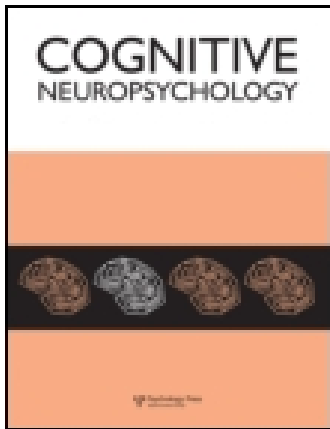


This article was downloaded by: [University of Virginia, Charlottesville]

On: 11 December 2014, At: 07:19

Publisher: Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office:
Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Cognitive Neuropsychology

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/pcgn20>

Localizing the deficit in a case of jargonaphasia

Andrew C. Olson ^a, Cristina Romani ^b & Liz Halloran ^a

^a University of Birmingham, UK

^b University of Aston, Birmingham, UK

Published online: 10 Apr 2007.

To cite this article: Andrew C. Olson, Cristina Romani & Liz Halloran (2007) Localizing the deficit in a case of jargonaphasia, *Cognitive Neuropsychology*, 24:2, 211-238, DOI: [10.1080/02643290601137017](https://doi.org/10.1080/02643290601137017)

To link to this article: <http://dx.doi.org/10.1080/02643290601137017>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

Localizing the deficit in a case of jargonaphasia

Andrew C. Olson

University of Birmingham, UK

Cristina Romani

University of Aston, Birmingham, UK

Liz Halloran

University of Birmingham, UK

We report the case of a neologistic jargonaphasic and ask whether her target-related and abstruse neologisms are the result of a single deficit, which affects some items more severely than others, or two deficits: one to lexical access and the other to phonological encoding. We analyse both correct/incorrect performance and errors and apply both traditional and formal methods (maximum-likelihood estimation and model selection). All evidence points to a single deficit at the level of phonological encoding. Further characteristics are used to constrain the locus still further. V.S. does not show the type of length effect expected of a memory component, nor the pattern of errors associated with an articulatory deficit. We conclude that her neologistic errors can result from a single deficit at a level of phonological encoding that immediately follows lexical access where segments are represented in terms of their features. We do not conclude, however, that this is the only possible locus that will produce phonological errors in aphasia, or, indeed, jargonaphasia.

Jargonaphasia is the term that has been given to a heterogeneous set of language disorders that are similar because patients produce fluent but unintelligible speech. Three broad types of jargonaphasia have been identified (Alajouanine, 1956; Perecman & Brown, 1985). In *semantic jargonaphasia* patients use real words, but in inappropriate combinations that render their speech unintelligible. *Neologistic jargonaphasia* is characterized by speech that is often normal and intelligible, but which includes distortions of intended words,

some of which are severe enough that the intended word is not apparent. Finally, *phonemic jargon* (or *undifferentiated jargon*) has been used to describe speech that is phonotactically regular, but where few or no real words can be identified. We report the single case of a patient who has fluent speech with many intact single words, but also with abstruse and target-related neologisms. In the broad typology from the literature she would be categorized as a neologistic jargonaphasic.

Correspondence should be addressed to Andrew Olson, School of Psychology, University of Birmingham, Edgbaston, Birmingham, B15 2TT (E-mail: olsonac@bham.ac.uk).

We would like to thank V.S. for all her help, cooperation and good humour throughout the study. Part of this research was done for the PhD dissertation of the third author. We are grateful for the comments of Alfonso Caramazza and three anonymous reviewers. Funding for patient transport was provided by the University of Birmingham, Behavioural Brain Sciences Centre.

Some neologistic errors are clearly related to their targets (target-related neologisms, e.g., *strawberry* → *strubury*). Others show no discernible relationship (abstruse neologisms, e.g., *suitcase* → *teligom*). We are concerned with the number and location of deficits responsible for these errors. To understand where they might arise, we briefly outline a cognitive architecture that we assume underlies speech production. We base our analysis on assumptions common to several prominent models derived from normal and impaired performance (e.g., Caramazza, 1997; Dell & O'Seaghdha, 1992; Levelt, Roelofs, & Meyer, 1999). These models do not always agree in their details, but mostly these details are not at issue here.

We assume that reading, naming, and repetition all access the same level of conceptual/semantic representation. Following the conceptual representation, there is a lexical level of representation where words have a unitary representation. This is the lemma level of Levelt et al. (1999), the p-lexeme level of Caramazza (1997), and the lexical nodes of Dell and O'Seaghdha (1992). This lexical level of representation does not include segmental information. There is some controversy about whether this level is amodal or modality specific (Caramazza, 1997; Roelofs, Meyer, & Levelt, 1998), but the data that we report here do not distinguish these alternatives.

Lexical representations must activate their corresponding sounds. In some models, access to phonological information is preceded by a level that is still lexical and represents information about word form (e.g., Levelt et al., 1999; see also discussion in Schwartz, Dell, Martin, Gahl, & Sobel, 2006). This level usually includes information about morphemes and metrical information. It is often ambiguous whether the following segmental layer is specifically *lexical*. This is the case because models have been developed almost exclusively for *word* processing, and do not deal with the issue of where lexical and nonlexical processing converge. For example, there is a segmental layer that is part of the form stratum in Levelt et al.'s (1999) lexical network. Presumably, however, this layer

might also be used when a participant is asked to repeat words that they are not familiar with (e.g., nonsense words or words from a foreign language).

Our current case does not help decide whether there is a segmental layer that is addressed by whole-word representations but is not accessed during nonlexical tasks. We make the minimal assumption that a segmental layer exists that is addressed by both whole-word representations (e.g., during picture naming) and nonlexical processes (e.g., during repetition and reading of nonwords). Lexical and nonlexical information must be integrated at this segmental level (for models with integration see Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Hanley, Dell, Kay, & Baron, 2004; Harm & Seidenberg, 2004; for neuropsychological data supporting integration see Hanley & Kay, 1997; Hanley, Kay, & Edwards, 2002; Hillis & Caramazza, 1991). All further speech processing is shared.

In the domain of phonological/articulatory processing, we follow proposals in the neuropsychological literature and test whether jargon-aphasic errors are produced by a memory component where phonological information is stored while articulation takes place and where it decays over a relatively short period of time (the phonological buffer, Bub, Black, Howell, & Kertesz, 1987; Caramazza, Miceli, & Villa, 1986; Shallice, Rumiat, & Zadini, 2000). We also assume that there may be more than one level of segmental representation. Phonological representations that are more central may be abstract in comparison to representations that are used to direct articulation (e.g., Goldrick & Rapp, 2007; Roelofs, 1999; Romani & Galluzzi, 2005).

Models generally agree about the distinction into conceptual, lexeme, and segmental levels. In the first part of our investigation we: (a) determine the number of deficits responsible for our patient's jargonaphasia, and (b) localize the deficit in relation to these broad distinctions. In the second part we localize the deficit more specifically.

One versus two deficits

All explanations of neologistic jargonaphasia postulate a postlexical deficit when there are target-related errors, since the phonemes of the target have clearly been available. Where accounts differ is over the source of abstruse errors. This is a long-standing question about jargonaphasia, embodied by the contrast between “conduction theory” and “anomia theory”. If there is only one deficit, the postlexical locus produces both the target-related and abstruse errors. Abstruse errors are simply more severe distortions of their targets (the “conduction theory”, Kertesz & Benson, 1970; more recent accounts describe a failure or weakness in activating phonological forms from a lexical representation, Kohn & Smith, 1994; Miller & Ellis, 1987; Robson, Pring, Marshall, & Chiat, 2003). An alternative is that abstruse neologisms result when a *separate* deficit prevents lexical access, and a random generator for syllables (Buckingham, 1981) or morphemes (Butterworth, 1979) fills in the response (the “anomia theory”, Buckingham, 1977; for a review of these approaches see Marshall, 2006).

To distinguish one- and two-deficit alternatives, we do three kinds of analysis. First, we look at the types of error that result from three different tasks: picture naming, reading, and repetition. Second, we analyse how the imageability, frequency, and length of the target word influence the number of errors made in the three tasks. Finally, we analyse the distribution of phonemes that are shared in targets and responses.

If there are two deficits, reading and repetition can be more accurate than naming and produce fewer abstruse errors, since reading and repetition can be improved by a contribution from nonlexical processes. A lack of difference is less revealing; fewer abstruse errors are only predicted when nonword reading and repetition are good.

If errors have two sources, the factors that cause lexical access to fail will be different from the factors that cause postlexical errors. As a result, abstruse errors should be influenced by

different factors from those influencing target-related errors. Imageability is thought to relate to semantic properties (imageable items have more semantic features, as in Harm & Seidenberg, 2004; imageable items have both verbal and visual codes, Paivio, 1991; imageable items are linked to more propositions, Jones, 1985; but see also Davies & Funnell, 2000). Imageability should be influential at conceptual and/or lexical levels, but less influential postlexically. It should influence abstruse errors, but not target-related errors. Word frequency and/or age of acquisition are also associated with lexical representations or with lexical phonological codes (Barry, Hirsh, Johnston, & Williams, 2001; Dell, 1988; Jescheniak & Levelt, 1994). Length effects, instead, are associated with postlexical representations based on segments or syllables, but not with unitary representations of words (e.g., Caramazza et al., 1986; Shallice et al., 2000). Target-related errors should be influenced by length.

Whether a particular factor influences just the level where it is represented or also subsequent levels depends on the degree of interaction between levels. We assume that there is some interaction due to cascade and/or feedback (Hillis & Caramazza, 1991; Rapp & Goldrick, 2000, 2004; Roelofs, 2004), but that the influence of a factor diminishes with distance from its source. Certainly, factors like imageability or word frequency do not always influence speech errors (for imageability: Bub et al., 1987; Shallice et al., 2000; for frequency: Aichert & Ziegler, 2004; Romani & Galluzzi, 2005).

If there is a single deficit, target-related and abstruse errors should be governed by the same influences, and these should only differ if the severity of abstruse errors introduces a difference. In addition, since a single source must be postlexical, all neologistic errors should show a clear effect of length.

One and two-deficit accounts predict different distributions of target–error overlap. If there are two deficits, the amount of target–error overlap should be bimodally distributed. The mode

produced by lexical access failures should be near zero, with no more overlap than that expected by chance since the response will have been generated by a random process (Buckingham, 1981; Butterworth, 1979). The mode produced by target-related errors should have greater than chance overlap.¹ A single deficit, instead, should produce a distribution of overlap with a single mode. Abstruse errors would be from the lower tail of what is, in fact, a continuous distribution.

We present an analysis of a single jargonaphasic patient. To anticipate our conclusions, the pattern of correct and erroneous responses across tasks, the comparison between characteristics of target-related and abstruse errors, and the distribution of phonemes preserved in errors clearly show that her neologistic errors result from a single postlexical deficit. After the case history and our analysis of the number of deficits we turn to the more detailed question of where the single deficit is located in the postlexical system.

Case history

V.S. was an 84-year-old right-handed woman. She spent much of her working life at the Cadbury chocolate factory until she retired. She was living alone, following the death of her husband. In April 1996 she was admitted to University Hospital, Birmingham, following two days of confusion and headaches. She was described at the time as exhibiting a mixed expressive and receptive dysphasia. A computed tomography (CT) scan revealed an extensive periventricular low-density change, a low-density region of the left parietal lobe, and a possible recent infarct. She rapidly regained independence on the ward, although her speech remained impaired. She continued to live independently. She attended speech therapy sessions for approximately two years and was attending a stroke club once a week.

