Published in final edited form as:

J Mem Lang. 2010 February 1; 62(2): 113–134. doi:10.1016/j.jml.2009.11.008.

Mrs. Malaprop's Neighborhood: Using Word Errors to Reveal Neighborhood Structure

Matthew Goldrick^{1,2}, Jocelyn R. Folk³, and Brenda Rapp¹

¹ Department of Cognitive Science, Johns Hopkins University

² Department of Linguistics, Northwestern University

³ Department of Psychology, Kent State University

Abstract

Many theories of language production and perception assume that in the normal course of processing a word, additional non-target words (lexical neighbors) become active. The properties of these neighbors can provide insight into the structure of representations and processing mechanisms in the language processing system. To infer the properties of neighbors, we examined the non-semantic errors produced in both spoken and written word production by four individuals who suffered neurological injury. Using converging evidence from multiple language tasks, we first demonstrate that the errors originate in disruption to the processes involved in the retrieval of word form representations from long-term memory. The targets and errors produced were then examined for their similarity along a number of dimensions. A novel statistical simulation procedure was developed to determine the significance of the observed similarities between targets and errors relative to multiple chance baselines. The results reveal that in addition to position-specific form overlap (the only consistent claim of traditional definitions of neighborhood structure) the dimensions of lexical frequency, grammatical category, target length and initial segment independently contribute to the activation of non-target words in both spoken and written production. Additional analyses confirm the relevance of these dimensions for word production showing that, in both written and spoken modalities, the retrieval of a target word is facilitated by increasing neighborhood density, as defined by the results of the target-error analyses.

Many theories of language processing assume that during the course of word perception and production lexical neighbors of a target word become active. Understanding which words are co-activated along with the target is of importance because it will shed light on the nature of the processes and representations involved in lexical access. Typically, neighbors are assumed to be non-target words that are related in form to the target. For instance, in many theories of perception, incoming acoustic or visual information serves to activate sublexical representations of word form (e.g., segments) which, in turn, activate all lexical representations that are at least partially consistent with the incoming form information (e.g., speech perception: Luce & Pisoni, 1998; written word recognition: Andrews, 1997). For example activation of the initial phoneme/d/spreads to all associated lexical representations: <DOG>.

Address for Correspondence: Matthew Goldrick, Department of Linguistics, Northwestern University, 2016 Sheridan Rd. Evanston, IL 60208 USA, Phone: (847) 467-7092, Fax: (847) 491-3770, matt-goldrick@northwestern.edu.

Portions of this work were presented at annual meetings of the Academy of Aphasia (Denver, 2001; Chicago, 2004) the Psychonomic Society (Orlando, 2001) and Architectures and Mechanisms of Language Processing (Turku, Finland, 2007).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

<DAD>, <DOT>, etc. Along similar lines, in word production, neighbors are also predicted to become active in theories in which form representations feed back to activate their associated lexical representations (e.g., speech production: Dell, 1986; written production: McCloskey, Macaruso, & Rapp, 2006). For example, the lexical representation <DOG> activates the phoneme/d/which, via feedback, activates <DAD>, <DOT>, etc. Note that according to theories that assume gradient activation, word representations are not restricted to being simply on or off but they can be partially activated. In these architectures neighbor vs. non-neighbor is best conceived of not as a binary distinction but as a dimension along which non-target words vary (stronger vs. weaker neighbors).

The properties of these partially activated non-target words are of theoretical interest because they reflect the processing principles and representational structure of the word production system. For example, some theories of speech production have claimed that the long-term memory representation of a word's form consists of segment representations in which identity and position are represented in a unitary manner (e.g., /k/-onset, /t/-coda; Dell, 1986). This representational structure predicts that only those non-target words that share segments in the same position will become active during production (e.g., <COPE> is a neighbor of <CAT>, but <PICK> is not). In contrast, other theories have claimed that segment identity and position are represented separately (e.g., Warker & Dell, 2006). These theories predict that non-target words with shared segments in different positions would also become partially active (e.g., <PICK> has a/k/that it shares with target <CAT>). Since the composition of the set of neighbors is determined by the representational and processing structure of the language processing system, an understanding of the characteristics of partially activated words provides a critical window into lexical representation and processing.

Most studies investigating neighborhood characteristics adopt the strategy of assuming that certain characteristics are shared between targets and neighbors and then test for the processing consequences of the proposed characteristic/s. For example, many studies of orthographic processing have assumed that visual recognition processes are structured such that all and only those words related by the substitution of a single letter within a position are partially activated during perception (e.g., Coltheart, Davelaar, Jonasson, & Besner, 1977). Similarly, other studies have assumed that in spoken perception (Vitevitch & Luce, 1998, 1999) and production (e.g., Vitevitch, 1997, 2002b) processes are structured such that all and only those non-target words related by the substitution, addition, or deletion of a single phoneme are partially activated. (Note that both of these definitions assume categorical definitions of neighborhood; see Luce & Pisoni, 1998, for a similar approach utilizing a gradient neighborhood metric in spoken perception.) After assuming this type of form-based neighborhood structure, these studies examine if the number of neighbors a word has (neighborhood density) correlates with some measure of word processing (e.g., does having a large number of neighbors facilitate or impair identification of a target word?). To the degree that such correlations are found, one gains confidence in the legitimacy of the definition of neighborhood structure assumed by the analysis. However, the neighborhood features assumed in different approaches may be highly correlated; in that case, a direct comparison of the proposals is necessary to ensure that a particular feature set provides the best available account of neighborhood structure (e.g., Davis & Taft, 2005; Newman, Sawusch, & Luce, 2005; Perea, 1998; Vitevitch, 2002a).

An alternative strategy that does not involve an a priori definition of neighborhood properties is to infer neighborhood structure by examining the properties of form-related lexical errors (e.g., Fay & Cutler, 1977; see below). These errors are non-semantic word errors in written (e.g., "lyric" written to dictation as LURID) or spoken ("mitten" named from a picture as "muffin") production. Such word substitutions have been reported in a number of situations: spontaneously (e.g., in malapropisms), in experimental error-inducing tasks and in cases of acquired neurological impairment (e.g., aphasia). When such an error is observed, it can be

assumed that the error-word was active during the processing of the target (and mis-selected); that is, that the error is a neighbor of the target. If this assumption is correct, the relationships between form-related lexical errors and their targets should reveal the principles governing word activation in the production system. In this manner, we are able to infer rather than assume definitions of neighborhood.

The work reported here examines form-related lexical errors arising in spoken and written language production. We first establish that these errors arise during the process of retrieving word-form information from long-term memory (i.e., in the phonological or orthographic lexicon). We then quantify the relationship between targets and errors along a number of feature dimensions. For each feature dimension we evaluate the significance of the observed relationship by means of a novel simulation-based statistical procedure that compares the observed values to those expected by chance under a range of hypotheses that make different assumptions about neighborhood properties. In this way we begin to identify the independent contribution to neighborhood activation of different lexical and form-related properties. Finally, we perform post-hoc analyses to examine how neighborhood density—defined on the basis of the significant neighborhood properties—influences word production accuracy.

Word production architecture

Before proceeding it will be helpful to lay out a framework within which the data and methodological points can be discussed (see Figure 1). Most theories of single word production assume that at least two processing stages are required to map from semantic to sub-lexical phonological or orthographic representations (Garrett, 1980; Levelt, 1992). The first stage is meaning-driven, corresponding to the selection of a syntactically appropriate word representation to express an intended meaning (e.g., in picture naming, mapping the semantic features [furry, feline, domesticated] to the noun <CAT>). Following Rapp and Goldrick (2000) we use the term ‘L-level’ to refer to this level of word representation and the process as L-level selection. These representations may or may not be specific to the written or spoken modality (see Caramazza, 1997; Roelofs, Meyer, & Levelt, 1998, for discussion). Be that as it may, the focus here is on the second form-based processing stage where the sound or letter information corresponding to the selected word is retrieved from long-term memory (e.g., the L-level representation <CAT> is mapped to its constituent phonemes /k//æ//t/ or its constituent letters C-A-T). We will refer to this stage of mapping from L-level to sub-lexical form representations as phonological or orthographic spell-out (referred to in Goldrick & Rapp, 2007, as lexical phonological processing). It should be noted that some theories posit additional stages intervening between L-level selection and phonological or orthographic spell-out (e.g., morphological encoding; Levelt, Roelofs, & Meyer, 1999).

Furthermore, additional post-lexical processes and representations are assumed in both spoken and written word production. With regard to spoken word production, it is widely assumed that the phonological information stored in long-term memory is subject to additional (post-lexical) phonological as well as articulatory processing prior to speech execution (e.g., in Levelt et al., 1999, such information is subject to prosodification, phonetic encoding, and articulatory planning; for a review of various proposals, see Goldrick & Rapp, 2007). In addition, one or more short-term memory or buffering processes are assumed to be involved in the course of post-lexical phonological processing (for a recent review see Shallice, Rumiat, & Zadini, 2000). Similar buffering processes are assumed to play a role in written word production; these maintain the activation of orthographic representations retrieved from the lexicon and are responsible for the serial selection of letters. These buffering processes are followed by more peripheral post-lexical processing responsible for letter-shape or letter-name production (see Tainturier & Rapp, 2001, for a recent review).

Most current theories assume that phonological and orthographic spell-out processes are implemented by spreading activation through weighted connections linking L-level representations of lexical items (or morphemes) with sub-lexical representations of phonological/orthographic form in long-term memory. For example, in Dell, Schwartz, Martin, Saffran, & Gagnon's (1997) proposal, a single 'lemma' node represents the lexical item <DOG>; it is connected to phoneme units representing /d/ in onset position, /a/ in vowel position, and /g/ in coda position. In such a system, retrieval of information from memory corresponds to the spreading of activation from nodes representing lexical items to phonological segment nodes. While most theories posit localist representations within such a system (Levelt, 1999), distributed representations are assumed by Plaut & Shallice (1993) in the context of semantically-mediated reading and by Graham, Patterson & Hodges (2000) for spelling.

Many theories make the additional assumption that activation spreads bi-directionally: not only do L-level units activate phonological segment nodes, but segment nodes in turn reactivate L-level representations (e.g., Dell, 1986 for spoken production; McCloskey et al., 2006 for written production; but see Levelt et al., 1999). (Note that it has been suggested that the segment-word feedback activation recruits the perception system; see Rapp & Goldrick, 2004, Roelofs, 2004a,b, for discussion.) This allows the lexical representations of words that share segments with the target (formal neighbors) to become active (e.g., during processing of target <DOG>, segment nodes /d/ and /a/ can activate the non-target lexical representation <DOT>; Vitevitch, 2002b; Dell & Gordon, 2003). Critically, if feedback mechanisms are absent, the lexical representations of formal neighbors will not become active; activation will be purely top-down, reflecting only semantic and grammatical properties of the target. Although some feedback mechanism is required to activate the lexical representations of formal neighbors, it has been argued that there are significant restrictions on the strength of feedback (Rapp & Goldrick, 2000). However, even restricted feedback is sufficient to allow for the activation of the lexical representation of formal neighbors, particularly during spell-out processes (for further discussion, see Goldrick, 2006; for supporting simulation results, see Rapp & Goldrick, 2000).

In addition to formal neighbors, most production theories also assume that L-level representations of semantic neighbors of the target become active during the course of L-level selection (e.g., Caramazza & Hillis, 1990; Roelofs, 1992). In theories with cascading activation between the L-level and subsequent levels of representation, these semantic neighbors will activate their corresponding sub-lexical form representations, allowing them to influence spell-out processes (e.g., Rapp & Goldrick, 2000; see below for further discussion).

Neighborhood properties investigated by examining target-error relations

The strategy pursued in this study involves examining the similarity relationships between target words and errors, assuming that words produced in error are co-active (neighbors) with the target. There have been a number of studies that have taken this approach and have reported target error similarity along a number of dimensions. The dimensions of similarity that have been examined can be characterized as lexical (frequency and grammatical category) or form-based (position-specific and/or position independent segmental overlap, segmental or syllabic length).

With regard to lexical properties, some studies of spontaneous speech errors (del Viso et al., 1991; Kelly, 1986) as well as errors arising in cases of acquired spoken language deficits (Blanken, 1990, 1998; Gagnon et al., 1997; Martin, Dell, Saffran & Schwartz, 1994) have reported that error responses are biased to be higher in frequency than their targets. However, a number of other studies have found no such effect (spontaneous errors: Harley & MacAndrew, 2001; Vitevitch, 1997; experimentally induced errors: Dell, 1990; aphasic errors

in spoken production: Best, 1996; written production: Romani, Olson, Ward, & Ercolani, 2002). In addition, a large number of studies have reported very high rates of syntactic category preservation in spoken errors produced both spontaneously (Abd-El-Jawad & Abu-Salim, 1987; Arnaud, 1999; Berg, 1992; del Viso et al., 1991; Fay & Cutler, 1977; Fromkin, 1971; Garrett 1975, 1980; Harley, 1990; Harley & MacAndrew, 2001; Leuninger & Keller, 1994; Nooteboom, 1969; Rossi & Defare, 1995; Silverberg, 1998; Stemberger, 1985) as well as subsequent to neurological impairment (Best, 1996; Berg, 1992; Blanken, 1990; Dell et al., 1997; Gagnon et al. 1997; Martin et al., 1994; but see Blanken, 1998). Some studies of written errors have reported that the tendency to preserve grammatical category is weaker in written than in spoken production—and is perhaps non-significant (spontaneous: Hotopf, 1980; aphasic errors: Romani et al. 2002).

