

**A critical review of the behavioral, neuroimaging, and electrophysiological studies  
of co-activation of representations during word production**

Nazbanou Nozari<sup>a,b</sup> and Svetlana Pinet<sup>a</sup>

*<sup>a</sup>Department of Neurology, Johns Hopkins University, Baltimore, USA; <sup>b</sup>Department of  
Cognitive Science, Johns Hopkins University, Baltimore, USA.*

Correspondences should be addressed to:

Nazbanou Nozari, MD, PhD

Address: 1629 Thames Street, Suite 350, Baltimore, MD 21231, USA

Phone: 443-287-1712

Fax: 410-955-0188

E-mail: nozari@jhu.edu

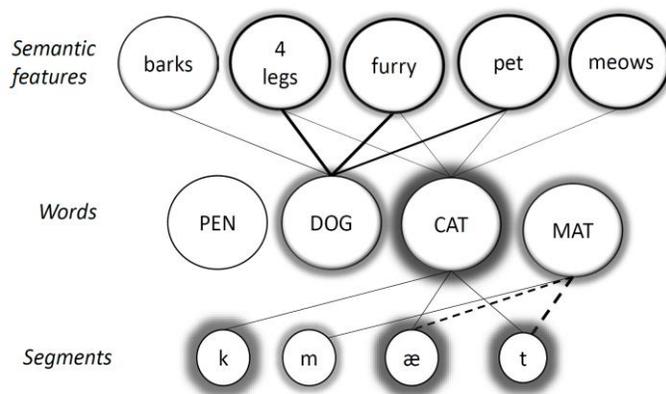
## **Abstract**

A large body of behavioral, neuroimaging, and electrophysiological studies have investigated the consequences of co-activation of representations during word production. Despite such an amazing body of empirical data, it remains unclear how the production system handles co-activated items. In this paper, we review this evidence in a systematic way, and point out three common problems in the interpretations attached to these data. We then discuss alternative approaches which might be more fruitful in understanding the links between such data and the processes of spreading activation, selection, monitoring and control in language production.

**Keywords:** word production; semantic interference; lexical competition; non-competitive selection; monitoring; cognitive control; N400; ERN; fMRI.

## Introduction

Although computational models of language production differ significantly in their specifics, they have a key feature in common: attempting to produce a word leads to the activation of more than just the representations corresponding to that specific word (e.g., Dell, 1986; Levelt, Roelofs, & Meyer, 1999). Instead, often several representations are simultaneously activated, either by receiving feedforward activation through shared semantic features (e.g., “cat”, “dog”, and “rabbit” are all furry four-legged pets), or, in models which allow feedback (e.g., Dell, 1986), by receiving feedback activation through shared segments (e.g., “cat”, “mat”, “hat” all have the same last two segments; see Fig. 1). In this paper, we discuss the consequences of the co-activation of multiple representations during word production. The review suggests that understanding these consequences and their underlying mechanisms is critical for understanding the basic dynamics of the production system, including spreading activation, cascading, selection, monitoring, and control. The first two sections review the behavioral and neural (comprising neuroimaging and electrophysiological) findings. The third section discusses the theoretical interpretations attached to the empirical findings in light of the strength of the underlying assumptions. The last section discusses possible solutions to the problems raised in the third section.



*Fig. 1. Schematic of the co-activation of multiple lexical items during word production using localist representations (e.g., Dell, 1986). Target is “cat”. Lexical nodes “dog” and “mat” are co-activated through the shared semantic features (darker solid connections), and shared segments (darker dashed connections) with “cat”.*

Although Fig. 1 uses a localist representation (i.e., a single node for a lexical concept) for the ease of demonstration, this is not a theoretical commitment. In fact, we use distributed representations to explain some of the consequences of representational overlap. The word “representation” is used in this work to refer to the content (neural and cognitive)—as opposed to processes—involved in producing sound from meaning during word production. For example, by a phonological representation of “cat”, we are referring to neural populations that encode the phonology of cat and their corresponding nodes in a computational model. The word “feature” is used in its broad sense to indicate a unit in the composition of a representation, and not in its narrower sense of physical features, e.g., a tail, shared between two members of the same semantic category. In agreement with the broad literature, we assume that semantic processing is carried out, in part, by reinstating sensory, motor, and emotional states associated with a concept. We also contend that, due to the non-systematic mapping between meaning and sound, an intermediate layer is needed for this mapping. However, we do not insist upon teasing apart “lexical concepts” from “lemmas” (see Kemmerer, 2018; Nozari, to appear; for an extensive discussion of this issue). This is not to imply that such a distinction does not exist, but that both of these representations constitute hidden layers between semantic features and segments, and thus serve a similar computational purpose of mediating the mapping between semantics and phonology. Importantly, both have also been localized to the temporal cortex (albeit more anterior vs. middle temporal gyri, respectively; Kemmerer, 2018), which, as will be reviewed, is critically implicated in both neuroimaging and EEG studies of co-activation of multiple semantically related representations.

### **Behavioral consequences of naming pictures in the presence of co-activated representations**

Several paradigms have been used to manipulate the number of co-activated representations in picture naming. Some of the more commonly used paradigms include Picture-Word Interference (PWI) paradigms (Schriefers, Meyer, & Levelt, 1990), priming paradigms (Lupker, 1988), continuous naming paradigms (Howard, Nickels, Coltheart, & Cole-Virtue, 2006), and blocked cyclic naming paradigms (Belke, Meyer, & Damian, 2005). In addition, certain indices, such as phonological neighborhood density (Sadat, Martin, Costa, & Alario, 2014) and name agreement (Alario et al., 2004) have been used to examine the effects of the co-activation of multiple items on picture naming (see Table 1 for a summary). It is worth mentioning that bilingualism has also been a fruitful avenue in understanding such consequences. But given the large number of studies in that literature and the complexities related to language and code switching, we leave the review of that evidence to another paper in the same issue focused on bilingualism. Instead, we confine our review to paradigms that have investigated the consequences of within-language co-activation of multiple representations during picture naming of objects.

In the PWI paradigm, a target picture is presented along with a distractor word (presented in written or auditory format) which must be ignored while the participant names the target picture. The typical finding in this paradigm is that the mere presence of a distractor interferes with the production of the target compared to when no distractor is present, manifesting as increased errors or longer response times (RTs; e.g., de Zubicaray & McMahon, 2009). Lexical distractors, even when they bear no semantic or phonological relationship to the target, i.e., “unrelated distractors” (e.g., “pen” for target “cat”) interfere with production more than nonlexical distractors such as a row of Xs (Hirschfeld, Jansma, Bölte, & Zwitserlood, 2008). The story is a little more complicated when the effect of distractors bearing different relationships to the target is compared against unrelated distractors. Phonologically-related distractors (e.g.,

“cup” for target “cat”) generally induce facilitation (Schriefers et al., 1990). The effect of semantically related targets has been argued to depend on the kind of semantic relationship between the target and the distractor. Categorically related distractors (e.g., “dog” for “cat”) often induce interference (e.g., Schriefers et al., 1990), whereas non-categorical relationships (e.g., part-whole relationships such as bumper-car; Costa, Alario, & Caramazza, 2005, and associations such as bone-dog; Alario, Segui, & Ferrand, 2000) have been reported to cause facilitation (see also Abel et al., 2009; de Zubicaray, Hansen, & McMahon, 2013).

These opposing effects appear to have different timelines: The facilitation induced by the non-categorical relationships is most robust in early stimulus onset asynchronies (SOAs), when the distractor is presented -450 to -100 ms before the target (or even earlier; see Python, Fargier, & Laganaro, 2018b). In comparison, the interference induced by the categorically related items is more robust when the distractor is presented around the same time as the target, i.e., SOA of -200 ms to 150 ms (Alario et al., 2000; La Heij, Dirx, & Kramer, 1990; Lupker, 1979). Finally, the effect of semantic distance between the target and the categorically related distractor has yielded opposite effects: Mahon, Costa, Peterson, Vargas, and Caramazza (2007) reported shorter RTs for semantically close vs. distant distractors, whereas Rose, Aristei, Melinger, and Abdel Rahman (2019) found the exact opposite pattern. In summary, the results of the PWI studies suggest that the co-activation of a lexical distractor (compared to no distractor or a non-lexical distractor) interferes with production; however, the nature of the relationship between target and the co-activated representation and the temporal gap between the two sometimes exacerbates, sometimes ameliorates such interference.

The paradigms we refer to as priming paradigms here are those in which production of a target is preceded by the production of a related compared to an unrelated word (see Table 1).

Unlike PWI, priming paradigms do not require suppression of a perceived word as an irrelevant item. Opposing effects of semantically related primes are best demonstrated in priming paradigms. While naming a picture shortly ( $< 2$  s) before another picture often facilitates naming (Biggs & Marmurek, 1990; Huttenlocher & Kubicek, 1983; Lupker, 1988), producing a semantically related word preceding the target by longer intervals ( $> 4$  s) generally leads to interference (Vitkovitch, Rutter, & Read, 2001; Wheeldon & Monsell, 1994). The possibility that facilitation may be primarily due to strategic expectations developed in participants about the nature of targets following semantically related primes has been recently ruled out by Wei and Schnur, (2019). These authors showed robust facilitation even with low ( $< 15\%$ ) probability that a target would be semantically related to the prime. Furthermore, the magnitude of facilitation was not modulated by either conscious awareness of semantic relations between targets and primes, or in later stages of the task where participants are expected to have had more time to realize that such relationships exist between primes and targets. These results point to the production-internal mechanisms not only for interference, as has been claimed before, but also for facilitation induced by the co-activation of semantically related items. Continuous naming paradigms are extended versions of priming paradigms, in which pictures from the same semantic category (e.g., animals) are presented, interleaved with unrelated pictures, without repetition. Since the gap between items is typically longer than a few seconds, they produce the typical interference effect observed in priming paradigms with long lag, discussed above. More specifically, RTs increase linearly as a function of the number of previously named semantically related pictures (Costa, Strijkers, Martin, & Thierry, 2009; Howard et al., 2006; Schnur, 2014).

In blocked cyclic naming paradigms, unlike the paradigms reviewed above, a small set of pictures are presented repeatedly in cycles (e.g., Belke et al., 2005; Schnur et al., 2009; see Table

1). Variations of this paradigm include repeated presentation of two (e.g., Nozari, Freund, Breining, Rapp, & Gordon, 2016) or more pictures without first cycling through all the pictures (e.g., O'Séaghdha & Frazer, 2014). Items in a block are either unrelated (e.g., cat, pen, bowl), categorically related (e.g., cat, dog, goat), associatively related (gladiator, chariot, shield), related in onset segments (cat, cup, cone), or related in non-onset segments (man, fan, can). The baseline for comparison is always the unrelated (often referred to as the heterogeneous) condition. In keeping with the PWI results, the effect of categorically related items on naming suggests interference (Belke et al., 2005; Nozari et al., 2016; Schnur, Schwartz, Brecher, & Hodgson, 2006; Schnur et al., 2009) in cycles after the first one. A transient facilitation for related items is often found on the first cycle (e.g., Schnur et al., 2009; Python, Fargier, & Laganaro, 2018a) which seems to have a strategic component (Belke, Shao, & Meyer, 2017). Also in line with the PWI results, the effect of onset-overlapping items on RTs is facilitatory, and is also subject to strategic and attentional influences (e.g., O'Séaghdha & Frazer, 2014). Interestingly, overlap in non-onset segments seems to induce interference, rather than facilitation (Breining, Nozari, & Rapp, 2018; Breining, Nozari, & Rapp, 2016; Nozari et al., 2016).

The effect of non-categorically-related items has been inconsistent. Abdel Rahman and Melinger (2007) found interference for non-categorically-related items (e.g., bee, honey, bee-keeper). In another study, they showed that ad-hoc categories, e.g., items related to a fishing trip, also induced interference as long as the theme had been established by presenting participants with an informative title, e.g., "fishing trip" (Abdel Rahman & Melinger, 2011; Aristei, Melinger, & Abdel Rahman, 2011). Contrary to these results, de Zubicaray, Johnson, Howard, and McMahon (2014) only found a transient facilitatory effect of thematically related over unrelated items on the first cycle, with no significant differences between the two conditions on

further cycles. The authors attributed the discrepancy between their findings and the findings of Abdel Rahman and colleagues to methodological problems in the latter studies, which they claim confounded thematic and categorical relationships. At this stage, more studies are needed to settle that debate.

In short, in the majority of cyclic naming studies, co-activation of two items leads to interference that is more pronounced in the related compared to the unrelated condition. Exceptions include facilitation in naming pictures with overlapping onsets, facilitation sometimes observed in the first cycle of the presentation of semantically related pictures, and possibly non-categorical semantic relationships. Of these, the third effect is far from robust. The first effect is almost certainly strategic, as onset facilitation disappears if the words with shared onsets are interleaved with words that have overlap with them in other segments (Breining et al., 2016). The origin of the second effect is less clear. In a meta-analysis of past studies, Belke (2017) speculated that semantic facilitation on the first cycle of the blocked cyclic naming task was observed when related cycles were grouped together, but not when they alternated with the unrelated cycles. But Shao, Roelofs, Martin, and Meyer (2015) found interference in the first cycle of a task with the alternating design, calling this interpretation into question. Replicating the findings of Navarrete, Del Prato, Peressotti, and Mahon (2014), Belke et al. (2017) found facilitation when sets were separated by pauses, but not when such pauses were removed. They took the sensitivity of the facilitation effect to imply a strategic origin, although it is not clear whether alternative explanations can be convincingly ruled out. Finally, Belke (2008) found that facilitation on the first cycle turned into interference if speakers were concurrently engaged in a working memory task which may leave less resources for them to use for strategic behavior.

Table 1. A summary of the behavioral results for the co-activation of two words during word production. See the text for details.

