyses conducted between the number of words produced at these peaks and word count revealed significant strong positive correlations (noun: $r = .76$, verb: $r = .89$, $p < .0001$). The VBCM results controlled for lesion volume and demographic variables (time after stroke onset, age, education; Figure 7A) revealed overlapping left frontal regions for verb (extend =
1579 voxels) and noun (extend = 950 voxels) retrieval within the lexical–semantic landscape. The overlap includes the left inferior frontal gyrus (pars opercularis), precentral gyrus, central opercular cortex, and IC. The cluster related to verbs extended to the postcentral gyrus; however, a direct contrast between noun versus verb retrieval within the lexical–semantic landscape revealed no differences in either direction.

The multivariate analyses yielded relatively similar results. The SVR-LSM results controlled for lesion volume and demographic variables (Figure 7B) identified significant clusters for both noun and verb retrieval within the lexical–semantic landscape that were only voxel-wise significant. The clusters involved the central opercular cortex. The cluster associated with verbs further extends to include the IC and the inferior frontal gyrus (pars opercularis). When lesion volume was removed from the model, the SVR-LSM produced significant clusters that were voxel-wise and cluster-wise significant for both nouns and verbs. The significant clusters associated with verb and noun retrieval within the lexical–semantic landscape were overlapping in the left inferior frontal gyrus (pars opercularis and pars triangularis), frontal and central opercular cortices, IC, and precentral and postcentral gyri. For noun retrieval, the cluster further extended to include the frontal orbital cortex, anterior superior temporal gyrus, and planum polare. Figure 7 shows a high degree of similarity between the VBCM and SVR-LSM results.

The PRoNTo approach revealed significant brain–behavior relationships for all models except for noun retrieval within the lexical–semantic landscape when controlled for lesion volume correction. There were no differences between models using the whole brain or restricted lesion space approaches. Specifically, for the models controlled for lesion volume and demographic variables, the cross-validated correlation coefficient for verb retrieval within the lexical–semantic landscape using the whole brain and restricted lesion territory was $r = .37$ and $r = .46$, respectively. On the other hand, the cross-validated correlation coefficient for the models without lesion volume correction using the whole brain was $r = .48$ and $r = .48$ for nouns and verbs, respectively, and

Figure 6. Contour maps for different aphasia subgroups representing lexical–semantic landscapes. Two-dimensional Frequency × Imageability landscapes showing the mean word count produced during storytelling narrative by (A) persons with fluent versus nonfluent aphasia, (B) high versus low performers on the semantic domain, and (C) high versus low performers on the phonological domain. $t$ Tests were used to show differences between the two groups in the third column, and significantly different parts of the spaces are shown in green ($p \leq .001$).
using the restricted lesion territory, it was $r = .51$ and $r = .54$ for nouns and verbs, respectively. For all significant models, we projected the beta weights to brain space and displayed the results in Figure 7C and D. The negative beta values from the PRoNTo results, in general, converged with results from both VBCM and SVR-LSM.

In conclusion, the convergent findings from the different lesion-symptom mapping approaches indicated that left frontal regions covering the inferior frontal gyrus (pars opercularis and pars triangularis), frontal and central opercular cortices, IC, and precentral gyrus are associated with word retrieval during connected speech within the lexical–semantic landscape.

**DISCUSSION**

Among the collection of variable symptoms in poststroke aphasia, at least some degree of reduced connected speech production is a very common feature. Accordingly, it is important to be able to assess and understand the nature of this core aphasiological symptom. Most language production assessment tools used in research and clinical examinations rely heavily on single-word naming and picture description, which is not an everyday form of communication and connected speech production. In this detailed investigation on a large number of persons spanning the full range of severity of chronic poststroke aphasia, we assessed the lexical–semantic properties of content words produced during connected speech. We explored this across different discourse genres to examine if the elicitation stimuli change the obtained results. We also compared the performance of the aphasia group to an age/education-matched group of neurotypical controls.