With regard to form-based dimensions of similarity, many studies have noted that, in speech production, non-semantically-related word errors tend to share phonological structure with the target (spontaneous errors: Arnaud 1999; Dell & Reich, 1981; del Viso et al., 1991; Harley & MacAndrew, 2001; Leuninger & Keller, 1994; aphasic errors: Best 1996; Blanken, 1990). Romani et al. (2002) noted a similar high degree of form overlap in spelling errors (see also McCloskey et al., 2006). While word production theories distinguish between position-specific vs. position-independent segmental representation, to our knowledge no studies have explicitly contrasted position-specific vs. position-independent overlap to determine which better accounts for the data. Nonetheless, many studies have noted that phonological overlap between targets and errors appears to favor the initial portions of the word, especially the first position (spontaneous errors: Harley, 1984; Fay & Cutler, 1977; aphasic errors: Gagnon et al., 1997; Martin et al. 1994; but see Miller & Ellis, 1987, and Harley, 1990). Some studies have suggested that the final position exhibits a higher rate of target-error overlap as well (spoken: Hurford, 1981; Tweney, Tkacz, & Zaruba, 1975; Silverberg, 1998; written: Hotopf, 1980; Wing & Baddeley, 1980). Another finding that has been reported is that form-related errors tend to share the length of targets. In spoken production, analyses have focused on preservation of the number of syllables in the target (spontaneous errors: Tweney, Tkacz, & Zaruba, 1975; Harley & MacAndrew, 2001; Fay & Cutler, 1977; Gagnon et al., 1997; Leuninger & Keller, 1994; Silverberg, 1998; errors in aphasia: Best, 1996; Biran & Freidmann, 2005; Blanken, 1990; Miller & Ellis, 1987). In written word production, Romani et al. (2002) reported that the written errors of a dysgraphic individual tended to preserve the number of syllables in the target.

In addition to investigations of dimensions of target-error similarity, studies have examined the relationship between accuracy/speed of word production and a word's neighborhood density. These studies have all defined density using purely form-based definitions (e.g., using density measures proposed by Coltheart et al., 1977, or Vitevitch & Luce, 1998). One finding is that pictures whose names are in dense vs. sparse lexical neighborhoods are named more quickly (Baus, Costa, & Carreiras 2008; Vitevitch, 2002b; Vitevitch, Armbrüster, and Chu, 2004; but see Jescheniak & Levelt, 1994, for a null result, and Vitevitch & Stamer, 2006, for a reversal of this effect in Spanish speakers). Roux & Bonin (2009) report that oral spelling latencies are shorter for words in dense vs. sparse neighborhoods. It is also reported that words in dense vs. sparse neighborhoods are less susceptible to speech errors (spontaneous: Vitevitch, 1997; experimentally induced: Vitevitch, 2002b; Stemberger, 2004; aphasic errors: Best 1995; Gordon 2002; Kittredge, Dell, Verkuilen, Schwartz, 2008; but see Newman & German, 2002, 2005) and are less likely to give rise to tip-of-the-tongue states (Harley & Bown, 1998; German & Newman, 2004; Vitevitch & Sommers, 2003). Similar results have been reported in recent studies of dysgraphia (Brunsdon, Coltheart, & Nickels 2005; Sage & Ellis 2004; see also Sage & Ellis, 2006). Going beyond simple form overlap, some studies have shown that frequency-weighted density also facilitates lexical phonological processing over and above overall neighborhood density (Baus et al., 2008; Newman & German, 2002; Vitevitch & Sommers 2003; but see Gordon, 2002). Vitevitch et al. (2004) also reported that the number

of neighbors sharing the initial portion of the target ('onset density') also influences picture naming independent of the influence of overall neighborhood density; however, this effect is inhibitory rather than facilitatory.

Challenges for target-error analysis

As we have already indicated, the guiding assumption of research that considers the relationship between targets and errors as a means of elucidating neighborhood structure is that non-semantically-related word errors arising in the course of phonological or orthographic spell-out occur because their L-level and/or form (e.g., phoneme, grapheme) representations were activated during the course of retrieving the target's form representation from long-term memory. Two important challenges facing this type of research are: (1) establishing that, in fact, errors arise during spell-out and (2) quantitatively evaluating the degree of overlap between targets and errors.

Establishing the processing locus of word errors

If the errors that are being analyzed in a particular study do not arise in the process of retrieving form information (e.g., phonemes or graphemes) from long-term memory, they will not be unambiguously informative regarding these processes and may instead reflect the structure/content of other cognitive processes.

In previous research this issue was addressed in a variety of ways. For example, exclusion of form-related word errors arising prior to spell-out has been primarily accomplished by excluding any errors that are semantically related to the target, regardless of degree of form overlap. As discussed above, most theories of production assume a semantically-based processing stage that selects a particular L-level node to express an intended concept. In single word production tasks such as picture naming, evidence suggests that purely formally related words are not strongly activated during this process. Although feedback from form to L-level serves to boost the activation of mixed semantic+form-related neighbors, it is not strong enough to significantly boost the activation of purely formally related words (see Goldrick, 2006, for a review). (See Ferreira & Humphreys, 2001, for a discussion of this issue in connected speech.) Therefore, restricting analyses to production errors that do not overlap semantically with the target (regardless of any formal overlap) should eliminate error responses arising prior to spell-out. (Note, however, that this criterion is conservative, as semantic errors can arise during lexical phonological/orthographic spell-out; we return to this issue below).

A more difficult problem has been to eliminate errors arising from processes subsequent to spell-out. As noted above, it is generally assumed that within both written and spoken modalities additional form-based processing is required before words can be articulated orally or manually. Errors arising at these later/more peripheral processing levels are likely to result in productions that are similar in the form to the target and, by chance, these productions may be words. For example, altering a single distinctive feature of "tab" can produce a nonword "pab," but a similarly close alteration of the target could produce the word "cab."

To exclude word errors arising from these later processes, many researchers have adopted form-based exclusionary criteria for errors to be analyzed. However, these criteria are currently insufficient for identifying word errors arising from a specific processing locus. For example, Fay & Cutler (1977) excluded all errors involving a single phoneme difference from the target word (e.g., single phoneme exchanges, anticipations, perseverations, omissions, additions). It is highly likely that this criterion is too inclusive as well as too exclusive. There is no reason to assume that processes subsequent to lexical spell-out cannot generate errors affecting multiple segments (e.g., two simultaneous feature specification failures could turn "cat" into "gad") and this criterion would fail to exclude these errors. Conversely, we cannot assume that

single phoneme errors could not arise during lexical spell-out (e.g., instead of “cat”, the phonological representation for “cad” could be accessed); this criterion would fail to include such errors.

Similar interpretative issues face Butterworth’s (1992) strategy of using the absence vs. presence of response variability to distinguish deficits affecting the lexicon vs. subsequent processes. For example, given the assumption that lexical phonological/orthographic spell-out is based on spreading activation (susceptible to stochastic noise), it is unclear why deficits to this process could not result in response variability (e.g., when random noise disrupts retrieval of the first phoneme of “cat,” retrieval processes may sometimes activate “bat” yet other times retrieve “hat”; see Rapp & Caramazza (1993) for further discussion of the diagnostic value of variability in the context of semantic deficits).

Therefore, form- or variability-based criteria cannot be relied on to distinguish form-related word errors arising within lexical phonological/orthographic spell-out versus those arising during subsequent processing. With no other means of establishing the locus of word errors, this issue makes it difficult to interpret the results of studies of word errors in spontaneous speech (Abd-El-Jawad & Abu-Salim, 1987; Arnaud, 1999; Berg, 1992; del Viso et al., 1991; Fay & Cutler, 1977; Garrett, 1980; Harley & MacAndrew, 2001; Leuninger & Keller, 1994; Nooteboom, 1969; Silverberg, 1998; Stemberger, 1985) and as well as those produced by relatively unselected groups of individuals with neurological impairments affecting some aspect of production processing (Dell et al., 1997; Gagnon, Schwartz, Martin, Dell & Saffran, 1997).

In this paper, we examine the performance of several individuals with neurological impairment. A considerable body of work has shown that brain damage can cause relatively specific functional deficits to particular stages of cognitive processing. We use functional theories of spoken and written production to generate “diagnostic criteria” that allow us to identify deficits affecting different stages of word production. This provides a principled method, that is independent of the error characteristics themselves, to establish that the non-semantic word errors produced by these individuals arose in the course of lexical phonological or orthographic spell-out.

Evaluating target-error similarity

Researchers have typically inferred that if targets and errors are highly similar along some dimension, then this dimension is represented or plays some role in word production. A key and complex challenge to this research strategy is to determine what counts as “highly similar.” For example, if 80% of targets and word errors share grammatical category, should we consider this to be a high or low degree of similarity? It should be clear that the significance of any such result can only be evaluated relative to an appropriate measurement of chance or a baseline. In this regard, it is critical to understand (although it not usually explicitly stated) that baseline or chance rates represent the rates expected under some null hypothesis. Therefore, the determination of chance/baseline requires articulating the alternative hypotheses under consideration. In this example, one hypothesis is that grammatical category is represented and participates in lexical spellout; the null hypothesis is that it does not. An appropriate baseline, therefore, should represent the degree of grammatical category overlap that is expected by chance when a word error is produced in a system in which language production process does not incorporate grammatical category.

One common method used to estimate baseline similarity rates is to evaluate the similarity between semantically related errors and their targets (e.g., Fay & Cutler, 1977; Harley & MacAndrew, 2001; Silverberg, 1998). This practice is based on two assumptions, both of which are problematic. First, such analyses assume the relationship between form and meaning is

completely arbitrary (deSaussure, 1910). However, some statistical analyses have suggested that semantically similar words are more likely to share phonological structure than words that are semantically dissimilar (Rapp & Goldrick, 2000; O'Toole, Oberlander, & Shillcock, 2001, Tamariz, 2005). Other research has suggested robust connections between particular sound sequences and meanings (see Vigliocco & Kita, 2006, for discussion and Bergen (2004) for evidence that such relationships influence language processing). This raises the possibility that semantically related word pairs may over-estimate baseline rates of formal similarity. Additionally, these correlations may not be constant across all aspects of phonological structure, introducing unknown biases into the similarity analyses. In this regard, Tamariz (2005) finds that consonantal similarity correlates positively with syntactic/semantic similarity, while vowel similarity is negatively correlated.

Second, this approach assumes semantic errors are an appropriate baseline because they do not arise at the same stage(s) of processing as non-semantically related word errors (i.e., they assume that all semantic errors arise prior to lexical phonological/orthographic spell-out). However, simulation analyses have demonstrated that in a system with cascaded activation the form level representations of words that are semantically related to a target word can become activated—leading to semantic errors during lexical spell-out (see Rapp & Goldrick, 2000). If that is the case, at least some semantic errors may be influenced by factors that contribute to the production of non-semantically word errors during lexical spell-out—providing another reason that they constitute a problematic baseline.

Other studies assume that appropriate baseline rates correspond to the rates at which any two randomly paired words are similar along the dimension of interest (frequency, grammatical category, etc.). This method has been implemented in a variety of ways including randomly repairing targets and errors (Berg, 1992; Dell & Reich, 1981) or extracting correctly produced words at random from the speech error corpus from which the word substitutions are drawn (Arnaud, 1999; Harley & MacAndrew, 2001). Gagnon et al. (1997) examined the properties of the entire set of CVC English words to estimate chance over the entire lexicon.

Although this method avoids some of the issues surrounding the use of semantic errors, it generally suffers from a different problem. The implicit null hypothesis that this approach implements is a language production system in which no specific factors influence the co-activation of target and non-targets. Of particular import in this context is this null hypothesis assumes there is no role for form-based similarity. If this is indeed the null hypothesis of interest, then the use of randomly paired words is appropriate. More typically, however, researchers actually do assume that form-based similarity is relevant and they are using the random-pairing approach to evaluate some additional candidate dimension of similarity (e.g., grammatical category, lexical frequency, etc.). However if there are correlations between form-based (phonological or orthographic) properties of words and other dimensions of representational structure, the use of randomly paired words will lead to an underestimation of baseline rates of similarity. For example, studies have shown that words that share phonological properties tend to also belong to the same syntactic category. In English and other languages, words within the same syntactic category tend to share certain phonological features (e.g., stress, length, vowel quality; Kelly, 1992; Shi, Morgan, & Allopenna, 1998).

The work we report on here evaluates the similarity relations observed in the target-error pairs against a number of null hypotheses, all of which assume that some sort of form-based similarity plays a role in word production. Rather than estimating baseline rates by selecting any word at random, in each of the statistical simulation analyses we consider words drawn from a subset of the lexicon. The makeup of this subset is determined in a “hypothesis-by-hypothesis” basis depending on characteristics of the null hypothesis against which the observed results are being evaluated. For example, one null hypothesis is that non-target words

are activated based solely in terms of position-independent segmental overlap with the target. In that analysis, we consider only those word pairs that have a certain degree of position-independent segmental overlap with the target and we then examine these word-pairs to estimate chance levels of similarity along the dimensions of interest (grammatical category, relative lexical frequency, etc.). By restricting our analysis to this particular subset, we are able to test a well-specified null hypothesis.

In fact, a similar technique that involved random selection from a set of form-related candidates was utilized by Hurford (1981) in his critique of the seminal Fay & Cutler (1977) study of malapropisms in spontaneous speech (see Cutler & Fay, 1982, for a reply). However, he estimated baseline rates using only a single set of randomly selected words while we generate many thousands of such sets to estimate the distribution of each measure of target-error similarity that would be expected under each null hypothesis examined.

The experimental section of the paper consists of four sections. First, we establish that the non-semantic errors produced by four individuals with acquired word production deficits arise (largely) within the phonological or orthographic spell-out process. Second, for each of the individuals, we quantify the relationship between targets and non-semantic errors along a number of lexical and form-based dimensions. Third, we compare observed rates of similarity to those expected by chance under seven different null hypotheses, using a statistical-simulation procedure. Finally, we consider whether naming accuracy is affected by neighborhood density as defined according to the lexical and form-based factors we found to be significant.

