What’s in a sentence? The crucial role of lexical content in sentence production in nonfluent aphasia

Paula Speer and Carolyn E. Wilshire
School of Psychology, Victoria University of Wellington, Wellington, New Zealand

This study investigated the effect of lexical content on sentence production in nonfluent aphasia. Five participants with nonfluent aphasia, four with fluent aphasia, and eight controls were asked to describe pictured events in subject–verb–object sentences. Experiment 1 manipulated speed of lexical retrieval by varying the frequency of sentence nouns. Nonfluent participants’ accuracy was consistently higher for sentences commencing with a high- than with a low-frequency subject noun, even when errors on those nouns were themselves excluded. This was not the case for the fluent participants. Experiment 2 manipulated the semantic relationship between subject and object nouns. The nonfluent participants produced sentences less accurately when they contained related than when they contained unrelated lexical items. The fluent participants exhibited the opposite trend. We propose that individuals with nonfluent aphasia are disproportionately reliant on activated conceptual–lexical representations to drive the sentence generation process, an idea we call the content drives structure (COST) hypothesis.

Keywords: Aphasia; Nonfluent aphasia; Broca’s aphasia; Sentence production; Lexical retrieval; Lexical competition.

In nonfluent aphasia, particularly Broca’s aphasia, speech is often sparse and fragmented. Utterances may consist of only one or two words at a time. Sentences, if they do occur, are limited to very basic subject–verb–object (SVO) structures and idiomatic phrases. Also, in some individuals, speech may be agrammatic, consisting entirely of content words, such as nouns and verbs, but lacking appropriate grammatical function words and inflections (Goodglass, Kaplan, & Barresi, 2001; Rochon, Saffran, Berndt, & Schwartz, 2000; Saffran, Berndt, & Schwartz, 1989). However, performance on single-word production tasks, such as picture naming, is often comparatively well preserved (e.g., Freedman, Martin, & Biegler, 2004; Schwartz & Hodgson, 2002; Scott...
& Wilshire, 2010; Williams & Canter, 1982). For example, Williams and Canter (1982) showed that individuals with Broca’s aphasia, as a group, were more accurate at producing the names of pictures when presented in isolation than when they were part of a larger scene, which had to be described using sentences. Interestingly, a group of individuals with Wernicke’s aphasia showed the exact opposite pattern. In a case study, Schwartz and Hodgson (2002) report a similar pattern for their participant M.P., who was diagnosed with nonfluent aphasia. She performed relatively well in standard picture naming (73% correct), but struggled to produce these very same words when asked to describe complex scenes which incorporated several pictures at once (23% correct).

Theoretical accounts of the sentence production impairment in nonfluent aphasia have focused primarily on individuals with classic “agrammatic” speech, in which the grammatical elements of speech are systematically omitted. This speech pattern is often attributed to a difficulty constructing an appropriate syntactic frame for a sentence—one that specifies the inter-relationships amongst sentence constituents, their sequential order, and also any necessary grammatical function words or inflections (e.g., Caramazza & Hillis, 1989; Goodglass, Christiansen, & Gallagher, 1994; see Bastiaanse & Jonkers, 2012, for a recent review). For example, Thompson and colleagues have proposed that individuals with this agrammatic profile have difficulty accessing information about verb argument structure (e.g., Lee & Thompson, 2004; Thompson, 2003; Thompson, Lange, Schneider, & Shapiro, 1997). Verb argument structure refers to the number and types of entities that can participate in the event described by the verb, which can range from one (e.g., The boy runs) to three (e.g., The boy gave the present to the girl). If this information is unavailable, an appropriate syntactic frame for the sentence cannot be constructed. Support for this proposal comes from the finding that individuals with Broca’s aphasia, considered as a broad group, are generally less accurate at producing verbs than nouns (e.g., Breedin, Saffran, & Schwartz, 1998; Chen & Bates, 1998; Miceli, Silveri, Villa, & Caramazza, 1984; Zingeser & Berndt, 1988, 1990) and are particularly poor at producing verbs that carry multiple arguments (e.g., Dragoy & Bastiaanse, 2010; Kim & Thompson, 2000; Thompson & Choy, 2009).

Other proposals characterize agrammatic speech in terms of the specific syntactic operations that are compromised. For example, Bastiaanse and van Zonneveld (2005) make a primary distinction between sentences that follow the “base” word order for the language (e.g., SVO sentences in English, such as The dog chased the cat) and “derived” sentences (e.g., wh- questions, such as Who is the dog chasing?). They predict that derived sentences involve additional syntactic computations and are therefore more likely to be failed by agrammatics, a prediction that has been supported in a number of studies (e.g., Bastiaanse & Thompson, 2003; Burchert, Meißner, & Bleser, 2008). Alternatively, Friedmann, Grodzinsky, and colleagues postulate a hierarchy of syntactic relations that need to be specified in a sentence (Friedmann, 2006; Friedmann & Grodzinsky, 1997). Some of these operate over low-level constituents, such as individual noun phrases (e.g., noun number agreement: those books), and others operate over larger units such as entire clauses (e.g., tense agreement: Yesterday, the boy wrote), or even across multiple clauses (e.g., The children whose work the teacher read were proud). A severe agrammatic impairment could compromise the specification of syntactic relations at all levels, even the lowest ones. A milder impairment, on the other hand, might impact on higher levels, but leave the lowest levels intact—for example, agreement within a noun phrase. This prediction has been supported in some studies (Friedmann & Grodzinsky, 1997; but see also Burchert, Swoboda-Moll, & de Bleser, 2005; Lee, Milman, & Thompson, 2008). Recently, Bastiaanse and colleagues have suggested that syntactic/thematic relations that span more than one sentence may be particularly severely compromised in agrammatism. Examples are pronominal references (e.g., The man was fat. He walked with a stick) and references to tense/time frames (e.g., I saw the doctor
yesterday. He *told* me to rest up). These types of relations may even operate across different speakers within the same discourse (e.g., *Q: Which of these cakes are gluten free? A: The pink ones are;* Bastiaanse, 2008; Dragoy, Stowe, Bos, & Bastiaanse, 2012). According to these various syntactic accounts, nonfluent—or more specifically, agrammatic—speech is seen as a selective difficulty establishing the structural relations amongst elements in the sentence. The processes involved in generating the actual lexical content, on the other hand, are considered to be relatively preserved.

Most of these grammatical accounts of nonfluent aphasia focus on issues of representation rather than process—that is, they aim to describe the types of relational representations that cannot be generated, while putting to one side the exact nature of the cognitive processes that are impaired. Consequently, they are not well equipped to deal with the kind of variability in performance that can occur in real-life speech production situations, in which time constraints and working memory demands vary widely. For example, there is considerable variability from case to case. For example, some nonfluent individuals omit function words consistently, while others succeed in producing them in some contexts but not others. Also, some individuals systematically omit all grammatical morphemes, but others omit mainly free-standing ones (e.g., *to, for*), and yet others omit mainly bound ones (e.g., *jump;* Miceli, Silveri, Romani, & Caramazza, 1989; Rochon et al., 2000). Also, there is often substantial variability within individuals (e.g., Bastiaanse, 1995; Beeke, Wilkinson, & Maxim, 2007; Hofstede & Kolk, 1994; Kolk, 2006, 2007; Kolk & Van Grunsven, 1985; Nespoulous, 2000; Sahraoui & Nespoulous, 2012). For example, Sahraoui and Nespoulous (2012) found that agrammatic participants were more likely to produce complete, grammatically well-formed sentences in a picture description task than they were in an open-ended interview (for additional supporting evidence, see Beeke, Maxim, & Wilkinson, 2008; Beeke, Wilkinson, & Maxim, 2003; Heeschen & Schegloff, 1999; Salis & Edwards, 2004).

Several processing-oriented accounts have been put forward to explain some of these sources of variability. For example, the framework of Kolk and colleagues emphasizes the importance of timing in sentence production. It suggests that a key problem in some individuals may be a lack of synchronicity between the activation of lexical elements and their associated “slots” in a syntactic frame (Hartsuiker & Kolk, 1998; Kolk, 1995, 2006; Kolk & Van Grunsven, 1985). If the activation for a given lexical element does not peak during the period in which its associated syntactic slot is available, the dominant syntactic constituent may be incomplete, or entirely omitted (Kolk, 1995). This lack of synchronicity may be due to a number of factors, including slow retrieval of syntactic frame information, slow retrieval of lexical content items, or fast decay of either of these types of representations (Kolk & Van Grunsven, 1985). Within this framework, between-participant variability may be attributed to different decay and/or retrieval functions in different individuals (Kolk, 2006). Within-participant variability may be attributed to differences in the timing demands imposed by different conversational contexts and to the specific structure and/or content of the sentences being planned. Importantly, this type of explanation is not limited to classic agrammatic speech—in fact, Kolk and colleagues suggest that at least some of the features of agrammatism are not themselves indications of an underlying grammatical impairment, but rather reflect the use of compensatory strategies designed to maximize the amount of information the speaker can produce in a limited time frame (see also de Roo, Kolk, & Hofstede, 2003; Ruiter, Kolk, & Rietveld, 2010; Salis & Edwards, 2004).

A crucial component of processing-oriented frameworks, such as that of Kolk and colleagues (e.g., Kolk, 2006; Kolk & Van Grunsven, 1985), is that sentence production success in real time depends not just on the utterance’s syntactic structure, but also on its lexical content—even when lexical content retrieval is not itself compromised in that individual. Even if the process of frame construction takes place relatively independently
of lexical content retrieval, the lexical content items still need to be available at the appropriate time for integration into the sentence plan. However, many models of normal sentence planning go further, suggesting that the processes of structure generation and lexical content retrieval can directly influence one another. In several such models, lexical content elements are capable of activating representations of their respective syntactic constituents and vice versa (e.g., J.K. Bock, 1986; Mackay, 1987; Stemberger, 1985). In others, lexical elements influence structural planning via an intermediate set of conceptual/thematic units (Chang, 2002; Chang, Dell, & Bock, 2006). In these kinds of models, a highly activated lexical item can have a direct impact on the syntactic structure of the resultant utterance. There is considerable supporting evidence for this view from studies of normal sentence production. For example, if one key lexical item is primed ahead of time (e.g., dog in The cat chased the dog), speakers are more likely to select a sentence plan that places this element early in the sentence, even when it means opting for a noncanonical structure (e.g., The dog was chased by the cat; K.J. Bock, 1986; Bock, 1987; Ferreira & Engelhardt, 2006; Konopka & Bock, 2009; Levelt & Kelter, 1982; Pickering & Branigan, 1998). If currently activated lexical content can drive the processes of structure generation in this way, or at least strongly support it, then individuals with structure generation difficulties may benefit particularly strongly from the ready availability of lexical content.

So far, however, lexical aspects of sentence production in nonfluent aphasia have received little attention. The theoretical focus has been squarely on syntactic and structural aspects of speech planning, and, consequently, most studies have explored the effects of syntactic structure on production while minimizing the influence of lexical content—for example, comparing different types of structures while keeping lexical content constant and/or minimizing the impact of lexical variables by using very easily accessible words. In this study, we take the opposite approach: We keep the syntactic structure of our target sentences constant and manipulate their lexical content. If we take the view that content and structure generation processes are interactive, then sentence production accuracy in nonfluent aphasia may be particularly powerfully influenced by the availability of lexical content.

