Going off the rails: Impaired coherence in the speech of patients with semantic control deficits

Paul HOFFMAN*[1]

Lucy COGDELL-BROOKE[2]

Hannah E THOMPSON[2]

[1] School of Philosophy, Psychology & Language Sciences, University of Edinburgh, UK

[2] School of Psychology, University of Surrey, UK

* Correspondence to:
Dr. Paul Hoffman
Centre for Cognitive Ageing and Cognitive Epidemiology,
Department of Psychology, University of Edinburgh,
7 George Square, Edinburgh, EH8 9JZ, UK
Tel: +44 (0) 131 650 4654
Email: p.hoffman@ed.ac.uk

Abstract

The ability to speak coherently, maintaining focus on the topic at hand, is critical for effective communication and is commonly impaired following brain damage. Recent data suggests that executive processes that regulate access to semantic knowledge (i.e., semantic control) are critical for maintaining coherence during speech. To test this hypothesis, we assessed speech coherence in a case-series of fluent stroke aphasic patients with deficits in semantic control. Patients were asked to speak about a series of topics and their responses were analysed using computational linguistic methods to derive measures of their global coherence (the degree to which they spoke about the topic given) and local coherence (the degree to which they maintained a topic from one moment to the next). Compared with age-matched controls, patients showed severe impairments to global coherence and milder impairments to local coherence, suggesting that semantic control deficits give rise to being "led up the garden path", i.e., one sentence automatically cueing another, with the topic becoming increasingly less relevant to the original question. Global coherence was strongly correlated with the patients' performance on tests of semantic control, with poorer ability to maintain top-down global coherence being associated with greater semantic control deficits. Other aspects of speech production were also impaired but were not correlated with semantic control deficits. This is the first study to investigate the impact of semantic control impairments within a naturalistic setting, and indicates patients with these impairments are likely to find maintaining focus in everyday conversation difficult.

Keywords: speech; discourse; aphasia; semantic control; coherence.

Introduction

Engaging in conversation is a complex cognitive activity, in which a speaker must identify the topic under discussion, generate a series of statements relevant to this subject and monitor their speech as the discourse unfolds to ensure that they remain on-topic. Discourse that successfully navigates these challenges is said to be coherent: it consists of a series of well-connected statements all related to a shared topic, making it easy to comprehend (Glosser and Deser, 1992; Foltz, 2007). Researchers often distinguish between global coherence (GC), the degree to which each utterance relates to the overall topic under discussion, and local coherence (local coherence), the degree to which adjoining utterances related meaningfully to one another (Kintsch and Vandijk, 1978; Glosser and Deser, 1992). Speech that frequently goes ”off the rails”, deviating greatly from the original topic, is said to be low in global coherence, while speech that shifts abruptly between subjects is low in local coherence.

The ability to speak coherently is frequently impaired following brain damage, even in patients who do not display classical language production deficits (Ellis *et al.*, 2016). Deficits to global coherence have been reported following stroke, most commonly in the context of left hemisphere damage (Christiansen, 1995; Rogalski *et al.*, 2010; Harris Wright and Capilouto, 2012; Barker *et al.*, 2017; Karaduman *et al.*, 2017). Poor coherence is also a characteristic of speech in Alzheimer’s disease (Glosser and Deser, 1991), traumatic brain injury (TBI) (Glosser and Deser, 1991; Marini *et al.*, 2014) and the frontal variant of frontotemporal dementia (Ash *et al.*, 2006). In addition, healthy ageing is associated with declines in global and local coherence. Older adults are more likely to produce tangential, off-topic utterances in conversation (Glosser and Deser, 1992; Arbuckle and Gold, 1993) and to provide irrelevant information when telling a story (Juncos-Rabadan *et al.*, 2005; Marini *et al.*, 2005) or describing an object (Long *et al.*, 2018). These impairments in coherence hinder effective communication and can be associated with reduced well-being and decreased satisfaction with social interactions (Gold *et al.*, 1988; Arbuckle and Gold, 1993; Pushkar *et al.*, 2000).

What underlying cognitive impairments give rise to impaired coherence in speech? Research has focused mainly on the role of domain-general cognitive and executive systems, rather than language-specific processes, in the planning and regulation of speech content

(Alexander, 2006; Kintz *et al.*, 2016; Barker *et al.*, 2017). In TBI patients, for example, poor coherence is observed in the context of minimal impairments to micro-linguistic aspects of speech, such as phonology, lexical and syntactic processing (Coelho, 2002; Marini *et al.*, 2014). In these patients, however, correlations have been reported between performance on the Wisconsin card-sorting test and errors of local and global coherence (Marini *et al.*, 2014) and more general impairments of narrative organisation (Coelho, 2002), suggesting that an executive deficit may underpin their difficulties in regulating their speech. Similarly, coherence deficits in frontotemporal dementia are present in patients with executive impairment but not in those with the non-fluent or semantic variants of the disorder (Ash *et al.*, 2006). In healthy individuals, age-related coherence declines have also been attributed to impaired executive function (North *et al.*, 1986; Gold *et al.*, 1988; Kintz *et al.*, 2016).

These results suggest that domain-general executive control processes are involved in the monitoring and selection of topics during speech. One view holds that poor coherence, at the global level in particular, results from a reduced ability to inhibit irrelevant information, such that people are less able to prevent irrelevant or off-topic ideas from becoming automatically activated and intruding into their discourse (Arbuckle and Gold, 1993; Marini and Andreetta, 2016). Recent work has developed this hypothesis by investigating the specific role of semantic control, defined as the top-down processes that control retrieval and selection of semantic knowledge (Jefferies, 2013; Lambon Ralph *et al.*, 2017; Hoffman *et al.*, 2018b). The role of semantic control has rarely been studied in the context of real world conversation. In recent work, however, Hoffman et al. (2018a) hypothesised that semantic control processes are critically involved in maintaining coherence because semantic knowledge is central to all propositional speech. This is obviously true at the lexical level, since the selection of words for production is determined by their meanings. But at the broader "message" level, the content of our speech is also determined by our general semantic knowledge about the topic under discussion. For example, if someone asks you how they would catch a train somewhere, you must access stored semantic knowledge about the typical characteristics of railway stations, the methods for buying tickets and so on. To ensure coherent discourse, this knowledge retrieval process must be regulated such that topic-relevant knowledge is used to drive speech production and any irrelevant associations that come to mind are avoided (e.g., retrieving information about trains might bring to mind other forms of transport not relevant to the discussion). For these reasons, we predicted that the

coherence of an individual's discourse is influenced by the efficiency of their semantic control processes, in addition to more general executive function.

Hoffman et al. (2018a) tested these predictors in a group of 60 healthy young and older adults. In line with previous studies, we found that the coherence of speakers was predicted by their performance on a test of domain-general executive function (the Trails test). Importantly, however, we also found that semantic control abilities were a significant independent predictor of coherence. Participants who produced more coherent speech were more skilled at selecting task-relevant semantic information in a semantic judgement task.

The results of Hoffman et al. (2018a) suggest that semantic control processes, particularly those governing the selection of task-relevant semantic knowledge, contribute to the maintenance of coherence during natural speech production. Deficits of semantic control may therefore be a key factor in explaining the coherence deficits frequently observed following brain damage. Here, we tested this hypothesis by investigating extended speech in a case-series of stroke aphasic patients who presented with clear deficits of semantic control. We collected samples of patients speaking about different topics and determined their coherence using previously-validated methods from computational linguistics (Hoffman *et al.*, 2018a). We predicted that patients would display impaired coherence, relative to healthy controls and that the severity of the semantic control impairment would predict the degree to which coherence was disrupted in individual patients. By assessing naturalistic speech production in patients with semantic control deficits for the first time, we therefore aimed to establish how these deficits impact on patients' everyday conversational ability.

