Target/error overlap in jargonaphasia

Olson, Andrew; Halloran, Elizabeth; Romani, Cristina

[Link to publication on Research at Birmingham portal](#)

Accepted Manuscript

Target/error overlap in jargonaphasia: The case for a one-source model, lexical and non-lexical summation, and the special status of correct responses

Andrew Olson, Elizabeth Halloran, Cristina Romani

Running head: TARGET/ERROR OVERLAP IN JARGONAPHASIA

Target/error overlap in jargonaphasia: The case for a one-source model,

lexical and non-lexical summation, and the special status of correct responses

Andrew Olson*, Elizabeth Halloran* & Cristina Romani**

*School of Psychology, University of Birmingham

**School of Life & Health Sciences, Aston University

Abstract

We present three jargonaphasic patients who made phonological errors in naming, repetition and reading. We analyze target/response overlap using statistical models to answer three questions: 1) Is there a single phonological source for errors or two sources, one for target-related errors and a separate source for abstruse errors? 2) Can *correct* responses be predicted by the same distribution used to predict errors or do they show a completion boost? 3) Is non-lexical and lexical information summed during reading and repetition? The answers were clear. 1) Abstruse errors did not require a separate distribution created by failure to access word forms. Abstruse and target-related errors were the endpoints of a single overlap distribution. 2) Correct responses required a special factor, e.g. a completion boost or lexical/phonological feedback, to preserve their integrity. 3) Reading and repetition required separate lexical and non-lexical contributions that were combined at output.

We present a study of target/response overlap in the spoken output of three jargonaphasic patients using data from reading, naming and repetition. We use statistical models to address three related questions. The first question concerns the striking phenomenon that gives jargonaphasia its name. Jargonaphasic patients sometimes make errors that are clearly related to the target word (target-related errors: e.g. strawberry > strewberry), but they also make errors that seemingly bear little relationship to the target (neologistic errors: e.g. suitcase > teligom). We ask if these errors have two sources—one for related errors, based on successful access to word forms, accompanied by occasional minor segmental errors, and a second, for abstruse errors, based on a failure to gain access word forms. In its clearest form, two sources would produce a bimodal distribution of overlap. The alternative, a single segmental source, would predict that some words will be relatively untouched by errors and others will be completely altered, but the majority will fall continuously between these extremes.

Our second question is related, but involves correct responses, which have not been a traditional concern of jargonaphasic studies. We ask if correct responses are predictable from the level of overlap that characterizes errors, as would be expected if correct responses are just those where all segments escape the error generating process unscathed.

Our third question concerns whether spoken responses are based on combining lexical and non-lexical information. There is extensive debate in the neuropsychological and modeling literature over the most appropriate architecture for the reading process and a more limited discussion of the same issues for repetition (e.g. see reading discussion in Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; for repetition see: Hanley, Dell, Kay, & Baron, 2004; Hanley & Kay, 1997; Hanley, Kay, & Edwards, 2002; Harm & Seidenberg, 2004; Zorzi, Houghton, & Butterworth, 1998

and associated references). We test the hypothesis that there are lexical and non-lexical routes to repetition and reading and that these sources sum together to produce a response (Alario, Schiller, Domoto-Reilly, & Caramazza, 2003; Bi, Han, Weekes, & Shu, 2007; Funnell, 1996; Hanley & Kay, 1997; Hanley et al., 2002; Hillis & Caramazza, 1991, 1995; Howard & Franklin, 1988; Miceli, Capasso, & Caramazza, 1994; Tree, Kay, & Perfect, 2005; Ward, Stott, & Parkin, 2000).

We test these theoretical alternatives by formalizing them in a hierarchically related set of mathematical models that predict the distribution of shared phonemes in targets and responses. We evaluate our set of models using formal model selection. Our case series, as a result, has methodological, as well as empirical implications. Formalising models and using model selection methods can be a good way to explore specific quantitative consequences of theories and to confront theories and data. Our results will illustrate a point that others have made before, most frequently in the domains of statistical theory and biological modelling (see, specifically, Burnham & Anderson, 2002): Model selection, which emphasizes the relative ability of a collection of models to account for data, and the interpretation of the values of the parameters required to fit the data, is a more appropriate perspective for comparing theories than winner-take-all methods based on hypothesis testing, where binary decisions sometimes reflect relatively minor differences in fit.

To situate our results in relationship to the existing psycholinguistic and neuropsychological literature, we will use terms and levels defined by a large literature on speech production (e.g. Caramazza, 1997; Dell, 1986; Garrett, 1980; Kempen & Huijbers, 1983; Levelt, 1989; Levelt, Roelofs, & Meyer, 1999; Rapp & Goldrick, 2000), but abstracting away from some of the details that differ between accounts. In common with most neuropsychological and psycholinguistic architectures, we assume there is a semantic/conceptual level that accesses a distinct

level where words are represented as unitary items (Caramazza & Hillis, 1990; Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Howard & Gatehouse, 2006; Levelt et al., 1999). This is the level of *word selection.* A unitary word level is needed to account for whole word errors in speech production (Garrett, 1975) and for effects of grammatical processes that apply to words and not their component parts (e.g. effects of grammatical class, gender, number: Badecker, Miozzo, & Zanuttini, 1995; Bock & Miller, 1991; Caramazza & Miozzo, 1997; Garrett, 1975; Henaff Gonon, Bruckert, & Michel, 1989; Miozzo & Caramazza, 1997a, 1997b; Vigliocco, Antonini, & Garrett, 1997). In some accounts there are two unitary word levels, connected to syntactic and phonological information (lemma and lexeme, respectively; Levelt et al., 1999; Vigliocco et al., 1997). In other accounts there is a single level (lexeme; Caramazza, 1997; Caramazza & Miozzo, 1997; Vigliocco et al., 1997). Our data do not speak to this issue, so we will refer to a unitary word level without prejudging whether a syntactic level is accessed *before* the level that is linked to phonemes (i.e. lemma before lexeme). Unitary word representations are connected, at the next level, to corresponding phonological segments. This is the level of *phoneme selection.* These are abstract phonemes, not completely specified for production, but specified for contrastive features (e.g. Voicing, which differentiates /p/ and /b/ in English would be specified, but not aspiration, which is not contrastive in English. The /p/ would not have aspiration specified at this level). Finally, there is a level where all phonetic dimensions are specified for articulation. These distinctions are widely shared in linguistics and psycholinguistics (Anderson, 1974; Goldrick & Rapp, 2007; Laganaro, 2012).

We assume that the connections between semantic information and word nodes, the word nodes themselves, and their connections with phonemes constitute the lexical representation of a word. We also assume that the abstract phoneme level that

is part of the lexical representation is also addressed by non-lexical conversion processes for reading or repetition through independent connections.

Our architecture for speech production is diagrammed in Figure 1. This is identical to the architecture described by Goldrick and Rapp (2007) except that they assume that lexical and non-lexical information converge later, at the level of *phonetic encoding*. In the interest of parsimony, we assume that summation occurs at the first available level (*phoneme selection* in Figure 1), so that phonetic encoding is not duplicated in lexical and non-lexical routes. One consequence is the phoneme level is accessed by both lexical and non-lexical processes in our architecture. To avoid the confusion that could result from giving this a lexical label (e.g. *lexical phonology*), we refer to this level as *phoneme selection* following Romani, Galluzzi, Bureca and Olson (2011).

Figure 1 about here

Note that there is some disagreement in the literature over what is considered *post-lexical* that is, naturally, related to which levels are considered to be part of lexicon. Some people do not consider the connections between word nodes and phonemes to be part of lexical representations, and, therefore consider phonological encoding to be a post-lexical level. We consider, instead, this level to be part of lexical representations because a word is characterized by its sequence of phonemes as much as by its meaning (see also, Goldrick and Rapp, 2007). To avoid potential confusion, we will not use the term *post-lexical* and, instead, use the term *post-access* to identify the levels after which a word node has been correctly activated/selected. We will describe the question of the locus of the neologistic errors as a contrast

between a source at lexical access and a post-access source.

Sources needed to explain target/response overlap. Two alternative accounts of jargonaphasic errors have been proposed. A one source account attributes target-related and neologistic errors to difficulties in phoneme selection which vary in severity (Kertesz & Benson, 1970; Kohn & Smith, 1994; Miller & Ellis, 1987; Olson, Romani, & Halloran, 2007; Robson, Pring, Marshall, & Chiat, 2003; Schwartz, Wilshire, Gagnon, & Polansky, 2004). The two types of errors result from the extremes of what is, in fact, a continuous distribution. According to this hypothesis, damage may impact some words strongly and others weakly, but errors with intermediate levels of target overlap should be frequent.

A two source account, instead, attributes neologistic and target-related errors to failures at two different loci: word selection and phoneme selection or, in other words, at lexical access and post-access, respectively (Buckingham, 1981; Buckingham & Kertesz, 1976). Neologistic errors arise because word nodes have either been lost or they have been disconnected from the phoneme level. A random source of phonemes "fills in" the response (Buckingham, 1981; Butterworth, 1979; Moses, Nickels, & Sheard, 2004; for Buckingham, the random source is syllables, for Butterworth, it is morphemes, for Moses, Nickels & Sheard it is previous responses; see Marshall, 2006). Target-related errors, instead, result from failure to access a limited number of the phonemes at phoneme selection, the same locus where errors arise in the one source account. This predicts a discontinuity in the distribution of errors with some showing substantial overlap with the target and others showing minimum overlap.

What is at issue here is not whether failures of word selection can occur at all. We assume that word selection failure <u>can</u> occur and that no response errors and semantic errors are examples of this. Our question is more specific. We ask whether

the *abstruse nonword* errors seen in jargonaphasia can also result from a word

selection failure.

Schwartz et al. (2004) present a general analysis of where nonword

phonological errors can arise in the speech architecture. They conclude that there is

no support for anything but a single locus at the level of phoneme selection, but they

define two other loci where errors *might* have originated. One is at the level of word

selection, in line with the jargonaphasia accounts mentioned above, but through a

mechanism different than "filling in." An abstruse neologism could result if an

incorrect word was selected and then distorted further by errors at phoneme selection

(p 160, "errors of compound origin"). Like errors resulting from "filling in," these

errors should have very low overlap with targets. A hallmark of this mechanism,

however, is that nonword errors should also be accompanied by a substantial number

of word or near-word errors (which result when the wrong word is selected but not

distorted enough to be unrecognizable).

An alternative, second source of nonword errors is at the level, after phoneme

selection. This is the source that Schwartz et al. (2004) explicitly argue against. They

find no evidence that errors with high and low target overlap differ in their serial order

effects, sensitivity to frequency, or sensitivity to length and conclude that all the errors

made by their 18 patients arise at phoneme selection, where phonemes are mis-

selected based on a correctly selected word.   Contrary to Schwartz et al., we believe

that nonword errors may also arise at a more peripheral/phonetic level and errors from

this level will have different characteristics (e.g. see Galluzzi, Bureca, Guariglia &

Romani, in press; Romani & Galluzzi, 2010; Romani, Olson, Semenza, & Grana,

2002). Our analyses here, however, do not consider this question (none of our

patients have more peripheral/phonetic impairments). We focus on whether abstruse

errors arise in phoneme selection (like target-related errors) or at a previous word selection level.

We have previously reported an analysis of target/error overlap for one of the patients that we also present here (VS, Olson et al., 2007). We found that it was not necessary to hypothesize a second source for her neologistic errors. Target/error overlap could be explained by a single distribution with an average overlap of about 50%. Here, we present new modeling results along with two other patients who also make neologistic errors, but who have different profiles from VS. One, in particular (JH), shows much less target/response overlap in naming than VS, and superficially, at least, seems to be a good candidate for a patient with a second, word-level, source of error.

Predicting correct responses. In addition to the source of *errors*, we will examine a related issue which has received little previous attention. We ask whether *all* responses-- *both correct and incorrect*--come from a single distribution. If there is a single segmental level where errors arise, the number of items that are completely free of errors should be predictable from the likelihood that any one phoneme is involved in an error. Correct responses will result when all phonemes emerge from the error generating process unscathed. We ask, therefore, whether all responses, both correct and incorrect, form a continuous distribution. The alternative is that correct responses are not a predictable function of error probabilities because correct responses benefit from an extra measure of cohesiveness or a lexical boost.