Paradigm	Procedure	Example studies	Finding
Picture-word Interference (PWI)	A picture is named while ignoring a (spoken or heard) distractor word.		
	- Lexical vs non-lexical distractors	Hirschfeld et al., 2008	Interference
	- Onset-related vs. unrelated distractors	Schriefers et al., 1990	Facilitation
	- Categorically-related vs. unrelated distractors	Schriefers et al., 1990	Interference
	- Semantically-related (non-categorical) vs. unrelated distractors	Alario et al., 2000; Costa et al., 2005	Facilitation
	- Close vs. far semantic neighbors	Mahon et al., 2007; Rose et al., 2019	Inconsistent
Priming paradigm	A picture is named after producing a related word (prime)		
	Short gap (< 2s) between prime and target	Biggs & Marmurek, 1990; Huttenlocher & Kubicek, 1983; Lupker, 1988	Facilitation
	Long gap (> 4s) between prime and target	Vitkovitch et al., 2001; Wheeldon & Monsell, 1994	Interference
Continuous naming paradigm	Multiple pictures from the same category are named without repetition, usually interleaved with pictures from other categories.	Costa et al., 2009; Howard et al., 2006; Schnur, 2014	Interference
Blocked cyclic naming paradigm	A small set of pictures that are in some related are named in cycles, with the same pictures repeating across cycles, usually (but not always) not interleaved with unrelated pictures.		
	- Onset-related vs. unrelated distractors	O'Séaghdha & Frazer, 2014)	Facilitation
	- Rhyme/coda-related vs. unrelated	Breining et al., 2016; Nozari et al., 2016	Interference
	- Categorically-related vs. unrelated distractors	Belke et al., 2005; Schnur et al., 2006; 2009	Interference
	- Semantically-related (non-categorical) vs. unrelated distractors	Abdel Rahman & Melinger, 2007, 2011; de Zubicaray et al., 2014)	Inconsistent
Simple picture naming	The influence of the properties of pictures that are hypothesized to contribute to the co-activation of other words is studied.		
	Phonological neighborhood density	Sadat et al., 2014	Interference
	Name agreement	Alario et al., 2000; Kan & Thompson-Schill, 2004	Interference

Finally, the consequences of co-activation of multiple items can be probed by the inherent properties of words. Two examples are name agreement and phonological neighborhood density. When controlled for other indices of lexical retrieval, pictures with low name agreement (e.g., couch, sofa) tend to have longer RTs compared to those with high name agreement (Alario et al., 2004; Kan & Thompson-Schill, 2004), pointing to a potential interference effect resulting from the co-activation of multiple items. Also, according to models of language processing such as the one depicted in Fig. 1, words with lots of phonological neighbors tend to co-activate more words during production than those with fewer neighbors. While the effect of phonological neighborhood density has long been debated, the most carefully controlled study together with a re-analysis of multiple prior studies has shown the net effect of phonological neighborhood density in neurotypical speakers to be interference (Sadat et al., 2014).

The behavioral consequences of co-activation of items in picture naming can be summarized as follows: a) Co-activation of other items has a definite effect on the production of the target item. b) While paradigm-specific effects undoubtedly exist, common trends can be seen across paradigms. c) Co-activation of semantically or phonologically related items induces both facilitation and interference effects (see also Abdel Rahman & Melinger, 2009; Belke et al., 2017). But there is disagreement about the origin of these opposing effects and how much they reflect underlying dynamics of the production system, as opposed to strategic effects. While facilitation has been viewed as a critical (Abdel Rahman & Melinger, 2009) or even the only real effect of co-activation of semantically related representations during the semantic-lexical activation phase (Mahon et al., 2007; Navarrete et al., 2014), others have viewed it as a strategic effect that is not consistently found (Belke, 2017; Belke et al., 2017). Collectively, the evidence suggests that while the facilitation often observed in the first cycle of the blocked cyclic naming

task may be strategic in origin, facilitation observed in other cases is unlikely to be strategic (e.g., Nozari, 2019; Wei & Schnur, 2019), and is instead a reflection of the core dynamics of spreading activation in the system (see also Abdel Rahman & Melinger, 2009). We return to this critical point when discussing the theoretical perspectives on the neural data.

## **Neural signatures of the co-activation of multiple representations**

### *Neuroimaging studies*

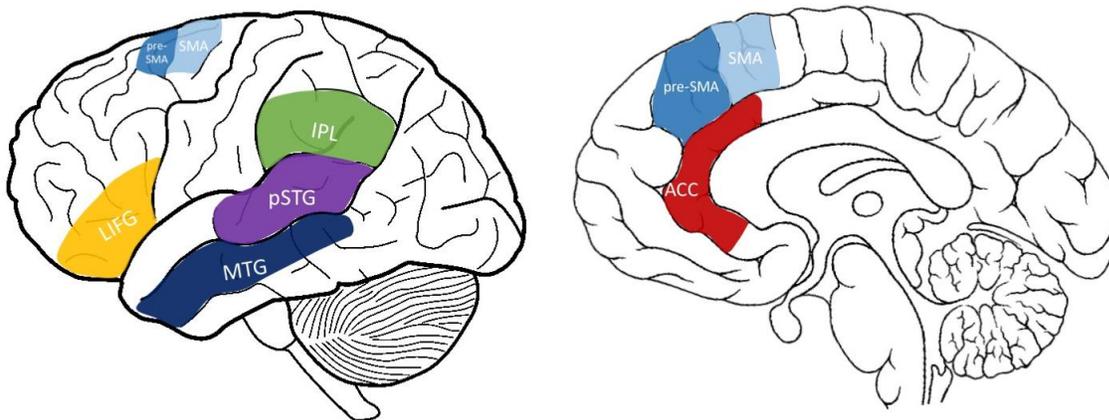
The main goal of neuroimaging studies is to understand the link between brain and behavior. A specific advantage of this technique is to identify brain regions that are involved in domain-specific operations (e.g., the occipital cortex is involved in visual processing, but not in word production), vs. those that are involved across a variety of tasks in different domains (e.g., the inferior prefrontal cortex is consistently implicated in situations where response selection is difficult, and this is not specific to a certain task or domain). The review below shows that both types of regions have been implicated in studies of co-activation of multiple representations in word production.

In one of the earliest neuroimaging studies of PWI, de Zubicaray, Wilson, McMahon, and Muthiah (2001) contrasted picture naming in the presence of semantic vs. neutral distractors (a row of Xs matched to the length of the semantic distractors). They found the following regions to show increased activation in the semantic compared to the neutral condition: the left middle temporal gyrus (MTG; BA21), left posterior superior temporal gyrus (pSTG; BA13), left anterior cingulate cortex (ACC; BA32), bilateral orbitomedial prefrontal cortex comprising right medial and left superior frontal gyri (BA 10), and left inferior parietal lobule (IPL: BA40). In another study, de Zubicaray, McMahon, Eastburn, and Pringle (2006) presented participants with

semantically related primes named from a definition (with intervening items between the definition and the target pictures), and showed that naming times were slowed down for related compared to unrelated distractors. The condition that caused greater behavioral interference was accompanied by increased activity in the left pMTG, ACC, and left inferior frontal gyrus (IFG), as well as left inferior prefrontal cortex (pars triangularis; BA 46), supplementary motor area (SMA; BA 8/32), and superior frontal gyrus (frontopolar cortex; BA10). Since then several other fMRI studies have investigated the neural correlates of co-activation of representations during production. A subset of the regions discussed above is consistently implicated in neuroimaging studies that contrast the co-activation of semantically related to unrelated items (de Zubicaray et al., 2013; de Zubicaray, McMahon, & Howard, 2015; de Zubicaray & McMahon, 2009; Heim, Friederici, Schiller, Rüschemeyer, & Amunts, 2009; Mechelli, Josephs, Ralph, McClelland, & Price, 2007; see Fig. 2). The most prominent of these are MTG, ACC, and LIFG.

*MTG.* Although changes to the BOLD signal in MTG are often reported in these studies, the direction of the change does not consistently predict the behavioral outcome. For example, in de Zubicaray et al. (2001, 2006; discussed above) as well as de Zubicaray et al. (2015), behavioral interference has been accompanied by *increased* activity in MTG. In fact, in a continuous naming paradigm, de Zubicaray et al. (2015) found a linear relationship between increased MTG activity and position of the picture in the sequence, which is in turn accompanied by cumulative semantic interference. In contrast, in a PWI study using auditory distractors that were categorically related to the targets (0 SOA), de Zubicaray and McMahon (2009) found behavioral interference accompanied by *decreased* BOLD signal in left middle and posterior parts of MTG. In keeping with this finding, another study by the same group, employing blocked cyclic naming, also found behavioral interference together with decreased MTG activity for

categorically related compared to the unrelated blocks (de Zubicaray et al., 2014). The opposite brain-behavior relationship was reported by Abel et al. (2009), who also used auditory distractors in PWI (albeit with a -200 ms SOA) and reported *increased* BOLD signal in the left MTG and left angular gyrus which accompanied behavioral *facilitation* for associatively related distractors. (That study failed to find a robust effect of categorically related distractors, or any neural differences pertaining to that comparison). The dissociation between the direction of changes to the BOLD signal and behavior is most obvious in de Zubicaray et al. (2013). In a PWI paradigm with written distractors, they found the expected behavioral interference induced by categorically related words (cow/rat), which was accompanied by a decrease in the BOLD signal in the left posterior MTG. The same study, however, reported behavioral facilitation for thematically related items (cow/pasture), which was also accompanied by a decrease in the BOLD signal in the left middle MTG.



*Fig. 2. Schematic of the neural regions commonly implicated in neuroimaging studies of co-activation of words during picture naming. ACC: anterior cingulate cortex; IPL: inferior parietal lobule; LIFG: left inferior frontal gyrus; MTG: middle temporal gyrus; pSTG: posterior superior temporal gyrus; pre-SMA: presupplementary motor area; SMA: supplementary motor area.*

ACC. In picture naming tasks with primes or distractors, ACC activation seems to be found when the target and competitor have close temporal overlap (e.g., de Zubicaray, McMahon, Eastburn, & Wilson, 2002; de Zubicaray et al., 2001; Heim et al., 2009) and not when the distractor is presented either too early (e.g., Abel et al., 2009) or too late (Spalek & Thompson-Schill, 2008). Generally speaking, ACC activation is observed under three circumstances (Barch, Braver, Sabb, & Noll, 2000): a) errors, b) when a prepotent response needs to be overcome, as in Stroop and go-nogo tasks, and c) when multiple underdetermined responses compete for selection, as in verb generation tasks when the noun is associated with multiple responses (e.g., ball → play, kick, throw, etc.), as opposed to a single dominant response (e.g., bell → ring). One unifying theme between these three conditions is that ACC is activated under situations of high conflict, i.e., when two or more representations compete for selection as a response, most likely in order to signal that the situation has a high chance of leading to an error (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Ullsperger, Fischer, Nigbur, & Endrass, 2014; Yeung, Botvinick, & Cohen, 2004). Compatible with this hypothesized monitoring role, ACC has been consistently found to show increased activation for representations that are harder to tease apart (de Zubicaray et al., 2001; Heim et al., 2009).

LIFG. A third region often implicated under conditions of co-activation of multiple representations in production is LIFG. In the broader literature, LIFG's activity is hypothesized to reflect the implementation of top-down control by biasing competition towards one (usually the more goal-compatible) of the two or more representations competing for selection (Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997; Thompson-Schill et al., 1998). In other words, while the ACC is hypothesized to play a role in the *detection of conflict*, LIFG is considered to be involved in the *resolution of conflict*. We have previously summarized a large

body of evidence pointing to a similar function for the LIFG in language comprehension and production (Nozari & Thompson-Schill, 2016). We thus confine this section to a discussion of some of the evidence most relevant to the current debate: Thompson-Schill et al. (1997) found greater LIFG activation during a verb generation task in response to nouns associated with multiple as opposed to a single dominant verb (see above under ACC). In the same vein, Kan and Thompson-Schill (2004) found increased LIFG activity for naming pictures with low vs. high name agreement, a finding that accompanied behavioral interference as described earlier. De Zubicaray and McMahon (2009) further found selective activation of the pars orbitalis of LIFG in semantically related compared to unrelated conditions. Increased LIFG activation has also been reported when semantically related items cause behavioral interference in blocked cyclic naming (Schnur et al., 2009) and continuous naming (de Zubicaray et al., 2015). While de Zubicaray et al. (2015) found greater activity of LIFG for later compared to earlier positions, a linear relationship between such activity and position was missing. A more recent study, however, found a linear increase in the BOLD response of LIFG as a function of the number of previously named semantically related pictures (Canini et al., 2016).

*Other regions.* In addition to the three regions above, a few other regions have been implicated in studies investigating the co-activation of multiple representations during production with some regularity. The most prominent of these are pSTG, left IPL and SMA and its most anterior portion, referred to as the pre-SMA. A decrease in the activation of the pSTG has been found to accompany the facilitation observed in the presence of phonologically related distractors in PWI (de Zubicaray et al., 2002; see also Bles & Jansma, 2008, who used picture–picture stimuli). Another region is the left IPL, especially the angular gyrus. It is undisputed that this region plays a role in language production. However, there is disagreement on whether its role is more

relevant to the processing of specific types of semantic information such as thematic relationships or information carried by verbs (Abel et al., 2009; Boylan, Trueswell, & Thompson-Schill, 2015; de Zubizaray et al., 2013; Schwartz et al., 2011) or to conflict resolution similar to LIFG (Noonan, Jefferies, Visser, & Lambon Ralph, 2013). Finally, left SMA/preSMA have been reported to show greater activation when co-activation of items led to longer RTs (de Zubizaray, Miozzo, Johnson, Schiller, & McMahan, 2012; de Zubizaray et al., 2001), but de Zubizaray et al. (2013) found decreased activity in left SMA for categorically related compared to unrelated items. Evidence suggests that pre-SMA and SMA represent higher-order information for planning movement sequences, such as the specific position of elements in a sequence and the intervals between the elements (e.g., Shima & Tanji, 2000), and as such are critical for the initiation of speech motor commands (e.g., Alario, Chainay, Lehericy, & Cohen, 2006). Empirical evidence on the differences between pre-SMA and SMA suggests a more abstract (cognitive) role for the pre-SMA as opposed to a more motoric role for the SMA. In implemented models of motor control of speech, such as GODIVA (Bohland, Bullock, & Guenther, 2010), pre-SMA's job is to encode abstract frames, while SMA's job is to initiate the planned speech acts. However, pre-SMA has also long been implicated in performance monitoring (see Ullsperger et al., 2014 for a review). This view has recently gained support from a study using intracranial recordings from the pre-SMA neurons in human subjects which found sensitivity of the pre-SMA neurons to errors even earlier than the ACC (Fu et al., 2019).