Case studies and characterization of production deficits

We analyze data from four individuals with acquired word production deficits; one with a spoken word production deficit and three with written word production deficits. We first present brief background information for each of them and then we present data that establishes that their deficits arise specifically in lexical phonological/orthographic spell-out.

Spoken production case

SP1 was a 62 year-old right handed man with three years of university education. He was employed as a jet-testing engineer prior to suffering an infarct in left parietal cortex as well as a lacunar infarct in the right basal ganglia. As a consequence he had difficulties in spoken and written language production, comprehension was largely intact.

Written production cases

WR1 was a 65 year-old right-handed woman with a high-school education. She worked in a clerical position prior to retirement. She suffered an infarct in left posterior parietal and superior temporal cortex. WR2 was a 21 year-old right-handed woman with a high-school education and some college coursework. She was involved in a motor vehicle accident that caused a severe closed head injury resulting in a subacute subdural hematoma in the left frontal and parietal lobes as well as contusions in the left posterior temporal and occipital lobes. Following their injuries, both WR1 and WR2 suffered mild spoken language difficulties (including the production of semantically and morphologically related words and phonological related words and nonwords) as well as significant spelling impairments. Spoken and written comprehension were intact. WR3 was a 65 year-old right handed woman with a high school education who worked as a home health aide. She suffered a stroke affecting the left hemisphere (additional data regarding her lesion are not available). She had significant impairments in both spoken and written language comprehension and production.

The analysis of neighborhood properties will involve comparing target-error similarity for all lexical errors that are not semantically related to the target words (e.g., naming a picture of a mitten as “muffin;” in spelling to dictation, writing LURID in response to target “lyric”). We will argue that these errors result from disruption to lexical spell-out processes (see Figure 1) and not from earlier semantic deficits or later deficits to more peripheral aspects of written or spoken motor planning and production. Only WR3 exhibits additional mild impairments affecting word semantics and orthographic buffering, yet we have included her because subsequent analyses show that her non-semantic lexical errors follow the same pattern as the other three individuals.

Evaluating lexical semantic processing

SP1’s score on the auditorily presented Peabody Picture Vocabulary Test–Revised (Dunn & Dunn, 1981)—which evaluates single word comprehension—was within the normal range (42nd percentile) and he made no errors on several other auditory comprehension tasks: an auditory word/picture verification task (with semantically and phonologically related foils; $N = 774$), the auditory comprehension subtests of the Boston Diagnostic Aphasia Exam and a synonym-matching test with abstract and concrete nouns ($N = 48$).

WR1 generally scored within the normal range on single word comprehension tasks. On the combined Imageability (test no. 5) and Morphology (6) auditory lexical decision tasks from the Psycholinguistic Assessments of Language Processing in Aphasia (PALPA) tests (Kay, Lesser, & Coltheart, 1992), WR1 was 93% correct (205/220). In addition, she was 95% correct (247/260) on an auditory word/picture verification task (which required her to correctly accept the pairing of an auditorily presented target word with its picture, and correctly reject the pairing of this picture with a semantically or phonologically related word). This score was at the low end of the normal range. WR2 made no errors on an auditory word/definition verification task (41/41 correct), and her performance was within the normal range of performance (97% correct) on the PALPA Imageability auditory lexical decision task (test no. 5). In contrast, WR3 exhibited some difficulties in comprehension, scoring 73% correct (186/254) on auditory word/picture verification. The vast majority of her errors (21% of responses) were semantic (accepting a picture of a CAT for the word DOG). However, this stands in contrast to her performance in spelling to dictation where semantic errors were rare (across tasks, fewer than 2% of total responses). It is unlikely, therefore that her non-semantic lexical errors in written production were the result of semantic or comprehension deficits (we return to this point below).

Peripheral input and output processes

Repetition is a task that allows us to evaluate both auditory perception (critical for ruling out an input locus for errors in writing to dictation for Cases WR1-3) as well as peripheral spoken production processes (critical for ruling out a peripheral motor planning and/or execution impairment in Case SP1).

SP1’s repetition of the picture names from 2 administrations of the Snodgrass and Vanderwart (1980) set was excellent (repetition: 99% segments correct; $N = 1976$) and contrasted significantly with his error rate in picture naming of the same words (93% segments correct; $\chi^2(1, N = 3962) = 72.8, p < .0001$). This contrast between good repetition and impaired naming is consistent with a deficit to lexical phonological spell-out. The reasoning is as follows: picture naming is a semantically mediated task requiring L-level selection and lexical phonological spell-out to gain access to phonological form. In contrast, repetition can be successfully performed using non-lexical acoustic-phonological conversion processes (Hanley, Dell, Kay & Baron, 2004; Hanley, Kay, & Edwards, 2002; McCarthy & Warrington, 1984). A selective

deficit to lexical phonological spell-out would be predicted to lead to impaired picture naming but spare repetition (Goldrick & Rapp, 2007).

In spelling to dictation, Cases WR1-3, were asked to repeat stimuli before spelling them. All three individuals were extremely accurate in the oral repetition component of the task, indicating that their spelling errors did not arise in processing the auditory input. (On the occasional trial where a repetition error occurred, spelling did not begin until the word target had been correctly repeated.)

Evidence that errors do not arise at a level of motor planning or execution for written forms is the finding that oral and written spelling to dictation are performed with comparable accuracy (see Rapp & Caramazza, 1997). This was the case for all three individuals: WR1: 66% correct written, 64% correct oral ($\chi^2(1, N = 524) = 0.008, p > .05$); WR2: 50% correct written, 46% correct oral ($\chi^2(1, N = 440) = 0.45, p > .05$) and WR3: 51% correct written, 58% correct oral ($\chi^2(1, N = 288) = 0.11, p > .05$).

Lexical processing

Up to this point we have ruled out pre- and post-lexical loci of disruption, favoring a locus of impairment in lexical spell-out by process of elimination. Positive evidence of a lexical locus is provided by a significant effect of lexical frequency. All four individuals were significantly more accurate on high vs. low frequency target words: Case SP1: high frequency 98% correct, low frequency 96% correct ($\chi^2(1, N = 1940) = 6.0, p < .02$); Case WR1: high frequency 78% correct, low frequency 53% correct ($\chi^2(1, N = 408) = 29.38, p < .05$); Case WR2: high frequency 64% correct, low frequency 36% correct ($\chi^2(1, N = 408) = 31.85, p < .05$); and Case WR3: high frequency 50% correct, low frequency 32% correct; $\chi^2(1, N = 338) = 10.65, p < .05$).

The absence of impairments to perceptual, semantic, or more peripheral production processes, coupled with the presence of a lexical frequency effect, indicates that for Cases SP1 and WR1 and 2 word production errors arose largely within lexical phonological or orthographic processing. As indicated earlier, WR3 is a more complex case with lexical processing constituting one among various disruption loci. (For additional information regarding Case WR1 see Folk, Rapp, & Goldrick, 2002 [referred to as case MMD]; for WR2 see Folk & Jones, 2004 [referred to as case JDO] and for SP1 see Goldrick & Rapp, 2007 [referred to as case CSS]).

Buffering

According to a number of word production theories, segments are buffered while awaiting additional specification and/or selection for production (for a recent review in spoken production, see Shallice et al., 2000; written production, Tainturier & Rapp, 2001). The hallmark characteristic of buffering deficits, either phonological or orthographic, is an effect of the number of buffered elements (segments or syllables) on accuracy (with segments in longer words produced less accurately than segments in shorter ones). In the case of written word production it has been argued that orthographic working memory processes can be distinguished from lexical orthographic processes such as L-level selection and orthographic spell-out and that, in fact, one set of processes may be selectively affected by neural injury (Tainturier & Rapp, 2001; but see Sage & Ellis, 2004). For Case SP1 length effects were examined in multiple administrations of the Snodgrass and Vanderwart (1980) set for spoken picture naming. Significant length effects for words 3–7 segments in length were found for both word ($\chi^2(4, N = 364) = 11.9, p < .02$) and segment accuracy ($\chi^2(4, N = 1628) = 17.9, p < .005$). For Cases WR1-3 length effects were examined in spelling to dictation for words 4–8 letters in length. For WR1 And WR2 length effects were absent whether performance was

evaluated by word or letter accuracy (WR1: word, $\chi^2(4, N = 140) = 4.11, p > .05$; letter: ($\chi^2(4, N = 840) = 1.24, p > .05$); WR2 word: ($\chi^2(4, N = 140) = 1.76, p > .05$); letter: ($\chi^2(4, N = 840) = 0.51, p > .05$). For Case WR3, although there was no significant effect of length on word accuracy ($\chi^2(4, N = 70) = 3.5, p > .05$) she did show a significant effect on letter accuracy ($\chi^2(4, N = 840) = 17.68, p < .005$).

In sum, Cases WR1 and WR2 exhibited a striking absence of length effects, as measured either by letter or word accuracy, consistent with a relatively selective lexical level disruption that does not implicate orthographic working memory. WR3 showed mild length effects, apparent only when measured by letter accuracy, suggesting additional disruption to the graphemic buffer. SP1, in contrast, showed robust effects of segment length. This may indicate an additional independent disruption of the phonemic buffer or it may be the consequence of the primary disruption to lexical spell-out processes. The latter would be the case if, in spoken production, there is not the same degree of functional independence between the lexical long-term memory and the working memory processes that can be seen in written word production.

Identifying errors arising from disruption to lexical spell-out

We have argued that the pattern of accuracy across tasks and the presence of lexical frequency effects indicate disruption to lexical processing in all four individuals. Given the functional architecture depicted in Figure 1, this could involve either L-level selection or Lexical spell-out. How to distinguish between them? Rapp & Goldrick (2000) simulated a word production architecture incorporating bi-directional interaction between L-level and segmental levels of representation (as in Figure 1). In these simulations, deficits limited to L-level selection resulted in the production of high rates of semantic errors with very few non-semantic errors (words or nonwords). This pattern was reliable across a range of accuracy levels (i.e., all simulated accuracy levels exceeding 40%). Importantly, production of high rates of non-semantic errors was only found when spell-out processes were damaged. With regard to word errors, this pattern of errors is consistent with the simulation findings that we have referred to earlier, indicating that non-semantic lexical representations are not strongly activated prior to spell-out. On this basis, we can assume that the non-semantic word errors produced in the four cases arose almost entirely from lexical spell-out. These are precisely the errors that will be used in the neighborhood analyses.

Table 1 reports the distribution of response types produced by the four subjects in spoken naming (SP1) or written spelling to dictation (WR1-3). For each individual the error distribution is consistent with a deficit to phonological/orthographic spell out. In spoken picture naming across a variety of sets of materials SP1 produced a total of 2386 responses which included phonologically related word and nonword errors as well as semantic (shirt → skirt; see Rapp & Goldrick, 2000, for further discussion) and morphological errors (particularly including a number of compound constituent substitutions such as butterfly → butterflyer; see Badecker, 2001, for further discussion). Given the possibility that the semantic and morphological errors arose in lexical selection, prior to lexical spell-out, we excluded these from further analysis (word accuracy on the remaining 1996 items was 92%). The data set used in the analyses reported below consisted of the 61 whole word substitutions (e.g., mitten → muffin) produced by SP1.

Cases WR1, WR2, and WR3 all produced word errors (e.g., “thaw” → T-H-O-U-G-H) as well as phonological plausible (e.g., “copy” → C-O-P-P-I-E; plausible spellings for each target phoneme were based on Hanna, Hanna, & Hodges, 1966) and other nonword errors (e.g., “deny” → D-E-N-O-C-K). It is worth noting that in writing to dictation, unlike spoken picture naming, both lexical as well as non-lexical phoneme-to-grapheme (PG) conversion processes contribute to the written response. (See Delattre, Bonin, & Barry (2006) and Rapp, Epstein, & Tainturier (2002) for recent reviews of data from neurologically intact and impaired

participants supporting the role of lexical processes in this task.) PG conversion processes may play a particularly important role in the face of lexical impairment (as in the cases reported here). The production of phonologically plausible errors constitutes evidence that these processes were indeed active in addition to the lexical ones. However, the fact that all errors were not phonologically plausible indicates that these PG processes were not fully functioning, an inference supported by the different levels of accuracy in nonword spelling exhibited by the three individuals (WR1: 78% of letters correct; WR2: 96% correct, WR3: 33% correct). As with SP1, the analyses below include only the form-based word errors (e.g., “poise” → P-A-U-S-E). Case WR1 produced 91 form-related word errors (overall word accuracy: 64%, N = 2430), while case WR2 produced 64 (overall word accuracy: 54%, N = 2453). Case WR3 produced a total of 96 lexical errors that were not semantically related to the target (overall word accuracy, excluding semantic and morphological errors: 50%, N = 332); these errors were used in the analyses below.

Evaluating target-error similarity

Similarity between targets and errors was evaluated along a number of dimensions: relative lexical frequency, grammatical category, extent of segmental overlap, the position of segmental overlap, and segmental and/or syllabic length. We detail methods for evaluating similarity below.

Lexical frequency

We estimated lexical frequency using the COBUILD counts in the CELEX word form database (Baayen, Piepenbrock & Gulikers, 1995; these counts collapse across multiple genres as well as spoken and written corpora). We collapsed all frequency counts across homophones or homographs given that single word responses cannot be unambiguously identified as a particular member of a homophone set (for reviews of the debate on the representation of lexical frequency for homophones see: Caramazza, Bi, Costa, & Miozzo, 2004; Jescheniak, Meyer, & Levelt, 2003).