V.S. spoke fluently but with frequent neologisms and phonemic paraphasias, which are the focus of our study. At points, her speech was

disorganized to the extent that neither the structure of the sentences nor the individual target words could be identified. Generally, however, one could recover the gist of what she was trying to say, and it was possible to carry out a relatively normal conversation with her. A sample of her spontaneous speech is reported in Appendix A. V.S. produced lexical errors in addition to neologisms. They were phonologically and sometimes semantically related to the target (semantic errors were really only apparent in naming). Her phonological production was sometimes unclear (she mumbled or lost intelligibility), but her speech generally sounded like normal English with a normal prosodic contour. She did not produce strange phonemes, and phonotactic constraints were respected. Clinically, V.S. appeared to have good language comprehension. V.S. was first seen by our laboratory at the University of Birmingham in May 1998. The data for the experimental section of this study were collected during the summer of 2001. V.S.'s language performance showed no evident change over this period.

Language evaluation

Tests of short-term memory, auditory input processing, and comprehension

A preliminary formal assessment of V.S.'s language skills was carried out using predominantly tests from Psycholinguistic Assessments of Language Processing in Aphasia (PALPA; Kay, Lesser, & Coltheart, 1992).

Phonological processing

Results for short-term memory, phonological discrimination, and lexical decision are presented in Table 1. V.S. suffered from a marked short-term memory impairment, which determined not only her poor digit span, but also influenced some other tests of discrimination and comprehension. Phonological discrimination was severely impaired. As expected, nonword minimal pairs were worse than word minimal pairs. When the task involved

¹ Technically, the prediction is of a mixture distribution, and we test this hypothesis. The clearest type of mixture would be a bimodal distribution.

Table 1. *Phonological STM, auditory discrimination, and lexical decision results for V.S.*

<i>Task</i>	<i>V.S. – correct (%)</i>	<i>Controls^a (%)</i>
Digit span (PALPA 13) = 2		
Auditory discrimination		
<i>Words – minimal pairs (PALPA 2)</i>		
same pairs	75	99
different pairs	56	97
Total	65	98
<i>Non-words – minimal pairs (PALPA 1)</i>		
same pairs	58	
different pairs	39	
Total	49	
<i>Minimal pairs with pictures (PALPA 4)</i>	73	98
Auditory lexical decision		
<i>Imageability and frequency (PALPA 5)</i>		
Words	88	
high/high	100	
high/low	90	
low/high	85	
low/low	75	
Nonwords	65	
Total	76	
<i>Morphology (PALPA 6)</i>		
Words	53	
Nonwords	63	
Total	58	
Visual lexical decision		
<i>Imageability and frequency (PALPA 25)</i>		
Words	97	
high/high	100	
high/low	100	
low/high	100	
low/low	80	
Nonwords	55	
Total	76	
<i>Morphology (PALPA 6)</i>		
Words	93	
Nonwords	53	
Total	73	

Note: PALPA = Psycholinguistic Assessments of Language Processing in Aphasia.

^aWhere available.

a word and a picture, performance was better but not yet normal.

To examine the influence of both frequency and imageability in auditory lexical decision and to quantify possible interactions, we used log-linear analysis of the number of correct and incorrect

items (indicated by G^2). V.S. was significantly better at judging high- than low-imageability items ($G^2 = 4.37, p = .04$) and better with high- than low-frequency words, although this last difference did not reach significance ($G^2 = 1.87, p = .17$). In visual lexical decision V.S. was very

Table 2. *Semantics and comprehension results for V.S.*

<i>Task</i>	<i>V.S.–correct (%)</i>	<i>Controls^a (%)</i>
Semantics		
<i>Pyramids and palm trees</i>		
Words	96	
Pictures	96	
Total	96	>90
<i>Semantic associates with words (PALPA 51)</i>		
High imageability	80	89
Low imageability	40	82
Total	60	85
<i>Auditory synonym judgements (PALPA 49)</i>		
High imageability	57	
Low imageability	40	
Total	48	
<i>Written synonym judgements (PALPA 50)</i>		
High imageability	90	
Low imageability	67	
Total	78	
Auditory comprehension		
<i>Spoken word–picture matching (PALPA 47)</i>	95	98
<i>BPVS</i>	96 = 10 years, 7 months	
Written comprehension		
<i>Written word–picture matching (PALPA 48)</i>	95	99
Sentence–picture matching (PALPA 55 & 56)		
<i>Auditory</i>		
Reversible	30	
Non-reversible	75	
Total	52	
<i>Written</i>		
Reversible	35	
Non-reversible	69	
Total	48	

Note: PALPA = Psycholinguistic Assessments of Language Processing in Aphasia.

^aWhere available.

poor, but better than chance ($G^2 = 44.8$, $p < .001$). She systematically misjudged nonwords as words, and she also failed to categorize two low-frequency/low-imageability words correctly.

Semantic processing

Results from tasks assessing semantic processing are reported in Table 2. V.S. did relatively well

with high-imageability words. She scored within the normal range on the pyramids and palm trees tests. On a test of semantic association with written words (which word is closest in meaning to the word *piano*? Answer: *organ*) she performed well on high-imageability words and poorly on low-imageability words, $\chi^2(1) = 5.00$, $p < .05$. Auditory synonym judgements were not different

from chance due to poor auditory discrimination and phonological memory, but, again, with written synonyms performance was good with high-imageability words and poor with low-imageability words, $\chi^2(1) = 4.81, p < .05$.

Consistent with her good performance with high-imageability words, V.S. performed well on word–picture matching tasks from PALPA. Her performance was normal with both spoken words and written words. Her BPVS score, however, was somewhat low (95% confidence interval = 9 years, 11 months to 11 years, 3 months). Finally, V.S. had poor sentence comprehension, particularly for nonreversible sentences.

Preliminary tests of naming, reading, and repetition

Results for preliminary tests of naming, reading, and repetition were obtained from PALPA, supplemented by additional lists that we constructed. V.S. was poor at all tasks requiring spoken production. She correctly named only 45% (18/40) of pictures in PALPA 53. Results for reading and repetition are reported in Table 3. In both tasks, V.S. was more accurate with high- than with low-imageability words (reading: $G^2 = 18.9, p < .001$; repetition: $G^2 = 14.1, p < .001$). Trends toward better performance with high- than with low-frequency words were not significant (reading: $G^2 = 2.77, p = .10$; repetition: $G^2 = 2.43, p = .12$).

Reading was influenced by phoneme length, but not syllable length: phoneme length, $\chi^2(3) = 26.04, p < .001$; syllable length, $\chi^2(1) = 1.00, ns$. Repetition was not influenced by either: phoneme length, $\chi^2(3) = 2.14, ns$; syllable length, $\chi^2(1) = 0.34, ns$, but the lack of a phoneme length effect is probably due to a floor effect. We explore length effects further in the experimental section.

Both in reading and in repetition, there were effects of grammatical class with some word forms being read or repeated better than others, but these effects did not form a coherent pattern.

V.S. was very poor at both reading and repeating nonwords. Even repetition of CVC/CVVC (where C is consonant, V is vowel) nonsense syllables was not possible. This minimizes the contribution that nonlexical input could provide to improve noisy lexical outputs, as we see later.

Table 3. Preliminary results for V.S.'s reading and repetition

Task	Correct (%)	
	Reading	Repetition
<i>Imageability and frequency (PALPA 31 and PALPA 9)</i>		
High/high	30	55
High/low	48	30
Low/high	8	10
Low/low	13	5
<i>Syllable length (PALPA 30 and PALPA 7)</i>		
1 syllable	63	63
2 syllables	50	50
3 syllables	38	50
<i>Phoneme length</i>		
3–4 phonemes	38	8
5–6 phonemes	8	10
7–8 phonemes	8	5
9–10 phonemes	3	3
<i>Grammatical class (PALPA 10)</i>		
Nouns	0	40
Verbs	15	27
Adjectives	25	33
Functors	5	7
<i>Nonwords (PALPA 36 and PALPA 8)</i>	0	8
<i>CVC/CVVC syllables</i>		
Words		15
Nonwords		0

Note: PALPA = Psycholinguistic Assessments of Language Processing in Aphasia.

The neologistic errors that V.S. makes in naming, reading, and repetition are the focus of our experimental study.

EXPERIMENTAL INVESTIGATION

Materials

We gave V.S. 990 items for reading, naming, repetition, and repetition with a picture present. The items came from four sources:

Snodgrass and Vanderwart (1980) pictures

The Snodgrass and Vanderwart (1980) picture set consists of 260 black-and-white line drawings. Published control data, however, are from

American participants, and some items are either not familiar (e.g., *football helmet*), or are given different names by British participants (e.g., *pants/trousers*). New control data were collected from 10 students at the University of Birmingham, ranging in age from 18 to 34 years. The most common name provided by the students was considered the target in the patient study. Items were excluded if more than 2 students failed to identify the item. Items were also removed if the most common name was the same as that given to another picture in the set. This left 236 items for the current study.

Boston Naming Test (Goodglass, Kaplan, & Weintraub, 1983)

The Boston Naming Test (BNT) included 60 black-and-white line-drawn objects that were graded in naming difficulty. Pictures were selected to eliminate items with alternative acceptable names.

Birmingham Object Recognition Battery (Riddoch & Humphreys, 1993): Picture naming, long version

This test included 76 black-and-white line-drawn objects. Half were animate objects and half inanimate. There were structurally similar and structurally dissimilar sets of objects. Within each set, half of the object names were high frequency, and half were low frequency.

Difficult Naming Test

This test included 238 line-drawn objects, including items that are not commonly found in conventional naming tests (e.g., *pagoda, kaleidoscope*).

Therapy study pictures

These items were 380 colour pictures selected for a therapy study, which is not reported here. The list contained items with one to four syllables, but the majority of the words had one syllable.

The complete set of 990 items was presented to V.S. in sessions extending over several months. Items were presented in only one modality during any one session. No change was evident in V.S.'s performance over the duration of the study. All of the items except those of the therapy study ($N =$

610) were also given to V.S. for repetition with a picture present to gauge the influence of V.S.'s input problems on the types of error made.

Responses were tape-recorded and transcribed in broad International Phonetic Alphabet (IPA). Transcriptions were checked several times against the tape. Segments were generally clearly articulated. When more than one response was given, only the first complete response was included in the analysis. Utterances with no downturn in intonation that were included in the following response or altered by a single phoneme in the following response were considered fragments and were not scored as complete responses (e.g., *de . . . desk*).

Error classification

We initially separated *lexical errors* from *nonword errors* according to whether the result was a real word. Nonword errors were then further divided into *target-related* and *abstruse* neologisms by an arbitrary criterion. We classified errors that shared more than 50% of the target's phonemes (without regard to position) as target-related neologisms. The rest were classified as abstruse neologisms.