To summarize, neuroimaging studies implicate several regions as relevant to situations where multiple items are co-activated during word production. Three of the more commonly reported regions are MTG, ACC and LIFG. Of these, ACC and LIFG have more consistently shown increased activation during production of related (compared to unrelated) items when

these relations have led to behavioral interference. Moreover, both have been implicated in general detection and resolution of conflict across multiple domains (see Nozari & Thompson-Schill, 2015 for a review). MTG results are mixed, but a non-negligible number of studies have found decreased activity of this region under conditions of co-activation of multiple items, whether the final outcome was facilitation or interference. Unlike the ACC and LIFG, this region is more closely linked to word-specific operations. The function of some of the other regions such as IPL and pre-SMA are less well understood in the broader literature, but have ties to the monitoring-control loop, similar to the ACC and LIFG.

### *Electrophysiological studies*

Unlike neuroimaging studies, the main advantage of electrophysiological studies is their high temporal resolution, which allows them to pick up changes in neural activity with millisecond precision. Thus, while fMRI studies are primarily concerned with the regions involved in processing multiple items during production, EEG studies are interested in the timeline of such effects. In this section, we attempt to summarize the findings of these studies. One of the earliest effects (within the first 100 ms of processing) has been the significantly greater positivity induced by a non-linguistic distractor compared to a variety of linguistic distractors, all of which patterned more or less similarly in this very early window (Hirschfeld et al., 2008). In the same study, a little later in the 120–220 ms window, differences between other distractor types emerged, with distractors related in surface features (e.g., “fur” for the target “dog”) inducing the lowest negative amplitude followed by semantic category members and unrelated distractors (which did not differ from one another), and finally the non-linguistic distractors which now showed the highest negative amplitude. Between 250–400 ms, again the non-linguistic distractors showed the greatest positivity (as in the earliest window), with non-significant

differences between the other three distractor types. Dell'Acqua et al. (2010) also found a sustained difference in the 50–200 ms window with higher positive amplitudes for related compared to unrelated distractors in the PWI paradigm. These studies show that the system could be sensitive not only to the co-activation of items, but also to the type of relationship between the target and the competitor starting earlier, or around 200 ms after the stimulus presentation. We call these effects “early effects”. Following this early window, several studies have reported effects within what is usually considered the N400 window in the comprehension literature. Finally, reports also exist of very late stimulus-locked effects, past the N400 window and around the time the verbal response is made. Not surprisingly, many studies have reported effects in multiple windows. These stimulus-locked effects are summarized in Table 2.

Findings in the early time window are difficult to compare, because some studies have identified and labeled components (e.g., N2, P2), while others have described effects over a time-window without specifically labeling components. The most striking effect in the early time window has been reported by Costa et al. (2009) in a continuous naming paradigm. These authors found a graded decrease to the N2 and a graded increase to the P3 amplitudes with ordinal position, an ERP effect which was correlated with the increase in the RTs (see Rose & Rahman, 2017, for a replication with close as opposed to distant semantic neighbors; but see Llorens, Trébuchon, Riès, Liégeois-Chauvel, & Alario, 2014, for an earlier failure to replicate). Using the blocked cyclic naming paradigm, Python et al. (2018a) reported a smaller P2 for the related compared to the unrelated blocks for cycles 2 and 3, where interference was observed. Studies that did not label their components have also found effects in the 200–450 ms window. Using blocked cyclic naming, Janssen, Carreiras, and Barber (2011) found an effect in this window, with related blocks producing smaller negative amplitudes than unrelated blocks

(Janssen, Hernández-Cabrera, van der Meij, & Barber, 2015; Wang, Shao, Chen, & Schiller, 2018). In a PWI paradigm comparing close and distant categorically related distractors to unrelated ones, Rose et al. (2019) found a lower amplitude associated with the closely related compared to unrelated distractors.

Compared to the early window, the findings in the N400 window have been reported more uniformly. PWI studies that compared semantically related to identity distractors have found more negative waveforms in the N400 time window for the semantically related distractors (Piai, Roelofs, & van der Meij, 2012; Shitova, Roelofs, Schriefers, Bastiaansen, & Schoffelen, 2017). It is important to note that the “identity” distractor does not activate a separate representation from that of the target. Thus, as far as the number of activated semantic-lexical representations is concerned, this comparison is closer to comparing picture naming with and without distractors. The results can thus be read as a more negative N400 in the presence of a second co-activated representation. When comparing the *nature* of the activated representations, findings consistently point to a lower-amplitude N400-like waveform for related compared to unrelated distractors, both when the effect of semantic distractors has been interference (Blackford, Holcomb, Grainger, & Kuperberg, 2012; Dell’Acqua et al., 2010; Rose et al., 2019; Wong et al., 2017) and when it has been facilitation due to the long negative SOA (Python et al., 2018b).

Table 2. A summary of stimulus-locked effects for co-activation of items in paradigms with picture naming.

	Study	Window/ component	Paradigm/ manipulation	Behavioral effect	EEG effect	Topography
<b>Early (starting before or around 200 ms post-stimulus)</b>						
	Costa et al., 2009	N2 [230–280], P3 [280–380]	continuous naming	semantic interference	Decrease in N2 amplitude and increase in P3 amplitude with serial position	parieto-occipital
	Llorens et al., 2014	N/A	continuous naming	cumulative semantic interference	Failure to replicate Costa et al. (2009) [No significant differences found]	N/A
	Janssen et al., 2011	[220–450]	blocked cycling naming	semantic interference	Less negative amplitudes for related items	fronto-central
	Python et al., 2018a	P2 [200–300]	blocked cycling naming	semantic interference in cycle 3	Lower P2 amplitude for related items in cycles 2 and 3	posterior
<b>N400 time window</b>						
	Blackford et al., 2012	N400 [325–600]	PWI with a backward mask	semantic interference, phonological facilitation	Decreased N400 for both identity and semantically related distractors; no effect of phonemic overlap (onset)	fronto-central
	Python et al., 2018b	[330–430]	PWI (SOA = -700)	semantic facilitation (both associative and categorical)	Less negative amplitude for both categorical and associative primes compared to unrelated	central
	Shitova et al., 2017	N400 [350–600]	PWI (SOA = 0)	semantic interference compared to identity distractor (no unrelated control)	More negative amplitude for semantic than identity distractors	central
	Piai et al., 2012	N400 [250–500]	PWI (SOA = 0)	semantic > unrelated > identity distractors	Same as Shitova et al. (2017); no difference between semantic and unrelated	central

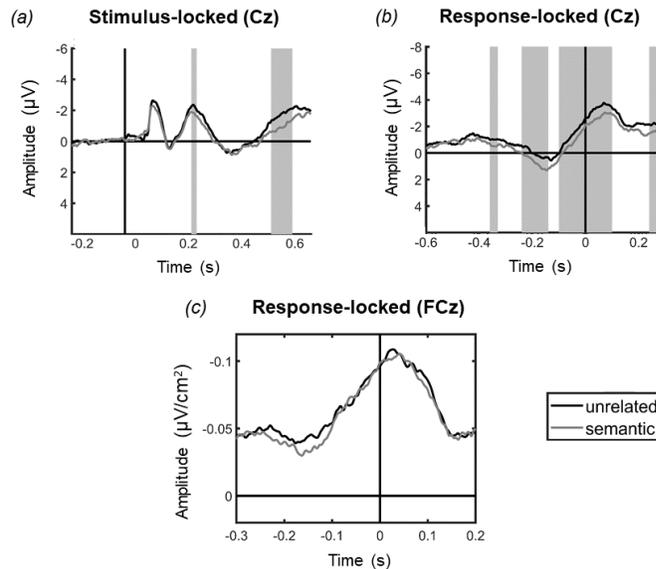
	Wong et al., 2017	[275-400]	PWI (SOA = 0)	semantic interference for categorical distractors, non-significant for associative distractors, marginal phonological facilitation	less negative amplitude for related than unrelated for both associative and categorical, no effect of phonological distractor	only electrode clusters shown
<b>Early + later effects</b>						
	Hirschfeld et al., 2008	[0-100], [120-220], [250-450]	PWI (SOA = -150)	semantic interference for categorical distractors, facilitation for nonlinguistic and feature distractor	[0-100]: lower positivity for all linguistic than nonlinguistic distractors. [120-220]: negativity amplitude in feature < category = unrelated < nonlinguistic [250-450]: lower positivity for all linguistic than nonlinguistic distractors	only Cz shown
	Dell'Acqua et al., 2010	[50-200], [200-450]	PWI (SOA = 0)	semantic interference and phonological facilitation	[50-200ms]: more positive amplitude for related items in semantic; RT effect correlated with ERP amplitude. [250-450ms]: less negative amplitude for related in both semantic and phonological conditions	central
	Rose et al., 2017	[100-150], [250-400], [500-600]	Continuous naming	Semantic interference for closely related words	100-150: increased P1 from 1 <sup>st</sup> to 2 <sup>nd</sup> serial position 250-400: higher positive amplitude with increasing serial position 500-600: lower positive amplitude with increasing serial position	Posterior (similar electrodes to Costa et al., 2009)
	Wang et al., 2018	[200-550], [350-500]	blocked cyclic naming	semantic interference and phonological facilitation	Lower amplitude for both semantic and phonological compared to unrelated blocks in both windows	only electrode clusters shown
	Rose et al., 2019	[228-292], [350-474]	PWI (SOA = 0)	semantic interference: closely related > distantly related > unrelated	No differences between distant and unrelated. Lower amplitude for close compared to either unrelated or distant in multiple time windows	posterior, central
	Janssen et al., 2015	[250-400], [500-750]	blocked cyclic naming	Facilitation for related compared to unrelated in cycle 1, interference for related in cycles 2-4	250-400: related less negative than unrelated. 500-750: related less positive than unrelated (cross over effect)	central, frontal

Finally, a few studies have reported delayed effects around the time of articulation. One prominent example is Janssen et al. (2015), who reported less positivity for related compared to the unrelated blocks in cycles 2, 3, and 4 of a blocked cyclic naming task (the same cycles which showed behavioral interference) in the 500-750 ms time window, primarily over anterior electrodes. In the continuous naming task, Rose and Abdel Rahman (2017) also found lower positive amplitudes with increasing serial position from 500-600 ms post-stimulus. The very late effects around the articulation time are usually investigated with response-locked ERPs. For example, Python et al. (2018a) found that related conditions yielded lower amplitudes than unrelated conditions around -100 ms before the vocal onset. Another set of studies that have looked at response-locked ERPs have been particularly interested in the difference between error and correct trials. This difference is reflected in the error related negativity (ERN; Falkenstein, Hoormann, Christ, & Hohnsbein, 2000), a fronto-central component around the response initiation with greater negativity for error compared to correct trials (Ganushchak & Schiller, 2008b; Masaki, Tanaka, Takasawa, & Yamazaki, 2001; Stéphanie Riès, Janssen, Dufau, Alario, & Burle, 2011). ERN has also been found in other language production tasks that do not involve picture naming, such as tongue-twisters (Acheson & Hagoort, 2014), SLIP task (Möller, Jansma, Rodriguez-Fornells, & Münte, 2007), copy typing (Kalfaoğlu, Stafford, & Milne, 2018) and phoneme substitution (Trewartha & Phillips, 2013).

To our knowledge, the only ERP study comparing the effects of semantic and unrelated conditions on both stimulus- and response-locked ERPs is (Python et al., 2018a; see Riès et al., 2017, for time frequency analyses of ECoG data). This study, however, time-locked the late effects not to the response itself but to 100 ms before articulation and did not explore the differences at and after the articulation time. A recent study in our lab investigated both stimulus-

and response-locked consequences of producing words in the context of either a semantically related or an unrelated word. Participants repeatedly named one of the two items in a block, causing both items to be highly activated throughout the block. ERP comparisons were always on the same target items (e.g., pie) when they appeared in the context of a semantically related item (e.g., cake) or an unrelated item (e.g., hair). Fig. 3a, shows stimulus-locked responses. In keeping with Costa et al. (2009), semantic similarity led to a less negative N2. Also in keeping with Janssen et al. (2015), the related condition produced less negativity in the late stimulus-locked window. Widespread differences were also found in the response-locked analyses (Fig. 3b). The two conditions differed not only in a window around 200 ms before the response onset, but also during response articulation.

After the Laplacian transformation, an ERN-like component (in terms of time, shape, and topography) was identified on FCz (Fig. 1c). When scaled to be comparable to Riès et al. (2011), the amplitude of the current ERN (for correct trials with co-activation) was two times higher than the ERN for correct trials (without co-activation) reported by Riès et al. (2011) but still lower than that of the error trials in that study. Interestingly, however, the latency and magnitude of this component was similar between semantic and unrelated conditions. This is, to our knowledge, the first demonstration of the continued effects of the co-activation of representations during production from the very early time window all the way to the articulation phase, using both stimulus-locked and response-locked analyses. To summarize, EEG studies have provided unequivocal evidence for the system's sensitivity to the co-activation of non-target items from very early all the way to the late stages of word production and even post-articulation. This literature has also attempted to link specific processes to narrow time windows using a serial model of language production. Next section assesses the validity of this approach.



*Fig. 3. ERPs for producing the same word in the context of semantically related (grey) vs. unrelated (black) representations. Stimulus-locked (panel a) and response-locked (panel b) ERPs are shown on the Cz electrode. Grey shaded area corresponds to periods of statistical significance, tested with non-parametric point-by-point Wilcoxon signed-rank test with permutations, corrected for a minimum duration of 20ms. Panel c shows Laplacian-transformed ERPs on the FCz electrode. The shape, timing, and topography of this component are compatible with the ERN. Average response time is 637 ( $\pm 164$ ) ms.*

## Deriving theoretical perspectives from empirical evidence

We have reviewed a rich body of behavioral, neuroimaging, and electrophysiological data addressing the consequences of co-activation of multiple representations in word production. A few general conclusions are warranted by these data: the results obtained using all three methods converge on the fact that presence of a non-target item changes the dynamic of processing in the word production system. Examination of the neural regions shows that both domain-specific and domain-general regions are involved in this altered dynamic. Moreover, inspection of the results of the electrophysiological studies shows that the change is not limited to a single point in the process and instead affects multiple windows of processing. In this section, we discuss some of the interpretations attached to the data and the problems associated with such interpretations.

A common approach to solving this problem has been to pinpoint an event to an underlying process by comparing its emergence in the EEG data against an assumed timeline of events in word production. This assumed timeline is often borrowed from Indefrey and Levelt (2004) or its upgraded form (Indefrey, 2011). According to this timeline, conceptual encoding is completed by 200 ms post-stimulus-onset, with lexical selection taking place between 200 and 275 ms post-onset, followed by phonological encoding and later articulatory buffering at around 455 ms after the picture onset. ERP differences obtained in other experiments are then compared to this timeline and labeled accordingly.