Across all three discourse genres, we found that, as expected, persons with aphasia produced a lower number of words than controls, they still sampled the same range of word frequency and imageability as that used by the neurotypical adults. This pattern was consistent across the entire range of persons with aphasia, including fluent and nonfluent subgroups, and also subgroups defined by different levels of semantic and phonological abilities. This provides evidence that persons with aphasia can produce words with the same lexical–semantic complexity as neurotypical adults, albeit at a reduced word quantity. Persons with aphasia also showed reduced quantity and diversity of retrieval for different word classes (nouns and verbs) compared to neurotypical adults across different discourse genres. Below, we discuss these results.

The Nature of Vocabulary Produced during Discourse in Poststroke Aphasia

It is striking that, although the quantity of vocabulary was reduced in fluent aphasia and even further in nonfluent aphasia, the pattern and types of words being produced were similar to those observed in neurotypical controls. This is not the case in all clinical populations; previous examinations of connected speech in semantic dementia have found that the sampling of the 2-D Frequency × Imageability landscape was not only lower than neurotypical adults but also...
The pattern of results with regard to the lexical–semantic properties of the words produced during discourse by the neurotypical group replicated previous findings from a different sample of neurotypical controls (Bird et al., 2000). In this study, persons with aphasia produced fewer verbs and nouns in terms of quantity and diversity than neurotypical adults, consistent with previous studies (Hodges, Patterson, Oxbury, & Funnell, 1992; Snowden, Goulding, & Neary, 1989). This degraded semantic input to the language production system leads to profound anomia (Lambon Ralph, McClelland, Patterson, Galton, & Hodges, 2001) and semantic degradation (Hoffman et al., 2014; Lambon Ralph, Graham, Ellis, & Hodges, 1998). Accordingly, it seems very likely that the low-frequency words used in connected speech simply have too little semantic input to drive their retrieval. By contrast, poststroke aphasia does not appear to involve the same type of core semantic degradation as that found in semantic dementia (Jefferies & Lambon Ralph, 2006) but instead reflects damage to other language components (Alyahya et al., 2020b). In terms of classical models of language production, semantic dementia would reflect impairment at the conceptualization phase, whereas the breakdown in poststroke aphasia is either after the semantic stage of the language production pathway and/or in other aspects related to fluent connected speech (e.g., in the mechanisms that control phrasal construction and the interface between lexical retrieval and sentence formation; Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Garrett, 1988). Preservation of conceptual input presumably means that persons with aphasia attempt to sample the entire breadth of the lexical–semantic landscape but the likelihood of any of these target words being produced is made (equally) less likely by post-semantic-production deficits. This conclusion would seem to align with an observation that many persons with aphasia state—namely, that they know what they want to say but cannot produce it.

The Relationship between Imageability, Frequency, and Word Class

The relationship of the lexical–semantic landscape but the likelihood of any of these target words being produced is made (equally) less likely by post-semantic-production deficits. This conclusion would seem to align with an observation that many persons with aphasia state—namely, that they know what they want to say but cannot produce it.

The Influence of Different Types of Discourse Elicitation Stimuli

The pattern of results with regard to the lexical–semantic properties of the words produced during discourse by the neurotypical group replicated previous findings from a different sample of neurotypical controls (Bird et al., 2000). In this study, persons with aphasia produced fewer verbs and nouns in terms of quantity and diversity than neurotypical adults, consistent with previous studies based on smaller sample sizes (Bastiaanse & Jonkers, 1998; Berndt et al., 1997). The present results also complement another study that indicated intact verb production with respect to semantic category and weight during connected speech in aphasia (Cruice, Pritchard, & Dipper, 2014). As with previous studies on a different patient group (i.e., semantic dementia; Hoffman et al., 2014; Bird et al., 2000), the distribution-based analysis provides a very clear and unique way to explore the results obtained from clinical populations. As well as revealing striking differences between poststroke aphasia and semantic dementia (see above), these analyses also reveal the inescapable relationship between word classes and lexical–semantic properties. Specifically, the lexical–semantic landscape contained many high-imageability nouns with mid-frequency (e.g., “phone, girl”) and high-frequency verbs with low imageability (e.g., “cook, change”) across all discourse genres. This is consistent with small-sample studies showing that patients with fluent and nonfluent aphasia used high-frequency light verbs in their narratives (Berndt et al., 1997) and fluent aphasia sampling word frequency in a similar manner to controls (Bastiaanse, 2011). The corollary of these differential distributions for nouns and verbs is that any apparent word-class effect needs to be treated with caution, as they could reflect the inherent differences driven by word imageability and/or frequency. This resonates with previous documents on the influence of semantic distinctions between nouns and verbs on single-word production in aphasia (e.g., Alyahya et al., 2018a; Bird et al., 2000).

