Oxford Handbooks Online

Connectionist Principles in Theories of Speech Production

Laurel Brehm and Matthew Goldrick

The Oxford Handbook of Psycholinguistics (2 ed.)
Edited by Shirley-Ann Rueschemeyer and M. Gareth Gaskell

Print Publication Date: Aug 2018 Subject: Psychology, Cognitive Psychology
Online Publication Date: Sep 2018 DOI: 10.1093/oxfordhb/9780198786825.013.16

Abstract and Keywords

This chapter focuses on connectionist modeling in language production, highlighting how core principles of connectionism provide coverage for empirical observations about representation and selection at the phonological, lexical, and sentence levels. The first section focuses on the connectionist principles of localist representations and spreading activation. It discusses how these two principles have motivated classic models of speech production and shows how they cover results of the picture-word interference paradigm, the mixed error effect, and aphasic naming errors. The second section focuses on how newer connectionist models incorporate the principles of learning and distributed representations through discussion of syntactic priming, cumulative semantic interference, sequencing errors, phonological blends, and code-switching.

Keywords: localist representations, spreading activation, distributed representations, picture-word interference, mixed error effect, syntactic priming, blends, code-switching

16.1 Introduction

IN psycholinguistics, speech production refers broadly to the processes mapping a message the speaker intends to communicate onto its form. If a speaker wishes to tell someone “The picture I’m looking at is an animal—a feline pet,” these processes allow the speaker to generate the spoken form “cat.” Psycholinguistic theories have focused on *formulation* processes—the construction/retrieval of a plan to produce an utterance. This plan specifies the phonological structure of the utterance (e.g., an accented syllable composed of three segments /k/ /ae/ /t/). Subsequent articulatory/motoric processes execute this plan, producing the actual movements of the speech organs. Theories of these post-formulation processes are not reviewed here (see Byrd & Saltzmann, 2003, for discussion).

Since the mid-1980s (e.g., Dell, 1986; MacKay, 1987; Stemberger, 1985) connectionist architectures have served as the dominant paradigm for characterizing theories of formulation processes. The first section of this chapter examines how two connectionist principles (localist representations and spreading activation) have influenced the development of speech production theories. The use of these principles in framing theories of speech production is discussed, followed by an illustration of how the principles have been used to account for three sets of empirical observations. Although this work has been quite successful in explaining a variety of empirical phenomena, it has failed to incorporate two principles that are central to connectionist research in many other domains: learning and distributed representations. The second section of the chapter reviews three examples of more recent work that incorporate these principles into theories of speech production. (p. 373)

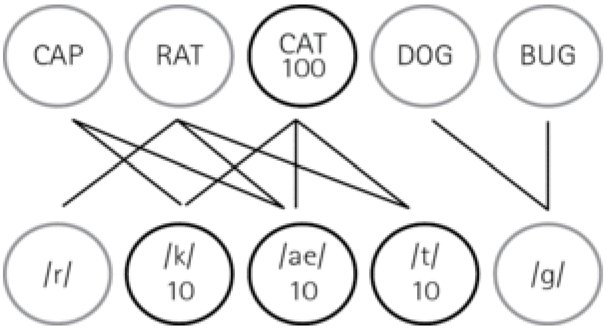
16.2 Spreading activation between localist representations

16.2.1 Localist connectionist principles

Two general connectionist processing principles (after Smolensky, 2000) have guided the bulk of connectionist research in speech production:

1. *Representations are activation patterns*. Mental representations are patterns of numerical activity.
2. *Processing is spreading activation*. Mental processes are transformations of activity patterns by patterns of numerical connections.

To instantiate the first principle, many connectionist speech production theories have assumed that different types of linguistic information are encoded using localist representations (see Page, 2000, for a detailed discussion of the use of such representational structures in connectionist networks). The two basic types of representations are illustrated in Figure 16.1. The first representational type, shown at the top of Figure 16.1, is strictly local; each linguistic object is represented by a single processing unit (e.g., each word has an independent unit such as <CAT>). The second representational type is feature-based (or “semi-local”). In such representations, a small, discrete group of processing units represents each linguistic object (e.g., each word is encoded by a small set of discrete phonemes such as /k/ /ae/ /t/).



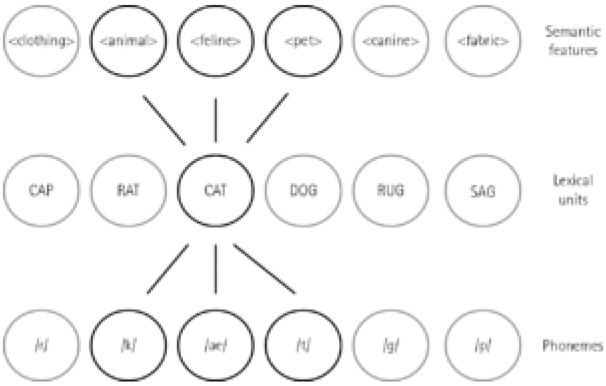
Click to view larger

Fig. 16.1 Illustration of spreading activation between strictly local (top layer) and featural (bottom layer) representations in speech production theories. Lines denote connections between units (here, all connection weights are set to 0.1). Numbers within units denote activation (units without numbers have zero activation).

To instantiate the second principle, the most basic element of processing in connectionist systems (localist as well as non-localist) is spreading activation. Suppose a numerical pattern of activity is imposed on some set of representational units (e.g., in Fig. 16.1, the word unit <CAT>'s activation is set to 100; all other word units are inactive). This activation can then be spread to other units via a set of weighted connections (e.g., in Fig. 16.1, <CAT> (p. 374) is linked to the phoneme units /k/ /ae/ /t/ by connections with weights of 0.1). The amount of activation a unit transmits to other units is simply the product of its activation and the weight on the connection between the units (e.g., 100 * 0.1). The activation of the target units is the sum of this incoming activation (e.g., 100 * 0.1 = 10 for each phoneme unit connected to <CAT>).

16.2.2 A generic localist connectionist framework

Following Rapp and Goldrick (2000), Figure 16.2 provides a generic representational and processing framework to illustrate how these two connectionist principles are instantiated within theories of single word production. First, three broad levels of linguistic structure are represented by numerical patterns of activity over localist representational units. At the top of the figure are semantic representations, specifying the meaning of lexical items in a particular language. Here, a set of semantic features represents each lexical concept (e.g., {animal, feline, pet} for lexical concept {CAT}). These representations provide an interface between more general (non-linguistic) conceptual processing and those processes that specify the linguistic form of an intended message. The bottom of the figure depicts phonological representations; stored, sublexical representations of the spoken form of lexical items. Here, a set of phonemes represents each word's form (e.g., /k/ /ae/ /t/ for the lexical item <CAT>). The relationship between these two representations is mediated by a lexical representation; here, a unitary word-size node (e.g., <CAT>).



Click to view larger

Fig. 16.2 A generic representational framework for speech production. The top layer represents word meaning; the middle, mediating lexical representations; and the bottom, sublexical representations of form. Lines show connections between the representational units for target "cat."

Most current theories of speech production (Garrett, 1980; Levelt, 1992 inter alia) assume that formulation processes are implemented via two stages of activation spreading between these localist representations. The first stage begins with activation of a set of (p. 375) semantic feature units; activation spreads from these units, and the stage ends with the selection of the most strongly activated lexical unit (discussed in more detail next). This corresponds to selecting a lexical item to express the intended message. The second stage begins with the selection of a lexical unit, which spreads activation throughout the production network. This stage ends with the selection of the most strongly activated phoneme units. This corresponds to the construction of an utterance plan for the selected lexical item. It is important to note that these two stages may not be strictly separated; they may interact and overlap in time (e.g., Dell, 1986; a discussion of the interactive mechanisms follows).

As shown by the description here, processing in localist connectionist architectures involves not only the simple spreading of activation between connected units, but also the *selection* of units at particular points in processing. This refers to processes that enhance the activation of units corresponding to one representation relative to that of other units (e.g., enhancement of a single lexical unit; enhancement of a set of phoneme units). By increasing the relative amount of activation that a unit (or group of units) can send on to other representational levels, this enhancement process allows the selected unit(s) to dominate subsequent processing. A variety of spreading activation mechanisms have been used to enhance selected representations. First, some theories propose that the selected representation's activation is simply boosted by adding extra activation to it (e.g., Dell, 1986; Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Rapp and Goldrick, 2000). For example, in Dell's (1986) theory, at selection points the most highly activated node (or nodes) has its activation boosted to a preset high level. The node is then much more active than its competitors, allowing it to dominate processing. The second selection mechanism involves inhibiting the activation of competitors (see Dell & O'Seaghdha, 1994, for a review; for recent discussion of contrasting views of the role of inhibition in production models, see Abdel Rahman & Melinger, 2009; Mahon & Caramazza, 2009). This mechanism is most often realized computationally via lateral inhibitory connections among units of a similar representational type (e.g., Harley, 1995).

