Which processes are crucial for effective sentence production? Behavioural and lesion-mapping evidence from chronic aphasia

Carolyn E. Wilshire and Paula Speer

School of Psychology, Victoria University of Wellington, New Zealand.

Correspondence to: Carolyn Wilshire, School of Psychology, Victoria University of Wellington, P.O. Box 600, Wellington, New Zealand.

Email: [Carolyn.Wilshire@vuw.ac.nz](mailto:Carolyn.Wilshire@vuw.ac.nz)

**Abstract**

In this study, we assessed eight individuals with aphasia on seven measures that have previously been associated with damage to anterior language structures. Voxel-based lesion-symptom mapping was used to explore their neural correlates. We identified two distinct clusters of measures. The first, *fluency and grammaticality*, included measures of speech rate, articulatory agility and grammatical word content. Poor scores on these measures were associated with damage to the left insula and central operculum. The second cluster, *context sensitivity,* included a measure of the difference in accuracy on subject and object nouns during sentence production, and also a measure of the semantic interference effect in the cyclic naming task. Poor scores were associated with damage to the left inferior and middle frontal gyri, and the postcentral and precentral gyri. Based on these results, we propose a two-factor account of the speech difficulties observed in nonfluent aphasia. The first factor is a general slowing of speech planning processes, most notably articulatory-motor programming, and is associated with damage to the left insula and central operculum. The second factor is a difficulty with top-down control of language, associated with left inferior and middle frontal gyrus damage. We suggest this latter difficulty is a consequence of problems generating and maintaining a robust message-level representation to drive the language element selection process.

**Keywords: Aphasia, Nonfluent aphasia, Sentence production, Left inferior frontal gyrus, Voxel-based lesion-symptom mapping.**

Which processes are crucial for effective sentence production? Behavioural and lesion-mapping evidence from chronic aphasia

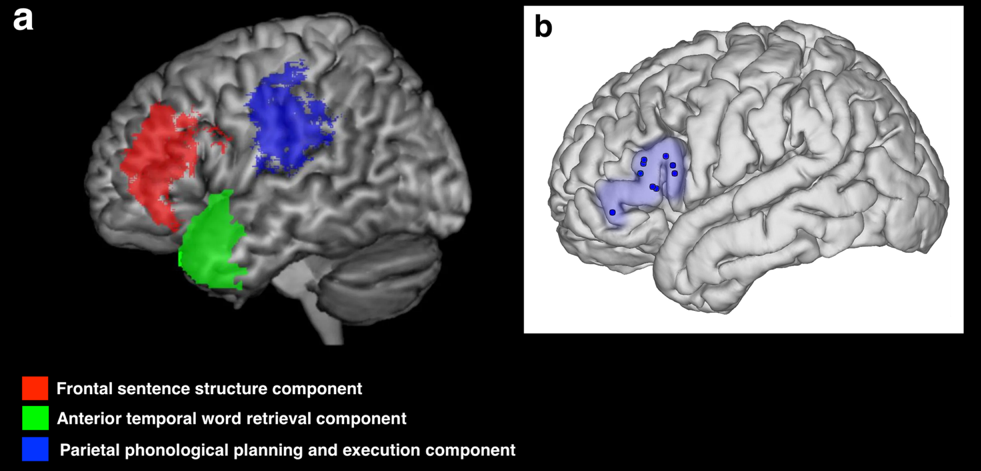
**1. Introduction**

Lesion mapping studies of neuropsychological samples have found that the ability to produce fluent, well-formed sentences is dependent upon the integrity of an extensive left perisylvian network. Poor scores on measures of speech fluency and grammatical wellformedness have been consistently associated with pathology in anterior language regions including the left inferior frontal gyrus (LIFG), the insula and precentral gyrus, and also in more posterior regions such as the inferior parietal lobe and the posterior superior temporal gyrus, and this is the case even after factoring out variance due to volume size (stroke: Basilakos et al., 2014; de Ouden et al., 2019; Ding et al., 2020; Halai et al. 2017, 2018; Lacey et al., 2017; Mirman et al., 2019; primary progressive aphasia: Rogalski et al., 2011). Indeed, the lesion maps obtained are remarkably similar, irrespective of whether the key measure emphasises fluency (Borovsky et al., 2007; Halai et al., 2017, 2018; Lacey et al., 2017; Rogalski et al., 2011; Wilson et al., 2010; Yourganov et al., 2016) or some aspect of grammatical wellformedness (e.g., the proportion of words produced in complete sentences during speech, or the prevalence of grammatical function words; den Ouden et al., 2019; Faroqi-Shah et al., 2014; Mirman et al., 2019; Rogalski et al, 2011; Wilson et al., 2010; but see Ding et al., 2020 for evidence for some dissociation between these two measures in large acute stroke sample).

The extensiveness of these lesion maps is perhaps not surprising, given that the production of fluent, well-formed speech is likely to depend upon a whole host of language processes that are not unique to sentences. In order to isolate the regions whose role is more specific to sentence production, some researchers have attempted to minimise demands on language processes not of central interest. For example, Rogalski et al. (2011) asked participants to order a series of word cards into a sequence that described a pictured action. Again using a pictured action description task, den Ouden et al (2019) presented participants with action scenes which were labelled with the key nouns and verb; participants had to generate the full sentence (see also Faroqi-Shah et al., 2014). However, the findings using such methods have been inconsistent. Rogalski et al. (2011) found that, in a sample of primary progressive aphasia patients, poor scores on their card ordering task were associated with reduced cortical thickness in an extensive region, including the left superior and inferior frontal gyri, and extending into the precentral and postcentral gyri and the supramarginal gyrus. Perhaps even more surprisingly, in a chronic stroke sample, den Ouden et al. (2019) found that poor scores on their lexically-supported sentence production task were associated with pathology restricted entirely to the left temporal lobe (most notably, the posterior superior temporal gyrus). It may be that these kinds of modified tasks encourage the use of rather idiosyncratic strategies that are not utilised in more naturalistic settings.

An alternative approach is to use more open-ended speech tasks, but factor out any variability associated with processes not of interest. Mirman et al. (2019) calculated a measure of the proportion of words produced in sentences during open-ended speech, and then factored out variance associated with scores on a measure of articulatory ability (the number of articulatory errors produced during a single word repetition task). Scores on this residualised measure were found to be associated with pathology largely restricted to the LIFG and middle frontal gyrus, a region the authors dubbed the *frontal sentence structure component*. As shown in Figure 1, this region showed no overlap with the areas that have previously been implicated in: a) lexical-phonological retrieval and phonological processing; and b) semantically-driven word retrieval.

Some converging support for Mirman et al.’s notion of a left prefrontal “sentence building” component comes from a recent electrocortical stimulation study of 14 awake craniotomy patients (Chang, Kurteff & Wilson, 2018). The patients had to describe pictured actions using simple sentences (e.g., *The boy is kissing the girl*). In seven of the 14 patients, stimulation of key sites in the LIFG (most commonly, the pars opercularis), elicited syntactic errors (e.g. “**the girl is kissing the…the boy is kissing the, the girl”)**, without compromising performance on counting, picture naming or repetition. Figure 1 shows the precise points of stimulation that elicited this effect.



**Figure 1**

Panel a) shows the location of the three components of the fluent speech production system proposed by Mirman et al. (2019). Panel b) shows the cortical sites where stimulation led to syntactic errors, but did not compromise counting, picture naming or repetition. The left inferior frontal gyrus is highlighted in pale blue (adapted from Chang, Kurteff & Wilson, 2018).

Taken together, this evidence suggests that the left lateral prefrontal cortex and/or its underlying white matter play a role in sentence generation specifically. However, it is not yet clear which specific cognitive processes are supported by these regions and why are they so critical for producing *sentences* in particular. To address this question, we must turn to cognitive models of sentence production. Most such models identify two broad classes of processes that are necessary for converting a conceptual message into a linear sequence of lexical elements: 1) retrieving the appropriate lexical items that denote the conceptual elements of the message to be expressed; and 2) constructing a structural representation of the sentence, which specifies the position of each lexical item, how it should be inflected and which grammatical function words it should contain (Bock, 1986; Garrett, 1976; Levelt et al., 1999; Dell, 1986; Ferreira & Slevc, 2007; Pickering. & Branigan, 1999; Stemberger, 1985). Early models viewed these two sets of processes as independent, their outputs being brought together at a relatively late stage (e.g., Garrett, 1976). More recent models allow for considerable interplay between the two. For example, some propose that when individual lexical items are retrieved/activated, they in turn activate representations of the syntactic roles they are capable of occupying (Bock, 1986, Stemberger, 1985). Consequently, many of the elements of the syntactic “frame” of a sentence are already specified as a result of lexical retrieval, and just need to be integrated into a full structural representation of the entire sentence. Some models assign particular importance to verbs in this respect, proposing that verb representations include extensive information about the types of syntactic structures they can be embedded into (e.g., the types of obligatory arguments they must take; Jackendoff, 2002; Pickering & Branigan, 1999). Consequently, in these models, much of the work of generating a sentence plan takes place during the retrieval of its main verb, and does not rely as heavily on a dedicated structure building process.

Several recent computational models go one step further: they propose that generating a syntactic plan for a sentence may be a relatively low-level process that is shaped primarily by learned sequential dependencies between different classes of words and other constituents (Chang, 2002; Chang, Dell & Bock, 2006; Dell & Chang, 2014). Chang et al. (2006) trained a simple recurrent network PDP model to generate structurally sound sentences based on learned dependencies between different word classes. The model was sensitive to both immediate and more distal dependencies. In this model, the selection of any given content word was constrained not only by the sequencing rules, but also by the message element(s) that are most highly activated at that time. Message elements consist of temporary pairings between the key conceptual elements that make up the intended message and their corresponding thematic roles (e.g., *cat-agent, chase-action, mouse-patient*). Within this type of framework, the most computationally demanding aspect of sentence planning is not generating a syntactic plan, but rather creating and maintaining a robust message-level representation to “drive” the sentence planning process.

Some of these cognitive perspectives call into question the idea that there may be a neuroanatomically distinct “sentence structure building” component. Instead, they raise the possibility that brain regions previously associated with sentence planning may perform an entirely different function. One possibility is that damage to left lateral prefrontal cortex compromises the ability to generate a robust representation of the conceptual proposition to be expressed. This idea would be consistent with evidence suggesting that the LIFG plays a crucial role in the formation of a message intention (e.g., Robinson et al., 2005; 2010).