The picture-supported storytelling narrative, which is not commonly used in research and clinical practice, boosted connected speech production including eliciting a greater quantity and diversity of words, both nouns and verbs, and the full range of word frequency and imageability. Indeed, the entire frequency and imageability ranges (including the word class distinction) were considerably higher for this elicitation stimulus. This production boost was found in both the neurotypical controls and person with aphasia (both fluent and nonfluent). These results are consistent with recent studies indicating the main effect of discourse elicitation stimuli in which the storytelling narrative had an advantage over other stimuli in eliciting greater quantity, diversity, words per minute, and propositional density (Alyahya et al., 2020b; Stark, 2019). We suspect that there are multiple factors that contributed to the success of storytelling narrative. First, in this study, persons with aphasia were not given a time limit for discourse production, and they took longer to narrate the story compared to the time spent on the other discourse stimuli. Presumably, this extended time led to more overt production (note that quantity does not necessarily follow duration: relative to neurotypical controls, the persons with aphasia produced less output in their storytelling narratives but spent much longer doing so). Second, the stimuli used in this study to elicit storytelling narrative had many characters and events and thus offer many more opportunities and topics for language production. Third, the storytelling narrative was picture supported, which could help as the storyboard provides direct prompts about items and actions to describe...
Neural Correlates of Word Retrieval during Connected Speech

The neural correlates associated with word retrieval (irrespective of the word class or lexical–semantic complexity) during connected speech mainly encompassed left frontal regions covering the inferior frontal gyrus, IC, fronto-central operculum, and pre-central gyrus. These regions were identified using both univariate and multivariate lesion-symptom mapping approaches. These findings are highly consistent with previous lesion studies and fMRI experiments in healthy controls, which associated these left frontal regions to fluency in aphasia, propositional speech production in neurotypical adults, and motor–speech planning (Alyahya et al., 2020b; Basilakos, Rorden, Bonilha, Moser, & Fridriksson, 2015; Blank, Scott, Murphy, Warburton, & Wise, 2002; Dronkers, 1996). A study that utilized network lesion-symptom mapping to explore the neural correlates of nouns and verbs during picture description in aphasia also found action word production to be supported by frontal networks (Gleichgerrcht et al., 2016).

From a methodological point of view, it is important to note the complementary differences between the interpretation of univariate and multivariate analyses (Hebart & Baker, 2018). Generally, univariate analyses assign beta values to voxels in a relatively transparent way, in which the strength and sign of these values indicate meaningful brain–behavior relationships. Therefore, it is easier to make inferences about localization of functions. However, there are practical challenges with univariate methods that must be accounted for, such as correcting for multiple comparisons. There are also theoretical concerns, including assumption of voxel independence and mislocalization of effects (DeMarco & Turkeltaub, 2018; Karnath, Sperber, & Rorden, 2018; Mah et al., 2014). On the other hand, multivariate approaches are inherently different. Specifically, they can be used to make formal behavioral predictions via mapping brain status to behavioral performance, and they are commonly used for encoding or decoding (Hebart & Baker, 2018; Naselaris, Kay, Nishimoto, & Gallant, 2011). One important limitation with multivariate approaches is how to interpret the feature weights (Hebart & Baker, 2018; Haufl et al., 2014). There are potential strategies that might aid improving interpretability, which include encoding methods (such as partial least squares and canonical correlation analysis) or bootstrap analyses (Kuceyeski et al., 2016), but this is nontrivial. Finally, multivariate decoding approaches typically require a large data set, as data are partitioned into training versus test sets for cross-validation. This can be practically challenging, particularly for patient studies where recruitment is difficult. In a recent simulation study (Sperber et al., 2019), it was suggested that a sample size of 100 participants is required to produce stable and reproducible beta parameter mapping, whereas the sample size for prediction of clinical outcomes peaked at 40 and was relatively stable from this point up to 100 participants. Given the differences between various brain-mapping approaches, it is striking that the multivariate analyses (both SVR-LSM and PRoNTo) converged with the VBCM findings in this study.