Grammatical category

We utilized the grammatical category information in the CELEX database to identify grammatical category overlap. For errors that were homophones or homographs, we assumed that a target and error shared grammatical category if any of the homophonic or homographic word forms shared the target’s grammatical category. In picture naming, target grammatical category was specified by the task. For spelling to dictation, targets were associated with the grammatical category of all homophonic or heterographic word forms.

Segmental overlap

We considered both position-independent and position-specific segmental overlap. We refer to the overlap index as SOI (for Segmental Overlap Index; based on the Phonological Overlap Index of Rapp & Goldrick, 2000). In the position-independent form of the analysis we define the SOI for two strings in the following way: the total number of segments (phonemes or letters) shared without regard to position, divided by the total number of segments in the two strings. For example, the position-independent SOI of the strings/kaet/and/taep/is 0.66 (4/6; phonemes/ae/and/t/are shared across the two strings).

For position-specific analyses, we followed previous analyses (Miller & Ellis, 1987; Schwartz, Wilshire, Gagnon, & Polansky, 2004) and collapsed segments of all words into 5 positions (after Wing & Baddeley, 1980). We did so as a common representational scheme for targets and errors facilitates their comparison, allowing the identification of coarse serial-position patterns. Details of the coding scheme are given in Appendix A. After assignment of all

segments of the target and error response words to the 5 positions, position-specific SOI was defined as: the total number of segments occurring in the same position in both target and error, divided by the total number of segments in the two strings. For example, the position-specific SOI of /kaet/ and /taep/ is 0.33 (2/6; in the coding scheme used here, /ae/ occurs in position 3 - the center position of five- in both of the strings). It is important to note that, according to this scheme, serial position *within* each position is not considered. For example, in the word “belt” the letters E and L are both assigned to position 3. Similarly, for the word “flip” L and I are both assigned to position 3. The overall position-specific SOI for these two words is 0.25 (2/8; L occurs in position 3 in each of the two strings).

In addition to considering overall position-specific overlap across the entire strings, we considered SOI for each position separately. Here also, serial order within a position was not considered. The SOI of a position was simply the number of segments shared by target and error within that position divided by the total number of segments in both target and error in that position. For example, for the pair “belt” and “flip,” the SOI for position 3 is 0.5 (2/4; L occurs in both the target and error in position 3).

Segmental length

For calculations involving phonemes, we relied on the primary CELEX transcription (with minor corrections for American pronunciation variants). Diphthongs (e.g., /o~/) and affricates (e.g., /t~/) were represented as single phonemes, and syllabic liquids and nasals (e.g., /n/ in “button”) were represented as schwa + consonant sequences.

We also evaluated syllabic length, although only for spoken word production. We report on these results only after considering those features shared by written and spoken modalities.

Results: Observed target-error similarity

Table 2 reports the observed relationships between targets and all non-semantic word errors for each of the dimensions of interest, for each of the 4 cases. Given these results, the question to be addressed in subsequent analyses is whether or not the observed similarity along these various dimensions supports the inference that the word production system is structured in a manner that is explicitly sensitive to one or more of these dimensions. For example, does the fact that 92% of target-error pairs in Case SP1 share grammatical category indicate that grammatical category is represented in the word production system? One alternative hypothesis is that this degree of grammatical category overlap occurs by chance in a system not explicitly sensitive to grammatical category, but merely as a result of the properties of the lexicon combined with other features that may be explicitly represented and play a role in word processing. To assess the significance of the observations in Table 2 we determined baseline (chance) rates expected under a null hypothesis in which the dimension of interest is not represented within the production system.

In the following sections, we evaluate the observed values by comparing them to baseline rates generated by seven different “null” hypotheses. Each of these represents a different word production architecture that incorporates certain features and excludes others. To recapitulate, the logic is as follows: if, for Theory +X –Y (incorporating feature X but not feature Y), we find that chance-generated overlap values for feature Y are lower than the observed overlap values for feature Y, then we conclude that “baseline” architecture is inadequate and that feature Y may indeed be explicitly represented in the system.

Statistical simulation evaluation of the observed relationships between targets and errors

Analysis methods

Appendix B details how the baseline rates or predictions for each architecture were determined. Briefly, for each case and each dimension of interest, a Monte Carlo method was used to generate the “predictions” of each null hypothesis/architecture. This was done by randomly pairing targets with other words creating target-pseudoerror pairs that matched the actual target-error pairs along the feature dimension/s represented in the specific architecture. To be clear, in this statistical analysis, the critical feature/s of the baseline architecture were implemented by means of the criteria used to sample the lexicon and generate target-pseudoerror sets. We then examined the degree to which the target-pseudoerror pairs were similar on the other non-represented dimensions; any similarity observed under those circumstances would be the result of chance factors as defined within that particular baseline architecture/null hypothesis. For example, Architecture 1 assumes that only position-independent segmental overlap influences the activation of a target word’s neighbors. To generate the predictions of this hypothesis with regard to target-error similarity along other dimensions, we first selected target-pseudoerror pairs that matched the position-independent SOI of the observed target-error pairs; we then considered their similarity along the other non-represented dimensions. For each architecture and for each individual we generated 10,000 sets of target-pseudo error pairs and carried out these evaluations.

If a null hypothesis/architecture is sufficient, the lexical and phonological properties of actual target-error pairs should not be significantly different from the randomly generated predicted values. To evaluate this, we compared the observed rates (Table 2) to the 95th percentile of the distribution of the chance rates predicted by each architecture (estimating the (one-tailed) cutoff value). If an observed rate exceeded the 95th percentile of the baseline distribution of the 10,000 sets of target-pseudoerror pairs, we concluded that the baseline architecture was inadequate and that the feature dimension may indeed be explicitly represented in the word production system.

Architecture 1: Position-independent segmental encoding only

As noted above, most production theories assume that form-related, non-target words are activated by feedback from the phonological representations they share with the target (e.g., during processing of target <DOG> or <BED>, feedback from the phoneme representation/d/ activates <DOT>). Our first analysis evaluated whether this assumption alone is sufficient to account for the full range of target-error relationships observed in Table 2.

Table 3 reports the predictions of Architecture 1 as estimated by the Monte Carlo simulations. The 95th percentile of the distribution predicted by Architecture 1 for each dimension of target-error overlap is provided along with the observed value for each case. The results reveal that, with few exceptions, the observed properties of the errors fall very close to or well outside the edge of the probability distribution predicted by this baseline theory. This provides clear support for the conclusion that a theory that assumes that only position-independent segmental overlap determines the activation of non-target words cannot account for the observed similarity between targets and errors along any of the other dimensions in any of the four cases.

Architecture 2: Position-specific segmental encoding only

The failure of Architecture 1 with regard to the within-position overlap values indicates that the word production processes involve segmental representations that encode some aspect of position. Architecture 2 assumes that segmental overlap is position-specific; that is, the

activation of non-target words is influenced by the position of the segments shared with the target. A new set of Monte Carlo simulations was carried out implementing this assumption by selecting target-pseudoerror pairs matched to the actual target-error pairs in terms of average within-position overlap. The results are given in Table 4.

This analysis reveals that a theory incorporating position-specific form overlap provides a better account of the data than a position-independent theory (Architecture 1). In particular this theory is able to account for the overlap of target-error pairs within serial positions 2–5 (excepting WR2, position 2). The average overlap values observed in these positions are well within the distribution predicted by this theory. This provides support for the position-specific segmental encoding assumptions of Architecture 2.

However, this theory is still unable to predict many other segmental and lexical properties of the errors. In particular, note that initial position overlap appears to exert an influence on non-target word activation over and above the average position-specific overlap for each word. These results indicate that word production theories must incorporate additional features in order to account for the activation of non-target words.

Architectures 3–6: Position-specific segmental encoding + X

While clearly superior to position-independent overlap, we have established that position-specific form overlap alone is still insufficient to account for the observed relationships between targets and errors. However, we cannot conclude that all the remaining features are represented in the spoken production system, as some may be intercorrelated. If that were the case, by including one additional dimension, the correlated dimensions would come along “for free.” In order to evaluate this possibility, we examined whether theories incorporating position-specific segmental encoding plus a single additional dimension would be sufficient to account for the observed similarity along the other dimensions. For example, Architecture 3 makes one additional assumption beyond the influence of position-specific overlap—higher frequency words are more active than lower frequency words. By incorporating this single additional factor into the theoretical architecture we can quite naturally expect to account for the relative frequency properties of the errors; however, the question we can address in this analysis is whether an architecture that includes these features can account for the other properties of the errors—not just their higher relative lexical frequency, but also their grammatical category overlap, overlap in first position, and segmental length. Given that the logic and analysis procedures are the same for the subsequent analyses, only a brief presentation of results will be necessary.

Architecture 3: Position-specific encoding + lexical frequency—The results (Table 5) reveal that, as would be expected, incorporating position-specific overlap as well as lexical frequency accounts for the average POI in positions 2–5 (excepting WR2, position 2) and for the relative frequency relationships between targets and errors, but not for the other lexical and phonological properties of errors. That is, this architecture does no better than Architectures 1 and 2 in accounting for the other properties of errors. For example, grammatical category overlap does not come along “for free” when lexical frequency is added to the processing system.

Architecture 4: Position-specific encoding + grammatical category—This set of Monte Carlo simulations examined the predictions of a theory incorporating both average position-specific form overlap and grammatical category (but no other factor), by matching targets-pseudoerror pairs with actual target-error pairs on these dimensions. The results (Table 6) indicate that this architecture was unable to account for any other properties of target-error relationships. That is, grammatical category is not intercorrelated with these other features to

such an extent that when a theory includes grammatical category membership as a dimension that determines co-activation with the target, other dimensions of similarity are automatically elevated.

Architecture 5: Position-specific segmental encoding + initial position identity

—This set of statistical simulations examined an architecture in which segmental identity in the initial position of a word plays a role above and beyond overlap in other positions (e.g., Shattuck-Hufnagel, 1992). Such a theory predicts that non-target words sharing phonemes in initial position with the target will be more active than words that exhibit comparable overlap within the other serial positions. The results of the Monte Carlo simulation of this architecture (as shown in Table 7) reveal that it cannot account for observed target-error overlap along the dimensions of frequency, length and grammatical category.

Architecture 6: Position-specific encoding + segmental length—This set of statistical simulations examined a theory in which non-target words sharing average position-specific overlap with the target as well as segmental length are more active than those differing in length. As the results of the Monte Carlo analysis (Table 8) indicate, this architecture was unable to account for any additional properties of target-error relationships for any of the cases.

As noted in the Introduction, previous studies of spoken production have suggested that target-error pairs tend to have the same number of syllables. This would seem to be especially relevant for Case SP1 with a spoken production deficit. To examine this possibility, we evaluated whether Architectures 1–6 predicted the observation that 85% of SP1's target-error pairs had the same number of syllables. That is, for each architecture we examined the degree to which the number of syllables in the target was predicted to be preserved in an error. The results indicate that none of the architectures could match the high degree of similarity observed in SP1's target-error pairs (95th tile: Architecture 1: 49%; Architecture 2: 69%; Architecture 3: 71 %; Architecture 4: 69%; Architecture 5: 64%; Architecture 6: 74%). (Note: a similar analysis in Goldrick and Rapp (2007) suffered from low power).

Architecture 7: Position-specific encoding + length in syllables

For SP1's data we also examined whether a theory that incorporated length in syllables (in addition to position-specific form overlap) would be sufficient to account for the observed similarity between targets and errors along the other dimensions of interest. The results are presented in Table 9. As before, this theory is unable to predict the degree of overlap that is observed in the actual target-error pairs along the dimensions that are not explicitly included within the theory.

Summary of statistical simulation evaluations

The analyses reveal that in the course of lexical phonological or orthographic spell-out, the activation of non-target words is driven not only by overall position-specific form overlap but also by: lexical (frequency and grammatical category) and form-based factors (overlap in initial position; length in terms of number of segments as well as number of syllables). These findings motivate a particular definition of neighborhood that we will refer to as the Lex-Form Composite as it combines both lexical and form-based factors in a "definition" of neighborhood. One expectation is that this composite should significantly predict the naming accuracy of words. We consider this possibility in the next analysis in which we examine naming accuracy for target words with high versus low density neighborhoods, as defined by the Lex-Form Composite.

Testing Lex-Form Composite: Effects of neighborhood density

To examine how the activation of non-target words as defined by the Lex-Form Composite affects naming accuracy, we considered the entire set of target words each individual attempted to name (not just the items that rise to non-semantic lexical errors as in the previous analyses). Specifically, we compared naming accuracy for words with many vs. few strongly activated non-target words (neighbors) as defined by the Lex-Form Composite.

To do so, we first identified, for each target word, the words in CELEX that Lex-Form might identify as being “strong neighbors.” Although we assume that neighborhood is a gradient notion for the purposes of this analysis we discretized it in the following manner. For each individual, for each target word, we identified the words in CELEX that: had a position-specific SOI with the target exceeding 70%; were higher in frequency than the target; shared the target’s grammatical category and 100% of segments in first position; and had segmental length identical to the target. (Additionally, for SP1, words were required to have the same number of syllables as the target.) We then created categories of high and low density target words. We considered all target words with 2 or more strongly activated neighbors to be high density words. To create a comparison category of low-density words, for each of these “high density” words, we identified a target word in each subject’s word set which, according to this implementation of the Lex-Form Composite should have no strongly activated neighbors. Each of these low density words was equal in segmental (and for SP1 also syllabic length) to a high density target, and high and low density target word log frequencies were matched as closely as possible. Table 10 reports the mean log frequencies of high and low density targets for each case. In no case were high and low density targets significantly different from one another ($ps > .85$).

