



HAL
open science

Lexical access in speech production: Psycho- and neurolinguistics perspectives on the spatiotemporal dynamics

Emilia Kerr, Bissera Ivanova, Kristof Strijkers

► **To cite this version:**

Emilia Kerr, Bissera Ivanova, Kristof Strijkers. Lexical access in speech production: Psycho- and neurolinguistics perspectives on the spatiotemporal dynamics. Robert Hartsuiker & Kristof Strijkers. Current issues in the Psychology of Language, Routledge press (Taylor & Francis), inPress. hal-03930891

HAL Id: hal-03930891

<https://hal.science/hal-03930891>

Submitted on 20 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accepted manuscript (uncorrected)

*To appear in the Current issues in the Psychology of Language. Routledge press
(Taylor & Francis), UK.*

LEXICAL ACCESS IN SPEECH PRODUCTION

Psycho- and neurolinguistic perspectives on the spatiotemporal dynamics.

Emilia Kerr, Bissera Ivanova & Kristof Strijkers

Aix-Marseille University, CNRS & LPL, 13001 Aix-en-Provence, France.

Corresponding author:

Kristof Strijkers

CNRS & Aix-Marseille Université (AMU)

Laboratoire Parole et Langage (LPL)

5 Avenue Pasteur

13001 Aix-en-Provence, France

Contact : kristof.strijkers@univ-amu.fr

Introduction

The speed and ease with which we produce words has puzzled researchers for decades. Uttering a single word comprises a great number of mental operations like conceptual selection ('choosing' the concept we are about to name), lexical retrieval (selecting the correct, and grammatically specified lemma for that concept), phonological encoding (retrieving the phonological form of the word), articulatory preparation and, lastly, producing the correct sequence of sounds that represent the intended word. All of this is processed by our brain in a few hundreds of milliseconds and is largely error-free. In other words, even though language production is an immensely complex psychomotor skill, we nonetheless manage to achieve it (apparently) quite effortlessly. Hence, our brain must be particularly efficient and successful in organizing the representations and dynamics underpinning our ability to speak. Understanding the nature of that organization has been a key research endeavor in the field of language production. In this chapter we will offer an overview of some of these potential architectures for the retrieval of the mental representation of words, better known as lexical access. To access a word is to retrieve it from the mental lexicon, a vast lexical storage that for an average adult language user comprises, according to different estimates, from 1000 to about 100,000 words (e.g., Levelt, 1989). In essence, for spoken word production, lexical access means coupling conceptual representations and their phonological forms. Below we will give an overview of four different types of word production models, which serve as a guide to highlight the different possible cognitive and neurobiological architectures that can support lexical access. First, we will review traditional serial and interactive theories (Part I and Part II), and then we will move onto some more recent models, namely dual-stream (Part III) and parallel (Part IV) models of lexical access in speech production.

I. The sequential (serial)¹ model of lexical access in word production.

The first model that we will cover is the serial model developed by Willem J.M. Levelt and colleagues (Levelt et al., 1999; see also Levelt, 1989) and later expanded into a neural model (Indefrey and Levelt, 2004; see also Indefrey, 2011). As the name suggests, one of the main principles is the sequentiality underlying lexical access during word production (see footnote 1). It advocates progressive step-by-step processing of each linguistic level before advancing to the next level in the hierarchy (Levelt et al., 1999) and considers functional specialisation² of brain areas that are involved in the representation of word components (Indefrey & Levelt, 2004). According to this model, the mental lexicon concerns an independent processing layer within the speech production architecture preceded in time by conceptual processing and followed by form encoding, housed in the left mid temporal gyrus (MTG) and functionally active at a specific point in time (roughly between 150 and 250 ms of processing). Below we will outline the main principles and spatiotemporal dynamics behind this model.

While most prior language production models were motivated by aphasic speech and speech error data (e.g., Dell, 1986; Garrett, 1975; Fromkin, 1971), the Levelt et al. model deviates from that tradition by being based on reaction time data of psycholinguistic experiments in healthy speakers (Glaser, 1992; Levelt, 1989; Levelt et al., 1991; Levelt &

¹ The terms seriality and sequentiality do not denote the same concept: seriality (or discreteness as also often associated historically with these types of models) means that stage B can only be initiated after stage A has been completed; sequentiality, on the other hands, allows some degree of overlap between stage A and B (i.e., like 'cascading' where a representational layer lower in the hierarchy can become activated prior to selection of a representation higher in the processing hierarchy), though stage A would still initiate well before stage B. Historically, the Levelt et al. (1999) model is serial (discrete). However, nowadays, most proponents of this model would agree with some degree of sequentially (cascading) instead strict seriality. For the present chapter, we group the serial and sequential models together (and thus also the historic division between serial and cascaded processing), since their differences are less relevant for present purposes.

² To avoid confusion, functional specialization means a given brain region X supports a function Y that other brain regions do not support, which is different from functional specificity where a given brain region X 'uniquely' supports a given function Y and nothing else (for discussion related to lexical access see: e.g., Indefrey, 2016; Strijkers & Costa, 2016).

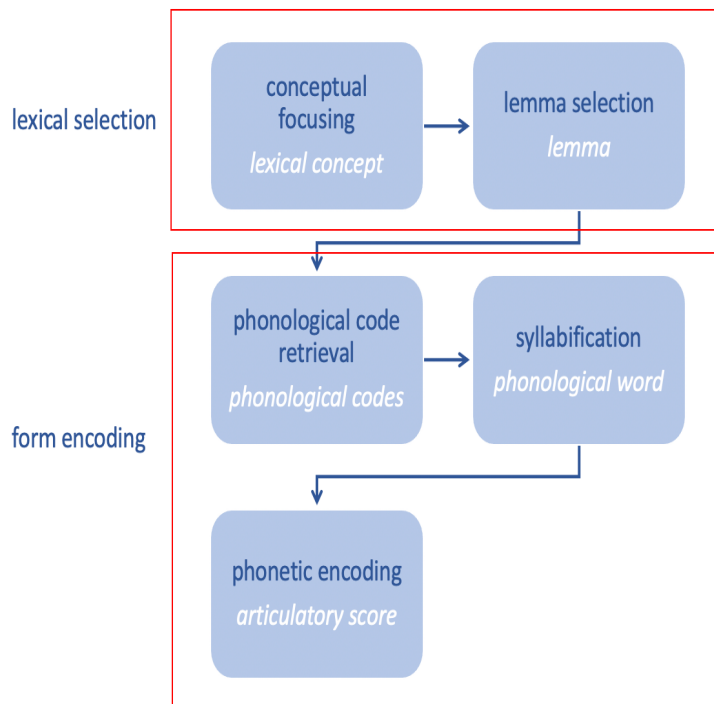


Figure 1. Schematic representation of the serial/sequential model language production model, where lexical selection concerns the process of translating a lexical concept into a lemma representation, and form encoding concerns the retrieval of the phonological and phonetic information of the activated lemma. (Based on Levelt, 2001).

Kelter, 1982; Meyer, 1992; Schriefers et al., 1990). Figure 1 represents a schematic overview of the word production system as proposed by Levelt and colleagues (1999; see also Levelt, 2001). The first step is lexical selection, that is, when a speaker selects an item from their mental lexicon. The way this works is that a speaker activates an intended concept s/he wants to utter (i.e., conceptual focusing), which in turn will activate word candidates in the mental lexicon

until an appropriate lemma, i.e., a morpho-syntactic representation of a lexical concept, is being singled out. During lexical selection there is thought to be competition between different lexical candidates due to spreading activation: That is, when activating an intended concept (e.g., CAT) related concepts (e.g., DOG) become (partially) activated as well because of their association with the target concept (spreading activation between strongly interconnected representations). This in turn will activate different lemma representations (e.g., cat and dog) which enter into a competitive process for selection, since after the lemma stage the system is thought to continue processing with a single representation, namely, the target lemma intended for articulation. The empirical input supporting these notions that lexical access is a competitive process where only a single lexical representation will be

selected for further phonological processing stems from semantic interference experiments, namely the well-known picture-word interference (PWI) paradigm (Glaser & Dünghoff, 1984; Levelt et al., 1991; Roelofs, 1992; Schriefers, Meyer & Levelt, 1990; Damian & Bowers, 2003; for review see: Abdel Rahman & Melinger, 2009; Piai et al., 2012) when a distractor word is either visually superimposed on the picture to be named or auditorily presented during stimulus (picture) presentation. The critical manipulation is that distractor words can either be semantically related to the target word or semantically unrelated. What is consistently shown is that semantically related distractors increase naming latency. That is, upon presentation of related distractors subjects are normally slower in naming than when an unrelated distractor is presented. This is because the speed of selecting a target item is proportional to the cumulative activation of all the lemmas that are competing for selection. Let us explain this with an example: when selecting the lemma ‘dog’, the lemma ‘cat’ is also activated for selection as a closely related conceptual item. If participants, when asked to name a picture of a dog, are presented with the word *cat* superimposed on that picture, the lemma ‘cat’ receives additional activation resulting in a longer naming latency. If, however, an unrelated word that is not part of the competitive selection process is presented over the picture, e.g., *hat*, it affects the response latency much less. These types of semantic interference effects are amongst the most cited evidence favouring the notion that lexical access is a competitive selection process (but see: e.g., Mahon et al., 2007; Dhooge & Hartsuiker, 2010).