The idea that sentence production may depend crucially on the availability of the lexical content elements is almost self-evident in its simplicity, yet the specifics of this dependency are rarely examined. While no previous studies have directly addressed these hypotheses (at least not to our knowledge), there are at least three studies in the literature that may speak to the issue of how lexical availability impacts on nonfluent aphasic sentence production. In the first study, participants were provided with written lexical prompts just prior to describing a picture scene in a sentence (e.g., The girl tickles the boy; Faroqi-Shah & Thompson, 2003). Arrows and other types of prompts were used to indicate the type of sentence to be produced (active or passive). A group of seven individuals with Broca’s aphasia failed to demonstrate reliably improved accuracy when the uninflected verb and key nouns were provided in written form on the picture. Unfortunately, the authors do not report individual data. Also, the authors scored the participants’ best effort instead of their first response, a procedure that may not reflect the challenges of online sentence production where time restrictions may not allow for multiple attempts at the sentence.

Two further studies provide some less formal evidence suggesting that heightened lexical availability may improve sentence production accuracy more generally. One examines a therapy software package called SentenceShaper®, which allows participants to record, replay, and even reorder words and phrases they want to produce (Linebarger, McCall, Virata, & Berndt, 2007; Linebarger, Schwartz, Romania, Kohn, & Stephens, 2000). In other words, it makes the relevant lexical content items more readily available and accessible during sentence construction. When using this treatment device, individuals with nonfluent aphasia have shown improvements in the length and the grammatical well-formedness of their sentences. For example, case D.D.’s description of a
scene changed from “Ohh! A fish! Ah, water and . . . ub mmmm and attendant, here, and bumped his head. Oh boy, oh my hand, my hand, my hand” to “The boy and the fishmonger is taking the fish. The boy hit his hand” (Linebarger et al., 2000, p. 422). Another recent therapy study trained a single individual with nonfluent aphasia to produce a specially selected set of nouns and verbs (Raymer & Kohen, 2006). After training, this person was more accurate at producing sentences featuring both the trained nouns and verbs than those featuring untrained items. Importantly, the training not only affected accuracy on the trained words themselves, but it also improved overall syntactic well-formedness, including the use of correct inflections and determiners.

In the current research, we explored the influence of lexical content on the production of simple SVO sentences in individuals with both fluent and nonfluent aphasia using a simple picture description task (e.g., The cat is chasing the dog). The simple canonical SVO structure was chosen to avoid floor effects: All of our participants could produce these sentences under at least some conditions. Further, the key nouns used in the sentences were well within each person’s vocabulary: Indeed, as we document below, although participants made very occasional errors on our key nouns during single naming, no single lexical item was consistently failed by any individual. Our classification of participants as nonfluent or fluent was done on the basis of standard diagnostic tools (the Boston Diagnostic Aphasia Examination; Goodglass et al., 2001), rather than on the incidence of agrammatic features, because we did not wish to make specific assumptions about the relationship between agrammatic features of speech and the underlying sentence production difficulties.

In Experiment 1, we systematically varied the frequency of the noun items to be included in the sentence. Based on the premise that individuals with nonfluent aphasia may rely abnormally heavily on the currently activated lexical elements to drive sentence structure planning, we hypothesized that the frequency of the lexical items, particularly those early in the sentence, would influence the production accuracy of the other sentence elements. In Experiment 2, we varied the semantic relationship between the subject and object nouns. It has been suggested that individuals with nonfluent aphasia cannot make effective use of a structural representation to maintain lexical elements in their correct order during sentence planning, and, consequently, those lexical elements are freer to compete with one another for selection to a particular position (see Freedman et al., 2004; Martin, Lesch, & Bartha, 1999). Therefore, in nonfluent aphasia, sentences containing semantically related subject and object nouns might be more prone to error overall than those containing unrelated ones.

**EXPERIMENT 1: EFFECTS OF LEXICAL FREQUENCY ON SENTENCE PRODUCTION**

Frequency is a variable known to influence the time course of lexical retrieval in normal individuals, and a robust effect of frequency has been demonstrated across a range of single-word processing tasks (e.g., Barry, Morrison, & Ellis, 1997; Oldfield & Wingfield, 1964). Recent research suggests that frequent words may be processed more rapidly at a number of stages, including the initial mapping from concept to lexical label and also the subsequent phonological specification of that label (Kittredge, Dell, Verkuilen, & Schwartz, 2008; Knobel, Finkbeiner, & Caramazza, 2008; Strijkers, Costa, & Thierry, 2010; see also Navarrete, Basagni, Alario, & Costa, 2006). Indeed, even when nouns are embedded into determiner + adjective + noun phrases, there is still a detectable noun frequency effect on phrase initiation time, suggesting that high noun frequency also reduces overall grammatical encoding time for a larger phrase (Alario, Costa, & Caramazza, 2002). Therefore, one simple way to manipulate the ease and speed with which lexical items can be made available for integration into a grammatical frame would be to vary their frequency of occurrence.
In Experiment 1, participants were asked to describe simple pictured events in a single SVO sentence (e.g., The cat is chasing the dog). We manipulated the lexical frequency of the key nouns. The picture stimulus disappeared as soon as the participant commenced their response. 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. The frequency of the first key noun in particular was predicted to have a powerful effect on the accuracy of the remainder of the sentence. In contrast, in fluent aphasia, the effect would be localized to the key nouns themselves.

Method

Participants

Controls. Eight older control participants aged between 63 and 84 years (\(M = 67.5, SD = 6.82\)), were recruited from the community. Three were males. All were native speakers of English, and none reported any significant neurological history.

Participants with aphasia. Nine participants with aphasia due to stroke were recruited from a register of volunteers willing to be contacted about research. All met the follow inclusion criteria: (a) Their stroke occurred more than 12 months prior to the commencement of the study; (b) their aphasia was classified as mild to moderate according to the Boston Diagnostic Aphasia Examination (BDAE) severity scale (Goodglass et al., 2001); and (c) they correctly named at least 50\% of the pictures in the Boston Naming test (long form; Goodglass et al., 2001). Further, all had normal or corrected-to-normal vision, and English was their native language. According to the BDAE, five of the participants were classified as having nonfluent aphasia (all had a diagnosis of Broca’s aphasia), and the remaining four as having fluent aphasia (two were diagnosed with anomic aphasia, one with conduction aphasia, and one with Wernicke’s aphasia). Table 1 presents diagnostic and background information for the aphasic participants and Table 2 summarizes their scores on key language tests and measures. Two points are worth noting here. First, speech rate, as assessed using the Quantitative Production Analysis (QPA; Saffran et al., 1989) was consistently lower for nonfluent than for fluent participants, and for the nonfluents, all scores were within one standard deviation of the mean reported by Rochon and colleagues for a group of individuals with Broca’s aphasia (Rochon et al., 2000). Second, on the QPA closed class ratio measure, scores for the nonfluent individuals varied widely; however, most nonfluent participants exhibited lower scores than the fluents (all but B.Y.). Graphical depictions of participants’ lesions, where available, are presented in Figure 1. Details of imaging methods and image preparation are given in Appendix A.

Materials

Norming of stimulus pictures. A set of 120 line drawings were drawn specifically for the study. Each displayed a scene that could be described in an SVO sentence (e.g., The pig is biting the bear). Examples are shown in Figure 2. The nouns depicted in the scenes were drawn from a pool of 18 monosyllabic nouns denoting people or animals. All items were early acquired—that is, under the age of seven according to Carroll and White (1973) and/or Morrison, Chappell, and Ellis (1997). The 18 nouns included nine low-frequency items (frequencies of 30 or less occurrences per million in the Subtitle Analysis Project; Brysbaert & New, 2009) and nine high-frequency items (frequencies in excess of 30 occurrences per million; see also Schnur, Schwartz, Brecher, & Hodgson, 2006, for a similar division).

Each target noun element appeared twice, in two different picture scenes, once as the subject (e.g., The pig is biting the bear) and once as the object (e.g., The bear is biting the pig). The noun elements in a given sentence were always from the same category: Animals only appeared with animals, and people only with other people. In all pictures, the subject was displayed on the left side of the picture. As shown in Figure 2, there
Table 1. Background, medical, and diagnostic information for each aphasic participant

<table>
<thead>
<tr>
<th>Participant</th>
<th>Age</th>
<th>Gender</th>
<th>Years post CVA</th>
<th>Lesion site/aetiology</th>
<th>BNT score (/60)</th>
<th>BDAE measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>B.Y.</td>
<td>59</td>
<td>male</td>
<td>37</td>
<td>Subarachnoid haemorrhage, subsequently operated upon, large lesion extending from anterior horn of L lateral ventricle to L parietal lobe</td>
<td>41</td>
<td>Broca’s</td>
</tr>
<tr>
<td>D.A.</td>
<td>71</td>
<td>male</td>
<td>11</td>
<td>Not known</td>
<td>53</td>
<td>Broca’s</td>
</tr>
<tr>
<td>J.G.</td>
<td>73</td>
<td>female</td>
<td>6</td>
<td>Isch. CVA, L MCA region</td>
<td>46</td>
<td>Broca’s</td>
</tr>
<tr>
<td>J.H.M.</td>
<td>52</td>
<td>female</td>
<td>10</td>
<td>Isch. CVA, Extensive L MCA</td>
<td>46</td>
<td>Broca’s</td>
</tr>
<tr>
<td>R.P.</td>
<td>66</td>
<td>male</td>
<td>9</td>
<td>Unspecified CVA, extensive L frontal and parietal and also right medial frontal infarct</td>
<td>38</td>
<td>Broca’s</td>
</tr>
<tr>
<td>N.P.</td>
<td>73</td>
<td>male</td>
<td>13</td>
<td>Isch. CVA, several foci in L occipital and temporal lobes</td>
<td>32</td>
<td>Anomia</td>
</tr>
<tr>
<td>S.T.R.</td>
<td>81</td>
<td>female</td>
<td>13</td>
<td>Isch. CVA, possibly multiple, infarcts in R occipital and L parietal lobe</td>
<td>53</td>
<td>Anomia</td>
</tr>
<tr>
<td>S.W.</td>
<td>82</td>
<td>female</td>
<td>4</td>
<td>Haem. CVA, L posterior temporal lobe</td>
<td>37</td>
<td>Wernicke’s</td>
</tr>
<tr>
<td>W.L.</td>
<td>64</td>
<td>male</td>
<td>2</td>
<td>Isch. CVA, L parietal and L posterior temporal lobe</td>
<td>40</td>
<td>Conduction</td>
</tr>
</tbody>
</table>

Note: Isch = ischaemic; haem = haemorrhagic; CVA = cerebrovascular accident; BNT = Boston Naming Test; BDAE = Boston Diagnostic Aphasia Examination; MCA = middle cerebral artery; L = left; R = right.
were four conditions representing the four different possible combinations of high- and low-frequency nouns, and 30 sentence exemplars for each condition. Fifteen different transitive verbs were depicted in the scenes, and their occurrence was balanced, each verb appearing twice in each of the four frequency conditions.

Forty-one psychology undergraduate students aged between 17 and 42 years (\(M = 19.63, SD = 5.47\)), all native English speakers, took part in the picture norming study for course credit. Eight participants were male. Each participant was given a booklet containing all 120 stimulus pictures, and beneath each picture they had to write down a sentence that best described what was happening in the picture. The criterion for selection of a picture for the main experiment was that it elicited the target construction (as defined above) in at least 80% of participants, not penalizing for substitutions of contextually appropriate alternative verbs that did not change the sentence’s grammatical structure (e.g., eating → biting). Since each target item appeared twice, in two different picture scenes—one in subject and once in object position—both pictures needed to reach this criterion to be selected.