For instance, based on the appearance of ERP differences starting after 300 ms post-stimulus-onset, Python et al. (2018b) concluded that the facilitatory effect of semantically related primes must be post-lexical, because lexical selection must be over by 275 ms (see also Python et al., 2018a). In contrast, studies that have reported semantically induced interference have taken relatively early effects around 200 ms post-picture-onset to indicate lexical competition (e.g., Costa et al., 2009; Rose et al., 2019; Wang et al., 2018). Together, these interpretations counterintuitively put lexical competition before semantic facilitation, which is often attributed to the dynamics of spreading activation preceding lexical selection (e.g., Abdel Rahman & Melinger, 2009; Nozari, 2019). In contrast, some have claimed that competition arises much later (after 500 ms post-picture-onset) in word production (Janssen et al., 2015), which, in turn, puts competition resolution way past phonological encoding and even articulation. One could blame the discrepancies in these claims on insufficient or discrepant data, but as reviewed earlier, there is plenty of evidence to suggest that co-activation of representations during word production induces effects at various stages from early to late (even post-articulation) phases. The problem,

then, must be in the interpretation of what these effects mean. Below, we lay out three problems in the approach discussed above.

1) Indefrey and Levelt (2004) is inarguably a tour de force, both in terms of being one of the first attempts to clearly link neural processes underlying word production to a well-formed cognitive model, and in terms of the number of studies reviewed. However, it must be remembered that the assumptions are made within the framework of Levelt et al.'s (1999) model, which is a serial model of word production with no cascading and no interactivity. Almost two decades later, there is now plenty of evidence that production processes include both cascading and interactivity (Baese-Berk & Goldrick, 2000; see Dell, Nozari, & Oppenheim, 2014, for a review of the cognitive evidence; see also Peramunage, Blumstein, Myers, Goldrick, & Baese-Berk, 2011, for neural evidence), at least in English (see Zhu, Damian, & Zhang, 2015, for arguments that some of these processes might be different in Chinese).

The modularity imposed by the Levelt et al.'s theoretical framework has an unequivocal bearing on the interpretation of the electrophysiological effects. For example, it must be assumed that phonological effects cannot possibly be observed at the same time as lexical effects, because phonology is only activated after lexical selection has been completed. Miozzo, Pulvermüller, and Hauk (2015) directly tested this assumption in a MEG study by investigating the effects of visual complexity, two indices of semantic retrieval, and word length as a proxy for phonological encoding. The findings revealed early activation of the visual cortex (~100 ms post-stimulus) followed by the perisylvian regions at around 150 ms, which showed sensitivity to both semantic and phonological variables. These results are incompatible with strict seriality of the production process, at least with the timeline proposed by Indefrey and Levelt (2004; see also Munding, Dubarry, & Alario, 2016 for a similar conclusion in reviewing the MEG data). More generally,

they call into question a serial timeline within which certain effects can be unambiguously attributed to certain processes without contamination from other processing levels.

2) Strict seriality aside, the model on which Indefrey and Levelt's (2004) reports are based assumes lexical selection by competition, thus findings of ERP differences that arise during the co-activation of semantically similar items are assumed to reflect competition at the lexical level. For example, in Maess, Friederici, Damian, Meyer, and Levelt (2002), the interference effect induced by the semantically related compared to the unrelated condition in a blocked naming task was accompanied by electrophysiological differences between the conditions in a 150-225 ms post-stimulus window. Assuming that this difference must reveal competition during selection, these findings are taken as reflecting the timing of lexical selection.

While lexical selection by competition is a reasonable assumption, it is far from undisputed (see Nozari & Hepner, 2018; 2019, and the commentaries). One topic of debate is whether interference or facilitation is the "real" behavioral effect, mostly in specific paradigms like blocked cyclic naming (see Navarrete et al., 2014, Belke et al., 2017 for examples). The main approach here has been to accumulate evidence for overt behavioral interference or facilitation by tweaking these paradigms. As far as this kind of evidence goes, our reading of the literature is that the empirical evidence for a net effect of interference in many cases is robust (see the various studies by Abdel Rahman and colleagues reviewed in earlier sections), and cannot be written off as spurious, as has been suggested by the proponents of non-competitive selection (e.g., Navarrete et al., 2014). The question now is whether robust interference necessarily means competitive selection. This brings us to the other main topic of debate, which is the link between the data and the underlying mechanism. To date, with the exception of PWI, the rest of the interference effects (either with repetition, as in blocked cyclic naming tasks, or

without repetition, as in continuous naming tasks) have been successfully modeled with non-competitive selection criteria (Oppenheim, 2018; Oppenheim, Dell, & Schwartz, 2010).

Therefore, the only advantage that competitive accounts of lexical selection currently hold over non-competitive accounts is their explanation of the findings derived from the PWI paradigm.

The weight given to this advantage depends on whether one views PWI as representing the true dynamics of word production or not, a topic that is hotly debated (see Oppenheim & Balatsou, 2019 vs. Roelofs, 2018).

A third angle relevant to the debate of competitive vs. non-competitive lexical selection is the use of the ERP data themselves to test the models. For example, Janssen et al. (2015) took the late timeline of their ERP effects to be evidence against incremental learning mechanisms proposed by the models of non-competitive selection, and in favor of conflict resolution accounts. But this late effect might very well reflect the top-down effect of the frontally mediated “booster” assumed in incremental learning models (Oppenheim et al., 2010). Similarly, Wang et al. (2018) claimed that their findings were evidence against the incremental learning model because they took the divergence of related vs. unrelated conditions at around 200 ms to represent lexical selection, and incremental learning was supposed to have its effect before lexical selection, i.e., during the mapping of semantic features to lexical items. But this claim is again based on an assumption derived from a serial model with competitive selection, the problems with which have been laid out in this and the previous point.

In short, while the interference effect at the behavioral level has been convincingly demonstrated, the argument that this interference can only be explained by competitive lexical selection is weak. Our goal is not to imply that competitive lexical selection has, by any means, been ruled out, but to emphasize that if, in truth, lexical selection is not competitive, evaluating

empirical data against a model that makes a strong assumption of competitive selection is bound to be misleading.

3) More generally, it is important to remember that the timelines that are used as assumptions have often been derived from the same paradigms that are now being tested against them, creating circular arguments. For example, using a blocked cyclic naming paradigm, the timeline of lexical selection was determined to be around 150-225 ms (Maess et al., 2002). This assumption is then used to interpret the findings in other blocked cyclic naming paradigms: when an effect of semantic similarity in the first cycle of the task appeared later than this window, it was attributed to phonological or monitoring processes (Python et al., 2018a). Ideally, assumptions should be derived independently of the specific paradigm they are tested on. The problem here is to find the right paradigm for developing solid assumptions that are generalizable to other paradigms. The PWI paradigm is a poor choice for this purpose, as it is a fairly complex task with multiple cognitive processes (reading or auditory comprehension) which are irrelevant to word production in everyday speech, and are also absent in other paradigms. Blocked cyclic naming paradigms avoid this problem, but they are not immune to strategic processes (Belke, 2008) which may alter the pure internal dynamics of word production. Studying the effects of different item properties (e.g., lexical frequency, length, phonological neighborhood density, etc.) in simple picture naming is a reasonable approach, as it is not contaminated by additional processes or wildly different strategies across participants, but it has the downside of comparing different items. Still, clear and replicable mapping of the timeline and duration of the effect of psycholinguistic variables during picture naming is a good step in developing more reliable assumptions about the loci of effects in more complex naming paradigms. As of now, such studies have been sparse and their findings have not been consistent

enough to recommend a clear timeline (Miozzo et al., 2015; Riès, Janssen, Burle, & Alario, 2013; Strijkers, Costa, & Thierry, 2010; Valente, Bürki, & Laganaro, 2014). Such a timeline could then be complemented with other naming tasks that compare the effect of semantically vs. phonologically related representations on the production of the same item in relatively simple naming paradigms, e.g., repeated naming of only two pictures with different relationships, although repetition priming and strategic effects must be taken into account for the latter.

### **A way forward**

In previous sections, we have reviewed the empirical evidence and touched upon some of the problems in the interpretations assigned to such evidence. The goal of this section is to propose a potential way for moving forward.

#### *The explanatory power of the dynamics of spreading activation*

As alluded to in several places above, the early divergences of ERP signatures for the related vs. unrelated conditions have often been taken to indicate lexical *selection*. Often overlooked, however, are the underlying dynamics of *spreading activation* as a potential reason for these early differences. Even though the importance of spreading activation has been highlighted in prior accounts, e.g., Abdel Rahman and Melinger's (2009) swinging network, such dynamics are often pushed aside in favor of the dynamics of lexical selection when explaining the neuroimaging and electrophysiological findings. The reasons are the dominance of net interference effects observed in the behavioral outcomes and the original proposal of Indefrey and Levelt (2004) that described the 200-275 ms post-stimulus window as the window for lexical selection. We have explained the problems associated with both of these assumptions.

The critical question here is, *Can* activation dynamics predict changes to the ERP and BOLD signals, and are those predictions compatible with the empirical data? Take, for example, a model with distributed lexical semantic representations, in which each representation consists of a number of features<sup>1</sup> (e.g., representations A, B, and C depicted in Fig. 4). Similarity is represented as overlap in features, as in A and C. Each feature is either turned on or off. Assume that a representation is activated when all its features are turned on. A simple way of representing the total activation in the system is by counting the number of the features that are on. All else being equal, this sum will be smaller in the case of co-activation of two related (A and C) compared to two unrelated (A and B) representations, because fewer features are activated in the former case. Therefore, simply assuming distributed lexical semantic representations across neural populations (an assumption that is biologically much more plausible than its localist counterparts) predicts differences in the related and unrelated conditions, arising merely due to the activation dynamics.

If this simple concept of the total activation were linked to the total activity of the neural population representing lexical items, one would expect a lower amplitude for the ERP component(s) which may represent such activity in the related compared to the unrelated conditions, and perhaps a decrease in the BOLD signal in areas related to semantic lexical processing (most likely middle MTG). Incidentally, both of these predictions find at least partial support in a number of empirical studies. For example, de Zubicaray and McMahon (2009) found decreased MTG activity in response to semantically related compared to unrelated conditions. Such a decrease was considered puzzling, so the authors speculated that it may be

---

<sup>1</sup> Note that by “features” we simply mean the units composing the representation of an item, and not physical features (see Introduction for details). We call them “features” in order to emphasize that, while they may not represent a real-life feature, they indeed capture the similarity between items.

related to the attenuation often observed in sensory regions for ignored stimuli in crossmodal tasks, or the modulation of processing based on how task-relevant the distracting stimuli are. While not impossible, the results may find a simpler explanation in the dynamics of activation.

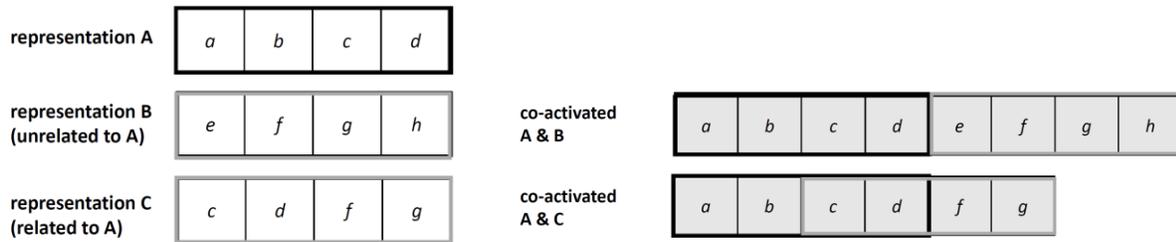


Fig. 4. Schematic of distributed semantic lexical representations for three items A, B, and C. A and B are unrelated, while A and C are related. Lower case letters denote “features”. Similar items A and C share two of their “features”. A feature is either “off” (shaded white), or “on” (light gray). For an item to be produced, all its features must be “on”. Total activation, counted as the number of “on” features, is lower for the related pair A and C, compared to the unrelated pair A and B.

Similarly, the component marked as N2 has been interpreted as marking lexical “selection” (e.g., Costa et al., 2009). This interpretation is close to the interpretation offered by the proponents of conflict-based ERN, who have proposed N2 as the stimulus-locked equivalent of the ERN in forced-choice tasks (Yeung et al., 2004). Both accounts maintain that N2 indexes the difficulty of separating two co-activated items, i.e., a process that is related to selection. If this interpretation were true, one would have expected a *larger* N2 in the related compared to the unrelated condition. Instead, in picture naming tasks it has consistently showed a *lower* amplitude for related items (Costa et al., 2009; see also Fig. 3a). Nevertheless, Costa et al. (2009) have interpreted the effect as indexing lexical competition, perhaps because the magnitude of gradual decrease in the P2-N2-P3 time-window correlated with the increase in the RTs. In later sections we discuss the brain-behavior relationship. For now, we simply open up the possibility

that the early EEG differences may be more easily explained by the dynamics of activation as opposed to selection.

The findings in the N400 window further encourage this interpretation. As explained in the earlier sections, numerous studies have reported a decrease in the amplitude of N400 in the related compared to the unrelated condition (Blackford et al., 2012; Python et al., 2018b, Wong et al., 2017). This decrease is even more meaningful than the early window, as the N400 is a relatively well understood component in the comprehension literature. Thus, if we accept, based on its timeline and topographical distribution, that the component recovered in this time window could be the counterpart of the N400 in comprehension (Blackford et al., 2012), then appealing to the dynamics of spreading activation for explaining the effects on the N400 gain further support from the comprehension literature (see below). It is important to note that we do not mean to advocate for this very simplified model as a realistic neural model. In fact, a careful balance of activation and inhibition derived from biologically plausible models may be necessary to capture the nuances in ERP data, as has been shown in a computational model of N400 in word reading (Cheyette & Plaut, 2017; Laszlo & Plaut, 2012). These authors showed that N400 in word reading can be elegantly modeled as the transient overactivation within distributed semantic representations by taking into account the balance of excitation and inhibition between and within cortical areas. The model not only produced the N400 in a stable time window, but also captured its sensitivity to a variety of factors in accordance with the empirical data, such as a higher amplitude for low frequency words as well as for semantically rich words, and a lower amplitude for repeated words (Cheyette & Plaut, 2017).