Another possibility is that lateral prefrontal regions support more general processes that are crucial for fluent, well-formed sentence production (for further discussion, see Nozari & Faroqi-Shah, 2017). For example, they may be involved in maintaining multiple lexical elements in an activated state, until those elements are ready to be integrated into larger, phrasal units (Freedman, Martin & Biegler, 2004; Martin, Lesch, & Bartha, 1999; Martin, Miller, & Vu, 2004; Martin & Romani, 1994; Martin, Shelton, & Yaffee, 1994). Alternatively, they may be crucial for managing competition between concurrently activated sentence elements, so that the correct element can be produced at the correct time (Schnur, 2017; Scott & Wilshire, 2012; Speer & Wilshire, 2013; Thothathiri, Schwartz & Thompson-Schill, 2010; Thothathiri, Gagliardi & Schwartz, 2012; Wilshire & McCarthy, 2002). There is some support for this second suggestion. Damage to anterior language regions, most notably the LIFG, has been associated with marked difficulties selecting a lexical target from amongst strongly activated competitors. For example, individuals with damage here may respond abnormally slowly on the Stroop colour-word task (Gläscher et al., 2012; Tsuchida & Fellows, 2013). They may also exhibit exaggerated semantic interference effects in picture naming tasks where the pictures are repeatedly sampled from a small set (“cyclic” naming tasks: Anders et al., 2017; Riès et al., 2014, 2015; Schnur et al., 2005, 2009). One study observed a steady decline in picture naming accuracy with each successive cycle through the set, a finding that was not observed on sets of unrelated pictures (Schnur et al., 2005, 2009; see also Anders et al., 2017). Damage to the LIFG specifically was associated with this pattern of performance.

Finally, effective sentence production also requires synchronisation of all relevant processes. For example, lexical elements need to be retrieved sufficiently rapidly to ensure that each is available at the moment when it needs to be integrated into the sentence plan (Hartsuiker & Kolk, 1998; Kolk & Van Grunsven, 1985; Kolk, 1995, 2006; see also Goldmann et al., 2001). The influence of phonological context on the sound errors of a speaker with Wernicke's aphasia. Brain and Language, 78(3), 279-307.). Delays arising further downstream – for example, during the construction of an articulatory-motor plan - may lead to similar consequences. Indeed, it has been suggested that outright grammatical errors, such as omissions of function words, might actually reflect a compensatory strategy designed to minimise the impact of processing delays on overall speech output, while maximising information content (De Roo, Kolk, & Hofstede, 2003; Ruiter, Kolk, & Rietveld, 2010; Salis & Edwards, 2004; Sahraoui & Nespoulous, 2012).

In sum, a number of quite diverse types of impairments might manifest themselves as disproportionately poor sentence production, and several of these have previously been associated with anterior language structures. So establishing the cognitive contribution of anterior regions like the middle and inferior frontal gyri is far from straightforward. These regions could be directly involved in “structure building” in some capacity. Alternatively, they could play a critical role in generating or maintaining propositional representations, in maintaining lexical information online during planning, or in managing competition between concurrently activated lexical representations. Finally, damage to these regions may simply create delays, which may impact disproportionately on sentence-level speech production. Not all of these possibilities are necessarily mutually exclusive.

In order to explore some of these possibilities in participants with aphasia, Speer and Wilshire (2013) used a task in which sentence structure was kept constant and only lexical content was varied. Five individuals with nonfluent aphasia (whose damage extended to the LIFG) and four individuals with fluent aphasia viewed pictured scenes like those in Figure 2 and had to describe them. The pictures disappeared from view at the commencement of their response. The pictures all consistently elicited simple subject-verb-object sentences when presented to healthy participants. All nouns were within the participants’ expressive vocabulary, as assessed on a single picture naming task. Speer and Wilshire found that nonfluent participants’ accuracy fell away very dramatically after production of the first subject noun: nouns in object position were produced much less accurately than those in subject position. Also, those with nonfluent aphasia were less accurate at producing the target sentence if it commenced with a low frequency subject noun than with a high frequency noun. Neither of these phenomena was observed in the fluent aphasia participants. This pattern of performance is consistent with a problem constructing or maintaining a robust message-level representation that is capable of “driving” selection of early sentence elements. Without this top-down control, the person’s performance becomes more heavily shaped by bottom-up factors, such as how rapidly and efficiently they can activate the candidate lexical elements.



**Figure 2**

Examples of the pictured scenes used in the sentence production tasks used in Speer and Wilshire (2013).

In a second study, Speer and Wilshire manipulated the degree of semantic relatedness between the subject and object nouns in the target sentence. The nonfluent participants were less accurate at producing sentences containing semantically related nouns than those with unrelated nouns, whereas the fluent participants showed a trend in the *opposite* direction, towards facilitation. A measure of this semantic relatedness effect was strongly correlated with a measure of the subject frequency effect in the first study, suggesting that both effects might arise from the same underlying impairment. The authors concluded that without a robust representation of the target message to be expressed – specifically, a representation of its propositional content - individuals cannot effectively manage competition between concurrently activated lexical items destined for that utterance.

Speer and Wilshire’s (2013) study primarily compared participants based on their broad aphasia classifications (fluent vs. nonfluent aphasia). They did not systematically examine the associations between performance and lesion location. Also, they did not examine whether the observed semantic relatedness effects were unique to sentence production contexts, or whether they reflected a more general difficulty managing competition. The current study aims to address these questions. First, we examine whether the individuals who exhibited semantic interference effects in Speer and Wilshire’s (2013) sentence production task were also vulnerable to semantic interference in a single picture naming task. Specifically, we used the blocked cyclic naming task, in which the pictures to be named are repeatedly sampled from a small pool (Belke, Meyer & Damian, 2005; Kroll and Stewart, 1994; Python Glize & Laganaro, 2018; Schnur et al., 2005, 2009; Wilshire & McCarthy, 2002). Following Schnur et al.’s (2005) observation that damage to the LIFG was associated with a marked semantic interference effect that increased across successive cycles through the picture list, we calculated a measure of this relatedness-by-cycle interaction for each individual participant. We then examined the interrelationships between these experimental measures and several commonly-used measures of speech fluency, articulatory quality and grammatical wellformedness. Second, using data from eight of the nine individuals studied, we examined the lesion correlates of these various measures.

## 2. Method

**2.1 Participants**

Eight individuals with aphasia were recruited from community stroke groups. All had suffered a stroke more than 12 months prior to the study, and were classified as having mild to moderate aphasia according to the Boston Diagnostic Aphasia Examination (BDAE-3) severity scale (Goodglass et al., 2001). To qualify for the study, participants needed to score at least 50% correct on the Boston Naming test (long form; Goodglass et al., 2001). Table 1 presents background information for each participant. Table S1 in the Supplementary materials provides additional neuropsychological test scores, and Figure S1 presents individual lesion maps.

**2.2 Behavioural Tasks**

**2.2.1. Elicitation of open-ended speech.** We obtained two samples of open-ended speech. The first was obtained using the BDAE-3 protocol, and consists of a sample of the individual’s open-ended conversational speech and also their description of a complex pictured scene (Goodglass et al., 2001). The second was obtained QPA protocol, and is based on the person’s recount of a well-known fairy tale (Saffran, Berndt & Schwartz, 1989).

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Initials | BY | DA | JG | JHM | RP | NP | STR | SW | WL |
| **General information:** |  |  |  |  |  |  |  |  |  |
| Age | 59 | 71 | 73 | 52 | 66 | 73 | 81 | 82 | 64 |
| Gender | M | M | F | F | M | M | F | F | M |
| Years post CVA | 37 | 11 | 6 | 10 | 9 | 13 | 13 | 4 | 2 |
| **BDAE measures**: | | | | | | | | | |
| Articulatory agility score | 50 | 40 | 40 | 10 | 30 | 70 | 85 | 50 | 70 |
| Fluency percentile | 60 | 12 | 23 | 15 | 25 | 93 | 100 | 88 | 100 |
| Boston Naming Test (/60) | 41 | 53 | 46 | 46 | 38 | 32 | 53 | 37 | 40 |
| Aphasia Diagnosis | BA | BA | BA | BA | BA | AA | AA | WA | CA |
| **Narrative speech (QPA) measures**: | | | | | | | | | |
| Words per minute | 67 | 19 | 69 | 48 | 22 | 103 | 153 | 96 | 126 |
| Proportion closed class items | 0.64 | 0.40 | 0.47 | 0.37 | 0.31 | 0.58 | 0.61 | 0.54 | 0.66 |
| **Sentence measures (from Speer &Wilshire, 2013)** | | | | | | | | | |
| Sentence context effect | -5.73 | -6.25 | -17.19 | -16.67 | -7.82 | 2.60 | -8.33 | -1.56 | -5.21 |
| Subject frequency effect | -10.85 | -7.17 | -16.73 | -15.13 | -13.36 | 2.33 | 9.62 | 7.97 | 9.56 |
| Sentence relatedness effect | -7.14 | -6.88 | -21.30 | -25.66 | -18.78 | 3.97 | 9.79 | 12.96 | 22.09 |
| **Blocked cyclic naming task measure** | | | | | | | | | |
| Relatedness x cycle interaction | -0.28 | 1.20 | -7.50 | -4.58 | 1.65 | 1.39 | 1.81 | 0.56 | -1.39 |

BA = Broca’s aphasia; WA = Wernicke’s aphasia; CA= conduction aphasia; AA = anomic aphasia

BDAE = The Boston Diagnostic Aphasia Examination (Goodglass et al., 2001).

QPA = The Quantitative Production Analysis (Saffran, Berndt & Schwartz, 1989).

**Table 1**

Background information for each participant and scores on key measures.

**2.2.2. Sentence production tasks.** Full details of these tasks are provided in Speer and Wilshire (2013), and will only be summarised here. In Speer and Wilshire’s Experiment 1, the stimuli consisted of 66 pictured scenes depicting simple interactions between two participants, which could be described using SVO sentences. The scenes had undergone a comprehensive norming procedure, and all had elicited at least 80% sentence naming agreement in a group of nonaphasic controls (not penalising for appropriate alternative verbs, e.g., eating -> *biting*). Examples of picture stimuli and their target sentences are shown in Figure 2. In all target sentences, the subject and object nouns were monosyllabic and of either high or low frequency. Every noun appeared once in a sentence as a subject noun and once as an object noun. To prevent syntactic priming, these 66 pictured scenes were randomly interspersed with an additional 66 filler pictures which elicited sentences with a different syntactic structure, and different noun elements.

In Speer and Wilshire’s Experiment 2, the stimuli consisted of 64 scenes similar to those used in Experiment 1, except that the semantic relationship between the subject and object nouns was systematically varied. In *related* sentences, both nouns were drawn from the same broad semantic category - either animals (e.g., *bear, skunk*) or real/imagined persons (e.g., *king, ghost*). In *unrelated* sentences, the subject and object nouns were drawn from different categories. A total of 28 related and 36 unrelated sentences passed our norming procedure. Word frequency was balanced across the related and unrelated sentences. To prevent syntactic priming, these pictured scenes were randomly interspersed with an additional 64 filler pictures of the kind used in Experiment 1.