With the activation of competitors greatly reduced, the target representation can dominate subsequent processing. A final prominent proposal for enhancing relative activation involves "gating" activation flow. In such systems, representations are not allowed to spread activation to other processing stages until they meet some activation-based response criterion (e.g., a threshold of activation: Laine, Tikkala, & Juhola, 1998; or a relative activation level sufficiently greater than that of competitors: Levelt, Roelofs, & Meyer, 1999). Since only selected representations can influence subsequent processes, they completely dominate processing at these levels.

These selection mechanisms detail how a representation comes to dominate processing. But how does the production system determine which representation to select? Generally, it is assumed that selection processes target a representation that is structurally appropriate. At the lexical level, words must be able to fit into the syntactic structure of the sentence being produced. When producing the head of a noun phrase, it is crucial that a noun (not a verb) be selected. At the phonological level, the selected segments must fit into the appropriate metrical structure. When producing the first segment of <CAT>, it is crucial that an onset consonant (not a vowel, nor a coda consonant such as /ng/) be selected. These structural influences are commonly incorporated into localist connectionist architectures by postulating distinct planning representations. One approach uses structural frames with categorically (p. 376) specified slots to guide selection (e.g., Dell, Burger, & Svec, 1997). Each frame activates its slots in the appropriate sequence. When a slot is active, it enhances the activation of all units within the specified category. This activation boost ensures that structurally appropriate units are selected. For example, at the lexical level, a structural frame for noun phrases would first activate a determiner slot, enhancing the activation of all determiners. Once the determiner has been selected, the frame would activate a noun slot, enhancing the activation of all noun units. This activation support ensures that the most highly activated noun (and not a verb) is selected for production.

It should be noted that the detailed structure of this generic architecture differs from that of many prominent localist connectionist theories. Although these details do not affect the account of the empirical results discussed next, they are briefly reviewed here due to their important implications for other aspects of speech production. First, note that this framework omits any representation of the grammatical properties of lexical items (e.g., grammatical category, number, gender, and so on) which play an important role in speech production (see Ferreira & Morgan, this volume, for further discussion). Second, many theories assume the existence of different numbers and types of localist representations in the production system. With respect to semantic representations, some proposals make use of unitary semantic concept nodes, not sets of features (e.g., {CAT}, instead of {animal, feline, pet}; see Levelt et al., 1999; Roelofs, 1992, for discussion). With respect to phonological representations, many theories assume that in addition to phoneme identity, multiple dimensions of phonological structure are represented (e.g., features, such as [-voice] for /k/; consonant/vowel structure (CVC), such as CVC for "cat"; and metrical structure such as location of stress; see e.g., Dell, 1988; Levelt et al., 1999). Finally, some theories assume that multiple levels of lexical representation are present (e.g., Dell, 1986, 1990; Levelt et al., 1999). A related debate concerns modality specificity: whether a given level of lexical representation is specific to the spoken modality (e.g., Caramazza, 1997) or shared across writing and speaking (e.g., Dell, Schwartz et al., 1997). Theories with two levels of lexical representations generally assume a distinction between modality independent lexical representations (typically referred to as lemmas, which link to grammatical information) and modality dependent representations (typically referred to as lexemes, which link to form information). Those with a single level either assume a single, amodal lexical representation (linking to both grammatical and form information), or distinct lexical representations for spoken and written production (which link to shared grammatical information but distinct form information). (For detailed discussions of the pros and cons of particular proposals for lexical representation(s), see Caramazza, 1997; Caramazza and Miozzo, 1997, 1998; Caramazza, Bi, Costa, & Miozzo, 2004; Caramazza, Costa, Miozzo, & Bi, 2001; Jescheniak, Meyer, & Levelt, 2003; Levelt et al., 1999; Rapp and Caramazza, 2002; Roelofs, Meyer, & Levelt, 1998.)

In spite of differences in the detailed structure of the system, this generic processing framework reflects two core assumptions shared by most speech production theories (see Wheeldon & Konopka, this volume, for a deeper review). First, it makes use of three processing levels that are shared across all current theories (conceptual, lexical, and phonological). Second, it adopts the general assumption (discussed previously) that formulation involves two stages of processing. These core assumptions are sufficient to frame the discussion of the empirical results discussed next. (p. 377)

16.2.3 Applying localist connectionist principles to empirical data

Localist representations and spreading activation mechanisms have been used to account for a wide variety of empirical phenomena. The discussion in this section uses three specific sets of observations to illustrate the influence of these principles on speech production theories. Table 16.1 provides an overview. First, accounts of the contrasting influence of semantic and phonological similarity in picture naming illustrate how connectionist representational principles have influenced production theories. The next section discusses how connectionist processing principles play a crucial role in the explanation of mixed error (p. 378) biases. The final section examines how neurobiologically inspired connectionist principles have been used to understand the consequences of neurological damage.

Table 16.1 Three sets of empirical observations that have been explained using connectionist principles in theories of speech production

Empirical phenomenon	***Connectionist account***

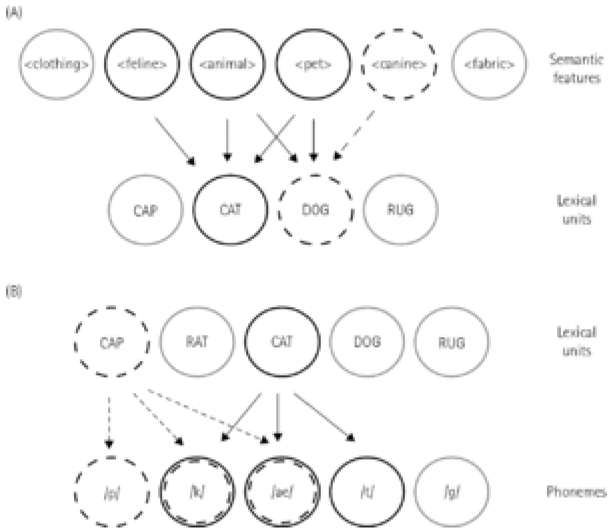
Semantic interference vs. phonological facilitation in picture naming.	*Effect of spreading activation depends on representational structure.*
In picture-word interference experiments, words in the same semantic category as the target interfere with picture naming more than unrelated controls. In contrast, words phonologically related to the target facilitate naming relative to controls.	Spreading activation from semantic representations leads to competition between strictly local lexical representations. Spreading activation from lexical representations converges on overlapping feature-based phonological representations.
Mixed error effect.	*Spreading activation allows processes at distinct representational levels to interact.*
Word errors that overlap with the target in both meaning and form (e.g., “cat”→“rat”) are more likely to occur than predicted based on the rates of purely semantic (e.g., “cat”→“dog”) and purely phonological (e.g., “cat”→“cab”) errors.	Cascading activation allows semantic neighbors to activate their phonological representations, making mixed errors more likely than purely phonological errors at the phoneme level. Feedback allows phonological representations to influence the activation of lexical representations, making mixed errors more likely than purely semantic errors at the lexical level.
Disruptions to speech production.	*Spreading activation between and/or within specific representational levels is disrupted by brain damage.*
Following brain damage, individuals produce varying distributions of error types in speech production.	Disrupting spreading activation lowers activation levels, allowing noise to overwhelm the target representation. Local damage provides a superior account of error patterns compared to global disruptions of processing.

16.2.3.1 Semantic interference versus phonological facilitation in picture naming

An important technique for studying speech production processes has been the picture-word interference task (for a historical overview of this research, and discussion of the importance of this paradigm in the development of theoretical accounts, see Levelt, 1999; Levelt et al., 1999). In this paradigm, participants are presented with pictures (typically, black and white line drawings) depicting common objects and asked to name them. At some point in time close to the presentation of the picture, an interfering stimulus is presented. Either a written word is superimposed on the picture, or an auditory stimulus is presented while participants look at the picture. Although participants are instructed to ignore the interfering stimulus, it can influence the time it takes them to initiate production of the picture's name. In particular, two distinct effects on naming latency are observed depending on the linguistic relationship between the interfering stimulus and the target. (Latencies are also influenced by the time difference between picture and word onset; these effects are not discussed here.)

First, semantic category relationships produce interference. In the seminal study of Schriefers, Meyer, & Levelt (1990), auditory distractor words from the same semantic category as the picture name slowed response times. If the word "dog" was presented prior to the presentation of a picture of a cat, the time to initiate the response "cat" was significantly slower (compared to trials where an unrelated word such as "mop" was presented). However, in contrast to interference from semantically related items, Schriefers et al. found that phonological relationships facilitate picture naming. If the word "cap" was presented at the same time or following presentation of target picture "cat," the response time was significantly faster compared to unrelated trials. Many studies have replicated the basic patterns of facilitation from similar-sounding words (see Starreveld, 2000, for a review) and inhibition from semantic category members (see Abdel Rahman & Melinger, 2009; Roelofs, 1992, for reviews; but see Costa, Alario, & Caramazza, 2005; Costa, Mahon, Savova, & Caramazza, 2003; Mahon & Caramazza, 2009; Mahon, Costa, Peterson, Vargas, & Caramazza, 2007, for discussion of alternative accounts).