**Composition of final stimulus materials.** A total of 66 pictures met the above criteria: 16 high-frequency (HF) subject and HF object, 18 low-frequency (LF) subject and LF object, 16 HF subject/LF object, and 16 LF subject/HF object (for a full list see Appendix B). Each target picture was rated by seven participants with regards to the plausibility of the event it depicted (using a 5-point Likert scale: 1 = very improbable,
Figure 1. Lesion maps for participants with nonfluent (Panel a) and fluent aphasia (Panel b) showing axial slices of the brain on a standard template (Rorden et al., 2012). Slices were selected according to representative display of individual lesions (corresponding Montreal Neurological Institute, MNI, Z coordinates are reported above each slice). Further details of imaging methods and image preparation are described in Appendix A. [To view this figure in colour, please see the online version of this journal.]
There were no significant differences of plausibility between the pictures from the different subject frequency conditions ($p = .71$) and object frequency conditions ($p = .50$).

To avoid priming of the SVO sentence structure (J.K. Bock, 1986), an additional 66 distractor pictures that elicited a different syntactic structure were interspersed amongst the target pictures, thus giving a total of 132 pictured events. None of the elements in the target pictures was featured in the distractor pictures. Three different fixed randomized versions of the experiment were created, each of which contained the same set of 132 pictured events, and differed only in their order.

In addition, 18 single pictures of the target elements depicted in the events were chosen from freely accessible picture pools. All had been previously used in our laboratory, and all had elicited name agreement of 80% or higher when tested on normal controls. These single pictures were used to create a single item naming pretest, to be administered prior to the main sentence production task in each session.

Figure 2. Examples of the pictured scenes from different experimental conditions.
Procedure
All participants were tested individually. Testing began with the pretest naming task. Each pictured object was presented individually in the centre of a laptop computer screen, and participants were asked to name it. Each naming trial began with a fixation cross, followed 1000 ms later by the target picture, accompanied by a tone. The picture remained on the screen until the response was completed.

Following the naming pretest, participants were informed that they would now see a series of pictured events and that they had to describe what was happening in each picture in one sentence. Instructions were presented both aurally and in written form. Each trial began with a fixation cross that remained on the screen for 1000 ms, followed immediately by the picture, accompanied by a tone. As soon as the participant began to vocalize their response, the experimenter manually pressed a key, and the picture immediately disappeared from the screen. If there was no response in 5000 ms, the picture automatically disappeared. Eight practice items were given prior to the experimental items in each session.

All participants completed all three versions of the task, in different testing sessions, separated by at least a week. Each session commenced with the pretest naming task. However, J.H.M. and D.A. were able to complete only half of each session per visit, so for them, the entire experiment was spread over six sessions; also, S.W. was unable to produce any of the target sentences once the picture had disappeared from the screen at speech onset, so the procedure was modified to enable her to view the picture throughout her response. All sessions were recorded using a digital tape recorder.

Response scoring
For the naming pretest, each individual’s first response to each picture was scored as either correct (identical to the target) or incorrect. Naming latencies were measured manually from the digitized recordings, from the onset of the tone that accompanied the picture until the onset of the first correct response. For the main sentence production task, only responses to the target pictures were scored; the filler pictures were not analysed. A response was scored as correct if it included the two target nouns and the target verb (or a permissible variant: See below), and these were incorporated into a thematically and grammatically correct sentence. For example, passives like “The bear is being bitten by the pig” or possessive phrases like “The pig is biting the bear’s leg” were not counted as errors. However, such responses were not included in the latency analysis. 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. In this analysis, substitution of the target noun with a pronoun (e.g., it instead of pig) was counted as an error. Verbs were scored using the same criteria, except that substitution of the target verb with a meaning-related one was allowed when it did not substantially change the meaning of the sentence (e.g., The pig is eating the bear instead of The pig is biting the bear).

Sentence initiation latencies were measured manually from the tone accompanying picture onset to speech initiation. Incorrect responses were not submitted to the latency analysis, except if the only error was a phonological error involving the verb or one or both nouns. Also, responses that were preceded by a comment about a previous item (e.g., ...“Oh, that was not a dog”) were excluded from the latency analysis. The latency data, particularly for aphasic participants, was highly positively skewed. Consequently, the data for all analyses were first trimmed of outliers (values 3 SDs outside each participant’s grand mean) and log transformed (see Biegler, Crowther, & Martin, 2008, for advantages of this procedure).

General statistical methods
Wherever possible, our analyses were conducted across items—that is, rather than collapsing across all trials of the same kind, each individual trial was treated as a separate measure. Accuracy data (correct/incorrect responses) were analysed using logistic regression, conducted using the SAS Genmod procedure. In this procedure, the
A regression model is built using generalized estimating equations (GEEs), which enable the researcher to model the effect of a repeated measure—in this case, participant and/or target item. In analyses at the group level, participant and target item were included as the repeated measures, while analyses at the individual level only included one repeated measure: target item. In all analyses, we employed the standard approach of removing nonsignificant predictor variables ($p > .10$) from the model in a stepwise manner, always beginning with the interactions and then moving on to the main effects. In some cases, when error rates were very low, and the vast majority of cells contained a zero, we supplemented the logistic regression analyses with an analysis of the aggregated error totals. Under these conditions, logistic regression has been shown not to detect effects reliably (Peduzzi, Concato, Kemper, Holford, & Feinstein, 1996). For these aggregated error analyses, we used the $Q'$ test (Michael, 2007), a non-parametric test for the analysis of small samples of nominal data. It is preferable to the chi-square test in that it does not assume independence of the measures.

For the analysis of latency data, we submitted the data for each item and each individual to a general linear mixed model analysis (also known as mixed effects models or linear mixed effects models, LMEs: Diggle, 1988) using the SAS Mixed procedure. In individual analyses, the model included only one random effect: item. In group analyses, the model also included participant as a random factor (or fixed effect; see Baayen, Davidson, & Bates, 2008, for more information on the application of this technique to psycholinguistic data). In models with two random factors, the factors were non-nested. In all mixed effects model analyses, we employed the standard approach of removing nonsignificant predictor variables ($p > .10$) from the model in a stepwise manner, always beginning with the interactions and then moving on to the main effects. Covariance structure selected for all models was variance components, which gave the best model fit in the majority of the cases.

**Results**

**Naming pretest**

Table 3 shows the incidence of naming accuracy and latency data for each of the three participant groups (controls, nonfluent, and fluent). None of the participants failed any item consistently across all three replications of the naming task.

<table>
<thead>
<tr>
<th>Table 3. Error rates and naming latencies in the pretests as a function of frequency for the three participant groups</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Experiment 1</th>
<th></th>
<th></th>
<th>Experiment 2</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>High Freq</td>
<td>Low Freq</td>
<td>Difference (low – high)</td>
<td>N</td>
<td>High Freq</td>
</tr>
<tr>
<td><strong>Total errors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonfluent</td>
<td>270</td>
<td>4</td>
<td>2</td>
<td>–2</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Fluent</td>
<td>216</td>
<td>5</td>
<td>11</td>
<td>6</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>Controls</td>
<td>432</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td><strong>Latency</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonfluent</td>
<td>na</td>
<td>1182</td>
<td>1628</td>
<td>446**</td>
<td>na</td>
<td>1109</td>
</tr>
<tr>
<td>Fluent</td>
<td>na</td>
<td>1413</td>
<td>1669</td>
<td>256(*)</td>
<td>na</td>
<td>1251</td>
</tr>
<tr>
<td>Controls</td>
<td>na</td>
<td>744</td>
<td>818</td>
<td>74</td>
<td>na</td>
<td>765</td>
</tr>
</tbody>
</table>

*Note: Experiment 1: Due to the very low error rates, logistic regression analyses of individual participants’ accuracy data could not be performed (the model could not be estimated). Individuals who showed a significant frequency effect on naming latencies were B.Y. ($p < .05$), J.H.M. ($p < .05$), R.P. ($p < .01$). Significant group by frequency interactions on naming latencies were found when comparing the nonfluent and the control groups ($p < .001$). Experiment 2: Individuals who showed a significant frequency effect on naming latencies were B.Y. ($p < .05$), J.H.M. ($p < .05$), N.P. ($p < .05$). Across groups, there were no significant group by frequency interactions. na = not applicable; na = not applicable.
Group analyses using logistic regression revealed no significant frequency effects for any of the three participant groups considered alone, and there were no significant group by frequency interactions. As can be seen in Table 3, the control group showed a trend towards longer naming latencies for low-frequency items, but this failed to reach significance. Each aphasic participant also revealed a consistent trend towards slower naming latencies for low- than for high-frequency nouns. The effect reached significance for the nonfluent group as a group and was marginally significant for the fluents as a group.

Comment. The aphasic individuals made very few errors in this simple picture naming task. Of course, their performance may not be representative of their word finding difficulties in everyday life: The words featured in our task had been practised many times. However, for the purposes of our study, we are able to conclude with confidence that all the target nouns featured in our sentences were “within the vocabulary” of each participant (that is, they could retrieve and produce each noun on at least some occasions). The analysis of naming latencies served as a manipulation check for the frequency manipulation. It revealed a consistent trend towards slower naming latencies for low- than for high-frequency nouns.

Main task
Response accuracy. Before addressing our key hypotheses, it is worth examining the overall patterns of noun production accuracy across sentence positions, irrespective of frequency. Panel (a) of Figure 3 shows the overall incidence of noun errors in each aphasic individual and in the control group as a function of their position in the sentence. Also shown for comparison is the incidence of errors on the same nouns during the pretest naming task. Controls’ noun production accuracy did not differ for subject and object positions ($p = .09$). In contrast, all nonfluent participants produced more errors on object nouns than on subject nouns. This difference was significant for the group as a whole, $\chi^2(1) = 43.76, p < .0001$, and also for every one of the nonfluent aphasics considered individually (see Figure 4 for significance levels for the individual subject analyses)\(^1\). Indeed, if we also consider their rates of noun errors in the pretest naming task, we observe a consistent linear pattern, in which errors were least common when the nouns were produced in isolation, moderately common when the noun appeared in subject position, and most common when the noun appeared in object position.

As a group, the fluent aphasics produced more errors on object than on subject nouns, $\chi^2(1) = 5.46, p < .05$, but this effect was not consistently observed across all participants. Finally, a direct comparison between the nonfluent and fluent groups revealed a significant group by error position interaction, $\chi^2(1) = 7.35, p < .01$, indicating that the nonfluents’ and fluents’ noun production accuracy was differently affected by sentence position.

Turning now to our key hypotheses, Figure 4 shows the overall incidence of incorrect sentence responses for each group and aphasic individual, broken down by subject noun frequency (a full breakdown of the error data by subject and object frequency is given in Appendix C). As illustrated in Figure 4, control participants produced more incorrect responses overall on sentences with a low-frequency subject than on those with a high-frequency subject, $\chi^2(1) = 7.10, p < .01$.\(^2\) The nonfluent group showed a difference in the same direction, but greatly exaggerated, $\chi^2(1) = 9.89, p < .01$. Direct comparison between the

\[^1\] The individual analyses were performed using repeated measures logistic regression, which included the predictor variables noun position (subject vs. object), subject noun frequency, and object noun frequency. The number of errors made in each error position on each sentence item served as the dependent variable, which ranged from zero (no errors) to three (errors in all three sessions). The distribution of the model was set to multinomial in order to capture the nonbinomial nature of the dependent variable.

\[^2\] All errors made by control participants were semantic/visual substitutions, out of which 34% were self-corrections [e.g., the dog (target: goat) oh goat is biting the cat].

Cognitive Neuropsychology, 2014 13
nonfluent and control groups using logistic regression failed to reveal a significant group by subject frequency interaction ($p = .65$). However, because of the extremely low error rates for the controls on this task, logistic regression may be inappropriate; indeed, an analysis of aggregated error scores using the $Q'$ test did reveal a significant interaction, $Q'(1) = 15.45, p < .001$, suggesting that the subject frequency effect may indeed be more marked in the nonfluent group than in controls.