Both tasks were administered as follows. First, participants completed a block of single picture naming trials in which they were asked to name individual pictures that depicted the nouns used in the target sentences. Then, in a separate session, they completed the sentence task. They were instructed to describe what was happening in each picture in one sentence. Each trial began with a fixation cross that remained visible for 1000ms, followed by the picture, accompanied by a tone. As soon as the participant began to vocalise their response, the experimenter manually pressed a key and the picture immediately disappeared from the screen. This was done in order to prevent a strategy of simple labelling of the pictured items, and also to promote advance planning of the entire sentence, thereby more closely replicating the demands of everyday speech. However, patient SW was unable to produce any of the target sentences once the picture had disappeared from the screen, so the procedure was modified to enable her to view the picture throughout her response.

Each participant completed each task three times in separate sessions, each time with the pictured scenes appearing in a different order, and each time preceded by eight practice items (for JHM and DA, who worked slowly and fatigued easily, the testing was spread over six sessions). A response was scored as correct if it included the two target nouns and the target verb (or a permissible variant), and these were incorporated into a thematically and grammatically correct sentence. Where there was more than one attempt at the sentence, only the first attempt was scored. Each of the constituent nouns was also scored correct or incorrect. Complete omissions of the target noun or substitution of the noun with a pronoun (e.g., *it* instead of *pig*) were counted as errors.

**2.2.3. Blocked cyclic naming task (BCNT).** This task was administered to participants within six months of the other tasks. The task is described in Scott and Wilshire (2010), and will only be summarised here.The stimuli consisted of 24 sets of six pictures depicting objects, plants or animals, all of which achieved name agreement of at least 80% when presented to healthy participants. In 12 of the sets, the items were from the same semantic category (e.g., all were animals, or all were items of furniture); in the remaining 12 sets, they were unrelated to each other. Each individual picture was used once in a semantically related set and once in an unrelated set.

Each of these six-picture sets formed the basis for a single block of naming trials. In these blocks, each picture was presented individually, then again in a different order, then again until the whole set had been cycled through four times. Each naming trial began with a warning beep, followed 100 ms later by the picture. The participant's response to the picture automatically triggered a voice activated timing device that initiated the next naming trial 700 ms later (thus, the response stimulus interval was 800ms).The 24 blocks of naming trials were presented in pseudorandom order over two testing sessions.

Prior to the task proper (and in a separate session), participants completed a standard naming task featuring the 72 picture targets that would subsequently appear in the experiment. If they produced any errors or alternative names (e.g., *crescent* for *moon*), they were told what name was usually given to the picture. A naming trial was scored as correct if the first response provided was the target noun. The only exception was when a participant systematically and consistently used a permissible alternative throughout all naming trials (for example, *mare* for *horse*).

**2.3. Behavioural measures obtained**

We calculated the following measures for each participant:

*1. Proportion of closed class items.* Based on the sample of narrative speech collected using the QPA protocol (the retelling of a fairy tale), we calculated the proportion of closed class items, as specified in that protocol.

*2. Speech rate.* Again, using the QPA narrative speech sample and its associated scoring protocol, we calculated the rate of speech in words per minute.

*3. Articulatory agility.* We rated the articulatory quality of the sample of open-ended speech collected using the BDAE protocol (Goodglass et al., 2001). This protocol instructs the rater to note instances where the person’s articulation sounded distorted, clumsy or effortful. The final rating is based on the frequency of such abnormalities. Ratings were obtained by an experienced assessor (CW)*.* The values used in our analyses were the scaled scores.

*4. Sentence context effect.* Using responses from Speer and Wilshire’s (2013) Experiment 1, we calculated the difference in the percentage of errors involving subject and object nouns respectively. A negative score indicates disproportionately poor object noun production.

*5. Sentence frequency effect.* Using participants’ responses from Speer and Wilshire’s (2013) Experiment 1, we calculated the difference in the percentage of errors involving sentences commencing with a low frequency subject and high frequency subject respectively. A negative score indicates disproportionately poor accuracy on sentences with a *low* frequency subject.

*6. Sentence relatedness effect.* Using participants’ responses from Speer and Wilshire’s (2013) Experiment 2, we calculated the difference in the percentage of incorrect responses for related and unrelated sentences. A negative score indicates poorer accuracy on related than on unrelated sentences, and a positive score indicates the opposite.

*7. BCNT relatedness by cycle interaction.* Using participant’s responses on the Blocked cyclic naming task (BCNT), we: a) calculated the difference in the percentage of correct naming trials on related and unrelated picture for each cycle; and b) determining the slope of the line that expressed the change in difference scores across successive cycles. Negative scores indicate increasing semantic interference across successive cycles.

**2.4. Lesion mapping methods**

All participants underwent a magnetic resonance imaging scan specifically for this study, within one year of being tested on the behavioural measures. Whole brain T1-weighted as well as T2-weighted structural scans with inversion recovery (FLAIR) were collected at 1.5 Tesla (T1 3D FFE: TR = 25 ms, TE = 4.6 ms, FOV = 252mm x 238mm, slice thickness = 1mm; Sagittal 3D FLAIR: TR = 4800ms, TE = 329ms, TI = 1660ms, FOV = 252mm x a250mm, slice thickness = 1mm). Each participant’s lesion map was manually segmented onto a 1x1x1mm T1-weighted structural image of their brain using Micron (Rorden et al., 2007), using the T2-weighted FLAIR image for additional guidance. Lesion boundaries were identified on the basis of visible differences in signal characteristics between the pathological and normal brain tissue. The scans and lesion maps were then normalised to a standard template in SPM8 (Ashburner et al., 2012), implemented in MATLAB 8.0 (the MathWorks Inc, Natick MA) using the MR segment-normalise algorithm (Rorden et al., 2012), which uses a variant of cost-function masking, and a standard template based on healthy individuals with a mean age of 65 years.

VLSM analyses were performed using the Lesymap package (Pustina, 2018). We performed a linear regression analysis at each voxel, in which the dependent variable was the language function score, expressed as a residual after regressing out variance shared with overall lesion size, and the independent variable was the pathological status of that voxel, adjusted for overall lesion size. To adjust for lesion size at the voxel level, we weighted the score for each voxel’s status according to the size of that individual’s entire region of pathology. This was done by assigning each damaged voxel the value of 1 (as per convention) and then dividing that value by the square root of the participant’s overall lesion volume (see DeMarco and Turkeltaub, 2018; Mirman et al., 2015). Voxels were only included in the analysis if it was lesioned in at least two participants.

We corrected for multiple comparisons using the continuous permutation-based FWER method (CFWER) described in Mirman et al. (2018). It has recently been suggested that standard permutation-based FWER, which defines the statistical threshold as the lowest value that is absent in 95% of the randomly permuted data sets (based on alpha = 0.05), may be overly conservative (Mirman et al., 2018). CFWER allows the researcher to loosen this threshold: for example, one can define the statistical threshold as the lowest value that is present in no more than five voxels in 95% of the randomly permuted data sets (*v* =5), or in no more than 10 voxels (*v* = 10), and so on. Mirman et al. (2018) recommend reporting results based on a range of values of v, so that the reader can assess the reliability of the voxel evidence for themselves. Each correction was calculated from a sample of 10,000 permutations. The Harvard-Oxford cortical atlas (Desikan et al., 2006) was used to determine the cortical localisation of voxels exceeding the significance threshold. The HCP1065 white matter atlas provided in FSL was used to determine overlap with the territories of the major association pathways. This atlas was developed from tractography data from 1065 healthy participants, and created using the XTRACT software package (Warrington et al., 2020).

**3. Results**

**3.1. Behavioural Results**

Table S2 presents the complete data for each participant on the blocked cyclic naming task. Table 1 summarises participants’ scores for each of the seven key behavioural measures. As predicted, all measures tended to be positively associated with one another, although not all associations were statistically significant for this small sample (see Table 3). It is of note that the sentence relatedness and the subject frequency measures were very highly correlated. So too were the QPA closed class, QPA speech rate and articulatory agility measures.

To further explore the relationships amongst these measures, we performed a principal component analysis with varimax rotation. Components with Eigenvalues above 1 were extracted. The purpose of this analysis was not to examine the components themselves (the sample was not large enough for this), but to identify clusters of measures that are likely to capture unique versus overlapping variability. The analysis resulted in two principal components, and the results are shown in Table 4. The closed class, speech rate and articulatory agility measures loaded primarily on the first component, which we will call *fluency and grammaticality*. Conversely, the sentence context effect and BCNT interaction measure loaded primarily on the second component, which we will call *context sensitivity*. The remaining measures loaded on both factors.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | QPA speech rate | QPA closed class propn | Articulation rating | Sentence context effect | Subject freq. effect | Sentence relatedness effect | BCNT relatedness x cycle |
| QPA speech rate | — | 0.820\*\* | 0.846\*\* | 0.294 | 0.813\*\* | 0.763\* | 0.134 |
| QPA closed class propn | 0.820\*\* | — | 0.815\*\* | 0.298 | 0.676\* | 0.766\* | 0.159 |
| Articulation rating | 0.846\*\* | 0.815\*\* | — | 0.457 | 0.818\*\* | 0.826\*\* | 0.460 |
| Sentence context effect | 0.294 | 0.298 | 0.457 | — | 0.749\* | 0.728\* | 0.785\* |
| Subject freq. effect | 0.813\*\* | 0.676\* | 0.818\*\* | 0.749\* | — | 0.963\*\*\* | 0.532 |
| Sentence relatedness effect | 0.763\* | 0.766\* | 0.826\*\* | 0.728\* | 0.963\*\*\* | — | 0.500 |
| BCNT relatedness \* cycle | 0.134 | 0.159 | 0.460 | 0.785\* | 0.532 | 0.500 | — |

\**p*<.05; \*\**p* < .01; \*\*\* *p* < .001.

**Table 3**

Correlation matrix for the seven key behavioural measures.