We then calculated naming accuracy for high and low density targets for each participant. The results of the density analysis are reported in Table 10; they indicate that for all individuals, as evaluated by segment and word accuracy, high density targets were more accurately produced than low density targets, although the differences were not always statistically significant. These results provide support for the functional relevance of the feature dimensions identified as significant in the set of simulation analyses.

General Discussion

Many theories of language production assume that during the process of preparing written or spoken words (targets) for production, the representations of non-target words—lexical neighbors—become active. An understanding of the relationships between targets and their neighbors provides a window into the representations and mechanisms of word production. In this research we examined non-semantic errors arising within lexical phonological or lexical orthographic spell-out in four individuals with acquired spoken or written word production impairments. The similarity between the intended target words and the errors produced was first quantified along several dimensions and then compared to the simulated predictions of theories that differ regarding the dimensions they posit influence the activation of a word and its neighbors. The analyses revealed strikingly similar patterns across individuals and modalities. In both lexical phonological and orthographic spell-out, the activation of non-target words is driven by: position-specific form overlap, lexical frequency, grammatical category, overlap in initial position, and length. Finally, post-hoc analyses suggest that, in both modalities, strongly activated neighbors facilitate the successful production of a target word.

Relationship of results to other theories of neighborhood structure

As indicated in the Introduction, most theories of neighborhood structure define neighbors along purely form-based dimensions. These studies report robust correlations between performance measures and these form-based definitions of neighbors. The findings we have reported here are not inconsistent with this literature; rather, they expand upon it providing evidence of the independent contribution of multiple lexical and form-based factors in the activation of a target word and its neighbors.

One of the challenges in this area of research is that many of the measures that have been considered in previous work to be critical in the activation of neighbors are inter-correlated—sometimes highly so. In the work we have reported on here we have been able to move forward some distance in establishing the independent contribution of the various dimensions we have examined. First, we have been able to distinguish between the role of overall segmental overlap and position-specific segmental overlap, with the results indicating that targets and their neighbors share position-specific segmental representation. Second, we have been able to distinguish between the role of the first segmental position and all other positions, finding that despite overall target/neighbor segmental similarity, there is an additional factor that is responsible for the greater degree of overlap between targets and neighbors in the initial position. That is, overall segmental similarity does not predict the high degree of overlap at the initial position, nor does the overlap in the first position fully account for similarity between targets and their neighbors at all other positions. Third, we were able to show that the combination of overall segmental overlap and one other single phonological or lexical factor (frequency, grammatical category, length, initial position) never predicted the degree of overlap observed along the remaining dimensions. That is, these factors were sufficiently independent of one another that when items matched along one dimension they did not automatically match along the others.

Given the high degree of intercorrelation among these various factors, we would expect that our composite measure of neighborhood density (which includes the various lexical and form-based dimensions we investigated) should correlate with other measures of neighborhood density used in the literature. For example, we expect that words that are identified by our Lex-Form Composite measure to have high density neighborhoods should also be likely to be identified as high density by other measures. In fact, this is what we find when we correlate the Lex-Form Composite ratings with two popular density measures—the number of words differing by a single segmental substitution (here, “Coltheart’s N;” Coltheart et al., 1977) and the number of words differing by the substitution, addition, or deletion of a single segment (here, “One Segment Edit Density;” Vitevitch & Luce, 1998). For example, for the target words used in WR1’s density analysis, significant positive correlations were found between the Lex-Form Composite rating and Coltheart’s N ($r = 0.40$, $t(330) = 7.9$, $p < .0001$) as well as between the Lex-Form Composite rating and One Segment Edit Density ($r = 0.45$, $t(330) = 9.1$, $p < .0001$). The relationship between the various density measures is not specific to the materials of this study. For example, Vitevitch et al.’s (2004) Experiment 3 found a significant influence of One Segment Edit Density on picture naming latencies; ‘dense’ words with many neighbors were named more quickly than ‘sparse’ words with few neighbors. Applying the Lex-Form Composite to the Vitevitch stimuli we find that this measure assigns a higher density (0.18) to Vitevitch et al.’s dense words relative to their sparse words (0.0; $t(42) = 2.2$, $p < .04$).

While it is clear that previous form-based measures of neighborhood density captured a significant amount of the variance in accuracy and reaction time data, it is also the case the results of our analyses clearly show that multiple dimensions are relevant in the activation of a target and its neighbors. What remains to be done is to determine the specific mechanisms by which these dimensions exert their influence and how their contributions are differentially

weighted throughout the production process. This will require additional empirical work—which should ideally draw on converging evidence from chronometric studies of picture naming as well as analyses of word production deficits.

Implications for theories of language production

A number of existing theories of word production provide mechanisms by which at least one or more of the factors we have identified can contribute to the co-activation of targets and their neighbors. We review these very briefly here and discuss how they could be integrated within a single production architecture. We also point out challenges provided by our findings to specific theories or aspects of theories of word production.

Lexical frequency

According to some accounts, both the L-level and the phoneme/grapheme representations of high frequency non-target words are more active than those of low frequency words. In some theories, this occurs because the strength of connections between L-level and form-level representations is modulated by lexical frequency (e.g., MacKay, 1987). Under such accounts, the L-level representations of high vs. low frequency non-target words will receive greater activation due to feedback. Alternatively, lexical frequency may influence the properties of L-level representations themselves (e.g., high frequency words may have higher resting activation levels: Dell, 1990). Here, the efficacy of feedback in activating non-target L-level representation is expected to be greater for high vs. low frequency non-target words. In this way, the higher vs. lower frequency neighbors of a target would become the most active and, in the case of disruption, be more likely to be produced (but see Jescheniak & Levelt (1994) for an architecture in which lexical frequency would not be expected to influence lexical phonological spell-out, and for which our results would represent a challenge).

Grammatical category

Under a number of accounts it is assumed that grammatical category information constrains L-level selection such that only/primarily syntactically appropriate words are selected for form-level encoding (Dell, 1986; Garrett, 1980; Levelt et al., 1999). In many spreading activation theories, this selection mechanism is implemented by using structural frames with categorically specified slots (see Dell, Burger, & Svec, 1997, for a review). These slots enhance the activation of all L-level units within the specified grammatical category, biasing selection to grammatically appropriate words. For example, during the course of noun phrase production, a structural frame would activate a noun slot, enhancing the activation of all L-level units corresponding to nouns. This activation boost ensures that the most highly activated L-level unit corresponding to a noun (and not a verb) is selected during production. If L-level selection is implemented in this manner, the activation of non-target words can be influenced by grammatical category. Feedback will easily enhance non-target L-level representations sharing the target's grammatical category as they have been pre-activated by the structural frames. Cascade from these pre-activated L-level representations will also serve to enhance the activation of form-level representations of these non-target words (for simulation results supporting this analysis, see Goldrick & Rapp, 2002; but see Dell, 1986 for an architecture in which grammatical category would not be expected to influence lexical phonological spell-out).

Serial order and position

Our analyses indicate that targets and neighbors tend to share segments in the same positions (at least coarsely defined). As briefly discussed above, many production theories assume that phonological/orthographic segmental information (phonemes, graphemes) is stored in a position-specific manner. For example, according to Dell (1986; see also Dell et al., 1997)

representations retrieved during lexical phonological spell-out contain position-specific representations of segments (e.g., the /k/ in “cope” is encoded by a distinct unit from the /k/ in “poke”). As discussed above, theories such as these predict that non-target words must share segments in the same position as the target in order to be co-activated. For example, a position-specific encoding of the /k/ in target “cat” would have a feedback connection to <COPE> but lack a link to words such as <POKE> (but see Warker & Dell (2006) for a proposal in which segment identity and order are independent).

Future work should aim to further refine claims regarding how position is encoded within lexical spell-out processes. As noted above our analyses have used rather coarse-grained notions of position, collapsing multiple phonemes/letters into 5 positions. This is consistent not only with position-specific representations (as proposed for phonological representations by Dell, 1986, and orthographic representations by McClelland & Rumelhart, 1981) but also those with representations that allow for partial overlap between the representations of phonemes/letters in different positions (see, e.g., Whitney (2001) and Fischer-Baum, McCloskey & Rapp (submitted) for theories of gradient representations of position in orthographic representations).

In addition to position-specific encoding of segments, some theories have accorded a special status to particular positions in the string. For example, Shattuck-Hufnagel (1992) proposed that consonants in the initial position of words form a distinct group within lexical phonological representations (allowing them to play a critical role in sequencing words for production). Other theories have not singled out initial segments, but have instead assumed that lexical phonological representations are retrieved sequentially (left-to-right; see O’Seaghdha & Marin, 2000, for a recent review of this proposal and related mechanisms). These mechanisms and representations would be consistent with our finding that initial positions are shared between target and neighbor at a higher rate than are other segments and also more than would be predicted by the overall similarity between target and neighbors.

Segmental and syllabic length

Most theories of spoken production assume that subsequent to lexical phonological spell-out segmental representations must be linked with wordshape frames that specify the metrical structure of words (e.g., consonant/vowel structure; syllabic organization; O’Seaghdha & Marin, 2000; Sevald, Dell, & Cole, 1995). Activation from the target wordshape frame may serve to enhance the activation of segmental representations consistent with the target structure. Under such accounts, feedback from phonological representations will then favor non-target L-level representations that share the target’s wordshape. This type of mechanism would generate a degree of similarity between target and neighbors in terms of length that is not predicted simply by overall segmental overlap.

Integrating these mechanisms within an architecture for spoken production

As reviewed in the previous sections, there are multiple mechanisms that are consistent with the various factors revealed by our analysis. In this section, we briefly sketch how one such set of mechanisms could be integrated into a single architecture. We focus on spoken production, but assume that similar principles apply within written production.

Figure 2 provides an illustration of this architecture for L-level selection and lexical spell-out processes. Building on the schematic in Figure 1, semantic, syntactic, L-level and sub-lexical form representations are instantiated via localist connectionist processing units. We assume that in addition to information regarding segmental identity (e.g., /k/, /ae/, /t/), sub-lexical form representations contain a prosodic frame organizing segments into a consonant-vowel (C/V) structure (e.g., Dell, 1988).

Processing begins via the activation of a set of semantic features corresponding to the intended target. This, in turn, activates the target's L-Level representation (here, CAT). Activation from both semantic and L-level representations contacts an appropriate syntactic frame (following proposals such as Dell et al., 1997; the frame is depicted here as the syntactic feature <noun>). This activates all L-level units within the specified grammatical category, biasing lexical selection to grammatically consistent representations. As a consequence the activation of L-level representations sharing the target's grammatical category is boosted.

L-level units also activate sub-lexical form representations. The strength of the connections between L-level units and their associated form representations varies with lexical frequency (e.g., MacKay, 1987; this is not depicted in the figure). During spell-out, prosodic frames (shown here as a syllable with a CVC frame) play a role parallel to syntactic frames (following Dell, 1988), biasing selection towards appropriate sub-lexical form units (shown here as position-specific segments). We assume that initial positions are privileged within these frames (Shattuck-Hufnagel, 1992). This is shown by the greater activation of the initial C unit within the frame and the corresponding boosted activation of the initial/k/unit.

Activation flow is bidirectional between L-level and sublexical form representations, allowing for the activation of form-related lexical neighbors (Rapp & Goldrick, 2000). This architecture allows the various form-related factors identified in this study to influence the activation of these neighbors. The activation of neighbors is driven by position-specific form representations, accounting for the tendency of neighbors to share segments within positions. The special status of the first position is accounted for by its privileged status within the prosodic frame. This boosts the activation of segments in initial position, leading to stronger feedback to L-level representations sharing the target's initial segment (contrast <HAT> vs. <CAP> in Figure 2). Finally, the tendency of neighbors to share length is also attributed to the influence of the prosodic frame. This boosts the activation of phonological representations that share the target's length, increasing feedback to their corresponding L-level representation (contrast <CAP> and <CAFE> in Figure 2).

The analyses here have suggested that words with many strongly activated neighbors are more accurately retrieved (consistent with previous research using both accuracy and reaction time measures). In this framework, this is attributed to the positive feedback loops between L-level and sublexical form representations. The target spreads activation to representations that share its formal properties; those neighbors that overlap on many dimensions become strongly activated. These neighbors send reciprocal activation to the representational elements they share with the target, enhancing the speed and accuracy with which they are retrieved (Dell & Gordon, 2003).

Modality-independent constraints on production

A striking result of the current study is that similar factors drive the activation of non-target words in both lexical phonological and lexical orthographic processing. Intuitively, this result may be somewhat unexpected. The physical manifestations of phonological structure are oral gestures and sounds, while orthographic structure is typically realized by manual gestures and visual symbols. Furthermore, unlike spoken production, lexical orthographic structure can be expressed in both visual and auditory modalities (i.e., written vs. oral spelling). However, viewed from the perspective of language production theories, the similarity across modalities is unsurprising. At the level at which these individuals experienced disruption—in the spell-out of long term memory representations of word sounds or spellings—both in the phonological (Goldrick & Rapp, 2007) and the orthographic (Tainturier & Rapp, 2001) representations are assumed to correspond to relatively abstract representations of segments. At these abstract levels of representation, differences in the ultimate format of output might be expected to not

exert a strong influence on processing. On a more speculative note, Dehaene & Cohen (2007) recently proposed that the parts of human cortex that are specialized for cultural domains (such as reading or arithmetic) are the product of “cultural recycling of cortical maps”. They argue that cultural skills recruit or “invade” pre-existing neural circuits that carry out computational functions that are similar to those required by the cultural skill. If written language has appropriated areas dedicated to spoken language it may, in so doing, have incorporated similar operating and representational principles. Interestingly, research on written and spoken word perception have yielded modality-specific differences such that studies of orthographic perception often document facilitatory effects of neighbors (Andrews, 1997; but see Rastle, 2007, for a recent review of conflicting findings) whereas studies of speech perception consistently show inhibitory effects of neighbors (Luce & Pisoni, 1998). A number of accounts of these conflicting patterns have been offered. Recently Magnuson, Mirman, & Strauss (2007) proposed that these contrasting patterns are a consequence of temporal differences in the input modalities (i.e., acoustic information is processed serially, while visual input is processed more in a more parallel fashion). Using an interactive-activation model of word perception, they show that serial input enhances competitive effects while parallel input enhances facilitatory effects.