For the fluent group, there was a trend toward a reverse subject frequency effect (fewer errors on sentences with low- than with high-frequency subjects), but this failed to reach significance for the group considered as a whole ($p = .33$). At the individual level, the frequency effect reached significance only for S.W. (see Figure 4). Direct
comparison of the fluent and the nonfluent groups revealed a significant group by subject frequency interaction, $\chi^2(1) = 8.04, p < .01$, indicating that subject frequency had a reliably different effects across the two aphasia groups.

Finally, to ensure that the subject frequency effects were not just due to localized errors on the subject nouns themselves, Figure 4 also gives overall sentence errors not penalizing for subject errors. The error patterns stay the same for all individuals and groups.

**Analyses of response latency.** The analysis of latency data served two purposes: First, it served as a manipulation check for our frequency manipulation. We expected that participants would take longer to produce sentences containing low-frequency elements, particularly when they appeared in subject position. Secondly, latency data may shed light on the sentence planning strategies of our participants. For example, if participants planned the entire SVO utterance in advance, including the subject and object noun, sentence initiation time may also be slower for sentences containing a low-frequency object noun. We would not expect such an effect if participants planned the sentence more incrementally. Furthermore, the time interval between the verb offset and the onset of the object noun may provide additional clues as to planning strategies. If time between verb offset and object noun onset is not influenced by object frequency, this would suggest that participants had already retrieved that noun prior to this time. In contrast, if participants planned the sentence more incrementally, they may show longer verb–object noun latencies for low-frequency object nouns. Finally, an interaction of subject noun and object noun frequency may indicate a temporal overlap between the retrieval of the subject noun and the object noun.

Before the latency data were analysed, outliers were removed. Collapsed across all individuals, this resulted in the removal of an average of 0.9% for the control group and 2.0% of data for the aphasic participants, considered together. An additional 2.8% of data points (across all groups) were lost due to technical problems or other disturbances (like background noise).
as a function of subject frequency condition for the three participant groups and also for aphasic individuals. At the group level, the findings were very similar for each of the three groups: Sentence initiation times were significantly influenced by subject frequency: controls, $F(1, 1337) = 16.21, p < .0001$; nonfluent, $F(1, 672) = 17.89, p < .0001$; fluent, $F(1, 440) = 10.02, p < .01$.

To perform comparisons between groups, we combined the relevant data (nonfluent + fluent and nonfluent + controls), included the new predictor variable group, and reran the analyses. The trend illustrated in Figure 5 towards more marked frequency effects for nonfluent than for controls was statistically reliable. The nonfluent group not only took longer overall to initiate sentences than did controls, $F(1, 2066) = 75.83, p < .0001$, but also, there was a significant group by subject frequency interaction, $F(1, 2066) = 12.60, p < .001$. The nonfluent and the fluent groups themselves did not differ significantly from one another ($p = .26$), and there was no significant group by subject frequency interaction ($p = .48$).

Object frequency had a considerably weaker effect on sentence initiation times than subject frequency. Indeed, at the group level, none of the three groups showed any reliable effect (controls, $p = .76$; nonfluent, $p = .70$; fluent, $p = .27$). Analyses across groups also failed to reveal any significant group differences (nonfluent vs. controls, $p = .43$; nonfluent vs. fluent, $p = .58$). At the individual level, fluent participants N.P. and S.W. showed significant effects of object frequency on sentence initiation times. Nonfluent participants B.Y., R.P., and J.G. revealed a nonsignificant trend in this direction. However, these individuals exhibited a pattern that suggested that much of their sentence planning was being delayed until the interval between the determiner and the subject noun (an effective strategy, given that every sentence begins with the same determiner “The”). When their data were reanalysed, measuring sentence initiation time from the onset of the tone until the onset of the subject noun, the effect of object frequency was marginally significant for R.P. ($p = .07$) and was approaching a trend for J.G. ($p = .17$), but not for B.Y. ($p = .75$).

When considering the time between offset of the target verb and onset of the object noun, the analysis of control participants revealed a significant main effect of object frequency on this measure, $F(1, 1232) = 28.05, p < .0001$. Curiously, there was also a significant interaction between subject and object frequency, $F(1, 1232) = 4.17, p < .05$, indicating that object frequency

Figure 5. Subject frequency effects on sentence initiation time depicted as the difference in mean latencies between sentences containing a low-frequency subject noun and sentences containing a high-frequency subject noun. Bars represent one standard error above and below the mean for each participant group. *$p < .10$; **$p < .05$; ***$p < .01$; ****$p < .001$. 

Cognitive Neuropsychology, 2014
had a stronger influence on verb–object noun initiation times when the sentence began with a high-frequency than with a low-frequency subject. The analysis failed to reveal any reliable object frequency effects for the nonfluent and the fluent group (\(p = .27\) and \(p = .55\), respectively), nor any significant interactions between subject and object frequency (\(p = .16\) and \(p = .14\), respectively). Comparison across groups failed to reveal any reliable differences in their overall sensitivity to object frequency (nonfluent vs. controls, \(p = .66\); nonfluents vs. fluents, \(p = .70\)).

At the individual level, there were only effects of object frequency for nonfluent participants B.Y. and J.H.M., and for J.H.M., a significant interaction between subject and object frequency, \(F(1, 103) = 4.39, \ p < .05\) (verb–object latencies were shortest when the sentence contained both a high-frequency object and a high-frequency subject and were longest when the sentence subject was of high frequency and the sentence object of low frequency).

**Comment**

Consistent with the latency analysis of single noun naming, all participants were faster to initiate sentences beginning with a high- than with a low-frequency subject, indicating that lexical retrieval of the subject noun had already commenced at the time of utterance initiation. However, not all participants showed an effect of object frequency, suggesting they may not have begun to retrieve the lexical label for the object element at that time. Indeed, controls showed no object frequency effect, which is consistent with previous studies of normal speakers, in which lexical elements have been found to influence sentence initiation times only when they appear in the initial noun phrase (e.g., Ferreira, 1991; Griffin, 2001; Martin, Miller, & Vu, 2004; Smith & Wheeldon, 1999). There was no significant effect of object frequency at the group level for either the nonfluent or the fluent participants. However, some specific individuals within each group did show trends in this direction, which reached statistical significance in two cases (fluent participants N.P. and S.W.). This pattern suggests a strategy of more extensive advance planning in these participants than in the other aphasic participants and controls.

In the latency analyses, we also examined the time interval between the verb offset and the onset of the object noun. In controls, this interval was reliably longer for low-frequency than for higher frequency object nouns, suggesting that object noun retrieval may still not have been complete at the time the verb was produced. Again, this pattern suggests an incremental, phrase-by-phrase planning strategy in controls. Few of the aphasic individuals showed this effect, but two exceptions were nonfluent cases B.Y. and J.H.M. Interestingly, for controls and for J.H.M., there was also a significant subject frequency by object frequency interaction, characterized by larger object frequency effects when the subject of the sentence was of high frequency. This latency pattern indicates a temporal overlap between the retrieval of the subject noun and the object noun, which was only observable when the subject noun was produced so quickly that retrieval of the lexical label for the object noun had not yet been completed.

**Discussion**

Experiment 1 yielded two important findings. First, our participants with nonfluent aphasia exhibited a very distinctive and consistent pattern in their noun production errors: They were less accurate at producing nouns in the context of a sentence than they were in producing nouns in isolation, and further they were less accurate at producing nouns in object position than in subject position. This pattern was not consistently observed in the fluent aphasia participants, whose performance across the three context conditions was more variable. Second, in SVO sentence production, the frequency of the subject noun had a powerful influence on the nonfluent’s sentence production accuracy, even when the accuracy measure excluded the subject noun itself. Again, this effect was not observed for the fluent aphasia participants. The control participants showed similar frequency effects in the sentence accuracy measures to those for the nonfluent individuals—however, not
surprisingly, of much smaller magnitude. In other words, subject noun accessibility appears to have a profound effect on downstream sentence production accuracy and well-formedness only in individuals with nonfluent aphasia. These findings support our initial hypothesis that the frequency of lexical target items—at least when they appear early in the sentence—has more generalized effects on sentence production accuracy in nonfluent than in fluent aphasia.

Within a Kolk-type framework (e.g., Kolk, 2006), both these features could be interpreted as an inability to retrieve early sentence elements in sufficient time to enable late-appearing lexical elements to be integrated into the current representation of the sentence frame. However, there was no direct evidence to suggest that lexical retrieval itself was any slower in our nonfluent than in our fluent aphasic participants: Single picture naming latencies did not differ significantly for the two groups, nor did overall sentence initiation times. Both participant groups were much slower than the controls, but both were slow to the same extent. An alternative possibility within the Kolk framework is that syntactic frame construction may itself be delayed, so that the activation levels of some lexical elements have already begun to decay by the time their corresponding frame elements become available. However, this would in fact predict higher accuracy for sentences containing low-frequency elements, since the slower words are retrieved, the greater their synchronicity with the delayed frame construction process. This was the opposite of what we found.

A more plausible explanation is that in nonfluent aphasia, the structure generation process itself is slow or ineffective, so these individuals are more heavily reliant on the representations of the lexical constituent elements themselves to support a structural plan for the sentence. This idea could in principle be accommodated within a more formal model of multiple word planning, such as that proposed by Randi C. Martin and colleagues (Freedman et al., 2004; Martin & Freedman, 2001; Martin et al., 1999, 2004; Martin & Romani, 1994; Martin, Shelton, & Yaffee, 1994). Martin and colleagues postulate a system, called semantic short-term memory, which operates to maintain ordered lexical elements in a temporarily heightened state of activation, for example during the planning of a larger utterance. Semantic short-term memory is itself contentless, consisting merely of a series of place-holders (or “slots”), which can be tied to particular content elements as needed. During production, representations of the lexical elements to be produced are connected to their respective place-holders. Interactive flow of activation between these units and the lexical elements ensures that information in the buffer is maintained over the course of production. Once it is time to produce a particular word, an attention shift is directed to the relevant “slot” unit, which boosts its activation and consequently that of its associated lexical element (Freedman et al., 2004).

This model is able to capture some of the ways that content and structure processes interact in speech—at least in the context of very simple utterances, such as noun phrases. Within this framework, inefficient binding of content elements to their respective “slots” and/or abnormally rapid decay of this information could result in a failure to produce the utterance correctly. Crucially, the effects of such an impairment would be modulated by lexical factors. Since there is a bidirectional flow of activation between elements and their “slots”, strongly activated lexical elements (such as those for common, frequently used words) may be capable of being maintained for longer than less strongly activated ones. This effect, which may be present but difficult to detect in normal healthy speakers, could be considerably exaggerated in individuals who are ineffective at binding lexical elements to their slots. Of course, this explanation remains speculative, since Martin and colleagues’ (e.g., Martin & He, 2004; Martin & Freedman, 2001) framework does not deal with utterances beyond a single phrase, nor does it explicitly address the question as to how frequency influences the structure-element binding process. Also, although several of the nonfluent cases reported here have shown other features consistent with an impairment to a semantic short-term memory, none were tested on those tasks
considered to be most diagnostic of this condition (namely, disproportionate performance on a category probe task, when compared with that on a rhyme probe task: see Freedman & Martin, 2001; Martin & He, 2004; Martin et al., 1994).

One final aspect of the results from Experiment 1 warrants some discussion. In this experiment, the stimulus picture disappeared from view as soon as the participant commenced their response. Our rationale for doing this was to promote advance planning of the entire sentence. However, it remains possible that in some participants, the intended conceptual message might have decayed from working memory before they had an opportunity to encode each element. If the nonfluent participants were more prone to this problem than the fluent participants, it could account for why they performed more poorly on sentences with low-frequency subject nouns. However, our results provide little support for this hypothesis. First, all the nonfluent participants were able of completing the task effectively and could correctly produce at least some of the sentences in all conditions; indeed, the only participant who could not complete the task under the original conditions was S.W., who had fluent (Wernicke’s) aphasia. Second, much of the delay induced by the subject noun frequency manipulation is likely to have occurred before the sentence was initiated, and before the picture had been removed from view. Third, there is direct evidence from a number of participants—both fluent and nonfluent—that lexical-level planning of the object element had begun in earnest even before they had begun to initiate their response. This would seem to preclude an explanation in terms of decay of the relevant conceptual representations. We return to this point in the discussion of Experiment 2 below.