Many connectionist theories of speech production have used localist representational principles to account for these effects. Specifically, these theories attribute contrasting effects of semantic and phonological distractors to differences in the structure of lexical and phonological representations (see, e.g., Levelt et al., 1999; Roelofs, 1992). Figure 16.1 illustrates the general properties of this account. As shown in Figure 16.3A, when a semantic distractor is presented, spreading activation from the target and competitor's semantic features diverges onto two distinct lexical representations (e.g., <CAT> and <DOG>). Because lexical representations are strictly local, this spreading activation increases the activation of competitor representations, slowing the selection of the target. As shown in Figure 16.3B, a different situation occurs for phonological distractors. Spreading activation from the target and competitor's lexical representations[1] converges onto shared phonemes (/k/, /ae/ (p. 379)); unshared phonemes receive activation from only one lexical representation. The activation of these shared phonemes can enhance the target's representation, speeding selection of the target's phonological structure compared to a case where no phonemes are shared. By assuming that the degree of localist representation for linguistic structure varies across levels, connectionist theories can account for the distinct patterns of semantic and phonological distractors.



Click to view larger

Fig. 16.3 Semantic interference and phonological facilitation in the picture-word interference task reflect the structure of localist representations. Dashed lines denote activation from the distractor word. (A) Semantic interference stems from competition between coactive unitary lexical representations. (B) Phonological facilitation arises due to the activation of overlapping feature-based phonological representations.

16.2.3.2 The mixed error effect

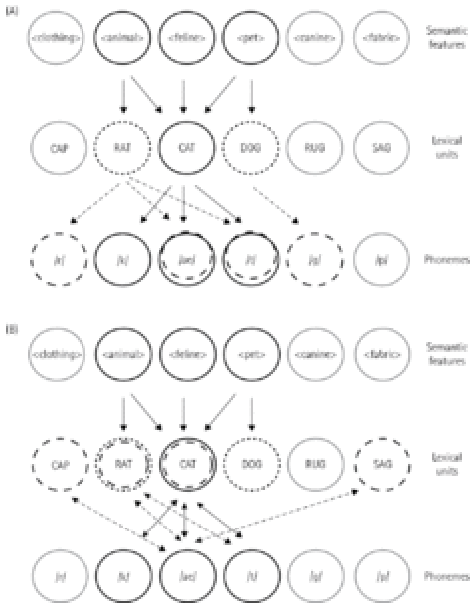
Errors in speech production are often classified in terms of their linguistic relationship to the target. Purely semantic errors (e.g., "cat"→ "dog") are similar in meaning, but not form; purely phonological errors (e.g., "cat"→"cap") share form, but not meaning. The term mixed error is generally used to refer to errors that overlap along both of these dimensions (e.g., (p. 380) "cat"→"rat"). Many studies have observed that mixed errors occur more often than would be predicted by the simple sum of the rates of purely semantic (e.g., "cat"→"dog") and purely phonological (e.g., "cat"→ "cap") errors. This has been observed in studies of spontaneous speech errors (e.g., Harley & MacAndrew, 2001), experimentally induced speech errors (e.g., Brédart & Valentine, 1992), and the production errors of many aphasic individuals (e.g., Rapp & Goldrick, 2000).

This result is unexpected under a fully discrete version of the two-stage framework of speech production discussed here. If we assume that the two stages have a strictly serial relationship, mixed errors should simply be the sum of (independently occurring) semantic and phonological errors. During the first stage, a lexical representation is selected solely based on the intended message. Both mixed and purely semantic competitors should therefore be equally active (e.g., for target "cat," <DOG> should be just as active as <RAT>). If processing is serial and discrete, during the second stage only the phonemes of the selected lexical item are activated. Both mixed and purely phonological competitors should therefore be equally active (e.g., /k/ /ae/ /p/ should be just as active as /r/ /ae/ /t/). Since at neither level of processing are mixed errors more likely than “pure” semantic or phonological errors, this discrete theory cannot account for the mixed error effect.

To produce the mixed error effect, many theories have relied on the connectionist principle of spreading activation. Specifically, the discrete architecture is enhanced by adding two spreading activation mechanisms (e.g., Dell, 1986). These are illustrated in Figure 16.4. The first is cascading activation (Fig. 16.4A). Cascade allows non-selected lexical representations to exert an influence on processing at the phonological level. For example, semantic neighbors (activated via spreading activation from semantic features) are allowed to activate their phonemes (e.g., <RAT> activates /r/). This activation boost makes mixed errors more likely than purely phonological errors (e.g., /r/ is more active than /p/, meaning that “rat” is more active than “cap”).

The second mechanism is feedback (Fig. 16.4B). Feedback systems allow activation from phonological representations to spread back to lexical representations (e.g., /ae/ /t/ activate <RAT>). This can combine with top-down activation from shared semantic features, boosting the activation of mixed competitors relative to that of purely semantic competitors (e.g., because it shares phonemes with the target, <RAT> is more active than <DOG>). By influencing the first stage of processing (i.e., the selection of a lexical item), feedback makes mixed error outcomes more likely to occur than purely semantic errors. The relative contributions and strength of cascading activation and feedback within the speech production system is a matter of some debate (see Goldrick, 2006; Rapp & Goldrick, 2000, for discussion).

Other theories have attributed the mixed error effect not to spreading activation within the production system but to the influence of response monitoring—what could be considered feedback from external processes. One such monitoring system is based in a perceptual loop (e.g., Levelt, 1983). This could halt speech prior to articulation, preventing some of the errors arising during formulation processes from being overtly produced. According to such accounts, since mixed errors are both phonologically and semantically like the target, they are less likely to be detected by the perceptual monitor than corresponding “pure” error types. Mixed errors are therefore more likely to be overtly produced, producing the mixed error effect (for discussion, see Levelt et al., 1999; Roelofs, 2004). **(p. 381)**



Click to view larger

Fig. 16.4 Interaction between levels of processing in speech production produces the mixed error effect. (A) Overlapping semantic features activate semantic neighbors of the target (depicted with dotted lines). Cascade allows these lexical units to activate their phonological representations (shown with dashed lines), producing an advantage for mixed errors. (B) Feedback allows phonological representations to activate lexical representations. Illustrated here is the first step of feedback: the target's phonological representation reactivates the target as well as its lexical neighbors (depicted by dashed lines). (Note that these lexical representations could then, in turn, activate their non-target phonological representations; e.g., SAG could activate /g/.) Feedback from the phonology of the target combines with activation of the target's semantic neighbors (shown by dotted lines), producing an advantage for mixed errors.

(p. 382) 16.2.3.3 Connectionist accounts of speech production impairments

As a consequence of brain damage, many individuals suffer from impaired speech production abilities (see, e.g., the contribution of Schwartz to this volume). Given that connectionist principles reflect (in part) neurobiological processing principles, connectionism may provide a very useful framework for understanding these impairments. Most commonly, researchers have conceptualized impaired speech production performance as reflecting the distortion of the spread of activation within the production network. Theories of damage can be broadly divided into two types: those that involve global alteration of spreading activation, and those that involve alterations that are specific to particular representational levels.

Global damage mechanisms.

In a series of papers, Dell, Martin, Saffran, Schwartz, and colleagues (Dell, Schwartz et al., 1997; Martin, Dell, Saffran, & Schwartz, 1994; Martin & Saffran, 1992; Martin, Saffran, & Dell, 1996; Schwartz & Brecher, 2000) proposed that global alterations to activation spreading could account for the range of patterns of impairment to speech production processes. They proposed two specific damage mechanisms. The first was a reduction of the ability of representational layers in the network to spread activation to one another (the “connection weight” parameter of Dell, Schwartz et al., 1997). If this type of activation spreading is reduced, less activation flows between representational levels. Due to lower levels of activation, noise on processing units can then overwhelm the representation of the correct response, leading to errors. The second mechanism involved a reduction of the ability of units to retain activation over time (“decay” in Dell, Schwartz et al., 1997). Typically, the activation of a unit at a given time step is not just determined by the activation flowing into it from other representational levels but also by its activation at previous time steps. (Note that this can be conceived of as a unit spreading activation back onto itself.) Increasing decay—that is, decreasing the amount of activation that units retain over time—can therefore serve to lower levels of activation, allowing random noise to disrupt the target and produce errors (for further discussion of the potential influence of decay on impairments to speech production, see Harley & MacAndrew, 1992; Wright and Ahmad, 1997).