In production, however, both written and spoken word processing are driven by the same input: amodal semantic (and syntactic) representations. This stands in contrast to perceptual processing which is inherently signal-driven and where physical differences between modalities may exert considerable effects on processing.

Conclusions

Considerable empirical work has examined the consequences of neighborhood density for both perception and production in both spoken and written modalities. However, far less attention has been given to the prior question of what makes a word a neighbor. We applied novel statistical simulation methods for evaluating the relationship between target words and errors, an approach that provided evidence for the multiple influences on word activation in production. Focusing our efforts on these questions should continue to contribute to the development of more comprehensive theories of the representational and processing mechanisms underlying both production and perception.

Acknowledgments

This research was supported in part by National Institutes of Health Grant DC007977 to MG and Grant DC006740 to BR, as well as the IGERT Program in the Cognitive Science of Language at Johns Hopkins University, National Science Foundation Grant 997280.

The authors would like to thank SP1, WR1, WR2, and WR3 for their participation.

References

- Abd-El-Jawad H, Abu-Salim I. Slips of the tongue in Arabic and their theoretical implications. *Language Sciences* 1987;9:145–171.
- Arnaud P.J.L. Target-error resemblance in French word substitution speech errors and the mental lexicon. *Applied Psycholinguistics* 1999;20:269–287.
- Andrews S. The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin and Review* 1997;4:439–461.
- Baayen, R.H.; Piepenbrock, R.; Gulikers, L. The CELEX Lexical Database (Release 2) [CD-ROM]. Philadelphia: Linguistics Data Consortium; 1995.
- Badecker W. Representational properties common to phonological and orthographic output systems. *Lingua* 1996;99:55–83.

- Badecker W. Lexical composition and the production of compounds: Evidence from errors in naming. *Language and Cognitive Processes* 2001;16:337–366.
- Baus C, Costa A, Carreiras M. Neighbourhood density and frequency effects in speech production: A case for interactivity. *Language and Cognitive Processes* 2008;23:866–888.
- Berg T. Prelexical and postlexical features in language production. *Applied Psycholinguistics* 1992;13:199–235.
- Bergen BK. The psychological reality of phonaesthemes. *Language* 2004;80:290–311.
- Best W. A reverse length effect in dysphasic naming: When elephant is easier than ant. *Cortex* 1995;31:637–652. [PubMed: 8750023]
- Best W. When racquets are baskets but baskets are biscuits, where do the words come from? A single case study of formal paraphasic errors in aphasia. *Cognitive Neuropsychology* 1996;13:443–480.
- Biran M, Friedmann N. From phonological paraphasias to the structure of the phonological output lexicon. *Language and Cognitive Processes* 2005;20:589–616.
- Blanken G. Formal paraphasias: A single case study. *Brain and Language* 1990;38:534–554. [PubMed: 2375981]
- Blanken G. Lexicalisation in speech production: Evidence from form-related word substitutions in aphasia. *Cognitive Neuropsychology* 1998;15:321–360.
- Brunsdon R, Coltheart M, Nickels L. Treatment of irregular word spelling in developmental surface dysgraphia. *Cognitive Neuropsychology* 2005;22:213–251.
- Butterworth B. Disorders of phonological encoding. *Cognition* 1992;42:261–286. [PubMed: 1582159]
- Caramazza A. How many levels of processing are there in lexical access? *Cognitive Neuropsychology* 1997;14:177–208.
- Caramazza A, Bi Y, Costa A, Miozzo M. What determines the speed of lexical access: Homophone or specific-word frequency? A reply to Jescheniak et al. (2003). *Journal of Experimental Psychology: Learning, Memory, and Cognition* 2004;30:278–282.
- Caramazza A, Hillis AE. Where do semantic errors come from? *Cortex* 1990;26:95–122. [PubMed: 2354648]
- Coltheart, M.; Davelaar, E.; Jonasson, JT.; Besner, D. Access to the internal lexicon. In: Dornic, S., editor. *Attention and performance VI*. Hillsdale, NJ: Erlbaum; 1977. p. 535–555.
- Cutler A, Fay D. One mental lexicon, phonologically arranged: Comments on Hurford's comments. *Linguistic Inquiry* 1982;13:107–113.
- Davis CJ, Taft M. More words in the neighborhood: Interference in lexical decision due to deletion neighbors. *Psychonomic Bulletin and Review* 2005;12:904–910. [PubMed: 16524009]
- de Saussure, F. *Third course of lectures on general linguistics*. Komatsu, E., editor; Harris, R., translator. Oxford: Pergamon Press; 1910/1993.
- Dehaene S, Cohen L. Cultural recycling of cortical maps. *Neuron* 2007;56:384–398. [PubMed: 17964253]
- Delattre M, Bonin P, Barry C. Written spelling to dictation: Sound-to-spelling regularity affects both writing latencies and durations. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 2006;32:1330–1340.
- del Viso S, Igoa JM, García-Albea JE. On the autonomy of phonological encoding: Evidence from slips of the tongue in Spanish. *Journal of Psycholinguistic Research* 1991;20:161–185.
- Dell GS. A spreading activation theory of retrieval in sentence production. *Psychological Review* 1986;93:283–321. [PubMed: 3749399]
- Dell GS. The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language* 1988;27:124–142.
- Dell GS. Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes* 1990;4:313–349.
- Dell GS, Burger LK, Svec WR. Language production and serial order: A functional analysis and a model. *Psychological Review* 1997;104:123–147. [PubMed: 9009882]
- Dell, GS.; Gordon, JK. Neighbors in the lexicon: Friends or foes?. In: Schiller, NO.; Meyer, AS., editors. *Phonetics and phonology in language comprehension and production: Differences and similarities*. New York: Mouton de Gruyter; 2003.

- Dell GS, Reich PA. Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior* 1981;20:611–629.
- Dell GS, Schwartz MF, Martin N, Saffran EM, Gagnon DA. Lexical access in aphasic and nonaphasic speakers. *Psychological Review* 1997;104:801–838. [PubMed: 9337631]
- Diaz-Empananza, I. Selecting the number of replications in a simulation study. Manuscript, University of the Basque Country. 1996. *Econometrics* 9612006, EconWPAhttp://econpapers.repec.org/paper/wpawuwpem/9612006.htm
- Dunn, LM.; Dunn, LM. Peabody picture vocabulary test-revised. Circle Pines, MN: American Guidance Service; 1981.
- Fay D, Cutler A. Malapropisms and the structure of the mental lexicon. *Linguistic Inquiry* 1977;8:505–520.
- Ferreira VS, Humphreys KR. Syntactic influence on lexical and morphological processing in language production. *Journal of Memory and Language* 2001;44:52–80.
- Fischer-Baum S, McCloskey M, Rapp B. Representation of letter position in spelling: Evidence from acquired dysgraphia. submitted. Manuscript submitted for publication.
- Folk JR, Jones AC. The purpose of lexical/sublexical interaction during spelling: Further evidence from dysgraphia and articulatory suppression. *Neurocase* 2004;10:65–69. [PubMed: 15849162]
- Fromkin VA. The non-anomalous nature of anomalous utterances. *Language* 1971;47:27– 52.
- Gagnon DA, Schwartz MF, Martin N, Dell GS, Saffran EM. The origins of formal paraphasias in aphasics' picture naming. *Brain and Language* 1997;59:450–472. [PubMed: 9299072]
- Graham NL, Patterson K, Hodges JR. The impact of semantic memory impairment on spelling: Evidence from semantic dementia. *Neuropsychologia* 2000;38:143–163. [PubMed: 10660226]
- Garrett, MF. The analysis of sentence production. In: Bower, GH., editor. *The psychology of learning and motivation: Advances in research and theory*. New York: Academic Press; 1975. p. 133-177.
- Garrett, MF. Levels of processing in sentence production. In: Butterworth, B., editor. *Language production (vol. I): Speech and talk*. New York: Academic Press; 1980. p. 177-220.
- German DJ, Newman RS. The impact of lexical factors on children's word finding errors. *Journal of Speech, Language & Hearing Research* 2004;47:624–636.
- Goldrick M. Limited interaction in speech production: Chronometric, speech error, and neuropsychological evidence. *Language and Cognitive Processes* 2006;21:817–855.
- Goldrick M, Rapp B. A restricted interaction account (RIA) of spoken word production: The best of both worlds. *Aphasiology* 2002;16:20–55.
- Goldrick M, Rapp B. Lexical and post-lexical phonological representations in spoken production. *Cognition* 2007;102:219–260. [PubMed: 16483561]
- Gordon JK. Phonological neighborhood effects in aphasia speech errors: Spontaneous and structured contexts. *Brain and Language* 2002;82:113–145. [PubMed: 12096871]
- Hanley JR, Dell GS, Kay J, Baron R. 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* 2004;21:147–158.
- Hanley JR, Kay J, Edwards M. Imageability effects, phonological errors, and the relationship between auditory repetition and picture naming: Implications for models of auditory repetition. *Cognitive Neuropsychology* 2002;19:193–206.
- Hanna, PP.; Hanna, JS.; Hodges, RE. *Phoneme-grapheme correspondences as cues to spelling improvement*. Washington, DC: US Government Printing Office; 1966.
- Harley TA. A critique of top-down independent levels models of speech production: Evidence from non-plan-internal errors. *Cognitive Science* 1984;8:191–219.
- Harley TA. Environmental contamination of normal speech. *Applied Psycholinguistics* 1990;11:45–72.
- Harley TA, Bown HE. What causes a tip-of-the-tongue state? Evidence for lexical neighbourhood effects in speech production. *British Journal of Psychology* 1998;89:151–174.
- Harley TA, MacAndrew SBG. Constraints upon word substitution speech errors. *Journal of Psycholinguistic Research* 2001;30:395–417. [PubMed: 11529522]
- Hotopf, N. Slips of the pen. In: Frith, U., editor. *Cognitive processes in spelling*. London: Academic Press; 1980. p. 287-307.

- Hurford JR. Malapropisms, left-to-right listing, and lexicalism. *Linguistic Inquiry* 1981;12:419–423.
- Jescheniak JD, Levelt WJM. Word frequency effects in spoken production: Retrieval of syntactic information and phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 1994;20:824–843.
- Jescheniak JD, Meyer AS, Levelt WJM. Specific-word frequency is not all that counts in speech production: Comments on Caramazza, Costa, et al. (2001) and new experimental data. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 2003;29:432–438.
- Kittredge AK, Dell GS, Verkuilen J, Schwartz MF. Where is the effect of lexical frequency in word production? Insights from aphasic picture naming errors. *Cognitive Neuropsychology* 2008;25:463–492. [PubMed: 18704797]
- Kelly, MH. Unpublished doctoral dissertation. Cornell University; Ithaca, NY: 1986. On the selection of linguistic options.
- Kelly MH. Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review* 1992;99:349–364. [PubMed: 1594729]
- Kelly MH. Indirect representation of grammatical class at the lexeme level. *Behavioral and Brain Sciences* 1999;23:49–50.
- Kohn SE, Smith KL. Distinctions between two phonological output deficits. *Applied Psycholinguistics* 1994;15:75–95.
- Leuninger, H.; Keller, J. Some remarks on representational aspects of language production. In: Hillert, D., editor. *Linguistics and cognitive neuroscience: Theoretical and empirical studies on language disorders*. Opladen: Westdeutscher Verlag; 1994. p. 83-110.
- Levelt WJM. Accessing words in speech production: Stages, processes, and representations. *Cognition* 1992;42:1–22. [PubMed: 1582153]
- Levelt WJM. Models of word production. *Trends in Cognitive Sciences* 1999;3:223–232. [PubMed: 10354575]
- Levelt WJM, Roelofs A, Meyer AS. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 1999;22:1–75. [PubMed: 11301520]
- Luce PA, Pisoni DB. Recognizing spoken words: The neighborhood activation model. *Ear & Hearing* 1998;19:1–36. [PubMed: 9504270]
- MacKay, DG. *The organization of perception and action: A theory for language and other cognitive skills*. New York: Springer-Verlag; 1987.
- Magnuson, JS.; Mirman, D.; Strauss, T. Why do neighbors speed visual word recognition but slow spoken word recognition?. Paper presented at *Architectures and Mechanisms for Language Processing (AMLaP)*; Turku, Finland. 2007.
- Martin N, Dell GS, Saffran EM, Schwartz MF. Origins of paraphasias in deep dysphasia: Testing the consequences of a decay impairment to an interactive spreading activation model of lexical retrieval. *Brain and Language* 1994;47:609–660. [PubMed: 7859057]
- McCarthy R, Warrington EK. A two-route model of speech production: Evidence from aphasia. *Brain* 1984;107:463–485. [PubMed: 6722512]
- McClelland JL, Elman J. The TRACE model of speech perception. *Cognitive Psychology* 1986;18:1–86. [PubMed: 3753912]
- McClelland JL, Rumelhart DE. An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review* 1981;88:375–407.
- McCloskey M, Macaruso P, Rapp B. Grapheme-to-lexeme feedback in the spelling system: Evidence from a dysgraphic patient. *Cognitive Neuropsychology* 2006;23:278–307.
- Miceli G, Capasso R. Spelling and dysgraphia. *Cognitive Neuropsychology* 2006;23:110–134.
- Miller, D.; Ellis, AW. Speech and writing errors in “neologistic jargonaphasia”: A lexical activation hypothesis. In: Coltheart, M.; Sartori, G.; Job, K., editors. *The cognitive neuropsychology of language*. Hove, UK: Lawrence Erlbaum; 1987. p. 253-271.
- Monaghan P, Chater N, Christiansen MH. The differential contribution of phonological and distributional cues in grammatical categorization. *Cognition* 2005;96:143–182. [PubMed: 15925574]
- Newman RS, German DJ. Effects of lexical factors on word naming among normal-learning children and children with word-finding disorders. *Language and Speech* 2002;43:285–317.