Lexical and non-lexical contributions to reading and repetition. There is a substantial and on-going debate about the organization of the general architecture for reading and both aphasic patients and computational models have contributed results (Coltheart et al., 2001; Harm & Seidenberg, 2004; Zorzi et al., 1998). Descriptive labels like *single-route* and *dual-route* aside, however, all alternative accounts

distinguish two mechanisms for output and, therefore, two potential sources of errors: one mapping letters to sounds (the grapheme-phoneme rule system in the Dual Route Cascaded, or DRC, model and the network linking orthography and phonology in the Triangle Model or Connectionist Dual Process, or CDP, model) and one mediated by semantic or word-based knowledge (the connection that runs through input and output lexicons in the DRC model, the "word-based" route through the hidden units in the CDP model and the semantically mediated route in the Triangle Model).

A consequence of having two contributions to output is that they must be reconciled. Either one process controls output completely, or the two contributions are combined to produce a single output. Here, we ask whether the distribution of target/error overlap shows evidence of a process that combines information at output. This contrast has received little explicit attention in computational modelling, although all of the major architectures have activation/competition mechanisms for combining information from more than one source at output (Coltheart et al., 2001; Harm & Seidenberg, 2004; Zorzi et al., 1998). There has been more discussion of this issue in the neuropsychological literature, where a summation mechanism has been explicitly proposed (Alario et al., 2003; Bi et al., 2007; Hillis & Caramazza, 1991, 1995; Howard & Franklin, 1988; Miceli et al., 1994; Nozari, Kittredge, Dell, & Schwartz, 2010; Tree et al., 2005; Ward et al., 2000).

If lexical and non-lexical information is not combined at output, there is a clear prediction about target/error overlap. When errors arise at the phoneme level, the lexical contribution will be the same in naming, reading and repetition. Reading and repetition also have a non-lexical route to output, however, which naming does not have. So, output in reading and repetition can be determined by *either* non-lexical information or by lexical information, but not by a combination of the two. If lexical and non-lexical information can be combined, instead, information from non-

lexical repetition or grapheme-phoneme conversion can fill in additional correct phonemes where this information is missing from the lexical route (indexed by naming). This means that overlap in word reading or repetition can be higher than the overlap predicted by naming alone (as an index of what is available from the lexical route) or by non-word reading/repetition alone.

Studies of repetition have directly addressed this question. Hanley and colleagues (Hanley & Kay, 1997; Hanley et al., 2002) reported two patients who repeated high imageability words better than low imageability words. Despite nearly identical levels of performance in picture naming, their repetition performance was very different and was related to their ability to repeat nonwords. Hanley et al. argued that these patterns required a non-lexical route in addition to the lexical route, with summation of outputs across the two routes. They argued that the patients' patterns would not be compatible with accounts that have a single lexical route (Dell et al., 1997; Foygel & Dell, 2000), and their theoretical claims were confirmed by modeling (Hanley et al., 2004).

Subsequently, Baron, Hanley, Dell and Kay (2008) found that a model with lexical and non-lexical routes to word repetition *overpredicted* success for a series of other patients who had relatively good nonword repetition. In other words, a non-lexical contribution boosted word repetition success above the level that was actually observed. Baron et al.'s conclusion was that patients may differ in the extent to which they are able to combine lexical and non-lexical information. Some patients augment lexical information with information from the non-lexical route and others do not. This could allow several alternatives. It could be that either individuals allow summation or they do not, in a binary fashion. Alternatively, lexical and non-lexical information could be weighted differently in different individuals, with some people allowing a larger non-lexical contribution than others. We will return to this in the

General Discussion. The analyses carried out by Hanley et al. (2004) and Baron et al. (2008) were based on the percentage of correct responses. Here we will test the predictions of a dual route/summation account based on a different and, possibly, more sensitive measure: target/error overlap.

Case Studies

Patient VS. Detailed case study information about VS is reported in Olson et al. (2007). We will summarise the main aspects of her performance here. Patient VS was an 84-year-old right-handed woman who had worked at the Cadbury chocolate factory. Following two days of confusion and headaches in 1996, she was found to have an extensive periventricular region of low density in the left parietal lobe with a possible recent infarct. Our data were collected from VS five years later, in 2001.

VS suffered from a marked short-term memory impairment, poor performance distinguishing phonological minimal pairs, but relatively good spoken word-picture matching. Visual lexical decision was relatively poor (91/120; 76%). She was within the normal range (50/52; 96%) on the Pyramids and Palm Trees test of semantics (Howard & Patterson, 1992), but in other tests performed well with high imageability words and poorly with low imageability words (synonym judgments, PALPA 50; high imageability, 27/30, 90%; low imageability, 20/30, 67%; $\chi^2(1)=3.5$, p=.06). Word-picture matching from PALPA (Kay, Lesser, & Coltheart, 1992) was normal (38/40; 95% for both written and spoken versions), but her BPVS (Dunn, Dunn, Whetton, & Pintilie, 1982) score was somewhat low (10 years 7 months, confidence interval = 9 years 11 months to 11 years 3 months). Our study will focus on her naming, reading and repetition. In spontaneous speech VS spoke fluently but with frequent neologisms and phonemic paraphasias.   At times, her sentences or words were hard

to understand, but generally one could recover the gist of what she was trying to say and a relatively normal conversation was possible. She made neologistic errors in all speech production tasks and nonword reading and repetition were poor. A summary of VS's performance on initial neuropsychological tests is presented in Table 1. Note that repetition is more impaired than naming, possibly because of her phonological discrimination difficulties.

Table 1 about here

Patient JH. JH was a 69-yr-old right-handed man. He was admitted to Solihull Hospital in December, 1999, with a loss of speech and a right-sided weakness. A CT scan showed an extensive low density area affecting the grey and white matter in the distribution of the left middle cerebral artery, suggestive of an infarct. The low density extended to involve the basal ganglia on the left.

Data for the current study was collected in conjunction with JH's speech therapist at Solihull Hospital and in JH's own home. A basic language assessment was carried out between February and August 2000. Most of the tests reported below were given towards the end of this period, with the exception of the minimal pairs tests and the Pyramid and Palm Trees. JH's performance remained stable while the data for the experimental section of this paper were collected in December 2000. JH's initial results are summarised in Table 1.

JH showed no clear difficulties with auditory discrimination, performed reasonably well on lexical decision tasks, but showed a mild to moderate impairment of comprehension. On a written synonym judgement task he performed at a similar level to VS (77%). On a semantic association task, he was better with high compared

to low imageability items (60% vs 40%; PALPA 51), but the difference was not statistically significant (χ^2 (1) = 1.2, p>0.05). On the Pyramids and Palm Trees test of semantics, he performed well with pictures (96%) (50/52) and slightly worse with written words (88%; χ^2(1)=1.2, p=.27).

JH made errors on all speech production tasks, but his naming and reading were worse than repetition. He was poor at naming pictures (0%) and reading items (5%) from the Boston Naming Test (Goodglass, Kaplan, & Weintraub, 1983) and he made neologistic errors. Repetition was better (37%) and errors had higher phonological overlap with their targets (reading vs. repetition, χ^2(1)=16.4, p<.001). JH correctly repeated 53% of nonwords, and 96% of words from PALPA 9. As with VS, JH made fewer errors in spontaneous speech than in speech production tasks. His speech was fluent. There were fewer neologistic errors than in controlled single word tasks and his speech was generally comprehensible.

To summarise, JH made errors on all speech production tasks, but reading and naming were affected more severely than repetition. Words were repeated better than nonwords, but many nonwords were repeated correctly, so the sublexical pathway was not completely impaired. Overall, his performance suggested severe problems with lexical access. His relatively mild comprehension problems were not serious enough to account for his severe impairment in naming and reading.

Patient JW. JW was a 74-yr-old right-handed man who had worked as a telecommunications clerk until he retired at the age of 62. He suffered a left CVA in October 1996. Unfortunately, no details from a CT scan were available to us. Data for the experimental part of this study were collected during the summer of 2001. Results of initial neuropsychological tests are summarised in Table 1.

JW's initial assessment indicated a very mild comprehension problem. He performed well on word–picture matching tasks (PALPA 47 and 48), correctly

matching all the pairs on both the auditory and written versions of the task. On the written synonym judgement task (PALPA 50) he matched more high imageability than low imageability pairs (28/30, 93% vs. 23/30, 77%), but the difference was not significant ($\chi^2(1)=2.1$, p=.15). On the auditory version (PALPA 49) he matched 83% and 73% of pairs ($\chi^2(1)=0.4$, p=.53). He showed a similar pattern on the word semantic association task (PALPA 51), where he correctly identified 73% and 53% of pairs ($\chi^2(1)=0.6$, p=.45). On the Pyramids and Palm Trees test he performed slightly better with three written words (94%) than with three pictures (86%), but the difference was not significant ($\chi^2(1)=1.0$, p=.32).

JW had only mild problems in auditory discrimination. He scored 90% in a minimal pairs task (PALPA 4). He also performed relatively well in auditory lexical decision (PALPA 5; 98% for words and 89% for nonwords). He was poor in naming (PALPA 53; 20% correct), reading (PALPA 31, 26% correct) and repetition (PALPA 9, 45% correct). Errors in repetition and reading were closer to the targets than in naming. He was also poor at nonword reading (PALPA 36) and nonword repetition (PALPA 8), only getting one item correct on each task. Errors were mainly incorrect nonword responses, with a few lexicalisations. JW's spontaneous speech was influenced by his word-finding difficulties and he often made several attempts at a target. Speech was slow, but well articulated. He had problems with sentence construction, which often made his conversation difficult to understand.

To summarise, JW made errors on all speech production tasks, but his errors in naming were more severe than in reading or repetition. Reading and repetition of nonwords was poor. Spontaneous speech was laborious and could be difficult to understand. He had mild problems with comprehension. These initial results suggest JW had a problem at the level of lexical access or phoneme selection. The contrast between naming and repetition was similar to an anomic pattern, suggesting a failure

in lexical access. However, anomic patients typically do not make target-related

nonword errors like JW.

Experimental Study

Our purpose is to compare the errors made by VS, JH and JW in naming,

reading and repetition to establish: 1) if more than one deficit is required to explain

the distribution of phoneme overlap between targets and errors; 2) if *correct* responses

and errors are part of a single distribution of phoneme overlap; and 3) if both lexical

and non-lexical contributions are needed to produce the overlap we see in reading and

repetition.

<u>Method</u>

Our test of picture naming, reading and repetition consisted of 380 items.

Picture naming was done to coloured pictures. Reading was done with individual

words on cards. Repetition was elicited by the experimenter (i.e. not based on a

recording). Items were presented to patients across several sessions, and individual

items were not repeated during a session. The number of items was not always the

same in each session. VS completed picture naming across two sessions. She

performed reading and repetition in the following two sessions, with half of the list

read and half repeated in each session. JH also named pictures across two sessions.

He also read and repeated in two subsequent sessions, with half the list being read and

half repeated on each occasion. JW named items across five sessions. Data for

reading and repetition were collected across two subsequent sessions, with half the list

being read and half repeated in each session. The distributions of frequencies and

lengths for the stimuli are displayed in Figure 2. Frequencies were taken from

CELEX (Baayen, Piepenbrock, & Van Rijn, 1995). Nonword reading and repetition

stimuli were taken from PALPA tests.

Figure 2 about here

Response analysis. Responses from all patients were tape recorded and transcribed in broad IPA. Transcriptions were checked several times against the tape. Only the first complete response was included in the analysis. Fragments with no downturn in intonation that were then included in the following response or that were altered by a single phoneme in the following response were not treated as the first complete response. Each error was classified as a nonword error (neologism), lexical error or other error. Lexical errors included semantic errors, formal errors, semantic+formal errors, unrelated lexical errors and visual errors. Formal errors were errors without an evident semantic relationship that shared at least one phoneme in the same relative position in the syllable, independent of length (e.g. hat > hen). Visual errors were errors where a misidentification of the target picture seemed to have occurred (pill > sweet). Unrelated lexical errors were word responses that shared neither semantics nor phonology with the target. Other errors included ambiguous or mixed errors with a lexical part and a nonword part (e.g. suitcase>seatbluma). Semantic errors, mixed errors with a semantic component, visual errors based on a misinterpretation of a picture, circumlocutions and no responses were excluded from the analyses of target/response overlap.

Note that excluding the errors with semantic and other identifiable sources is necessary to properly evaluate where *nonword* errors arise. We assume that semantic errors arise at word selection (e.g. when the target is unavailable and a semantically-related alternative, active based on the semantic representation, is produced instead; Caramazza & Hillis, 1990; Vigliocco et al., 1997). If these errors are included in the analysis, they will raise the number of unrelated errors, but not because a failure at the level of word selection leads to selection of an unrelated set of phonemes (see the

analysis in Goldrick & Rapp, 2007 for a similar decision and motivation). In fact, the presence of semantic errors (or errors with other clear sources like perseverations) along with nonword errors would only complicate interpretation if a two-source model were favoured. As we will see, this will not be an issue for our results.