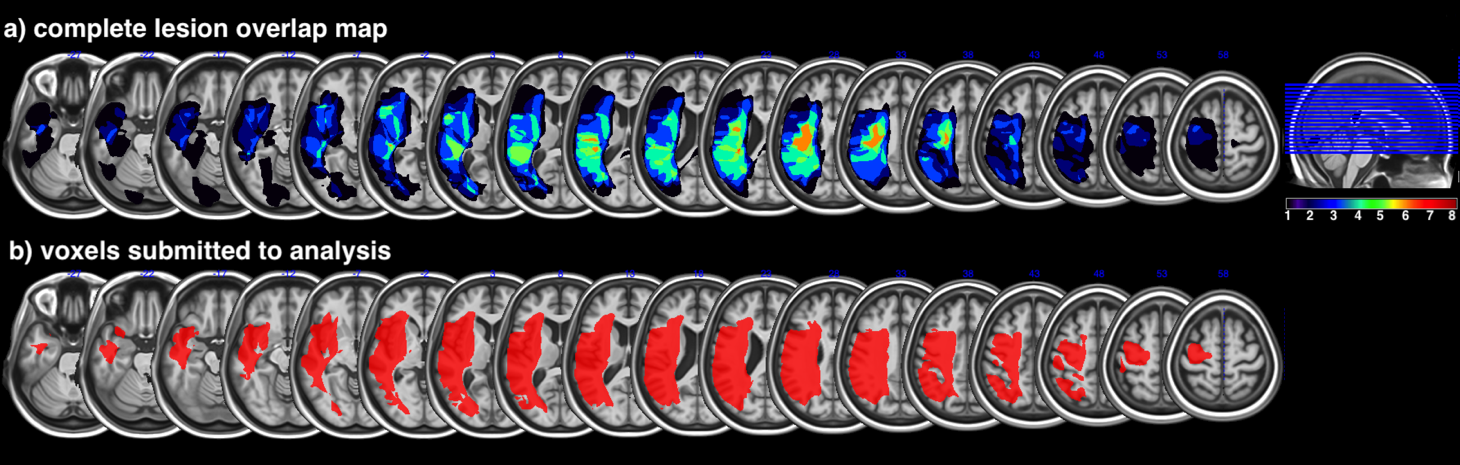
|  |  |  |
| --- | --- | --- |
|  | Component | |
|  | *Fluency & grammaticality* | *Context sensitivity* |
| QPA speech rate | 0.958 |  |
| QPA closed class proportion | 0.925 |  |
| Articulatory agility rating | 0.873 |  |
| Sentence context effect |  | 0.922 |
| Subject frequency effect | 0.774 | 0.577 |
| Sentence relatedness effect | 0.797 | 0.544 |
| BCNT relatedness x cycle |  | 0.928 |
| Total variance accounted for | 54.8% | 35.1% |

**Table 4**

Principal component analysis for the key behavioural measures. Loadings with absolute values of less than 5 are not shown.

**3.2. Lesion analysis results**

Panel a of Figure 2 displays the lesion overlap map for all eight participants who took part in the analysis, projected onto our template for healthy older controls. No voxel was lesioned in more than six out of eight people. Panel b displays the voxels that met our criterion for inclusion in the VLSM analysis – that is, those where at least two participants had damage. Table S2 in the supplementary materials describes the localisation of these voxels according to our cortical and white matter atlases. To summarise, the map covered substantial portions of left temporal, parietal, frontal and insular cortex and its underlying white matter (including the majority of the territories of the arcuate fasciculus, the middle longitudinal fasciculus the superior longitudinal fasciculus, as well as substantial portions of the inferior fronto-occipital fasciculus, the inferior longitudinal fasciculus, and the frontal aslant tract). However, it excluded a sizeable region in the ventral temporal lobe, the rostral prefrontal cortex, and medial portions of both the frontal and parietal lobes, and most of the left occipital lobe.



**Figure 2**

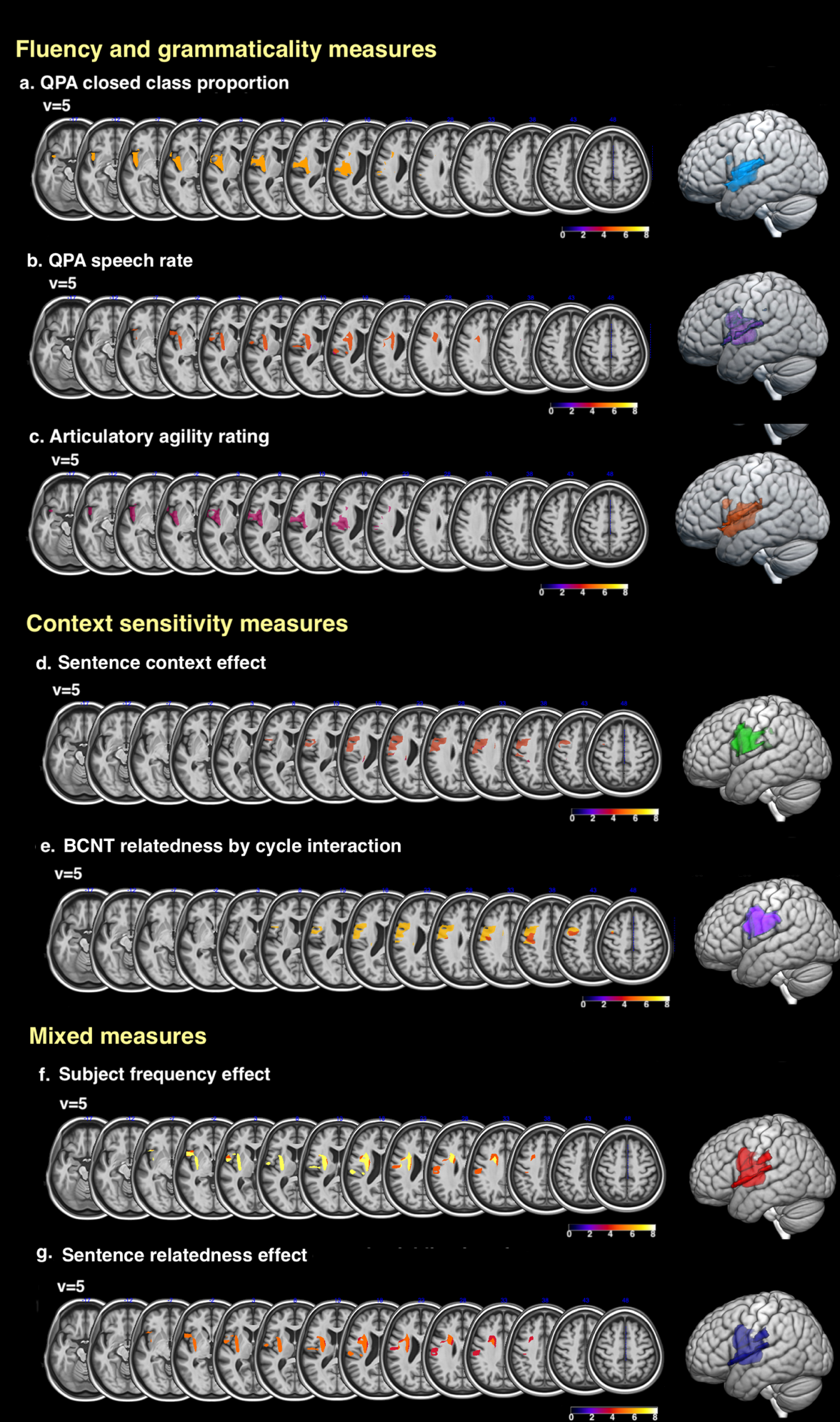
Panel a shows the complete lesion overlap map for all participants, colour-coded according to the number of cases with damage at each voxel, and projected onto our template for healthy older controls (Rorden et al., 2012). One participant had an accompanying right posterior lesion, which is not visible on this map (patient STR: see Figure S1 in the supplementary materials). Panel b shows the voxels that met our criterion for inclusion in VLSM analysis (that is, a minimum of 2 participants had lesions involving that voxel).

Turning now to the VLSM results, none of the analyses revealed significant voxels using the standard method of correction for familywise error rate (*v*=1). Below we summarise the results based on a less stringent correction method: CFWER with the value of v set to 5.

**3.2.1. Fluency and grammaticality measures.** Panels a, b and c of Figure 3 displays the voxels where pathology was significantly associated with impaired scores on the three measures that loaded primarily on our *fluency and grammaticality* factor: the QPA proportion of closed class words, the QPA speech rate, and the BDAE articulatory agility measure. Table 4 summarises the localisation of these voxels with respect to our reference atlases. The maps for the closed class measure and the articulatory agility measure were identical, and were centred on the insular cortex and the central opercular cortex, with only minimal extension into the LIFG. The map for speech rate was smaller, but was also primarily centred on the insular cortex and the central opercular cortex, again with minimal extension into the LIFG. With respect to the HCP1065 white matter atlas, both maps showed substantial overlap with the territory of the superior longitudinal fasciculus (bundle 3), and the frontal aslant tract, but the speech rate map showed greater frontal aslant tract involvement, and also some encroachment into the arcuate fasciculus.

**3.2.2. Context sensitivity measures.** Panels d and e of Figure 3 displays the voxels where pathology was associated with high scores on the two measures that loaded primarily on our *context sensitivity* factor: the sentence context effect measure and the BCNT relatedness by cycle interaction. Table 4 summarises the localisation of these voxels. The map for both measures encompassed significant portions of the pars opercularis of the inferior frontal gyrus, the middle frontal gyrus, the precentral gyrus and the postcentral gyrus. The map for the BCNT relatedness by cycle interaction was larger had greater extension into the postcentral gyrus than the sentence context effect map, but the two maps were otherwise very similar. With respect to white matter territories, both maps overlapped substantially with of the territories of the bundles 2 and 3 of the superior longitudinal fasciculus, and to a lesser extent with the frontal aslant tract.