- Newman RS, German DJ. Lifespan effects of lexical factors on oral naming. *Language and Speech* 2005;48:123–156. [PubMed: 16411502]
- Newman RS, Sawusch JR, Luce PA. Do postonset segments define a lexical neighborhood? *Memory & Cognition* 2005;33:941–960.
- Nickels, L. *Spoken word production and its breakdown in aphasia*. United Kingdom: Psychology Press; 1997.
- Nooteboom, SG. The tongue slips into patterns. In: Sciarone, AG.; van Essen, AJ.; van Raad, AA., editors. *Leyden studies in linguistics and phonetics*. The Hague: Mouton; 1969. p. 114-132. reprinted in V. A. Fromkin (ed.) (1973) *Speech errors as linguistic evidence* (pp. 144-156). The Hague: Mouton
- O'Seaghdha PG, Marin JW. Phonological competition and cooperation in form-related priming: Sequential and nonsequential processes in word production. *Journal of Experimental Psychology: Human Perception and Performance* 2000;26:57–73. [PubMed: 10696605]
- O'Toole, M.; Oberlander, J.; Shillcock, R. The age-complicity hypothesis: A cognitive account of some historical linguistic data. *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*; Hillsdale, NJ: Lawrence Erlbaum; 2001. p. 716-719.
- Perea M. Orthographic neighbours are not all equal: Evidence using an identification technique. *Language and Cognitive Processes* 1998;13:77–90.
- Rapp B, Caramazza A. On the distinction between deficits of access and storage: A question of theory. *Cognitive Neuropsychology* 1993;10:113–141.
- Rapp B, Caramazza A. From graphemes to abstract letter shapes: Levels of representation in written spelling. *Journal of Experimental Psychology: Human Perception and Performance* 1997;23:1130–1152. [PubMed: 9269731]
- Rapp B, Epstein C, Tainturier MJ. The integration of information across lexical and sublexical processes in spelling. *Cognitive Neuropsychology* 2002;19:1–29.
- Rapp B, Goldrick M. Discreteness and interactivity in spoken word production. *Psychological Review* 2000;107:460–499. [PubMed: 10941277]
- Rapp B, Goldrick M. Feedback by any other name is still interactivity: A reply to Roelofs' comment on Rapp & Goldrick (2000). *Psychological Review* 2004;111:573–578.
- Rastle, K. Visual word recognition. In: Gaskell, MG., editor. *The Oxford handbook of psycholinguistics*. Oxford: Oxford University Press; 2007. p. 71-87.
- Roelofs A. A spreading-activation theory of lemma retrieval in speaking. *Cognition* 1992;42:107–142. [PubMed: 1582154]
- Roelofs A. The WEAVER model of word-form encoding in speech production. *Cognition* 1997;64:249–284. [PubMed: 9426503]
- Roelofs A. Error biases in spoken word planning and monitoring by aphasic and nonaphasic speakers: Comment on Rapp and Goldrick (2000). *Psychological Review* 2004a;111:561–572. [PubMed: 15065924]
- Roelofs A. Comprehension-based versus production-internal feedback in planning spoken words: A rejoinder to Rapp and Goldrick (2004). *Psychological Review* 2004b;111:579–580.
- Roelofs A, Meyer AS, Levelt WJM. A case for the lemma-lexeme distinction in models of speaking: Comment on Caramazza and Miozzo (1997). *Cognition* 1998;69:219–230. [PubMed: 9894405]
- Romani C, Olson A, Ward J, Ercolani MG. Formal lexical paraphasias in a single case study: How “masterpiece” can become “misterpieman” and “curiosity” “suretoy. *Brain & Language* 2002;83:300–334. [PubMed: 12387800]
- Rossi M, Degare EP. Lapsus linguae: Word errors or phonological errors? *International Journal of Psycholinguistics* 1995;11:5–38.
- Roux S, Bonin P. Neighborhood effects in spelling in adults. *Psychonomic Bulletin & Review* 2009;16:369–373. [PubMed: 19293109]
- Sage K, Ellis AW. Lexical influence in graphemic buffer disorder. *Cognitive Neuropsychology* 2004;21:381–400.
- Schwartz MF, Wilshire CE, Gagnon DA, Polansky M. Origins of nonword phonological errors in aphasic picture naming. *Cognitive Neuropsychology* 2004;21:159–186.

- Sevaid CA, Dell GS. The sequential cuing effect in speech production. *Cognition* 1994;53:91–127. [PubMed: 7805353]
- Sevaid CA, Dell GS, Cole JS. Syllable structure in speech production: Are syllables chunks or schemas? *Journal of Memory and Language* 1995;34:807–820.
- Shallice T, Rumiat RI, Zadini A. The selective impairment of the phonological output buffer. *Cognitive Neuropsychology* 2000;17:517–546.
- Shattuck-Hufnagel S. The role of word structure in segmental serial ordering. *Cognition* 1992;42:213–259. [PubMed: 1582157]
- Shi R, Morgan JL, Allopenna P. Phonological and acoustic bases for earliest grammatical category assignment: A cross-linguistic perspective. *Journal of Child Language* 1998;25:169–201. [PubMed: 9604573]
- Silverberg, NB. Unpublished doctoral dissertation. University of Arizona; Tucson, AZ: 1998. Word form parameters for lexical retrieval in language production.
- Snodgrass JG, Vanderwart M. A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory* 1980;6:174–215. [PubMed: 7373248]
- Stemberger, JP. An interactive activation model of language production. In: Ellis, AW., editor. *Progress in the psychology of language*. Vol. 1. Hillsdale, NJ: Erlbaum; 1985. p. 143–186.
- Stemberger JP. Neighbourhood effects on error rates in speech production. *Brain and Language* 2004;99:413–422. [PubMed: 15172557]
- Tainturier, M-J.; Rapp, B. The spelling process. In: Rapp, B., editor. *The handbook of cognitive neuropsychology: What deficits reveal about the human mind*. Philadelphia: Psychology Press; 2001. p. 263–289.
- Tamariz, M. Configuring the phonological organization of the mental lexicon using syntactic and semantic information. In: Bara, BG.; Barsalou, L.; Bucciarelli, M., editors. *Proceedings of the 27nd annual conference of the cognitive science society*; Hillsdale, NJ: Lawrence Erlbaum; 2005. p. 2145–2150.
- Tweney RD, Tkacz S, Zaruba S. Slips of the tongue and lexical storage. *Language and Speech* 1975;18:388–396. [PubMed: 1228358]
- Vigliocco G, Harsuiker RJ. The interplay of meaning, sound, and syntax in sentence production. *Psychological Bulletin* 2002;128:442–472. [PubMed: 12002697]
- Vigliocco G, Kita S. Language-specific properties of the lexicon: Implications for learning and processing. *Language and Cognitive Processes* 2006;21:790–816.
- Vitevitch MS. The neighborhood characteristics of malapropisms. *Language and Speech* 1997;40:211–228. [PubMed: 9509578]
- Vitevitch MS. Influence of onset density on spoken-word recognition. *Journal of Experimental Psychology: Human Perception and Performance* 2002a;28:270–278. [PubMed: 11999854]
- Vitevitch MS. The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory and Cognition* 2002b;28:735–747.
- Vitevitch MS, Ambrüster J, Chu S. Sublexical and lexical representations in speech production: Effects of phonotactic probability and onset density. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 2004;30:514–529.
- Vitevitch MS, Luce PA. When words compete: Levels of processing in spoken word recognition. *Psychological Science* 1998;9:325–329.
- Vitevitch MS, Luce PA. Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language* 1999;40:374–408.
- Vitevitch MS, Sommers M. The facilitative influence of phonological similarity and neighborhood frequency in speech production in younger and older adults. *Memory & Cognition* 2003;31:491–504.
- Vitevitch MS, Stamer MK. The curious case of competition in Spanish speech production. *Language and Cognitive Processes* 2006;21:760–770. [PubMed: 18802492]
- Warker JA, Dell GS. Speech errors reflect newly learned phonotactic constraints. *Journal of Experimental Psychology: Learning, Memory and Cognition* 2006;32:387–398.

- Whitney C. How the brain encodes the order of letters in a printed word: The SERIOL model and selective literature review. *Psychonomic Bulletin & Review* 2001;8:221–243. [PubMed: 11495111]
- Wing, AM.; Baddeley, AD. Spelling errors in handwriting: A corpus and a distributional analysis. In: Frith, U., editor. *Cognitive processes in spelling*. London: Academic Press; 1980. p. 251-285.
- Ziegler JC, Muneaux M, Grainger J. Neighborhood effects in auditory word recognition: Phonological competition and orthographic facilitation. *Journal of Memory and Language* 2003;48:779–793.

Appendix A: Coding of Serial Position

Segments in words of varying length (numbered from left-to-right) were assigned to 5 common positions (after Wing & Baddeley, 1980) as follows:

Word length	Assignment of segments to position				
	1	2	3	4	5
1	1				
2	1				2
3	1		2		3
4	1		2,3		4
5	1	2	3	4	5
6	1	2	3,4	5	6
7	1	2,3	4	5,6	7
8	1	2,3	4,5	6,7	8
9	1	2,3	4,5,6	7,8	9
10	1,2	3,4	5,6	7,8	9,10
11	1,2	3,4	5,6,7	8,9	10,11
12	1,2	3,4,5	6,7	8,9,10	11,12
13	1,2	3,4,5	6,7,8	9,10,11	12,13
14	1,2	3,4,5	6,7,8,9	10,11,12	13,14
15	1,2,3	4,5,6	7,8,9	10,11,12	13,14,15

N.B.: 15 segments were sufficient to cover all targets and errors as well as the set of relevant entries in CELEX.

Appendix B: Determining Predictions of Each Architecture

Architectures 1+2: Matching for SOI

We used the CELEX database as our simulated lexicon. We selected this as the most complete and current representation of the lexicon of English speakers. Our chance rates are therefore based on the assumption that CELEX is an accurate specification of the content of each individual's lexicon. See Kittredge et al. (2008) for an alternative analysis method that does not rely on sampling from a specified lexicon.

To generate baseline rates for any architecture that assumes segmental overlap alone influences the activation of neighbors for each target-error pair (e.g., bus-buzz) we identified a pool of potential pseudoerrors. Specifically, for each target-error pair we identified in CELEX all words that had within 10% of the same degree of segmental overlap of as the actual target-error pair (e.g., for the position independent architecture, for the target-error pair “bus-buzz” this would include words such as base, buzz, cuffs, such). Note, of course, that because targets

and pseudoerrors were matched to the actual target-error, the set of candidate pseudoerrors includes the actual error.

Using a Monte Carlo method, we then compared the overlap/similarity rates of the actual target-error pairs to the distribution of rates corresponding to sets randomly drawn from the pseudoerror pool. On each iteration of the simulation, for each target word, a random pseudoerror was selected from its corresponding pseudoerror set. The random pairing of each target with a pseudoerror was repeated ten thousand times to provide an estimate of the probability distribution predicted by the baseline hypothesis. Following the method of Diaz-Emparanza (1996), this number of random pairings should be very likely to provide a highly accurate estimate (within 0.5%) of the true 95th percentile of the baseline distribution. The difference between observed overlap rates and chance-generated rates was then evaluated by comparing the observed rates of overlap to the 95th percentile for the target-pseudoerror distribution (estimating the (one-tailed) cutoff for rates generated by chance).

Architectures 3–7: Matching along additional dimensions

To examine the predictions of theories incorporating SOI plus an additional dimension of structure, the generation of target-pseudoerror sets was modified. In addition to selecting pseudoerrors that matched the actual target-error SOI (within 10%), each target was (probabilistically) paired with a pseudoerror that matched along an additional dimension of structure. The probability of selecting a pseudoerror that matched along this additional dimension was set to reflect the probability observed in the actual target-error pairs. For example, for Case SP1, the relative probability of selecting pseudoerrors higher vs. lower in frequency than the target was set so that in 10,000 random target-pseudoerror pairings, the mean of the resulting distribution of target-pseudoerror pairs (58%) was quite similar to what was observed in SP1's errors (57%). The resulting target-pseudoerror sets were then evaluated along the other dimensions of lexical and phonological structure that were not explicitly implemented.

To determine the appropriate probability levels for each architecture, we first determined the baseline probability that target-pseudoerror pairs would match along this additional dimension of structure (assuming that all pseudoerrors matching the target-error SOI had an equal probability of being selected). The relative probability of selecting pseudoerrors that matched this additional dimension of structure was then increased and the probability of target-pseudoerror pairs matching was recalculated. This was repeated until the probability that target-pseudoerror pairs would match along this dimension of structure was approximately equal to the rate observed in the actual target-error pairs.

More specifically, we estimated the baseline probability that for some target t a randomly selected word from the pseudoerror set E_t would match the target along an additional dimension

of structure by:
$$\frac{\sum_{i \in E_t} w_i}{\sum_{i \in E_t} c_i}$$

where w_i is the weight of pseudoerror i ; $c_i = 1$ for pseudoerrors matching the additional dimension of structure and 0 otherwise. Initially, all pseudoerrors had an equal weight of 1; the baseline probability is therefore simply the proportion of errors that exhibit the particular relationship to the target. The baseline probability across the entire set was simply the average baseline probability over all targets.