Lexical errors were subdivided into formal errors (*monkey* → *money*), semantic errors (*lettuce* → *potato*), mixed semantic/formal errors (*foot* → *boot*), unrelated lexical errors (*latch* → *shake*), visual errors (e.g., a pill being mistaken for a sweet), and other errors. Formal errors shared at least one phoneme in the same relative position in the syllable (*bat* → *ben*). This method of categorizing formal errors is likely to overestimate how often formal errors truly occur, but this was conservative in the current context where we are focusing on nonword errors. Unrelated lexical errors were words with no obvious relationship to the target. The category, *other lexical errors*, included morphological errors, perseverations, abbreviations, circumlocutions, and any error where the source was ambiguous (e.g., an error that was visually and semantically related to the target). The category, *other errors*, distinct from lexical and nonword errors, included targets that elicited no response, circumlocutions, and errors that contained both lexical and nonword parts (e.g., *suitcase* → *seat/blamə/*).

We are interested in localizing the source of V.S.'s abstruse and target-related nonword errors. We do not claim that this is the only language process that has been affected by her stroke. Her documented input problems and the presence of semantic errors show that her overall profile is affected by other deficits. These other deficits, however, do not prevent us from localizing her nonword errors.

Lexical access versus postaccess deficit

Error types

A postlexical deficit predicts essentially equivalent performance in reading, repetition, and naming. If problems with lexical access contribute to the pattern, reading and repetition could be better than naming, and the types of error should be different, with more circumlocutions and no response errors in naming.

Table 4 reports the different types of errors made by V.S. across tasks. There were some minor differences. Repetition produced the most errors and the largest proportion of phonological lexical errors. This is not surprising given V.S.'s

documented input problem. In line with this explanation, the number of errors and the number of phonological lexical errors decreased when a picture was present during repetition.

V.S. made more semantic and visual errors when naming pictures. This is also not surprising. A picture that does not elicit the appropriate word may elicit a semantically or visually related name. Across all tasks, however, nonword errors were more frequent than lexical errors, abstruse neologisms were common, and no responses and ambiguous/mixed errors were rare (justifying the clinical categorization *jargonaphasic*).

Discussion. Despite some differences due to task, the overall similarity of the pattern across tasks is clear. This pattern demonstrates, first, that V.S. has a speech production problem that is largely independent of the type of input used to elicit a response. In addition, the similarity of errors across tasks is a first indication that most of her errors are due to a post-access impairment.

These results by themselves, however, are not strong enough to exclude another source for

Table 4. *Error types in reading, naming, repetition, and repetition with a picture present*

<i>Error type</i>	<i>Naming (N = 990)</i>		<i>Reading (N = 990)</i>		<i>Repetition (N = 990)</i>		<i>Repetition with picture (N = 610)</i>	
	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>
<i>Nonword errors</i>								
Related neologisms	150	22	259	44	293	37	139	37
Abstruse neologisms	177	26	124	21	218	28	122	33
Total	327	48	383	65	511	65	261	70
<i>Lexical Errors</i>								
Phonological	86	13	171	29	218	28	64	17
Semantic	65	9	3	1	7	1	7	2
Semantic/phonological	3	0	4	1	1	0	9	2
Visual errors	15	2	2	0	0	0	0	0
Unrelated	85	12	12	2	29	4	19	5
Other	61	9	5	1	6	1	4	1
Total	315	46	197	33	261	33	103	28
<i>Other</i>								
Amb/mixed	17	2	9	2	10	1	7	2
No response	18	3	0	0	0	0	1	0
Circumlocution	9	1	0	0	1	0	0	0
Grand total	686		589		783		372	

neologistic errors. On the one hand, there were more abstruse responses in naming than in reading and repetition. This is consistent with a second deficit at lexical access, but a one-deficit account could explain these errors by saying that some visual or semantic errors (which do occur independently) were further distorted by the single deficit that produced the neologisms, rendering them abstruse when compared to the original target. At the same time, we have noted that patterns in naming, reading, and repetition were broadly similar, consistent with a single deficit. However, V.S. was also poor in reading and repeating nonwords. If reading and repetition do not benefit from nonlexical input, they would be expected to be similar to naming even if there were two deficits, so, again, the result is not definitive. The distribution of error types contributes to the overall picture we develop, but it cannot be used to differentiate a two-deficit from a one-deficit locus.

If V.S.'s abstruse neologisms were produced when lexical access fails, they should be influenced by word frequency and imageability but not length. A single deficit, instead, predicts effects of length even for abstruse errors. Effects of frequency, imageability and length could be somewhat stronger for abstruse neologisms, since, by definition, these are the errors that are more severely affected by the deficit, but abstruse and target-related errors should be influenced in similar ways by these factors.

Effects of imageability, frequency, and length

Ratings were taken from the MRC Psycholinguistic Database (Wilson, 1988). Average values for the targets of correct responses and target-related and abstruse neologisms are reported in Table 5. The targets generating correct responses were more imageable, of higher frequency, and shorter than targets generating errors. The effect of imageability approached significance in the three-factor one-way analysis of variance (ANOVA) for naming and reading ($p = .07$ and $.08$) and was significant for repetition ($p = .001$). Targets of related and abstruse neologisms were generally not different in post hoc tests of each variable (all Tukey's HSD

$p > .49$). The only exception was that targets of related neologisms were shorter than targets of abstruse neologisms in reading (Tukey's HSD $p < .001$).

Imageability, frequency, and length are inter-related variables. To judge their independent contribution, we entered the factors in a simultaneous logistic regression predicting correct and incorrect responses.

In naming, log frequency and imageability were clearly significant (both $p = .003$), and phoneme length was nearly significant ($p = .062$). In reading, all three factors were clearly significant (log frequency and phoneme length, $p < .001$, and imageability, $p = .003$). In repetition, log frequency and imageability were clearly significant ($p < .001$), while phoneme length was not ($p = .58$).

Discussion. Effects of imageability and frequency can be predicted by both single-deficit and two-deficit accounts. A single deficit only predicts these effects, however, if the deficit follows the lexicon closely enough that lexical variables can still be influential at the level that is affected (e.g., through cascade or interactive processing).

Effects of length, instead, are only consistent with a post-access deficit. Crucially, we found that effects of length were not different for abstruse and target-related errors in two out of three tasks. In all three tasks, moreover, abstruse neologisms were affected by length—that is, they were produced to targets that were longer than those eliciting correct responses. These results are contrary to the hypothesis that the abstruse neologisms are the result of lexical access failures.

Word-initial phonemes

Production of word fragments, or positional effects, more generally, are not consistent with a complete failure to access a lexical representation. However, it has been noticed that normal speakers in a tip-of-the-tongue state (assumed to be due to failure to access phonological information after successful lemma access) can sometimes have access to a word's initial phoneme (Brown & McNeill, 1966; Caramazza & Miozzo, 1997;

Table 5. Imageability, frequency and length of targets for correct responses and target-related and abstract neologisms

	Imageability	N	95% confidence interval	Log frequency	N	95% confidence interval	Length	N	95% confidence interval
<i>Naming</i>									
correct	593	260	588–599	1.33	286	1.26–1.40	3.83	304	3.64–4.02
target-related neologisms	582	98	575–590	0.87	127	0.76–0.98	5.08	150	4.81–5.35
abstruse neologisms	589	106	581–596	0.82	145	0.71–0.92	5.45	177	5.20–5.70
ANOVA	$F(2, 461) = 2.7, MSE = 1,531, p = .07$			$F(2, 555) = 42.3, MSE = 0.39, p < .001$			$F(2, 628) = 60.2, MSE = 2.84, p < .001$		
<i>Reading</i>									
correct	592	341	588–596	1.30	377	1.24–1.36	3.79	401	3.63–3.95
target-related neologisms	587	160	581–592	0.87	220	0.78–0.95	5.31	259	5.11–5.50
abstruse neologisms	581	61	572–591	0.72	98	0.59–0.84	6.19	124	5.91–6.47
ANOVA	$F(2, 559) = 2.5, MSE = 3,540, p = .08$			$F(2, 692) = 51.1, MSE = 0.4, p < .001$			$F(2, 781) = 138.7, MSE = 2.5, p < .001$		
<i>Repetition</i>									
correct	598	173	593–603	1.28	198	1.19–1.37	4.17	206	3.92–4.43
target-related neologisms	585	203	580–590	0.95	255	0.87–1.03	4.88	293	4.66–5.09
abstruse neologisms	586	136	580–592	0.97	182	0.87–1.06	5.11	218	4.86–5.36
ANOVA	$F(2, 509) = 6.9, MSE = 1,249, p = .001$			$F(2,632) = 6.9, MSE = 0.4, p < .001$			$F(2, 714) = 14.5, MSE = 3.5, p < .001$		

Harley & Bown, 1998). If V.S.'s abstruse neologisms arise from a deficit at the level of lexical access, they might preserve initial phonemes more often than other phonemes in the word.

If initial phonemes have special early availability, they should not be preserved more often than other positions just because they are first, and performance declines across word position (e.g., see the decreasing accuracy function reported by Ward & Romani, 1998). Instead, preserved initial phonemes should literally be out of line with performance on other positions. To evaluate this, we fit a regression line to Positions 2–7 (where there were substantial numbers of exemplars) and tested whether the first position was outside the range predicted by the residuals from the other positions. The prediction is that initial phonemes may be preserved even when lexical access fails, and, therefore, it only applies to abstruse errors. For completeness and comparison, however, we present results from both types of neologism.

Results are presented in Figure 1. The initial position was significantly preserved in only one case: target-related errors in naming ($z = 4.27, p < .001$). In all other cases the initial position was not out of line with the other positions, or was out of line in the wrong direction (naming: abstruse neologisms, $z = .95, p = .34$; reading: target-related neologisms, $z = 0.76, p = .44$; abstruse neologisms, $z = 1.57, p = .12$; repetition: target-related neologisms, $z = 0.63, p = .53$; abstruse neologisms, $z = 3.31, p = .001$, initial position less often preserved).

Discussion. There was no evidence that V.S. systematically preserved initial phonemes. The only result showing preservation involved target-related errors where there is clear evidence of access to other segmental information. The pattern is certainly not consistent with abstruse errors having an earlier source than target-related errors.

Distribution of preserved phonological information in errors

When we made an initial categorization of errors we arbitrarily defined abstruse errors as those that included less than 50% of the target phonemes.

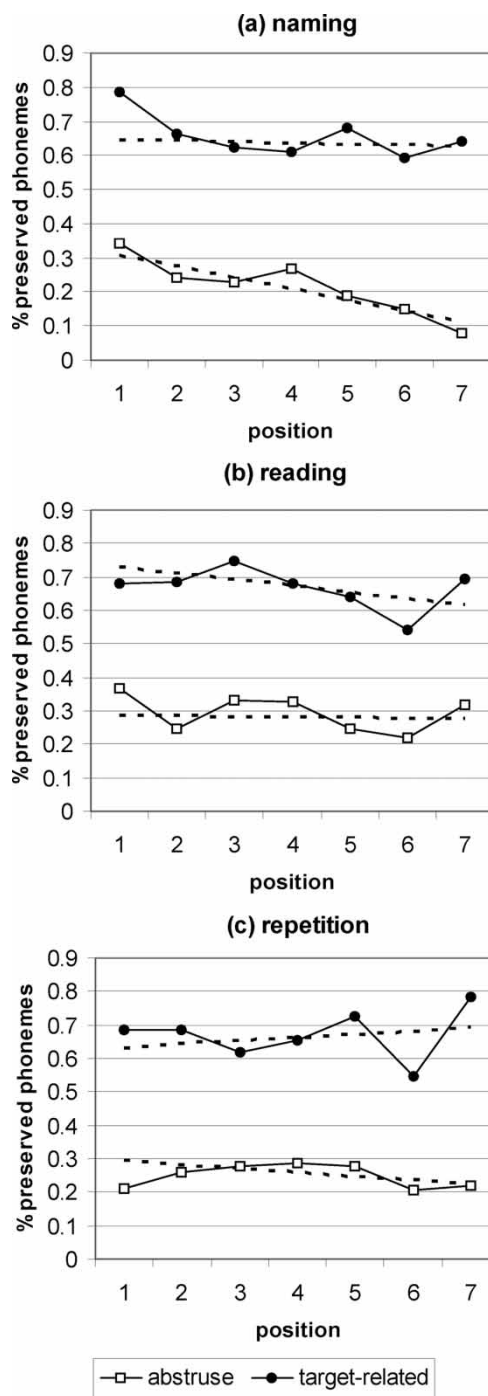


Figure 1. The percentage of phonemes in targets that are preserved in responses according to target position. Dotted lines are regression fits to Positions 2–7.