To test the ability of these two mechanisms of global damage to account for aphasic naming patterns, Dell, Schwartz et al. (1997) constructed a simulation of the formulation processes of English speakers. For 21 individuals with aphasia, the connection strength and decay parameters of this simulation were globally adjusted to see if the simulation could reproduce their error patterns. Specifically, for each of the 21 patients, the simulation’s parameters were globally altered so that it matched (as closely as possible) the patient’s relative proportion of: correct responses; phonologically related (e.g., cat → rat) and unrelated (e.g., cat → dog) semantic errors; phonologically related (e.g., cat → cap) and unrelated (e.g., cat → rug) word errors; and non-word errors (e.g., cat → zat). The results of this parameter-fitting procedure provided some quantitative support for the global damage theory. The simulation was able to fairly closely approximate the individual error distributions (but see Ruml & Caramazza, 2000, for a criticism of the simulation's fit to the data, and Dell, Schwartz, Martin, Saffran, & Gagnon, 2000, for a response to these criticisms).

Not only was the global damage simulation able to reproduce the patients' error patterns, but the parameter fits used to account for the error distributions were able to derive novel predictions about patient performance. As just discussed, the presence of a mixed error (p. 383) effect requires the presence of spreading activation between phonological and lexical representations (either due to lexical to phonological cascade or phonological to lexical feedback). If an individual's error pattern is fit by reducing connection strength, the spreading activation theory of mixed errors predicts an associated reduction of the mixed error effect. Consistent with this prediction, Dell et al. (1997) found that as a group, individuals whose pattern was fit by high connection weights showed a significant mixed error effect, while individuals whose pattern was fit by low connection weights did not.

Local damage mechanisms.

Other theoretical accounts of neurologically impaired speech production have proposed that deficit patterns result from distinct disruptions to specific processes (see, e.g., Ruml, Caramazza, Capasso, & Miceli, 2005, for discussion). Connectionist theories have realized this claim in several diverse ways. One proposal simulates neurological damage by increasing the strength of noise at particular representational levels (Laine et al., 1998; Rapp and Goldrick, 2000). Increased noise can overwhelm target's activation at a particular processing level, producing errors. Another proposal uses localist instantiations of disruption to lexical selection processes (e.g., reducing the amount by which the activation of the selected representation is enhanced: Goldrick & Rapp, 2002; Harley & MacAndrew, 1992; Rapp & Goldrick, 2000; or manipulations of the threshold for lexical selection: Dell, Lawler, Harris, & Gordon, 2004; Laine et al., 1998). Disrupting selection interferes with the normal flow of activation in the production system, leading to phonological and semantic errors.

More recent work within Dell, Schwartz et al.'s two-step model framework implements localist damage by independently weakening the strength of connections between semantic and lexical vs. lexical and phonological levels (Foygel & Dell, 2000; see also Harley & MacAndrew, 1992). Weakening connection strength produces errors by lowering activation levels, allowing noise to overwhelm the activation of the target. This model accounts for the novel predictions made by the parameter fits of Dell, Schwartz et al. (1997) and later work from the same group (Dell, Martin, & Schwartz, 2007; Schwartz, Dell, Martin, Gahl, & Sobel, 2006) shows local damage to have a small but reliable advantage over global mechanisms for capturing word naming errors (see also Hanley, Dell, Kay, & Baron, 2004).

Finally, and perhaps most problematic for global damage proposals, local damage can account for empirically observed error patterns that simply cannot be produced by global damage. Rapp and Goldrick (2000) reviewed the performance of two individuals with deficits to formulation processes (i.e., their comprehension and articulation were intact;

their deficits were in mapping messages onto form). These individuals produced only semantic errors in picture naming. As shown by a number of studies (cited next), this pattern of only semantic errors cannot be produced by simulations incorporating global damage. Similarly, Caramazza, Papagno, and Ruml (2000) review cases where individuals with formulation deficits produce only phonologically related errors (see Goldrick, 2016, for discussion). Global damage simulations also fail to produce this pattern of performance. Global damage predicts that "pure" error patterns should never occur—damage always results in the production of a mixture of error types (e.g., not just semantic errors, but phonologically related word and non-word errors as well). In contrast, simulations with local damage can account for these patterns of errors (so long as there is an appropriate degree of interaction between representational levels; see Goldrick & Rapp, 2002; Rapp & Goldrick, 2000; for discussion). For more detailed qualitative and quantitative critiques of global damage theories, (p. 384) see: Caramazza et al. (2000); Cuetos, Aguado, and Caramazza (2000); Foygel and Dell, (2000); Goldrick (2011); Hanley, Dell, Kay, and Baron (2004); Rapp and Goldrick (2000); Ruml et al. (2005); Ruml, Caramazza, Shelton, and Chialant, (2000); Walker and Hickok (2016). This large body of work leads to the conclusion that impairments to speech production processes are the consequence of local, not global disruptions to processing.

16.3 Distributed representations: Learning and processing

16.3.1 Connectionist principles outside the traditional localist framework

As noted in the introduction, the work reviewed in the previous section differs in two ways from the bulk of connectionist research in other domains. First, these are mainly localist networks which assume that connection weights (specifying how activation spreads in the production system) are largely fixed to values set by the simulation designer. In contrast, learning has played a crucial role in other domains of connectionist research (e.g., Elman et al., 1998). The process of learning is in fact seen as a third general principle of connectionist theories (after Smolensky, 2000).

> **3.** *Learning is innately guided modification of spreading activation by experience.* Knowledge acquisition results from the interaction of:
> - **a.** innate learning rules
> - **b.** innate architectural features
> - **c.** modification of connection strengths with experience

A second divergence is that the research reviewed in the previous section makes use of localist representations, whereas most connectionist research assumes that mental representations are highly distributed patterns of activity as evidenced by the title of the seminal connectionist work *Parallel Distributed Processing* (PDP; see Rumelhart, McClelland, & the PDP Research Group, 1986); for recent reviews, see special issues of *Frontiers in Psychology* (Mayor, Gomez, Chang, & Lupyan, 2014) and *Cognitive Science* (Rogers & McClelland, 2014). In such approaches, the first principle of connectionist processing can be reformulated as:

1. *Representations are distributed activation patterns*. Mental representations are highly distributed patterns of numerical activity.

In fact, learning and distributed representations are often closely connected in connectionist architectures. Many connectionist networks learn using error correction algorithms. In these simulations, the designer specifies the structure of input and output representations and a learning algorithm. The network is then trained using a set of examples pairing input and output patterns (e.g., the network is taught to map the pattern <animal, feline, pet> to /k/ /ae/ /t/). To allow networks to learn complex input-output mappings, many connectionist (p. 385) theories assume the presence of additional internal representations. These are realized using “hidden” units that mediate the relationship between the input and output units (much like the lexical level in Figure 16.2). The structure of these representations is not prespecified in the simulation design. Instead, the representations (i.e., the response patterns of the hidden units) develop over the course of learning the mapping between input and output representations (most prominently via the method of backpropagation of error; see Rumelhart, Durbin, Golden, & Chauvin, 1996; Rumelhart, Hinton, & Williams, 1986, for overviews). Of particular relevance here is that these learned internal representations are often highly distributed (see, e.g., Plaut, McClelland, Seidenberg, & Patterson, 1996). Rather than a single unit or a small discrete set of units responding to input patterns, inputs to these trained networks evoke a highly distributed pattern of activity over the hidden units. In this way, learning and distributed representations are often intertwined in connectionist theories.

These two principles, so crucial to connectionist accounts in other domains, were not incorporated into the localist architectures discussed in the first section. This may in part be a historical artifact. The highly influential model of Dell (1986; the foundation of work such as Dell, Schwartz et al., 1997) was grounded in localist models developed in the early 1980s (Dell’s 1980 thesis, as well as McClelland & Rumelhart, 1981). Such work predated the foundational work in the learning-centered PDP approach (Rumelhart et al., 1986) It may also be due to the nature of the problem: for example, it is more tractable to pose questions regarding relative degrees of interactivity in a network with designer-specified vs. learned connection weights (e.g., Rapp & Goldrick, 2000). Recent work has begun to bridge this gap; the remainder of this chapter considers several examples in detail. The application of connectionist learning mechanisms to problems in sentence production is reviewed first, followed by discussions of processing and selection in distributed representational structures.

16.3.2 Learning and syntactic priming

The term syntactic priming is used here to refer to the observation that speakers repeat the same syntactic structures in successive utterances (this is also referred to as structural priming in the sentence production literature). A typical experimental paradigm for inducing this effect has participants repeat a prime sentence aloud and then describe (on a subsequent trial) a picture depicting an event. Many studies have found that participants' picture descriptions tend to reflect the structure of the prime sentence. For example, if participants repeat a passive prime sentence (e.g., "The building manager was mugged by a gang of teenagers"), they are more likely to describe subsequent pictures using passive constructions (e.g., "The man was stung by a bee") compared to active constructions (e.g., "The bee stung the man"). This priming is syntactic in that it does not appear to rely on the prime and target sentences overlapping in other aspects of linguistic structure such as lexical semantics, argument structure, or prosody, nor does it require explicit memory for the previous utterance (see Pickering & Ferreira, 2008, for a review of the paradigm and basic results).