To match the properties of the observed target-error pairs, the weight of all pseudoerrors matching the target along this additional dimension of structure was increased from 1.0 by 0.05

increments until the predicted probability met or exceed that observed in the actual target-error pairs (the weight of all pseudoerrors not matching the target along this additional dimension of structure was held constant at 1.0).

Once this new weighting had been determined, we utilized a Monte Carlo method similar to that used for Architectures 1 + 2. On each of the simulation's 10,000 iterations, for each target word a random pseudoerror was selected from its corresponding pseudoerror set (utilizing the new weighting determined above). Observed overlap rates were then compared to the 95th percentile for the target-pseudoerror distribution.

Note that this analysis assumes a categorical distinction between pseudoerrors that do vs. do not exhibit a particular relationship to the target. These were defined as follows. For Architecture 3, pseudoerrors were divided into words higher vs. lower in frequency than the target. For Architecture 4, pseudoerrors either did or did not share target grammatical category. For Architecture 5, pseudoerrors were divided into high vs. low overlap categories where high overlap words shared more than 90% of the target's segments in first position (this was necessary as we were using the 5-position scheme that could result in multiple letters sharing a position). Finally, for Architectures 6 and 7, the segmental or syllabic length of pseudoerrors was either equal to or different from the target length.

Figure 1.

Functional organization of spoken and written production. Levels of representation are depicted within text boxes; processes mapping between representational levels are labeled to the right. During L-level selection, semantic and syntactic representations guide the selection of word-sized representations at the L-level (which may or may not be shared across modalities). During Lexical orthographic or phonological spell-out, L-level representations map to modality-specific sub-lexical representations stored in long-term memory (e.g., graphemes/letters vs. phonemes). Post-lexical processes map these long-term memory representations to articulatory/motor plans.

Figure 2.

Architecture for L-level selection and lexical spell-out in spoken production, incorporating various mechanisms to account for the features contributing to activation of formal neighbors (illustrated for target CAT). Degree of activation is shown via thickness of lines for each representation unit; dashes denote low activity levels. Syntactic frames (shown as feature nodes) bias the activation of lexical nodes, enhancing activation of nodes sharing the target's grammatical category. Prosodic frames (shown as consonant C and vowel V nodes) enhance the activation of position-specific sub-lexical representations sharing the target's length and provide a strong boost to initial positions. Note: some features are omitted, including: variation in the strength of L-level to sub-lexical form representations (reflecting lexical frequency) and connections from semantic features to syntactic frames (reflecting semantic constraints on syntactic structure).

Table 1

Distribution of responses in spoken picture naming (SP1) and spelling to dictation (WR1-3).

	Correct	Semantic/Morphologically related word	Non-semantically related word	Nonword	PPE
SP1	82%	6%	5%	7%	n/a
WR1	64%	0%	5%	21%	10%
WR2	54%	0%	4%	11%	32%
WR3	34%	4%	27%	65%	4%

Note: SP1's response distribution is taken from the Snodgrass & Vanderwart (1980) picture set. PPE = phonologically plausible error.

Table 2

Relationships between targets and errors

Target-Error Relationship	SP1	WR1	WR2	WR3
Average position-independent SOI	0.63	0.77	0.73	0.64
% pairs w/error frequency > target	57%	53%	50%	48%
% pairs w/shared grammatical category	92%	81%	73%	78%
Average target-error SOI within position	1	0.77	0.92	0.78
	2	0.17	0.51	0.48
	3	0.46	0.50	0.58
	4	0.28	0.36	0.20
	5	0.49	0.55	0.59
% pairs w/= segment length	49%	63%	63%	39%

Table 3

Comparison of observed values with predictions from **Architecture 1: position-independent segmental encoding only**. The 95th percentile of this distribution estimates the (one-tailed) cutoff for significant differences from rates predicted by this architecture.

		SP1		WR1		WR2		WR3	
Target-Error Relationship		observed	Architecture 1: 95 th %ile	Observed	Architecture 1: 95 th %ile	observed	Architecture 1: 95 th %ile	observed	Architecture 1: 95 th %ile
% pairs w/error frequency > target		57% **	31%	53% **	30%	50% **	23%	48% **	20%
% pairs w/shared grammatical category		92% **	59%	81% **	47%	73% **	52%	78% **	48%
Average target-error SOI within position	1	0.77 **	0.27	0.92 **	0.35	0.78 **	0.33	0.81 **	0.28
	2	0.17	0.18	0.51 **	0.28	0.48 **	0.23	0.15 **	0.12
	3	0.46 **	0.28	0.50 **	0.30	0.58 **	0.35	0.52 **	0.29
	4	0.28 **	0.23	0.36 **	0.23	0.20 *	0.20	0.06	0.12
	5	0.49 **	0.32	0.55 **	0.26	0.59 **	0.29	0.28 *	0.26
% pairs w/= segment length		49% **	25%	63% **	62%	63% (p < .07)	63%	39% **	47%

* indicates the estimated probability of the observed value given the baseline distribution (* = estimated probability <.05, ** = estimated probability <.01).

Table 4

Comparison of observed values with predictions from **Architecture 2: position-specific segmental encoding only**. The 95th percentile of this distribution estimates the (one-tailed) cutoff for significant differences from rates predicted by this architecture.

		SP1		WRI		WR2		WR3	
Target-Error Relationship		observed	Architecture 2: 95 th %ile	observed	Architecture 2: 95 th %ile	observed	Architecture 2: 95 th %ile	observed	Architecture 2: 95 th %ile
% pairs w/error frequency > target		57% **	36%	53% **	35%	50% **	28%	48% **	24%
% pairs w/shared grammatical category		92% **	71%	81% **	62%	73% **	61%	78% **	62%
Average target-error SOI within position	1	0.77 **	0.51	0.92 **	0.67	0.78 **	0.65	0.81 **	0.52
	2	0.17	0.31	0.51	0.55	0.48 *	0.47	0.15	0.22
	3	0.46	0.52	0.50	0.62	0.58	0.67	0.52	0.58
	4	0.28	0.40	0.36	0.50	0.20	0.42	0.06	0.22
	5	0.49	0.64	0.55	0.62	0.59	0.63	0.28	0.51
% pairs w/= segment length		49% **	41%	63% **	45%	63% **	52%	39% **	33%

* indicates the estimated probability of the observed value given the baseline distribution (* = estimated probability < .05, ** = estimated probability < .01).

Table 5

Comparison of observed values with predictions from **Architecture 3: position-specific segmental encoding + lexical frequency**. The 95th percentile of this distribution estimates the (one-tailed) cutoff for significant differences from rates predicted by this architecture.

Target-Error Relationship	SP1		WR1		WR2		WR3	
	observed	Architecture 3: 95 th %ile	observed	Architecture 3: 95 th %ile	observed	Architecture 3: 95 th %ile	observed	Architecture 3: 95 th %ile
% pairs w/error frequency > target (<i>matched</i>)	57%	67%	53%	59%	50%	58%	48%	54%
% pairs w/shared grammatical category	92% **	75%	81% **	65%	73% **	64%	78% **	65%
1	0.77 **	0.49	0.92 **	0.68	0.78 **	0.66	0.81 **	0.52
2	0.17	0.31	0.51	0.54	0.48 **	0.45	0.15	0.22
3	0.46	0.53	0.50	0.62	0.58	0.68	0.52	0.56
4	0.28	0.40	0.36	0.49	0.20	0.41	0.06	0.21
5	0.49	0.63	0.55	0.64	0.59	0.61	0.28	0.51
% pairs w/= segment length	49% **	44%	63% **	49%	63% **	56%	39% *	37%

* indicates the estimated probability of the observed value given the baseline distribution (* = estimated probability <.05, ** = estimated probability <.01).

Table 6

Comparison of observed values with predictions from **Architecture 4: position-specific segmental encoding + grammatical category**. The 95th percentile of this distribution estimates the (one-tailed) cutoff for significant differences from rates predicted by this architecture.

Target-Error Relationship	SP1		WR1		WR2		WR3	
	observed	Architecture 4: 95 th %ile	observed	Architecture 4: 95 th %ile	observed	Architecture 4: 95 th %ile	observed	Architecture 4: 95 th %ile
% pairs w/error frequency > target	57% **	39%	53% **	40%	50% **	30%	48% **	26%
% pairs w/shared grammatical category (<i>matched</i>)	92%	97%	81%	87%	73%	81%	78%	84%
1	0.77 **	0.47	0.92 **	0.65	0.78 **	0.64	0.81 **	0.50
2	0.17	0.30	0.51	0.54	0.48 *	0.47	0.15	0.22
3	0.46	0.51	0.50	0.61	0.58	0.67	0.52	0.55
4	0.28	0.40	0.36	0.50	0.20	0.42	0.06	0.22
5	0.49	0.69	0.55	0.69	0.59	0.66	0.28	0.55
% pairs w/= segment length	49% **	43%	63% **	50%	63% **	55%	39% *	35%

* indicates the estimated probability of the observed value given the baseline distribution (* = estimated probability <.05, ** = estimated probability <.01).

Table 7

Comparison of observed values with predictions from **Architecture 5: position-specific segmental encoding + initial position identity**. The 95th percentile of this distribution estimates the (one-tailed) cutoff for significant differences from rates predicted by this architecture.

SP1			WRI		WR2		WR3	
Target-Error Relationship	observed	Architecture 5: 95 th %ile	observed	Architecture 5: 95 th %ile	observed	Architecture 5: 95 th %ile	observed	Architecture 5: 95 th %ile
% pairs w/error frequency > target	57% **	34%	53 % **	37%	50% **	30%	48% **	25%
% pairs w/shared grammatical category	92% **	66%	81% **	59%	73% **	59%	78% **	59%
1 (<i>matched</i>)	0.77	0.84	0.92	0.37	0.78	0.84	0.81	0.95
Average	0.17	0.28	0.51	0.59	0.48 *	0.47	0.15	0.21
target-error	0.46 *	0.45	0.50 *	0.96	0.58	0.62	0.52 **	0.45
SOI within								
position	0.28	0.35	0.36	0.53	0.20	0.38	0.06	0.16
	0.49	0.53	0.55 *	0.55	0.59 *	0.57	0.28	0.35
% pairs w/= segment length	49% **	39%	63% **	44%	63% **	53%	39% **	32%

* indicates the estimated probability of the observed value given the baseline distribution (* = estimated probability <.05, ** = estimated probability <.01).

Table 8

Comparison of observed values with predictions from **Architecture 6: position-specific segmental encoding + segment length**. The 95th percentile of this distribution estimates the (one-tailed) cutoff for significant differences from rates predicted by this architecture.

Target-Error Relationship	SP1			WR1			WR2			WR3		
	Observed			Architecture 6: 95 th %ile			observed			Architecture 6: 95 th %ile		
	Architecture 6: 95 th %ile			observed			observed			observed		
% pairs w/error frequency > target	57% **	38%	53% **	39%	50% **	30%	48% **	25%				
% pairs w/shared grammatical category	92% **	71%	81% **	66%	73% **	64%	78% **	64%				
Average target-error SOI within position	1 0.77 **	0.49	0.92 **	0.64	0.78 **	0.65	0.81 **	0.50				
	2 0.17	0.32	0.51	0.57	0.48	0.49	0.15	0.23				
	3 0.46	0.54	0.50	0.64	0.58	0.68	0.52	0.57				
	4 0.28	0.42	0.36	0.51	0.20	0.43	0.06	0.23				
	5 0.49	0.63	0.55	0.64	0.59	0.63	0.28	0.51				
% pairs w/= segment length (matched)	49%	57%	63%	69%	63%	70%	39%	46%				

* indicates the estimated probability of the observed value given the baseline distribution (* = estimated probability <.05, ** = estimated probability <.01).

Table 9

Comparison of observed values with predictions from **Architecture 7: position-specific segmental encoding + length in syllables** for case SP1. The 95th percentile of this distribution estimates the (one-tailed) cutoff for significant differences from rates predicted by this architecture.

SP1		
Target-Error Relationship	Observed	Architecture 7: 95 th percentile
% pairs w/error frequency > target	57% **	36%
% pairs w/shared grammatical category	92% **	71%
Average target-error SOI within position	1	0.77 **
	2	0.17
	3	0.46
	4	0.28
	5	0.49
% pairs w/= segment length	49% *	49%
% pairs w/= number of syllables (<i>matched</i>)	85%	92%

* indicates the estimated probability of the observed value given the baseline distribution (* = estimated probability <.05, ** = estimated probability <.01).

Table 10

Influence of neighborhood density on production accuracy.

	Average log frequency		Segment accuracy		Word accuracy	
	High density	Low density	High density	Low density	High density	Low density
SP1	1.02	1.02 $t(308) = 0.02, p > .90$	98%	97% $\chi^2(1, N = 1420) = 3.65, p < .06$	96%	92% $\chi^2(1, N = 310) = 2.50, p < .12$
WR1	0.81	0.80 $t(330) = 0.13, p > .85$	92% **	87% $\chi^2(1, N = 1694) = 10.3, p < .005$	72% *	58% $\chi^2(1, N = 332) = 6.41, p < .02$
WR2	.083	0.83 $t(394) = 0.10, p > .90$	93% **	86% $\chi^2(1, N = 1932) = 21.98, p < .0001$	71% **	55.6% $\chi^2(1, N = 396) = 10.46, p < .005$
WR3	1.14	1.15 $t(106) = 0.16, p > .85$	70%	65% $\chi^2(1, N = 468) = 1.65, p < .20$	67% *	43% $\chi^2(1, N = 108) = 6.31, p < .02$