EXPERIMENT 2: EFFECTS OF NOUN SEMANTIC RELATEDNESS ON SENTENCE PRODUCTION

In Experiment 1, the semantic relationships between the subject and object nouns were not specifically manipulated, but kept constant. However, according to recent research, individuals with nonfluent aphasia may be particularly susceptible to interference effects when they are planning to produce two semantically related words within the same utterance (e.g., Freedman et al., 2004; Scott & Wilshire, 2010). For example, Freedman and colleagues showed that two individuals with nonfluent aphasia were significantly slower to name noun phrases containing semantically related nouns (e.g., nose and mouth) than they were to name phrases containing unrelated nouns. This effect was not observed in a comparison case with fluent aphasia. Freedman and colleagues (2004) postulated that this semantic interference effect reflects the poor functioning of the semantic short-term memory buffer. When this buffer, which normally maintains ordered lexical information during phrase production planning, is operating ineffectively, the activated representations of the nouns planned for the phrase will compete more with each other for selection. This effect will be particularly evident if the words are also semantically related, since semantically related items will tend to further activate one another via their shared semantic representations.

A more informal account for such effects is that a robust structural representation of the target utterance helps to manage the naturally competitive influence of one word on another when both have similar selection constraints. If this representation is absent, lexical items may be freer to compete for selection for the same position in the utterance plan.

In Experiment 2, we explored whether our participants with nonfluent aphasia might show exaggerated semantic interference effects over SVO sentences, where the two key lexical items are in different noun phrases. Specifically, their accuracy may be poor for sentences containing two related nouns than for those containing unrelated nouns. Further, initiation times may also be slower for sentences containing related elements, at least for those individuals who plan their sentences substantially in advance. These effects would not be expected to occur in individuals with fluent aphasia, who might actually show a facilitatory semantic priming effect when the two nouns share meaning.
Method

Participants
The participant group for this study consisted of the same nine individuals with aphasia and eight older controls as those who took part in Experiment 1.

Materials

Norming of new stimulus pictures. Fifty-four new line drawings depicting simple agent–object events were drawn specifically for this study. The drawings depicted combinations of the same verb and noun elements as those that appeared in Experiment 1, except that dog and fox were not used, and snake was replaced with skunk (a noun with a similar lexical frequency that also met our previous age of acquisition criterion). In these drawings, both subject and object nouns were always from the same frequency group, according to the criteria used in Experiment 1. Using the same procedure as that in Experiment 1, these new pictures were then presented to 33 new psychology undergraduate students, aged 18 to 27 years (M = 18.72, SD = 1.65), 12 of whom were male. Again, all were native English speakers. Thirty-eight of these new pictures met the response agreement criterion of 80% (as detailed in Experiment 1) and were selected. These pictures were combined with 26 pictures from Experiment 1 to create a final set of 64 pictures in total.

Composition of final stimulus materials. The 64 final stimulus pictures were composed of: (a) 28 “related” pictures, which depicted a subject and an object noun from the same broad category—either animals (e.g., bear, skunk) or real/imagined persons (e.g., king, ghost); and (b) 36 “unrelated” pictures depicting subject and object nouns from different categories. Noun frequency was balanced so that half of the related pictures depicted two high-frequency nouns, and the other half depicted two low-frequency nouns, and the same was true for the unrelated pictures (see Appendix B for a complete list of the target sentences). Again, the target pictures were combined with 64 additional filler pictures depicting various sentence structures, and none featured any elements displayed in the target sentences (over 90% of those pictures contained either nonspecific inanimate nouns, such as man, boy, woman, or inanimate nouns, such as book, ball). Again, the entire set of pictures was administered three times across three different testing sessions, each time in a different, fixed randomized order. Practice items were the same as those for Experiment 1.

Similar to Experiment 1, a pretest single picture naming task was also constructed, in which each of the 16 noun elements that appeared in the scenes was depicted in isolation. All pictures were drawn from freely accessible picture pools, all of them reaching name agreement of 80% or higher in previous norming studies.

Procedure, scoring, and statistical analysis
The administration, scoring procedures, and statistical analyses were identical to those used in Experiment 1.

Results

Naming pretest
As for Experiment 1, accuracy on the naming pretest was close to ceiling in all participants. Again, no target noun was failed consistently by any of the participants across the three sessions of the pretest naming (see Table 3). As in Experiment 1, there was a trend across all groups and individuals towards shorter latencies on high-frequency words, although the frequency effects were generally flatter. Indeed, at the group level, it reached significance only for the nonfluent group, F(1, 154) = 6.00, p < .05.

Main task

Analysis of response accuracy. Panel (b) of Figure 3 plots noun production accuracy for each individual and group as a function of utterance context. It can be seen that the pattern observed in Experiment 1 was replicated in Experiment 2—that is, nonfluent showed an incremental pattern of noun errors, which was lowest for nouns produced in isolation and highest for nouns produced in object position within the SVO sentences.
Turning now to the results that are directly relevant to our hypotheses, Figure 6 shows the overall percentage of sentence errors for each group and aphasic individual, broken down by relatedness condition. The control group exhibited a small but reliable trend towards more errors on sentences concerning semantically related nouns than unrelated ones, \( \chi^2(1) = 5.67, p < .05 \). The participants with nonfluent aphasia, considered as a group, showed a considerably exaggerated effect in the same direction, \( \chi^2(1) = 4.69, p < .05 \). A combined logistic regression analysis of the data from the nonfluent and control groups failed to reveal a significant group by semantic relatedness interaction \( (p = .52) \). However, due to the low overall error rates for controls, logistic regression may not be sufficiently sensitive to reveal any genuine interactions. Consequently, the aggregated data for each group were reanalysed using the \( Q' \) test, and a significant group by relatedness interaction was obtained, \( Q'(1) = 17.34, p < .0001 \).

In contrast, the fluent group actually exhibited a trend in the opposite direction, producing fewer errors on sentences containing related nouns than on those containing unrelated ones, \( \chi^2(1) = 5.75, p < .05 \). Indeed, a direct comparison between the fluent and nonfluent groups revealed a significant group by relatedness interaction, \( \chi^2(1) = 10.14, p < .01 \).

Figure 6 also shows the percentage of subject and object noun errors as a function of relatedness. It can be seen from the figure that the differential semantic relatedness effects exhibited by the nonfluent and the fluent groups were also evident when noun accuracy was considered on its own.

Finally, in order to check for any potential cumulative semantic interference effects across trials, we compared sentence and noun accuracy for the first and second halves of each session. No significant effects of overall trial position (first vs. second half) were found (sentence accuracy: nonfluent, \( p = .80 \); fluent, \( p = .29 \); control, \( p = .99 \); noun accuracy: nonfluent, \( p = .58 \); fluent, \( p = .88 \); control, \( p = .43 \)), indicating that semantic relatedness effects have not operated across the experimental trials.

**Analysis of response latency.** Latency data were trimmed of outliers in the same way as for Experiment 1.

Figure 7 shows the sentence initiation times for each group and each individual. When collapsed across all individuals, this led to a removal of an average of 2.0% of data for the aphasic groups and 1.0% for the control group. Due to technical problems and other disturbances, an additional 2.2% of data points (across all groups) were lost.
aphasic participant as a function of relatedness (see Appendix C for more detailed descriptive statistics for each key latency measure). When considered as a group, none of the three participant groups showed a main effect of relatedness on sentence initiation times (controls, $p = .42$; nonfluent, $p = .91$; fluenst, $p = .85$). A direct comparison of the data for the nonfluent and the control participant groups failed to reveal any differential effects of relatedness across groups ($p = .87$). A direct comparison of the nonfluent and fluent groups also failed to reveal any such effect ($p = .49$).

At the individual level, only one participant (S.T.R.) showed a significant effect of relatedness: Fluent participant S.T.R.’s sentence initiation times were significantly faster when the sentence contained two semantically related nouns than when it contained two unrelated nouns. There were no effects of relatedness for any of the other aphasic individuals. However, when this was examined at the descriptive level, it stands out that three nonfluent participants—and none of the fluent participants—showed a trend towards slower initiation times for sentences containing meaning related nouns. In fact, the only nonfluents not to show a trend in this direction were B.Y. and D.A., both of whom showed no evidence of advance planning in Experiment 1. However, these differences only exist on the descriptive level and are not backed up by inferential statistics.

Finally, the effect of target picture position was investigated in order to consider potential cumulative interference effects. As for the accuracy analysis, no significant effect of trial position (1st vs. 2nd half) on sentence initiation time was found (nonfluent, $p = .08$; fluenst, $p = .52$; controls, $p = .69$).

**Discussion**

In this experiment, we manipulated the semantic relatedness of the subject and object nouns to be produced in SVO sentences. Our primary hypothesis, that individuals with nonfluent aphasia would be less accurate at producing sentences containing semantically related nouns, was supported. Importantly, the semantic interference effect was specific to individuals with nonfluent aphasia: Our participants with fluent aphasia actually showed a trend in the opposite direction towards semantic facilitation. This finding is in line with
previous studies of multiple single-word naming, which have shown considerably smaller semantic interference effects (e.g., Freedman et al., 2004; Schnur et al., 2006), or even facilitatory effects (see, e.g., Laine & Martin, 1996) in fluent cases. Controls showed a similar, but greatly attenuated effect to those of the nonfluent individuals, which is also consistent with previous studies (e.g., Freedman et al., 2004).

In addition, as was the case in Experiment 1, the nonfluent participants were considerably more accurate at producing nouns in isolation than within a sentence, and, inside the sentence itself, they were more accurate at producing subject nouns in than object nouns. Again, these effects were consistently observed only in individuals with nonfluent aphasia.

Curiously, although semantic relatedness had a robust effect on our nonfluent participants’ overall sentence accuracy, their sentence initiation times were not reliably influenced by this variable. This observation suggests that a large portion of the interfering effects of semantic relatedness may have occurred during production and/or planning of the second, object, noun, rather than during planning of the first, subject, noun. Of course, we would only expect an effect of semantic relatedness on early planning if the speaker plans the object noun substantially in advance. In Experiment 1, only three of the five nonfluent participants showed any evidence of advance planning of this type: J.G. and R.P. were the only ones to show any suggestions of an object frequency effect on their sentence and/or subject initiation times, while J.H.M. showed more marked effects of object frequency on verb–object noun initiation time when the subject noun was of high frequency, indicating a temporal overlap of subject and object noun retrieval. These were the very same individuals who showed a trend towards slower initiation times for sentences containing semantically related nouns in the current experiment.

It should be emphasized that the semantic interference effects we observed in our nonfluent participants involved different noun phrases. Previous studies of individuals with nonfluent aphasia have obtained semantic interference effects only when the related elements occurred within the same noun phrase (e.g., nose and ear; Freedman et al., 2004). Indeed, Martin and colleagues drew directly on such findings to support an incremental phrase-by-phrase model of sentence planning, in which speakers lexically encode only a single phrase at a time (see especially Martin & Freedman, 2001; Martin & He, 2004).}

Two aspects of the current findings are relevant for reconciling these apparently inconsistent findings. The first concerns the directionality of the effects observed here: Evidence from our latency analyses indicates that much of the relatedness effect might reflect perseveratory, rather than anticipatory interference—that is, elements early in the sentence appeared to interfere with those later in the sentence to a greater extent than vice versa. If this is the case, then interference effects might plausibly extend beyond the “window” in which advance planning occurs. Second, it is very possible that the scope of planning varies according to current task demands. In our study, advance planning of the entire sentence was actively encouraged by removing the stimulus picture from view at the time of response initiation. Further, the noun phrases in our target sentences were extremely simple: They consisted only of a single noun plus determiner. These would be just the kinds of conditions that might engender advance planning across more than one phase at a time.