We include other types of lexical errors, instead, because their source is more ambiguous. Formally related errors, in particular, have a high level of overlap by definition, and may only be formal errors by chance. We include unrelated lexical errors, which have low overlap, to be conservative about excluding errors. The percentages of different error types included and excluded from the following analyses are reported in Table 2.

Table 2 about here

There are several things to note in Table 2. Responses included in the analysis are a substantial majority of the errors in all cases. The number of included responses for naming is lower (69-74%) because pictures do not identify their word targets unambiguously. Beyond this, there are some important differences by patient or task.

VS makes more errors in repetition than naming and more errors in naming than reading. She makes a substantial number of formal lexical errors in all tasks, and more in reading and repetition than naming, but neologisms are always the largest category of error. VS produces more formal lexical errors in repetition, probably as a consequence of her input difficulties (see Dell, Martin, & Schwartz, 2007 for an analysis showing that input problems increase formal but not nonword errors). We return to this below.

JH gets virtually nothing completely correct in naming and reading, but shows dramatically better performance in repetition. In naming and reading his errors are

virtually always neologisms, but in repetition he makes more formal lexical errors.

JW also shows task differences. He is better at reading and repetition than naming. He makes some formal lexical errors in all tasks.

Across patients, naming produces more *unrelated* lexical errors, reading and repetition, generally more *related* lexical errors. It will be worth keeping this in mind when we discuss summation across routes below.

Model parameters. The statistical models that we compare will be based on target/response overlap. To identify the number of segments that were preserved in responses, we counted the number of phonemes from targets that also appeared in responses (for example, if the target was /pə□pə/ (*paper*), and the response was / pi□pəd/ an overlap of 3/4 was counted and if the response was / pi□də/ the overlap was 2/4). Because errors involving movement of phonemes would still require phonemes to have been selected based on a correctly activated word, we do not consider order in the scoring of overlap. Movement errors, in any case, are not a major feature of the corpus.

The data we used for modelling was a matrix with columns for different word lengths and rows for numbers of phonemes preserved. For example, the first column might tabulate values for 3-phoneme words, with the proportion that had no phonemes preserved in row 1 and proportions for 1, 2, or 3 phonemes preserved in subsequent rows. We compared models based on their ability to predict these overlap matrices for different tasks.

We will describe the model parameters in relationship to the three questions that we set out in the Introduction. The first question asked about the number of sources for target-related and abstruse errors. If there is a single locus, it will be at the level of phoneme selection, and only one probability will determine the level of overlap in all tasks. This is the general segmental probability (*GSP*) that a phoneme

from the target appears in the response because it has been successfully selected.

The same process that produces errors could also occasionally produce correct phonemes by chance (chance segmental probability *ChSP*). If a phoneme is not preserved through phoneme selection, it may still appear in a response because, for example, it is picked at random from a set of weakly activated phonemes. To estimate the probability of a phoneme appearing correctly by chance, we randomly re-paired all patient targets and errors a large number of times (1000 times) and calculated the probability that phonemes were preserved by fitting a binomial probability to the overlap distribution created by re-pairing. The binomial fit to the overlap distribution was always extremely good (adjusted R-squared values in the region of .99). This method of calculating chance takes into account any idiosyncratic bias patients have for particular segments and also the frequency with which segments occur in the stimulus list. Across patients, the chance probability was pretty stable (see Figure 3 for a typical example). Note that this probability was set by re-pairing, and was not an adjustable parameter (it will not, therefore, appear in tables of adjustable parameters, e.g. Tables 4, 5 and 7).

Figure 3 about here

A one locus model. Our simplest model has only two probabilities for modelling overlap in word tasks: an adjustable parameter for general segmental probability correct and a fixed parameter for segmental chance probability. These parameters are used to derive overlap matrixes for the different patients. The probability that different numbers of phonemes appear in a response (*p(phoneme correct)*) is derived using the binomial distribution and the following equation for word repetition, reading and naming:

(1) *p(phoneme correct) = GSP + ((1-GSP) * ChSP)*

This equation derives matrix 1 in Figure 4a.

Figure 4 about here

The equations for *nonword* repetition and reading use different parameters that are specific for each of the non-lexical conversion routes. This reflects the possibility that the non-lexical routes *could* activate phoneme selection to different degrees, depending on the level of damage to these routes. Therefore:

(2) *p(phoneme correct) = NLSP_Rep + ((1-NLSP_Rep)*ChSP)*

for non-word repetition, and

(3) *p(phoneme correct) = NLSP_Rd + ((1-NLSP_Rd * ChSP)*

for non-word reading.

A good fit using these equations would be consistent with a single source model, with errors arising only at phoneme selection.

A two-locus model. The possibility that high overlap and low overlap errors arose at *different* loci was modelled by adding a parameter that allowed a proportion of responses to result from word selection failure (*WSF*). When word selection fails, no phonemes are activated through phoneme selection and the whole response is determined by random activation in the phoneme layer. Therefore, the distribution of overlaps is determined by a mixture of two matrixes: matrix one, described above, plus a second matrix where overlap is only at chance levels (see Figure 4b). The proportion of responses where word selection fails and, therefore, the proportion of responses that use matrix two rather than matrix one is modelled by the WSF parameter. The two matrices are combined as follows:

(4) *matrix3 = WSF * [matrix2$_{chance}$] + (1-WSF)* [matrix1$_{GSP}$]*

Note that the *WSF* parameter modulates the proportion of responses where *all* of the phonemes are selected by chance. This is different from the probability that a single phoneme is correct by chance, which is not a variable parameter, as described above. The *WSF* parameter allows a portion of responses to have very low overlap and others to have higher overlap. At the extremes, this would create a bimodal distribution.

For non-word tasks, a similar logic applies, but, instead of word selection failure, there is a probability that the non-lexical routes for reading or repetition fail for a whole response. Thus, instead of the *WSF* parameter, there are the parameters, *NLF_Rd* and *NLF_Rep* for whole-response failures in reading and repetition. This allows results for word and nonword tasks to be independent. Low overlap could occur in some tasks but not others. The two source models, therefore, have the adjustable parameters: *WSF, NLF_Rd* and *NLF_Rep* in addition to the parameters *GSP, NLSP_Rd, NLSP_Rep* used by the one source models.

Completion boost. Our second question was whether the number of *correct* responses was accurately predicted by the error probability for each phoneme, or if it was higher. A higher value implies a mechanism that increases the probability that *all* phonemes in a word are correct together. To implement this mechanism we need to combine our basic matrix of segmental probabilities (matrix one in Figure 4a) with a matrix (labelled matrix four) where all the phonemes are correct and then modulate how often we use this matrix with a new parameter (Completion Boost: *CB*, see Figure 4c). The equation to combine matrices in this model is:

(5) $matrix5 = CB * [matrix4_{complete\_overlap}] + (1-CB)* [matrix1_{GSP}]$

The model involving a completion boost is the mirror image of a two source model where a *WSF* parameter modulates the contribution of a matrix where phoneme overlap is at chance levels. The CB parameter models the proportion of responses

that are all correct, over and above the number predicted by the general segmental probability. Figure 4c shows a one source model with a completion boost. The question of whether CB is necessary is orthogonal to the question of whether we need a source of low overlap errors at word selection and using both matricies is possible. Using both matrices together would test a model with both word selection failures (generating low overlap responses) and a completion boost (generating completely correct responses).

The way we implemented the completion boost is theoretically neutral regarding the source of this effect. The boost could happen at the level of word selection, through the interaction between word nodes and phonemes, or at the level of phoneme selection only. In the model by Dell (1986), when words reach the activation threshold for selection, they receive an extra activation boost which increases the probability that all phonemes will be selected correctly. If a boost is lacking, unboosted responses will be more prone to phoneme selection errors, and this will separate correct responses from errors.

Alternatively, a completion boost could arise from feedback between words and phonemes. Feedback from correct phonemes will increase correct word activation which, in turn, will produce more activation of correct phonemes. Errors disrupt this reinforcing mechanism and increase separation between correct responses and errors. Finally, a completion boost could arise from a chaining component where activation of phonemes later in a sequence depends on correct activation of previous phonemes. When phonemes are correct at the beginning of a response, this increases the probability that all phonemes will be correct. We will return to these issues in the General Discussion.

A summation model. The last question we introduced above asks if lexical and non-lexical information is combined to produce a response in reading and

repetition. This model includes parameters derived from *non-word* repetition and reading -- *NLSP_Rd* and *NLSP_Rep* --to make predictions for *word* repetition and reading. Thus *p(phoneme correct)* for word repetition with summation*:*

(6) *p(phoneme correct)= GSP + ((1-GSP) * NLSP_Rep) + ((1-GSP)*(1-NLSP_Rep)*ChSP)*

and for word reading with summation:

(7) *p(phoneme correct) = GSP + ((1-GSP) * NLSP_Rd) + ((1-GSP)*(1-NLSP_Rd)*ChSP)*

Equations 6 and 7 implement the idea that if a phoneme has insufficient activation from the word level it may still be selected if it reaches sufficient activation as a result of input coming from the non-lexical route. A summation account allows us to model patients who do very poorly in naming, but better in reading or repetition, as long as their non-word reading or repetition support this. If the summation model is *not* correct, overlap in *word* reading, naming and repetition will be unrelated to what happens in *non-word* tasks.

How the adjustable parameters in our model map onto theoretical questions is summarised in Table 3. Combinations of parameters allow for more complex models (e.g. completion boost with summation). Although the seven adjustable parameters and the fixed chance parameter are set individually for each patient, these parameters are not varied for each task. For example, the general segmental probability has one value for all tasks. However, the non-lexical segmental probability for reading influences only word reading (when there is summation) and non-word reading. The completion boost parameter influences word reading, repetition and naming, but not the non-word tasks. The seven parameters, for the most complex model, and fewer parameters for each simpler model, are set once to predict all tasks.

Table 3 about here

Model selection.   We will compare models using Akaike's Information Criterion (AIC; Akaike, 1973) . The AIC measure balances model fit and parsimony (as counted by the number of parameters in the model). AIC measures fit by calculating -2 * sum of the log likelihood of the data for any specific model. A model that fits well makes the observed data more likely, producing a *lower* log likelihood value (better models have lower AIC values). The measure that implements parsimony is derived from the number of parameters and is added to the likelihood, so models with more parameters have their AIC value increased by more than models with fewer parameters). A more complex model, with more parameters, usually produces a better fit, if only because it allows the model to fit variability that is due to noise. The AIC measure is designed to answer the question of whether the increased fit justifies the increased complexity. This balance is important in order to arrive at models which will generalise well to new data rather than overfitting a particular sample (and producing a model that is poor for inference).

AIC is not a hypothesis testing measure. It does not divide models into "winners" and "losers" that are "significantly" worse. There are well known problems with approaching model comparison in this way (see Burnham & Anderson, 2002). Instead, model selection is based on the different amounts of support that the data provide for each model. It encourages attention to relative differences (in some situations one model is clearly ahead of the others, in others alternatives are nearly equivalent), to the meaning of parameter values and to the importance of parameters across a set of models.   Analyses were carried out with the open source statistical package R (http://www.R-project.org).   The models are ranked in terms of their AIC

scores: The lower the score, the better the model. Differences in AIC, rather than absolute values, are meaningful for comparison. There is no fixed scale or cutoff, but, as a rule of rule-of-thumb, Burnham and Anderson (2002) advise that when two models differ by less than 2, they are substantially equivalent. Differences between 4 and 7 offer considerably more support for the model with the lower AIC value and differences greater than 10 offer clear for the model with the lower AIC value.

Results

The overlap distributions for each patient and each task are shown in Figures 5-7. Appendix 1 lists all the models that were evaluated and their parameters. Tables 4, 5 and 7 list the best fitting models for each patient. The preferred model is highlighted in grey. Poorly fitting models are not included.

Patient VS. The observed and predicted distributions of overlap for patient VS are plotted in Figure 5. Models are reported in Table 4. R-squared values measuring the fit between predicted and observed overlap across all tasks are reported along with AIC and model parameter values. R-squared values for the best fitting models are around 90%, indicating a good fit between predicted and observed values.

Figure 5 and Table 4 about here

Consistent with our previous analysis, a two source account of jargonaphasia was not supported. The second ranked model was only 1.8 AIC units away from the simpler model and so essentially equivalent. The second ranked model *did* include a parameter for word selection failure, but the weight of this parameter was very low (0.6% of responses would be based on purely chance levels of overlap). No model included a substantial contribution from word selection failures. Our new analyses, however, also show that a non-lexical route substantially contributes to reading and to

a lesser extent to repetition. Overlapping phonemes that are *not* produced by *lexical* segmental processing have a probability of nearly 51% of being produced through non-lexical processing in reading. This parameter predicts non-word reading and the increase in overlap for *word* reading above the level predicted by naming.