Here we evaluate whether there is, indeed, a bimodal (mixture) distribution of target/error overlap that justifies the division of errors into separate types. One mode would be provided by cases where lexical access succeeds, and target information is clearly present, and another mode by cases where lexical access fails, and target information is only present at chance levels. Buckingham and Kertesz (1976), in fact, report a bimodal distribution of target/error overlap for a Wernicke's aphasic. A unimodal distribution, with no clear division between target-related and abstruse errors, instead, would be consistent with a single impairment that affects some words more severely than others.

We modelled the observed percentage of phoneme overlap using either a single normal distribution (reflecting a single deficit affecting some words more severely than others), or a mixture of a normal distribution and a distribution based on chance overlap (reflecting lexical access failures). To evaluate what the level of chance overlap was, we paired each neologistic error with all targets of neologisms except its own. Target/error overlap was always the number of response phonemes that also appeared in the target (or rematched target) as a proportion of the total number of target phonemes.

Since the targets make some degrees of overlap more likely than others (e.g., if all words were three phonemes long, only overlaps of 0, .33, .67, and 1.0 would be possible), the model distributions had to be adjusted for the possible levels of overlap offered by the stimuli. This is why the fit based on the normal distribution does not have the expected Gaussian shape (Figure 2).

Fits were calculated by maximum likelihood (fitting details given in Appendix B), and models were compared according to the Akaike Information Criterion (AIC; Akaike, 1973; Burnham & Anderson, 2002). Since the one-deficit model is a more restricted version of the two-deficit model (where there is no contribution of lexical-access failure), the two-deficit version is guaranteed to provide at least a marginally better fit. However, the AIC measure balances fit and parsimony by rewarding fit and penalizing the

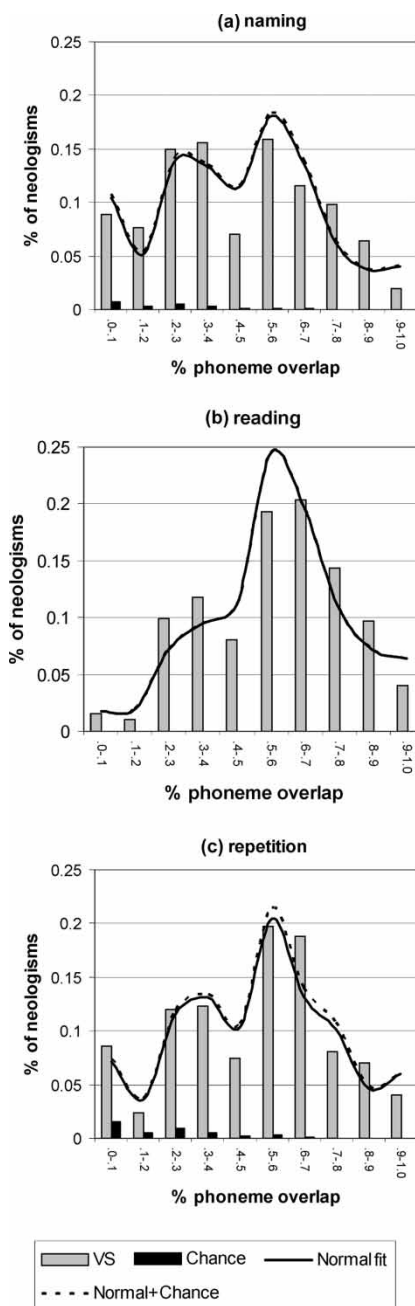


Figure 2. The distribution of phonemes in targets also present in V.S.'s errors for naming, reading, and repetition. Fits to a normal distribution, and the best mixture of chance and normal distributions are shown. The size of the displayed chance distribution (black bars) is the size of the contribution of the chance distribution in the best chance + normal fit.

extra parameters that the more complex model includes. Thus, a simpler model may be preferred, even if its fit is numerically lower. The model with the smaller AIC is preferred, and models within 2 AIC units of each other are considered roughly equivalent (Burnham & Anderson, 2002). We always report the difference from the model with the smallest AIC value (labelled Δ AIC, Burnham & Anderson, 2002), since the absolute size of AIC values is not informative.

Results are presented in Figure 2. A normal distribution always provided a very good fit to V.S.'s distribution of phoneme overlap, and the predicted and observed distributions were never significantly different ($-2 \ln \log$ likelihood ratio naming = 0.09, reading = 0.07, repetition = 0.06, all $p > .99$; see Cowan, 1998, for testing fits using log likelihood). The mixture of chance and normal distributions never provided a better model according to the AIC (i.e., the single distribution always had the lowest AIC, and differences for the mixture distribution were: naming, Δ AIC = 2; reading, Δ AIC = 2; repetition, Δ AIC = 2). In addition, even in the models with a second deficit, the contribution of a lexical deficit (the contribution of the chance distribution to the mixture distribution) was always very low (naming, 2%; reading, <0.1%; repetition, 4%).

Discussion. These results show that our criterion for dividing target-related and abstruse errors arbitrarily divides what is, in fact, a continuous distribution of target/error overlap. What looks like some degree of bimodality in V.S.'s raw distribution of phonemes (see the range 0.2–0.4 in Figure 2), is entirely the product of the opportunities for overlap presented by the stimuli. There was no evidence that failures of lexical access contributed substantially. The results indicate that V.S.'s neologistic errors arise from a single post-access deficit that affects some words more severely than others.

This does not mean that lexical access failure is never a factor in V.S.'s errors. We know, for example, that V.S. makes lexical errors. It is very likely, given the locus that we have identified, that some lexical errors will be further distorted

postlexically to produce abstruse neologisms. These certainly are failures to access the correct lexical representation, and we know they occur. What our analyses show, however, is that failures of lexical access do not account for a substantial number of V.S.'s abstruse neologisms.

Similarly, our results do not mean that all of V.S.'s language performance can be explained by a single deficit. She has input problems, and we have noted that she makes some lexical errors that cannot be explained by the locus responsible for the neologistic errors. It is certain that other deficits would be required to explain all of her language performance. What we have shown is that additional deficits are not required to explain her neologistic errors (i.e., her jargonaphasia).

There is a contrast between our results and Buckingham and Kertesz's (1976) report of a bimodal distribution of target–error overlap. Schwartz, Wilshire, Gagnon, and Polansky (2004), who also report a unimodal distribution for errors from a group of patients, suggest that the contrast with Buckingham and Kertesz (1976) has a methodological source. They note that targets in spontaneous speech are not always apparent. More errors may be called abstruse because the experimenter was not successful in identifying the target. Studies using tasks where the target is clear (like ours and hers) will report fewer abstruse errors.

While task differences could be a source of differences between our results and those of Buckingham and Kertesz (1976), it is also true that the two cases could have involved different numbers of deficits. Identifying a single source in one patient does not preclude a second source in other cases. We return to this issue in the General Discussion.

Interim discussion

At this point our results clearly favour a single post-access deficit of phonological encoding and not two deficits. Error types were similar across reading, repetition, and naming. Abstruse and target-related errors were influenced by the same variables: imageability and frequency, and, crucially, length. Abstruse errors did not preserve initial phonemes as tip-of-the-tongue states

might have predicted. Finally, the overall distribution of phoneme overlap was fitted very closely by a single normal distribution, consistent with a single deficit at the level of phonological encoding.

Having established that V.S.'s neologistic speech errors are largely produced by a single deficit beyond the level where words have unitary representations, however, does not locate it precisely in the speech production system. The deficit could be at the level of phonological encoding that immediately follows lexical access, it could be at the level of the phonological buffer, or it could be at the level of articulatory encoding. We have already mentioned that the combined effects of frequency, imageability, and length suggest a locus where lexical variables still have some influence. In the following analyses, however, we investigate, in more detail, possible post-access loci.

Postlexical loci

Compared to the broad architectural distinctions referenced above, there is somewhat less agreement about the levels of representation involved in postlexical processing and about the characteristics of those levels. Our discussion, therefore, is more speculative when addressing postlexical organization. The strategy that we adopt is to consider predictions made by various proposals, which do not always agree or cross-reference one another. We can show clearly which aspects of V.S.'s results are consistent with each of these proposals. The full set of results converges on a probable specific locus for her deficit.

We examine two general dimensions of distinction in postlexical processing. First, we distinguish memory representations from representations that are not involved in short-term retention. Second, we compare abstract phonological representations and representations used to guide articulatory programming.

Memory representations: The phonological buffer

Models of speech processing derived from neuropsychological cases have proposed that there is a short-term store, the phonological buffer, that keeps phonemes from the lexicon or from

nonlexical conversion active while motor planning for articulation takes place. Deficits to the phonological buffer have been reported for patients with relatively mild phonological output problems (Bub et al., 1987; Caramazza et al., 1986; Shallice et al., 2000), but there is no reason why a more severe deficit with the same characteristics could not lead to jargonaphasia.

The errors reported for buffer patients have had several characteristics. Buffer errors were affected by frequency and length, matching V.S.'s pattern. Substitution errors involved phonologically related segments. We examine the relatedness of substitutions errors below. Buffer patients have shown no influence of imageability. V.S. was influenced by imageability, but imageability also strongly affected V.S.'s performance in lexical decision and semantic tasks. The deficit necessary to explain these tasks may contribute to the pattern in naming, reading, and repetition and account for this particular discrepancy. In this section we examine the most characteristic aspect of a retention deficit: the connection between error rate and word length (Caramazza et al., 1986).

It is uncontroversial that information in a short-term memory component should be subject to decay over time (e.g., Baddeley, 1990). The probability that a segment is correct should depend on the amount of time it has spent in the buffer. As a result, long words should be incorrect more often than short words. A simple effect of length, however, does not require decay. For example, if every fifth phoneme produced an error, this would be a constant rate of error not subject to decay. Three-phoneme words would often be correct, but no six phoneme words would be correct. Instead, decay is only implicated when the rate of error per phoneme increases with length.

Reports of buffer patients have not usually tested whether the per-phoneme probability of error increases for longer words, and a cursory inspection of the published results does not allow this to be decided post hoc. This is clearly an important prediction, however, which is diagnostic of a memory component.