Finally, the results from Experiment 2 provide further evidence against an account of nonfluent aphasic performance in terms of rapidly decaying conceptual representations. Clearly, in order to obtain semantic relatedness effects, there must be substantial activation of both relevant conceptual representations in the first place. Further, most accounts require that the relevant lexical representations have also received substantial activation (Bloem & La Heij, 2003; Bloem, van den Boogaard, & La Heij, 2004; Levelt, Roelofs, & Meyer, 1999; Rahman & Melinger, 2009, 2011; Roelofs, 1992). Put simply, the semantic interference effects themselves provide further evidence that the difficulty in nonfluent
aphasia is not due to abnormally fast decay of the object’s conceptual representation, but rather to effects occurring during linguistic specification of the message.

**Key determinants of the experimental effects**

So far in this paper, we have simply dichotomized our participants into fluent and nonfluent groups based on their broad BDAE diagnosis and compared the two groups. However, it may be interesting to go beyond this broad dichotomization and consider what aspects of language processing are most strongly associated with the key effects observed in Experiments 1 and 2. As can be seen in Table 2, we also tested our participants on a number of other potentially relevant tasks. Three particularly crucial ones are: (a) the QPA rate of speech measure, which is essentially a quantitative measure of overall fluency; (b) the QPA closed class ratio, which indexes the extent to which agrammatic features are present in the participant’s speech; and (c) the reversible sentences score on the Philadelphia Comprehension Battery (Saffran et al., 1988), a measure of the degree of syntactic comprehension impairment. Other aspects of language function that might be relevant include: verbal short-term memory (which we measured using the digit span task), object naming accuracy (measured using the Boston Naming Test, BNT (Goodglass et al., 2001)), and verb retrieval (measured by calculating the difference between object and action naming scores on the Druks and Masterton Object and Action Naming test (Druks & Masterson, 2000)).

Considering first the subject frequency effect reported in Experiment 1, we further explored this effect by creating a qualitative measure of its magnitude in each participant: we simple calculated the difference in the number of incorrect sentences with high- and low-frequency subjects, respectively (a high positive score indicates a strong subject frequency effect). Simple correlational analysis revealed that both speech rate and closed class ratio were significantly correlated with the magnitude of this subject frequency effect (QPA speech rate: \( r = -0.812, n = 9, p < 0.01 \); QPA proportion of closed class items: \( r = -0.617, n = 9, p < 0.05 \)). However, both these measures were also highly correlated with one another, and, indeed, a multiple regression analysis including both as predictors of the subject frequency effect showed neither to be a significant unique predictor of the subject frequency effect, once the contribution of the other had been taken into account. (The \( p \) value associated with the partial correlation coefficient for QPA speech rate was .11, and for QPA closed class ratio it was .94). None of the other measures examined—digit span, naming accuracy, reversible sentence comprehension, and action versus object naming—was significantly correlated with the subject frequency effect measure. Of course, with a sample size of nine, our power to reveal significant associations is low. However, it is perhaps worth noting that not only was the correlation between digit span and the subject frequency effect nonsignificant, but there was not even a trend in the direction that would be expected if verbal short-term memory limitations contributed to the effect (\( r = 0.222, n = 9, p = 0.57, ns \)).

We repeated the same set of analyses using an operationalized measure of the extent of the semantic relatedness (Experiment 2), which we obtained by calculating the difference in accuracy on sentences containing semantically related and unrelated nouns, respectively. This measure is itself powerfully correlated with the subject frequency effect measure from Experiment 1 (\( r = 0.963, n = 9, p < 0.0001 \)). So, not surprisingly, a similar pattern of results was obtained: Both speech rate and closed class ratio were significantly correlated with the relatedness effect measure (QPA speech rate: \( r = 0.763, n = 9, p < 0.05 \); QPA closed class ratio: \( r = 0.766, n = 9, p < 0.05 \)), but of course, as we saw above, they were both high intercorrelated themselves. None of the other measures was significantly correlated with the relatedness measure. Again, there was not even a trend towards a negative relationship between digit span and the size of the relatedness effect (\( r = 0.191, n = 9, p = 0.62, ns \)).
GENERAL DISCUSSION

Most studies of sentence production in nonfluent aphasia have examined accuracy across different types of sentence structures, while minimizing the influence of lexical content. In this study, we have done the opposite: We have systematically examined the influence of lexical content on sentence production while keeping syntactic structure constant. Experiment 1 demonstrated that the frequency of noun elements in the sentence, particularly early elements, has a marked downstream effect on overall sentence production accuracy in nonfluent aphasia. This effect was observed even though the nouns themselves were within every participant’s range of vocabulary, and there were few errors involving these early nouns themselves. Also, the effect was unique to individuals with nonfluent aphasia: It was not observed in our participants with fluent aphasia. Experiment 2 demonstrated that the production of SVO sentences is less accurate in nonfluent aphasia if the subject and the object noun elements are semantically related than when they are unrelated. Again, this effect was unique to the nonfluent participants; the fluent participants actually showed a trend in the opposite direction. These findings provide strong evidence that lexical variables have a particularly powerful effect on sentence production accuracy in nonfluent aphasia.

Not only do our findings indicate that lexical content influences sentence production in nonfluent aphasia, but, conversely, they also show that utterance structure strongly influences lexical content retrieval in these individuals. That is, in nonfluent aphasia, the chances of correctly producing a particular lexical element depend heavily on the utterance context in which it must appear. Both Experiments 1 and 2 found that noun production in isolation (in the naming pretest) was very accurate, but accuracy dropped substantially when the same nouns had to be produced in the context of a sentence, especially when they appeared in object position. This pattern was again unique to the individuals with nonfluent aphasia.

In our primary analyses, we classified participants into two groups based on their fluency diagnosis on the BDAE. Further correlation analyses using quantitative measures of language performance broadly supported these group effects, by demonstrating that the participant’s speech rate, as measured on the QPA, was a significant predictor of both the subject frequency effect in Experiment 1 and the relatedness effect in Experiment 2. The QPA closed class ratio, which provides a quantitative measure of the extent to which the speech shows agrammatic features, was tightly related to speech rate and, again, was a significant predictor of both experimental effects—indeed, it is not possible to tease the effects of these two measures apart. Interestingly, there was no evidence of a relationship between our key experimental measures and verbal short-term memory capacity (as measured by the digit span task). This suggests that verbal short-term memory imitations, at least as measured in the digit span task, contribute little, if at all, to the key effects observed here.

Taken together, the findings from our study suggest that a full understanding of the sentence production difficulties in nonfluent aphasia can only be achieved if we consider the way in which lexical retrieval interacts with structural aspects of sentence planning. We propose an explanation for these findings within a framework that allows for two-way interaction between the process responsible for the generation of structure and content, respectively (e.g., J.K. Bock, 1986; Chang et al., 2006; Mackay, 1987; Stemberger, 1985). Within this type of framework, the sentence planning difficulty in nonfluent aphasia would be characterized as an inability to generate a robust, stable structural representation of the sentence being planned. In this situation, the bidirectional interplay between structure generation and lexical element retrieval processes becomes particularly crucial. Specifically, the construction of the sentence plan relies disproportionately heavily on activation input from the constituent lexical elements. Sentence production success in nonfluent aphasia therefore depends particularly heavily on the ability to successfully
and efficiently retrieve content words. We call this the *content drives structure* (COST) hypothesis. The COST hypothesis offers a plausible explanation as to why sentence production accuracy in nonfluent aphasia is more susceptible to manipulations of lexical content than it is in fluent aphasia. Individuals with fluent aphasia can rely on robust structure building processes to guide sentence production, so a lexical retrieval failure has only a localized effect and does not impact on the remainder of the sentence. Furthermore, the COST hypothesis can account for why the production of single words is relatively well preserved in individuals with nonfluent aphasia, since single-word production would not require any coordination of syntactic and lexical elements, and lexical competition is reduced to a minimum.

The semantic short-term memory model of Martin and colleagues offers a useful framework for elaboration of this hypothesis (e.g., Freedman et al., 2004; Martin et al., 1999, 2004). These researchers suggest that in nonfluent aphasia, there may be weak or inefficient binding of lexical elements to their structural positions within the phrase. This leads to increased production failures, particularly on phrases containing multiple words with similar selection constraints, where structure binding is crucial to resolve order. Importantly, the likelihood of such failures depends on the lexical elements themselves. For example, a strongly activated lexical element in turn strongly activates its corresponding structural representation in semantic short-term memory, thereby mitigating the effects of the impairment. Moreover, if elements are only weakly bound to their positions in semantic short-term memory, then they will be freer to compete with one another for selection to a given position (provided, of course, they have similar selection constraints, so are both plausible candidates for the same position). This competition might be particularly intense if the words are also semantically related, as they were in Experiment 2. However, Martin and colleagues’ framework deals only with planning within a single phrase. In our study, we were able to demonstrate semantic interference effects that span more than a single phrase, so to account for our results, we need a framework that considers sentence planning more broadly.

One model of aphasic sentence production that is compatible with our proposal is the *division of labour* (DoL) model (Gordon & Dell, 2003; see also Dell, Oppenheim, & Kittredge, 2008). DoL is a connectionist model in which the activation of different lexical elements, including determiners, nouns, and verbs, depends on the combined input from two types of nodes: *conceptual–semantic* and *syntactic–sequential* nodes. The semantic–conceptual input nodes activate the lexical element that represents the correct meaning of a sentence element, whereas the syntactic–sequential input nodes ensure that the right word in the utterance is produced at the right time, activating lexical elements that are congruent with the structure that needs to be produced. Gordon and Dell (2003) used a simple learning algorithm to train the connection strength between these two sets of input nodes and various types of lexical items. After training, nouns and semantically rich verbs (e.g., *fly*) were primarily activated by semantic–conceptual input nodes, whereas determiners and semantically impoverished verbs (e.g., *have*) were primarily activated by syntactic–structural input nodes.

Further, by lesioning the syntactic-sequential nodes in the DoL model, Gordon and Dell (2003) were able to simulate the consequences of selective damage to the structure building process. The resultant model, where lexical elements were primarily reliant on top-down activation from semantic–conceptual input nodes, produced an agrammatic pattern of speech, characterized by the selective omission of determiners and semantically impoverished verbs. Conversely, the pattern of fluent but lexically empty speech was simulated by reducing the efficiency of conceptual–semantic activation of sentence elements. Importantly, when the syntactic–sequential nodes were

---

5 Conversely, the lesioning of semantic input led to a fluent anomic speech pattern, represented by a good production of determiners and semantically impoverished verbs, but a lack of nouns and semantically rich verbs (Gordon & Dell, 2003).
lesioned, enhanced lexical competition also occurred, because all the planned sentence elements became concurrently activated, and no structural representation was in place to enhance activation of the right word for the right position. This is exactly what we observed in Experiment 2. The DoL model only simulated subject–verb sentences, but presumably this effect would be considerably intensified for sentence elements that share the same word class (e.g., subject and object nouns), and particularly so if they also share semantic features.