**3.2.3. Mixed measures.** Panels f and g of Figure 3 display the voxels where pathology was significantly associated with disproportionately poor scores on our two remaining measures: the subject noun frequency effect and the noun semantic relatedness effect. As summarised in Table 4, the voxel maps for these measures overlapped almost completely, and incorporated substantial



**Figure 3**

VLSM maps showing lesion correlates of low scores on the seven key spoken language measures. The maps are projected onto our healthy older control template (Rorden et al., 2012). Also shown is the approximate localisation of significant voxels projected onto a surface rendered left lateral view of the MNI152 template using MRIcroGL software (Rorden, 2014).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Fluency and grammaticality measures** | | **Context sensitive measures** | | **Mixed measures** |
|  | **QPA Closed class propn/ articulatory agility\*** | **QPA speech rate** | **Sentence context effect** | **BCNT relatedness by cycle** | **Subject frequency effect/sentence relatedness effect\*** |
| Total significant voxels | 13659 | 11079 | 14631 | 21058 | 15627 |
| IFG, pars opercularis | 451 (8.6) | 287 (5,4) | **2825 (53.6)** | **2890 (54.8)** | 408 (7.7) |
| Middle frontal gyrus | - | - | **2583 (51.2)** | **3177 (63.0)** | 517 (10.2) |
| Frontal Operculum Cortex | 435 | 681 (18.0) | - | - | 799 (20.0) |
| Insular Cortex | **5160 (33.7)** | **2564 (16.8)** | - | - | **2599 (17.0)** |
| Central Opercular Cortex | **4973 (48.7)** | **1954 (19.2)** | 316 (3.1) | 370 (3.6) | **2062 (20.2)** |
| Precentral Gyrus | 726 (3.2) | 494 (2.1) | **6470 (28.1)** | **8537 (37.1)** | 1000 (4.3) |
| Postcentral Gyrus | 337(3.2) | - | **1962 (18.5)** | **3968 (37.4)** | **1644 (15.5)** |
| SMG, anterior division | - | - | - | 361 (6.6) | 373 (6.8) |
| Parietal Operculum Cortex | 317 (4.9) | 311 (4.8) | - | - | - |
| Planum Polare | 334 (8.5) | - | - | - | - |
| Temporal Pole | 492 (17.3) | - | - | - | - |
| SLF 2 | - | - | **2198 (16.9)** | **2953 (22.7)** | 423 (3.2) |
| SLF 3 | 985 (7.7) | 673 **(**5.3**)** | **2036 (16.0)** | **2248 (17.6)** | 1089 (8.5) |
| Frontal aslant tract | 319 (6.4) | **1208 (24.1)** | 900 (17.9) | **1520 (30.3)** | **1828 (36.5)** |
| Arcuate fasciculus | - | 550 (5.1) | 347 (3.2) | 417 (3.9) | 635 (5.9) |

\* The maps for these measures were identical.

IFG = inferior frontal gyrus; SMG -= supramarginal gyrus; SLF = superior longitudinal fasciculus

**Table 5**

Localisation of voxels reaching the threshold for statistical significance in each analysis, based on the Harvard-Oxford cortical atlas, and the HCP-1065 tractography atlas (association tracts). Titles that combined two measures separates by a slash indicate that the maps for those two measures were identical. The table presents the total number of voxels in the statistical map that overlapped with each territory. The figures in parentheses express these figure as a percentage of the total voxels submitted to analysis for that region. Regions where fewer than 200 voxels reached threshold are not shown. Figure in bold are regions containing a minimum of 1,000 voxels for that map, and that implicated at least 15% of the voxels submitted to analysis for that region. All voxels were located within the left hemisphere. The value of v used in the CFWER correction method was 5.

portions of the insular cortex, the central operculum, and the postcentral gyrus. With respect to white matter territories, there was substantial encroachment into the territory of the frontal aslant tract and the superior longitudinal fasciculus (bundle 3).

**3.2.4. Exploring indices of “posterior” aphasia**. All of our key behavioural measures we have examined are those associated with nonfluent aphasia or with anterior language structures. Our observations would support stronger conclusions if we were able to show that measures predicted to involve more posterior regions yielded distinctly different lesion maps. To address this question, we examined the statistical maps associated with two behavioural indices that have been previously associated with damage to left posterior language structures. Speer and Wilshire (2013) found that several of their participants with more posterior lesions actually showed a semantic facilitation effect in sentence production – that is, they performed *more* accurately on sentences containing semantically related nouns than on those containing unrelated nouns. We created an index of this semantic facilitation effect by calculating the inverse of the values for the semantic relatedness effect. Table S4 summarises the localisation of voxels that were significantly associated with this effect, after CFWER correction based on *v*= 5. The majority of significant voxels were located within the temporooccipital part of the middle temporal gyrus, and there was some overlap with the territories of the arcuate fasciculus and the inferior longitudinal fasciculus.

To examine an alternative “posterior” behavioural index, we created a measure of the repetition priming effect across cycles on the blocked cyclic naming task. Using a similar cyclic naming task to our own, Python et al. (2018) observed a characteristic repetition priming effect in a group of individuals with left middle temporal gyrus lesions: their naming latencies became faster as the cycles progressed, irrespective of whether the picture sets were related or unrelated. To create an index of this effect, we calculated the average percentage change in naming latencies for each participant with each successive cycle through the picture set, irrespective of relatedness condition. 1 A VLSM analysis based on this measure did not reveal any significant voxels after applying a CFWER correction based on *v*= 5. Table S3 provides the individual latency data for each participant, as well as the values for our repetition priming index, and Table S4 summarises the localisation of voxels that were significantly associated with this effect, after CFWER correction based on *v*= 10. This analysis identified a small region within the inferior temporal lobe (the posterior inferior temporal gyrus, and the anterior and posterior portions of the temporal fusiform gyrus), with extension into the territory of the inferior longitudinal fasciculus and the IFOF.

**4. Discussion**

This study investigated the associations between seven key measures in people with aphasia that have previously been associated with nonfluent aphasia and/or damage to anterior language regions. Not surprisingly, all the measures showed a trend toward positive associations, but there was evidence for two distinct clusters. The first cluster, *fluency and grammaticality*, included three measures of open-ended speech: the proportion of closed class words, speech rate and examiner-rated articulatory agility. The association between measures of speech rate and grammatical function word production is consistent with previous research (e.g., Nozari & Faroqi-Shah, 2017; but see Ding et al., 2020, for a large group study demonstrating some dissociation). The second cluster, *context sensitivity,* included a comparative measure of subject and object noun accuracy on a constrained sentence production task and also an index of semantic interference in the cyclic naming task. The remaining two measures shared overlap with both these clusters. These findings, taken together, suggest that the variance in our seven key measures is likely to be underpinned by at least two distinct types of cognitive difficulties.

In the second part of the study, we examined the lesion correlates of these measures. The sample size was small, so statistical power was low, and coverage was limited to the central territory of the left middle cerebral artery, so these findings must be treated with caution. However, after factoring out the effects of overall lesion volume, there were several findings of interest, which survived a relatively stringent correction for multiple comparisons. The measures in our fluency and grammaticality cluster were associated with damage to a central left hemisphere region encompassing the left insula and central operculum. There was little involvement of the inferior or middle frontal gyri, despite adequate power to detect such effects, as indicated by their involvement in other VLSM analyses. The left insula has previously been implicated in measures of fluency and speech rate (e.g., Borovsky et al., 2007; Ding et al., 2020; Halai et al., 2017, 2018), and also in more specific measures of articulatory-motor planning that focus on the phonetic quality of speech and/or its error characteristics (Baldo et al., 2011; Henseler, Regenbrecht & Obrig, 2014; Itabashi et al., 2016; but see also Mirman et al., 2019, for contrasting findings using a picture naming task). However, unlike several previous studies that have also implicated the LIFG in speech rate and grammatical content measures (Ding et al., 2020; den Ouden et al., 2019), we found little evidence of LIFG involvement in any of the measures in this cluster. Subcortically, the territory of the superior longitudinal fasciculus (bundle 3) was implicated in all three maps, but the map for speech rate extended had the greatest extension into the frontal aslant tract. Our finding, of greater frontal aslant tract involvement in a generic speech rate measure than in a more specific measure of speech articulatory quality, is consistent with the view that this tract is involved in processes that lie upstream from articulatory-motor planning (Chernoff et al., 2019).

Turning now to the two measures in our *context sensitivity* cluster, both were associated with damage to a more anterior region encompassing the left middle frontal gyrus, the LIFG (pars opercularis) and with extension into the postcentral and precentral gyri. Importantly, the left insula and central operculum were not heavily implicated, suggesting they index distinctly different cognitive impairments when compared to the fluency and grammatically measures. Subcortically, the maps for the context sensitivity measures overlapped extensively with the territory of the superior longitudinal fasciculus and to a lesser extent, that of the frontal aslant tract. The observed involvement of the LIFG in our cyclic naming measure is consistent with previous studies that have found associations between LIFG damage and the cumulative semantic interference effect in this paradigm (Anders et al., 2017; Schnur et al., 2005, 2009). The maps for the remaining “mixed” measures tended to be broader, extending across both areas identified in the maps for the two main clusters, but subcortically, encroaching more substantially on the territory of the frontal aslant tract.

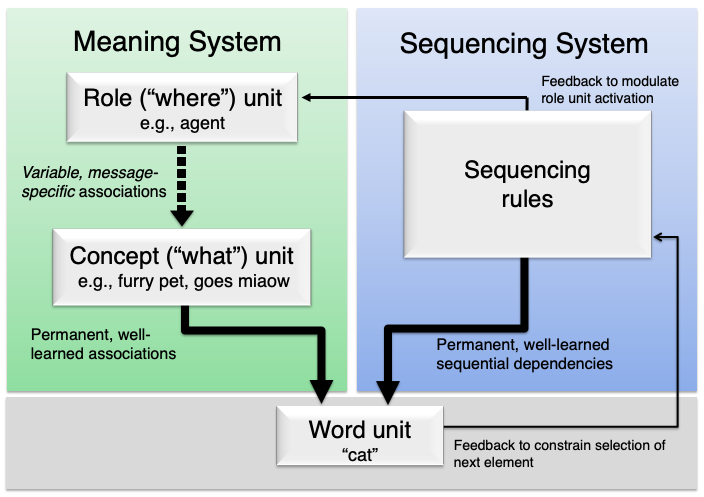
Finally, to test the specificity of the lesion-behaviour associations obtained in this study, we examined the lesion maps for two measures that have previously been associated with posterior left hemisphere damage. Neither map shared any overlap with those of our measures of interest. These findings suggest that the lesion maps observed in our main analyses were highly specific to the measures examined.

Our behavioural and lesion data, considered as a whole, suggest there are likely to be at least two cognitive difficulties contributing to the speech abnormalities observed in nonfluent aphasia. The first is a general slowing of speech planning process, associated with damage to central left hemisphere structures such the left insula and the central operculum. Given that that behaviours associated with this difficulty include reduced speech articulatory quality, and given that the left insula has previously been implicated in articulatory-motor planning specifically (Baldo et al., 2011; Henseler, Regenbrecht & Obrig, 2014; Itabashi et al., 2016), we suggest that a major contributor to these planning delays is impaired articulatory-motor programming. The finding of a closed association between grammatical function word omission and reduced speech rate/poor articulatory quality is consistent with the notion that “agrammatic” speech may be an adaptation to such a timing limitation, rather than a selective deficit in its own right. The idea of agrammatism as an adaptive strategy is not new (Hartsuiker & Kolk, 1998; Kolk & Van Grunsven, 1985; Kolk, 1995, 2006; De Roo et al., 2003; Kolk & Heeschen, 1996; Ruiter et al., 2010; Salis & Edwards, 2004; Sahraoui & Nespoulous, 2012). However, accounts vary as to the nature of the underlying processing limitation that gives rise to this strategy. The current findings suggest strongly that a major source of that limitation may be at the level of articulatory-motor planning specifically.