There is also substantial support for the special status of correct responses. The preferred model predicts that 28% of responses are in a special category where all of the phonemes are correct. Not just the preferred model, but each of the top three models includes this parameter, and each gives it virtually the same value.

Note that there is some evidence of the impact of VS's mild auditory deficit in the model fits. The model over-predicts the number correct for repetition and under-predicts the number correct for reading and naming (see Figure 5). This is what we would expect if repetition is affected by impaired input processing, but the model must predict the number correct for all three tasks together. The impact on the rest of the overlap distribution is less strong. This is consistent with results from Dell et al. (2007) where they found that impaired input processing affected the number correct particularly strongly (patients made more formal errors and got fewer items correct) but had little impact on nonword errors.

Finally, the top two models included a parameter for complete failure of the non-lexical route for nonword reading and repetition (around 24% of responses in each task). Again, VS's input processing problems and a reduced STM span (PALPA digit span of 2) may explain the value of this parameter in repetition. Difficulties with phoneme-grapheme conversion could explain failures in nonword reading. It is worth noting, however, that estimates based primarily on nonword reading or repetition need to be treated with some caution. While the lexical tasks are supported by a large corpus, our nonword data are based on a much smaller sample. Despite some uncertainty about these parameters, it is reassuring to know that changes to them

do not drastically alter the other parameter estimates. Even when failure of the non-lexical routes is excluded (Model 3 in Table 4), the parameters for general segmental probability of error, completion boost, word selection failure and non-lexical contributions to reading and repetition are virtually unchanged.

Patient JH. Observed and predicted overlap distributions are plotted in Figure 6 and model results are presented in Table 5.   Patient JH has strikingly different levels of overlap in repetition compared to reading and naming. The preferred model has only segmental level variables (the general segmental parameter + the summation parameter), but provides a very good fit to his data (adjusted r-squared = .97). The two models with lower AIC do include parameters for a completion boost or word selection failures, but their contributions are very small. Moreover, these models differ from the preferred model by less than 2 AIC units.

Despite JH's very low levels of overlap in naming and reading, the general segmental probability of overlap is not zero, but about 15% (before the chance contribution). Chance overlap on its own fits less well, which is why word selection failure does not account for JH's naming and reading errors. Better performance in repetition than reading is accounted for by differences in non-lexical contributions. The non-lexical probability that segments are preserved is 9% for reading and 81% for repetition (which means repetition is largely non-lexical, but not completely; see columns five and six of the preferred model in Table 5). This provides strong evidence that a non-lexical route contributes to word repetition (Hanley et al., 2004; Hanley & Kay, 1997; Hanley et al., 2002).

Figure 6 and Table 5 about here

Note that our model fits are not an artefact of the way we calculated overlap.

We use the number of shared segments divided by the number of target segments. This measure does not explicitly factor in response length and one may worry that the degree of overlap is inflated as a result. Our measure, however, does not bias our model fits or conclusions. The most fundamental reason is because our chance measure has the same basis. If overlap is higher because response length is not factored out, this is also the case in the chance measure. The systematic contribution *above the level of chance* will not be affected. A second reason that response length is not a concern is because our data are not vulnerable to a length bias. Two patients, VS and JW, do not systematically produce responses that are longer or shorter than the target. In fact, the overlap calculated with respect to target length only (number of phonemes preserved/target length) and with respect to target length plus response length (2*number of phonemes preserved / target length + response length) are nearly identical (Table 6).

JH does produce long responses in naming, but bias would be a more serious concern if the level of overlap in these responses was high. Instead, overlap is low. Moreover, we found that the pattern remains very similar whether or not length is taken into account: overlap is always low but still higher than expected by chance. We calculated bootstrap confidence intervals for the chance data (since they are not normally distributed) and found that observed overlap exceeded chance overlap no matter which measure was used, confirming the model results (Table 6). The reason we have not adopted the target plus response length measure for modelling is that it does not produce a smooth distribution appropriate for a modelling (the numerator takes values that are multiples of two, but the denominator takes continuous values).

Table 6 about here

Patient JW. JW's overlap distributions are plotted in Figure 7 and model results are shown in Table 7. Patient JW combines the main points that we have noted for patients VS and JH. Both lexical and segmental variables are needed to fit his pattern and the resulting model provides a very good fit (adjusted r-squared = .97). Like VS, JW requires a substantial contribution from a completion boost parameter (56% of responses are completely correct). There is a reasonably high segmental probability that phonemes from the target will appear in the response (.3), but, like JH, JW also requires a large contribution from non-lexical processing. Non-word reading adds a contribution of 0.5, non-word repetition adds a contribution of 0.7. These contributions explain JW's better performance in reading and repetition compared to naming.

Figure 7 and Table 7 about here

The models for JW, like those for VS, have parameters for whole response failures from the non-lexical routes. This is clearer in nonword repetition (proportion whole route failures = .3) than in nonword reading (proportion failures = .1). JW had some problems with phonological input processing (PALPA same/different minimal pairs: 82%; minimal pairs with pictures: 85%), but these were milder than for VS (65% and 73% correct respectively). Despite this, JWs estimate for the proportion of non-lexical route whole-response failures is higher. Removing non-lexical failure parameters gave a larger change in AIC for JW than for VS (ΔAIC = 16.5 vs 8.1). JW, then, provides evidence that the non-lexical route may fail, independent of input problems. We note this, however, with the same caution that we mentioned for VS. Nonword results were based on small samples and, thus, they can only be considered preliminary.

Effects of frequency, length and concreteness. Effects of frequency, length and concreteness were examined using binomial regression. In line with our previous modelling of the probability of overlap, we used the proportion of preserved phonemes out of the total number of target phonemes as our dependent variable. Models with a term for concreteness were always better, so we modelled only items with concreteness ratings (smallest N=240). We predicted preserved phonemes starting from a model with terms for length, frequency and concreteness and their interactions and then proceeding to examine simpler models in the set. Since no single model was dominant in the model outcomes (the highest Akaike weight of any model never exceeded 0.45 on a scale of 0-1 and a single dominant model would require a value around 0.9, Burnham & Anderson, 2002), we measured the importance of frequency, length and concreteness by summing the Akaike weights for each model that contained the term of interest. If a variable is present in all models that have good support, the sum of the Akaike weights for all models containing the variable will be high. If a variable is only present in models with poor support, the summed value will be low. These values measure the importance of variables across the full set of models; a more robust method than one based on a single model, when, as in our case, the best model may be essentially equivalent to several other models in the same set.

Table 8 about here

Results reported in Table 8 show that lexical variables (frequency and concreteness) and the variable associated with segment selection (length) wereimportant across the full set of models, consistent with the combination of lexical and non-lexical contributions in our overlap models. Both types of analyses--our

statistical models of the probability of overlap and the binomial regression analysis using phonemes preserved/not preserved—converged. Both showed that lexical and non-lexical factors influenced outcome. Our results were similar to those reported by Nozari et al. (2010) who found that lexical variables (especially frequency) and segmental variables (length) both contribute to modelling nonword errors in a group of patients. They also found that the effect of frequency did not change with task (naming vs. repetition). We reported results for each task, because there was always some effect of task in our data, but when we compared models with and without a task X frequency interaction, this was never significant, consistent with the Nozari et al. results.

Across models, the absolute level of fit was somewhat low (the maximum value for Nagelkerke's R^2 for our models was .125). This is partly due to the nature of binomial data. The models will predict exactly the same proportion of preserved phonemes for a unique level of length, frequency and concreteness, but we clearly do not expect this level of consistency for individual items (e.g. a given *p(phoneme correct)* would not be expected to produce *exactly the same* level of overlap in all words of one length). Despite this, all models that were favoured were clearly different from the null model, except for JH reading, where no model fit well. Across models, length made a greater contribution to repetition than lexical variables in both JH and JW. This is consistent with a non-lexical contribution to repetition. In sum, both lexical and segmental factors were important in our results, consistent with our more specific models of the speech architecture using overlap.

Syllabic simplifications. In a series of studies, Romani, Galluzzi and colleagues have shown that, in Italian, patients who make larger numbers of phonetic errors, and, therefore, have articulatory planning difficulties, also systematically simplify syllabic structures when they make nonword errors. Patients with low

numbers of phonetic errors did not make errors that systematically simplified syllable structure (Galluzzi et al., in press; Romani & Galluzzi, 2005, 2010; Romani et al., 2002). To test whether our patients made errors that systematically simplified syllable structure, we compared the syllabic complexity of targets and errors, using only those errors where the target and error had the same syllable length. This avoids the problem of aligning targets and responses when syllable lengths differ.

Syllables were compared first on CV structure. The best/simplest syllable was considered to be CV. Each change to this template was considered a complication (e.g. CV>CVC, CV>V, CV>CCV etc.). If the CV structure was preserved we then looked at sonority. There is some debate over the details of the definition of sonority (e.g. see Goldsmith, 1990), but there is broad agreement on a sonority scale, where vowels are considered to be highest in sonority, followed by glides (G), liquids (L), nasals (N) and obstruents (O). The simplest syllable should have a maximal rise in sonority in the onset of a syllable, whereas in coda should have a minimal decline (Clements, 1990). Our scoring follows the methods described in the Romani et al. papers cited above.

Table 9 about here

Numbers of simplifications and complications are presented in Table 9. Where decreases in complexity exceeded increases, we tested the significance of the difference using χ^2. There was only one case where simplifications exceeded complications. VS made more simplification than complication errors in reading. However, this pattern was not seen in the other tasks or in the overall totals for VS's errors. Her pattern is not consistent with simplifications that result from a deficit at the level of articulatory planning. Neither JH nor JW simplified syllabic structures

with their errors. These results, together with the results from the analysis of frequency, concreteness and length support a deficit at the level of phoneme selection rather than articulatory planning.

General Discussion

We have presented results from three patients who made primarily non-word errors in speech production tasks, some of which were far from their targets. Considering the mix of target-related and abstruse nonword errors in their responses, these patients would be categorised as jargonaphasics. We have constructed statistical models to account for distributions of target/response overlap in naming, repetition, reading, nonword reading and nonword repetition. We posed three general questions. 1) Was there evidence of two sources for jargonaphasic responses: one with high to medium target/response overlap and one with only chance overlap? 2) Was response overlap continuous between errors and correct responses, such that correct responses could be predicted directly from the overlap in errors? 3) Was there evidence that information was combined across lexical and non-lexical routes for reading and repetition? Using statistical models we obtained clear answers to each of these questions.

Our first question was whether failure to select a word-level representation could produce abstruse nonword errors. The question was *not* whether a failure of word selection could happen at all or produce errors. When word selection fails to pick the intended word, patients may produce semantically related or unrelated *words* or not respond at all. There is no question that these errors occur as a result of failures in word selection. The question addressed by this study was whether *abstruse nonword* errors may *also* result from word selection failures, or whether these errors arise at the subsequent stage of phoneme selection, where a variable number of

phonemes from the selected word may be preserved. We found no support for the hypothesis that low overlap errors resulted from word selection failure, and, therefore, no support for a second source of errors in jargonaphasia. Where word selection failure was implicated in successful models, the proportion of predicted failures was very small and there was always a competing model that was within a difference of 2 AIC units that did not include a parameter for word selection failure.

Our conclusion that patients did not require a word-level source to explain abstruse nonword responses is in agreement with our earlier analysis that only examined VS's neologistic errors (Olson et al., 2007). There, as well, we found that a single source of error at phoneme selection could account for the level of overlap in both abstruse and target-related nonword errors.

Our statistical models were not designed to directly distinguish sources of error at phoneme selection and articulatory planning, but we found no evidence of articulatory difficulties in our patients. They showed no tendency to simplify target phonology, contrary to what has been reported by Romani and colleagues for patients with an articulatory-planning deficit (Galluzzi et al., in press; Romani & Galluzzi, 2005, 2010; Romani et al., 2002).

Moreover, we found lower overlap in naming than in repetition and/or reading. This is consistent with a phoneme selection problem which can benefit from a non-lexical input in repetition and reading, but not with an impairment downstream from phoneme selection, where the probability of correct phoneme production should be the same across tasks.[1] Finally, all patients showed effects of lexical variables

---

[1] Our results also showed that the converse pattern--equivalent performance in naming and reading and/or repetition-- does not, in itself, support a deficit at a level that is more peripheral than phoneme selection. JH, for example, does very poorly in both reading and naming, but this does not mean that he has a deficit that after the level of phoneme selection. A difference with naming is only predicted when non-word reading or repetition are good enough to boost word reading or repetition

(frequency and/or concreteness) across tasks. The influence of these variables should be stronger at the level of phoneme selection than at further levels (see Goldrick and Rapp, 2007 for similar claims but related to formal errors).