When the target lemma is selected, it triggers the next step, form encoding, which comprises the retrieval of the morpho-phonological form of the selected lemma, structuring those speech sounds in the appropriate order (syllabification) and generating the appropriate articulatory (motor) commands. In contrast to lexical selection, during form encoding there is no competition since only a single lexical representation is selected for phonological and

phonetic processing (e.g., Levelt et al., 1991; the only exception being synonyms when the phonological codes for both synonymous words are activated simultaneously, e.g., Jescheniak & Schriefers, 1998). To show this empirically, PWI experiments also show that when distractor words are phonologically related to the target word (e.g., *cap* for the target *CAT*) the opposite effect is observed, that is, phonological facilitation, which results in shorter naming latencies (e.g., Lupker, 1982; Glaser & Dünghoff, 1984; Glaser & Glaser, 1989; Meyer & Schriefers, 1991; Cutting & Ferreira, 1999; Ferreira & Pashler, 2002; Jescheniak et al., 2003). The rationale here is thus the following: with semantic distractors we observe naming interference because lexical selection is competitive, but with phonological distractors we observe faster naming latencies as there is no competition anymore (that has been resolved at the lexical level) and what matters is the overlap in sounds between target and distractor (but see: e.g., Starreveld & La Heij, 1996; Cutting & Ferreira, 1999; Bowers & Damian, 2003; Bloem & La Heij, 2003; Navarrete & Costa, 2005). And even though the exact nature of semantic interference and phonological facilitation in picture naming remains a debated issue (e.g., Mahon et al., 2007; Roelofs & Piai, 2013; Navarrete et al., 2014; Abdel Rahman & Melinger, 2019; Mahon & Navarrete, 2019; Runnqvist et al., 2019), the context effects on lexical processing as assessed with the PWI paradigm represent one of the most used and important approaches to assess the nature of the mental lexicon (for recent reviews see: e.g., de Zubicaray & Piai, 2019; Nozari & Pinet, 2020), and is the corner stone paradigm that has informed the serial model of lexical access (e.g., Levelt et al., 1999; Meyer et al., 2019).

Furthermore, PWI experiments were also of importance to obtain initial chronometric evidence on the time course of lemma selection and form encoding. In their classic PWI experiment, Schriefers and colleagues (1990; see also Levelt et al., 1991) explored both semantic interference and phonological facilitation within the same study while manipulating

the stimulus-onset asynchrony (SOA). SOA (here) refers to the time between the presentation of the (auditory) distractor and the target picture to name: the distractor could either be presented prior to the target (e.g., -150 ms SOA), at the same time (0 ms SOA), or after target presentation (e.g., +150 ms SOA). The authors showed that the semantic interference effect was maximal when the distractor was presented before the picture target (-150 ms SOA), while the phonological facilitation effect was maximal when distractor and target were presented at the same time (0 ms SOA) or when the distractor was presented after the target (+150 ms SOA). From this result, it was concluded that lexical access and phonological encoding are two dissociable processing stages with a temporal delay between them of around 150 ms. This sequence of events was furthermore successfully simulated by Roelofs in a computational model (e.g., 1992; 1997), adding to the dominant view at that time for a temporal segregation of approximately 100-150 ms between the initiation of lemma selection and the start of form encoding (but see e.g., Alario et al., 2000; Bloem & La Heij, 2003, Costa et al., 2005; Strijkers & Costa, 2011).

Given the enormous impact of the serial/sequential model developed by Levelt et al. (1999), and its very precise functional and temporal predictions to go from a concept to the utterance of that concept, it offered an ideal blueprint to implement at the neural level and link its different, sequential processing stages to their respective spatial and temporal brain correlates. In order to do so, Indefrey and Levelt (2004) performed a meta-analysis of most available neuroscientific data on picture naming at that point to identify those brain regions and their respective time course of cortical activation involved in the different word production components of the Levelt et al. model (1999). First, to identify the brain regions that are reliably active during word production they looked at 82 production experiments and 26 perception experiments from neuroimaging localisation studies (see Table 2 in Indefrey & Levelt, 2004). By contrasting the patterns of brain activity found in all 108 experiments,

Indefrey and Levelt propose that the set of brain areas reliably found for both the picture naming and word generation tasks can be regarded as the core set of brain areas responsible for word production. These include 11 areas in the left hemisphere (posterior inferior frontal gyrus, ventral precentral gyrus, supplementary motor area, mid and posterior superior and middle temporal gyri, posterior temporal fusiform gyrus, anterior insula, thalamus, and medial cerebellum) and four in the right hemisphere (mid superior temporal gyrus, medial and lateral cerebellum, and the supplementary motor area). Next, the authors mapped these regions to the different word production components as proposed in the Levelt et al. (1999) model (see Figure 2): mid temporal regions linked to lexical access, superior temporal regions involved in phonological encoding, the inferior frontal gyrus for speech segmentation (syllabification), motor regions associated with phonetics and articulation, and finally the superior temporal cortex for speech monitoring (for more details on the neurobiological basis of speech monitoring see: Runqvist, 2022 in this handbook). The result of the meta-analysis and its association with the sequential model of word production thus suggest a functional specialisation of brain areas (see footnote 2): each brain area is responsible for a particular kind of linguistic computation (i.e., conceptual, lexical, phonological, articulatory) during word production. In this view the linguistic computations underpinning speech planning are neurally discrete and the different brain areas ‘communicate’ hierarchically: the output of one becomes the input of another. This hierarchical conceptualisation of word processing therefore copies the key property of the Levelt et al. (1999) model, namely the serial/sequential activation of brain regions linked to a specific word production component (see also later extensions: e.g., Indefrey, 2011; and also: WEAVER++/ARC model (Roelofs, 2014; 2018)).

Importantly, Indefrey and Levelt (2004) provided their model also with temporal estimates, hereby portraying not only the spatial components of word production, but also its

temporal dynamics. This is an important addition in comparison to all previous brain language models of language production which solely focused on where in the brain language production processes may come about, but not when. In order to do so, Indefrey and Levelt adopted the following strategy: First, they estimated time-windows when each word production process would happen in the course of speech planning, and next they compared

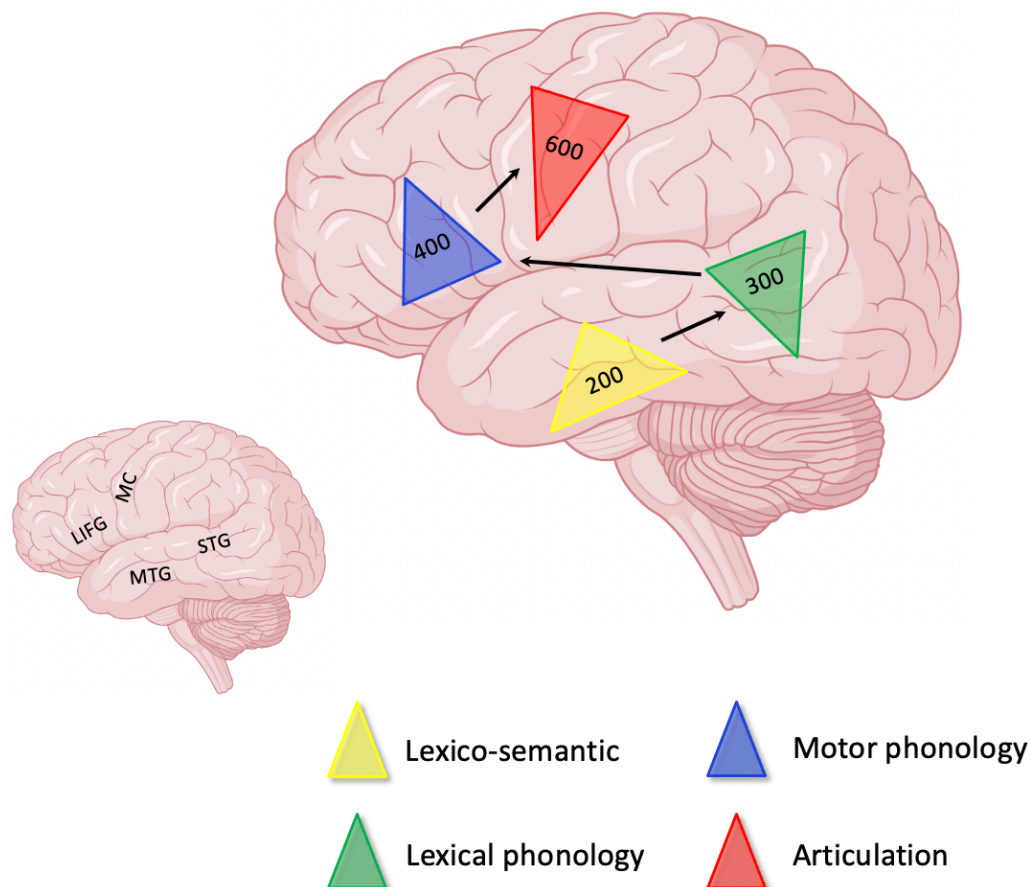


Figure 2. Schematic representation in time and space of the sequential brain language model of word production. See text for explanations (based on Indefrey & Levelt, 2004; and adapted from Strijkers & Costa, 2016).

those temporal estimates with the results of magnetoencephalography (MEG) studies. With regard to the first step, the authors relied on ERP data of object categorization studies (e.g., Thorpe et al., 1996; Hauk et al., 2007; Johnson & Olshausen; 2005; Schmitt, Münte & Kutas, 2000) to estimate that conceptual processing (of an image) takes about 150-200 ms. With that

as starting point, they subsequently added the chronometric estimates as assessed with the PWI studies mentioned above and their computational simulations (e.g., Roelofs, 1992; 1997). The following temporal picture emerged for word production (assuming an average naming latency of 600 ms): 0-175 ms = conceptual processing; 175 – 250 ms = lexical selection; 250 – 330 ms = phonological encoding; 330 – 455 ms = syllabification; 455 – 600 ms = articulatory preparation. Armed with these time-windows associated with different word production components, in the second step, they looked at MEG studies of object naming. The big advantage of MEG is that one obtains both temporal and spatial data. In this manner, Indefrey and Levelt (2004) explored the above-mentioned time-windows in three MEG studies of language production (Levelt et al., 1998; Maess et al., 2002; Salmelin et al., 1994), and checked which brain regions that were maximally activated in that specific functional time-window corresponded to one of the brain regions defined in their meta-analysis. This resulted in the following spatiotemporal map of word production (see Figure 2): Lexico-semantic processing taking place around 200 ms after picture presentation in the MTG. Word form (lexical phonology) encoding of the selected lemma manifesting around 300 ms after picture presentation in the posterior STG. Syllabification (motor phonology) emerging after 400 ms in the IFG. And finally, activating the motor commands necessary to move the articulators taking place around 600 ms after stimulus onset in the pre- and post-central gyri.