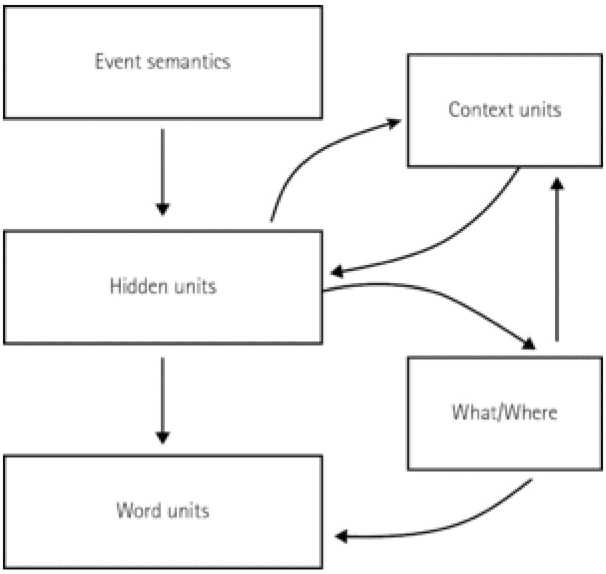
What processing mechanism gives rise to this effect? As noted here, many connectionist theories assume that some activation persists on representational units over time (e.g., Dell, Schwartz et al.'s (1997) decay parameter). One view of syntactic priming is that it is influenced by this persistence; representational units (such as slots in a structural frame) are preactivated by previous productions, allowing them to be more quickly and easily retrieved (p. 386) (e.g., Branigan, Pickering, & Cleland, 1999). However, since units retain only a fraction of their activation, smaller and smaller amounts of activation persist across time steps. The influence of this mechanism is therefore necessarily limited in time. In contradiction to this prediction, Bock and Griffin (2000) found that syntactic priming effects can persist across extremely long lags (e.g., 10 intervening sentences; but see Branigan et al., 1999, for evidence of decay). They interpreted this as support for an alternative account of syntactic priming based on implicit learning. According to this view, syntactic priming is a consequence of learning processes which make longer-term adjustments to the sentence production system—learning processes that might be an extension of the abilities that allowed us to acquire language in the first place Importantly, learning has a natural interpretation within connectionist architectures. In the third connectionist principle detailed here, learning is seen as the adjustment of connection weights. Instead of relying solely on persistent activation, the system can rely on experience-driven changes to the way in which activation flows[2].

A learning-based account of syntactic priming has been examined in simulation experiments by Chang and colleagues (e.g., Chang, 2002; Chang, Dell, & Bock, 2006; Chang, Dell, Bock, & Griffin, 2000). They utilized the simple recurrent network architecture (Elman, 1990; Jordan, 1986), setting up separate pathways for processing meaning and sequencing words that both contribute to the incremental production of a sentence. A simplified version of Chang et al. (2006)'s network is depicted in Figure 16.5.

To produce an utterance, activation of the message units is fixed to a pattern representing a sentence's meaning. The message system relies upon distributed semantic representations that separate event semantics (representing argument number, tense, and aspect) and lexical semantics; lexical semantics are themselves composed of event roles (*where*) and lexical semantics (*what*). The intended message and the learned semantics give rise to the model's sequencing system, such that the selection and sequencing of words for production is based on thematic roles assigned to the message.

At each time step, the activation of the hidden units in the sequencing system (the learned internal representations just discussed) is influenced by this message representation and by a set of context units that are a copy of the hidden units' activation pattern from the previous time step. This recurrence of hidden unit activation patterns allows previous states of the network to influence processing; in effect, providing the network with a memory for what has been already said. The combination of memory for the past and a top-down message allows the model to be flexible in instantiating alternative plans, allowing incremental production of utterances and allowing the model to produce syntactic alternations (e.g., the dative and active/passive alternations) which have multiple possible word orders.

Since the model's internal representations are sensitive to previous states, and since there are separate components for messages, structures, and words, the network can be trained to produce novel sequences of outputs by comparing model predictions to an externally generated utterance (see Elman, 1990; Jordan, 1986, for further discussion). In this case, the network learns to activate, in sequence, the word units corresponding to the intended sentence (e.g., first activating <THE>, then <CAT>, then <WALKS>) with training using the (p. 387) backpropagation algorithm mentioned here. Chang et al. (2006) argue that these changes to hidden units allow the network to acquire language in a human-like fashion, providing evidence that it can simulate a variety of phenomena in child language acquisition.



Click to view larger

Fig. 16.5 Simple recurrent network architecture of Chang et al. (2006). The meaning system is comprised of two parts: event semantics represent propositional structure; what/where units represent the binding between concepts (what); and their event structure roles (where). Word units represent the words that make up the sentence the network produces (the network is trained to activate sequences of word units). A set of hidden units is used to mediate the mapping between these representations. To allow for the production of word sequences, context units (containing copies of the hidden unit activations and what/where units from the previous time step) are allowed to influence the activation of the hidden units.

To simulate syntactic priming, the network received further training corresponding to prime sentences, changing the flow of activation within the network as fast-changing weights were updated. Following this additional training, the simulation was tested using new message inputs. In response to these inputs, the stimulation tended to produce the same structure as the prime sentence, replicating the syntactic priming effect. This influence extended across long intervening lags (e.g., 10 sentences) and across differing prepositions and different tenses, showing that the learning-based theory can account for Bock and Griffin’s (2000) results.

These results illustrate how the third principle of connectionist architectures (experience-driven modification of connection weights) can serve as the basis for a theoretical account for speech production behavior, capturing generalizations about changes to the language system in the short term (priming) and across the long term (acquisition). Such generalizations form the foundation of emerging work, where incorporation of a learning mechanism serves to capture other aspects of language production.

(p. 388) One line of this work focuses on sequencing in production, asking how items (words, phonemes) are selected in the right order. With respect to sentence production, Gordon and Dell (2003) and Dell, Oppenheim, and Kittredge (2008) use syntactic frame constraints (a “syntactic traffic cop”) to govern selection and ordering of lexical items; this model can be trained using error-based learning to mirror typical and aphasic production. Rohde (2002) instantiates a model that is conceptually like the Chang et al. model discussed here; prediction and learning mechanisms provide a bridge between production and comprehension and this allows the model to account for a variety of classic psycholinguistic findings. Models incorporating error-based learning have also been highly successful in word-level production: Warker and Dell (2006) model the acquisition of novel phonotactic constraints; Gupta and Tisdale (2009) model the learning of novel phonological sequences. In both models, error-based learning provides strong coverage for empirical phenomena.

The interplay between short-term and long-term learning is another area of ongoing work. Oppenheim, Dell, and Schwartz (2010) use error-based learning in a simple network to elegantly account for two empirical phenomena: cumulative semantic interference, where retrieval of a word from a set of semantic competitors becomes more difficult over time, and repetition priming, where the second repetition of a word is faster and less-error prone. On the theoretical side, such generalizations about the nature of adaptation of the language system over time also provide the basis for unified theories of language mechanisms (e.g., the *P-Chain*; Dell & Chang, 2013).

16.3.3 Selection in distributed representations

As noted in the first section, localist connectionist architectures commonly incorporate categorically specified planning representations that guide selection of content units (e.g., a noun phrase frame guides selection of a lexical unit representing a determiner <THE> followed by a unit representing a noun <CAT>). Theories making use of learned internal representations (such as the simple recurrent network just described) often eschew such explicit planning representations (for “frame”-less approaches to phonological processing, see Dell, Juliano, & Govindjee, 1993; Gupta & Dell, 1999). Alternative approaches explored in recent work do not eliminate distinct planning representations but do utilize more distributed representations of structure. Such models require attention to the problem of selection over distributed, componential representations. Next, we walk through two frameworks that involve distributed structural representations and use them to highlight the role of selection in capturing several empirical phenomena. The first framework is an oscillator model based in control signal theory (Harris, 2002; Hartley & Houghton, 1996; Vousden, Brown, & Harley, 2000); the second is Gradient Symbolic Computation (Cho & Smolensky, 2016; Goldrick & Chu, 2014; Goldrick, Putnam, & Schwarz, 2016; Smolensky, Goldrick, & Mathis, 2014).

16.3.3.1 Oscillator models

Vousden et al. (2000) focused on the selection of sublexical phonological structure (e.g., selecting onset /k/, vowel /ae/, and coda /t/ for target <CAT>). They posit that selection is controlled by a distributed representation of syllable structure generated by a set of oscillators (based on a more general theory of serial order proposed by Brown, Preece, (p. 389) and Hulme, 2000). A set of repeating oscillators sweep through the same series of values during each syllable, just as on a clock a minute hand sweeps through the same digits every hour (e.g., in every syllable, 15 minutes past represents “onset,” 30 minutes past represents “vowel,” 45 minutes past represents “coda”). This repeating component represents structural similarity across syllables. “Non-repeating” oscillators (i.e., oscillators with extremely long periods) take on distinct values for each syllable, allowing their system to represent the distinction between syllables. This is similar to the hour hand on a clock, which allows one to distinguish 3:30 from 4:30.