We used two methods to compare models with and without a changing probability of phoneme

recall. The first method derived the probability that each phoneme was correct from the percentage of words of different lengths that were completely correct. The second method used both completely correct responses and errors. In both cases, we compared a model where the probability of recall per phoneme was constant with a model where this probability could change as a linear function of word length.

Measuring the influence of length using the number of words completely correct (Method 1) can be rather insensitive when errors have a wide range of overlap with targets. Errors with a single substitution are lumped together with errors that have nothing in common with the target. For our second method, we counted the number of target phonemes that were preserved in responses as a function of word length. Phonemes could be preserved anywhere in responses (not necessarily in the correct position). We called this $p(\text{inclusion})$. This probability was based both on words that were completely correct and on errors. We fitted the observed data, comparing a constant $p(\text{inclusion})$ with a $p(\text{inclusion})$ that varied with word length.

Probability correct as a function of word length. For a word to be correct, all of its phonemes must be correct, so the probability of a word being correct is $p(\text{phoneme correct})^L$, where L is the number of phonemes, and we assume that the probability for each phoneme is independent of the others. The proportion correct that we fitted was based on the total number of correct responses out of correct responses and neologisms. We excluded errors that did not clearly arise at the segmental level, since these errors may not be sensitive to length (e.g., semantic errors). We used maximum likelihood to compare a model with a constant probability correct per segment to a model with a linear change in $p(\text{correct})$ —that is, $p(\text{word correct}) = p(\text{phoneme correct})^L$, and $p(\text{phoneme correct}) = \text{constant} + \text{slope} * L$.

Results are shown in Figure 3. There were different patterns in different tasks. In naming, correct/incorrect responses were modelled well

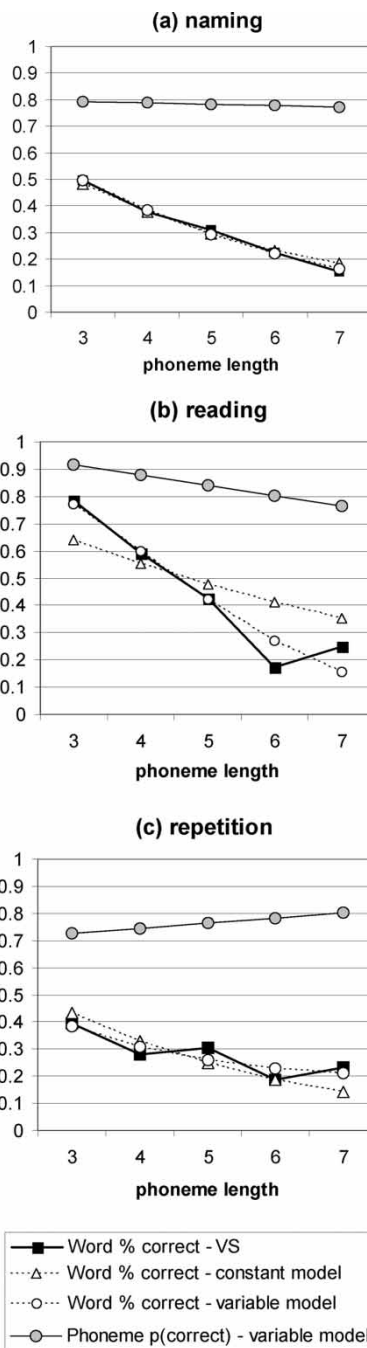


Figure 3. The observed and modelled percentage of words correct for words of different lengths based on models with a constant and variable probability that each phoneme is correct. Grey circles represent the fitted probability that each phoneme is correct in the variable model.

by a constant probability correct per phoneme—the best model was constant $p(\text{correct}) = .78$; $p(\text{correct})$ for the variable model at length 3 = .79; decrease for each phoneme = .004, $\Delta\text{AIC} = 2$. In reading, instead, there was clear evidence that the probability that each phoneme would be correct decreased with longer words: $p(\text{correct})$ for variable model at length 3 = .92, decrease for each phoneme = .04; constant $p(\text{correct}) = .86$; $\Delta\text{AIC} = 37$. In repetition, the model with a variable $p(\text{correct})$ was slightly better, but the probability increased with word length rather than decreasing: $p(\text{correct})$ for variable model at length 3 = .72, increase for each phoneme = .02; constant $p(\text{correct}) = .76$; $\Delta\text{AIC} = 3$.

Probability of inclusion as a function of word length. Our $p(\text{inclusion})$ counted the number of target phonemes that appeared anywhere in errors as a function of the number of target phonemes. We analysed data from words with three to seven phonemes, where there were sufficient numbers for a reasonable estimate.

Results based on phoneme overlap also favoured a constant model for naming and repetition: naming: constant $p(\text{inclusion}) = .44$; starting value for variable $p(\text{inclusion})$, that is, $p(\text{inclusion})$ at length 3, = .44, change in $p(\text{inclusion})$ per phoneme = .004, $\Delta\text{AIC} = 2$; repetition: constant $p(\text{inclusion}) = .50$; starting value for variable $p(\text{inclusion}) = .49$, change in $p(\text{inclusion})$ per phoneme = .007, $\Delta\text{AIC} = 1$. For reading, instead, a model where $p(\text{inclusion})$ decreased with increasing length was better: starting value for variable $p(\text{inclusion}) = .62$, change in $p(\text{inclusion})$ per phoneme = -.02; constant $p(\text{inclusion}) = .56$; $\Delta\text{AIC} = 3$. For reading, the estimated $p(\text{inclusion})$ for 3-phoneme words was .62, and for 7-phoneme words it was .52.

Discussion. Reading produced the only results consistent with a deficit in a short-term memory component since the probability that target phonemes were preserved in responses changed with word length. Length did not change the $p(\text{inclusion})$ in repetition or naming. These differences across tasks are not consistent with a deficit at the level of the phonological buffer, which should affect

all tasks. In addition, a buffer account does not highlight reading as the task where memory effects should be strongest. Reading is the only task where a segmental representation, in the form of letters, is present throughout.

Central versus peripheral phonological representations

Most linguistic and psycholinguistic theories of phonological processing invoke more than one level of representation between the lexicon and the commands to the articulators, with an abstract phonological representation being translated into a representation that can guide the separate components of the vocal tract (glottis, tongue, velum, etc.). The distinction between underlying and surface representations (e.g., Chomsky & Halle, 1968), between representations based on abstract segments and those based on features (e.g., Roelofs, 1999), between representations that are phonological and those that are phonetic (Anderson, 1981), and between representations that are underspecified and fully specified (Anderson & Lightfoot D.W., 2002) share this general flavour. Keeping to areas of agreement, a full featural representation is generally thought to be an appropriate precursor to motor planning, since many features are directly related to the independent parts of the vocal tract that are manipulated to produce speech—for example, (+voice), (+nasal), (+alveolar), and so on. Beyond this, agreement over details is scarce. There is little consensus over how information is organized and distributed between levels, or even over how levels should be described and distinguished (e.g., can/how should phonological and phonetic levels be distinguished? e.g., Anderson, 1981; Anderson & Lightfoot, 2002; Ohala, 1990; are there underlying and surface representations? e.g., Burzio, 2002; Chomsky & Halle, 1968).

One distinction that V.S. can address is between accounts that include an abstract segmental level, where phonemes are unitary (Dell, 1986; Goldrick & Rapp, 2007; Levelt, 1989; Roelofs, 1994, 1999; Shattuck-Hufnagel & Klatt, 1979; Stemberger, 1982) and those where segments are always represented in terms of features. Roelofs (1999) specifically motivates a level where unitary

phonemes do not have featural content using evidence from speech errors and priming. He notes that the majority of speech errors involve anticipation, perseveration, exchange, shift, addition, or deletion of single segments, but that unambiguous feature movement errors are rare (clear blue sky → glear plue sky; citing evidence from Shattuck-Hufnagel & Klatt, 1979). Specifically, movement of single features is less common than movement errors involving segments that differ by more than one feature. In addition, movement errors are more common when the phoneme that moves ends up next to the same phoneme in the error as in the intended phrase (e.g., *bit pin* → *pit pin*, where /p/ is next to /i/ in the target and in the error, is more likely than *bet pin* → *pet pin*, where /p/ is next to /i/ in the target, but /e/ in the error). To produce this effect the context must be identical. Phonemes that share features but are not identical do not increase movement errors. So, it seems that speech errors manipulate and respond to segments, but they do not manipulate features independent of segments.

Similarly, in a set of form preparation experiments, Roelofs (1999) found that speech latencies were only speeded when items that could be prepared ahead of time shared whole segments, not when they shared everything except a single feature. Identity allowed advance preparation, but nearly complete feature overlap did not. Roelofs takes his results together with those from speech errors to support a model where there is a level where phonemes are abstract units without featural content.

Certainly Roelofs's results require that phonemes be completely specified before they speed up preparation times. Similarly, speech errors suggest that features cannot move freely during speech errors. Neither of these results, however, requires that features be absent from the segmental representation. We return to this issue in the General Discussion. Despite some ambiguity about results that have been used to argue for unitary segments, the theoretical account itself makes some clear predictions. A level with unitary segments predicts that some patients will freely

substitute phonemes without regard to features. According to this account, if V.S. makes phonologically related substitutions, she must have a deficit at a more peripheral level. Alternatively, phonemes may always have featural content. Then related errors could arise at one of several levels.

Phonological relatedness of substitution errors

To see whether the substitution errors that V.S. made were phonologically related, we identified unambiguous substitution errors using an algorithm for calculating string edit distance (or Levenshtein distance, see Gusfield, 1997) that was modified by the first author to include transposition errors. The algorithm identifies equivalent sets of transformations that turn a target into its neologistic error—for example, *spider* → *spoon* could be the result of (a) substituting *o* for *i*, substituting *n* for *d*, and deleting *er*, or (b) deleting *id*, substituting *o* for *e*, and substituting *n* for *r*, among other possibilities. We used only substitutions that were formally unambiguous according to Levenshtein distance and constructed a confusion matrix for each task.

The confusion matrices are presented in Figure 4. In all three tasks, consonant and vowel substitutions showed a strong tendency to remain within class. In reading and naming, consonant substitutions also usually preserved the manner of articulation. In repetition, this tendency was weaker.

Discussion. Substitution errors involved segments related by features. We did not find evidence for a level of unitary segments. Either V.S. makes errors at a peripheral level where features are represented, or segments are never unitary. This one result does not allow us to decide the issue, so we turn now to other predictions associated with more peripheral levels.

Articulatory programming

There is no reason to exclude, a priori, that neologisms could arise during articulatory planning in addition to phonological encoding. Given her pattern of substitution errors, V.S. could have trouble at a stage where phonemes are converted

(a) naming

		responses			
		vowel	liquid+nasal	stop	fricatives
targets	vowel	76	8	9	3
	liquid+nasal	16	21	4	4
	stop	6	8	25	7
	fricatives	1	3	6	9

(b) reading

		responses			
		vowel	liquid+nasal	stop	fricatives
targets	vowel	50	2	0	0
	liquid+nasal	1	13	5	2
	stop	1	11	26	14
	fricatives	0	3	4	7

(c) repetition

		responses			
		vowel	liquid+nasal	stop	fricatives
targets	vowel	59	6	3	0
	liquid+nasal	2	21	4	3
	stop	1	38	32	19
	fricatives	1	13	17	2

Figure 4. Confusion matrices for unambiguous substitution errors (consonant area of the matrix is grey).

into articulatory programmes rather than the stage where phonemes are selected based on a word form. The effects of word frequency and imageability and V.S.'s fluent speech and well-formed phonemes argue against this, but an articulatory deficit could contribute to V.S.'s errors even if frequency and imageability effects had a different source, and phonetic errors are notoriously difficult to detect without instrumental analyses. Accounts of articulatory programming, however, make additional predictions for articulatory-level deficits.