We have assigned the errors made by our patients to a level of phoneme selection which is clearly post lexical access. Whether this level should be considered a lexical level is more debatable and beyond the scope of the present study. As highlighted in the Introduction, what is meant by a post-lexical locus is different in different accounts. Goldrick and Rapp (2007) draw a distinction between abstract phonological representations and a subsequent level of phonological encoding where all dimensions are specified. In the Levelt et al., (1999), architecture, syllabification is outside the mental lexicon, and, therefore, post-lexical, while in Romani et al. (2011) syllabification is part of fully specified lexical representations. Although our results do not speak directly to the precise boundary between lexical and non-lexical representations, they indicate a number of properties that a level of phoneme selection must have. Our results show an influence of lexical dimensions such as frequency and concreteness, which indicate that this level is either part of lexical representation proper, or close enough to be influenced through cascade. Our results, however, also show non-lexical influences, through summation of information from sub-lexical routes. Our proposal, where there is a combination of lexical and non-lexical influences at this level, is more efficient than alternatives that have additional abstract phonological levels. It means, however, that this level would not be exclusively "lexical." We can't rule out the possibility that there is more than one abstract phonological level--an initial lexical level and a subsequent post-lexical, but still abstract, level—but our results do not require this.

---

through summation. Instead, his non-lexical route for reading functions poorly, making little contribution to boost word reading.

Although we didn't find that errors in our patients arose at the level of word selection or articulatory planning, we cannot exclude these levels as *potential* sources of error in jargonaphasia. The results from Italian patients (Romani & Galluzzi, 2005, 2010; Romani et al., 2002), in fact, show clearly that there are systematic differences between kinds of patients who make nonword errors, and these result from different levels of deficits, including articulatory planning (see also Buchwald & Miozzo, 2011). Similarly, it may be too soon to say that deficits at word selection can *never* contribute to jargonaphasia, even though our results and those of Schwartz et al. (2004) did not support this source. What is clear is that quantitative modelling is necessary to decide the issue and the simple observation of some near and some far responses is not sufficient. Our patients were superficially quite different, but on this question their model results were similar.

A second important aspect of our results was the need to combine contributions from lexical and non-lexical routes in order to model response overlap in reading and repetition. Models where responses were based exclusively on a lexical or a non-lexical route were clearly inferior to models where lexical and non-lexical information was combined (Alario et al., 2003; Bi et al., 2007; Hanley et al., 2004; Hanley & Kay, 1997; Hanley et al., 2002; Hillis & Caramazza, 1991, 1995; Miceli et al., 1994; Nozari et al., 2010; Tree et al., 2005; Ward et al., 2000). A summation account is also consistent with other aspects of our patients' performance. In particular, JW makes a number of *unrelated* lexical errors in naming, but almost exclusively *formally related* lexical errors in reading and repetition, as would be expected if non-lexical information about the form of targets constrains the set of lexical errors that are possible. The summation account has often been supported by noting a higher number of semantic errors in naming (see the reading and naming performance of the original Hillis & Caramazza, 1991, patient along with others cited

above) and the increased percentage correct in reading and repetition (e.g. Hanley & Kay, 1997; Hanley et al., 2002). Here we contribute to this support by modelling the overlap between targets and responses.

Our evidence for summation converges nicely with results from a different type of analysis reported by Nozari et al. (2010). They found, in an analysis of 59 aphasic patients, that frequency had a comparable effect on the likelihood of making an error in repetition and naming, pointing to a lexical component of both tasks. They also found, however, that nonword errors were less likely in repetition than naming, pointing to a contribution from the non-lexical route that boosts accuracy. Our analysis and theirs, one based on analysis of overlap between targets and responses and the other based on an analysis of number correct, converge on the same conclusion: lexical and non-lexical information is combined to determine a response.

Our model assumes that input is combined constructively, but not destructively. That is, if a phoneme is not activated enough via lexical processing it may reach threshold through non-lexical input, but non-lexical processing does not *interfere* with lexical processing. Theoretically, a non-lexical input could add noise to the phoneme level and disrupt, as well as improve, performance. In our patients, performance was generally *better* in the tasks with non-lexical input which means they are not a strong test of this possibility. VS's repetition was worse than naming, but this is more likely to reflect poor auditory input than disruption from a non-lexical route.

Similarly to us, Nozari et al. (2010) did not find disruptive effects of summation. They were measuring the influence of summation on frequency effects and found no attenuation of the frequency effect when there was a non-lexical contribution to recognition. This is consistent with a summation that leaves the lexical contribution intact. Exploring other models of summation would be

particularly useful with future patients that have different characteristics (good input

processing but poor non-lexical processing) and with a larger sample of nonword data.

We noted in the introduction that Baron et al. (2008) suggested that

individuals may differ in whether they combine lexical and non-lexical information.

An extension of the modelling approach introduced here could allow this to be

explored more fully. First, it would be possible to see, with a larger sample of

patients, if the population divides into patients who require summation and patients

who do not. Second, for patients that require summation, it would be possible to add

a parameter that weights the contribution of lexical and non-lexical information in

reading and repetition. Looking at the population of weights, it will be clear whether

weights are concentrated at only the high and low ends of the specturm (associated

with a binary distinction), or distributed widely (associated with variable weighting).

Finally, a novel and important aspect of these results was the larger than

expected numbers of completely correct responses in two of the three patients.[2] The

best model for both VS and JW included a large contribution of what we have called a

completion boost that allowed correct responses to exceed the number predicted by

the level of overlap in the errors.[1] We have mentioned different possible

interpretations of a completion boost. One possibility is that it occurs at the level of

word selection, as hypothesized in the original Dell (1986) model. When words are

selected from a semantic specification, they receive a jolt of activation (a selection

---

[2] Note that given our measure, it is possible for responses to preserve all the target phonemes but not be correct (when errors rearrange phonemes or insert erroneous phonemes). There were, however, only a small numbers of these errors. The great majority of cases with complete overlap were correct responses (VS naming: 115/120; VS reading: 187/197; JW naming: 127/132; JW reading: 268/278; JW repetition: 314/318). The exception was for VS's repetition errors, where a somewhat larger proportion of these responses were errors (53/74 all correct). Even in this case, however, the number of words completely correct was still a clear majority of the errors. Since the model must use a single parameter to set the number of completely correct responses in all tasks together, the parameter is driven mostly by responses which are, indeed, completely correct.

boost) which further separates them from competitors and reduces the possibility that phonemes from competing words are wrongly selected. The separation of correct responses from errors could arise if some responses get the selection boost (responses destined to be correct), but others, as a result of damage, do not. Without the selection boost, cascading activation would still activate phonemes for output, but more errors would be expected. This account does re-introduce a factor that influences errors at the level of word selection (presence or absence of the completion boost), but it is a different source than word selection failures as the errors themselves arise at phoneme selection. A failure of word selection means *no* phonemes are activated successfully, leading to only chance-level overlap. A failure of a completion boost increases the probability of errors, but would not lead to only chance-level overlap and abstruse errors. In all other accounts of a completion boost, the determinants of nonword errors are at the level of phoneme selection and increased numbers of all-correct responses arise from interaction or chaining mechanisms.

Feedback interaction between word and segmental levels (e.g. see Dell, 1986; Rapp & Goldrick, 2000) would increase probability that *all* phonemes are correct. When the correct segments are activated, they feed activation back to the correct item at the word level which, in turn, produces a corresponding increase of activation at the segmental level. When incorrect segments are active at the segmental level, instead, they feed activation back to words that are potential competitors of the target, which, in turn, activate segments that are not in the target word, increasing the likelihood of errors. Interactive feedback instantiates a recurrent loop that pushes apart the level of overlap found in correct responses and errors.

Another similar mechanism is compound chaining. Here, each segment produced correctly provides a context that helps to drive correct production of the next segment (see discussion in Botvinick & Plaut, 2006; Goldberg & Rapp, 2008).

When an error occurs, this disrupts the context that should trigger the next segment, making a subsequent error more likely. If another error follows, the context is further disrupted leading to an even greater chance of error. Both feedback and chaining introduce an all-or-none element into production. If everything proceeds smoothly, a correct response is produced. Once errors begin to happen, further errors are more likely, introducing a gap between the overlap seen in errors and correct responses.

How to distinguish these different accounts of a completion boost is not easy. For example, our data show a stronger separation between correct responses and errors in word than nonword tasks. VS (in reading) and JW (in reading and repetition) have many completely correct responses in word tasks and fewer in nonword tasks. This, however, is consistent with all of the described accounts because existing words will provide a stronger selection jolt, they are necessary for a feedback loop between words and phonemes and stored representations are required for chaining. Future studies should use different kinds of evidence to adjudicate between these accounts, looking at positional effects or effects of previous errors. For example, chaining predicts that probability correct would increase with number of previously correct phonemes and decrease as a function of previous errors. Feedback predicts an influence of neighbourhood density on errors. What is important here is that we have demonstrated the need for some type of mechanism which increases the probability of completely correct responses.

In sum, our data have shown the utility of putting theoretical alternatives into an explicit form that can be captured by statistical models and then using model selection procedures to select the best alternative. This process provides more information and gives a more nuanced view of the data than binary hypotheses testing procedures. We found no evidence that jargonaphasic errors came from two sources: one based on random phonological material based on failures of word selection and

another based on failures of phoneme selection. A single segmental source was sufficient to explain both low and high overlap errors.

Our data also clearly required that information from lexical and non-lexical processing be combined to produce a response. We showed that nonword performance could not be judged only by the relatively crude measure of percentage correct. Even when nonword reading or repetition was not particularly good, as measured by percentage correct, nonword target/response overlap helped to explain why word reading and repetition produced higher overlap than naming.

Finally, we showed that errors and correct responses did not form a continuous distribution. Correct responses were too numerous to be from the same overlap distribution as errors. This requires a 'completion' effect which pushes apart the overlap in errors and completely correct responses. This effect could arise from a jolt at lexical selection, word/segment interaction or compound chaining. More data and theoretical exploration will be needed to clarify and distinguish these alternatives.

References

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.

Alario, F. X., Schiller, N. O., Domoto-Reilly, K., & Caramazza, A. (2003). The role

of phonological and orthographic information in lexical selection. *Brain and*

Language, 84(3), 372-398.

Anderson, S. R. (1974). *The organisation of phonology*: Academic Press.

Baayen, R. H., Piepenbrock, R., & Van Rijn, H. (1995). *The CELEX Lexical*

Database (CD-ROM): Linguistic Data Consortium.

Badecker, W., Miozzo, M., & Zanuttini, R. (1995). The 2-stage model of lexial

retrieval -- Evidence from a case of anomia with selective preservation of

grammatical gender. *Cognition, 57*, 193-216.

Baron, R., Hanley, J. R., Dell, G. S., & Kay, J. (2008). Testing single- and dual-route

computational models of auditory repetition with new data from six aphasic

patients. *Aphasiology, 22*(1), 62-76.

Bi, Y. C., Han, Z. Z., Weekes, B., & Shu, H. (2007). The interaction between

semantic and the nonsemantic systems in reading: Evidence from Chinese.

Neuropsychologia, 45(12), 2660-2673.

Bock, K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology, 23*, 45-

93.

Botvinick, M. M., & Plaut, D. C. (2006). Short-term memory for serial order: A

recurrent neural network model. *Psychological Review, 113*(2), 201-233.

Buchwald, A., & Miozzo, M. (2011). Finding levels of abstraction in speech production: Evidence from sound production impairment. *Psychological Science, 22,* 1113-1119.

Buckingham, H. W. (1981). Where do neologisms come from? In J. Brown (Ed.), *Jargonaphasia* (pp. 39-62). New York: Academic Press.

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* (2 ed.). New York: Springer.

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*(1), 177-208.

Caramazza, A., & Hillis, A. E. (1990). Where do semantic errors come from? *Cortex, 26*(1), 95-122.

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*(3), 309-343.

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: Cambridge University Press.

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*(1), 204-256.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93*, 283-321.

Dell, G. S., Martin, N., & Schwartz, M. F. (2007). A case-series test of the interactive two-step model of lexical access: Predicting word repetition from picture naming. *Journal of Memory and Language, 56*(4), 490-520.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review, 104*(4), 801-838.

Dunn, L. M., Dunn, L. M., Whetton, C. W., & Pintilie, D. (1982). *British Picture Vocabulary Scale* (1 ed.). Windsor: NFER-Nelson.

Foygel, D., & Dell, G. S. (2000). Models of impaired lexical access in speech production. *Journal of Memory and Language, 43*(2), 182-216.