This distributed representation of structure is then used to control selection of phonological content. The time-varying oscillator states (both repeating and non-repeating) are combined to generate a dynamic control signal. The system learns a set of weights[3] on connections associating control signal states to phonological structures (following the clock analogy, this means learning that 3:15 corresponds to /k/, 3:30 to /ae/, and so on). During retrieval, the appropriate control signal is provided to the system; the oscillators then automatically generate the sequence of control signal states that cue retrieval of the stored phonological sequence with a winner-take-all selection algorithm.

This proposal shares many properties with localist connectionist planning frames. Both frameworks assume a division between structure and content with categorically specified structural representations (e.g., the repeating oscillator states are predefined to be the same across all syllables). This allows both frameworks to account for structural similarity effects on speech errors, where segments in similar positions are more likely to interact than those in dissimilar positions (e.g., onset consonants are more likely to interact with onset consonants as compared to those in coda; Vousden et al., 2000). By assuming categorically specified structural representations, the effect can be explained as a consequence of representational overlap between segments in similar positions. For example, in virtue of their shared structural representations, onset /k/ will be more like onset /g/ than coda /g/. This similarity leads to a greater likelihood of segments interacting in errors.

Despite the properties shared by the two frameworks, there are important distinctions. As noted by Vousden et al. (2000), the oscillator mechanism provides an explicit account of how successive states of the planning representation are generated—oscillators will cycle through their states automatically, just like a clock that has been wound up will automatically cycle through the minutes of each hour. In contrast, many localist frame-based theories have failed to provide detailed sequencing mechanisms (but see Dell, Burger, et al., 1997).

A second difference stems specifically from properties of distributed representations. As shown by Vousden et al. (2000), speech errors are influenced by distance—all else being equal, closer segments are more likely to interact with one another than more distant segments, meaning that errors occur from the improper selection of an element appearing within a certain time window of the target. This phenomenon is a natural consequence of the use of distributed representations. Vousden et al.'s control signal specifies slots in the planning representation using a time-varying signal. The time-dependence of this signal entails that slots that are temporally close will also have a similar structure. For example, consider a three-syllable word such as "subjective" using the clock face analogy just discussed. Each (p. 390) syllable will be associated with a distinct state of the hour hand on the clock (e.g., "sub" will be 4, "jec" will be 5, and "tive" will be 6), while their internal segments are associated with distinct states of the minute hand (e.g., "s" will be 4:15, "u" will be 4:30, and so on). Because these states are generated by time-varying oscillators, the temporally close first and second syllables will be associated with closer values on the hour hand (e.g., 4, 5) than the first and third syllables (4, 6). This overlap means that errors will be more likely to occur between the first and second syllables than between the first and third. In contrast, localist frame units do not represent similarity in time as an inherent component of the representation. In many of these theories, slots in planning representations are specified by discrete, atomic units equal in similarity, allowing the system to represent, for example, the distinction between the onsets of the first, second, and third syllables but not to encode the fact that the first and second are produced closer in time than the first and third.

Though to date, the control signal theory has been used only for word-level production, the computational principles are domain-general and should transfer to other levels of production. For example, the example of a three-phoneme word <CAT> has a parallel to a three-constituent sentence such as [$_{S}$ [$_{NP}$ Mary] [$_{VP}$ [$_{V}$ loves] [$_{NP}$ John]]]. A similar model could be implemented to produce this sentence. As in the phonological model, a set of long-period oscillators could represent the order of elements within a string (here: word order, vs. phoneme order). As in the phonological model, oscillators with a shorter period could represent elements belonging to the same class (here: "Mary" and "John," vs. /k/ and /t/). Then, the combination of oscillators instantiates hierarchical structure, as in the multisyllabic example ("subjective"). In the example [$_{S}$ [$_{NP}$ Mary] [$_{VP}$ [$_{V}$ loves] [$_{NP}$ John]]], one-hour oscillators might represent subject NP and VP, while 30-minute oscillators might represent V and object NP. As such, applying sequencing mechanisms from control signal theory to other levels of language production is likely to be a promising area for future research.

In principle, then, control signal theories incorporate the positive aspects of frame-based representations (i.e., categorically specified slots, accounting for positional similarity effects) while increasing their empirical coverage (i.e., accounting for distance effects in errors). This increased empirical coverage can be directly attributed to relying on a connectionist processing principle—distributed representations—during selection of linguistic structure.

16.3.3.2 Gradient Symbolic Computation

Another recent formalism, Gradient Symbolic Computation (GSC; Smolensky, Goldrick, & Mathis, 2014) incorporates distributed representations of both structural positions (as in the oscillator representations of syllable structure) and the elements that fill such structural positions (the sounds occupying a syllable position; the words occupying a syntactic position). This allows for graded activation in all aspects of linguistic representational structure. While this increased representational power allows GSC to capture a variety of empirical phenomena (as discussed next), it also requires a novel approach to selection. The question is how to allow graded activation while regulating it to match the observed limited levels of graded activation in language production.

Unlike the localist selection mechanisms reviewed in §2.2 (e.g., increasing activation to a target, inhibiting activation to competitors), selection in the GSC architecture arises via optimization of a *quantization* constraint that pushes outcomes toward discrete states. This operates in parallel with constraints on structure retrieval and planning. Quantization starts (p. 391) low in the beginning of a simulation, allowing the network to enter in to intermediate processing states with graded activation of a variety of possible outcomes. For example, during lexical access, the network might begin with the desired onset /k/ /ae/ and would activate the various words beginning with those sounds (cat, cab, cap . . .) to graded degrees. Similarly, during syntactic planning, the network might begin with the words "The cat" and then would activate the various sentences beginning with those words ("The cat naps," "The cat chases the mouse" . . .) to graded degrees. Over the course of processing, the strength of quantization is increased. This pushes the model toward a state that discretely selects one of these outcomes out of the many possibilities.

Although the quantization constraint pushes the system *toward* a discrete outcome, it is crucially violable. Allowing these non-discrete final states lets GSC capture some novel empirical phenomena. For example, it has been shown that speech errors retain acoustic and articulatory properties of the original target. When a /k/ is produced in error instead of the target /t/ (i.e., top kop → *k*op kop), it is distinguishable from a /k/ produced as a correct target; the error /k/ exhibits articulatory movements specific to the target /t/ (e.g., Goldrick & Chu, 2014). Goldrick and Chu (2014) analyze this phenomenon using the GSC framework. In their model, when errors occur there is graded coactivation of the target (e.g., /t/) and the error outcome (/k/), reflecting the influence of planning constraints that prefer target properties be retained during production. When this blended phonological plan is mapped onto articulation, the result is a blend of the two representations—a response that might be dominated by one representation (the error) but still retains aspects of another (the original target). In essence, coactivation pushes the final state slightly away from a discrete outcome, though errors and correct targets alike remain identifiable as tokens of (discrete) English phonemes (i.e., they are very close to one phonological representation and distant from another). This limited gradience—or in other words, violable discreteness—is captured inherently by the GSC architecture.

Another phenomenon that might best be described by a non-discrete selection mechanism is code-switching, where bilingual speakers use two languages within a single sentence (see Kroll & Gollan, 2014 for a recent review). Goldrick et al. (2016) outline a GSC analysis of an extreme case of code-switching—doubling constructions. In these, a word and its translation equivalent both appear in a single sentence. In languages that have different constraints on word ordering, the repeated word tokens surround a point of commonality between the two languages. For example, English uses a subject-verb-object (SVO) structure, while Tamil uses a subject-object-verb (SOV) structure. An English-Tamil doubling construction might take the form $S_{English}\ V_{English}\ O_{English}\ V_{Tamil}$, fulfilling the local constraints of the two languages (VO for English; OV for Tamil) in a global blend. An observed example is the utterance "They *gave* me a grant *koḍutaa,"* which has two synonymous verbs (the English "*gave*" and the Tamil "*koḍutaa*") sandwiching the objects "me a grant." Goldrick et al. analyze this as a blend of an SVO *and* an SOV sentence, resulting from planning constraints (preferring the presence of both verbs in the utterance) overriding the preferences of quantization.

These results show how GSC's quantization constraint provides a novel means of specifying how speakers regulate the degree of coactivation of linguistic representations. GSC preserves the ability to account for structure-sensitive language processing while also allowing for fully distributed representations of structural positions and the elements that fill them; these parallel considerations give GSC the power to describe previously uncaptured data within a connectionist framework.