Conclusion

We explored the properties of content words produced during connected speech across different discourse genres in persons with poststroke aphasia and compared their performance to an age/education-matched group of neurotypical controls. The results provided evidence that persons with aphasia can produce words with the same lexical–semantic complexity as neurotypical adults, albeit at a reduced word quantity. Across all three discourse genres, we found that, although the quantity of words produced by persons with aphasia was lower than that by neurotypical controls, they still sampled the same range of word frequency and imageability as that used by the neurotypical adults. This pattern was consistent across the entire ranges of the aphasia group, including subgroups of fluent and nonfluent aphasia and those with high versus low semantic and phonological abilities. Findings from this study also indicated that caution should be taken when selecting discourse elicitation stimuli, as different stimuli lead to differences in the lexical–semantic properties of the produced discourse. Picture-supported storytelling narrative, which is not commonly used in research and clinical practice, can boost connected speech production including eliciting a greater quantity and diversity of words across all classes and the full range of word frequency and imageability. For clinical assessments, this is important as the other commonly utilized elicitation stimuli might sometimes
underestimate the production abilities and the range of vocabulary used by patients. For research investigations, maximizing discourse output will inevitably make the assessments more sensitive and better able to grade differences between patients. Neuroimaging findings from both univariate and multivariate approaches revealed shared left frontal regions in association with word retrieval, irrespective of their word class and lexical–semantic complexity, during connected speech. This would imply that interventions such as noninvasive brain stimulation can target these left frontal brain regions to enhance word retrieval regardless of the word’s class or complexity.

Acknowledgments

We are indebted to all participants and carers for their contribution in this research study. We thank Blanca De Dios Perez for her help in obtaining the frequency and imageability values. R. S. W. A. was supported by King Fahad Medical City fellowship. M. A. L. R. was supported by an ERC Advanced grant (GAP: 670428) and MRC intramural funding (MC_UU_00005/18).

Reprint requests should be sent to Reem S. W. Alyahya, Communication and Swallowing Disorders Department, King Fahad Medical City, P.O. Box 59046, Riyadh 11525, Kingdom of Saudi Arabia, or via e-mail: rsalyahya@gmail.com.

Author Contributions

Reem S. W. Alyahya: Conceptualization; Formal analysis; Investigation; Methodology; Validation; Visualization; Writing—Original draft; Writing—Review & editing. Ayaj D. Halai: Conceptualization; Formal analysis; Methodology; Validation; Writing—Review & editing. Paul Conroy: Conceptualization; Methodology; Validation; Writing—Review & editing. Matthew A. Lambon Ralph: Conceptualization; Methodology; Validation; Writing—Review & editing.

Funding Information

Matthew A. Lambon Ralph, ERC Advanced Grant, grant number: 670428. Matthew A. Lambon Ralph, MRC Intramural Funding, grant number: MC_UU_00005/18. Reem S. W. Alyahya, King Fahad Medical City (https://dx.doi.org/10.13039/501100011680).

Diversity in Citation Practices

A retrospective analysis of the citations in every article published in this journal from 2010 to 2020 has revealed a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience (JoCN)* during this period were M(an)/M = .408, W(oman)/M = .335, M/W = .108, and W/W = .149, the comparable proportions for the articles that these authorship teams cited were M/M = .579, W/M = .243, M/W = .102, and W/W = .076 (Fulvio et al., JoCN, 33:1, pp. 3–7). Consequently, JoCN encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article’s gender citation balance.

REFERENCES