Funnell, E. (1996). Response biases in oral reading: An account of the co-occurrence of surface dyslexia and semantic dementia. *Quarterly Journal of Experimental Psychology Section A-Human Experimental Psychology, 49*(2), 417-446.

Galluzzi, C, Bureca, I., Guariglia, C., & Romani, C. (in press). Phonological simplifications, apraxia of speech and the interaction between phonological and phonetic processing. *Neuropsychologia.*

Garrett, M. F. (1975). The analysis of sentence production. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 9). New York: Academic Press.

Garrett, M. F. (1980). Levels of production in sentence processing. In B. Butterworth (Ed.), *Language production (Volume 1)* (pp. 177-220). London: Academic Press.

Goldberg, A. M., & Rapp, B. (2008). Is compound chaining the serial-order mechanism of spelling? A simple recurrent network investigation. *Cognitive Neuropsychology, 25*(2), 218-255.

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: Basil Blackwell.

Goodglass, H., Kaplan, E., & Weintraub, S. (1983). *The Boston Naming Test*. Philadelphia: Lea & Febiger.

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*(2-4), 147-158.

Hanley, J. R., & Kay, J. (1997). An effect of imageability on the production of phonological errors in auditory repetition. *Cognitive Neuropsychology, 14*(8), 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*(3), 193-206.

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*(3), 662-720.

Henaff Gonon, M. A., Bruckert, R., & Michel, F. (1989). Lexicalization in an anomic patient. *Neuropsychologia, 27*, 391-407.

Hillis, A. E., & Caramazza, A. (1991). Mechanisms for accessing lexical representations for output - Evidence from a category-specific semantic deficit. *Brain and Language, 40*(1), 106-144.

Hillis, A. E., & Caramazza, A. (1995). Converging evidence for the interaction of semantic and sublexical phonological information in accessing lexical representations for spoken output. *Cognitive Neuropsychology, 12*(2), 187-227.

Howard, D., & Franklin, S. (1988). *Missing the meaning: A cognitive neuropsychological study of the processing of words by an aphasic patient.* Cambridge, MA: MIT Press.

Howard, D., & Gatehouse, C. (2006). Distinguishing semantic and lexical word retrieval deficits in people with aphasia. *Aphasiology, 20*(9), 921-950.

Howard, D., & Patterson, K. (1992). *Pyramids and Palm Trees: A Test of Semantic Access from Words and Pictures*. Bury St Edmunds, Suffolk: Thames Valley Test Company.

Kay, J., Lesser, R., & Coltheart, M. (1992). *Psycholinguistic Assessments of Language Processing in Aphasia*. Hove: Psychology Press.

Kempen, G., & Huijbers, P. (1983). The lexicalization process in sentence production and naming - indirect election of words. *Cognition, 14*(2), 185-209.

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*(1), 75-95.

Laganaro, M. (2012). Patterns of impairments in AOS and mechanisms of interaction between phonological and phonetic encoding. *Journal of Speech Language and Hearing Research, 55*(5), S1535-S1543.

Levelt, W. J. M. (1989). *Speaking*. Cambridge, Mass.: 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.

Marshall, J. (2006). Jargon aphasia: What have we learned? *Aphasiology, 20*(5), 387-410.

Miceli, G., Capasso, R., & Caramazza, A. (1994). The interaction of lexical and sublexical processes in reading, writing and repetition. *Neuropsychologia, 32*(3), 317-333.

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*. London: Erlbaum Associates Ltd.

Miozzo, M., & Caramazza, A. (1997a). On knowing the auxiliary of a verb that cannot be named: Evidence for the independence of grammatical and phonological aspects of lexical knowledge. *Journal of Cognitive Neuroscience, 9*, 160-166.

Miozzo, M., & Caramazza, A. (1997b). Retrieval of lexical-syntactic features in tip-of-the-tongue states. *Journal of Experimental Psychology-Learning Memory and Cognition, 23*(6), 1410-1423. doi: 10.1037//0278-7393.23.6.1410

Moses, M. S., Nickels, L. A., & Sheard, C. (2004). Disentangling the web: Neologistic perseverative errors in jargon aphasia. *Neurocase, 10*(6), 452-461.

Nozari, N., Kittredge, A. K., Dell, G. S., & Schwartz, M. F. (2010). Naming and repetition in aphasia: Steps, routes, and frequency effects. *Journal of Memory and Language, 63*(4), 541-559.

Olson, A. C., Romani, C., & Halloran, L. (2007). Localizing the deficit in a case of jargonaphasia. *Cognitive Neuropsychology, 24*(2), 211-238.

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review, 107*(3), 460-499.

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(1), 109-124.

Romani, C., & Galluzzi, C. (2005). Effects of syllabic complexity in predicting

accuracy of repetition and direction of errors in patients with articulatory and

phonological difficulties. *Cognitive Neuropsychology, 22*(7), 817-850.

Romani, C., & Galluzzi, C. (2010). Aphasic errors and phonological universals:

Contrasting patterns in patients with and without articulatory difficulties. In C.

Papagno (Ed.), *48th Academy of Aphasia Proceedings* (6 ed., pp. 80-81).

Amsterdam: Elsevier.

Romani, C., Galluzzi, C., Bureca, I., & Olson, A. (2011). Effects of syllable structure

in aphasic errors: Implications for a new model of speech production.

Cognitive Psychology, 62(2), 151-192.

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*(4), 541-567.

Schwartz, M. F., Wilshire, C. E., Gagnon, D. A., & Polansky, M. (2004). Origins of

nonword phonological errors in aphasic picture naming. *Cognitive*

Neuropsychology, 21(2-4), 159-186.

Tree, J. J., Kay, J., & Perfect, T. J. (2005). "Deep" language disorders in nonfluent

progressive aphasia: An evaluation of the "summation" account of semantic

errors across language production tasks. *Cognitive Neuropsychology, 22*(6),

643-659.

Vigliocco, G., Antonini, T., & Garrett, M. F. (1997). Grammatical gender is on the tip

of Italian tongues. *Psychological Science, 8*, 314-317.

Ward, J., Stott, R., & Parkin, A. J. (2000). The role of semantics in reading and spelling: evidence for the 'summation hypothesis'. *Neuropsychologia, 38*(12), 1643-1653.

Zorzi, M., Houghton, G., & Butterworth, B. (1998). Two routes or one in reading aloud? A connectionist dual-process model. *Journal of Experimental Psychology-Human Perception and Performance, 24*(4), 1131-1161.

Table 1.

Results of initial neuropsychological testing for patients VS, JH and JW.

	VS		JH		JW	
	N	%	N	%	N	%
phonological discrimination						
Same-different minimal pairs-words (PALPA 2)	47/72	(65)	70/72	(97)	59/72	(82)
Minimal pairs-pictures (PALPA 4)	29/40	(73)	37/40	(93)	34/40	(85)
single word comprehension						
Auditory lexical decision (PALPA 5)	122/160	(76)	145/160	(91)	149/160	(93)
Visual lexical decision (PALPA 25)	91/120	(76)				
Spoken word-picture matching (PALPA 47)	38/40	(95)	31/40	(78)	40/40	(100)
Written word-picture matching (PALPA 48)	38/40	(95)	35/40	(88)	40/40	(100)
Pyramids & Palm Trees-pictures	50/52	(96)	50/52	(96)	45/52	(87)
Pyramids & Palm Trees-words	50/52	(96)	46/52	(88)	49/52	(94)
Written synonym judgements (PALPA 50)	47/60	(78)	46/60	(77)	51/60	(85)
Auditory synonym judgements (PALPA 49)	29/60	(48) *chance*			47/60	(78)
speech production						
Picture naming (PALPA 53)	18/40	(45)			8/40	(20)
Word reading (PALPA 31)	12/80	(15)			21/80	(26)
Word repetition (PALPA 9)	20/80	(25)	77/80	(96)	36/80	(45)
Boston Naming Test - naming			0/40	(0)		
Boston Naming Test - reading			3/40	(8)		
Boston Naming Test - repetition			22/60	(37)		
Nonword reading (PALPA 36)	0/24	(0)			1/24	(4)
Nonword repetition (PALPA 8)	1/30	(3)			1/30	(3)
Nonword repetition (PALPA 9)			42/80	(53)		

Table 2.

Number correct and errors of different types in naming, reading and repetition for patients VS, JH and JW. Numbers in parentheses are percentages.

	Patient VS						Patient JH						Patient JW					
	naming		reading		repetition		naming		reading		repetition		naming		reading		repetition	
correct	127	(33)	187	(49)	54	(14)	1	(0)	5	(1)	244	(64)	127	(33)	268	(71)	314	(83)
% correct as a proportion of included errors		(45)		(49)		(15)		(0)		(1)		(64)		(49)		(71)		(84)
included errors																		
formal lexical	37	(10)	90	(24)	116	(31)	10	(3)	19	(5)	51	(13)	36	(9)	52	(14)	30	(8)
unrelated lexical	33	(9)	8	(2)	15	(4)	23	(6)	14	(4)	1	(0)	21	(6)	1	(0)	0	(0)
neologisms	85	(22)	93	(24)	186	(49)	247	(65)	331	(87)	83	(22)	77	(20)	57	(15)	32	(8)
included subtotal	282	(74)	378	(99)	371	(98)	281	(74)	369	(97)	379	(100)	261	(69)	378	(99)	376	(99)
excluded errors																		
semantic or semantic+other	49	(13)	2	(1)	4	(1)	6	(2)	2	(1)	0	(0)	62	(16)	1	(0)	1	(0)
visual or visual+phonological	12	(3)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	5	(1)	0	(0)	0	(0)
perseveration	23	(6)	0	(0)	4	(1)	0	(0)	9	(2)	0	(0)	15	(4)	0	(0)	0	(0)
no response	6	(2)	0	(0)	0	(0)	2	(1)	0	(0)	0	(0)	29	(8)	0	(0)	0	(0)
ambiguous/mixed/circumlocution	8	(2)	0	(0)	1	(0)	91	(24)	0	(0)	1	(0)	8	(2)	1	(0)	3	(1)
excluded subtotal	98	(26)	2	(1)	9	(2)	99	(26)	11	(3)	1	(0)	119	(31)	2	(1)	4	(1)

Table 3.

List of adjustable parameters, the unit they apply to and the hypothesis that each tests.   Phoneme parameters are in white and whole response parameters are in grey.

Parameter	Description	Relevant unit	Hypothesis tested	Model needing this parameter
GSP	general per phoneme probability of error	phoneme	Segmental parameter used in all models, but main parameter in a one-locus model	All; but esp one locus model
WSF	Proportion of responses where there is word selection failure	whole response	Is there a second source of error? Errors have only chance overlap with target	Two locus model (word tasks)
NLF_Rd	Proportion of responses where there is failure of non-lexical reading route	whole response	Is there a second source of error for non-word reading?	Two locus model (for nonword reading)
NLF_Rep	Proportion of responses where there is failure of non-lexical repetition route	whole response	Is there a second source of error for non-word repetition?	Two locus model (for nonword rep)
CB	Proportion of responses where all phonemes are correct as a result of a completion boost	whole response	Are there more than the expected number of correct responses because of a completion boost?	Completion boost
NLSP_Rd	Per phoneme probability of error -- non-lexical route reading	phoneme	When this parameter contributes to _word_ reading, it tests a summation account. Otherwise, it only applies to NW reading.	Summation-reading
NLSP_Rep	Per phoneme probability of error -- non-lexical route for repetition	phoneme	When this parameter contributes to _word_ repetition, it tests a summation account. Otherwise, it only applies to NW repetition	Summation-repetition

Table 4.

Model fit and selection results for Patient VS. Model selection was based on ΔAIC and models are ordered by their differences from the minimal model.   Adjusted r-squared values were based on observed and predicted overlap in tables summarising the number preserved target segments for each word length. Models from that do not appear here had a ΔAIC value greater than the value for the worst fitting model shown here. The preferred model is highlighted in grey.

Patient VS

Model	AIC	ΔAIC	Adjusted R-squared	GSP	NLSP_Rd (summation in reading)	NLSP_Rep (summation in rep)	WSF	CB (completion boost)	NLF_Rd (NL route failure-reading)	NLF_Rep (NL route failure-rep)
Summation for reading and rep Completion boost Failure of non-lexical routes for reading/rep	399.7	0.0	0.9021	0.2609	0.5085	0.1681	-	0.2768	0.2356	0.2542
Summation for reading and rep Completion boost Second locus at word selection failure Failure of non-lexical routes for reading/rep	401.5	1.8	0.9033	0.2656	0.5160	0.1664	0.0061	0.2747	0.2470	0.2580
Summation for reading and rep Completion boost	407.8	8.1	0.8972	0.2598	0.4932	0.1395	-	0.2822	-	-
Completion boost Second Locus at word selection failure Failure of non-lexical routes for reading/rep	500.4	100.7	0.8107	0.4682	-	-	0.0643	0.2987	0.0027	0.3664

Table 5.