(p. 392) 16.4 Conclusions: Connectionist principles in speech production theories

Connectionist principles have had a profound impact on speech production research. For three decades, production theories have framed their discussion of behavioral data using two assumptions: mental representations are numerical patterns of activity; and processing is spreading activation between these representations. This has not only allowed specific accounts of a variety of empirical phenomena (as illustrated here) but has also supported the development of unified theories of single word production (e.g., WEAVER++; Levelt et al., 1999). As documented in the second section, more recent work has examined how speech production phenomena can be accounted for by using connectionist principles that are quite prominent in other empirical domains (learning and distributed representations). Importantly, much of this new research is cumulative in that it attempts to build on the insights of previous localist approaches. For example, in both syntax (Chang, 2002; Chang et al., 2006) and phonology (Harris, 2002), many distributed, learning-based theories have incorporated the localist theories' distinction between mechanisms that control sequencing (e.g., structural frames) and mechanisms specifying representational content, where distributed architectures which lack this distinction can have great difficulty accounting for the empirical data. A challenge for future work is to determine the crucial features of localist connectionist theories of production and how best to incorporate them within a dynamic, distributed representational framework. One such framework that we believe holds promise for instantiating structure and representational similarity in an integrated way is GSC (Smolensky et al., 2014).

Acknowledgment

Supported by a grant from the National Science Foundation (NSF) (BCS1344269). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

References

Abdel Rahman, R., & Melinger, A. (2009). Semantic context effects in language production: A swinging lexical network proposal and a review. *Language and Cognitive Processes*, *24*, 713–34.

Barry, C., Hirsh, K. W., Johnston, R. A., & Williams, C. L. (2001). Age of acquisition, word frequency, and the locus of repetition priming of picture naming. *Journal of Memory and Language*, *44*, 350–75.

Bock, K., & Griffin, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General*, *129*, 177–92.

(p. 393) Branigan, H. P., Pickering, M. J., & Cleland, A. A. (1999). Syntactic priming in written production: Evidence for rapid decay. *Psychonomic Bulletin & Review*, *6*, 635–40.

Brédart, S., & Valentine, T. (1992). From Monroe to Moreau: An analysis of face naming errors. *Cognition*, *45*, 187–223.

Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review*, *107*, 127–81.

Byrd, D., & Saltzmann, E. (2003). Speech production. In: Arbib, M. (Ed.), *The Handbook of Brain Theory and Neural Networks*, 2nd Edition (pp. 1072–6). MIT Press, Cambridge, MA.

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, *14*, 177–208.

Caramazza, A., Bi, Y., Costa, A., Miozzo, M. (2004). 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*, *30*, 278–82.

Caramazza, A., Costa, A., Miozzo, M., & Bi, Y. (2001). The specific-word frequency effect: Implications for the representation of homophones in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 1430–50.

Caramazza, A., & Miozzo, M. (1997). The relation between syntactic and phonological knowledge in lexical access: Evidence from the "tip-of-the-tongue" phenomenon. *Cognition*, *64*, 309–43.

Caramazza, A., & Miozzo, M. (1998). More is not always better: A response to Roelofs, Meyer, and Levelt. *Cognition*, *69*, 231–41.

Caramazza, A., Papagno, C., & Ruml, W. (2000). The selective impairment of phonological processing in speech production. *Brain and Language*, *75*, 428–50.

Chang, F. (2002). Symbolically speaking: A connectionist model of sentence production. *Cognitive Science*, *26*, 609–51.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, *113*(2), 234–72.

Chang, F., Dell, G. S., Bock, K., & Griffin, Z. M. (2000). Structural priming as implicit learning: A comparison of models of sentence production. *Journal of Psycholinguistic Research*, *29*, 217–29.

Cho, P.-W., & Smolensky, P. (2016). Bifurcation analysis of a Gradient Symbolic Computation model of incremental processing. In: Papafragou, A., Grodner, D., Mirman, D., & Trueswell, J. (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 1487–92). Cognitive Science Society, Austin, TX.

Costa, A., Alario, F.-X., & Caramazza, A. (2005). On the categorical nature of the semantic interference effect in the picture-word interference paradigm. *Psychonomic Bulletin & Review*, *12*, 125–31.

Costa, A., Mahon, B., Savova, V., & Caramazza, A. (2003). Levels of categorization effect: A novel effect in the picture-word interference paradigm. *Language and Cognitive Processes*, *18*, 205–33.

Costa, A., Miozzo, M., & Caramazza, A. (1999). Lexical selection in bilinguals: Do words in the bilingual's two lexicons compete for selection? *Journal of Memory and Language*, *41*, 365–97.

Cuetos, F., Aguado, G., & Caramazza, A. (2000). Dissociation of semantic and phonological errors in naming. *Brain and Language*, *75*, 451–60.

Dell, G. S. (1980). *Phonological and Lexical Encoding in Speech Production: An Analysis of Naturally Occurring and Experimentally Elicited Speech Errors* [Unpublished doctoral dissertation]. University of Toronto, Toronto, ON.

(p. 394) Dell, G. S. (1986). A spreading activation theory of retrieval in sentence production. *Psychological Review*, *93*, 283–321.

Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language*, *27*, 124–42.

Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes*, *4*, 313–49.

Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review*, *104*, 123–47.

Dell, G. S., & Chang, F. (2013). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *369*, 20120394.

Dell, G. S., Juliano, C., & Govindjee, A. (1993). Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive Science*, *17*, 149–95.

Dell, G. S., Lawler, E. N., Harris, H. D., & Gordon, J. K. (2004). Models of errors of omission in aphasic naming. *Cognitive Neuropsychology*, *21*, 125–45.

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

Dell, G. S., & O'Seaghdha, P. G. (1994). Inhibition in interactive activation models of linguistic selection and sequencing. In: Dagenbach, D., & Carr, T. H. (Eds.), *Inhibitory Processes in Attention, Memory, and Language* (pp. 409–53). Academic Press, San Diego, CA.

Dell, G. S., Oppenheim, G. M., & Kittredge, A. K. (2008). Saying the right word at the right time: Syntagmatic and paradigmatic interference in sentence production. *Language and Cognitive Processes*, *23*, 583–608.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, *104*, 801–38.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (2000). The role of computational models in neuropsychological investigations of language: Reply to Ruml and Caramazza (2000). *Psychological Review*, *107*, 635–645.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179–211.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1998). *Rethinking Innateness*. MIT Press, Cambridge, MA.

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

Garrett, M. F. (1980). Levels of processing in sentence production. In: Butterworth, B. (Ed.), *Language Production: Speech and Talk*, *Vol. I* (pp. 177–220). Academic Press, New York, NY.

Goldrick, M. (2006). Limited interaction in speech production: Chronometric, speech error, and neuropsychological evidence. *Language and Cognitive Processes*, *21*, 817–55.

Goldrick, M. (2011). Theory selection and evaluation in case series research. *Cognitive Neuropsychology*, *28*, 451–65.

Goldrick, M. (2016). Integrating SLAM with existing evidence: Comment on Walker and Hickok (2016). *Psychonomic Bulletin & Review*, *23*, 648–52.

Goldrick, M., & Chu, K. (2014). Gradient co-activation and speech error articulation: Comment on Pouplier and Goldstein (2010). *Language, Cognition and Neuroscience*, *29*, 452–8.

Goldrick, M., Putnam, T., & Schwarz, L. (2016). Coactivation in bilingual grammars: A computational account of code mixing. *Bilingualism: Language and Cognition*, *19*, 857–76.

(p. 395) Goldrick, M., & Rapp, B. (2002). A restricted interaction account (RIA) of spoken word production: The best of both worlds. *Aphasiology*, *16*, 20–55.

Gordon, J. K., & Dell, G. S. (2003). Learning to divide the labor: An account of deficits in light and heavy verb production. *Cognitive Science*, *27*, 1–40.

Gupta, P., & Dell, G. S. (1999). The emergence of language from serial order and procedural memory. In: MacWhinney, B. (Ed.), *Emergentist Approaches to Languages*, pp. 447–81. Erlbaum, Mahweh, NJ.

Gupta, P., & Tisdale, J. (2009). Does phonological short-term memory causally determine vocabulary learning? Toward a computational resolution of the debate. *Journal of Memory and Language*, *61*, 481–502.

Hanley, J. R., Dell, G. S., Kay, J., & Baron, R. (2004). Evidence for the involvement of a nonlexical route in the repetition of familiar words: A comparison of single and dual route models of auditory repetition. *Cognitive Neuropsychology*, *21*, 147–58.

Harley, T. A. (1995). Connectionist models of anomia: A comment on Nickels. *Language and Cognitive Processes*, *10*, 47–58.

Harley, T. A., & MacAndrew, S. B. G. (1992). Modelling paraphasias in normal and aphasic speech. In: *Proceedings of the 14th Annual Meeting of the Cognitive Science Society* (pp. 378–83). Lawrence Erlbaum Associates, Hillsdale, NJ.