Syllable frequency

Levelt and Wheeldon (1994) claim that articulation is mediated by a library of syllable-sized

articulatory programmes (see also Aichert & Ziegler, 2004). This syllabary stores precomputed programmes for frequently used syllables to reduce the computational load required for fluent speech. In addition, syllables are good units to use for an articulatory store because most language-specific phonetic variation occurs within syllables. Most phonetic variation would be precomputed, rather than computed on-line (Levelt et al., 1999).

A frequency-based syllabary predicts that errors will concentrate on words with lower frequency syllables. Looking at where errors actually occur, the frequency of syllables produced incorrectly should be lower than the frequency of syllables

produced correctly. Aichert and Ziegler (2004), in fact, found that syllable frequency did affect error rates in patients with apraxia of speech. When errors occur, high-frequency syllables that are still available in the syllabary should substitute for lower frequency syllables from the targets.

To examine syllable frequency we used the CELEX corpus (Baayen, Piepenbrock, & Van Rijn, 1995). We counted syllables according to the number of times they were used, not just the number of words they occurred in, and we used log frequency for comparisons. Syllables that did not occur were given a log frequency of 0.

Table 6 reports syllable frequencies for syllables in correctly produced words and in the target-related and abstruse errors. We also report the frequencies of syllables in the targets of words that resulted in errors (labelled *targets of errors*).

Contrary to expectations, target syllables leading to correct responses were less frequent than target syllables leading to errors: naming, $F(2, 1106) = 3.63, p = .03$; reading, $F(2, 1389) = 11.37, p < .001$. The exception was in repetition, where syllable frequencies did not differ: repetition, $F(2, 1324) = 1.94, p = .14$.

Also contrary to expectations, syllables from neologisms were always of lower frequency than syllables from their targets: naming: target-related, $t(559) = 3.02, p = .003$; abstruse, $t(615) = 4.83, p < .001$; reading: target-related, $t(957) = 3.34, p = .001$; abstruse, $t(511) = 2.31, p = .02$;

repetition: target-related errors, $t(982) = -5.97, p < .001$; abstruse errors, $t(693) = -5.46, p < .001$.

V.S. produced a substantial number of syllables that did not occur in the CELEX corpus. Even without these included, however, the syllable frequency of targets and errors never differed: naming: target-related, $t(572) = -0.79, p = .43$; abstruse, $t(634) = -0.16, p = .88$; reading: target-related, $t(1007) = -1.44, p = .15$; abstruse, $t(569) = -1.32, p = .19$; repetition: target-related, $t(1006) = 0.70, p = .49$; abstruse errors, $t(694) = -0.71, p = .48$ (note that degrees of freedom in the t tests are corrected for unequal variances, so that degrees of freedom in the tests excluding non-existent syllables can be greater than those in tests including them).

Discussion. V.S. produced errors that included a number of syllables that did not occur in CELEX, and she was not especially poor with low-frequency syllables. Her pattern was not consistent with a partially damaged syllabary. Since syllables from errors were often less frequent than syllables from targets, syllable frequency was clearly not a determining factor in her errors.

If a syllabary were eliminated, rather than just damaged, no effect of syllable frequency would be predicted, since all syllables would need to be computed. Loss of the syllabary should affect the fluency of speech, however, since the syllabary is motivated by the efficiency of using stored rather

Table 6. *The frequency of target syllables producing correct responses and target-related and abstruse neologisms, along with the frequency of the syllables produced in errors*

	Naming	Reading	Repetition
<i>Correct</i>	3.57	3.49	3.7
<i>Target-related neologisms</i>			
Target	3.70	3.74	3.77
Errors: occurring syllables only	3.78	3.84	3.72
Errors: including nonoccurring syllables	3.36	3.48	3.29
<i>Abstruse neologisms</i>			
Target	3.76	3.77	3.65
Errors – occurring syllables only	3.74	3.90	3.71
Errors – including non-occurring syllables	3.25	3.52	3.11

than computed structures (Levelt et al., 1999). V.S.'s fluent speech was not consistent with complete loss of a syllabary whose function is to make speaking more efficient.

Syllable complexity

Studies of fluent and nonfluent Italian aphasics have shown that a clear pattern of syllabic simplifications accompanies deficits of articulatory planning and not more central phonological deficits (Romani & Calabrese, 1998; Romani & Galluzzi, 2005; Romani, Olson, Semenza, & Grana, 2002). If V.S. made errors that systematically simplified syllable structures, this would suggest a deficit of articulatory planning.

Following Romani and Galluzzi (2005), we measured syllabic complexity using both the complexity of syllabic templates and the sonority of segments that make up syllables. The best template is universally considered to be CV (Clements, 1990; Goldsmith, 1990; Romani & Calabrese, 1998). We considered any single change to this template to be a complication.

Syllable templates V, CVC, and CCV were considered more complex than CV. Syllable templates VC, CCVC, and CVCC were considered even more complex.

Within each syllable template, syllables differ in complexity according to the relative sonority of their segments. According to Clements (1990), the simplest syllable has the biggest sonority contrast in onset and the smallest contrast in coda. Vowels have the highest sonority, followed by glides, liquids, nasals, and obstruents. We used these principles to score errors within each template type. For example, if a nasal was substituted for an obstruent in a CVC syllable onset, an increase in complexity was scored. If a nasal was substituted for an obstruent in a CVC syllable coda, a decrease was scored.

Only neologisms preserving syllable length were scored, since corresponding syllables in targets and errors could be identified. Results are shown in Table 7. We subdivided results according to task, type of neologism, and whether or not the syllable template was altered. In each case we asked

Table 7. *Percentage of syllables from errors that increased, decreased, or did not change syllable complexity when compared to their targets*

<i>Task</i>	<i>Syllable template</i>	<i>N</i>	<i>Change in complexity</i>		
			<i>Increase (%)</i>	<i>Equal (%)</i>	<i>Decrease (%)</i>
<i>Naming</i>					
Related	Unchanged	100	11	82	7
	Changed	80	36	18	46
Abstruse	Unchanged	40	13	75	13
	Changed	125	43	18	38
<i>Reading</i>					
Related	Unchanged	206	7	81	12
	Changed	146	28	27	45
Abstruse	Unchanged	69	22	62	16
	Changed	149	32	20	48
<i>Repetition</i>					
Related	Unchanged	213	16	69	15
	Changed	187	50	17	34
Abstruse	Unchanged	84	24	48	29
	Changed	197	45	17	38

whether syllable complexity changed systematically in one direction (i.e., simplified or complicated syllable structure). Statistical significance was assessed using log-linear analysis. Errors that did not either increase or decrease complexity were not included in the analysis.

In an overall analysis that included task as a factor, there was a significant interaction between task and direction of change ($G^2 = 4.2, p = .04$). Different tasks produced different patterns of simplification/complication, so collapsing across tasks could not be justified. As a result, we analysed data from each task separately. In naming, there were an equal number of complications and simplifications (terms including the factor *direction of change* were never significant, all $p > .34$). Repetition and reading produced opposite results. In reading there were significantly more simplifications ($G^2 = 9.4, p = .002$), but in repetition there were significantly more complications ($G^2 = 4.1, p = .04$). The pattern of changes did not depend on whether the syllable template had been altered (all interactions involving *direction of change* and *template altered*, $p > .19$), nor did it depend on the type of neologism (all interactions involving *direction of change* and *type of neologism*, $p > .22$).

Discussion. V.S. did not consistently simplify syllable structures across tasks. She did make more simplification errors in reading, but an effect limited to reading is not consistent with a deficit of articulatory planning. Contrary to the early claims by Blumstein (1973), we found that not all errors from fluent speech simplified syllables. These results contribute to an emerging picture of differentiated deficits affecting speech processing (consistent with the claims of Romani & Galluzzi, 2005; Romani et al., 2002). Phonological errors do not have uniform characteristics in aphasic patients, and they do not arise from a single locus.

GENERAL DISCUSSION

In this paper we have addressed one of the traditional questions raised by jargonaphasia. We

have asked whether the neologistic errors made by our patient arise from a single postlexical deficit or from a deficit affecting lexical access in addition to a postlexical deficit. We have shown that the neologistic errors characteristic of jargonaphasia can arise from a single post-access impairment in spite of the striking contrast between neologisms that have nearly everything and almost nothing in common with their targets (e.g., *session* → *fession*; *analogy* → *fnichal*). The abstruse errors that V.S. produces do not result from a complete failure of lexical access. Instead, they are words that are more severely affected by the same transformations that produce target-related errors. The fact that target-related and abstruse neologisms occurred in similar proportions across tasks, that both types of error were sensitive to frequency, imageability, and length, and that the percentage of preserved phonemes showed no discontinuity all points to a single underlying deficit.

Our conclusions regarding the source of the neologistic errors match proposals made by several earlier studies (Cohen, Verstichel, & Dehaene, 1997; Hillis, Boatman, Hart, & Gordon, 1999; Miller & Ellis, 1987; Robson et al., 2003). In particular, Robson et al. (2003) draw similar conclusions based on a patient who makes abstruse errors, which, nevertheless, share more phonemes than expected by chance with their targets. Our study extends the case for a single source by marshalling converging evidence across tasks, by showing consistent effects of influential factors across error types, and by formally testing the possibility that the distribution of overlap was a mixture of target-related and unrelated errors.

We have evaluated whether the errors made by V.S. were consistent with loci that have been implicated in patients who make nonword errors: the phonological buffer, an articulatory syllabary, and articulatory planning. However, a series of results suggested that V.S.'s deficit is immediately postlexical. The rate of error per phoneme did not increase for longer words as it should have done if errors involved a memory store (a phonological buffer). V.S. had fluent speech, her errors did not

use especially frequent syllables, and they did not systematically simplify syllable structure. These characteristics are not consistent with an articulatory planning deficit. Instead, V.S. did make substitutions that were related by features. Put together with our other results, this implies that V.S. had trouble accessing segmental representations at the immediately postlexical level, and that these representations are not unitary, but have a complex featural structure.

As we anticipated above, our evidence is not necessarily in conflict with the characteristics of movement errors from speech corpora (Shattuck-Hufnagel & Klatt, 1979), where features do not move independent of their segments. In autosegmental theories of phonological representation, a complex representation of segments is linked to a skeletal position that encodes order. Under this system of representation, movement errors, because they are errors of order, necessarily involve the skeletal tier. Features do not typically have independent links to the skeleton, but have a hierarchical internal structure (a feature geometry) with a single link to the skeletal tier. Thus, it is natural for movement errors to involve segments rather than features. Movement involves breaking the single link to the skeleton and relinking at another position. The more general theoretical point (independent of an assumption of autosegmental representations) is that representing features does not guarantee, by itself, that they will be able to move independent of the segments they define.