Interestingly, the model came on the heel of a popular model of N400 which attached a more complex evaluative interpretation to N400, namely that of implicit semantic prediction

error (Rabovsky & McRae, 2014). While Cheyette and Plaut's (2017) model was able to account for all the findings accounted for by Rabovsky and McRae's (2014) model, it had the advantage of actually generating an N400, and accounting for findings like the presence of the N400 in response to isolated words and nonwords that are harder to reconcile with a prediction error account. In other words, a model which simply linked N400 to the dynamics of bottom-up activation in the system (Kutas & Federmeier, 2011) proved not only to be sufficient, but perhaps even superior to a model which attributed N400 to more complex processes akin to integration (e.g., Brown & Hagoort, 1993). A parallel may very well exist in word production.

This suggestion may raise two questions: 1) What about early phonological effects, such as those reported by Miozzo et al. (2015), appearing as early as semantic effects? 2) What about the behavioral interference? The first question can be answered by accepting cascading in the system. In models without cascading, activation can only propagate to a later layer, e.g., phonemes, once competition has been resolved at the lexical layer and a single representation has been selected. Models with cascading, on the other hand, allow for activation in later layers before selection in a previous layer is complete. As such, multiple lexical representations would be able to simultaneously activate multiple phonological representations, creating concurrent signatures of lexical-semantic and phonological effects. The basic computational arguments for cascading have been discussed in Dell (1986) and summarized in Dell et al. (2014), so we will not repeat them all here. As independent evidence, Baese-Berk and Goldrick (2009) found significantly longer voice onset times (VOTs) to the same phoneme when the word had a minimal pair neighbor when the initial consonant was voiced (e.g. cod → god) compared to when it did not (e.g., cop → \*gop; see also Goldrick & Blumstein, 2007, for an extension of cascading to later processes such as articulatory phonetic encoding). This and similar reports

show that co-activation at the lexical level has downstream consequences up to articulation. There is thus no theoretical contradiction between attributing differences in ERPs to activation dynamics in the semantic-lexical space in the same window where phonological effects have been observed, as activation flow is likely to also reach later representations to some degree.

Answering the second question is more difficult, mainly because the relationship between EEG and behavior is not always straight forward. In comprehension, higher N400 amplitude is accompanied by longer RTs for low frequency words (N400 $\uparrow$ , RT $\uparrow$ ), longer RTs for non-repeated words (N400 $\uparrow$ , RT $\uparrow$ ), but shorter RTs for semantically rich words (N400 $\uparrow$ , RT $\downarrow$ ). Therefore, one cannot simply predict the change in behavior by looking at the change in the N400 amplitude. This, however, does not imply that the relationship is arbitrary. Cheyette and Plaut's (2017) model was able to predict such a dissociation because the changes to the activation in the semantic space concern not only the target, but also its neighbors (orthographic neighbors in their case, since the stimuli were written words). For example, low frequency words generate a larger N400 because they are less effective in suppressing their orthographic neighbors (which then activate their own semantic representations), but the activation of those neighbors does not facilitate the production of low frequency words. Semantically richer words, on the other hand, cause greater activation of their corresponding representations in the semantic space (hence the larger N400), and this greater activation directly benefits their access.

Very similar problems exist in production, and may be solved by similar solutions. For example, a counterintuitive effect in the PWI paradigm is that low-frequency distractors tend to interfere more with production than high-frequency distractors (Catling, Dent, Johnston, & Balding, 2010). According to Cheyette and Plaut's (2017) model, the reason is that low-frequency distractors are less successful in suppressing their neighbors (orthographic for written

and phonological for auditory distractors). The model thus predicts more activation in the semantic-lexical space because of the unsuccessfully suppressed neighbors. If we take the activity in MTG to index the activation dynamics, as argued above, then the prediction finds support in a neuroimaging study by de Zubicaray et al. (2012), who found increased activity in the left MTG for low-frequency distractors. This higher activation, however, is not due to the activation of the target response, and does not help performance.

These examples demonstrate the utility of an implemented computational framework for understanding the change to the neural signatures of co-activation of representations. Importantly, given similar implementations from the comprehension side, it is not implausible that ERP changes up to the N400 window, as well as the corresponding BOLD signal changes, reflect the dynamics of spreading activation in word production (see also Blackford et al., 2012). The question remains: Where does the interference come from? The next section addresses this issue.

*The critical question about selection: Where?*

Even if some of the neuroimaging and electrophysiological findings can be explained simply by the overlap during the activation of semantic lexical representations, any cognitive or neural theory of word production must ultimately explain why the co-activation of multiple representations during production often leads to interference. The simple answer is that, somewhere in the process, there is a bottleneck. The critical question is, *where?* The most obvious candidate is the very last stage of production, i.e., articulation. One word—and not two or three—can be articulated at any given time, so by the time the speaker gets to the articulation phase, it seems that selection must have occurred. But there are several stages before articulation, and selection may take place at any (or all) of those stages. Lexical selection has taken such a

monumental place in the word production literature because it is the first place in the production process where selection becomes meaningful: If the system successfully weeds out the non-target representations at this stage, production can move swiftly and cleanly to later stages and avoid errors on most occasions. The fact that native adult speech is in fact efficient and error-free in most cases has further supported an early selection mechanism at the lexical level. Moreover, some accounts deem lexical selection a good locus for the input from the syntactic frame (e.g., Gordon & Dell, 2003).

The importance of a selection mechanism is even more apparent in an interactive system with cascaded activation. Once semantic-lexical representations are activated, they send activation to their corresponding downstream representations (e.g., phonemes, articulatory phonetic features, motor plans, etc.), which in turn activate several potential responses in lower layers of the production system. The interactivity in the system activates additional representations through feedback, or intensifies the previously activated representations, further adding to the growing list of possible candidate responses. In such a system, it is vital to rein in the spread of activation relatively early in the system in order to give one response its best shot. This could be achieved in one of two ways: 1) Activated items impose some degree of inhibition on other representations. This so-called *lateral inhibition* emerges from the internal dynamics of the spreading activation within the production system, and naturally tunes production towards a winner. 2) An external system intervenes and biases the competition towards one of the activated representations. Lateral inhibition is a very reasonable assumption, given its prevalence in biological systems, and we assume it also exists in language production. In the next section on monitoring and control, we elaborate on the second possibility while discussing the monitoring

control loop. Immediately relevant to the current discussion, however, is the question, When is the production system allowed to proceed towards producing a response?

The answer to this question constitutes the critical debate between competitive and non-competitive theories of lexical selection. Competitive accounts of lexical selection maintain that a *relative threshold*, i.e., a certain difference between the activation of two lexical items, is necessary before production can proceed (e.g., Roelofs, 2018). In the case of related items, their shared features contribute to smaller differences in the activation levels of the two representations, so this distance is initially smaller. We have referred to this as “conflict” (Nozari, Dell, & Schwartz, 2011; Nozari & Hepner, 2018). When conflict is higher, resolving it takes longer, hence the interference observed in the related vs. unrelated conditions. The upside is that, since other representations have been mostly suppressed, there will be little chance for competition in later stages. The behavioral prediction is straightforward: relatively preserved accuracy in the presence of related co-activated items, but longer RTs.

In contrast, non-competitive accounts propose that production moves towards generating a response as soon as a lexical item passes an *absolute threshold* (e.g., Mahon et al., 2007). Here, since the related items boost the activation of one another, it is probable that one of them reaches the absolute threshold even earlier (hence the proposed facilitation effect). The downside is that others are not far behind (since a relative threshold has not been imposed). So they have also activated their downstream representations with a force that may be comparable to that of the target. The behavioral prediction is less straightforward than the competitive models: There is no slow down at the lexical selection level as a function of co-activation of related items (in fact, there is often a speed-up). The final RTs thus depend on how competition is resolved at later stages: If an equally liberal criterion is applied to phonological/phonetic/motor selection, then the

net effect should be facilitatory. But it is also possible that later stages of processing enforce stricter criteria, in which case one could still observe behavioral interference. Importantly, the locus of such interference would not be at the stage of lexical selection.

A third possibility is a flexible threshold, called the *selection criterion*, determined by the task goals and the ability of the production system to meet them (Nozari & Hepner, 2018; 2019). If the task goals emphasize accuracy, the system sets a conservative criterion, such that production would only proceed if conflict is very small. This would mimic a competitive selection model. If, on the other hand, speed is emphasized over accuracy, or lowering conflict with additional delay is unlikely (as in individuals with damaged production systems), a liberal criterion could be set.

The problem should be clear by now. While competitive models allow a small time window for the resolution of competition arising from semantic similarity, non-competitive models allow such competition to trickle down to lower layers and be resolved at any of those. Flexible criterion models accommodate both, and allow competition resolution from the lexical layer all the way to motor representations. Thus in non-competitive and flexible criterion models unresolved lexical competition may manifest at a much later stage, e.g., in longer VOTs (Baese-Berk & Goldrick, 2009). This may account for the wide range of effects, from very early to very late, observed in various ERP paradigms. In the next section, we discuss the ramifications of this persistent co-activation of items in production for monitoring and control processes.

### *Monitoring and control*

In the previous section, we discussed the highly likely existence of lateral inhibition mechanisms that automatically prune non-target representations and naturally contribute to the selection

process. For lateral inhibition to successfully suppress other representations in favor of the target, the target must already have stronger activation than the non-target representations (otherwise it will be the representation which is suppressed through lateral inhibition). There are three situations where this is not the case: a) Trials with high levels of noise in which the non-target representations erroneously gain a lot of activation. These often lead to errors. b) Situations in which a prepotent response needs to be overcome in order for the target to be produced (e.g., Stroop), and c) Situations in which multiple responses are more or less equally activated, without a strong preference for a single response. Recall that these three situations are identical to those described as the functions of ACC.

In the above situations, the internal dynamics of the production system alone may be insufficient to insure the desired behavioral outcome (i.e., accurate and timely production of the target word). Therefore, external intervention might be needed to regulate the system in order to ensure the production of the correct response. Determining when such intervention is required is what we consider to be the job of a monitoring system (Nozari, 2018). Note that this view differs from the classic view of monitoring as a system that detects speech errors (e.g., Levelt, 1983). While we concur that error detection is an important aspect of monitoring, we argue that the role of the monitor in the production system, like any other goal-oriented self-regulating system, is primarily to regulate the system in so as to *prevent errors*. In this broader view, monitoring requires access to the internal dynamics of the production system in order to detect error-prone situations before they lead to errors. One influential candidate for such a monitoring system is the conflict-based monitor (Botvinick et al., 2001).

The concept of the conflict-based monitor has strong ties to the main theme of this paper, namely the co-activation of representations during production. The closer the activation of these

representations, the higher the conflict. This higher conflict makes the suppression of other candidates in favor of one candidate more difficult, so the internal dynamics of lateral inhibition may no longer be sufficient to guarantee clean selection, leading to an increased probability of errors. Due to this close correspondence between conflict and error probability, conflict-based accounts view the amount of conflict as valuable information for the monitor to determine when control is required. Based on such information, the monitor then signals the need for top-down control in order to decrease conflict and facilitate clean selection.

In terms of neural correlates, neuroimaging studies of conflict monitoring and control outside of the language production literature have identified medial and lateral prefrontal areas as critical regions. Most noteworthy is the involvement of the ACC (Kerns et al., 2004), pre-SMA (Garavan, Ross, Kaufman, & Stein, 2003), and LIFG (Thompson-Schill et al., 1997). There are still debates about the function of each of these regions, but the evidence in support of conflict monitoring (and subsequent implementation of control), although not undisputed, is still strong (e.g., Ullsperger et al., 2014). Under this view, these regions, together with domain-specific representations, create monitoring-control loops that apply similar domain-general principles to provide regulation over specific domains (see Freund & Nozari, 2018, for a discussion on the balance of evidence for domain-general vs. domain-specificity of control). Specifically, one of the functions attributed to ACC is to monitor conflict arising between domain-specific representations (the monitoring leg), and to communicate with the lateral prefrontal areas including LIFG to adjust top-down control in accordance with task goals and in proportion to the need determined by the amount of conflict (the control leg). In terms of EEG correlates, conflict monitoring has been most consistently linked to the ERN (Botvinick et al., 2001). Although the exact mechanism by which ERN is produced is, like the function of the area from which it

originates (i.e., the ACC), debated, sensitivity to response conflict is still one of the most well-supported accounts of the ERN (see Ullsperger et al., 2014 for alternative accounts). Other probable indices of conflict outside the production literature include N2 and N450 (see below).

Conflict detection has been shown to be a viable mechanism for monitoring in language production, and has found empirical support in the error detection data from individuals with aphasia (Nozari et al., 2011), children (Hanley, Cortis, Budd, & Nozari, 2016), and L2 production in bilinguals (Nozari, Martin, & McCloskey, accepted with minor revisions). As reviewed in the neuroimaging section, many picture naming studies have consistently reported the activation of ACC and LIFG under situations of co-activation. These areas, together with domain-specific regions such as MTG, form the monitoring-control loop hypothesized to regulate word production.

Picture naming studies have typically also reported more negative ERN on error compared to correct productions (e.g., Ganushchak & Schiller, 2008; Masaki et al., 2001; Riès et al., 2011). If, however, ERN really indexes conflict, as opposed to error occurrence, then one must also find it on correct but high-conflict trials. Prior studies outside of the language domain have claimed that N2 is the equivalent of ERN on correct trials (Huster, Enriquez-Geppert, Lavalée, Falkenstein, & Herrmann, 2013; Yeung et al., 2004). As noted earlier, however, the direction of the N2 changes in the picture naming paradigms has been the opposite of what is predicted by the accounts that view N2 and ERN as representing the same information; while ERN is more negative for errors, N2 has been shown to be less negative for the potentially high-conflict semantically related trials (e.g., Costa et al., 2009; see also Fig. 3a). It is worth pointing out that the comparability of N2 and ERN outside of the area of language production are often based on two-choice button press tasks that are much simpler than word production. We must

also point out that many different accounts of N2 exist, some of which maintain that what is observed as N2 cannot be accounted for by a single theory (Folstein & Petten, 2008).