Method

Participants: Seven patients with deficits of semantic control were recruited from Yorkshire, Surrey, Sussex and Manchester community stroke groups. We refer to these as stroke aphasia (SA) patients. All suffered a left-hemisphere stroke at least four years previously. In line with the inclusion criteria adopted by Jefferies and Lambon Ralph (2006), patients were selected to show difficulties accessing semantic knowledge in both verbal and non-verbal tasks. Two other SA patients attempted to take part in the study but were unable to complete the speech elicitation task. Their speech was considerably less fluent than the patients who successfully

completed the task (on the Cookie Theft description, they produced 9 and 12 words per minute, cf. a mean of 61 words per minute in the patients who completed the task). We also included one right-hemisphere stroke patient as a control case. This individual suffered a right-hemisphere stroke but was not aphasic and exhibited no evidence of semantic deficits. Age, years of education and time since stroke for all patients are reported in Table 1. All patients provided informed written consent to participate in this study. Ethical approval was obtained from the University of Surrey Ethics Committee (UEC/2016/090/FHMS). It also received ethical approval from the NRES Committee Yorkshire and The Humber (ref: 12/YH/0323).

The characteristics of speech in the patients were compared with control data from healthy older adults, taken from Hoffman et al. (2018a). The control sample comprised all participants reported by Hoffman et al. (2018a) who were aged between 60 and 75 (N=12). The control group were matched to the SA patients for age (SA mean = 63.0 years; control mean = 67.4 years; $t(17) = 1.44$, $p = 0.17$) and years of education (SA mean = 13.1 years; control mean = 13.5 years; $t(17) = 0.31$, $p = 0.76$). MRI scans were available for five of the seven SA patients (see Table 1 for description of lesion locations and Supplementary Figure 1 for scan images). Patients presented with damage to a range of left-hemisphere sites. This is in line with previous studies showing that similar semantic control deficits can arise from either left prefrontal or posterior temporal and parietal damage (Jefferies and Lambon Ralph, 2006; Noonan *et al.*, 2010; Thompson *et al.*, 2018).

Background neuropsychological tests: The SA patients were examined on a range of general neuropsychological assessments, including forwards and backwards digit span (Wechsler, 1987), dot counting and number location from the Visual Object and Space Perception (VOSP) battery (Warrington and James, 1991), the Coloured Progressive Matrices test of non-verbal reasoning (Raven, 1962), The Trail Making Test part A and B (Reitan, 1992) and the Brixton Test of Spatial Anticipation (Burgess and Shallice, 1997).

Semantic processing was assessed using a number of standard tests. This included three components of the Cambridge 64-item semantic test battery (Bozeat *et al.*, 2000): spoken word–picture matching using 10 semantically-related response options, picture naming, and the picture version of the Camel and Cactus Test (a four alternative-forced-choice test tapping semantic associations). The 96-item three alternative-forced-choice test

synonym judgement task was also run (Jefferies *et al.*, 2009). The object use test (Corbett *et al.*, 2011) involved selecting an object for a task (e.g., BASH A NAIL INTO WOOD). The target was either the canonical tool (e.g., HAMMER), or a non-canonical option (e.g., BRICK), presented among a set of five unrelated distractors. The cookie theft picture description test was run (Goodglass, 1983), giving a measure of verbal fluency for each patient. Additionally, category and letter fluency tasks were run, using eight categories (animals, fruit, birds, breeds of dog, household objects, tools, vehicles, types of boat) and three letters (F, A and S).

Tests of semantic control: Semantic control was assessed using a 2 x 2 manipulation of semantic control demands in two different tasks (Hoffman, 2018) (following Badre *et al.*, 2005). In the first task, participants made semantic decisions based on global semantic association. They were presented with a probe word and asked to select its closest semantic associate from either two or four alternatives (see Supplementary Figure 2 for examples). The strength of relationship between the probe and target was manipulated: the associate was either strongly associated with the probe (e.g., *town-city*) or weakly associated (e.g., *iron-ring*). The Weak Association condition was assumed to place greater demands on semantic control, specifically the *controlled retrieval* of semantic information, as automatic spreading of activation in the semantic network would be insufficient to identify the correct response (Badre and Wagner, 2007).

In the second task, participants were asked to match items based on specific features. At the beginning of each block, participants were given a feature on which to base their decisions (e.g., colour). On each trial, they were given the name of an object and were asked to select another item that was most similar on the specified feature. We manipulated the semantic congruency of the probe and target. On Congruent trials, the probe and target shared a pre-existing semantic relationship, as well matching on the currently relevant feature (e.g., *cloud-snow* are semantically related in addition to matching in colour). In contrast, on Incongruent trials the probe and target shared no meaningful relationship, other than their similarity on the specified feature (e.g., *salt-dove* are both typically white but otherwise semantically unrelated). Moreover, on Incongruent trials one of the foils had a strong semantic relationship with the probe, but did not match on the currently relevant feature (*salt-pepper*). Incongruent trials placed high demands on semantic control, and particularly on *semantic selection* processes, for two reasons. First, because there was no pre-

existing semantic relationship between probe and target, participants could only identify the correct response if they focused selectively on the pre-specified features of the items and not on their other semantic properties. Second, participants were required to inhibit the strong but irrelevant relationship between the probe and foil.

Speech elicitation: Procedures for obtaining speech samples was based on those used by Hoffman et al. (2018a) with healthy participants, but were adapted for use with aphasic patients. Hoffman et al. presented participants with 14 different prompts to elicit speech on different topics. To avoid patient fatigue, we used only six prompts here:

1. Describe the steps you would need to take if going somewhere by train.
2. What do the police do when a crime has been committed?
3. Which is your favourite season and why?
4. Do you think it's a good idea to send people to live on Mars?
5. What sort of things do you have to do to look after a dog?
6. Why do people go to Scotland on holiday?

In the study with healthy participants, written prompts were presented on a computer screen and participants spoke about each one for a fixed period of 60s. Here, the prompts were read aloud by the researcher and no time limit was imposed on responses. We did not use a time limit because we expected speech rate to be reduced in our aphasic patients and therefore that they would produce significantly less speech than controls if only allowed to speak for 60s. Instead, patients were encouraged to speak for as long as chose to about each subject. If they provided little information spontaneously, the researcher encouraged them to elaborate with a neutral comment like “Can you say anything else about that?”. To ensure that potential differences between groups in quantity of speech could not account for our results, we recorded the length of each response in number of words and included this as a covariate in all analyses (see Results). In addition, we performed a second analysis to control for differences in the time available to produce a response. To do this, we computed coherence values using only speech produced in the first 60s of each patient response (thus matching the time period to the control group). We then repeated our main analysis of coherence effects using these values (see Results).

Speech processing and computation of coherence: Spoken responses were digitally recorded for later transcription. Non-lexical fillers (umm, ah etc.) were not transcribed and pauses were not marked. However, all lexical items were transcribed, including lexical fillers (e.g., "like", "I mean"). We also included incomplete or aborted phrases ("We wanted to…we always went to the shops"), repetitions of words or phrases ("But it's not, it's not got air there") and comments about the task itself or the patients' language problems ("I don't know what the word is"). However, analyses that excluded such utterances produced very similar results to those reported here. Our main dependent variables were computed measures of global and local coherence (GC and LC), as described below. A number of other speech markers were also computed and were included in supplementary analyses (see Statistical Analyses).