The dual-path model of Chang and colleagues (Chang, 2002; Chang et al., 2006), illustrated in Figure 8, incorporates this division of labour idea within a broader connectionist framework, which can deal with more complex sentences. This model is complex, but its most essential aspects can be summarized quite simply. In the model, the sequencing system represents learned information about the sequencing rules that apply to particular word classes, which are acquired through extended exposure to a range of sentences. The meaning system represents conceptual knowledge about entities and actions. When a speaker wishes to express a message, temporary associations are created between the representations of the key concepts and their roles in the intended message (e.g., cat–agent, chase–action, mouse–patient). To produce a given sentence, the sequencing and the meaning systems must work together. The model produces sentences one word at a time, based on the most activated role unit, which in turn depends upon two factors: (a) how strongly activated its associated concept is at that time; and (b) which grammatical roles are permissible for production given the previous word output. Consequently, both meaning and structural information combine to influence the selection of the next upcoming word.

Many of the systems incorporated into the dual-path model develop with repeated experience: For example, the concepts represented in the meaning system and their associated lexical labels are learned through practice, and so too are the sequencing rules that control which word classes can be legally produced at each point in a sentence. However, the associations between concepts and their roles must be established anew for each conceptual message, since any one concept can potentially be associated with more than one thematic role, depending upon the message. This process of concept–role assignment might therefore be particularly susceptible to impairment after anterior brain damage. We might therefore postulate that the sentence production impairment in nonfluent aphasia consists of an inability to effectively associate concepts to their roles and/or maintain these associations long enough to effectively drive the sentence planning process (an idea not unlike that proposed in some earlier accounts: see, e.g., Saffran, Schwartz, & Marin, 1980). Since role units are the locus at which information from the meaning and sequencing systems converges, these two systems will then become effectively decoupled. Instead, the speaker may fall back on the meaning system to drive word selection, and competition for selection of a particular word will be won by whichever concept/word is most strongly activated at the time of word selection. In other words, utterance production becomes much more heavily driven by current activation patterns in the conceptual/lexical network and less constrained by relational and structural information.

In addition, without the activation boost provided by the sequencing system, the process of word
selection may also be considerably slower and more susceptible to factors that modulate levels of activation in conceptual and lexical representations—such as conceptual salience and frequency of occurrence. An interesting direct prediction of this account is that any variable that reduces the activation differential between coactivated concepts and their associated labels—such as semantic similarity—will have a particularly detrimental effect on production latency and/or accuracy in these cases. This is exactly what we found in Experiment 2 in our participants with nonfluent aphasia.

As discussed earlier, a complete account of nonfluent aphasic sentence production also needs to explain the observed variability in performance between and within individuals. In the COST framework, the phenomenon of between-participant variability can be explained in terms of variability of the degree of impairment to the structure generation system and/or how this interacts with other constraints imposed by the aphasic impairment (for example, slowed motor–articulatory planning). The framework also offers a means for explaining within-participant variability, since success on any given utterance will depend upon the degree to which conceptual–lexical content can be used to help support sentence planning and may be further modulated by the degree of contextual support available (for example, a picture scene) and the specific time demands of the current conversational context.

Implications for theories of normal sentence production

Models of sentence production differ in the extent to which they assume an interaction between structure and content processes. In some models, structure generation and content retrieval processes occur completely independently (e.g., Garrett, 1975). In other models, there is an interactive flow of information between structure generation and lexical selection processes during sentence planning (e.g., Bock & Eberhard, 1993; MacKay, 1987; Stemberger, 1985). In a third class of models, there are distinct syntactic and conceptual components, which independently contribute to the activation of candidate lexical elements (e.g., Chang, 2002; Chang et al., 2006; Dell et al., 2008; Gordon & Dell, 2003). The findings from this study provide further evidence to suggest a powerful interplay between content and structure building processes during speech production. It therefore favours models of the latter two classes.

Further, research into sentence planning has so far focused heavily on grammatical/structural aspects of sentence planning, but the issues raised in the present study suggest that greater attention may need to be paid to the message level and how it relates to the structure actually produced. That is, rather than comparing different types of grammatical structures (e.g., active versus passive), while keeping the message constant, we may need to pay more consideration to the match between message and structure. For example, in most situations, where the agent of the key action is the highly salient (e.g., The warrior threw down his spear), production of an active structure may require fewer resources than that of a passive structure. However, in those cases where the patient of the action is most salient (e.g., A puppy was trapped under the falling debris), a passive structure may require fewer resources. Disorders that selectively impair different aspects of the speech planning process, of the kind studied here, offer a promising tool for further exploring the interaction between message level and sequencing/structure building processes.

Another important theoretical issue concerns the relationship between the structural planning processes outlined here and language control more generally. The two appear to be very closely related. Many individuals with nonfluent aphasia exhibit abnormalities even when producing single words. For example, picture naming accuracy often declines dramatically if the stimulus pictures are repeatedly sampled from a small set of semantically related items (Hsiao, Schwartz, Schnur, & Dell, 2009; Schnur et al., 2006; Scott & Wilshire, 2010; Wilshire & McCarthy, 2002). In fact, at least four of our current nonfluent
participants (J.H.M., B.Y., J.G., and R.P.) exhibit these very same effects (see, e.g., Scott & Wilshire, 2010; Wilshire, Bareham, & Scott, 2009). Indeed, in virtually all cases so far reported, there seems to be a powerful association between single-word interference phenomena and the degree of impairment to sentence-level production (see, e.g., Biegler et al., 2008; Schnur et al., 2006; Wilshire & McCarthy, 2002), and both types of deficits appear to have common lesion correlates—most notably, left inferior frontal gyrus (sentence production: Borovsky, Saygin, Bates, & Dronkers, 2007; Kling, 2007; single-word production: Schnur et al., 2009).

To account for the interference effects observed in single-word production tasks, some researchers have proposed the existence of an “activation biasing” mechanism, most likely located in the left inferior frontal gyrus, which normally operates to resolve competition in favour of the representation(s) most consistent with current goals, but which is impaired in these cases (e.g., Dell et al., 2008; Novick, Trueswell, & Thompson–Schill, 2010; Novick, Kan, Trueswell, & Thompson–Schill, 2009; Robinson, Shallice, Bozzali, & Cipolotti, 2010; Schnur et al., 2009; Scott & Wilshire, 2010; Wilshire & McCarthy, 2002). Noting the sentence production difficulties that usually accompany this profile, some researchers have further suggested that sentence production may present a special case were levels of lexical competition are unusually high, because it essentially involves the activation and retrieval of several different word targets concurrently. Others have suggested that the word sequencing demands imposed during sentence production may tax the hypothesized activation biasing mechanism particularly heavily (Thothathiri, Schwartz, & Thompson–Schill, 2010).

Drawing on the framework of Chang et al. (2006), we tentatively suggest that both the single-word and sentence-level phenomena observed in these cases arise from damage to an anteriorly located mechanism that operates to favour the activation of goal-consistent representations over others. In sentence production, the goal consists of a thematic representation of the sentence elements and their interrelationships, which operates to bias selection at each critical juncture in favour of the element that best meets the current thematic and grammatical requirements. This is not to say that such individuals are unable to formulate thematic relationships, but simply that they are less effective at utilizing thematic communication goals to drive language production. This proposal offers a way of reconciling the involvement of the left inferior frontal gyrus in both lexical competition resolution and in sentence planning and production more generally.

Limitations and suggestions for future research

One important limitation to our study was that syntactic structure was kept constant (simple SVO sentences), and only lexical content was manipulated. Therefore, we cannot say whether the lexical effects observed here would extend to the production of other types of syntactic structures, nor how structural variables might interact with lexical ones. This, in fact, is an avenue for future research. However, designing such studies will be challenging. Sentence structure and content are often confounded (the more complex the sentence structure, the more lexical content it contains), so the researcher would be limited in the types of structures he or she could compare. Another, even bigger problem is that structure effects in and of themselves need not necessarily indicate a structure building deficit: Caplan and Hanna (1998) were able to show that the order of difficulty of various syntactic structures (including intransitive, active transitive, passive transitive, active dative, and passive dative) was very similar in Broca’s and Wernicke’s aphasia. Finally, it may be difficult to elicit some kinds of sentence structures, especially if one wants to avoid the use of linguistic and other types of prompts.

Implications for the treatment of sentence production difficulties in nonfluent aphasia

Understanding the interplay between structural planning and content retrieval in nonfluent
aphasia is also important from a more clinical point of view. For example, Linebarger and colleagues’ treatment device, the SentenceShaper®, specifically supports the retrieval and/or maintenance of key lexical elements during sentence planning (Linebarger et al., 2007; Linebarger et al., 2000). This support has been found to result in an increase in well-formed sentences in many cases with nonfluent aphasia—even though no syntactic assistance is provided. These findings are in line with the COST framework, which predicts that activation from lexical elements will support a more elaborated construction of the syntactic frame. Taken together, our findings suggest that a greater attention to lexical support in treatment of nonfluent aphasia can result in significant gains.

REFERENCES


Hulleman, J., & Humphreys, G. W. (2007). Maximizing the power of comparing single cases against a control sample: An argument, a program for making comparisons, and a worked example from the pyramids and palm trees test. Cognitive Neuropsychology, 24, 279–291. doi:10.1080/02643290601181890


Wilshire, C. E., Bareham, C., & Scott, R. (2009, October). Lexical competition effects in nonfluent aphasia: Converging evidence from three different tasks. Poster presented at the annual meeting of the Academy of Aphasia, Boston, USA.


APPENDIX A

Details of imaging methods and image preparation

All participants underwent a magnetic resonance imaging (MRI) scan specifically for this study, within one year after being tested in both experiments. Whole brain T1-weighted as well as T2-weighted structural scans with inversion recovery (fluid-attenuated inversion recovery, FLAIR) were collected at 1.5 tesla (T1 3D fast field echo, FFE: time to repetition, TR = 25 ms; echo time, TE = 4.6 ms; field of view, FOV = 252 mm × 238 mm; slice thickness = 1 mm; sagittal 3D FLAIR: TR = 4800 ms; TE = 329 ms; inversion time, TI = 1660 ms; FOV = 252 mm × 250 mm; slice thickness = 1 mm).

The lesions were manually drawn onto the T1-weighted structural image using MRIcron (Rorden, 2007), while consulting the T2-weighted FLAIR image as additional guidance. Subsequently, the scans and the lesions were normalized using SPM8 (Ashburner et al., 2008; see also http://www.fsl.ion.ucl.ac.uk/spm/), implemented in MatLab (The Mathworks Inc, 2008). The lesions were then overlaid onto a template based on healthy elderly individuals (with a mean age of 65 years, Rorden, Bonilha, Fridriksson, Bender, & Karnath, 2012) to roughly account for the age group of our participants.