An account that only represents features at a level immediately prior to articulation has some more general drawbacks. The phonological alternations that are a major concern of generative phonology must frequently be written with reference to features (e.g., the alternations between the *-z*, *-s*, *-es* forms of the English plural). These alternations are not usually described to be at the level of articulatory implementation. Having features connected only to articulation necessitates some other treatment for these phenomena (or requires that they be located at a more peripheral level).

Our results point to a deficit of phonological encoding, but they do not distinguish whether

segmental representations received only weak activation from lexical representations or were, themselves, damaged. One might hope to be able to do better. In the architecture we have assumed, the phonological level is activated from the lexicon, but also from nonlexical processes for reading and repetition. When a patient's reading, naming, and repetition are similar, it would be natural to assume that damage is at the phonological level, the level that is shared by these tasks. Conversely, if naming were poorer than reading and repetition, one might be tempted to conclude that access to phonemes from the lexicon was impaired, but the phonemes themselves were spared. Reading and repetition would be better than naming because nonlexical processes boost weak lexical activation (Hanley et al., 2004; Hanley et al., 2002; Hillis & Caramazza, 1995).

Under these assumptions, we would conclude that V.S. has damage to the phonological representations themselves. We cannot be sure that V.S.'s damage is at this level, however, because the logic we have just outlined is faulty. V.S.'s nonword reading and repetition were poor. With little or no nonlexical contribution, reading, naming, and repetition would all be based on lexical activation of phonological representations, even if damage affects access to these representations rather than the representations themselves.

Similar performance across tasks is not uniquely identified with damage at the phoneme level itself. Conversely, performance can differ across tasks even if there is damage at the phoneme level. When phonemes are activated by both lexical and nonlexical processes, even damaged representations have a better chance of being produced correctly than when there is only one source of activation. Thus, reading and repetition could still be better than naming. Our case is a specific example of the more general problem, which others have identified, of distinguishing damage affecting "storage" from damage affecting "access" to the stored information (Forde & Humphreys, 1997; Rapp & Caramazza, 1993).

We have mentioned that our conclusions regarding the number and locus of deficits in V.S. are similar to those presented in several

other single-case studies. They are also similar to the conclusions reached by Schwartz et al. (2004) based on group data. It may be surprising, then, that our conclusions regarding the possible sources of nonword errors are very different from those presented by Schwartz et al. The Schwartz et al. (2004) study used the naming errors of a mixed group of aphasic patients (conduction aphasics, Wernicke's aphasics, and anomic patients) to argue that there is a single source for all aphasic nonword errors. We agree with Schwartz et al. that nonword errors (and for us, jargonaphasia, specifically) can be produced by a deficit at the level where segments are addressed by the lexicon. We cannot, however, use our data to argue that all nonword aphasic errors must have this source.

The data from V.S. and from other patients of varying severity who make phonological errors (e.g., Caramazza et al., 1986; Miller & Ellis, 1987; Robson et al., 2003; Shallice et al., 2000) make it clear that there are methodological problems with the overlap analysis that Schwartz et al. (2004) use to support their conclusions. Our demonstration that a single deficit can be responsible for a wide range of target-error overlap makes interpreting data from a group of patients problematic. Schwartz et al. claim that since they failed to find a bimodal distribution of target/error overlap, all patients have a single type of deficit. Unfortunately, a bimodal distribution would be difficult to see in a group under most circumstances. Bimodality would be obscured, even if there were both lexical and postlexical sources for errors unless, by chance, the group only included patients with deficits to lexical access and patients with very mild impairments to phonological encoding. In other words, patients like V.S., with single deficits of intermediate severity, are just those that would make bimodality difficult to observe in a group (by implication, it is also possible, although unlikely, for bimodality to arise as an artefact in a group of patients who all have a single deficit; this will happen if, by chance, members of the group have only severe or mild deficits, e.g., contrast the patients reported by Caramazza et al., 1986, and Shallice et al.,

2000, with the patient reported by Robson et al., 2003).

Schwartz et al. (2004) report that 2 of their 18 patients showed some evidence of bimodality at the individual level. Whether the evidence from these patients would be strong enough to support a two-deficit account would require formal analysis, but it is clear that the critical tests would have to take place at the level of the individuals and not the group.

The point we have made by contrasting our case study and the Schwartz et al. group study also follows from putting our results in the wider context of the single-case literature devoted to patients who make phonological errors. Documenting an impairment at phonological encoding for one patient does not rule out the involvement of lexical access in other patients who make phonological errors. V.S. does not rule out the possibility that an intact phonological encoding system, which is not directed by lexical representations but by previous responses or left-over activation, will produce abstruse neologisms (see proposals with different flavours in Buckingham, 1981; Butterworth, 1979; Moses, Nickels, & Sheard, 2004; Stenneken, Hofmann, & Jacobs, 2005).

Other postlexical levels might also produce jargonaphasia. There are patients who make target-related nonword errors whose characteristics either clearly do not match or may not match those of V.S. (e.g., Caramazza et al., 1986; Cohen et al., 1997; Romani & Calabrese, 1998; Romani & Galluzzi, 2005; Romani et al., 2002; Shallice et al., 2000). One might propose that jargonaphasia could result from any deficit that produces segmental errors, as long as the deficit was sufficiently severe. The range of postlexical levels that could produce jargonaphasia would then extend from phonological encoding and the phonological buffer, through to articulatory planning and implementation.

Our own view is that while the phonological buffer is a possible locus, articulatory planning and implementation are not. Abstruse errors require that phonological material be filled into a response, and mechanisms to do this must be

available where jargonaphasia arises. This means that jargonaphasia will arise postlexically where phonemes are being selected or temporarily stored. Our prediction is that, at articulatory planning and implementation, nonword errors will remain related, and increasing severity will produce increasingly simplified responses (e.g., /tə/), incomplete responses, or severely dysfluent responses, but not unrelated nonwords.

Our view, then, is that the sources of abstruse neologisms are probably limited, not to a single level, but to those levels that interact directly with phonological encoding. Comparisons between V.S. and single case studies of patients with deficits localized to other levels, however, show that all nonword errors are unlikely to have the single source that Schwartz et al. (2004) claim. The dysfluent patients reported by Romani and Galluzzi (2005), for example, systematically simplified syllable structure. Since these patients had deficits covering a wide range of severity (from near 0 to over 60% correct in repetition), a general factor like “level of impairment” will not explain why they are clearly different from V.S.²

What our results show is that detailed case studies that examine both the types of target that produce aphasic errors (e.g., their frequency, concreteness, and length) and the errors themselves will be rewarded in the effort to identify the deficit or combination of deficits that produce neologistic errors. We have demonstrated that a variety of formal methods (maximum likelihood estimation, model selection, edit distance) can be productively applied to patient data to make accounts explicit and testable. We hope that our example and others using similar quantitative methods (e.g., Cohen & Dehaene, 1998; Graham, Patterson, & Hodges, 2001; McCloskey, Macaruso, & Rapp, 2005; Rapp & Goldrick, 2000; Shallice et al., 2000) will encourage these kinds of analysis to be applied more widely to patient data as theoretical accounts become more articulated.

REFERENCES

- Aichert, I., & Ziegler, W. (2004). Syllable frequency and syllable structure in apraxia of speech. *Brain and Language*, 88, 148–159.
- Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Second International Symposium on Information Theory* (pp. 267–281). Budapest: Akademiai Kiado.
- Alajouanine, T. (1956). Verbal realization in aphasia. *Brain*, 79, 22.
- Anderson, S. R. (1981). Why phonology isn't “natural”. *Linguistic Inquiry*, 12, 493–539.
- Anderson, S. R., & Lightfoot, D. W. (2002). *The language organ*. Cambridge, UK: Cambridge University Press.
- Baayen, R. H., Piepenbrock, R., & Van Rijn, H. (1995). *The CELEX Lexical Database [CD-ROM]*. Philadelphia: University of Pennsylvania, Linguistic Data Consortium.
- Baddeley, A. (1990). *Human memory: Theory and practice*. Hove, UK: Lawrence Erlbaum Associates Ltd.
- Barry, C., Hirsh, K. W., Johnston, R. A., & Williams, C. L. (2001). Age of acquisition, word frequency, and the locus of repetition priming of picture naming. *Journal of Memory and Language*, 44, 350–375.
- Blumstein, S. E. (1973). *A phonological investigation of aphasic speech*. The Hague, The Netherlands: Mouton.
- Brown, R., & McNeill, D. (1966). The “tip of the tongue” phenomenon. *Journal of Verbal Learning and Verbal Behavior*, 5, 325–337.
- Bub, D., Black, S., Howell, J., & Kertesz, A. (1987). Speech output processes and reading. In M. Coltheart, G. Sartori, & R. Job (Eds.), *The cognitive neuropsychology of language*. Hove, UK: Lawrence Erlbaum Associates Ltd.
- Buckingham, H. W. (1977). Conduction theory and neologistic jargon. *Language and Speech*, 20, 174–184.
- Buckingham, H. W. (1981). Where do neologisms come from? In J. Brown (Ed.), *Jargonaphasia*. New York: Academic Press.

²In addition, it is unclear why Schwartz et al. apparently allow normal slips of the tongue to have a source after the level of phonological encoding but exclude this level as a possible source of aphasic errors (see their treatment of movement errors, Schwartz et al. 2004, p. 161).