Measures of GC and LC were generated using an automated computational linguistic approach, first described by Hoffman et al. (2018a). Analyses were implemented in R; the code is publicly available and can easily be applied to new samples (https://osf.io/8atfn/). Our approach made use of latent semantic analysis (LSA) (Landauer and Dumais, 1997). LSA provides the user with vector-based representations of the meanings of words, which can be combined linearly to represent the meanings of whole passages of speech or text (Foltz *et al.*, 1998). Using similar methods to other researchers (Elvevag *et al.*, 2007; Foltz, 2007), we used these representations to characterise the content of each speech sample and to quantify its coherence. Coherence calculated in this way has high internal reliability and test-retest reliability and is highly correlated with human ratings of coherence (Hoffman *et al.*, 2018a).

The computation process is illustrated in Supplementary Figure 3. Our overall strategy was to divide each speech sample into smaller windows (of 20 words each) and to use LSA to generate vector representations of the semantic content of each window. Coherence was then assessed with a moving window approach. LC was assessed by measuring the similarity of the vector for each window with that of the patient's previous window. Therefore, in common with other researchers (Elvevag *et al.*, 2007; Foltz, 2007), we define LC as the degree to which adjoining utterances convey semantically related content. A low LC value would be obtained if a patient switched abruptly between topics during their response. We used the cosine between vectors as the measure of their similarity, so a value of 0 indicates no semantic relationship between windows and 1 indicates identical content.

GC was assessed by comparing the semantic content of each window with a vector representing the prototypical semantics produced by healthy participants to the same prompt (GC). To generate this prototypical representation, we took all the responses made by healthy older adults in Hoffman et al. (2018a) and computed an LSA representation for each one. We then averaged these to give a composite vector that represented the typical semantic content that people produced when responding to the prompt (for further details of the LSA space and averaging procedure, see Hoffman *et al.*, 2018a). For example, the composite vector for the prompt "Describe the steps you would need to take if going somewhere by train" would be similar to the vectors for *train, railway, ticket, station* and so on, as these words were frequently used in responses to this prompt. The GC for each window was defined as the similarity between its vector and the composite vector. Therefore, GC was a measure of how much a patient's response matched the typical semantic content of responses to that prompt. A low GC value would be obtained if a patient tended to talk about other topics that were semantically unrelated to the topic being probed. Thus, our measure of GC captured the degree to which patients maintained their focus on the topic under discussion, in line with the definition used by other researchers (Glosser and Deser, 1992; Wright *et al.*, 2014).

Once GC and LC had been calculated, the window moved one word to the right and the process was repeated, until all windows had been assessed. GC and LC values were averaged across windows to give overall values for each response.

<u>Statistical analyses:</u> Group analyses compared performance in the SA group (not including the non-aphasic RH patient) with the age-matched healthy control group. At the group level, analyses were performed using linear mixed effects (LME) models. Accuracy on the semantic control task was analysed using a generalised LME model with a logit link function. The model had a 2 x 2 x 2 (group x task x control demands) factorial structure and included random intercepts for participants and test items, as well as all random slopes for all factors varying within-unit (Barr *et al.*, 2013). The significance of fixed effects was assessed using a likelihood ratio test to compare the full model with a reduced model identical except for the removal of the effect of interest (Barr *et al.*, 2013).

To test for group differences in coherence, LME models were estimated separately for GC and LC values, with a factorial manipulation of group (SA vs. controls, again excluding the

RH patient) and including response length as a control predictor (since longer responses may be more likely to deviate off-topic; Hoffman *et al.*, 2018a). These models included random intercepts for participants and prompts and random by-probes slopes for the effect of group. Effects of group were assessed using a t-test with Satterthwaite approximation of degrees of freedom, as there were a smaller number of observations here which could cause likelihood ratio tests to be anti-conservative (Huber *et al.*, 2015). To determine whether the coherence scores of individual patients differed from the control group, we repeated the LME analyses, each time comparing the control group to a single patient. This method extends the modified t-test approach to the LME framework allowing us to take item effects into account when testing for impairments (Huber *et al.*, 2015). We also computed correlations between coherence measures and semantic task performance in the patient group.

Finally, we calculated seven other markers of the lexical-semantic properties of speech: the mean frequency, concreteness, age of acquisition, semantic diversity and length of words produced, type:token ratio and the proportion of closed-class words (see Supplementary Materials for full details). To analyse these data, we performed a principal components analysis to reduce the individual measures to four latent factors (Following the same method as our previous studies; Hoffman *et al.*, 2018a; Hoffman, 2019). Consistent with previous work, the four factors appeared to index specificity of semantic content, complexity of vocabulary, coherence and lexical diversity (see Table 4 for factor loadings). We performed LME analyses to the scores on each factor to determine whether patients differed from controls on each factor, and tested whether correlations with semantic performance were present.

Data availability: The data reported in this paper are available at https://osf.io/cvuqz/

Results

Background neuropsychological tests: Scores on the background neuropsychological tests are presented in Table 2. In all tables and figures, SA patients are ordered by the severity of their semantic control impairment (i.e., mean performance on the high control conditions of the semantic control task). SA patients showed evidence of multimodal semantic impairments on

background testing. Semantic impairment in SA is typically accompanied by domain-general executive impairments (Jefferies and Lambon Ralph, 2006; Thompson *et al.*, 2018) and this was also the case here. Patients frequently showed impairment on tests of general executive function, such as the Brixton test of spatial anticipation, the Trails test part B and Raven's progressive matrices. All SA patients performed below the normal range on at least one of these tests, with the exception of SA1 and SA2 (who also showed the most intact performance on tests of semantic control; see below). The right-hemisphere patient, RH1, performed well on the semantic tasks, displaying mild impairment only on the synonym judgement task.

Tests of semantic control: Semantic control was assessed using a forced-choice semantic judgement task with a 2 x 2 design that manipulated the type of semantic judgement (global vs. specific feature) and the need for controlled semantic processing. Accuracy on this task is presented in Figure 1. The data were analysed with generalised mixed effects models. SA patients were significantly impaired on the task as a whole, compared with the healthy control group ($B = 0.58$, $se = 0.22$, $p = 0.004$). There was a main effect of the semantic control manipulation ($B = -0.95$, $se = 0.20$, $p < 0.001$), indicating poorer performance in the high control conditions. Importantly, there was also a marginal interaction between control demands and group ($B = -0.39$, $se = 0.19$, $p = 0.062$), suggesting that manipulating the need for semantic control had a larger effect on the patients than on the controls. There was no effect of task and no interactions with other factors, suggesting that the two different manipulations of control in this experiment (controlled retrieval vs. semantic selection) had similar effects on our participants.

Speech elicitation: Basic information on the speech samples collected from patients is provided in Table 3. We presented either five or six prompts to each patient, depending on the time available for testing. For SA3, we were only able to analyse data from four prompts for analysis because in the other two cases she failed to produce enough words to provide reliable coherence data (7 and 17 words only). The mean duration of patients' responses varied between 62 and 115s (in contrast, a 60s time limit was imposed on controls). Despite this, patients produced fewer words on average than controls ($t(17) = 2.69$, $p = 0.016$) because they spoke at a much slower rate ($t(17) = 4.69$, $p < 0.001$). Examples of patients' responses are given are provided as Supplementary Materials. They appeared to frequently deviate from

the topic about which they were asked, behaviour which we analyse formally in the next section.