In summary, the sequential model of lexical access in word production proposes that lexical access is a hierarchically organised process which begins with a concept and moves step-by step in a feedforward manner through linguistic levels of representation until a motor command is carried out to articulate the utterance. The core of this model postulates that the two parts of the lexical access system, i.e., lexical selection and form encoding, serving two different functions (selecting a target lexical item, and producing an articulatory score for that

selected item, respectively), are segregated in time and space. That is, according to the model, information spreads in a feedforward manner, where each step is neurally realised at a well-defined time window by a distinct brain area, which specialises in one linguistic function only. This model is considered to be a detailed mapping of the neural dynamics supporting word production and to be a valuable tool and theoretical basis for producing precise predictions and hypotheses and testing them (e.g., Aristei et al., 2011; Christoffels et al., 2007; Habets et al., 2008; Koester & Schiller, 2008; Hulten et al., 2009; Hanulova et al., 2011; Dell'Acqua et al., 2010; Laganaro et al., 2009; 2012; Sahin et al., 2009; Piai et al., 2014; Fargier & Laganaro, 2017). Other researchers, however, have questioned the model's strictly sequential and highly localised properties (e.g., Strijkers & Costa, 2016; Munding, Dubarry & Alario, 2016), and alternative models of lexical access have been proposed; some of which we will discuss below. Nevertheless, and in spite that certain properties of the sequential model have led to much debate, most of the alternatives are directly built on Levelt's model and still have many aspects in common. The latter is remarkable in itself, especially when taken into account that there is no model in language production that has been as extensively tested as the sequential model of lexical access.

2. Interactive models.

The next model we will discuss is the one proposed by Dell in 1986 (see also: Dell & O'Seaghdha, 1992; Dell et al., 1997; Dell et al., 2007; Dell et al., 2013), which introduced the notion of interactivity in lexical access (see also: e.g., Dell & Reich, 1977; Harley, 1984), and inspired many following models and much research in psycho- and neurolinguistics (e.g., Caramazza, 1997; Rapp & Goldrick, 2000; Vigliocco & Hartsuiker, 2002; Pickering & Garrod, 2004; 2013; Harley, 2008; Strijkers & Costa, 2011; Ueno et al., 2011; Strijkers, 2016; Walker & Hickok, 2016; Abdel Rahman & Melinger, 2019; Roelofs & Ferreira, 2019;

Nozari & Pinet, 2020). The central characteristic of this model is that while just like sequential models it assumes linguistic representations are organized hierarchically in functionally dissociable processing layers, it allows for interactivity between those levels, meaning that information can flow bidirectionally between processing layers (see Figure 3). In this manner, and compared to the serial view on lexical access as discussed before, an interactive account of lexical access is somewhat more flexible in that there is some degree of processing overlap in both time and space. Below we will first detail the computational principles of Dell's interactive lexical model further and subsequently give an overview of its proposed neural implementation.

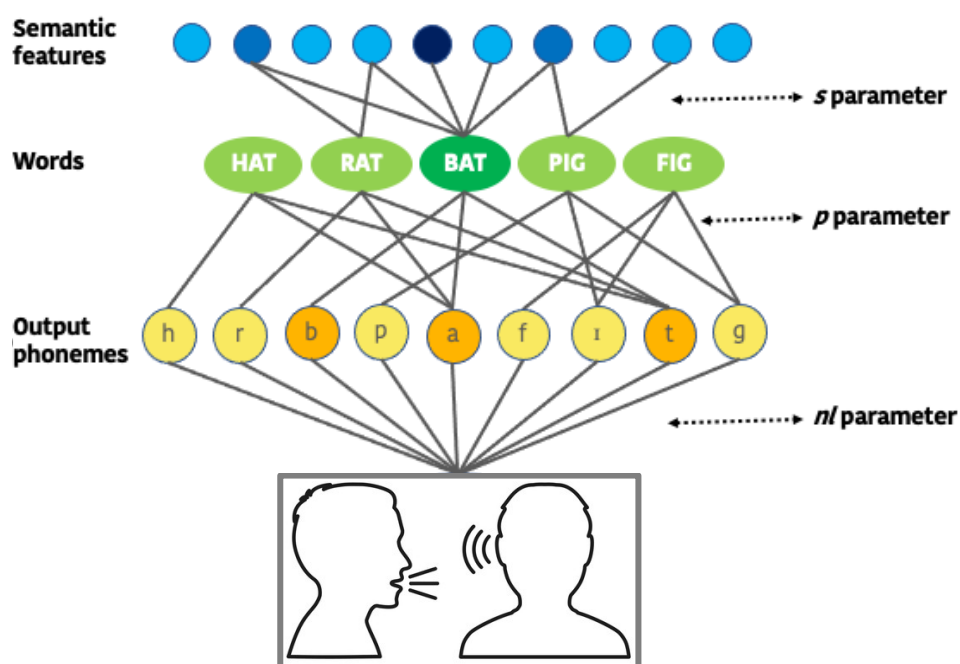


Figure 3: Schematic illustration of the two-step interactive activation model of Dell. The two steps of the model during word production are from the semantic (features) to the lexical level (words), and from the lexical to the phonological level (output phonemes). Weights refer to the parameters s , p , and nl in the computational model that represent the connections between the linguistic levels (see main text). Adapted from Dell et al. (2013).

Interactive lexical access in the Dell-model is, as in the serial model (Levelt et al., 1999; but see Caramazza, 1997), regarded as a two-step process where activation flows through the lexical system in the following manner: The first step is initialised by a jolt of activation at the conceptual level ("Semantic features" in Fig. 3), continues as spreading activation through the network and activates a number of relevant units at the lexico-syntactic level ("Words"). Units at this level are called lemmas and are abstract symbols that unify the semantic-syntactic representation of a word. The first step is finalised when the lemma with the highest level of activation is selected, regulated via an 'insertion' rule that checks whether the lemma is semantically and syntactically suitable (e.g. "swim" if a verb is required and "swimming" if a noun), and activation of other lemmas is inhibited. The second step of word production begins with a jolt of activation spreading from the lemma to the phonological level ("Output phonemes"), activating the phonological units corresponding to the selected lemma representation. At this stage, beyond for example differences in compositionality (e.g., Roelofs, 1997b), the previously discussed serial model and Dell's interactive activation model are quite similar in their dynamics of lexical access. Crucially, however, interactive models propose that information travels bi-directionally in the lexical network. That is, while in the serial model (Levelt et al., 1999) lexical access is entirely feedforward, in Dell's model (1986) activation can flow back between the representational layers (see Figure 3). This means that even if linguistic levels are globally modular and initially become activated in a sequential manner, they are locally interactive, displaying temporal overlap in the activation of lexical representations: when level x is most active, there will be some activity too in levels $x+1$ and $x-1$ (Dell & O'Seaghdha, 1992). For example, spreading activation from the concept BAT will activate the lemmas "bat" and "pig" because they are categorically related. Then, spreading activation from the lemma "bat" will activate the phonological units "b", "a" and "t", and those phonological units will in their turn send activation back to those lemmas

associated with the phonological features, hereby ‘boosting’ the activation of the target lemma “bat” and reducing the activation linked to the semantically related word “pig” (see Figure 3). Or put differently, within such interactive framework, the intended lexical entry for speech receives two sources of activation: the bottom-up driven feedforward activation from the concept a speaker wishes to utter and the feedback driven activation from the phonological level of processing allowing to “check” that the activated sounds indeed correspond to the word we wish to convey³.