Another ERP waveform, called N450, is reported in several studies investigating differences between congruent and incongruent conditions, especially in the Stroop paradigm. Although varying significantly in shape, the general finding is that the component appears 350–500 ms post-stimulus-onset, and has greater negative amplitude (or, depending on the shape, less positive amplitude) for incongruent than congruent trials (e.g., Perlstein, Larson, Dotson, & Kelly, 2006; Swick & Turken, 2002; West, Jakubek, Wymbs, Perry, & Moore, 2005). The link to monitoring and control comes from its recovery from the Stroop paradigm, as well as source localization studies claiming the source of N450 to be the ACC (Liotti, Woldorff, Perez, & Mayberg, 2000; West, 2003), although neither of these studies had individual MRIs or a sufficient number of electrodes to allow for reliable source localization. Recall that effects of co-activation of items have been reported in many word production studies in this time window. Since congruent trials in Stroop map onto the same representation, while incongruent trials map onto two different representations, the general pattern of N450 matches the N400-like effects in PWI studies with identity vs. semantic distractors (Piai et al., 2012; Shitova et al., 2017): The addition of a new representation (the non-identity distractor) induces more negativity on N450 and the ERPs observed within the N400 window in PWI, blocked cyclic and continuous naming paradigms, suggesting a possible common origin.

This interpretation would also naturally fit with the interpretation that the ERPs in this time window represent the degree of the activation of unique features, hence the lower amplitude of ERPs for related compared to unrelated conditions in this window discussed earlier. Findings of one study, however, report an opposite effect. Ganushchak and Schiller (2008a) had

participants monitor for a specific phoneme in the name of a picture (e.g., /b/ in “nose”) while being presented with semantically related (e.g., “ear”), or unrelated (e.g., “window) auditory distractors. In this experiment, the N450 was found to be more negative for the related than the unrelated conditions. While compatible with the “conflict” interpretation, this result is different from the differences observed in the N400 window in picture naming studies involving related and unrelated items. The difference may reflect the different nature of the phoneme monitoring and picture naming tasks, or these ERP effects may simply have different origins. The general point we wish to make by comparing N450 and other ERP waveforms in the N400 window is that concepts of co-activation and conflict are closely intertwined. In many cases, e.g., comparison of the identity vs. related distractors, it is hard to determine which consequences of co-activation ERP differences index: more or less general activation in the system (relevant to activation dynamics), or more or less conflict between representations (relevant to selection dynamics).

While N2 and N450 have, in some accounts, been taken as ERN counterparts, the most straightforward prediction regarding the ERN is that correct but high conflict trials should show a clear ERN, similar to error trials<sup>2</sup>. In our experiment, the results of which are depicted in Fig. 3, we did not have a baseline without the strong co-activation of any items (i.e., isolated picture naming). All we can say, therefore, is that an ERN-like component with the same shape, timeline, and topography is observed on the FCz (Fig. 3c), and comparison with previous reports (Riès et al., 2011) suggests that its magnitude is twice as large as the correct trials, compatible

---

<sup>2</sup> Yeung et al. (2004) argue that conflict on correct trials should be observable before a response is made, while conflict on error trials should be most prominent after the error. It is critical to remember that this argument depends on the nature of the task. Errors that result from premature responding would have a different profile from those that have occurred after long deliberations and ultimately failure of conflict resolution. Accordingly, different tasks are associated with different relative speeds for error and correct responses. There is no reason to assume that forced-choice tasks and picture naming have a similar profile in this respect.

with a conflict-based account. This component, however, did not show sensitivity to semantic relations. It is thus possible that this specific component is only capturing co-activation of late-stage representations, such as motor programs (Ganushchak & Schiller, 2008a; Guenther, 2016). Its insensitivity to similarity at the higher levels also suggests that the activation of motor representations is not entirely bottom-up. Especially in paradigms in which a small set of pictures are repeatedly named, it is likely that motor programs are strategically kept activated, in both related and unrelated conditions, in order to facilitate production (see Belke, 2008, for a discussion of strategic effects in such paradigms).

Note, however, that the absence of an influence of semantic similarity on this specific component does not preclude a late influence of semantic factors on production. A glance at the results of Ganushchak and Schiller (2008a), who reported an influence of semantic similarity on the ERN, also suggests that the reported effect, while greater for related than unrelated trials, may not have been on what is typically labeled the ERN. Our own data, depicted in Fig. 3, panel b, also show statistically significant differences between the related and unrelated conditions around the time of articulation on Cz. The direction of the effect, however, is the opposite of that reported in Ganushchak and Schiller, with less negative amplitude for the related condition, although we only analyzed correct trials (the task was simple and did not produce enough errors to analyze), while Ganushchak and Schiller (2008a) only compared error trials. In short, effects of semantic similarity are clearly visible on late response-locked ERPs around the time of articulation. However, evidence of semantic similarity affecting the ERN is, at best, weak.

In summary, since conflict is an indirect index of co-activation of representations, any ERP changes due to such co-activation could potentially be taken as an index of conflict. As reviewed in the previous sections, such changes manifest throughout the production process. But

before they are assigned a “monitoring” interpretation, other explanations, like the dynamics of spreading activation, must be ruled out. Importantly, there is no reason to assume that monitoring “begins when the first segment of the phonological word is encoded”, as has been assumed in prior studies (Indefrey & Levelt, 2004; Python et al., 2018b).

Finally, apart from transient adjustments in control (e.g., Freund & Nozari, 2018), the monitoring-control loop most likely also mediates implicit learning processes that modify the system towards optimized performance in accordance with task goals, e.g., by changing the selection criterion over time. Understanding the effect of such processes on the ERP data requires studying the changes to the behavior and the accompanying signal over the course of the experiment.

## **Summary and conclusions**

This review highlights the existence of an amazing body of behavioral, neuroimaging, and electrophysiological data addressing the consequences of the co-activation of multiple representations during word production. At the same time, it points out non-trivial problems in the interpretations assigned to such data, especially the electrophysiological data. We discussed three such problems: using a linear timeline for interpreting events in a cascaded system, adopting the controversial assumption of competitive lexical selection, and the appropriateness of tasks from which assumptions have been extracted. The assumption of a non-serial system, where cleanly separable windows can no longer be assigned to individual operations, do not imply a dead end in research, but they do call for a shift in perspective when interpreting the neural data. Our recommendation in this regard is to separate mandatory from optional processes. Spreading activation in the semantic lexical space is a mandatory process for *any*

production attempt. Competitive lexical selection may not be. Thus the former must be given precedence in interpretation of the empirical data. Only if a difference cannot be accounted for by a mandatory process should it be considered in light of optional processes. Put concretely, a difference in the EEG signal between conditions with low and high competition in the early time window may be much better explained with how the spread of activation differs between the two conditions, as opposed to indexing competitive selection. Moreover, a neurally plausible model of spreading activation must include both excitation and inhibition, the latter of which provides a natural segue into the question of selection.

This approach, in turn, offers new ways for testing the predictions of competitive vs. non-competitive accounts. Since, according to non-competitive accounts, non-target items that have activations close to the target do not need to be strongly suppressed, these accounts naturally predict differences between semantically related and unrelated items to propagate down to the lower layers of the system and still be detectable around the articulation time. On the other hand, competitive accounts of selection, at least in their strong form, maintain that the activation of non-target items must be much lower than the target item before production can continue. Therefore, observing significant differences between semantically related and unrelated items around the articulation time is not naturally predicted by these accounts, and must be explained as a separate effect with a separate mechanism. One is then forced to define the nature of that mechanism, because, as pointed out in the monitoring and control section, monitoring mechanisms that are usually appealed to in order to explain late ERP effects are closely related to the activation levels of the competing items.

While the complexities laid out in this review make the interpretation of the neural data more complicated than current practices, we have discussed examples that demonstrate the utility

of biologically plausible computational models in understanding the brain-behavior relationship. We believe that developing such models is a necessary step in fully understanding the consequences of co-activation of multiple representations in word production.

### **Acknowledgments**

This work was funded in part by the NSF grant 1631993 awarded to N.N., and in part by the Therapeutic Cognitive Neuroscience Fund. We would like to thank Stephanie Riès and Marta Kutas for valuable discussions.

### **References**

- Abdel Rahman, R., & Melinger, A. (2007). When bees hamper the production of honey: Lexical interference from associates in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(3), 604–614. <https://doi.org/10.1037/0278-7393.33.3.604>
- 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(5), 713–734. <https://doi.org/10.1080/01690960802597250>
- Abdel Rahman, R., & Melinger, A. (2011). The dynamic microstructure of speech production: Semantic interference built on the fly. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(1), 149–161. <https://doi.org/10.1037/a0021208>
- Abel, S., Dressel, K., Bitzer, R., Kümmerer, D., Mader, I., Weiller, C., & Huber, W. (2009). The separation of processing stages in a lexical interference fMRI-paradigm. *NeuroImage*, 44(3), 1113–1124. <https://doi.org/10.1016/j.neuroimage.2008.10.018>

- Acheson, D. J., & Hagoort, P. (2014). Twisting tongues to test for conflict-monitoring in speech production. *Frontiers in Human Neuroscience*, 8.  
<https://doi.org/10.3389/fnhum.2014.00206>
- Alario, F.-X., Chainay, H., Lehericy, S., & Cohen, L. (2006). The role of the supplementary motor area (SMA) in word production. *Brain Research*, 1076(1), 129–143.  
<https://doi.org/10.1016/j.brainres.2005.11.104>
- Alario, F.-X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (2004). Predictors of picture naming speed. *Behavior Research Methods, Instruments, & Computers*, 36(1), 140–155. <https://doi.org/10.3758/BF03195559>
- Alario, F.-X., Segui, J., & Ferrand, L. (2000). Semantic and associative priming in picture naming. *The Quarterly Journal of Experimental Psychology: Section A*, 53(3), 741–764.
- Aristei, S., Melinger, A., & Abdel Rahman, R. (2011). Electrophysiological Chronometry of Semantic Context Effects in Language Production. *Journal of Cognitive Neuroscience*, 23(7), 1567–1586. <https://doi.org/10.1162/jocn.2010.21474>
- Baese-Berk, M., & Goldrick, M. (2009). Mechanisms of interaction in speech production. *Language and Cognitive Processes*, 24(4), 527–554.  
<https://doi.org/10.1080/01690960802299378>
- Barch, D. M., Braver, T. S., Sabb, F. W., & Noll, D. C. (2000). Anterior Cingulate and the Monitoring of Response Conflict: Evidence from an fMRI Study of Overt Verb Generation. *Journal of Cognitive Neuroscience*, 12(2), 298–309.  
<https://doi.org/10.1162/089892900562110>

- Belke, E. (2008). Effects of working memory load on lexical-semantic encoding in language production. *Psychonomic Bulletin & Review*, *15*(2), 357–363.  
<https://doi.org/10.3758/PBR.15.2.357>
- Belke, E. (2017). The Role of Task-Specific Response Strategies in Blocked-Cyclic Naming. *Frontiers in Psychology*, *7*. <https://doi.org/10.3389/fpsyg.2016.01955>
- Belke, E., Meyer, A. S., & Damian, M. F. (2005). Refractory effects in picture naming as assessed in a semantic blocking paradigm. *The Quarterly Journal of Experimental Psychology Section A*, *58*(4), 667–692. <https://doi.org/10.1080/02724980443000142>
- Belke, E., Shao, Z., & Meyer, A. S. (2017). Strategic origins of early semantic facilitation in the blocked-cyclic naming paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(10), 1659–1668. <https://doi.org/10.1037/xlm0000399>
- Biggs, T. C., & Marmurek, H. H. (1990). Picture and word naming: Is facilitation due to processing overlap? *The American Journal of Psychology*, *103*(1), 81–100.  
<https://doi.org/10.2307/1423260>
- Blackford, T., Holcomb, P. J., Grainger, J., & Kuperberg, G. R. (2012). A funny thing happened on the way to articulation: N400 attenuation despite behavioral interference in picture naming. *Cognition*, *123*(1), 84–99. <https://doi.org/10.1016/j.cognition.2011.12.007>
- Bles, M., & Jansma, B. M. (2008). Phonological processing of ignored distractor pictures, an fMRI investigation. *BMC Neuroscience*, *9*(1), 20. <https://doi.org/10.1186/1471-2202-9-20>
- Bohland, J. W., Bullock, D., & Guenther, F. H. (2010). Neural Representations and Mechanisms for the Performance of Simple Speech Sequences. *Journal of Cognitive Neuroscience*, *22*(7), 1504–1529. <https://doi.org/10.1162/jocn.2009.21306>

- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652.  
<https://doi.org/10.1037/0033-295X.108.3.624>
- Boylan, C., Trueswell, J. C., & Thompson-Schill, S. L. (2015). Compositionality and the angular gyrus: A multi-voxel similarity analysis of the semantic composition of nouns and verbs. *Neuropsychologia*, *78*, 130–141. <https://doi.org/10.1016/j.neuropsychologia.2015.10.007>
- Breining, B. L., Nozari, N., & Rapp, B. (2018). Learning in complex, multi-component cognitive systems: Different learning challenges within the same system. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, No Pagination Specified-No Pagination Specified. <https://doi.org/10.1037/xlm0000630>
- Breining, B., Nozari, N., & Rapp, B. (2016). Does segmental overlap help or hurt? Evidence from blocked cyclic naming in spoken and written production. *Psychonomic Bulletin & Review*, *23*(2), 500–506. <https://doi.org/10.3758/s13423-015-0900-x>
- Brown, C., & Hagoort, P. (1993). The Processing Nature of the N400: Evidence from Masked Priming. *Journal of Cognitive Neuroscience*, *5*(1), 34–44.  
<https://doi.org/10.1162/jocn.1993.5.1.34>
- Canini, M., Rosa, P. A. D., Catricalà, E., Strijkers, K., Branzi, F. M., Costa, A., & Abutalebi, J. (2016). Semantic interference and its control: A functional neuroimaging and connectivity study. *Human Brain Mapping*, *37*(11), 4179–4196.  
<https://doi.org/10.1002/hbm.23304>
- Catling, J. C., Dent, K., Johnston, R. A., & Balding, R. (2010). Age of acquisition, word frequency, and picture–word interference. *The Quarterly Journal of Experimental Psychology*, *63*(7), 1304–1317. <https://doi.org/10.1080/17470210903380830>