APPENDIX B

Stimulus material for Experiments 1 and 2

<table>
<thead>
<tr>
<th>Noun</th>
<th>Frequency rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>High frequency</td>
<td>(≥ 30 occ. per million)</td>
</tr>
<tr>
<td>Bear&lt;sup&gt;ab&lt;/sup&gt;</td>
<td>57.41</td>
</tr>
<tr>
<td>Cat&lt;sup&gt;ab&lt;/sup&gt;</td>
<td>66.33</td>
</tr>
<tr>
<td>Dog&lt;sup&gt;c&lt;/sup&gt;</td>
<td>192.84</td>
</tr>
<tr>
<td>Ghost&lt;sup&gt;ab&lt;/sup&gt;</td>
<td>36.59</td>
</tr>
<tr>
<td>Horse&lt;sup&gt;ab&lt;/sup&gt;</td>
<td>92.88</td>
</tr>
<tr>
<td>King&lt;sup&gt;d&lt;/sup&gt;</td>
<td>129.52</td>
</tr>
<tr>
<td>Nurse&lt;sup&gt;c&lt;/sup&gt;</td>
<td>44.98</td>
</tr>
<tr>
<td>Pig&lt;sup&gt;ab&lt;/sup&gt;</td>
<td>39.14</td>
</tr>
<tr>
<td>Queen&lt;sup&gt;ab&lt;/sup&gt;</td>
<td>54.69</td>
</tr>
<tr>
<td>Low frequency</td>
<td>(&lt; 30 occ. per million)</td>
</tr>
<tr>
<td>Clown&lt;sup&gt;ab&lt;/sup&gt;</td>
<td>15.82</td>
</tr>
<tr>
<td>Cow&lt;sup&gt;ab&lt;/sup&gt;</td>
<td>25.51</td>
</tr>
<tr>
<td>Fox&lt;sup&gt;c&lt;/sup&gt;</td>
<td>21.61</td>
</tr>
<tr>
<td>Goat&lt;sup&gt;ab&lt;/sup&gt;</td>
<td>10.53</td>
</tr>
<tr>
<td>Maid&lt;sup&gt;ab&lt;/sup&gt;</td>
<td>22.82</td>
</tr>
<tr>
<td>Nun&lt;sup&gt;ab&lt;/sup&gt;</td>
<td>6.96</td>
</tr>
<tr>
<td>Sheep&lt;sup&gt;ab&lt;/sup&gt;</td>
<td>22.35</td>
</tr>
<tr>
<td>Skunk&lt;sup&gt;b&lt;/sup&gt;</td>
<td>3.25</td>
</tr>
<tr>
<td>Snake&lt;sup&gt;a&lt;/sup&gt;</td>
<td>13.43</td>
</tr>
<tr>
<td>Witch&lt;sup&gt;ab&lt;/sup&gt;</td>
<td>27.65</td>
</tr>
</tbody>
</table>

Note: occ = occurrence; frequency ratings according to the Subtitle Analysis Project (Brysbaert & New, 2009).

<sup>a</sup>Nouns used in experiment 1. <sup>b</sup>Nouns used in Experiments 1 and 2. <sup>c</sup>Nouns used in Experiment 2.
**Table B2. Experiment 1: Target pictures with corresponding experimental condition**

<table>
<thead>
<tr>
<th>High-frequency subject/ high-frequency object</th>
<th>High-frequency subject/ low-frequency object</th>
<th>Low-frequency subject/ high-frequency object</th>
<th>Low-frequency subject/ low-frequency object</th>
</tr>
</thead>
<tbody>
<tr>
<td>The dog is licking the cat</td>
<td>The cat is chasing the sheep</td>
<td>The sheep is shooting the pig</td>
<td>The fox is shooting the cow</td>
</tr>
<tr>
<td>The pig is lifting the cat</td>
<td>The pig is shooting the sheep</td>
<td>The fox is chasing the dog</td>
<td>The snake is biting the fox</td>
</tr>
<tr>
<td>The pig is biting the bear</td>
<td>The nurse is kissing the witch</td>
<td>The sheep is chasing the cat</td>
<td>The fox is chasing the goat</td>
</tr>
<tr>
<td>The ghost is shooting the king</td>
<td>The horse is licking the fox</td>
<td>The fox is kicking the bear</td>
<td>The cow is shooting the fox</td>
</tr>
<tr>
<td>The queen is pushing the nurse</td>
<td>The bear is kicking the fox</td>
<td>The goat is kicking the horse</td>
<td>The goat is licking the cow</td>
</tr>
<tr>
<td>The ghost is washing the queen</td>
<td>The nurse is shooting the clown</td>
<td>The fox is kicking the pig</td>
<td>The fox is biting the snake</td>
</tr>
<tr>
<td>The cat is licking the dog</td>
<td>The king is pushing the maid</td>
<td>The snake is kissing the bear</td>
<td>The witch is kissing the nun</td>
</tr>
<tr>
<td>The king is shooting the ghost</td>
<td>The dog is chasing the fox</td>
<td>The witch is washing the nurse</td>
<td>The goat is chasing the fox</td>
</tr>
<tr>
<td>The horse is chasing the bear</td>
<td>The bear is kissing the snake</td>
<td>The maid is hitting the ghost</td>
<td>The sheep is kicking the cow</td>
</tr>
<tr>
<td>The cat is lifting the pig</td>
<td>The ghost is hitting the maid</td>
<td>The witch is kissing the nurse</td>
<td>The nun is kissing the witch</td>
</tr>
<tr>
<td>The bear is chasing the horse</td>
<td>The pig is licking the snake</td>
<td>The maid is pushing the king</td>
<td>The cow is kicking the sheep</td>
</tr>
<tr>
<td>The horse is kicking the pig</td>
<td>The nurse is washing the witch</td>
<td>The clown is shooting the knight</td>
<td>The maid is lifting the nun</td>
</tr>
<tr>
<td>The queen is washing the ghost</td>
<td>The ghost is pushing the clown</td>
<td>The sheep is brushing the goat</td>
<td>The snake is brushing the goat</td>
</tr>
<tr>
<td>The nurse is pushing the queen</td>
<td>The bear is lifting the goat</td>
<td>The goat is lifting the bear</td>
<td>The nun is lifting the maid</td>
</tr>
<tr>
<td>The pig is kicking the horse</td>
<td>The dog is biting the goat</td>
<td>The fox is licking the horse</td>
<td>The goat is brushing the sheep</td>
</tr>
<tr>
<td>The bear is biting the pig</td>
<td>The horse is kicking the cow</td>
<td>The goat is biting the dog</td>
<td>The maid is washing the maid</td>
</tr>
</tbody>
</table>

**Table B3. Experiment 2: Target pictures with corresponding experimental condition**

<table>
<thead>
<tr>
<th>Related high frequency</th>
<th>Related low frequency</th>
<th>Unrelated high frequency</th>
<th>Unrelated low frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>The pig is lifting the cat</td>
<td>The goat is licking the cow</td>
<td>The ghost is lifting the horse</td>
<td>The sheep is kicking the witch</td>
</tr>
<tr>
<td>The pig is biting the bear</td>
<td>The witch is kissing the nun</td>
<td>The pig is kicking the nurse</td>
<td>The sheep is lifting the cow</td>
</tr>
<tr>
<td>The ghost is shooting the king</td>
<td>The sheep is kicking the cow</td>
<td>The king is kicking the cat</td>
<td>The goat is kicking the clown</td>
</tr>
<tr>
<td>The queen is pushing the nurse</td>
<td>The sheep is kicking the cow</td>
<td>The cat is kicking the king</td>
<td>The sheep is lifting the cow</td>
</tr>
<tr>
<td>The ghost is washing the queen</td>
<td>The sheep is kicking the cow</td>
<td>The cat is tickling the ghost</td>
<td>The sheep is kicking the cow</td>
</tr>
<tr>
<td>The king is shooting the ghost</td>
<td>The sheep is kicking the cow</td>
<td>The pig is shooting the queen</td>
<td>The sheep is kicking the cow</td>
</tr>
<tr>
<td>The horse is chasing the bear</td>
<td>The sheep is kicking the cow</td>
<td>The cat is washing the queen</td>
<td>The sheep is kicking the cow</td>
</tr>
<tr>
<td>The cat is lifting the pig</td>
<td>The sheep is kicking the cow</td>
<td>The ghost is chasing the pig</td>
<td>The sheep is kicking the cow</td>
</tr>
<tr>
<td>The bear is chasing the horse</td>
<td>The sheep is kicking the cow</td>
<td>The nurse is kicking the pig</td>
<td>The sheep is kicking the cow</td>
</tr>
<tr>
<td>The horse is kicking the pig</td>
<td>The sheep is kicking the cow</td>
<td>The nurse is kicking the pig</td>
<td>The sheep is kicking the cow</td>
</tr>
<tr>
<td>The queen is washing the ghost</td>
<td>The sheep is kicking the cow</td>
<td>The nurse is kicking the pig</td>
<td>The sheep is kicking the cow</td>
</tr>
<tr>
<td>The nurse is pushing the queen</td>
<td>The sheep is kicking the cow</td>
<td>The nurse is kicking the pig</td>
<td>The sheep is kicking the cow</td>
</tr>
<tr>
<td>The pig is kicking the horse</td>
<td>The sheep is kicking the cow</td>
<td>The nurse is kicking the pig</td>
<td>The sheep is kicking the cow</td>
</tr>
<tr>
<td>The bear is biting the pig</td>
<td>The sheep is kicking the cow</td>
<td>The nurse is kicking the pig</td>
<td>The sheep is kicking the cow</td>
</tr>
</tbody>
</table>
APPENDIX C

Supplementary data for Experiments 1 and 2

Table C1.  Experiment 1: Percentage of errors for participants with aphasia and the three participant groups for the different frequency conditions

<table>
<thead>
<tr>
<th>Participant</th>
<th>Overall sentence errors</th>
<th>Subject errors</th>
<th>Object errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonfluent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B.Y.</td>
<td>15.6</td>
<td>26.5</td>
<td>25.5</td>
</tr>
<tr>
<td>D.A.</td>
<td>11.5</td>
<td>18.6</td>
<td>13.5</td>
</tr>
<tr>
<td>J.G.</td>
<td>32.3</td>
<td>49.0</td>
<td>36.5</td>
</tr>
<tr>
<td>J.H.M.</td>
<td>42.7</td>
<td>57.8</td>
<td>53.1</td>
</tr>
<tr>
<td>R.P.</td>
<td>39.6</td>
<td>52.9</td>
<td>39.2</td>
</tr>
<tr>
<td>Group mean</td>
<td>26.1</td>
<td>42.6</td>
<td>32.3</td>
</tr>
<tr>
<td>Fluent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N.P.</td>
<td>22.9</td>
<td>20.6</td>
<td>19.8</td>
</tr>
<tr>
<td>S.T.R.</td>
<td>30.2</td>
<td>20.6</td>
<td>18.8</td>
</tr>
<tr>
<td>S.W.</td>
<td>35.4</td>
<td>27.5</td>
<td>32.3</td>
</tr>
<tr>
<td>W.L.</td>
<td>29.2</td>
<td>17.7</td>
<td>28.1</td>
</tr>
<tr>
<td>Group mean</td>
<td>27.1</td>
<td>22.4</td>
<td>23.8</td>
</tr>
<tr>
<td>Controls</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group mean</td>
<td>5.1</td>
<td>9.2</td>
<td>5.1</td>
</tr>
</tbody>
</table>

Note: freq. = frequency.

Table C2.  Experiment 2: Geometric means of latencies for participants with aphasia in the nonfluent, fluent, and control groups

<table>
<thead>
<tr>
<th>Participants</th>
<th>Speech onset latency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Relatedness</td>
</tr>
<tr>
<td></td>
<td>Related</td>
</tr>
<tr>
<td>Nonfluent</td>
<td></td>
</tr>
<tr>
<td>B.Y.</td>
<td>2970</td>
</tr>
<tr>
<td>D.A.</td>
<td>3483</td>
</tr>
<tr>
<td>J.G.</td>
<td>2999</td>
</tr>
<tr>
<td>J.H.M.</td>
<td>2230</td>
</tr>
<tr>
<td>R.P.</td>
<td>7150</td>
</tr>
<tr>
<td>Group mean</td>
<td>3482</td>
</tr>
<tr>
<td>Fluent</td>
<td></td>
</tr>
<tr>
<td>N.P.</td>
<td>2369</td>
</tr>
<tr>
<td>S.T.R.</td>
<td>2311</td>
</tr>
<tr>
<td>S.W.</td>
<td>2867</td>
</tr>
<tr>
<td>W.L.</td>
<td>2356</td>
</tr>
<tr>
<td>Group mean</td>
<td>2466</td>
</tr>
<tr>
<td>Controls</td>
<td></td>
</tr>
<tr>
<td>Group mean</td>
<td>1167</td>
</tr>
</tbody>
</table>