Importantly, beyond offering a mechanism on how our lexical system can avoid misselecting words, the interaction between levels allows for explaining common speech errors that people make. In fact, the original model of Dell (1986) was constructed exactly for that purpose. Indeed, while overall, we make surprisingly few speech errors when speaking, when we do make an error it seldomly is random, but instead follows linguistic constraints (e.g., Fromkin, 1973; Garret, 1975; Dell, 1986; Rapp & Goldrick, 2000). In this manner, speech errors have been an important source of information to understand the architecture underpinning language production (for more recent overviews: e.g., Goldrick, 2011; Dell et al., 2014; Runnqvist, 2022 this handbook). At the level of lexical access, three types of errors have been particularly informative: semantic, phonological and mixed errors. Semantic errors refer to saying for example “pig” instead of the intended “bat” and are explained in the model because of the spreading activation from related concepts (i.e., other animals) (see Figure 3). Phonological errors refer to saying for example “hat” instead of “bat” and come about in the second phase of activation in the model, namely when the selected lexical item activates overlapping phonological units. Finally, and most importantly here, the mixed error effect refers to errors which are both semantically and phonologically related, like saying “rat”

³ While the Dell and colleagues’ model does not include inhibitory connections, other interactive models do, suggesting this a biologically plausible mechanism for limiting interaction and cascading within the system. For more detail look at Harley (1993) and Schade and Berg (1992).

instead of “bat”. Furthermore, these mixed errors occur more frequently than predicted in light of semantic or phonological errors in isolation (e.g., Dell & Reich, 1981; Harley, 1984; Martin et al., 1997; Goldrick & Rapp, 2002; Ferreira & Griffin, 2003); an observation which is assumed to be a consequence of interactivity. That is, keeping with the above example, it is more likely that the target "bat" will be substituted by "rat" than by "pig", because “rat” will receive activation from two sources, namely the spreading activation (from the target concept “bat”) from the semantic layer and the feedback activation (interactivity) from the phonological layer (because it has two phonological units that overlap with the phonological units of “bat”), while “pig” will only receive activation from one source, namely the spreading activation from the semantic layer (see Figure 3). In other words, due to the interactivity, mixed errors are readily explained in Dell’s model (1986; Dell et al., 2013). In contrast, in discrete and serial models like that of Levelt et al. (1999), such mixed error effects are particularly difficult to explain, since the holistic word forms do not affect the lexical representations, and thus mixed errors should not happen.

Another important behavioral observation that has typically been cited to support interactivity concerns the lexical bias effect (Baars et al., 1975; Dell & Reich, 1981; Humphreys, 2002; Hartsuiker, et al., 2005; 2006; Nooteboom, 2005; Nozari & Dell, 2009; Runnqvist et al., 2016; 2021). The lexical bias effect is the tendency of phonological errors to also be real words, for example people are more likely to substitute "bat" with "hat" than with "lat". Interactive models explain this because phonological units follow phonological rules in a language, they will favour the grouping together of legal over illegal phonological activations. Since in an interactive model activated phonological units feedback to the lexical system, it logically follows that a speech error will more likely be a word than a non-word for the simple fact that words are represented in our mental lexicon, while non-words are not. Hence, when making a phonological speech error like a phoneme substitution (e.g., changing

the first phoneme of a word), within an interactive system it is predicted that the substitution will more likely result in an actual word (“hat” instead of “lat”), because it receives activation from the lexical layer, while a nonword does not. Similarly as for the mixed error effect, in a serial model (Levelt et al., 1999), this is not necessarily predicted (but see: e.g., Roelofs, 2004).

Finally, behavioral evidence for the interactive nature of the language production system also comes from aphasia studies (e.g., Nozari et al., 2010). For example, in a study by Jefferies et al. (2006) aphasic patients were asked to repeat words while the processing load on the phonological or semantic system was manipulated. The results demonstrated that the semantic manipulation effect was larger in the phonologically straining than the phonologically undemanding task for the phonetically impaired group. The phonetic manipulation effect however was larger in the semantically straining than the semantically undemanding task for the semantically impaired group. This suggests that the phonological system plays a more important role in repetition when the semantic system is impaired and vice versa, providing evidence for interactivity between semantics and phonology in brain damaged speakers.

In terms of neural implementation, the notion of interactivity can fit different neuroanatomical architectures (e.g., Indefrey, 2011; Strijkers & Costa, 2011; Hickok, 2012). For example, it could be consistent with the previously described Indefrey and Levelt (2004) model if some more temporal and spatial flexibility between the different brain regions (linked to distinct functional word components) is allowed. However, while the Indefrey and Levelt (2004) model suggest a single feedforward processing pathway, going from mid temporal via superior temporal towards frontal brain regions, Dell and colleagues have linked their specific interactive model to a dual-route architecture (Dell et al., 2013; see also Nozari et al., 2010; Ueno et al., 2011), hereby integrating established neuroanatomical ideas from the

perception literature (e.g., Hickok & Poeppel, 2000; 2004; 2007; Scott & Johnsrude, 2003; Rauschecker & Scott, 2009). In this manner, Dell et al. (2013) suggest that lexical access in an interactive model is realized by two partially distinct brain networks, a more ventral stream and a more dorsal stream (roughly corresponding to the red and blue patches in Figure 4). Previous research implicates the dorsal stream in phonological processing as there is evidence that phonological form retrieval takes place in the pSTG (e.g. de Zubicaray et al., 2002; Graves et al., 2008; Edwards et al., 2010), phonological short-term memory is processed by temporo-parietal and inferior parietal regions (e.g. Buschsbaum et al., 2011), and phonologically-related errors stem from the dorsal pathway (e.g. Cloutman et al., 2009; Duffau et al., 2008; Schwartz et al., 2012). The ventral stream, on the contrary, has been implicated in more lexical-semantic processing as evidence suggests that lemma processing takes place in the MTG and pITG (e.g. Damasio et al., 1996; Graves et al., 2007; de Zubicaray & McMahon, 2009; de Zubicaray et al., 2015; Ries et al. 2017), objects and events are represented in the ATL and AG (e.g. Binder & Desai, 2011), and damage to the ventral stream, including the MTG, ITG, ATL, AG and IFG results in semantically-based word retrieval difficulties (e.g. Antonucci et al., 2008; DeLeon et al., 2007; Schnur et al., 2009).

Dell and colleagues (2013) further confirmed and extended this two-step neural architecture to brain damaged aphasic speakers via voxel-based lesion parameter mapping. The study used the interactive activation computational model to simulate the individual error pattern of 103 aphasic patients in three different tasks: word production, word repetition, and non-word repetition. Parameters corresponding to the connections between linguistic levels were adjusted so that the model can account for the variability in errors for all three tasks. The lesion status of a voxel was then used to predict the three critical parameters: s representing the ability to map from the conceptual to the lexical level, p the mapping from the lexical to the phonological level, and nl the mapping from auditory input to the

phonological level (see Figures 3 and 4). While the p and nl parameters shared a significant proportion of their brain maps, for the s and p parameters overlap was much less pronounced

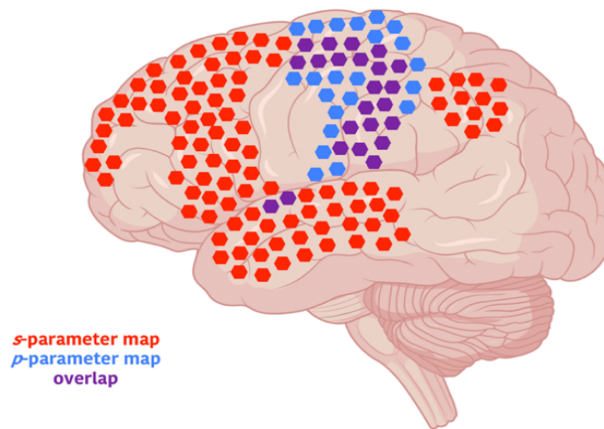


Figure 4: The maps of two of the parameters from the interactive computational model suggest that semantic and lexical processing on the one hand and phonological and articulatory processing on the other have only a few brain areas in common (based on Dell et al., 2013).

(see Figure 4), suggesting that the brain areas responsible for realizing the lexical and phonological levels of word production are at least in part neurally distinct.

Based on these lesion mapping results (as well as the neuroimaging in healthy speakers) Dell and colleagues (2013) proposed the following neural implementation of the interactive two-step model: The first step, semantic encoding and lemma access, is realised by a rather large left-lateralised network, including the aSTG, aMTG, temporal pole, MFG, IFG, TPJ, and AG (red and purple locations in Figure 4). The second step, the access of the phonological form of a word, is implemented more dorsally and more posteriorly (blue and purple locations in Figure 4), including regions such as the STG, TPJ, the planum temporale and the pre- and post-central gyri. In other words, while the neural implementation of the two-step interactive activation model by Dell et al. (2013) shares with the Indefrey and Levelt model (2004) the notion that lexico-semantic and phonological processing are achieved by largely dissociable neural circuits, these circuits themselves are much less localized compared

to the Indefrey and Levelt model, recruiting an extensive network of brain regions in frontal, temporal and parietal cortex both for the lexical and phonological layers of processing.

Regarding the time course of lexical access, as mentioned earlier, interactive models assume that the first pass activation of each representation proceeds sequentially, and thus propose a temporal dynamic which initially can mimic the time course proposed by the Indefrey and Levelt model (2004): lexical processing around 175ms and phonological form encoding some 100ms later around 250-300ms (Dell & O'Seaghdha, 1991; 1992). But in contrast to the serial model, more than one layer can be active at same time, given that when the interactivity (feedback) kicks in, (at the least) adjacent processing layers should display an overlapping time-course. Specifically, for the implementation of interactive lexical access in a dual-route neuroanatomical structure as envisioned by Dell and colleagues (2013), speech planning would thus first trigger ventral brain regions linked to lexico-semantic knowledge (see Figure 4), some 100ms later more dorsal brain regions linked to the phonological knowledge associated with the intended word for speech (see Figure 4), and subsequently feed the dorsal activity back to the ventral brain regions, allowing for the interactivity between the lexico-semantic and phonological representations a speaker is about to utter.