- Cheyette, S. J., & Plaut, D. C. (2017). Modeling the N400 ERP component as transient semantic over-activation within a neural network model of word comprehension. *Cognition*, *162*, 153–166. <https://doi.org/10.1016/j.cognition.2016.10.016>
- 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*(1), 125–131. <https://doi.org/10.3758/BF03196357>
- Costa, A., Strijkers, K., Martin, C., & Thierry, G. (2009). The time course of word retrieval revealed by event-related brain potentials during overt speech. *Proceedings of the National Academy of Sciences*, *106*(50), 21442–21446. <https://doi.org/10.1073/pnas.0908921106>
- de Zubicaray, G. I., Hansen, S., & McMahon, K. L. (2013). Differential processing of thematic and categorical conceptual relations in spoken word production. *Journal of Experimental Psychology: General*, *142*(1), 131–142. <https://doi.org/10.1037/a0028717>
- de Zubicaray, G. I., Johnson, K., Howard, D., & McMahon, K. (2014). A perfusion fMRI investigation of thematic and categorical context effects in the spoken production of object names. *Cortex*, *54*, 135–149. <https://doi.org/10.1016/j.cortex.2014.01.018>
- de Zubicaray, G. I., McMahon, K., Eastburn, M., & Pringle, A. (2006). Top-down influences on lexical selection during spoken word production: A 4T fMRI investigation of refractory effects in picture naming. *Human Brain Mapping*, *27*(11), 864–873. <https://doi.org/10.1002/hbm.20227>
- de Zubicaray, G. I., McMahon, K., & Howard, D. (2015). Perfusion fMRI evidence for priming of shared feature-to-lexical connections during cumulative semantic interference in

- spoken word production. *Language, Cognition and Neuroscience*, 30(3), 261–272.  
<https://doi.org/10.1080/01690965.2013.848990>
- de Zubicaray, G. I., & McMahon, K. L. (2009). Auditory context effects in picture naming investigated with event-related fMRI. *Cognitive, Affective, & Behavioral Neuroscience*, 9(3), 260–269. <https://doi.org/10.3758/CABN.9.3.260>
- de Zubicaray, G. I., McMahon, K. L., Eastburn, M. M., & Wilson, S. J. (2002). Orthographic/Phonological Facilitation of Naming Responses in the Picture–Word Task: An Event-Related fMRI Study Using Overt Vocal Responding. *NeuroImage*, 16(4), 1084–1093. <https://doi.org/10.1006/nimg.2002.1135>
- de Zubicaray, G. I., Miozzo, M., Johnson, K., Schiller, N. O., & McMahon, K. L. (2012). Independent Distractor Frequency and Age-of-Acquisition Effects in Picture–Word Interference: fMRI Evidence for Post-lexical and Lexical Accounts according to Distractor Type. *Journal of Cognitive Neuroscience*, 24(2), 482–495.  
[https://doi.org/10.1162/jocn\\_a\\_00141](https://doi.org/10.1162/jocn_a_00141)
- de Zubicaray, G. I., Wilson, S. J., McMahon, K. L., & Muthiah, S. (2001). The semantic interference effect in the picture-word paradigm: An event-related fMRI study employing overt responses. *Human Brain Mapping*, 14(4), 218–227.  
<https://doi.org/10.1002/hbm.1054>
- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93(3), 283–321. <https://doi.org/10.1037/0033-295X.93.3.283>
- Dell, G. S., Nozari, N., & Oppenheim, G. M. (2014). Lexical access: Behavioral and computational considerations. In V. Ferreira, M. Goldrick, & M. Miozzo (Eds.), *The*

- Oxford Handbook of Language Production* (pp. 88-104). Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199735471.013.014>
- Dell'Acqua, R., Sessa, P., Peressotti, F., Mulatti, C., Navarrete, E., & Grainger, J. (2010). ERP Evidence for Ultra-Fast Semantic Processing in the Picture–Word Interference Paradigm. *Frontiers in Psychology, 1*. <https://doi.org/10.3389/fpsyg.2010.00177>
- Falkenstein, M., Hoormann, J., Christ, S., & Hohnsbein, J. (2000). ERP components on reaction errors and their functional significance: a tutorial. *Biological Psychology, 51*(2), 87–107. [https://doi.org/10.1016/S0301-0511\(99\)00031-9](https://doi.org/10.1016/S0301-0511(99)00031-9)
- Folstein, J. R., & Petten, C. V. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology, 45*(1), 152–170. <https://doi.org/10.1111/j.1469-8986.2007.00602.x>
- Freund, M., & Nozari, N. (2018). Is adaptive control in language production mediated by learning? *Cognition, 176*, 107–130. <https://doi.org/10.1016/j.cognition.2018.03.009>
- Fu, Z., Wu, D.-A. J., Ross, I., Chung, J. M., Mamelak, A. N., Adolphs, R., & Rutishauser, U. (2019). Single-Neuron Correlates of Error Monitoring and Post-Error Adjustments in Human Medial Frontal Cortex. *Neuron, 101*(1), 165-177.e5. <https://doi.org/10.1016/j.neuron.2018.11.016>
- Ganushchak, L. Y., & Schiller, N. O. (2008a). Brain Error–monitoring Activity is Affected by Semantic Relatedness: An Event-related Brain Potentials Study. *Journal of Cognitive Neuroscience, 20*(5), 927–940. <https://doi.org/10.1162/jocn.2008.20514>
- Ganushchak, L. Y., & Schiller, N. O. (2008b). Motivation and semantic context affect brain error-monitoring activity: An event-related brain potentials study. *NeuroImage, 39*(1), 395–405. <https://doi.org/10.1016/j.neuroimage.2007.09.001>

- Garavan, H., Ross, T. J., Kaufman, J., & Stein, E. A. (2003). A midline dissociation between error-processing and response-conflict monitoring. *NeuroImage*, *20*(2), 1132–1139.  
[https://doi.org/10.1016/S1053-8119\(03\)00334-3](https://doi.org/10.1016/S1053-8119(03)00334-3)
- Goldrick, M., & Blumstein, S. E. (2007). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes*. <https://doi.org/10.1080/01690960500181332>
- 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), 1–40.  
[https://doi.org/10.1207/s15516709cog2701\\_1](https://doi.org/10.1207/s15516709cog2701_1)
- Guenther, F.H. (2016). *Neural Control of Speech*. Cambridge, MA: MIT Press.  
<http://dx.doi.org/10.7551/mitpress/10471.001.0001>
- Hanley, J. R., Cortis, C., Budd, M.-J., & Nozari, N. (2016). Did I say dog or cat? A study of semantic error detection and correction in children. *Journal of Experimental Child Psychology*, *142*, 36–47. <https://doi.org/10.1016/j.jecp.2015.09.008>
- Heim, S., Friederici, A. D., Schiller, N. O., Rüschemeyer, S.-A., & Amunts, K. (2009). The determiner congruency effect in language production investigated with functional MRI. *Human Brain Mapping*, *30*(3), 928–940. <https://doi.org/10.1002/hbm.20556>
- Hirschfeld, G., Jansma, B., Bölte, J., & Zwitserlood, P. (2008). Interference and facilitation in overt speech production investigated with event-related potentials: *NeuroReport*, *19*(12), 1227–1230. <https://doi.org/10.1097/WNR.0b013e328309ecd1>
- Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: experimental and computational studies. *Cognition*, *100*(3), 464–482.  
<https://doi.org/10.1016/j.cognition.2005.02.006>

- Huster, R. J., Enriquez-Geppert, S., Lavallee, C. F., Falkenstein, M., & Herrmann, C. S. (2013). Electroencephalography of response inhibition tasks: Functional networks and cognitive contributions. *International Journal of Psychophysiology*, *87*(3), 217–233.  
<https://doi.org/10.1016/j.ijpsycho.2012.08.001>
- Huttenlocher, J., & Kubicek, L. F. (1983). The source of relatedness effects on naming latency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*(3), 486–496.  
<https://doi.org/10.1037/0278-7393.9.3.486>
- Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1), 101–144. <https://doi.org/10.1016/j.cognition.2002.06.001>
- Indefrey, Peter. (2011). The Spatial and Temporal Signatures of Word Production Components: A Critical Update. *Frontiers in Psychology*, *2*. <https://doi.org/10.3389/fpsyg.2011.00255>
- Janssen, N., Carreiras, M., & Barber, H. A. (2011). Electrophysiological effects of semantic context in picture and word naming. *NeuroImage*, *57*(3), 1243–1250.  
<https://doi.org/10.1016/j.neuroimage.2011.05.015>
- Janssen, N., Hernández-Cabrera, J. A., van der Meij, M., & Barber, H. A. (2015). Tracking the Time Course of Competition During Word Production: Evidence for a Post-Retrieval Mechanism of Conflict Resolution. *Cerebral Cortex*, *25*(9), 2960–2969.  
<https://doi.org/10.1093/cercor/bhu092>
- Kalfaoğlu, Ç., Stafford, T., & Milne, E. (2018). Frontal theta band oscillations predict error correction and posterror slowing in typing. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(1), 69–88. <https://doi.org/10.1037/xhp0000417>

- Kan, I. P., & Thompson-Schill, S. L. (2004). Selection from perceptual and conceptual representations. *Cognitive, Affective, & Behavioral Neuroscience*, 4(4), 466–482.  
<https://doi.org/10.3758/CABN.4.4.466>
- Kemmerer, D. (2018). From blueprints to brain maps: the status of the Lemma Model in cognitive neuroscience. *Language, Cognition and Neuroscience*, 0(0), 1–32.  
<https://doi.org/10.1080/23273798.2018.1537498>
- Kerns, J. G., Cohen, J. D., MacDonald, A. W., Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior Cingulate Conflict Monitoring and Adjustments in Control. *Science*, 303(5660), 1023–1026. <https://doi.org/10.1126/science.1089910>
- Kutas, M., & Federmeier, K. D. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology*, 62(1), 621–647. <https://doi.org/10.1146/annurev.psych.093008.131123>
- La Heij, W., Dirx, J., & Kramer, P. (1990). Categorical interference and associative priming in picture naming. *British Journal of Psychology*, 81(4), 511–525.  
<https://doi.org/10.1111/j.2044-8295.1990.tb02376.x>
- Laszlo, S., & Plaut, D. C. (2012). A neurally plausible Parallel Distributed Processing model of Event-Related Potential word reading data. *Brain and Language*, 120(3), 271–281.  
<https://doi.org/10.1016/j.bandl.2011.09.001>
- Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, 14(1), 41–104.  
[https://doi.org/10.1016/0010-0277\(83\)90026-4](https://doi.org/10.1016/0010-0277(83)90026-4)
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(1), 1–38.  
<https://doi.org/10.1017/S0140525X99001776>

- Liotti, M., Woldorff, M. G., Perez, R., & Mayberg, H. S. (2000). An ERP study of the temporal course of the Stroop color-word interference effect. *Neuropsychologia*, *38*(5), 701–711. [https://doi.org/10.1016/S0028-3932\(99\)00106-2](https://doi.org/10.1016/S0028-3932(99)00106-2)
- Llorens, A., Trébuchon, A., Riès, S., Liégeois-Chauvel, C., & Alario, F.-X. (2014). How familiarization and repetition modulate the picture naming network. *Brain and Language*, *133*, 47–58. <https://doi.org/10.1016/j.bandl.2014.03.010>
- Lupker, S. J. (1979). The semantic nature of response competition in the picture-word interference task. *Memory & Cognition*, *7*(6), 485–495. <https://doi.org/10.3758/BF03198265>
- Lupker, S. J. (1988). Picture naming: An investigation of the nature of categorical priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 444–455. <https://doi.org/10.1037/0278-7393.14.3.444>
- Maess, B., Friederici, A. D., Damian, M., Meyer, A. S., & Levelt, W. J. M. (2002). Semantic Category Interference in Overt Picture Naming: Sharpening Current Density Localization by PCA. *Journal of Cognitive Neuroscience*, *14*(3), 455–462. <https://doi.org/10.1162/089892902317361967>
- 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*(3), 503–535. <https://doi.org/10.1037/0278-7393.33.3.503>

- Masaki, H., Tanaka, H., Takasawa, N., & Yamazaki, K. (2001). Error-related brain potentials elicited by vocal errors: *Neuroreport*, *12*(9), 1851–1855.  
<https://doi.org/10.1097/00001756-200107030-00018>
- Mechelli, A., Josephs, O., Ralph, M. A. L., McClelland, J. L., & Price, C. J. (2007). Dissociating stimulus-driven semantic and phonological effect during reading and naming. *Human Brain Mapping*, *28*(3), 205–217. <https://doi.org/10.1002/hbm.20272>
- Miozzo, M., Pulvermüller, F., & Hauk, O. (2015). Early Parallel Activation of Semantics and Phonology in Picture Naming: Evidence from a Multiple Linear Regression MEG Study. *Cerebral Cortex*, *25*(10), 3343–3355. <https://doi.org/10.1093/cercor/bhu137>
- Möller, J., Jansma, B. M., Rodriguez-Fornells, A., & Münte, T. F. (2007). What the Brain Does before the Tongue Slips. *Cerebral Cortex*, *17*(5), 1173–1178.  
<https://doi.org/10.1093/cercor/bhl028>
- Munding, D., Dubarry, A.-S., & Alario, F.-X. (2016). On the cortical dynamics of word production: a review of the MEG evidence. *Language, Cognition and Neuroscience*, *31*(4), 441–462. <https://doi.org/10.1080/23273798.2015.1071857>
- Navarrete, E., Del Prato, P., Peressotti, F., & Mahon, B. Z. (2014). Lexical selection is not by competition: Evidence from the blocked naming paradigm. *Journal of Memory and Language*, *76*, 253–272. <https://doi.org/10.1016/j.jml.2014.05.003>
- Noonan, K. A., Jefferies, E., Visser, M., & Lambon Ralph, M. A. (2013). Going beyond inferior prefrontal involvement in semantic control: evidence for the additional contribution of dorsal angular gyrus and posterior middle temporal cortex. *Journal of Cognitive Neuroscience*, *25*(11), 1824–1850.

- Nozari, N. (2018). How Special Is Language Production? Perspectives From Monitoring and Control. In *Psychology of Learning and Motivation* (Vol. 68, pp. 179–213).  
<https://doi.org/10.1016/bs.plm.2018.08.006>
- Nozari, N. (to appear). Neural basis of word production. In L. Gleitman, J. Trueswell, & A. Papafragou (Eds.), *Oxford Handbook of Mental Lexicon*. Oxford, UK: Oxford University Press.
- Nozari, N. (2019). The dual origin of semantic errors in access deficit: activation vs. inhibition deficit. *Cognitive Neuropsychology*, *0*(0), 1–23.  
<https://doi.org/10.1080/02643294.2019.1587397>
- Nozari, N., Dell, G. S., & Schwartz, M. F. (2011). Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive Psychology*, *63*(1), 1–33. <https://doi.org/10.1016/j.cogpsych.2011.05.001>
- Nozari, N., Freund, M., Breining, B., Rapp, B., & Gordon, B. (2016). Cognitive control during selection and repair in word production. *Language, Cognition and Neuroscience*, *31*(7), 886–903. <https://doi.org/10.1080/23273798.2016.1157194>
- Nozari, N., & Hepner, C. R. (2018). To select or to wait? The importance of criterion setting in debates of competitive lexical selection. *Cognitive Neuropsychology*, *0*(0), 1–15.  
<https://doi.org/10.1080/02643294.2018.1476335>
- Nozari, N., & Hepner, C. R. (2019). To select or to wait? Response to the commentaries. *Cognitive Neuropsychology*. <https://doi.org/10.1080/02643294.2019.1632280>
- Nozari, N., Martin, C. D., & McCloskey, N. (accepted with minor revisions). Is repairing speech errors an automatic or a controlled process? Insights from the relationship between error and repair probabilities in English and Spanish. *Language, Cognition, & Neuroscience*.