Speech coherence: Our formal assessment of coherence involved using previously-validated computational linguistics methods to quantify the global coherence (GC) and local coherence of each response (LC). Mean GC and LC for each patient is shown in Table 3, with SA and Control means in Figure 2. LME analyses of these data (including length of responses as a covariate) revealed that SA patients were significantly less globally coherent than controls ($t(8.4) = 4.78$, $p = 0.001$) and, to a lesser extent, less locally coherent ($t(12.5) = 2.21$, $p = 0.047$). At the level of individual patients, all patients except SA1 showed significantly impaired GC scores relative to controls ($p < 0.05$). SA2 and SA7 also exhibited LC scores that were significantly lower than the control group ($p < 0.05$).

Correlations between the coherence measures, response length and scores on the high-control semantic conditions are shown in Figure 3. There was a strong tendency for patients who performed better on the high-control semantic tasks to score more highly for GC in speech (Incongruent Features: $r = 0.75$, $p = 0.03$; Weak Associations: $r = 0.55$, $p = 0.16$). In contrast, semantic control performance did not predict LC values, nor did it predict the average length of responses.

Other characteristics of speech: Finally, we computed a range of other speech markers to determine the degree to which semantic control impairment affected other aspects of speech content (see Supplementary Materials for details). Following the same method as previous studies (Hoffman *et al.*, 2018a; Hoffman, 2019), we performed a principal components analysis on these data, which identified four factors corresponding to distinct aspects of speech production (see Table 4 for factor loadings):

1. Semantic specificity, the tendency to produce highly concrete, low frequency words with low semantic diversity
2. Complexity of vocabulary, the degree to which participants produced long, low frequency, late-acquired words
3. Coherence, which was strongly related to the GC and LC measures analysed above
4. Lexical diversity, the degree to which participants produced a wide range of different lexical items and open-class words.

Table 4 also shows the mean factor scores for each group and the effects of group, assessed using LME models of the same form used to analyse GC and LC. In addition to differing from controls on the latent coherence factor, SA patients also tended to use less complex vocabulary than controls and to exhibit less lexical diversity. Correlations between patients' performance on tests of semantic control and their scores on the latent speech factors are shown in Figure 4. The coherence factor showed the strongest correlations with semantic control ability, replicating the main analysis of GC. However, there was also a weaker tendency for patients with more semantic control ability to use more complex vocabulary. This suggests that semantic control ability influences the coherence of speech more than it does other aspects of speech content (in line with Hoffman et al.'s (2018a) findings in healthy participants).

Discussion

Maintaining coherence when speaking is a critical skill which is thought to depend on the ability to retrieve and select currently-relevant knowledge from semantic memory (known as semantic control). We investigated the coherence of connected speech in a case-series of aphasic stroke patients with deficits of semantic control. We predicted that poor control over their access to semantic knowledge would have detrimental effects on their coherence. In line with our prediction, patients showed gross impairments to global coherence: compared to controls, their spoken responses were less relevant to the topics we asked them to speak about. They showed milder deficits in local coherence, indicating that they were also less able to maintain a coherent topic from moment to moment when speaking. Moreover, we found that global coherence was correlated with performance on tests of semantic control, with the most impaired patients exhibiting the lowest global coherence values. As one would expect in an aphasic group, other aspects of speech production were also impaired but none showed such a strong correlation with semantic control abilities. Our findings provide converging support for the idea that semantic control processes are central to maintaining coherence in speech and indicate that poorly coherent speech is a major issue for patients with deficits of semantic control.

Our patients exhibited greater impairments to global coherence than to local coherence. In other words, they were mostly able to maintain semantic relationships between consecutive utterances, but they were poor at ensuring their utterances were related to the topic at hand. The result was that their verbal output often had the character of an undirected stream of consciousness: they began coherently but over time, they drifted away from the topic they had been asked about (see Supplementary Materials for examples). Our interpretation of this behaviour is that *automatic* semantic retrieval processes are relatively intact in these cases. When speaking, they successfully activate a chain of meaningful associations which drive their speech output. However, in order to maintain focus on the topic at hand, top-down control processes must act on this information retrieved from semantic memory, to ensure that information selected for production is relevant to the topic under discussion. Disruption to these selection processes appear to be responsible for the patients' global coherence deficits.

A distinction has often been drawn between two distinct forms of semantic control: *semantic selection* of task-relevant knowledge vs. *controlled retrieval* of less salient knowledge (Badre and Wagner, 2007). Our account of the SA patients' coherence deficits holds that the selection element plays the key role in maintaining global coherence. This is consistent with our previous study in healthy participants (Hoffman *et al.*, 2018a) and with an fMRI study, again in healthy older people, indicating that high coherence during speech is correlated with increased activation in the pars triangularis (BA45) region of left inferior frontal gyrus (Hoffman, 2019). This area is strongly implicated in semantic selection (Thompson-Schill *et al.*, 1997; Badre *et al.*, 2005; Badre and Wagner, 2007). In the present study, however, we found that tasks probing selection and controlled retrieval showed similar correlations with global coherence. SA patients typically show parallel impairments to both of these aspects of semantic control, making it difficult to tease apart contributions of each in this patient group. However, the converging evidence across studies suggests that poor selection of semantic information is at the root of coherence impairments.

Another outstanding question is the degree to which the semantic selection processes we measured here overlap with executive functions in other cognitive domains. This is not an issue we were able to address directly in this study because patients did not complete comprehensive assessments of executive function in other domains. However, other studies in SA patients indicate that general executive deficits typically co-occur with, and are

correlated with, semantic control impairments (Jefferies and Lambon Ralph, 2006). Furthermore, patients who present with a general dysexecutive syndrome also show the hallmarks of a semantic control disorder (Thompson *et al.*, 2018). It is not clear the degree to which these co-occurring deficits indicate a common functional system, or are simply a consequence of concomitant damage to distinct but neurally proximate systems. However, recent work in healthy individuals suggests that semantic selection ability patterns closely with performance on non-semantic executive tasks, while controlled retrieval is more closely related to semantic abilities (Hoffman, 2018). Thus, it is possible that inhibition of irrelevant semantic information is closely related to inhibitory functions in other cognitive domains.

For the present study, we deliberately used a relatively unconstrained speech elicitation task. Patients were given a topic to speak about and were required to structure their own response with minimal support from the experimenter. We chose this method to provide a strong test of patients' ability to regulate their own speech output. However, it is worth noting that SA patients show positive effects of increasing task constraints across a range of semantically-driven tasks and we would expect to see similar effects here (Noonan *et al.*, 2010; Corbett *et al.*, 2011). For example, in everyday life the patient's conversational partners may play an important role in providing verbal and non-verbal cues that direct them back towards topic-relevant discourse. There is a parallel here with studies in healthy older adults, in which coherence impairments are greatest when verbal topic prompts are used and less prominent when participants are asked to describe pictures or comic strips, where there is a visual reminder of the subject matter throughout (James *et al.*, 1998; Wright *et al.*, 2014). One way to improve coherence in this patient group may therefore be to provide external cues that maintain their focus on the current topic.

In conclusion, the unique contribution of this research is to show the nature of impairments to everyday conversation that semantic control impairments bring. In particular, SA patients have difficulties maintaining global coherence, where their unregulated retrieval of semantic information leads them "off the rails" and progressively away from the topic at hand. Better understanding of these processes will be critical in uncovering the root causes of conversational difficulties experienced by aphasic patients and will be useful in guiding rehabilitation in hospital, clinics and in the home.

Acknowledgements

We are grateful to the patients for generously offering their time to take part in this study.