To summarize, lexical access in an interactive model (e.g., Dell, 1986; Dell et al., 2013) has some very similar assumptions to lexical access in a serial/sequential model (e.g., Levelt et al., 1999; Indefrey & Levelt, 2004), because it considers lexical access to involve sequential steps of processing which work with different kinds of representations, stored in different brain areas. Specifically, in this view, words in the brain have lexical and phonological representations, which are independent of each other and activation flows from the concept through the lemma and onto the phonological form in functionally distinct time steps. In contrast to the serial/sequential model of lexical access, however, these lexical and

phonological layers are “globally modular, but locally interactive” (cf. Dell & O’Seaghdha, 1992), in that they are much more distributed in our brain, and can have a functional and temporal influence upon each other through feedback interactivity.

3. Dual-stream Feedback Models

The two previous sections were devoted to the traditional models of lexical access in speech production which have been of enormous impact on the neurocognitive research in language production. Thanks to their pioneering role, the field has advanced much, in particular with regard to the integration of the cognitive mechanisms of word production at the level of the brain, allowing the development of novel brain language models on lexical access. In the final two sections of the present chapter we will describe two such novel brain language models, which lend many insights from the traditional serial and interactive models, but also extend and/or differ from them in important ways.

The first model that we are going to discuss, the hierarchical state feedback control model (HSFC), has been developed by Hickok (2012; see also Hickok, 2014; Walker & Hickok, 2016) and is part of a larger integrated theory on speech processing, the dual-route

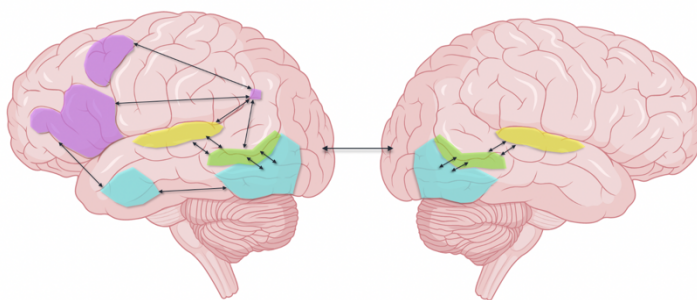


Figure 5. The dual-stream model of speech processing. Bilateral auditory regions in the dorsal STG (yellow) and STS (green) are engaged in the early stages of speech processing, which later diverges into two streams: the ventral stream responsible for speech comprehension (blue) and a left dominant dorsal stream involved in speech production (purple). (based on and adapted from Hickok & Poeppel, 2007).

model (see Figure 5; Hickok &

Poeppel, 2000; 2004; 2007).

The dual-route architecture, which we already mentioned in the previous section with

regard to a specific neural

implementation of Gary Dell’s

(1986) interactive activation

model of lexical access, is

borrowed from research in the visual domain where this processing mechanism, namely, the division of labour between the ventral and dorsal streams, has been well established and demonstrated empirically (e.g., Milner & Goodale, 1993; Ungerleider & Mishkin, 1982). The dual-route model of speech processing proposes two pathways, or streams, that underlie the neuroanatomy of language processing: the ventral stream, concentrated in the superior and middle parts of the temporal lobe, is responsible for the comprehension of speech, and the dorsal stream, involving structures in the posterior planum temporale and posterior frontal lobe, is engaged in speech production through sensory-motor integration (more on this below) (see Figure 5). While originally the dual-route architecture was developed to explain language comprehension (Hickok & Poeppel, 2000; 2004; 2007), here we will focus on the recent extension to language production (Hickok, 2012; 2014). This model serves as a neat bridge with the prior two sections in this chapter, since it borrows many properties of the sequential and interactive models discussed before but embeds it within the neuroanatomical principle of dual-stream processing and adds the concepts of feedback control and predictive processing (see for similar psycholinguistic feedback/predictive models of language production e.g., Pickering & Garrod, 2013; Dell & Chang, 2014).

The HSFC model attempts to bridge two traditions of speech processing research – psycholinguistics on the one hand (e.g., Dell, 1986; Levelt et al., 1999), and motor control theories on the other (e.g., Gracco & Lofqvist, 1994; Guenther, 2006; Golfinopoulos et al., 2010; Tourville & Guenther, 2011; Perkell, 2012). Drawing from the previous psycholinguistic theories (e.g., Dell, 1986; Levelt et al., 1999), the model agrees that the main stages of word production comprise conceptual and lexical selection, phonological encoding and articulatory preparation. These stages are considered to be implemented in a hierarchical manner where lexico-semantic processes give rise to phonological processes (Hickok, 2012; 2014; 2019). Neurally, this hierarchy is organized in the following regions: lexical selection

occurs in the middle temporal regions and phonological processing takes place in the posterior STG (word forms), inferior parietal and inferior frontal and premotor cortex (motor phonology) (see Figure 6).

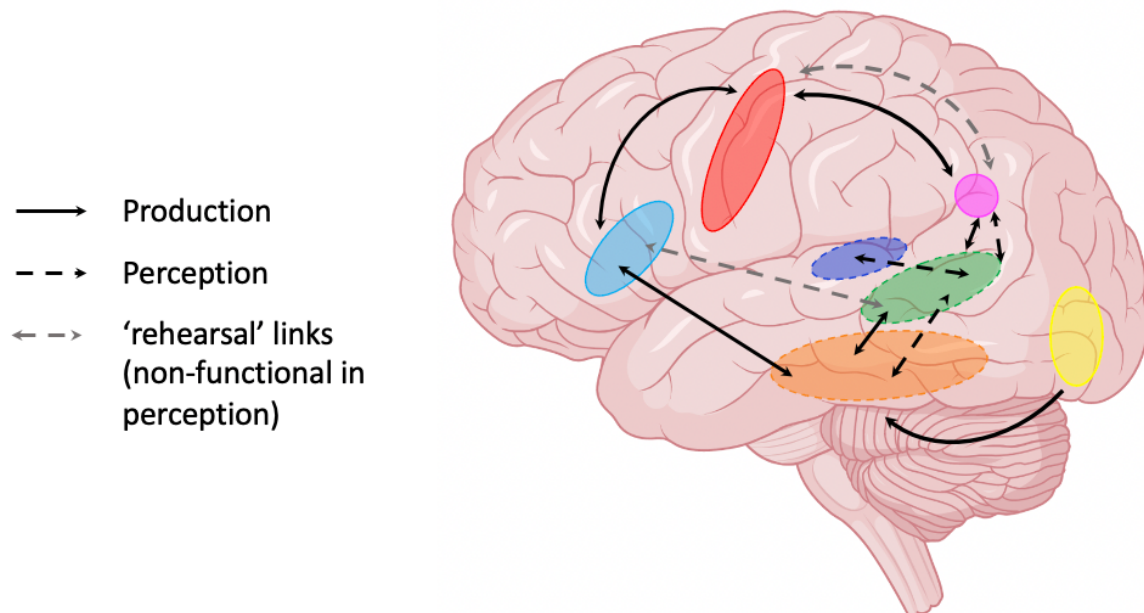


Figure 6. Simplified schematic representation of Hickok's dual-stream HSFC model, where lexical representations (lemmas) are triggered (for example from a visual information -yellow area- we wish to utter) in the (posterior) MTG (orange) and activate two streams of processing: (1) towards inferior frontal regions (blue) in order to retrieve the motor syllable representations linked to the intended lexical representation for speech, and thereafter towards the motor cortex (red) to activate the motor phoneme programs for articulation; at the same time (2) a more dorsal stream is activated from the lexical level towards superior temporal regions (green and dark blue) to activate auditory syllable predictions (and thereafter towards supramarginal gyrus to activate somatosensory phoneme predictions). Crucially, the auditory syllable/phoneme targets are predictions of what the sensory outcome (i.e., efference copy) of the articulation of the syllables/phonemes would sound like. These syllable representations in temporal brain regions are then used to check the syllable representations in frontal brain regions (via area Spt and the cerebellum): when they overlap, articulation is triggered, when they don't overlap, reprocessing is required to correct speech planning towards the intended syllable representations. In this manner, the HSFC thus integrates prediction (efference copies) and feedback control (checking mechanism between motor and auditory representations) within its hierarchical neuroanatomical dual-stream architecture. Finally, note that the time-course estimates are purely illustrative (the HSFC does not make explicit temporal predictions) in that the model agrees with a sequential progression of activation from higher towards lower levels of processing as in interactive or serial psycholinguistic models (Figure roughly based and adopted from Hickok, 2012; 2014).

The main idea adopted from the motor control theory is the notion of feedback control which aids in achieving an action with minimal errors between a planned action and the sensorimotor output (e.g., Fairbanks, 1954; Shadmehr & Krakauer, 2008; Tian & Poeppel, 2010). Within motor control theories, it has been suggested that a smooth implementation of the feedback control system is solved through an internal model of the body. In essence, this means that the system builds a predictive model that controls the motor commands sent to the effector and the most recent state of that effector. This control mechanism is implemented via an efference copy of a given motor command being sent to the internal model. This allows the brain to detect and correct errors almost immediately through predicting the consequences of a motor command before its actual implementation, that is, before feedback occurs (Shadmehr et al., 2010; Shadmehr & Krakauer, 2008; Wolpert et al., 1995). The idea of the internal predictive model attempts to explain why we generally have little trouble and make very few to no errors in our everyday actions including speech. Moreover, the cortical regions that have been shown to be engaged in motor control overlap with those that are proposed to be actively involved in phonological processes in psycholinguistic models of speech production. These regions include IFG, somatosensory-motor cortex, premotor cortex, posterior STG and temporal-parietal junction (Guenther, 2006; Houde & Nagarajan, 2011) (see Figure 6).