The second factor that we propose contributes to speech difficulties in nonfluent aphasia is an impairment affecting the top-down control of language production, secondary to damage involving more anterior language regions, including the LIFG and possibly also the middle inferior frontal gyrus. We propose that this impairment is responsible for the marked effects of context sensitivity that were associated with damage to these regions. Numerous authors have previously suggested that anterior language structures, including the LIFG, may play an “activation biasing” role in speech production, that becomes crucial when there is competition for selection amongst language representation, or where there are few external constraints on what should be selected (Belke & Stielow, 2013, Belke, 2017; Novick et al., 2009, 2010; Riès et al., 2015; Robinson et al., 2005, 2010, 2012; Schnur et al., 2005, 2009; Scott & Wilshire, 2010; Thothathiri, et al., 2010; 2012; Wilshire & McCarthy, 2002; but see Anders et al., 2017; Hsiao et al., 2009; Oppenheim, Dell & Schwartz, 2010, for a slightly different proposal). As to the *nature* of this activation modulating mechanism, one hypothesis is that it is closely connected with the processes involved in generating a message-level communication “intention” that is capable of driving the selection of language representations (Robinson et al., 2005, 2010, 2012; Barker Nelson & Robinson, 2020).

This type of account can provide an integrated explanation of the context sensitivity effects we observed here in both single word and sentence production. We propose that the key problem underlying these effects is an inability to generate and/maintain a robust message-level representation to drive the selection of appropriate content words in all speech production contexts. As a consequence, the speaker has greater difficulty selecting a target lexical element from amongst other, concurrently activated representations.

**4.1. A tentative model.** A useful theoretical framework in which to situate this two-factor proposal is the sentence production model of Chang and colleagues (Chang, 2002; Chang et al., 2006; see also Dell & Chang, 2014), which is illustrated in Figure 5. This model proposes that there are two primary sets of representations that are important for sentence planning. The *meaning system* comprises the knowledge we possess about concepts, represented within a network of hierarchically organised units representing their semantics, their lexical labels and so forth. Here we will use the term *content element* to describe the network associated with a given concept. In contrast to the meaning system, the *sequencing system* represents learned information about word sequencing rules, that are coded in the form of sequential dependencies between different word classes (this part of the model learns by a recurrent PDP network, so it is sensitive not only to immediate transitional probabilities but also dependencies than span several words). In addition to these systems that represent generalised conceptual knowledge and sequential dependencies respectively, the model further includes a set of *role (“where”) units,* which represent the types of thematic roles that can be expressed in given proposition (e.g., *agent, patient, action)*, and which can be dynamically linked to concept units to represent a given proposition.



**Figure 5**

Simplified illustration of Chang et al.’s (2006) dual-path model of sentence production. The figure illustrates some of the units that are active just prior to selection of the word “cat” in the sentence “*The cat chased the mouse*”.

The process of speech planning begins with the generation of a message-level representation: the key content elements to be included in the message are activated and are bound to their appropriate role units (e.g., cat –> *agent*, chase –> *action*, mouse –> *patient*), and other important relational information is attached where relevant (for example, if both *mouse* and *cat* are known entities they will be tagged as definite, and if the event happened in the past, *chase* will be tagged as past tense). The sequencing system will then select the first word, based on the information coded in the message-level representation – specifically, which content-element. Role pairing is most strongly activated - and on its knowledge of the sequential tendencies that apply to different classes of words. The sequencing system ensures that the appropriate grammatical morphemes are also inserted.

There are two useful features of this model for our current purposes. The first is its sequential nature: the process of word selection occurs sequentially. The second feature is that the model views syntactic encoding as a relatively low-level process, supported by knowledge of the sequential regularities of language which has been massed from thousands of exposures to different sequences, making it highly robust to localised damage. The aspect of sentence planning that demands the greatest cognitive control is the generation and maintenance of the message-specific links between concept and role units, which must remain active throughout the process of sentence element selection.

**4.2. Application to current findings.** Within this framework, we might propose that the generation and maintenance of a set of content element-role linkings is supported by anterior language structures, including the LIFG, and that this process can be compromised by damage to these regions. This will seriously compromise the ability to generate novel propositional speech, and will also lead to specific difficulties on more constrained tasks of sentence and word production. The absence of strong content element-role linkages will mean that word units are selected largely on the basis of which content elements happen to be most highly activated, and whether they are permissible for selection based on the dependencies coded within the sequencing system. In noun production tasks such as picture naming, the grammatical class of the target word is fixed, so the competition for selection will be won by whichever noun-compatible content element is most strongly activated at the time. It has been suggested that repeated naming of a small set of semantically related pictures is likely to induce high levels of competition, particularly as the naming cycles progress (Howard et al., 2006; Roelofs, 2018; Schnur et al, 2006; but see Oppenheim et al., 2010, for an alternative account that emphasises learning effects). If this is indeed the case, then such effects are likely to be a sensitive marker of difficulties at this level.

In constrained sentence production tasks like the ones used by Speer and Wilshire (2013), successful selection of the first content element will again depend largely on which concept is most strongly activated in the mind of the speaker at that time – most probably, the agent of the action to be described. A complete, robust, message-level representation may not be required at this stage. However, selection of content elements for later sentence positions will depend crucially on the robustness of the message-level representation. Without such a representation to “drive” content element selection, the system becomes more vulnerable to interference from non-target content elements that are also currently active – which may include those that have recently been produced, those being considered for later production, and perhaps also those that share semantic similarities with active content elements. Hence, this framework offers a plausible account as to why those with difficulty generating message-level representations might be more successful at producing the first content word in the utterance (in this instance, the agent/subject noun) than its later elements.

In this study, we observed two further patterns of performance that were significantly more common following damage to the LIFG/middle frontal gyrus (and also following insula/central operculum damage, which we will discuss below). This first was poorer accuracy on sentences commencing with a low frequency subject noun. Without a robust top-down signal from the message level, the sentence planning system becomes more vulnerable to current patterns of activation in the meaning system, which are subject to decay. Therefore, any factor that introduces planning delays, particularly early in the process, will be likely to lead to selection failures downstream. The second pattern we found to be associated with LIFG/middle frontal gyrus damage was poorer accuracy on sentences that contained semantically related nouns. An explanation for this phenomenon falls out neatly from our proposed framework: without a robust message-level representation that ties concepts to their roles, content elements will be more likely compete with one another for selection at a given position. Consequently, any manipulation that enhances the degree of competition between two elements - in this case, increasing their semantic relatedness - will tend to lead to more sentence production failures.

We now turn to our second proposed factor contributing to sentence production difficulties – the “timing” factor - and consider how it might be conceptualised within this framework. A central feature of the Chang model is that sentence planning proceeds one word at a time. If we assume that the message-level representation is a set of temporary pairings that decay over time, then any difficulty that introduces substantial delays may result in the premature decay of the message-level representation. In this scenario, the system responsible for generating a message-level representation itself is itself functioning normally, but the planning process becomes derailed due to downstream delays. We have suggested here that damage to central left hemisphere structures, including the insula, may have just this effect, possibly because it severely compromises articulatory-motor programming, thereby introducing significant delays into the sentence planning process. We further suggested that speakers may attempt to minimise the impact of such delays by omitting grammatical function words. But this in itself might not always be sufficient to manage the problem. Indeed, the two patterns considered above – poorer accuracy on sentences beginning with a low subject noun, and poorer accuracy on sentences containing semantically related nouns – were observed not only in individuals with damage to anterior language structures, but also following damage to the insula/central operculum. In other words, these patterns may be a common endpoint of two very different kinds of impairments.

Finally, and perhaps most importantly, the proposed two-factor account we have set out above offers a framework for understanding variability in the speech profiles of different individuals with nonfluent aphasia. The classic “dynamic aphasia” profile, of relatively well-formed speech but difficulties selecting appropriate language units in an unconstrained setting, would reflect a difficulty at the conceptual/propositional level. In contrast, the nonfluent/agrammatic profile, consisting of slowed speech in the context of articulatory abnormalities, in the absence of context-specific noun production problems, would be attributed to our proposed timing difficulty. A third group of individuals are likely to have a combination of these two problems, and given the proximity of their neural substrates, this combination might be quite common (for a similar proposal, see Robinson et al., 2005, 2010, 2012).

**4.3. Limitations and conclusions**. There are a number of things to bear in mind when evaluating these findings. The first is the small sample size, which limits the conclusions we can draw from the VLSM analyses in particular. There was likely to be insufficient power to detect effects in some regions, and no coverage at all of others. Also, the various differences we observed in the VLSM maps for different measures could have been driven by just one or two individuals; we clearly need to see whether these findings bear out in a larger participant group. There were several other limitations to the VLSM analysis worth mentioning. Due to the large dimensions of the lesion maps and their heavily overlapping distributions, there was saturation in some regions (for example, five out of the eight cases had more than 10% damage to the insula, and that was true for the central operculum in six of the eight cases). Also, there was a lack of *independent* coverage of some regions. For example, the degree of damage to the central operculum was correlated with that of the insula, and similar high correlations were observed between the middle frontal gyrus and the of the LIFG (pars opercularis), and between the precentral and postcentral gyri respectively (see Table S4 for a full correlation matrix). Consequently, we can say little about the about the independent contribution of these regions to the functions under investigation, and must keep our region descriptions broad (see Ding et al, 2020, for further discussion of this problem).

Nevertheless, the study provides some valuable insights into sentence planning and its impairment in aphasia. Perhaps most importantly, it demonstrates the usefulness of assessing individuals on a wide array of language measures designed to target different aspects of the speech planning process. In particular, we were able to show how supplementing more “naturalistic” measures of speech with those obtained from more constrained sentence and word production tasks can be helpful in teasing apart the different cognitive mechanisms underpinning speech problems in nonfluent aphasia. Further, our finding of a close behavioural and anatomical relationship between grammatical content word omission and other aspects of speech fluency and articulatory quality raises the possibility that apparent “grammatical” abnormalities in nonfluent speech may not reflect a grammatical problem at all. Our tentative model may be able to account for a wide range of features of nonfluent aphasic speech without proposing any kind of syntactic difficult or “structure generator” at all. The findings make a strong case for shifting our research focus away from grammatical aspects of nonfluent aphasia, towards the r selection of conceptual elements, including phenomena that straddle the divide between sentence and single word production. Perhaps most importantly, if we are o gain a fuller understanding of the brain regions that are crucial for sentence production, and the specific cognitive processes they support, we need to be guided by a richer theoretical framework that identifies the various cognitive processes that contribute to this task and the likely consequences of impairment to each.