Funding

PH was supported by The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross council Lifelong Health and Wellbeing Initiative (MR/K026992/1). Funding from the Biotechnology and Biological Sciences Research Council (BBSRC) and Medical Research Council (MRC) is gratefully acknowledged. H.E.T. received support from the Wellcome Trust [ref: 105624] through the Centre for Chronic Diseases and Disorders (C2D2) at the University of York.

References

Alexander MP. Impairments of procedures for implementing complex language are due to disruption of frontal attention processes. Journal of the International Neuropsychological Society 2006; 12(2): 236-47.

Arbuckle TY, Gold DP. Aging, inhibition, and verbosity. J Gerontol 1993; 48(5): P225-32.

Ash S, Moore P, Antani S, McCawley G, Work M, Grossman M. Trying to tell a tale: Discourse impairments in progressive aphasia and frontotemporal dementia. Neurology 2006; 66: 1405-13.

Badre D, Poldrack RA, Pare-Blagoev EJ, Insler RZ, Wagner AD. Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. Neuron 2005; 47(6): 907-18.

Badre D, Wagner AD. Left ventrolateral prefrontal cortex and the cognitive control of memory. Neuropsychologia 2007; 45(13): 2883-901.

Barker MS, Young B, Robinson GA. Cohesive and coherent connected speech deficits in mild stroke. Brain and language 2017; 168: 23-36.

Barr DJ, Levy R, Scheepers C, Tily HJ. Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of memory and language 2013; 68(3): 255-78.

Bozeat S, Lambon Ralph MA, Patterson K, Garrard P, Hodges JR. Non-verbal semantic impairment in semantic dementia. Neuropsychologia 2000; 38(9): 1207-15.

Brysbaert M, Warriner AB, Kuperman V. Concreteness ratings for 40 thousand generally known English word lemmas. Behavior research methods 2014; 46(3): 904-11.

Burgess P, Shallice T. The Hayling and Brixton tests. Suffolk, England: Thames Valley Test Company; 1997.

Christiansen JA. Coherence violations and propositional usage in the narratives of fluent aphasics. Brain and Language 1995; 51(2): 291-317.

Coelho CA. Story narratives of adults with closed head injury and non-brain-injured adults: Influence of socioeconomic status, elicitation task, and executive functioning. Journal of Speech, Language, and Hearing Research 2002; 45(6): 1232-48.

Corbett F, Jefferies E, Lambon Ralph MA. Deregulated semantic cognition follows prefrontal and temporoparietal damage: Evidence from the impact of task constraint on non-verbal object use. Journal of Cognitive Neuroscience 2011; 23: 1125-35.

Covington MA, McFall JD. Cutting the Gordian knot: The moving-average type–token ratio (MATTR). Journal of quantitative linguistics 2010; 17(2): 94-100.

Ellis C, Henderson A, Wright HH, Rogalski Y. Global coherence during discourse production in adults: a review of the literature. International journal of language & communication disorders 2016; 51(4): 359-67.

Elvevag B, Foltz PW, Weinberger DR, Goldberg TE. Quantifying incoherence in speech: An automated methodology and novel application to schizophrenia. Schizophrenia research 2007; 93(1-3): 304-16.

Foltz PW. Discourse coherence and LSA. Handbook of latent semantic analysis. Mahwah (NJ): Erlbaum; 2007. p. 167-84.

Foltz PW, Kintsch W, Landauer TK. The measurement of textual coherence with latent semantic analysis. Discourse Processes 1998; 25(2-3): 285-307.

Glosser G, Deser T. Patterns of Discourse Production among Neurological Patients with Fluent Language Disorders. Brain and Language 1991; 40(1): 67-88.

Glosser G, Deser T. A comparison of changes in macrolinguistic and microlinguistic aspects of discourse production in normal aging. J Gerontol 1992; 47(4): P266-72.

Gold D, Andres D, Arbuckle T, Schwartzman A. Measurement and correlates of verbosity in elderly people. Journal of gerontology 1988; 43(2): P27-P33.

Goodglass H. The assessment of aphasia and related disorders (2nd Edition). Philadelphia: Lea & Febiger; 1983.

Harris Wright H, Capilouto GJ. Considering a multi-level approach to understanding maintenance of global coherence in adults with aphasia. Aphasiology 2012; 26(5): 656-72.

Hoffman P. An individual differences approach to semantic cognition: Divergent effects of age on representation, retrieval and selection. Scientific reports 2018; 8(1): 8145.

Hoffman P. Reductions in prefrontal activation predict off-topic utterances during speech production. Nature communications 2019; 10(1): 515.

Hoffman P, Lambon Ralph MA, Rogers TT. Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. Behavior Research Methods 2013; 45(3): 718-30.

Hoffman P, Loginova E, Russell A. Poor coherence in older people's speech is explained by impaired semantic and executive processes. eLife 2018a; 7: e38907.

Hoffman P, McClelland JL, Lambon Ralph MA. Concepts, control and context: A connectionist account of normal and disordered semantic cognition. Psychological Review 2018b; 125: 293-328.

Huber S, Klein E, Moeller K, Willmes K. Comparing a single case to a control group–applying linear mixed effects models to repeated measures data. Cortex 2015; 71: 148-59.

James LE, Burke DM, Austin A, Hulme E. Production and perception of" verbosity" in younger and older adults. Psychology and aging 1998; 13(3): 355.

Jefferies E. The neural basis of semantic cognition: Converging evidence from neuropsychology, neuroimaging and TMS. Cortex 2013; 49: 611-25.

Jefferies E, Lambon Ralph MA. Semantic impairment in stroke aphasia versus semantic dementia: a case-series comparison. Brain 2006; 129(Pt 8): 2132-47.

Jefferies E, Patterson K, Jones RW, Lambon Ralph MA. Comprehension of concrete and abstract words in semantic dementia. Neuropsychology 2009; 23: 492-9.

Juncos-Rabadan O, Pereiro AX, Rodriguez MS. Narrative speech in aging: quantity, information content, and cohesion. Brain Lang 2005; 95(3): 423-34.

Jung J, Cloutman LL, Binney RJ, Lambon Ralph MA. The structural connectivity of higher order association cortices reflects human functional brain networks. Cortex 2016.

Karaduman A, Göksun T, Chatterjee A. Narratives of focal brain injured individuals: A macro-level analysis. Neuropsychologia 2017; 99: 314-25.

Kintsch W, Vandijk TA. Toward a Model of Text Comprehension and Production. Psychological Review 1978; 85(5): 363-94.

Kintz S, Fergadiotis G, Wright HH. Aging effects on discourse production. In: Wright HH, editor. Cognition, Language and Aging. Amsterdam/Philadelphia: John Benjamins Publishing Company; 2016. p. 81-106.

Kuperman V, Stadthagen-Gonzalez H, Brysbaert M. Age-of-acquisition ratings for 30,000 English words. Behavior Research Methods 2012; 44(4): 978-90.

Lambon Ralph MA, Jefferies E, Patterson K, Rogers TT. The neural and computational bases of semantic cognition. Nature Reviews Neuroscience 2017; 18: 42-55.

Landauer TK, Dumais ST. A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. Psychological Review 1997; 104: 211-40.

Long MR, Horton WS, Rohde H, Sorace A. Individual differences in switching and inhibition predict perspective-taking across the lifespan. Cognition 2018; 170: 25-30.

Marini A, Andreetta S. Age-related effects on language production: A combined psycholinguistic and neurocognitive perspective. In: Wright HH, editor. Cognition, Language and Aging. Amsterdam/Philadelphia: John Benjamins Publishing Company; 2016. p. 55-79.