- Nozari, N., & Thompson-Schill, S. L. (2016). Chapter 46 - Left Ventrolateral Prefrontal Cortex in Processing of Words and Sentences. In G. Hickok & S. L. Small (Eds.), *Neurobiology of Language* (pp. 569–584). <https://doi.org/10.1016/B978-0-12-407794-2.00046-8>
- Oppenheim, G. M. (2018). The paca that roared: Immediate cumulative semantic interference among newly acquired words. *Cognition*, *177*, 21–29.  
<https://doi.org/10.1016/j.cognition.2018.02.014>
- Oppenheim, G. M., & Balatsou, E. (2019). Lexical competition on demand. *Cognitive Neuropsychology*, 1–4. <https://doi.org/10.1080/02643294.2019.1580189>
- 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*(2), 227–252. <https://doi.org/10.1016/j.cognition.2009.09.007>
- O’Séaghdha, P. G., & Frazer, A. K. (2014). The exception does not rule: Attention constrains form preparation in word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(3), 797–810. <https://doi.org/10.1037/a0035576>
- Peramunage, D., Blumstein, S. E., Myers, E. B., Goldrick, M., & Baese-Berk, M. (2011). Phonological Neighborhood Effects in Spoken Word Production: An fMRI Study. *Journal of Cognitive Neuroscience*, *23*(3), 593–603.  
<https://doi.org/10.1162/jocn.2010.21489>
- Perlstein, W. M., Larson, M. J., Dotson, V. M., & Kelly, K. G. (2006). Temporal dissociation of components of cognitive control dysfunction in severe TBI: ERPs and the cued-Stroop task. *Neuropsychologia*, *44*(2), 260–274.  
<https://doi.org/10.1016/j.neuropsychologia.2005.05.009>

- Piai, V., Roelofs, A., & van der Meij, R. (2012). Event-related potentials and oscillatory brain responses associated with semantic and Stroop-like interference effects in overt naming. *Brain Research, 1450*, 87–101. <https://doi.org/10.1016/j.brainres.2012.02.050>
- Python, G., Fargier, R., & Laganaro, M. (2018a). ERP evidence of distinct processes underlying semantic facilitation and interference in word production. *Cortex, 99*, 1–12. <https://doi.org/10.1016/j.cortex.2017.09.008>
- Python, G., Fargier, R., & Laganaro, M. (2018b). When Wine and Apple Both Help the Production of Grapes: ERP Evidence for Post-lexical Semantic Facilitation in Picture Naming. *Frontiers in Human Neuroscience, 12*. <https://doi.org/10.3389/fnhum.2018.00136>
- Rabovsky, M., & McRae, K. (2014). Simulating the N400 ERP component as semantic network error: Insights from a feature-based connectionist attractor model of word meaning. *Cognition, 132*(1), 68–89. <https://doi.org/10.1016/j.cognition.2014.03.010>
- Riès, S. , Dhillon, R. K., Clarke, A., King-Stephens, D., Laxer, K. D., Weber, P. B., ... Knight, R. T. (2017). Spatiotemporal dynamics of word retrieval in speech production revealed by cortical high-frequency band activity. *Proceedings of the National Academy of Sciences, 114*(23), E4530–E4538. <https://doi.org/10.1073/pnas.1620669114>
- Riès, S. , Janssen, N., Burle, B., & Alario, F.-X. (2013). Response-Locked Brain Dynamics of Word Production. *PLOS ONE, 8*(3), e58197. <https://doi.org/10.1371/journal.pone.0058197>
- Riès, S. , Janssen, N., Dufau, S., Alario, F.-X., & Burle, B. (2011). General-Purpose Monitoring during Speech Production. *Journal of Cognitive Neuroscience, 23*(6), 1419–1436. <https://doi.org/10.1162/jocn.2010.21467>

- Roelofs, A. (2018). A unified computational account of cumulative semantic, semantic blocking, and semantic distractor effects in picture naming. *Cognition*, *172*, 59–72.  
<https://doi.org/10.1016/j.cognition.2017.12.007>
- Rose, S. B., Aristei, S., Melinger, A., & Abdel Rahman, R. (2019). The closer they are, the more they interfere: Semantic similarity of word distractors increases competition in language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(4), 753–763. <https://doi.org/10.1037/xlm0000592>
- Rose, S. B., & Rahman, R. A. (2017). Semantic similarity promotes interference in the continuous naming paradigm: behavioural and electrophysiological evidence. *Language, Cognition and Neuroscience*, *32*(1), 55–68.  
<https://doi.org/10.1080/23273798.2016.1212081>
- Sadat, J., Martin, C. D., Costa, A., & Alario, F.-X. (2014). Reconciling phonological neighborhood effects in speech production through single trial analysis. *Cognitive Psychology*, *68*, 33–58. <https://doi.org/10.1016/j.cogpsych.2013.10.001>
- Schnur, T. T. (2014). The persistence of cumulative semantic interference during naming. *Journal of Memory and Language*, *75*, 27–44. <https://doi.org/10.1016/j.jml.2014.04.006>
- Schnur, T. T., Schwartz, M. F., Brecher, A., & Hodgson, C. (2006). Semantic interference during blocked-cyclic naming: Evidence from aphasia. *Journal of Memory and Language*, *54*(2), 199–227. <https://doi.org/10.1016/j.jml.2005.10.002>
- Schnur, T. T., Schwartz, M. F., Kimberg, D. Y., Hirshorn, E., Coslett, H. B., & Thompson-Schill, S. L. (2009). Localizing interference during naming: Convergent neuroimaging and neuropsychological evidence for the function of Broca's area. *Proceedings of the*

*National Academy of Sciences*, 106(1), 322–327.

<https://doi.org/10.1073/pnas.0805874106>

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(1), 86–102. [https://doi.org/10.1016/0749-596X\(90\)90011-N](https://doi.org/10.1016/0749-596X(90)90011-N)

Schwartz, M. F., Kimberg, D. Y., Walker, G. M., Brecher, A., Faseyitan, O. K., Dell, G. S., ... Coslett, H. B. (2011). Neuroanatomical dissociation for taxonomic and thematic knowledge in the human brain. *Proceedings of the National Academy of Sciences*, 108(20), 8520–8524. <https://doi.org/10.1073/pnas.1014935108>

Shao, Z., Roelofs, A., Martin, R. C., & Meyer, A. S. (2015). Selective inhibition and naming performance in semantic blocking, picture-word interference, and color-word Stroop tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(6), 1806–1820. <https://doi.org/10.1037/a0039363>

Shima, K., & Tanji, J. (2000). Neuronal Activity in the Supplementary and Presupplementary Motor Areas for Temporal Organization of Multiple Movements. *Journal of Neurophysiology*, 84(4), 2148–2160. <https://doi.org/10.1152/jn.2000.84.4.2148>

Shitova, N., Roelofs, A., Schriefers, H., Bastiaansen, M., & Schoffelen, J.-M. (2017). Control adjustments in speaking: Electrophysiology of the Gratton effect in picture naming. *Cortex*, 92, 289–303. <https://doi.org/10.1016/j.cortex.2017.04.017>

Spalek, K., & Thompson-Schill, S. L. (2008). Task-dependent semantic interference in language production: An fMRI study. *Brain and Language*, 107(3), 220–228.

<https://doi.org/10.1016/j.bandl.2008.05.005>

- Strijkers, K., Costa, A., & Thierry, G. (2010). Tracking Lexical Access in Speech Production: Electrophysiological Correlates of Word Frequency and Cognate Effects. *Cerebral Cortex*, 20(4), 912–928. <https://doi.org/10.1093/cercor/bhp153>
- Swick, D., & Turken, A. U. (2002). Dissociation between conflict detection and error monitoring in the human anterior cingulate cortex. *Proceedings of the National Academy of Sciences*, 99(25), 16354–16359. <https://doi.org/10.1073/pnas.252521499>
- Thompson-Schill, S. L., D’Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences*, 94(26), 14792–14797. <https://doi.org/10.1073/pnas.94.26.14792>
- Thompson-Schill, S. L., Swick, D., Farah, M. J., D’Esposito, M., Kan, I. P., & Knight, R. T. (1998). Verb generation in patients with focal frontal lesions: A neuropsychological test of neuroimaging findings. *Proceedings of the National Academy of Sciences*, 95(26), 15855–15860. <https://doi.org/10.1073/pnas.95.26.15855>
- Trevartha, K. M., & Phillips, N. (2013). Detecting self-produced speech errors before and after articulation: an ERP investigation. *Frontiers in Human Neuroscience*, 7. <https://doi.org/10.3389/fnhum.2013.00763>
- Ullsperger, M., Fischer, A. G., Nigbur, R., & Endrass, T. (2014). Neural mechanisms and temporal dynamics of performance monitoring. *Trends in Cognitive Sciences*, 18(5), 259–267. <https://doi.org/10.1016/j.tics.2014.02.009>
- Valente, A., Bürki, A., & Laganaro, M. (2014). ERP correlates of word production predictors in picture naming: a trial by trial multiple regression analysis from stimulus onset to response. *Frontiers in Neuroscience*, 8. <https://doi.org/10.3389/fnins.2014.00390>

- Vitkovitch, M., Rutter, C., & Read, A. (2001). Inhibitory effects during object name retrieval: The effect of interval between prime and target on picture naming responses. *British Journal of Psychology*, *92*(3), 483–506. <https://doi.org/10.1348/000712601162301>
- Wang, M., Shao, Z., Chen, Y., & Schiller, N. O. (2018). Neural correlates of spoken word production in semantic and phonological blocked cyclic naming. *Language, Cognition and Neuroscience*, *33*(5), 575–586. <https://doi.org/10.1080/23273798.2017.1395467>
- Wei, T., & Schnur, T. T. (2019). Being fast or slow at naming depends on recency of experience. *Cognition*, *182*, 165–170. <https://doi.org/10.1016/j.cognition.2018.09.012>
- West, R. (2003). Neural correlates of cognitive control and conflict detection in the Stroop and digit-location tasks. *Neuropsychologia*, *41*(8), 1122–1135. [https://doi.org/10.1016/S0028-3932\(02\)00297-X](https://doi.org/10.1016/S0028-3932(02)00297-X)
- West, R., Jakubek, K., Wymbs, N., Perry, M., & Moore, K. (2005). Neural correlates of conflict processing. *Experimental Brain Research*, *167*(1), 38–48. <https://doi.org/10.1007/s00221-005-2366-y>
- Wheeldon, L. R., & Monsell, S. (1994). Inhibition of Spoken Word Production by Priming a Semantic Competitor. *Journal of Memory and Language*, *33*(3), 332–356. <https://doi.org/10.1006/jmla.1994.1016>
- Wong, A. W.-K., Chiu, H.-C., Wang, J., Cao, J., Wong, S.-S., & Chen, H.-C. (2017). An early locus of associative and categorical context effects in speech production: evidence from an ERP study using the picture–word interference paradigm. *Language, Cognition and Neuroscience*, *32*(10), 1305–1319. <https://doi.org/10.1080/23273798.2017.1355060>

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The Neural Basis of Error Detection: Conflict Monitoring and the Error-Related Negativity. *Psychological Review*, *111*(4), 931–959. <https://doi.org/10.1037/0033-295X.111.4.931>

Zhu, X., Damian, M. F., & Zhang, Q. (2015). Seriality of semantic and phonological processes during overt speech in Mandarin as revealed by event-related brain potentials. *Brain and Language*, *144*, 16–25. <https://doi.org/10.1016/j.bandl.2015.03.007>

## Footnotes

1. Note that by “features” we simply mean the units composing the representation of an item, and not physical features (see Introduction for details). We call them “features” in order to emphasize that, while they may not represent a real-life feature, they indeed capture the similarity between items.

2. Yeung et al. (2004) argue that conflict on correct trials should be observable before a response is made, while conflict on error trials should be most prominent after the error. It is critical to remember that this argument depends on the nature of the task. Errors that result from premature responding would have a different profile from those that have occurred after long deliberations and ultimately failure of conflict resolution. Accordingly, different tasks are associated with different relative speeds for error and correct responses. There is no reason to assume that forced-choice tasks and picture naming have a similar profile in this respect.

## Figure captions

Fig. 1. Schematic of the co-activation of multiple lexical items during word production using localist representations (e.g., Dell, 1986). Target is “cat”. Lexical nodes “dog” and “mat” are co-activated through the shared semantic features (darker solid connections), and shared segments (darker dashed connections) with “cat”.

Fig. 3. Schematic of the neural regions commonly implicated in neuroimaging studies of co-activation of words during picture naming. ACC: anterior cingulate cortex; IPL: inferior parietal lobule; LIFG: left inferior frontal gyrus; MTG: middle temporal gyrus; pSTG: posterior superior temporal gyrus; pre-SMA: presupplementary motor area; SMA: supplementary motor area.

Fig. 3. ERPs for producing the same word in the context of semantically related (grey) vs. unrelated (black) representations. Stimulus-locked (panel a) and response-locked (panel b) ERPs are shown on the Cz electrode. Grey shaded area corresponds to periods of statistical significance, tested with non-parametric point-by-point Wilcoxon signed-rank test with permutations, corrected for a minimum duration of 20ms. Panel c shows Laplacian-transformed ERPs on the FCz electrode. The shape, timing, and topography of this component are compatible with the ERN. Average response time is 637 ( $\pm 164$ ) ms.

Fig. 4. Schematic of distributed semantic lexical representations for three items A, B, and C. A and B are unrelated, while A and C are related. Lower case letters denote “features”. Similar items A and C share two of their “features”. A feature is either “off” (shaded white), or “on” (light gray). For an item to be produced, all its features must be “on”. Total activation, counted as the number of “on” features, is lower for the related pair A and C, compared to the unrelated pair A and B.