**Footnotes**

1. Latencies were first log transformed (natural log), and then outliers were removed (values that lay more than 2.5 standard deviations above or below the mean for that relatedness x cycle combination). All means and all statistical analyses reported in this paper, including the measures submitted to VLSM analysis, were based on the logged latencies.

**Acknowledgements**

We thank all our participants for their time spent to support our research. Thank you to Jacob Cameron and Angie Slocombe for help and advice regarding the magnetic resonance imaging (MRI) scans.

This research was supported by a Marsden Project grant to C.E. Wilshire [grant number VUW0505].

**References**

Anders, R., Riès, S., Van Maanen, L., & Alario, F. X. (2017). Lesions to the left lateral prefrontal cortex impair decision threshold adjustment for lexical selection. *Cognitive Neuropsychology, 34(1-2)*, 1-20. https://doi.org/10.1080/02643294.2017.1282447

Ashburner, J., Barnes, G., Chen, C., Daunizeau, J., Flandin, G., Friston, K., ... Phillips, C. 1199 (2012). *SPM8 manual.* Functional Imaging Laboratory, Institute of Neurology.

Baldo, J. V., Wilkins, D. P., Ogar, J., Willock, S., & Dronkers, N. F. (2011). Role of the precentral gyrus of the insula in complex articulation. *Cortex, 47(7)*, 800-807. https://doi.org/10.1016/j.cortex.2010.07.001

Barker, M. S., Nelson, N. L., & Robinson, G. A. (2020). Idea formulation for spoken language production: The interface of cognition and language. Journal of the International *Neuropsychological Society, 26(2)*, 226-240. https://doi.org/10.1017/S1355617719001097

Basilakos, A., Fillmore, P. T., Rorden, C., Guo, D., Bonilha, L., & Fridriksson, J. (2014). Regional white matter damage predicts speech fluency in chronic post-stroke aphasia. *Frontiers in Human Neuroscience, 8, 845*. https://doi.org/10.3389/fnhum.2014.00845

Belke, E. (2017). Effects of lesions to the left lateral prefrontal cortex on task-specific top-down biases and response strategies in blocked-cyclic naming. *Cognitive Neuropsychology, 34(1-2)*, 26-32. https://doi.org/10.1080/02643294.2017.1329200

Belke, E., & Stielow, A. (2013). Cumulative and non-cumulative semantic interference in object naming: Evidence from blocked and continuous manipulations of semantic context. *Quarterly Journal of Experimental Psychology, 66(11)*, 2135-2160. https://doi.org/10.1080/17470218.2013.775318

Belke, E., Meyer, A. S., & Damian, M. F. (2005). Refractory effects in picture naming as assessed in a semantic blocking paradigm. *The Quarterly Journal of Experimental Psychology, 58(4),* 667-692. https://doi.org/10.1080/02724980443000142

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology, 18(3)*, 355-387. https://doi.org/10.1016/0010-0285(86)90004-6

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

Catani, M., Mesulam, M. M., Jakobsen, E., Malik, F., Martersteck, A., Wieneke, C., ... & Rogalski, E. (2013). A novel frontal pathway underlies verbal fluency in primary progressive aphasia. *Brain, 136(8)*, 2619-2628. https://doi.org/10.1093/brain/awt163

Chang, E. F., Kurteff, G., & Wilson, S. M. (2018). Selective interference with syntactic encoding during sentence production by direct electrocortical stimulation of the inferior frontal gyrus. *Journal of Cognitive Neuroscience, 30(3)*, 411-420. https://doi.org/10.1162/jocn\_a\_01215

Chang, F. (2002). Symbolically speaking: A connectionist model of sentence production. *Cognitive Science, 26(5),* 609-651. https://doi.org/10.1207/s15516709cog2605\_3

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review, 113(2)*, 234. doi:10.1037/0033-295X.113.2.234https://doi.org/10.1037/0033-295X.113.2.234

Chernoff, B. L., Sims, M. H., Smith, S. O., Pilcher, W. H., & Mahon, B. Z. (2019). Direct electrical stimulation of the left frontal aslant tract disrupts sentence planning without affecting articulation. *Cognitive Neuropsychology, 36(3-4)*, 178-192. https://doi.org/10.1080/02643294.2019.1619544

de Roo, E., Kolk, H., & Hofstede, B. (2003). Structural properties of syntactically reduced speech: A comparison of normal speakers and Broca's aphasics. *Brain and Language, 86*, 99-115. https://doi.org/10.1016/S0093-934X(02)00538-2

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. Psychological *Review, 93(3)*, 283. https://doi.org/10.1037/0033-295X.93.3.283

Dell, G. S., & Chang, F. (2014). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369(1634)*, 20120394. https://doi.org/10.1098/rstb.2012.0394

DeMarco, A. T., & Turkeltaub, P. E. (2018). A multivariate lesion symptom mapping toolbox and examination of lesion-volume biases and correction methods in lesion-symptom mapping. *Human Brain Mapping, 39(11)*, 4169-4182. https://doi.org/10.1002/hbm.24289

den Ouden, D. B., Malyutina, S., Basilakos, A., Bonilha, L., Gleichgerrcht, E., Yourganov, G., ... & Fridriksson, J. (2019). Cortical and structural‐connectivity damage correlated with impaired syntactic processing in aphasia. *Human Brain Mapping*, *40(7)*, 2153-2173. https://doi.org/10.1002/hbm.24514

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., ... Albert, M. S. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage, 31 (3), 968-980. https://doi.org/10.1016/j.neuroimage.2006.01.021

Ding, J., Martin, R. C., Hamilton, A. C., & Schnur, T. T. (2020). Dissociation between frontal and temporal-parietal contributions to connected speech in acute stroke. *Brain, 143(3*), 862-876. https://doi.org/10.1093/brain/awaa027

Druks, J., & Masterson, J. (2000). An object and action naming battery. Hove: Psychology Press.

Dunn, L. M., & Dunn, L. M. (1997). *The Peabody Picture Vocabulary Test* (3rd edition). Pearson. https://doi.org/10.1037/t15145-000

Faroqi-Shah, Y., Kling, T., Solomon, J., Liu, S., Park, G., & Braun, A. (2014). Lesion analysis of language production deficits in aphasia. *Aphasiology, 28(3)*, 258-277. https://doi.org/10.1080/02687038.2013.853023

Ferreira, V. S., & Slevc, L. R. (2007). Grammatical encoding. In G. Gaskell (Ed.), *Oxford Handbook of Psycholinguistics* (pp. 453-469). Oxford, UK: OUPhttps://doi.org/10.1093/oxfordhb/9780198568971.013.0027

Freedman, M. L., Martin, R. C., & Biegler, K. (2004). Semantic relatedness effects in conjoined noun phrase production: Implications for the role of short-term memory. *Cognitive Neuropsychology, 21*, 245-265https://doi.org/10.1080/02643290342000528

Fuster, J. M. (2001). The prefrontal cortex-an update: time is of the essence. *Neuron, 30(2)*, 319-333. https://doi.org/10.1016/S0896-6273(01)00285-9

Garrett, M. F. (1976). Syntactic processes in sentence production. In R. J. Wales & E. C. T. Walker (Eds.), *New approaches to language mechanisms* (pp. 231-256). North-Holland.

Gläscher, J., Adolphs, R., Damasio, H., Bechara, A., Rudrauf, D., Calamia, M., ... & Tranel, D. (2012). Lesion mapping of cognitive control and value-based decision making in the prefrontal cortex. *Proceedings of the National Academy of Sciences, 109(36),* 14681-14686. https://doi.org/10.1073/pnas.1206608109

Goldmann, R. E., Schwartz, M. F., & Wilshire, C. E. (2001). The influence of phonological context on the sound errors of a speaker with Wernicke's aphasia. *Brain and Language, 78(3)*, 279-307. https://doi.org/10.1006/brln.2001.2468

Goodglass, H., Kaplan, E., Barresi, B. (2001). *The Boston Diagnostic Aphasia Examination* (Third ed). Lippincott Williams & Wilkins

Halai, A. D., Woollams, A. M., & Ralph, M. A. L. (2017). Using principal component analysis to capture individual differences within a unified neuropsychological model of chronic post-stroke aphasia: Revealing the unique neural correlates of speech fluency, phonology and semantics. *Cortex, 86*, 275-289. https://doi.org/10.1016/j.cortex.2016.04.016

Halai, A. D., Woollams, A. M., & Ralph, M. A. L. (2018). Triangulation of language-cognitive impairments, naming errors and their neural bases post-stroke. *NeuroImage: Clinical, 17,* 465-473. https://doi.org/10.1016/j.nicl.2017.10.037

Hartsuiker, R. J., & Kolk, H. H. J. (1998). Syntactic facilitation in agrammatic sentence production. Brain and Language, 62, 221 - 254. https://doi.org/10.1006/brln.1997.1905

Henseler, I., Regenbrecht, F., & Obrig, H. (2014). Lesion correlates of patholinguistic profiles in chronic aphasia: Comparisons of syndrome-, modality-and symptom-level assessment. *Brain, 137, 918-930*. https://doi.org/10.1093/brain/awt374

Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: Experimental and computational studies. *Cognition, 100(3)*, 464-482. https://doi.org/10.1016/j.cognition.2005.02.006

Hsiao, E. Y., Schwartz, M. F., Schnur, T. T., & Dell, G. S. (2009). Temporal characteristics of semantic perseverations induced by blocked-cyclic picture naming. *Brain and Language, 108,* 133-144. https://doi.org/10.1016/j.bandl.2008.11.003

Itabashi, R., Nishio, Y., Kataoka, Y., Yazawa, Y., Furui, E., Matsuda, M., & Mori, E. (2016). Damage to the left precentral gyrus is associated with apraxia of speech in acute stroke. *Stroke, 47*, 31-36. https://doi.org/10.1161/STROKEAHA.115.010402

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford University Press,

Kay, J., Lesser, R., & Coltheart, M. (1992). PALPA: *Psycholinguistic assessments of language processing in aphasia*. Hove: Psychology Press.