Marini A, Boewe A, Caltagirone C, Carlomagno S. Age-related differences in the production of textual descriptions. J Psycholinguist Res 2005; 34(5): 439-63.

Marini A, Zettin M, Galetto V. Cognitive correlates of narrative impairment in moderate traumatic brain injury. Neuropsychologia 2014; 64: 282-8.

Noonan KA, Jefferies E, Corbett F, Lambon Ralph MA. Elucidating the nature of deregulated semantic cognition in semantic aphasia: Evidence for the roles of the prefrontal and temporoparietal cortices. Journal of Cognitive Neuroscience 2010; 22: 1597-613.

North AJ, Ulatowska HK, Macaluso-Haynes S, Bell H. Discourse performance in older adults. The International Journal of Aging and Human Development 1986; 23(4): 267-83.

Pushkar D, Basevitz P, Arbuckle T, Nohara-LeClair M, Lapidus S, Peled M. Social behavior and off-target verbosity in elderly people. Psychology and aging 2000; 15(2): 361.

Raven JC. Coloured progressive matrices sets A, AB, B. London: H. K. Lewis; 1962.

Reitan RM. Trail Making Test: Manual for administration and scoring. Tucson, AZ: Reitan Neuropsychology Laboratory; 1992.

Rogalski Y, Altmann LJ, Plummer-D'Amato P, Behrman AL, Marsiske M. Discourse coherence and cognition after stroke: A dual task study. Journal of communication disorders 2010; 43(3): 212-24.

Thompson-Schill SL, D'Esposito M, Aguirre GK, Farah MJ. Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. P Natl Acad Sci USA 1997; 94: 14792-7.

Thompson HE, Almaghyuli A, Noonan KA, barak O, Lambon Ralph MA, Jefferies E. The contribution of executive control to semantic cognition: Convergent evidence from semantic aphasia and executive dysfunction. Journal of neuropsychology 2018.

Van Heuven WJ, Mandera P, Keuleers E, Brysbaert M. SUBTLEX-UK: A new and improved word frequency database for British English. The Quarterly Journal of Experimental Psychology 2014; 67(6): 1176-90.

Warrington EK, James M. The Visual Object and Space Perception battery. Bury St. Edmunds, Suffolk: Thames Valley Test Company; 1991.

Wechsler D. Wechsler memory scale: Revised (WMS-R). New York: Psychological Corporation; 1987.

Wright HH, Koutsoftas AD, Capilouto GJ, Fergadiotis G. Global coherence in younger and older adults: Influence of cognitive processes and discourse type. Aging, Neuropsychology, and Cognition 2014; 21(2): 174-96.

Xiang H-D, Fonteijn HM, Norris DG, Hagoort P. Topographical functional connectivity pattern in the perisylvian language networks. Cerebral cortex 2009; 20(3): 549-60.

Table 1: Patient demographic and lesion information

Patient	Age	Years of education	Years post stroke	Lesion location
RH1	56	16	8	Right frontal and temporal
SA1	59	13	8.5	Left frontal and superior temporal
SA2	53	13	7	Left hemisphere; no scan
SA3	78	11	4.5	Left frontal and parietal
SA4	59	11	8	Left frontal and parietal
SA5	74	16	4.5	Left frontal and parietal
SA6	54	16	4.5	Left hemisphere; no scan
SA7	64	12	5.5	Left posterior/ventral temporal and occipital

Table 2: Scores on background neuropsychological tests

	Normative mean (cut-off)	RH1	SA1	SA2	SA3	SA4	SA5	SA6	SA7
Cookie theft (words per minute)		NT	38	58	54	60	77	41	97
Category fluency (8)	115.1 (76)	18[a]	69*	NT	80	14*	NT	46*	24*
Letter fluency (F,A,S)	41.1 (19)	17[b]	12*	NT	16*	3*	NT	18*	19*
Semantic Cognition									
Picture naming /64	62.3 (59)	64	46*	14/16*	56*	60	NT	13/16*	21*
Word-picture matching /64	63.8 (62)	64	63	63	64	62	11/16*	15/16*	46*
Synonym judgement /96	94.5 (91)	88*	81*	87*	78*	66*	64*	81*	81*
Picture semantic association /64	59.1 (51)	56	61	60	53	45*	49*	58	44*
Canonical Object use /37	-	-	37	36	37	37	32	25	30
Non-canonical object use /37	36.0 (34)	-	32*	32*	26*	34	17*	12*	12*
Cognition and Executive									
Brixton test of spatial anticipation /54	39 (30)	21*	39	45	31	24*	27*	41	29*
Trail-making test Part A /24	23.9 (23)	24	24	24	24	24	23	24	23
Trail-making test Part B /23	23.5 (21)	22	21	23	19*	1*	16*	14*	7*
Raven's coloured progressive matrices /36	29.7 (20)	30	33	29	31	19*	30	36	29
Forward digit span	6.8 (4)	7	6	3*	5	6	2*	3*	8
Backward digit span	4.8 (3)	4	4	NT	3	2*	1*	3	4
VOSP dot counting /10	9.9 (9)	10	10	NT	10	10	NT	10	8*
VOSP number location /10	9.4 (7)	10	8	NT	5*	10	NT	10	8

*Control data were obtained from published norms. Minimum control score indicates cut-off below which performance is considered abnormal (two standard deviations below the mean if no other threshold was provided). Patient scores below this level are indicated by *. For some of the semantic tests, shortened 16-item versions were administered to some patients. [a] = one category "animal", [b] = one letter "S". NT = not tested. VOSP = Visual Object and Space Perception battery.*

Table 3: Characteristics of speech samples

	RH1	SA1	SA2	SA3	SA4	SA5	SA6	SA7	SA mean	Control mean
Number of samples	6	5	6	4	6	5	6	6	--	--
Response duration (s)	64	94	115	81	81	82	84	62	85.5 (16.2)	60 (0)
Response length (words)	173	80	93	91	91	164	54	137	101.4 (37.0)	136.7 (20.8)
Speech rate (words per minute)	160	51	48	67	68	120	38	134	75.1 (37.0)	136.7 (20.8)
Global coherence	0.46	0.45	0.38	0.36	0.35	0.21	0.37	0.35	0.35 (0.07)	0.50 (0.03)
Local coherence	0.27	0.28	0.24	0.30	0.30	0.29	0.30	0.23	0.28 (0.03)	0.31 (0.03)

SA = stroke aphasia.

Table 4: Factor loadings in principal components analysis of speech characteristics, and mean factor scores in each group

Loadings with absolute values >0.4 are shown in bold. SA = stroke aphasia.