According to HSFC, motor control is implemented at the phonological level of processing which comprises two main components - a motor-phonological (output) and an auditory-phonological (input) component. This division is an essential difference compared to prior psycholinguistic models and their neural implementations we discussed above. Since the model ascertains that internal feedback is auditory (Burnett et al., 1998; Houde & Jordan, 1998; Stuart et al., 2002), the split of the phonological level into two is justified via the logic of feedback control: the motor system executes an act (in our case, a speech act), which is

evaluated against previously formed predictions about the sensory consequences of that act; in case of an error, the difference between the prediction and the resulted output is assessed for error correction. This, according to the HSFC model, dictates the necessity for the dual organization of the phonological level – the motor component responsible for execution of a speech act and the auditory component that is involved in the prediction and monitoring of that act (see Figure 6).

An important body of evidence in favour of this view concerns conduction aphasia. This language impairment traditionally manifests in poor phonemic planning, that is, patients normally produce fluent speech and have preserved speech comprehension, but they often make phonemic errors, or paraphasias, which they sometimes are conscious of but find hard to correct. It has been difficult to pinpoint the exact link between these symptoms and prior language production models since none of them seem to find a coherent explanation (for a more detailed review see Baldo et al., 2008; Buchsbaum et al., 2011; Hickok, 2014; 2019), and this is where the dissociation between the motor and auditory phonological systems comes into play. If such a dissociation is assumed, then the symptoms can be explained by a disrupted connection between the auditory (sensory) and motor systems: both the motor and sensory systems are not impaired, hence, speech fluency and comprehension are intact but due to the disconnection between the two systems, the sensory system no longer affects the motor one, which results in frequent paraphasias. Moreover, lesion location for conduction aphasia corresponds to the area Spt, a region in the posterior Sylvian fissure at the parietal-temporal junction (Buchsbaum, et al., 2001; Hickok, et al., 2009), which is hypothesized to be the mediating interface between the auditory and motor phonological systems.

Other data, from healthy speakers, that finds an elegant explanation in the HSFC model concerns neuroimaging data related to (1) auditory suppression and to (2) speech monitoring. Concerning (1), auditory suppression refers to the phenomenon that the auditory cortex's

response to one's own speech is reduced compared to the speech of others (e.g., Houde et al., 2002). The fact that this phenomenon seems to occur ultra-rapidly (already within 100 ms of processing; e.g., Heinks-Maldonado et al., 2006) fits well with the notion of forward prediction as implemented in Hickok's HSFC model: That is, because one can predict one's own speech better than that of others, the neural response in the auditory cortex is attenuated since there is a good match between the action (motor) and its consequence (auditory) and no further reprocessing is necessary. Concerning (2), and recent neuroimaging finding that goes well with the predictions of the HSFC, are studies where the speech monitoring system is investigated by using a slip task (e.g., Runnqvist et al., 2016; 2021). In a slip-task (e.g., Motley et al., 1982) researchers attempt to experimentally induce speech errors by priming people to make phoneme substitutions (e.g., barn door -> darn boor). Runnqvist and colleagues investigate this phenomenon with TMS and fMRI for correct trials (that is, trials where the participant didn't make the error -the phoneme substitution- but the likelihood of making such error was high; put differently, conditions that tax the speech monitoring system). The authors observed the involvement of the right cerebellum and superior temporal brain regions (amongst a relevant set of other brain regions: e.g., Runnqvist et al., 2016; 2021). Given the cerebellum's role in forward modelling (e.g., Blakemore et al., 2001; Ito, 2008), the link of right cerebellar activity and left superior temporal regions to monitor phoneme substitutions fits neatly with the predictions of the HSFC (Hickok, 2012; 2014).

In summary, lexical access in the HSFC shares some key elements with prior serial and interactive models of speech production, but also adds some additional features that predict different spatial and temporal dynamics underpinning the mental lexicon. With regard to the similarities, one of the key elements of the HSFC model is that it assumes a hierarchical and sequential, interactive relationship between different stages of processing (i.e., lexico-semantic and phonological steps) (see Figure 6). Furthermore, the model shares with the

Indefrey and Levelt model (2004) the notion of functionally specialized processing regions in the brain, with a clear division of labor between temporal and frontal brain regions (note that this aspect is different from the Dell et al. model (2013; see also Ueno et al., 2011), where there was also functional segregation of brain regions, but not necessarily in function of temporal versus frontal brain structures). It also shares with the sequential and interactive models of lexical access the hierarchical structure and by consequence functionally distinct time-course of activation for different word components (see Figure 6). However, by incorporating a dual-stream architecture and predictive feedback control, some key differences with sequential and interactive models become apparent as well: For one, in terms of temporal dynamics, after the activation of words at the lemma level (mid temporal brain regions) a (more or less) parallel activation time-course for frontal motor syllables and temporal auditory syllables is predicted. Second, at the functional spatial level, given the important role predictive feedback control play in the model, inferior frontal and superior temporal brain regions engage in a novel functional dynamic, namely one where their cross-talk serves a checking mechanism for speech production. In this manner, the HSFC model has a more task-dependent structure where a brain region's functional role can dynamically shift in function of task (and become asymmetrical, for example, between production and perception). Another attractive and novel feature of the HSFC model is that it suggests a generalized neuroanatomy across the language modalities, namely that of a dorsal-ventral organizational structure.

4. The Parallel Assembly Model

The last model we will describe proposes the notion of fast parallel processing where lexico-semantic and phonological-articulatory stages occur in the same temporal windows and simultaneously integrate distributed cortical activations. This model differs from

previous models in that it does not assume a hierarchical structure underpinning the different linguistic components making up a word, but rather the full integration of all word components into a single functional whole; the word as the neural Gestalt of language in the brain. Another interesting difference this model has compared to the previous ones is that its basis is inspired by system neuroscience theory rather than psycholinguistics and neuropsychology. That is, while the previous models were mainly guided by reaction time data (Levelt et al., 1999), speech error patterns (Dell, 1986; Dell et al., 2013) or patient data (Hickok, 2012; 2014), the parallel assembly model took the reverse approach and is mainly driven by neurophysiological processing principles. More concretely, the driving principle behind the model is Hebbian-based learning, which in broad terms states that ‘what fires together wires together’ (Hebb, 1949), and which Friedemann Pulvermüller used to develop a model of how our brain would represent words (Pulvermüller, 1999; and later: Pulvermüller, 2001; 2005; 2018; Pulvermüller & Fadiga, 2010). Recently this Hebbian-based assembly model of language was then adopted to the issue of lexical access in speech production (Strijkers, 2016; Strijkers & Costa, 2016).

In short, Hebbian-based learning (or assembly coding) means that those neural representations that are active at the same time will bind together into a single functional unit (e.g., Hebb, 1949; Braitenberg, 1978; Singer & Gray, 1995; Fries, 2005; Buzsaki, 2010; Singer, 2013). Or, put differently, if for a given event X there is temporal correlation (or coherence) between neural population A and neural population B, and this temporal correlation occurs often, then for that specific event X those two neural populations A and B will bind together in a novel, overarching neural assembly C that can reflect event X as a whole. Translating Hebb’s postulate to words, the idea is that since the meaning and sounds of a given word always co-occur (e.g., when speaking about a ‘ball’, the semantic features and sound features of ‘ball’ will always be active at the same time), through Hebbian-like