Model fit and selection results for Patient JH. Models from Table 4 that do not appear here had a ΔAIC value greater than the value for the worst fitting model shown here. The preferred model is highlighted in grey.

Patient JH

Model	AIC	ΔAIC	Adjusted R-squared	GSP	NLSP_Rd (summation in reading)	NLSP_Rep (summation in rep)	WSF	CB (completion boost)	NLF_Rd (NL route failure-reading)	NLF_Rep (NL route failure-rep)
Summation for reading and rep Completion boost	306.5	-	0.9739	0.1492	0.0894	0.8109	-	0.0073	-	-
Summation for reading and rep Second locus at word selection failure	307.8	1.3	0.9732	0.1556	0.0874	0.8119	0.0018	-	-	-
Summation for reading and rep	307.9	1.4	0.9727	0.1553	0.0872	0.8108	-	-	-	-
Summation for reading and rep Completion boost Failure of non-lexical routes for reading/rep	310.5	4.0	0.9739	0.1485	0.0904	0.8111	-	0.0071	0.0000	0.0000
Summation for reading and rep Second locus at word selection failure Failure of non-lexical routes for reading/rep	311.8	5.3	0.9729	0.1560	0.0869	0.8113	0.0017	-	0.0000	0.0000
Summation for reading and rep Completion boost Second locus at word selection failure Failure of non-lexical routes for reading/rep	312.4	5.9	0.9744	0.1498	0.0892	0.8120	0.0016	0.0073	0.0000	0.0000
Completion boost Second locus at word selection failure Failure of non-lexical routes for reading/rep	1144.7	838.2	0.4547	0.6336	-	-	0.3923	0.1387	0.3486	0.0000

Table 6.

Comparing overlap calculated in two different ways. Confidence intervals for chance overlap and observed mean overlap calculated with regard to target length only (shared length/target length) and target length plus response length (2 * shared length) / (target length + response length).

Patient	Overlap method	Naming Confidence interval for chance overlap lower upper	Observed mean	Reading Confidence interval for chance overlap lower upper	Observed mean	Repetition Confidence interval for chance overlap lower upper	Observed mean
VS	Target length only	0.13 - 0.16	0.53	0.13 - 0.16	0.82	0.13 - 0.16	0.53
	Target + response length	0.12 - 0.15	0.52	0.12 - 0.16	0.86	0.12 - 0.15	0.52
JH	Target length only	0.14 - 0.18	0.27	0.15 - 0.18	0.34	0.15 - 0.18	0.86
	Target + response length	0.12 - 0.15	0.17	0.12 - 0.15	0.3	0.12 - 0.15	0.86
JW	Target length only	0.12 - 0.16	0.67	0.12 - 0.16	0.86	0.12 - 0.16	0.92
	Target + response length	0.12 - 0.16	0.67	0.12 - 0.15	0.86	0.12 - 0.15	0.92

Table 7.

Model fit and selection results for Patient JW. Models from Table 4 that do not appear here had a ΔAIC value greater than the value for the worst fitting model shown here. The preferred model is highlighted in grey.

Patient JW

Model	AIC	ΔAIC	Adjusted R-squared	GSP	NLSP_Rd (summation in reading)	NLSP_Rep (summation in rep)	WSF	CB (completion boost)	NLF_Rd (NL route failure-reading)	NLF_Rep (NL route failure-rep)
Summation for reading and rep Completion boost Failure of non-lexical routes for reading/rep	379.7	0.0	0.9689	0.3029	0.4993	0.7044		0.5628	0.0988	0.2916
Summation for reading and rep Completion boost Second locus at word selection failure Failure of non-lexical routes for reading/rep	381.3	1.6	0.9699	0.3063	0.5037	0.7179	0.0025	0.5581	0.1054	0.3036
Summation for reading and rep Completion boost	396.2	16.5	0.9594	0.3001	0.4566	0.5968		0.5976		
Completion boost Second Locus at word selection failure Failure of non-lexical routes for reading/rep	477.4	97.7	0.9270	0.6064			0.0502	0.6254	0.0393	0.0951

Table 8. Summed Akaike weights across all models, predicting preserved phonemes using word frequency, concreteness and phoneme length. Weights are based on AIC values for each model and summed across the full model set. The scale is from 0-1 and higher weights indicate stronger support. Nagelkerke's R^2 values are pseudo-r squared measures of the fit between models and observed data.

Task	VS				JH				JW			
	Frequency	Length	Concreteness	Nagelkerke's R^2 for best model	Frequency	Length	Concreteness	Nagelkerke's R^2 for best model	Frequency	Length	Concreteness	Nagelkerke's R^2 for best model
Naming	0.993	0.648	0.692	0.069	0.492	0.762	0.809	0.025	0.824	0.707	0.757	0.052
Reading	0.996	0.995	0.957	0.125	-	-	-	0.006	0.949	0.984	0.603	0.079
Repetition	0.849	0.829	0.908	0.057	0.494	0.992	0.576	0.034	0.477	0.961	0.889	0.069

Table 9. The percentage of errors that decrease, increase or do not change syllable complexity. Only responses with the same number of syllables as targets were scored. χ^2 values are reported when simplifications (decreases) exceeded complications (increases).

		Increase		Equal		Decrease		χ^2	
VS	Naming	25	(34)	36	(49)	12	(16)	.	
	Reading	22	(17)	69	(53)	40	(31)	5.22	p=.02
	Repetition	75	(36)	83	(39)	53	(25)	.	
	Total	122	(29)	188	(45)	105	(25)	1.27	p=.26
JH	Naming	34	(57)	12	(20)	14	(23)	.	
	Reading	50	(33)	44	(29)	57	(38)	0.46	p=.50
	Repetition	14	(12)	91	(78)	12	(10)	.	
	Total	98	(30)	147	(45)	83	(25)	.	
JW	Naming	15	(21)	43	(60)	14	(19)	.	
	Reading	14	(17)	50	(62)	17	(21)	0.29	p=.59
	Repetition	9	(20)	26	(59)	9	(20)	.	
	Total	38	(19)	119	(60)	40	(20)	0.05	p=.82

Appendix 1. The set of models that were evaluated for each patient. Ticks indicate parameters that were estimated.

General segmental prob	Non-lexical segmental prob - reading	Non-lexical segmental prob - repetition	Word selection failure	Completion boost	Non-lexical whole resp failure - reading	Non-lexical whole resp failure - repetition
√	√	√	√	√	√	√
√	√	√		√	√	√
√	√	√	√		√	√
√	√	√		√		
√	√	√	√			
√			√	√	√	√
√				√	√	√
√			√		√	√
√	√	√				
√						

Figure Captions

Figure 1. The speech production architecture that forms the basis for the model tests.

Figure 2. Distributions of word frequency and length for stimulus items.

Figure 3. Typical distribution of chance overlap. This is taken from VS results, but others were very similar (JH binomial probability = .144, JW binomial probability = .148).

Figure 4. One-locus, two-locus and completion boost models and their parameters.

Figure 5. Overlap distributions for patient VS. The model values are taken from the preferred model (see Table 5).

Figure 6. Overlap distributions for patient JH. The model values are taken from the preferred model (see Table 6).

Figure 7. Overlap distributions for patient JW. The model values are taken from the preferred model (see Table 7).
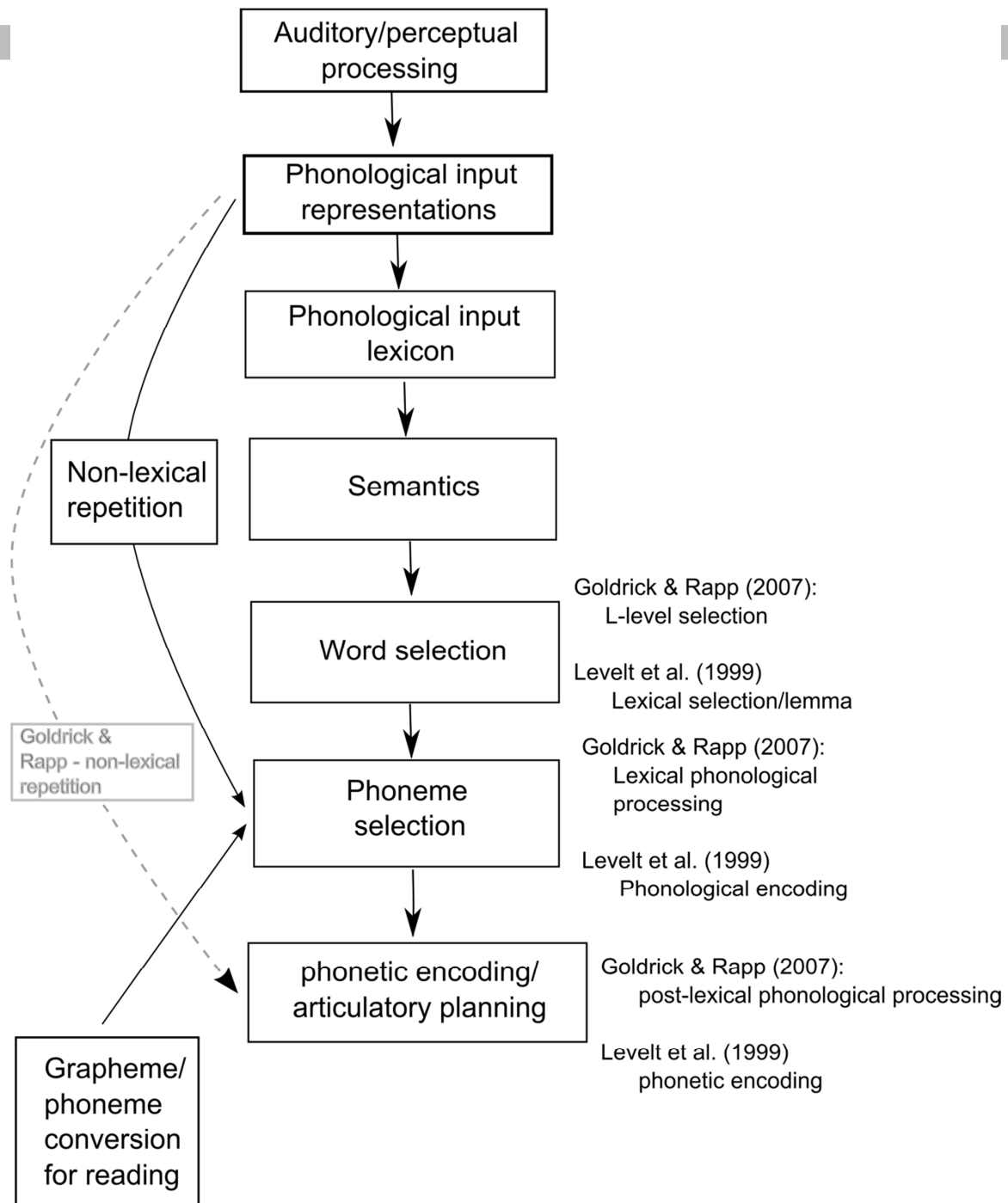
Figure 1.