- Buckingham, H. W., & Kertesz, A. (1976). *Neologistic jargon aphasia*. Amsterdam: Swets and Zeitlinger.
- Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). New York: Springer.
- Burzio, L. (2002). Surface-to-surface morphology: When your representations turn into constraints. In P. Boucher (Ed.), *Many morphologies*. Somerville, MA: Cascadilla Press.
- Butterworth, B. (1979). Hesitation and the production of verbal paraphasias and neologisms in jargon aphasia. *Brain & Language*, *8*, 133–161.
- Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, *14*, 177–208.
- Caramazza, A., Miceli, G., & Villa, G. (1986). The role of the (output) phonological buffer in reading, writing, and repetition. *Cognitive Neuropsychology*, *3*, 37–76.
- Caramazza, A., & Miozzo, M. (1997). The relation between syntactic and phonological knowledge in lexical access: Evidence from the “tip-of-the-tongue” phenomenon. *Cognition*, *64*, 309–343.
- Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. Cambridge, MA: The MIT Press.
- Clements, G. N. (1990). The role of the sonority cycle in core syllabification. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology 1*. Cambridge, UK: Cambridge University Press.
- Cohen, L., & Dehaene, S. (1998). Competition between past and present—assessment and interpretation of verbal perseverations. *Brain*, *121*, 1641–1659.
- Cohen, L., Verstichel, P., & Dehaene, S. (1997). Neologistic jargon sparing numbers: A category-specific phonological impairment. *Cognitive Neuropsychology*, *14*, 1029–1061.
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*, 204–256.
- Cowan, G. (1998). *Statistical data analysis*. Oxford, UK: Clarendon Press.
- Davies, P. D., & Funnell, E. (2000). Semantic representation and ease of predication. *Brain and Language*, *73*, 92–119.
- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*, 283–321.
- Dell, G. S. (1988). The retrieval of phonological forms in production—tests of predictions from a connectionist model. *Journal of Memory and Language*, *27*, 124–142.
- Dell, G. S., & O’Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition*, *42*, 287–314.
- Forde, E. M. E., & Humphreys, G. W. (1997). A semantic locus for refractory behaviour: Implications for access-storage distinctions and the nature of semantic memory. *Cognitive Neuropsychology*, *14*, 367–402.
- Goldrick, M., & Rapp, B. (2007). Lexical and post-lexical phonological representations in spoken production. *Cognition*, *102*(2), 219–260.
- Goldsmith, J. A. (1990). *Autosegmental and metrical phonology*. Oxford, UK: Basil Blackwell.
- Goodglass, H., Kaplan, E., & Weintraub, S. (1983). *The Boston Naming Test*. Philadelphia: Lea & Febiger.
- Graham, N. L., Patterson, K., & Hodges, J. R. (2001). The emergence of jargon in progressive fluent dysgraphia: The widening gap between target and response. *Cognitive Neuropsychology*, *18*, 343–361.
- Gusfield, D. (1997). *Algorithms on strings, trees, and sequences: Computer science and computational biology*. Cambridge, UK: Cambridge University Press.
- Hanley, J. R., Dell, G. S., Kay, J., & Baron, R. (2004). Evidence for the involvement of a nonlexical route in the repetition of familiar words: A comparison of single and dual route models of auditory repetition. *Cognitive Neuropsychology*, *21*, 147–158.
- Hanley, J. R., & Kay, J. (1997). An effect of imageability on the production of phonological errors in auditory repetition. *Cognitive Neuropsychology*, *14*, 1065–1084.
- Hanley, J. R., Kay, J., & Edwards, M. (2002). Imageability effects, phonological errors, and the relationship between auditory repetition and picture naming: Implications for models of auditory repetition. *Cognitive Neuropsychology*, *19*, 193–206.
- Harley, T. A., & Bown, H. E. (1998). What causes a tip-of-the-tongue state? Evidence for lexical neighbourhood effects in speech production. *British Journal of Psychology*, *89*, 151–174.
- Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review*, *111*, 662–720.
- Hillis, A. E., Boatman, D., Hart, J., & Gordon, B. (1999). Making sense out of jargon—a neurolinguistic and computational account of jargon aphasia. *Neurology*, *53*, 1813–1824.
- Hillis, A. E., & Caramazza, A. (1991). Mechanisms for accessing lexical representations for output—evidence from a category-specific semantic deficit. *Brain and Language*, *40*, 106–144.

- Hillis, A. E., & Caramazza, A. (1995). Converging evidence for the interaction of semantic and sub-lexical phonological information in accessing lexical representations for spoken output. *Cognitive Neuropsychology*, *12*, 187–227.
- Jescheniak, J. D., & Levelt, W. J. M. (1994). Word-frequency effects in speech production—retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning Memory and Cognition*, *20*, 824–843.
- Jones, G. V. (1985). Deep dyslexia, imageability, and ease of predication. *Brain and Language*, *24*, 1–19.
- Kay, J., Lesser, R., & Coltheart, M. (1992). *Psycholinguistic assessments of language processing in aphasia*. Hove, UK: Psychology Press.
- Kertesz, A., & Benson, D. F. (1970). Neologistic jargon. *Cortex*, *6*, 362–386.
- Kohn, S. E., & Smith, K. L. (1994). Distinctions between two phonological output deficits. *Applied Psycholinguistics*, *15*, 75–95.
- Levelt, W. J. M. (1989). *Speaking*. Cambridge, MA: MIT Press.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioural and Brain Sciences*, *22*, 1–38.
- Levelt, W. J. M., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary. *Cognition*, *50*, 239–269.
- Marshall, J. (2006). Jargon aphasia: What have we learned? *Aphasiology*, *20*, 387–410.
- McCloskey, M., Macaruso, P., & Rapp, B. (2005). Grapheme-to-lexeme feedback in the cognitive spelling system. *Cognitive Neuropsychology*, *23*, 278–307.
- Miller, D., & Ellis, A. (1987). Speech and writing errors in “neologistic jargonaphasia”: A lexical activation hypothesis. In M. Coltheart, G. Sartori, & R. Job (Eds.), *The cognitive neuropsychology of language*. Hove, UK: Lawrence Erlbaum Associates Ltd.
- Moses, M. S., Nickels, L. A., & Sheard, C. (2004). Disentangling the web: Neologistic perseverative errors in jargon aphasia. *Neurocase*, *10*, 452–461.
- Ohala, J. J. (1990). There is no interface between phonology and phonetics. *Journal of Phonetics*, *18*, 153–171.
- Paivio, A. (1991). Dual coding theory—retrospect and current status. *Canadian Journal of Psychology—Revue Canadienne de Psychologie*, *45*, 255–287.
- Perecman, E., & Brown, J. (1985). Varieties of aphasic jargon. *Language Sciences*, *7*, 176–215.
- Rapp, B., & Caramazza, A. (1993). On the distinction between deficits of access and deficits of storage—a question of theory. *Cognitive Neuropsychology*, *10*, 113–141.
- Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review*, *107*, 460–499.
- Rapp, B., & Goldrick, M. (2004). Feedback by any other name is still interactivity: A reply to Roelofs (2004). *Psychological Review*, *111*, 573–578.
- Riddoch, M. J., & Humphreys, G. W. (1993). *Birmingham Object Recognition Battery*. Hove, UK: Lawrence Erlbaum Associates Ltd.
- Robson, J., Pring, T., Marshall, J., & Chiat, S. (2003). Phoneme frequency effects in jargon aphasia: A phonological investigation of nonword errors. *Brain and Language*, *85*, 109–124.
- Roelofs, A. (1994). On-line versus off-line priming of word-form encoding in spoken word production. In A. Ram & K. Eiselt (Eds.), *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society* (pp. 772–777). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Roelofs, A. (1999). Phonological segments and features as planning units in speech production. *Language and Cognitive Processes*, *14*, 173–200.
- Roelofs, A. (2004). Comprehension-based versus production-internal feedback-in planning spoken words: A rejoinder to Rapp and Goldrick (2004). *Psychological Review*, *111*, 579–580.
- Roelofs, A., Meyer, A. S., & Levelt, W. J. M. (1998). A case for the lemma/lexeme distinction in models of speaking: Comment on Caramazza and Miozzo (1997). *Cognition*, *69*, 219–230.
- Romani, C., & Calabrese, A. (1998). Syllabic constraints in the phonological errors of an aphasic patient. *Brain & Language*, *64*, 83–121.
- Romani, C., & Galluzzi, C. (2005). The effects of syllabic complexity on correct repetition in a group of aphasic patients. *Cognitive Neuropsychology*, *22*, 817–850.
- Romani, C., Olson, A., Semenza, C., & Grana, A. (2002). Patterns of phonological errors as a function of a phonological versus an articulatory locus of impairment. *Cortex*, *38*, 541–567.
- Schwartz, M. F., Dell, G. S., Martin, N., Gahl, S., & Sobel, P. (2006). A case-series test of the interactive two-step model of lexical access: Evidence from picture naming. *Journal of Memory and Language*, *54*, 228–264.
- Schwartz, M. F., Wilshire, C. E., Gagnon, D. A., & Polansky, M. (2004). Origins of nonword phonological errors in aphasic picture naming. *Cognitive Neuropsychology*, *21*, 159–186.

- Shallice, T., Rumiati, R. I., & Zadini, A. (2000). The selective impairment of the phonological output buffer. *Cognitive Neuropsychology*, *17*, 517–546.
- Shattuck-Hufnagel, S., & Klatt, D. H. (1979). The limited use of distinctive features and markedness in speech production: Evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior*, *18*, 41–55.
- Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 174–215.
- Stemberger, J. P. (1982). The nature of segments in the lexicon: Evidence from speech errors. *Lingua*, *56*, 235–259.
- Stenneken, P., Hofmann, M., & Jacobs, A. M. (2005). Patterns of phoneme and syllable frequency in jargon aphasia. *Brain and Language*, *95*, 221–222.
- Ward, J., & Romani, C. (1998). Serial position effects and lexical activation in spelling: Evidence from a single case study. *Neurocase*, *4*, 189–206.
- Wilson, M. D. (1988). The MRC Psycholinguistic Database: Machine readable dictionary, Version 2. *Behavioural Research Methods, Instruments and Computers*, *20*, 6–11.

APPENDIX A

Sample of spontaneous speech

'Cos my garden's always being mashed. It's gone to [weʃd]. [splɒŋ] like nobody. But I mean if they was [grɒp] properly I should see [ðegfɒm] in there quite a lot. Ooh quite nice. They're [bihɒp] push if you only look. My garden was a nice garden once. I wish somebody would clean it up. You know, it's gone to a hole. I mean 'cos I use to do it [mɔːsəl] but I can't, I can't do it. When I had, when I first done it, it was quite nice. I used to do it, doing it my [weɪn] and beans and everything . . . all my own garden.

APPENDIX B

Calculating the model estimates for unimodal and mixture models of overlap

A normal distribution of target/error overlap that took into consideration the opportunities for error from the stimuli was constructed by tabulating all the possible overlaps for each of the targets involved in neologistic errors. For example, if a three-letter target was involved in an error, we would tabulate possible target–error overlaps of 0, .33, .66, and 1. Accumulating these overlaps for all targets gave a distribution of the opportunities for overlap provided by the stimuli. This was multiplied by a normal distribution with a given mean and standard deviation to produce a measure of normally distributed overlap given the characteristics of the targets.

Two models were compared. A normal distribution by itself and a normal distribution plus the distribution of chance overlap. The distribution of chance overlap was calculated by pairing each target with all the neologisms except its own and calculating the phoneme overlap. For the combined normal + chance distribution, the weight given to normal and chance distributions was a parameter of the model. The two models were, then:

1. Normal only:

$$p(\text{overlap } x) = \text{normal}(x, \text{mean}_u, \text{SD}_i) \times \text{opportunities}(x)$$

with parameters: $\text{mean}_u, \text{SD}_i$.

2. Normal + chance:

$$p(\text{overlap } x) = [(1 - wt) \times \text{normal}(x, \text{mean}_u, \text{SD}_i)] \times \text{opportunities}(x) \times [(wt \times \text{chance}(x))]$$

with parameters: $\text{mean}_u, \text{SD}_i, wt$.

3. The parameter wt was the relative weight assigned to the chance distribution, and $(1 - wt)$ the weight assigned to the normal distribution. A large weight would mean that responses with no more than a random number of target phonemes contributed substantially to predicting the observed data and a small weight would mean that responses without target information contributed little to the observed data.
4. The log likelihood of the histogram of V.S.'s data (using 10 equally spaced bins) was calculated for each model (according to Cowan, 1998), and the parameters of the model were optimized numerically (using the `constrOptim` function from R Development Core Team (2006). R: *A language and environment for statistical computing*. <http://www.R-project.org>). The AIC was calculated from the maximum likelihood estimate of the optimized model (Burnham & Anderson, 2002):

$$AIC = -2 \log(\text{likelihood value}) + 2(\text{number of parameters})$$

If the observed distribution of target/error overlap came from a mixture of responses where lexical access failed and responses where lexical access succeeded, the normal + chance model should be better according to the AIC, and the chance distribution should contribute a meaningful amount to predicting the results (i.e., the wt parameter should be large enough to contribute meaningfully to the overall pattern). Neither of these was the case.