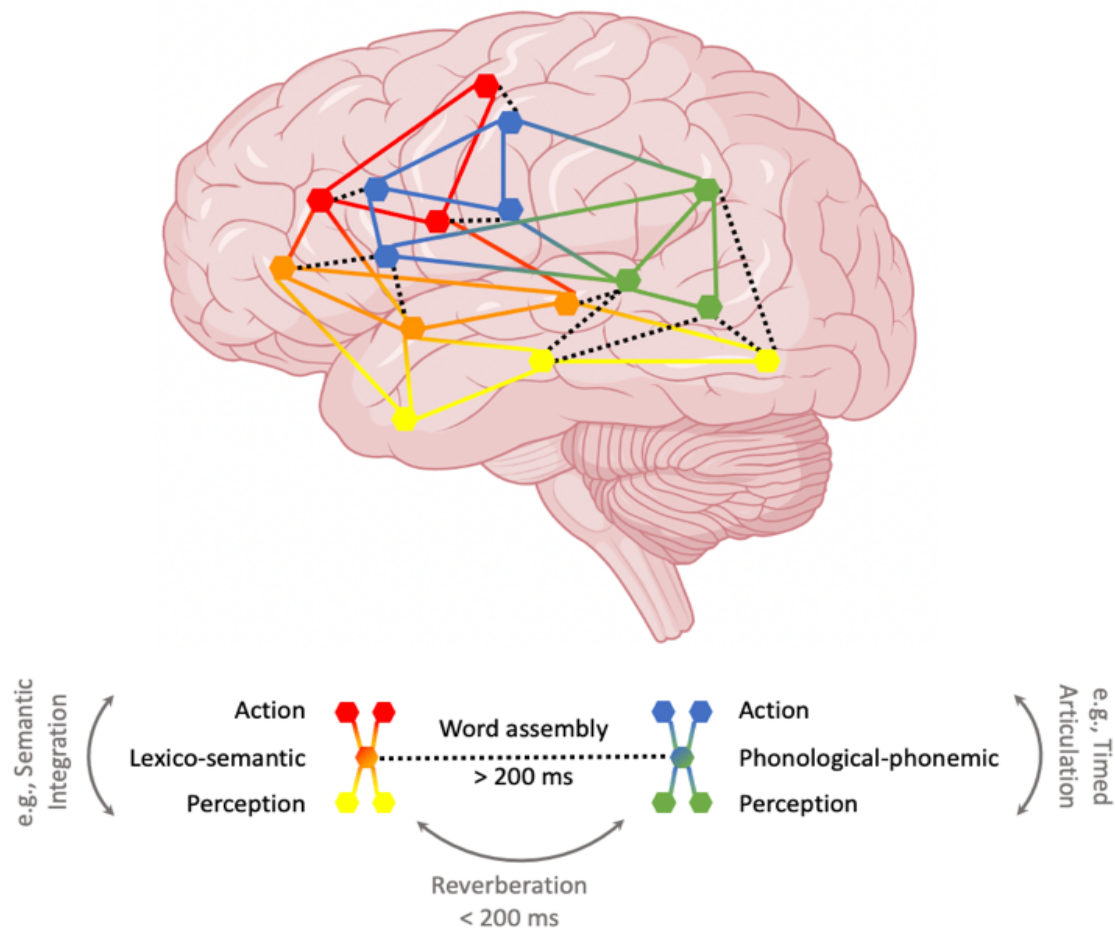


Figure 7. A schematic visualisation of the parallel assembly model. A widely distributed lexico-semantic network embedded in action (red) – perception (yellow) circuits and a widely distributed phonological-phonemic network embedded action (blue) – perception (green) circuits form a word assembly which ignites as a whole within the first 200 ms of processing. After ignition activity may remain active in the whole word assembly or reverberate in specific parts of the assembly to generate well-timed (sequential) spatiotemporal dynamics (adapted from Strijkers & Costa, 2016)

learning they form a single functional unit capable to reflect a word in its totality: a word assembly (e.g., Pulvermüller, 1999; 2005; 2018; Pulvermüller & Fadiga, 2010). Note that the starting point is not so different from the previously discussed models, namely a system dedicated to meaning-related processes and a to sound-related processes that are (partially) separated and independent from one another. The difference emerges during development where, according to Pulvermüller’s Hebbian model, the key to word learning lies in binding together meaning and sounds that form a coherent word. Viewing words in this way results in a substantially different spatiotemporal dynamic of lexical access when adopted to word

production compared to the previously discussed serial and interactive theories (see Figure 7) (Strijkers, 2016; Strijkers & Costa, 2016): (a) the time course of lexical access would be parallel instead of sequential; (b) the spatial recruitment would involve distributed networks both for lexico-semantic and phonological knowledge, instead of localized brain regions or processing streams for the lexico-semantic (lemmas) representations on the one hand and phonological representations on the other (see Figure 7).

First, we will look at the temporal dynamics and the main differences between the parallel assembly and sequential or interactive models. Traditional language production theories suggest that the firing of the lower-level cells (input) spreads to the firing of the higher-level cells (output), thus constructing a hierarchy of activations that occur in a sequential manner, i.e., segregated in time with differences between lexico-semantic and phonological activation often estimated in the range of 100 ms (e.g., Dell & O'Sheagha, 1992; Indefrey & Levelt, 2004; Indefrey, 2011). Note, as mentioned above for models incorporating interactive (and the same holds for cascading) properties (e.g., Hickok, 2012; 2014; Dell et al., 2013), that some overlapping (and in that sense “parallel”) activation is predicted in such models. However, and crucially, this overlapping activation occurs at a later point in time after initial sequential activation (for example, in an interactive model, first there is sequential activation of the hierarchical layers before they can start interacting with each other and overlap occurs). This is different from the parallel dynamics envisioned by Hebbian-based assembly models where the parallel activation of the meaning and sounds of a word occurs immediately during the first pass activation. Indeed, given that in a parallel assembly model words are reflected as integrated functional units, neuronal firing triggers the near-synchronous ignition of a word as a whole, because lexico-semantic and phonology are bound together in a single parallel distributed processing network.

While there is quite convincing data (some of which presented in the sections above) that speech production involves a sequential component (for an overview see: e.g., Indefrey & Levelt, 2004; Indefrey, 2011; 2016), whether sequential dynamics is enough has been questioned (e.g., Strijkers & Costa, 2011; 2016), and many of the data used to assess the temporal dynamics did either not rely on immediate and overt speech production (e.g., Strijkers & Costa, 2011) or assessed only a single word production component at a time (e.g., Strijkers & Costa, 2016). Therefore, Strijkers et al. (2017) conducted a MEG study on overt picture naming, and to explore the spatiotemporal dynamics of word production, they manipulated within the same experiment lexico-semantic (words with either lower or higher frequency: e.g., stool vs. table) and articulatory-acoustic properties (minimal pair words with either an initial labial or coronal speech sound: e.g., Monkey vs. Donkey). The crucial finding in this study is that the obtained spatiotemporal patterns of activation correlated with both the words' frequency and initial phonemes during early stages of language production. More precisely, between 160-240 ms after stimulus onset activity in the left inferior frontal and middle temporal gyri was modulated by the words' lexical frequency, and in the same time window initial phoneme-specific dissociations (labial vs. coronal) were observed in the sensorimotor cortex and the superior temporal gyrus. In other words, this study showed that the phonological component is accessed alongside the lexical word properties (for more evidence favoring a parallel processing view in speech production: e.g., Feng et al., 2021; Miozzo et al., 2015; Ries et al., 2017; Strijkers et al., 2010).

Nevertheless, while such results pose serious problems for purely serial processing theories, it remains debated whether they fit word production models with a sequential time-course. That is, the type effects presented above where a lexico-semantic and phonological variable both emerge within (roughly) 250 ms of processing, may still reflect a sequential activation over hierarchically distinct processing layers (lexico-semantic → phonological),

but simply where activation spreads from one layer onto the other in a faster manner than previously assumed, namely in 10s of ms instead of 100s of ms (e.g., Mahon & Navarrete, 2016; Strijkers et al., 2017). To assess this, Fairs et al. (2021) approached the issue differently and compared the time-course of lexico-semantic (targeted through lexical frequency) and phonological word processing (targeted through phonotactic frequency) during both production and perception. The logic here was the following: Given the hierarchical nature of sequential models, they predict the reverse time-course between production and perception. In production the lexical frequency effect, linked to lexico-semantic word knowledge, should emerge before the phonotactic frequency effect, linked to phonological word knowledge, while in perception the reverse should happen (regardless of whether this happens fast in 10s of ms or slow in 100s of ms). In contrast, according to a parallel assembly model, a word ignites as a whole both in production and perception and thus for both language modalities the lexical and phonotactic frequency effects should manifest simultaneously. The latter is

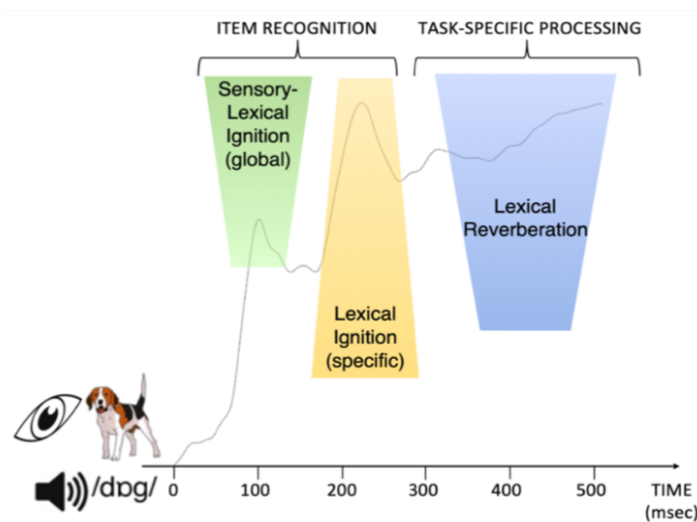


Figure 8. Schematic overview of word production processing according to the parallel assembly model. First, up to about 300 ms lexical ignition takes place where the word assembly becomes active simultaneously. Second, task- and stimulus-specific reverberations occur after 300 ms of processing (taken from Fairs et al. (2021)).

indeed the result found by Fairs et al. (2021), where for both production and perception lexical frequency and phonotactic frequency variables modulated the ERP effects in the early time windows (74-145 ms and 186-287 ms). Differences between the language modalities occurred only in a later time-window (316-369 ms after stimulus onset).

The results by Fairs et al. (2021) are particularly intriguing because they demonstrate the parallel effects (early on) and the sequential effects (later on) within the same study. Indeed, their data showed that after 300 ms of processing lexical and phonotactic frequency only modulated word production not perception, hereby showing modality and task-specific reverberations at later stages of processing. In this manner, a parallel assembly model explains the temporal dynamics underpinning lexical access in the following manner (see Figure 8): After early initial sensorial activation in response to the input, (1) the first linguistic activation emerges roughly between 75-150 ms after onset denoting the start of “globally” activating potential words fitting the initial sensory analyses (this ultra-rapid lexical access may be achieved via prediction); (2) next, this initial “global word space” activation is further refined and delineated within roughly 150-250 ms leading to the ignition and recognition of the specific lexical item associated with the input; (3) finally, roughly after 300 ms, slower, sequential-like reverberation upon the activated word assembly takes effect in order to embed the recognised target word into the proper linguistic and task context to be able to perform the intended behaviour, and modality-specific processing effects become visible. Note the last point is key, because it highlights that the parallel assembly model of language production does not invalidate all the studies reporting sequential activation time-courses observed in word production experiments, but rather offers an alternative interpretation to the same data. In particular, rather than reflecting the initial activation of word components in a sequential manner, the observed sequential effects reflect sensitivity to the later task- and language-specific processes upon a parallelly retrieved word representation (see e.g., Strijkers, 2016; Strijkers et al., 2017; Fairs et al., 2021). In a similar vein, dissociations between lexico-semantics and phonology in patients or as observed for speech errors (e.g., Brehm & Goldrick, 2016), would in this model be due to problems during reverberatory processing rather than initial word retrieval (for more details see: e.g., Strijkers,

2016; Strijkers & Costa, 2016b). In sum, this temporal dynamic of lexical access is markedly different from those described for the models in the previous sections in that there is not a single time frame (serial or interactive) where a given linguistic component is active, but (at the least) two functionally distinct time frames: a parallel linked to word activation (i.e., ignition) and a sequential one linked to task-specific operations upon that ignited word assembly (i.e., reverberation) (see Figures 7 and 8).