Measure	Factor 1: Semantic specificity	Factor 2: Vocabulary	Factor 3: Coherence	Factor 4: Lexical diversity
Noun frequency	**-.69**	**-.53**	.06	-.01
Noun semantic diversity	**-.89**	-.08	-.01	.04
Noun concreteness	**.82**	-.32	.05	.08
Noun age of acquisition	-.06	**.93**	-.09	-.13
Noun phonemic length	.01	**.77**	.14	.11
Global coherence	.07	.16	**.82**	.03
Local coherence	-.04	-.12	**.93**	-.06
% closed class	-.15	.17	.04	**-.92**
Type:token ratio	-.26	.33	.01	**.61**
Mean factor scores by group:				
Controls mean (s.d.)	.10 (.31)	.42 (.64)	.50 (.27)	.47 (.39)
SA mean (s.d.)	-.07 (1.12)	-.32 (.62)	-.41 (.35)	-.37 (.64)
Effect of group	$t(14.1) = 1.07$, $p = 0.302$	$t(11.5) = 3.57$, $p = 0.004$	$t(7.6) = 4.21$, $p = 0.003$	$t(18.5) = 2.54$, $p = 0.020$

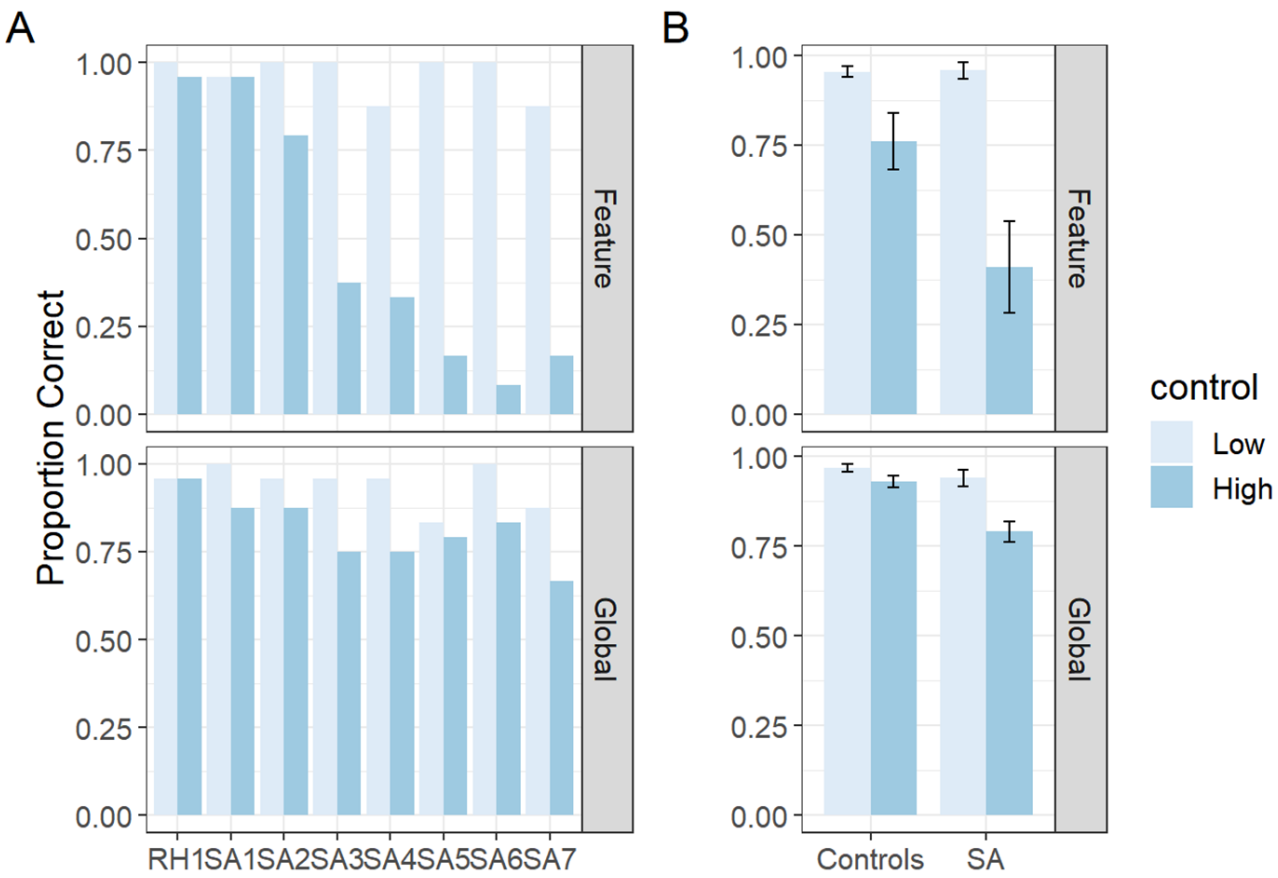
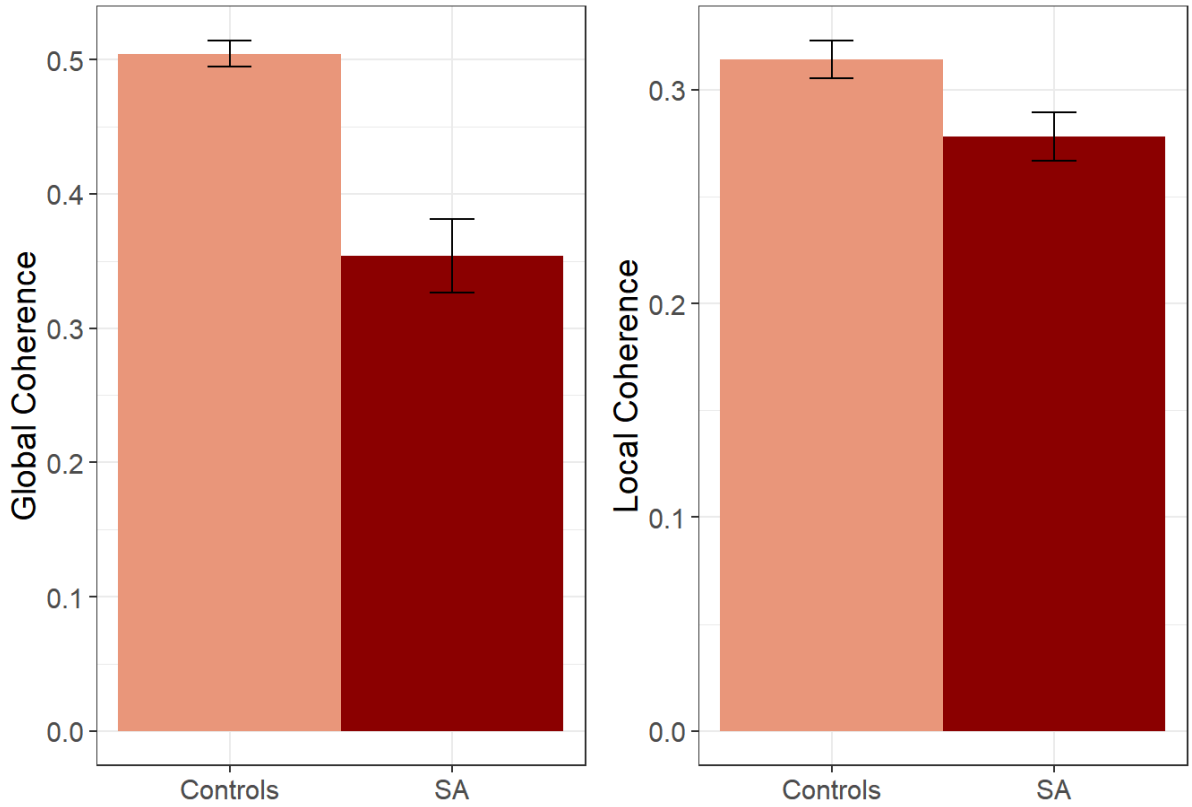


Figure 1: Performance on semantic control tasks (A) Performance in each patient, (B) means for SA patients and controls. Bars indicate one standard error of the mean. SA = stroke aphasia.



<u>Figure 2: Mean coherence levels for SA patients and controls.</u> Bars indicate one standard error of the mean. SA = stroke aphasia.