Kolk, H. H. J. (1995). A time-based approach to agrammatic production. *Brain and Language, 50,* 282-303. https://doi.org/10.1006/brln.1995.1049

Kolk, H. H. J. (2006). How language adapts to the brain: An analysis of agrammatic aphasia. In L. Progovac (Ed.), *The syntax of nonsententials: Multi-disciplinary perspectives* (pp. 229 - 258). London: John Benjamins. https://doi.org/10.1075/la.93.11kol

Kolk, H. H. J., & Van Grunsven, M. M. F. (1985). Agrammatism as a variable phenomenon. *Cognitive Neuropsychology, 2*, 347 - 384. https://doi.org/10.1080/02643298508252666

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language, 33(2)*, 149. https://doi.org/10.1006/jmla.1994.1008

Lacey, E. H., Skipper-Kallal, L. M., Xing, S., Fama, M. E., & Turkeltaub, P. E. (2017). Mapping common aphasia assessments to underlying cognitive processes and their neural substrates. *Neurorehabilitation and Neural Repair, 31(5)*, 442-450. https://doi.org/10.1177/1545968316688797

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22(1)*, 1-75. https://doi.org/10.1017/S0140525X99001776

Martin, R. C., & Romani, C. (1994). Verbal working memory and sentence comprehension: A multiple-components view. *Neuropsychology, 8*, 506 - 523. https://doi.org/10.1037/0894-4105.8.4.506

Martin, R. C., Lesch, M., & Bartha, M. (1999). Independence of input and output phonology in word processing and short-term memory. *Journal of Memory and Language, 41*, 3 - 29. https://doi.org/10.1006/jmla.1999.2637

Martin, R. C., Miller, M., & Vu, H. (2004). Lexical-semantic retention and speech production: Further evidence from normal and brain-damaged participants for a phrasal scope of planning. *Cognitive Neuropsychology, 21*, 625 - 644. https://doi.org/10.1080/02643290342000302

Martin, R. C., Shelton, J., & Yaffee, L. (1994). Language processing and working memory: Neuropsychological evidence for separate phonological and semantic capacities. *Journal of Memory and Language, 33*, 83 - 111. https://doi.org/10.1006/jmla.1994.1005

Mirman, D., Chen, Q., Zhang, Y., Wang, Z., Faseyitan, O. K., Coslett, H. B., & Schwartz, M. F. (2015). Neural organization of spoken language revealed by lesion- symptom mapping. *Nature Communications, 6*, 6762. https://doi.org/10.1038/ncomms7762

Mirman, D., Kraft, A. E., Harvey, D. Y., Brecher, A. R., & Schwartz, M. F. (2019). Mapping articulatory and grammatical subcomponents of fluency deficits in post-stroke aphasia. *Cognitive, Affective, and Behavioral Neuroscience, 19(5)*, 1286-1298. https://doi.org/10.3758/s13415-019-00729-9

Mirman, D., Landrigan, J. F., Kokolis, S., Verillo, S., Ferrara, C., & Pustina, D. (2018). Corrections for multiple comparisons in voxel-based lesion-symptom mapping. *Neuropsychologia, 115,* 112-123. https://doi.org/10.1016/j.neuropsychologia.2017.08.025

Novick, J. M., Kan, I. P., Trueswell, J. C., & Thompson- Schill, S. L. (2009). A case for conflict across multiple domains: Memory and language impairments following damage to ventrolateral prefrontal cortex. *Cognitive Neuropsychology, 26*, 527-567. https://doi.org/10.1080/02643290903519367

Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2010). Broca's area and language processing: Evidence for the cognitive control connection. *Language and Linguistics Compass, 4*, 906-924. https://doi.org/10.1111/j.1749-818X.2010.00244.x

Nozari, N., & Faroqi-Shah, Y. (2017). Investigating the origin of nonfluency in aphasia: A path modeling approach to neuropsychology. *Cortex, 95*, 119-135. https://doi.org/10.1016/j.cortex.2017.08.003

Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. *Cognition, 114(2)*, 227-252. https://doi.org/10.1016/j.cognition.2009.09.007

Pickering, M. J. & Branigan, H. P. (1999). Syntactic priming in language production. *Trends in Cognitive Sciences, 3*, 136-141. https://doi.org/10.1016/S1364-6613(99)01293-0

Pustina, D. (2018). *LESYMAP: Lesion to Symptom Mapping in R*. R package version 0.0.0.9201

Python, G., Glize, B., & Laganaro, M. (2018). The involvement of left inferior frontal and middle temporal cortices in word production unveiled by greater facilitation effects following brain damage. *Neuropsychologia, 121,* 122-134. https://doi.org/10.1016/j.neuropsychologia.2018.10.026

Riès, S. K., Greenhouse, I., Dronkers, N. F., Haaland, K. Y., & Knight, R. T. (2014). Double dissociation of the roles of the left and right prefrontal cortices in anticipatory regulation of action. *Neuropsychologia, 63*, 215-225. https://doi.org/10.1016/j.neuropsychologia.2014.08.026

Riès, S. K., Karzmark, C. R., Navarrete, E., Knight, R. T., & Dronkers, N. F. (2015). Specifying the role of the left prefrontal cortex in word selection. *Brain and Language, 149*, 135-147. https://doi.org/10.1016/j.bandl.2015.07.007

Robinson, G., Shallice, T., Bozzali, M., & Cipolotti, L. (2010). Conceptual proposition selection and the LIFG: neuropsychological evidence from a focal frontal group. *Neuropsychologia, 48(6)*, 1652-1663. https://doi.org/10.1016/j.neuropsychologia.2010.02.010

Robinson, G., Shallice, T., Bozzali, M., & Cipolotti, L. (2012). The differing roles of the frontal cortex in fluency tests. *Brain, 135(7)*, 2202-2214. https://doi.org/10.1093/brain/aws142

Robinson, G.A., Shallice, T., & Cipolotti, L. (2005). A failure of high level verbal response selection in progressive dynamic aphasia. *Cognitive Neuropsychology, 22*, 661-694. doi: 10.1080/ 02643290442000239https://doi.org/10.1080/02643290442000239

Roelofs, A. (2018). A unified computational account of cumulative semantic, semantic blocking, and semantic distractor effects in picture naming. *Cognition, 172*, 59-72. https://doi.org/10.1016/j.cognition.2017.12.007

Rogalski, E., Cobia, D., Harrison, T. M., Wieneke, C., Thompson, C. K., Weintraub, S., & Mesulam, M. M. (2011). Anatomy of language impairments in primary progressive aphasia. *Journal of Neuroscience, 31(9)*, 3344-3350. https://doi.org/10.1523/JNEUROSCI.5544-10.2011

Rorden, C., Bonilha, L., Fridriksson, J., Bender, B., & Karnath, H. O. (2012). Age- 1550 specific CT and MRI templates for spatial normalization. *Neuroimage, 61*, 1551 957-965. https://doi.org/10.1016/j.neuroimage.2012.03.020

Ruiter, M. B., Kolk, H. H., & Rietveld, T. C. (2010). Speaking in ellipses: The effect of a compensatory style of speech on functional communication in chronic agrammatism. *Neuropsychological Rehabilitation, 20*, 423-458. https://doi.org/10.1080/09602010903399287

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

Saffran, E. M., Schwartz, M. F., Linebarger, M. C., Martin, N., & Bochetto, P. (1988). *The Philadelphia comprehension battery*. Unpublished test battery.

Sahraoui, H., & Nespoulous, J. L. (2012). Across-task variability in agrammatic performance. *Aphasiology, 26(6)*, 785-810. https://doi.org/10.1080/02687038.2011.650625

Salis, C., & Edwards, S. (2004). Adaptation theory and non-fluent aphasia in English. *Aphasiology, 18*, 1103-1120. https://doi.org/10.1080/02687030444000552

Schnur, T. T. (2017). Word selection deficits and multiword speech. *Cognitive Neuropsychology, 34(1-2)*, 21-25. https://doi.org/10.1080/02643294.2017.1313215

Schnur, T. T., Lee, E., Coslett, H. B., Schwartz, M. F., & Thompson-Schill, S. L. (2005). When lexical selection gets tough, the LIFG gets going: A lesion analysis study of interference during word production. *Brain and Language, 95(1*), 12-13. https://doi.org/10.1016/j.bandl.2005.07.008

Schnur, T. T., Schwartz, M. F., Kimberg, D. Y., Hirshorn, E., Coslett, H. B., & Thompson-Schill, S. L. (2009). Localizing interference during naming: Convergent neuroimaging and neuropsychological evidence for the function of Broca's area. *Proceedings of the National Academy of Sciences, 106(1),* 322-327. https://doi.org/10.1073/pnas.0805874106

Scott, R. M., & Wilshire, C. E. (2010). Lexical competition for production in a case of nonfluent aphasia: Converging evidence from four different tasks. *Cognitive Neuropsychology, 27*, 505 - 538. https://doi.org/10.1080/02643294.2011.598853

Speer, P., & Wilshire, C. E. (2013). What's in a sentence? The crucial role of lexical content in sentence production in nonfluent aphasia. *Cognitive Neuropsychology, 30(7-8)*, 507-543. https://doi.org/10.1080/02643294.2013.876398

Stemberger, J. P. (1985). An interactive activation model of language production. In A. Ellis (Ed.), *Progress in the psychology of language* (Vol. 1, pp. 143-186). Mahwah NJ: Erlbaum.

Thothathiri, M., Gagliardi, M., & Schwartz, M. F. (2012). Subdivision of frontal cortex mechanisms for language production in aphasia. *Neuropsychologia, 50(14)*, 3284-3294. https://doi.org/10.1016/j.neuropsychologia.2012.09.021

Thothathiri, M., Schwartz, M. F., & Thompson-Schill, S. L. (2010). Selection for position: The role of left ventrolateral prefrontal cortex in sequencing language. *Brain and Language, 113(1)*, 28-38. https://doi.org/10.1016/j.bandl.2010.01.002

Tsuchida, A., & Fellows, L. K. (2013). Are core component processes of executive function dissociable within the frontal lobes? Evidence from humans with focal prefrontal damage. *Cortex, 49(7)*, 1790-1800. https://doi.org/10.1016/j.cortex.2012.10.014

Warrington, S., Bryant, K. L., Khrapitchev, A. A., Sallet, J., Charquero-Ballester, M., Douaud, G., ... & Sotiropoulos, S. N. (2020). XTRACT-Standardised protocols for automated tractography in the human and macaque brain. *NeuroImage, 116923*. https://doi.org/10.1016/j.neuroimage.2020.116923

Wilshire, C. E., & McCarthy, R. A. (2002). Evidence for a context-sensitive word retrieval disorder in a case of nonfluent aphasia. *Cognitive Neuropsychology, 19(2)*, 165-186. https://doi.org/10.1080/02643290143000169

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

Yourganov, G., Fridriksson, J., Rorden, C., Gleichgerrcht, E., & Bonilha, L. (2016). Multivariate connectome-based symptom mapping in post-stroke patients: networks supporting language and speech. *Journal of Neuroscience, 36(25)*, 6668-6679. https://doi.org/10.1523/JNEUROSCI.4396-15.2016