Harley, T. A., & MacAndrew, S. B. G. (2001). Constraints upon word substitution speech errors. *Journal of Psycholinguistic Research*, *30*, 395–417.

Harris, H. D. (2002). Holographic reduced representations for oscillator recall: A model of phonological production. In: Gray, W. D., & Schunn, C. D. (Eds), *Proceedings of the 24th Annual Meeting of the Cognitive Science Society* (pp. 423–8). Lawrence Erlbaum Associates, Hillsdale, NJ.

Hartley, T., & Houghton, G. (1996). A linguistically constrained model of short-term memory for nonwords. *Journal of Memory and Language*, *35*, 1–31.

Jescheniak, J. D., Meyer, A. S., & Levelt, W. J. M. (2003). 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*, *29*, 432–8.

Jordan, M. I. (1986). *Serial Order: A Parallel Distributed Processing Approach*. Institute for Cognitive Science Report 8604. University of California, San Diego, CA. Reprinted (1997) in Donahoe, J. W., & Dorsel, V. P. (Eds.), *Neural-Network Models of Cognition: Biobehavioral Foundations*, pp. 221–77. Elsevier Science Press, Amsterdam.

Kroll, J. F., & Gollan, H. (2014). Speech planning in two languages: What bilinguals tell us about language production. In: Goldrick, M., Ferreira, V., & Miozzo, M. (Eds.), *The Oxford Handbook of Language Production*, pp. 165–81. Oxford University Press, Oxford.

Laine, M., Tikkala, A., & Juhola, M. (1998). Modelling anomia by the discrete two-stage word production architecture. *Journal of Neurolinguistics*, *11*, 275–94.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, *14*, 41–104.

Levelt, W. J. M. (1992). Accessing words in speech production: Stages, processes, and representations. *Cognition*, *42*, 1–22.

Levelt, W. J. M. (1999). Models of word production. *Trends in Cognitive Sciences*, *3*, 223–32.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–38.

MacKay, D. G. (1987). *The Organization of Perception and Action: A Theory for Language and Other Cognitive Skills*. Springer-Verlag, New York, NY.

(p. 396) Mahon, B. Z., & Caramazza, A. (2009). Why does lexical selection have to be so hard? Comment on Abdel Rahman and Melinger’s swinging lexical network proposal. *Language and Cognitive Processes*, *24*, 735–48.

Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., & Caramazza, A. (2007). Lexical selection is not by competition: A reinterpretation of semantic interference and facilitation effects in the picture-word interference paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 503.

Martin, N., Dell, G. S., Saffran, E. M., & Schwartz, M. F. (1994). 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*, *47*, 609–60.

Martin, N., & Saffran, E. M. (1992). A computational study of deep dysphasia: Evidence from a single case study. *Brain and Language*, *43*, 240–74.

Martin, N., Saffran, E. M., & Dell, G. S., (1996). Recovery in deep dysphasia: Evidence for a relation between auditory-verbal STM capacity and lexical errors in repetition. *Brain and Language*, *52*, 83–113.

Mayor, J., Gomez, P., Chang, F., & Lupyan, G. (2014). Connectionism coming of age: Legacy and future challenges. *Frontiers in Psychology*, *5*, 187.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, *88*, 375–407.

Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. *Cognition*, *114*, 227–52.

Page, M. (2000). Connectionist modeling in psychology: A localist manifesto. *Behavioral and Brain Sciences*, *23*, 443–512.

Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: A critical review. *Psychological Bulletin*, *134*(3), 427.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, *103*, 56–115.

Rapp, B., & Caramazza, A. (2002). Selective difficulties with spoken nouns and written verbs: A single case study. *Journal of Neurolinguistics*, *15*, 373–402.

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review*, *107*, 460–99.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*, 107–42.

Roelofs, A. (2004). Error biases in spoken word planning and monitoring by aphasic and nonaphasic speakers: Comment on Rapp and Goldrick (2000). *Psychological Review*, *111*, 561–72.

Roelofs, A., Meyer, A. S., & Levelt, W. J. M. (1996). Interaction between semantic and orthographic factors in conceptually driven naming: Comment on Starreveld and La Heij (1995). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 246–51.

Roelofs, A., Meyer, A. S., & Levelt, W. J. M. (1998). A case for the lemma-lexeme distinction in models of speaking: Comment on Caramazza and Miozzo (1997). *Cognition*, *69*, 219–30.

Rogers, T., & McClelland, J. (Eds.) (2014). 2010 Rumelhart Prize Special Issue Honoring James L. McClelland. [Special issue] *Cognitive Science*, *38*(6), 1286–315.

Rohde, D. L. T. (2002). *A Connectionist Model of Sentence Comprehension and Production* [Unpublished PhD thesis]. School of Computer Science, Carnegie Mellon University, Pittsburgh, PA.

(p. 397) Rumelhart, D. E., Durbin, R., Golden, R., & Chauvin, Y. (1996). Backpropagation: The basic theory. In: Smolensky, P., Mozer, M. C., & Rumelhart, D. E. (Eds.), *Mathematical Perspectives on Neural Networks*, pp. 533–66. Lawrence Erlbaum Associates, Mahwah, NJ.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In: Rumelhart, D. E., McClelland, J. L., & the PDP Research Group (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Vol. 1, Foundations* (pp. 318–62). MIT Press, Cambridge, MA.

Rumelhart, D. E., McClelland, J. L., & the PDP Research Group (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Vol. 1, Foundations*. MIT Press, Cambridge, MA.

Ruml, W., & Caramazza, A. (2000). An evaluation of a computational model of lexical access: Comment on Dell et al. (1997). *Psychological Review*, *107*, 609–34.

Ruml, W., Caramazza, A., Capasso, R., & Miceli, G. (2005). Interactivity and continuity in normal and aphasic language production. *Cognitive Neuropsychology*, *22*, 131–68.

Ruml, W., Caramazza, A., Shelton, J. R., & Chialant, D. (2000). Testing assumptions in computational theories of aphasia. *Journal of Memory and Language*, *43*, 217–48.

Schriefers, H, Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language*, *29*, 86–102.

Schwartz, M. F., & Brecher, A. (2000). A model-driven analysis of severity, response characteristics, and partial recovery in aphasics' picture naming. *Brain and Language*, *73*, 62–91.

Schwartz, M. F., Dell, G. S., Martin, N., Gahl, S., & Sobel, P. (2006). A case-series test of the interactive two-step model of lexical access: Evidence from picture naming. *Journal of Memory and Language*, *54*, 228–64.

Smolensky, P. (2000). Grammar-based connectionist approaches to language. *Cognitive Science*, *23*, 589–613.

Smolensky, P., Goldrick, M., & Mathis, D. (2014). Optimization and quantization in gradient symbol systems: A framework for integrating the continuous and the discrete in cognition. *Cognitive Science*, *38*, 1102–38.

Starreveld, P. (2000). On the interpretation of phonological context effects in word production. *Journal of Memory and Language*, *42*, 497–525.

Stemberger, J. P. (1985). An interactive activation model of language production. In: Ellis, A. W. (Ed.), *Progress in the Psychology of Language, Vol. 1* (pp. 143–86). Erlbaum, Hillsdale, NJ.

Vousden, J. I., Brown, G. D. A., & Harley, T. A. (2000). Serial control of phonology in speech production: A hierarchical model. *Cognitive Psychology*, *41*, 101–75.

Walker, G. M., & Hickok, G. (2016). Bridging computational approaches to speech production: The semantic–lexical–auditory–motor model (SLAM). *Psychonomic Bulletin & Review*, *23*, 339–52.

Warker, J. A., & Dell, G. S. (2006). Speech errors reflect newly learned phonotactic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 387.

Wright, J. F., & Ahmad, K. (1997). The connectionist simulation of aphasic naming. *Brain and Language*, *59*, 367–89.

Notes:

([1]) An additional source of activation from word distractors is via sublexical conversion procedures that directly activate phonological representations from orthographic or acoustic input (e.g., Roelofs, Meyer, & Levelt, 1996). In fact, Costa, Miozzo, & Caramazza (1999) argue that these sublexical processes drive the phonological facilitation effect. Regardless of the source of the activation, the presence of facilitation (as opposed to inhibition) derives from the use of feature-based localist representations (such that target and distractor overlap in structure).

([2]) Note that this account is also capable of using persistent activation effects to account for other priming effects that occur only over short lags. However, it does not currently specify why different effects have different priming lags (e.g., in single word production, why repetition priming is found over long lags while semantic priming is not; Barry, Hirsh, Johnston, & Williams, 2001).

([3]) See Harris (2002) for discussion of the limitations of Vousden et al.'s (2000) method and a distributed associative memory proposal for more efficiently storing the relationship between control signals and phonological structure.

Laurel Brehm

Laurel Brehm, Max Planck Institute for Psycholinguistics, the Netherlands

Matthew Goldrick

Matthew Goldrick, Northwestern University, USA