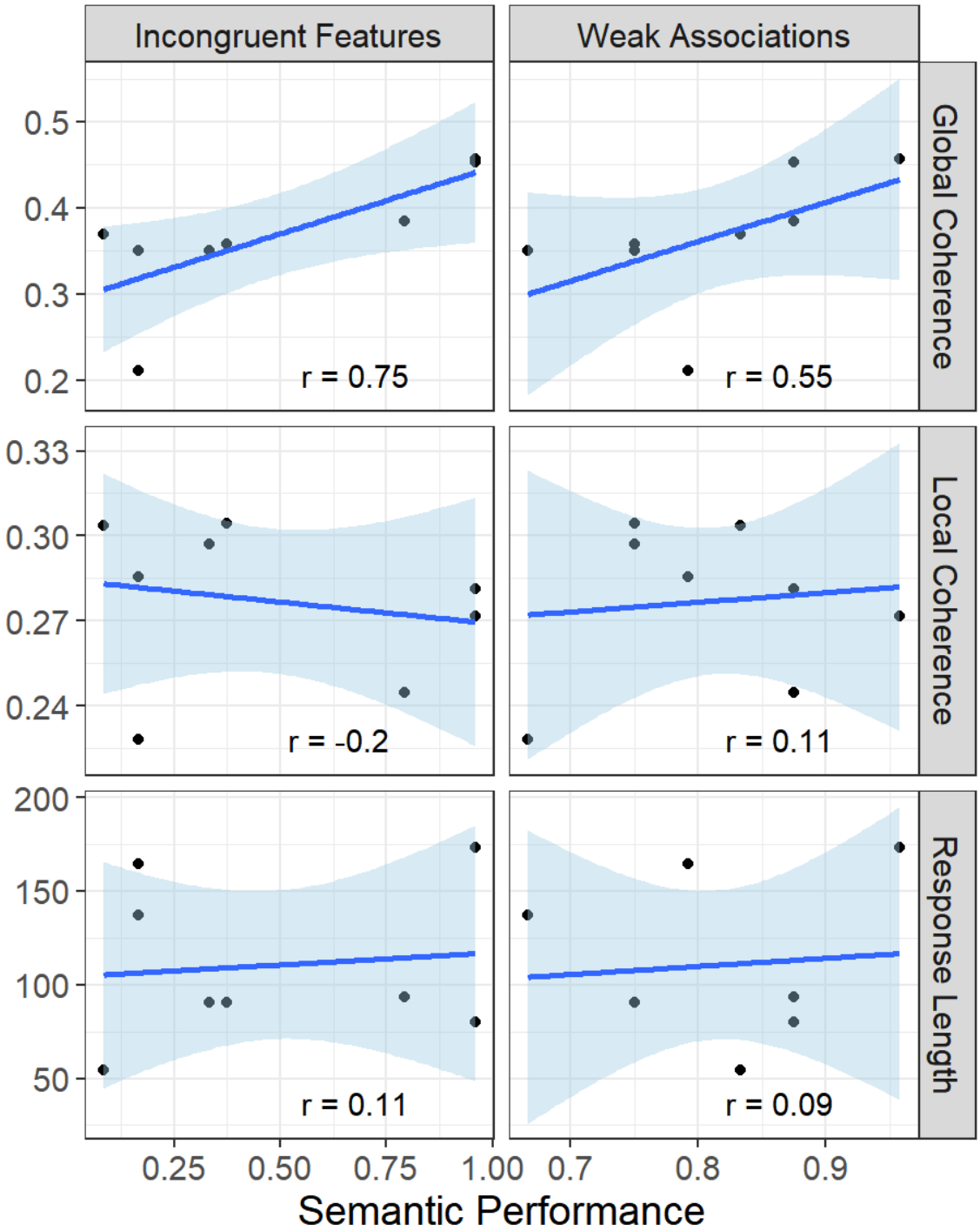


Figure 3: Relationships between coherence, response length and semantic control performance in the patients

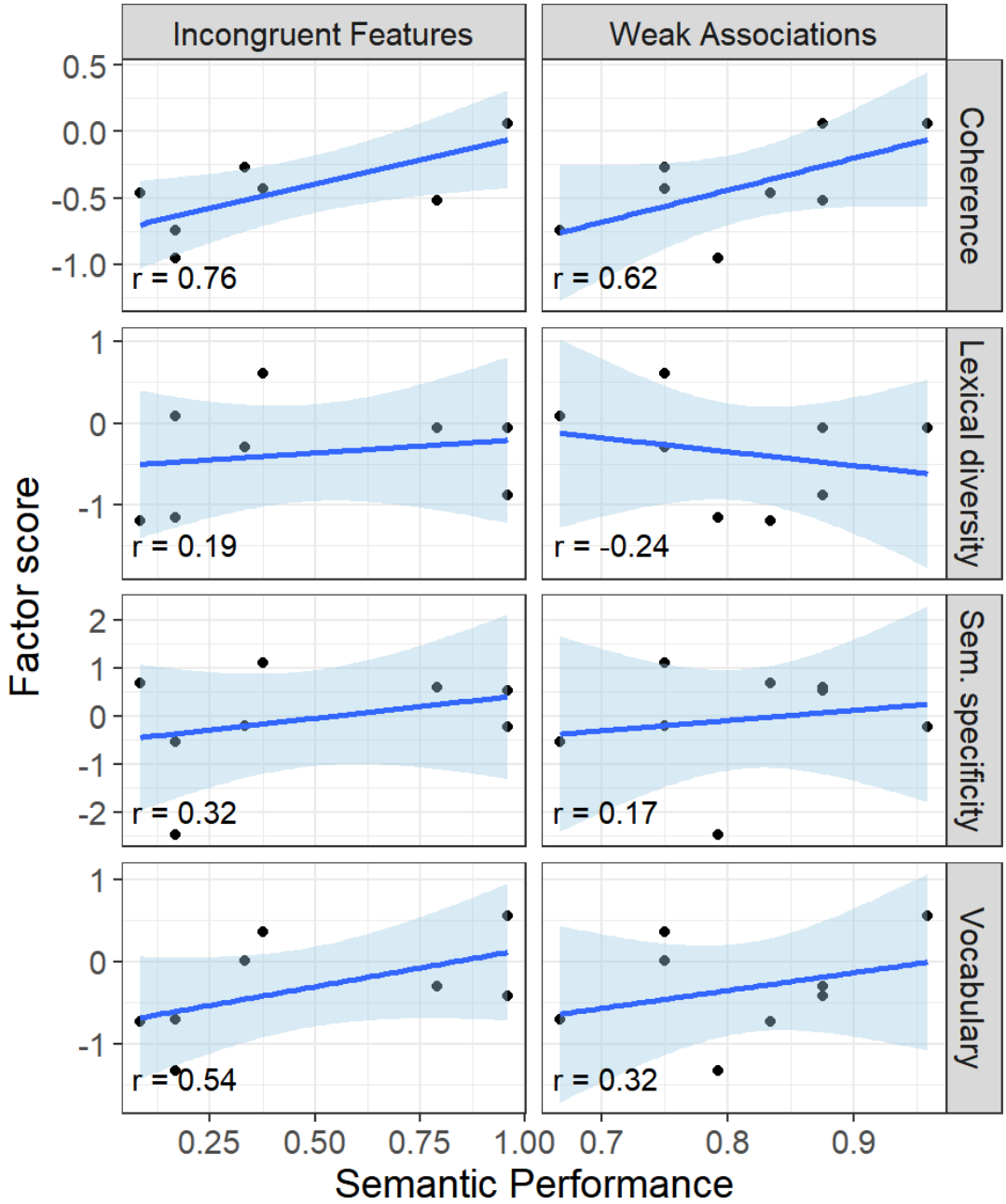


Figure 4: Relationships between semantic control performance and latent speech factors in the patients