Figure 2.
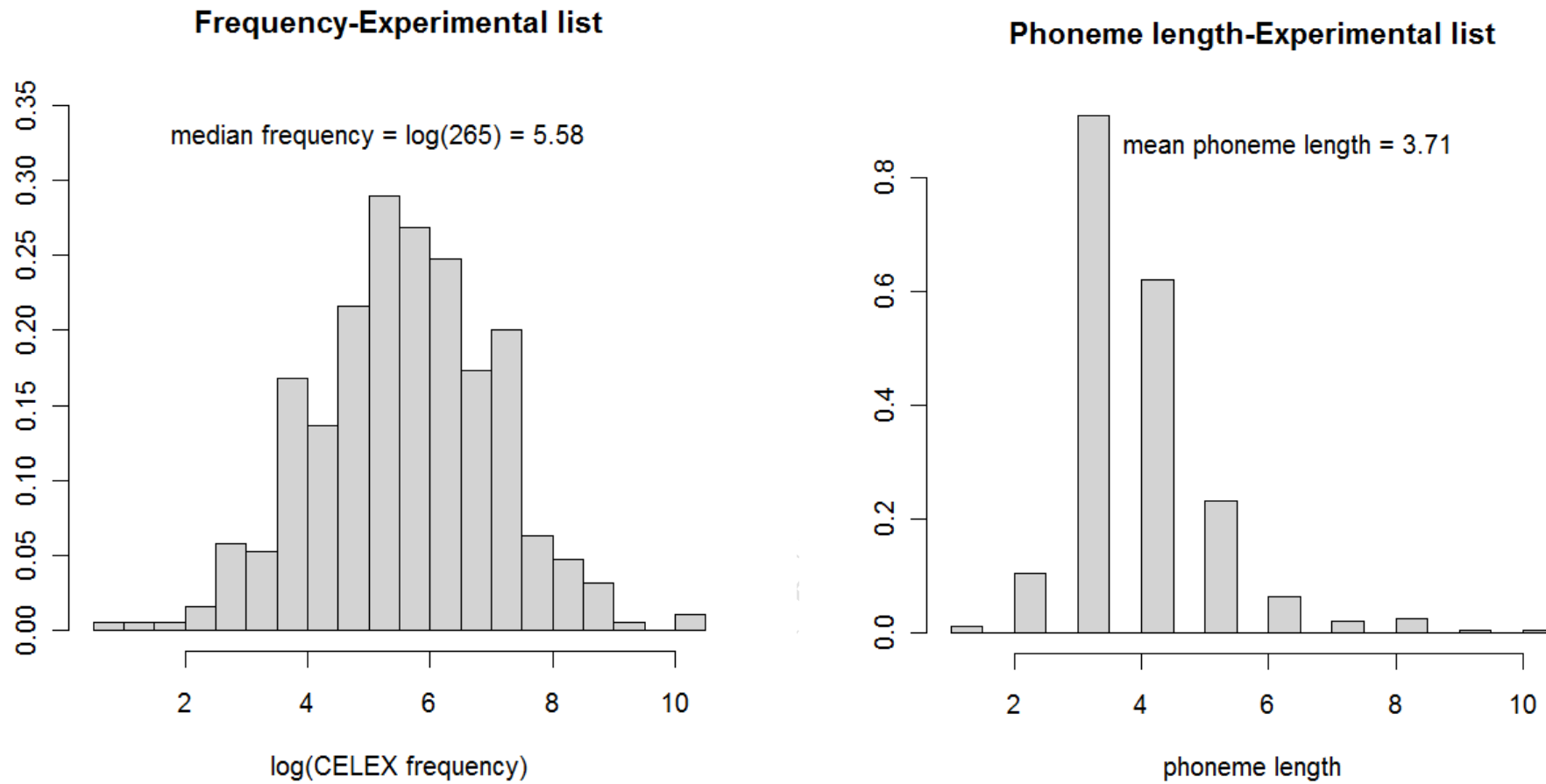
Figure 3.
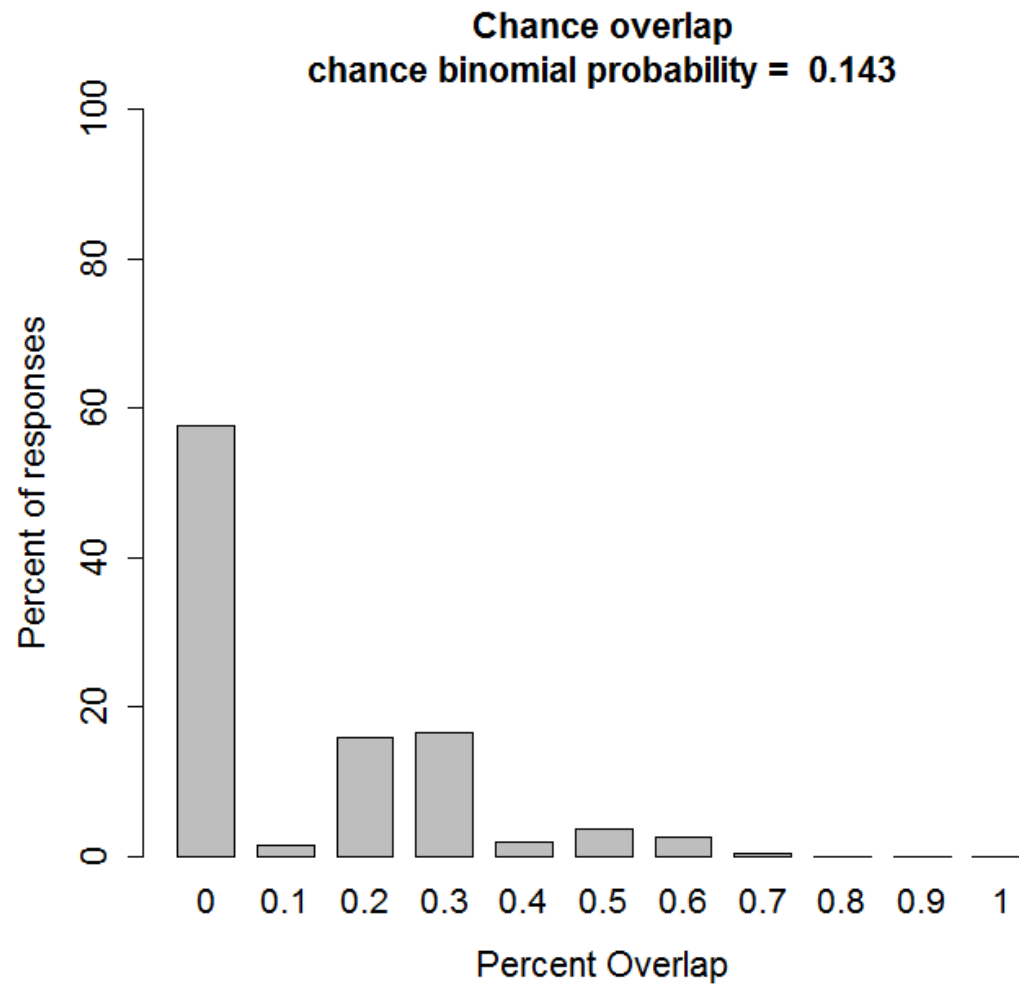
Predicted overlap matrix

(a) one-locus model:

General Segmental Probability
of phoneme preserved

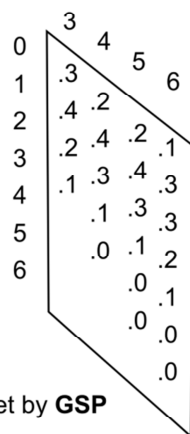**adjustable parameter *GSP***

binomial
distribution

Matrix 1
General segmental overlap
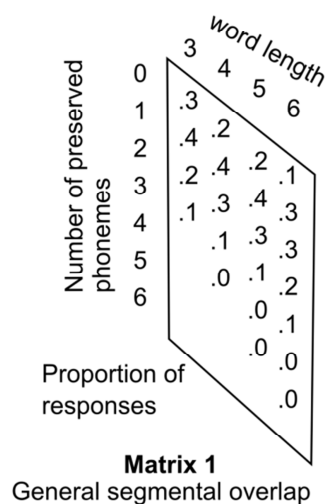
(b) two-locus model:
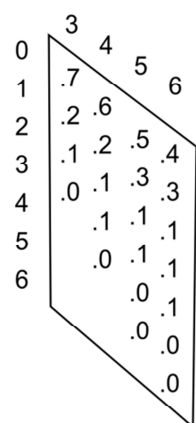
| Proportion of word selection failures | | |
| adjustable parameter *WSF* | | |

X
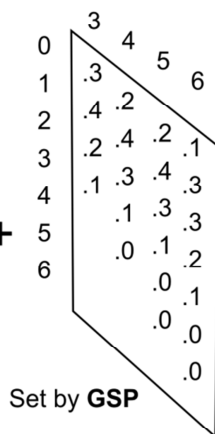
+

Set by **GSP**

X | 1- WSF |

=

Matrix 2
Chance only overlap

Matrix 1
General segmental overlap

Matrix 3
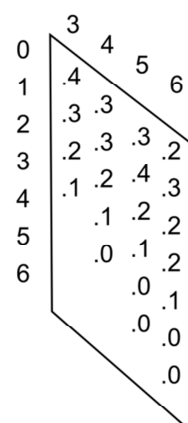
(**c**) **completion boost model:**
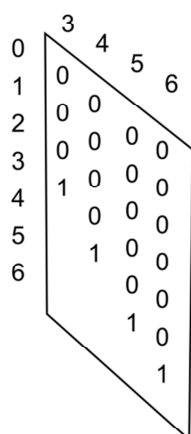
| Proportion of items that get a completion boost | | |
| adjustable parameter *CB* | | |

X

+

Set by **GSP**

X | 1- CB |

=

Matrix 4
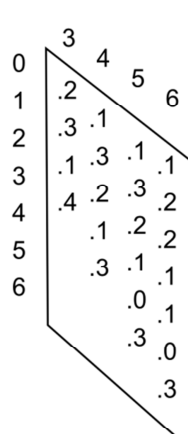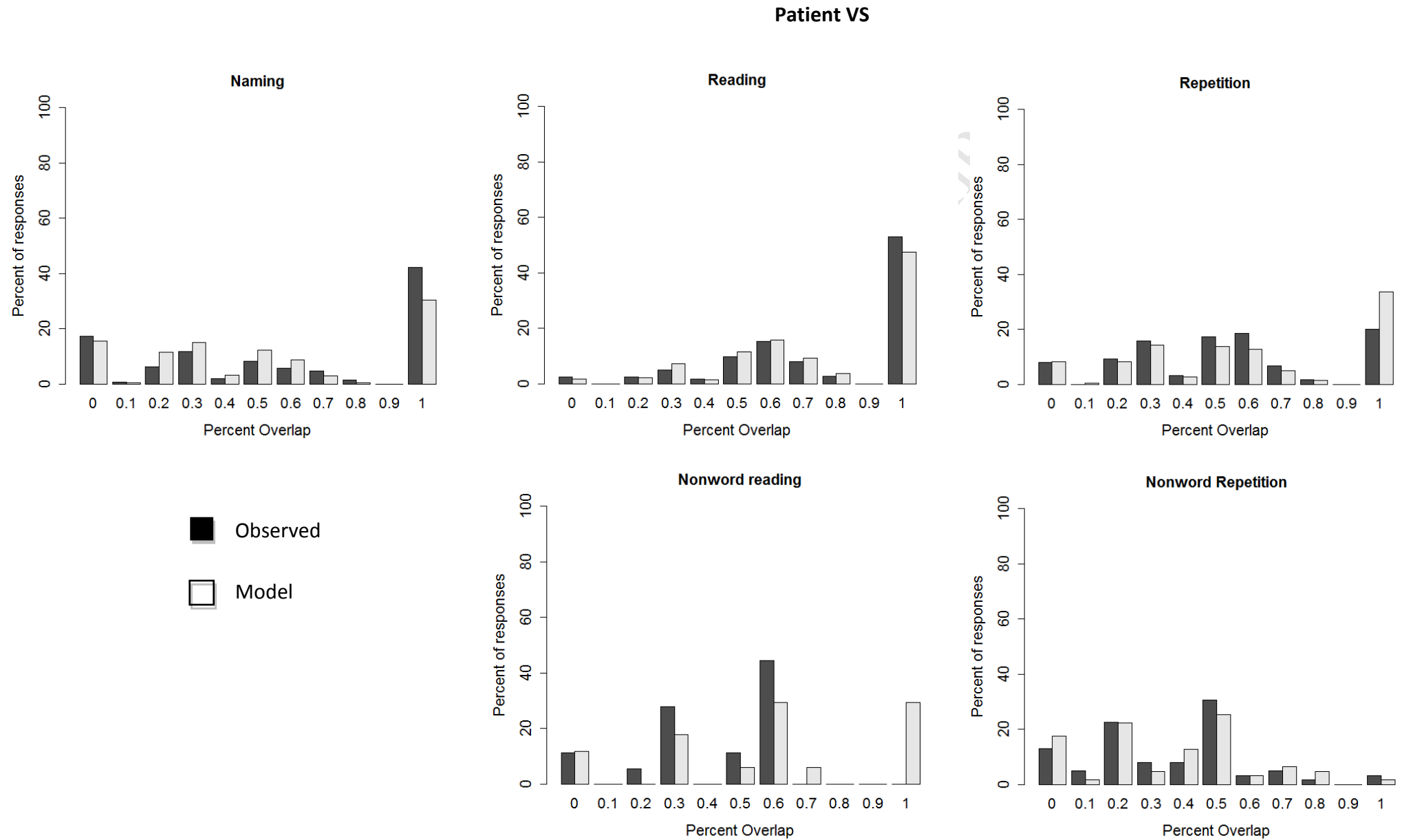all items correct

Matrix 1
General segmental overlap

Matrix 5

Figure 4.

Figure 5.

Figure 6.

Figure 7.