Another important difference between prior models and the parallel assembly model lies in the distinct language-to-brain mapping (e.g., Pulvermüller & Fadiga, 2010; Strijkers, 2016; Strijkers & Costa, 2016; Pulvermüller, 1999; 2018). While in the previous models that we discussed the link between brain localization and linguistic function was mainly driven by a one-to-one (for example, lexical representations in MTG, phonological representations in STG, etc.; e.g., Indefrey & Levelt, 2004), or one-to-many relationship (for example, lexical processing in a ventral stream, phonological processing in a dorsal stream; e.g., Hickok, 2012; 2014; Dell et al., 2013), in the parallel assembly model mapping is many-to-many. This means that the neural organization of linguistic components during word production (lexico- semantics and phonology) is reflected in distributed frontotemporal and parietal circuits that integrate sensorimotor networks (see Figure 7). The reason why different word components will map onto different overlapping and distributed neural networks in parallel assembly models is similar as to why time-course of lexical access manifests rapidly and simultaneously in these models, namely because of the Hebbian neural binding between consistently co-activated input (perception) and output (production) (see Figure 7). This is the reason why an assembly model can explain ‘embodied-like’ responses during language perception (for example, the word ‘kick’ activating the leg-region in the motor cortex) (e.g., Watkins et al., 2003; Hauk et al., 2004; Pulvermüller et al., 2006; Carota et al., 2012; Dreyer et al., 2015), and why during object naming Strijkers and colleagues (2017) observed parallel

distributed frontotemporal networks activated in response to both lexical and phonological properties (see also: e.g., Munding et al., 2016; Ries et al., 2017), with (just as in perception) feature-specific topographies in the sensorimotor cortex (e.g., bilabial phonemes linked to lip motor cortex and alveolar phonemes linked to tongue motor cortex; see also Fairs et al., 2021).

In summary, contrary to the previous models we have discussed, lexical access for word production in a parallel assembly model does not assume there is a localizable mental lexicon that is activated at a single, specific point in time. In fact, according to the neural assembly view on language, in general, and word production, specifically, there is no lexical layer of processing. Instead, lexical representations in this model are the binding of meaning-features and sound-features across time and across neural space, resulting in integrated word assemblies. The prediction that there is no specific lexical layer, and thus no dedicated stage of lexical access, is quite a big conceptual difference with the prior models. Nevertheless, it does not mean either that we move away from a representationalism view on cognition and language, but rather that the unit of processing is the ‘word’, the word as the Gestalt of language processing.

Conclusion

In this chapter we have outlined the main theories on lexical access during language production. Lexical access as a process of binding conceptual semantic information with its phonological form has been analyzed from different perspectives: as a serial/sequential hierarchical system (e.g., Levelt, 1999; Levelt & Indefrey, 2004; Indefrey, 2011), as an interactive system that allows more communication between its levels (e.g., Dell, 1986; Hickok, 2012; Dell et al., 2013), and as a parallel non-hierarchical system (e.g., Pulvermuller, 1999; 2018; Strijkers, 2016; Strijkers & Costa, 2016). The development of these theories

spans a few decades of research, and the topic still remains a vital area of research with many exciting issues to resolve. Traditional serial and interactive accounts on word production have played an essential role in advancing the field by providing the first clear models where the main components of word production were identified. Recently, the advancements of research methods and imaging techniques have allowed to investigate word production in more fine-grained detail and ecologically valid conditions. This, in turn, has triggered some reconsideration of past ideas and postulates that have been considered undeniable, especially the hierarchical and serial nature of speech production processing.

References

References

- Abdel Rahman, R., & Melinger, A. (2009). Semantic context effects in language production: A swinging lexical network proposal and a review. *Language and Cognitive Processes, 24*(5), 713–734.
- Alario, F.-X., Segui, J., & Ferrand, L. (2000). Semantic and associative priming in picture naming. *Q. J. Exp. Psychol. A 53A*, 741–764.
- Antonucci, S.M., Beeson, P.M., Labiner, D.M. and Rapcsak, S.Z., (2008). Lexical retrieval and semantic knowledge in patients with left inferior temporal lobe lesions. *Aphasiology, 22*(3), 281-304.
- 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.
- Baars, B. J., Motley, M. T., & MacKay, D. G. (1975). Output editing for lexical status in artificially elicited slips of the tongue. *Journal of Verbal Learning & Verbal Behavior, 14*(4), 382–391.
- Baldo, J.V., Klostermann, E.C., Dronkers, N.F. (2008). It's either a cook or a baker: patients with conduction aphasia get the gist but lose the trace. *Brain and Language, 105*(2), 134–140.
- Binder, J.R. and Desai, R.H., (2011). The neurobiology of semantic memory. *Trends in cognitive sciences, 15*(11), 527-536.
- Blakemore, S.-J., Frith, C. D. Wolpert, D. M. (2001) The cerebellum is involved in predicting the sensory consequences of action, *Neuroreport, 12*(9), 1879-1884.
- Bloem, I. & LaHeij, W. (2003). Semantic facilitation and semantic interference in word translation: implications for models of lexical access in language production. *J. Mem. Lang.* 48, 468–488.

- Braitenberg, V. (1978). Cell assemblies in the cerebral cortex. In *Theoretical approaches to complex systems* (pp. 171-188). Springer, Berlin, Heidelberg.
- Brehm, L., & Goldrick, M. (2016). Empirical and conceptual challenges for neurocognitive theories of language production. *Language, Cognition and Neuroscience*, 31(4), 504-507.
- Buchsbaum, B., Hickok, G., Humphries, C. (2001). Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cognitive Science*, 25, 663–678.
- Buchsbaum, B.R., Baldo, J., Okada, K., Berman, K.F., Dronkers, N., D'Esposito, M. and Hickok, G., (2011). Conduction aphasia, sensory-motor integration, and phonological short-term memory—an aggregate analysis of lesion and fMRI data. *Brain and language*, 119(3), 119-128.
- Burnett, T.A., Freedland, M.B., Larson, C.R., Hain, T.C. (1998). Voice F0 responses to manipulations in pitch feedback. *J. Acoust. Soc. Am.*, 103(6), 3153–3161.
- Buzsáki, G. (2010). Neural syntax: cell assemblies, synapses, and readers. *Neuron*, 68 : 362-385.
- Caramazza, A., (1997). How many levels of processing are there in lexical access?. *Cognitive neuropsychology*, 14(1), 177-208.
- Christoffels, I. K., Firk, C., & Schiller, N.O. (2007). Bilingual language control: An event-related brain potential study. *Brain Research*, 1147, 192–208.
- Cloutman, L., Gottesman, R., Chaudhry, P., Davis, C., Kleinman, J.T., Pawlak, M., Herskovits, E.H., Kannan, V., Lee, A., Newhart, M. and Heidler-Gary, J., (2009). Where (in the brain) do semantic errors come from?. *Cortex*, 45(5), 641-649.
- Coltheart, M., Rastle, K., Perry, C., Langdon, R. and Ziegler, J., (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological review*, 108(1), 204.
- Costa, A., Alario, F.-X. & Caramazza, A. (2005). On the categorical nature of the semantic interference effect in the picture–word interference paradigm. *Psychon. Bull. Rev.* 12, 125–131.
- Cutting, J. C., & Ferreira, V. S. (1999). Semantic and phonological information flow in the production lexicon. *Journal of experimental psychology. Learning, memory, and cognition*, 25(2), 318–344.
- Damasio, H., Grabowski, T.J., Tranel, D., Hichwa, R.D. and Damasio, A.R., (1996). A neural basis for lexical retrieval. *Nature*, 380(6574), 499-505.
- Damian, M., & Bowers, J. (2003). Locus of semantic interference in picture-word interference tasks. *Psychonomic Bulletin & Review*, 10, 111-117.
- 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.
- de Zubicaray, G.I., McMahon K.L. & 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.

- de Zubicaray, G.I., McMahon, K.L. (2009). Auditory context effects in picture naming investigated with event-related fMRI. *Cognitive, Affective, & Behavioral Neuroscience* 9, 260–269.
- DeLeon, J., Gottesman, R.F., Kleinman, J.T., Newhart, M., Davis, C., Heidler-Gary, J., Lee, A. and Hillis, A.E., (2007). Neural regions essential for distinct cognitive processes underlying picture naming. *Brain*, 130(5), 1408-1422.
- Dell, G. S., & Chang, F. (2013). The P-chain: relating sentence production and its disorders to comprehension and acquisition. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 369(1634), 20120394.
- Dell, G. S., & O’Seaghdha, P. G. (1991). Mediated and Convergent Lexical Priming in Language Production: A Comment on Levelt et al. (1991). *Psychological Review*, 98(4), 604–614.
- Dell, G. S., & O’Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition*, 42(1–3), 287–314.
- Dell, G. S., & Reich, P. A. (1977). A model of slips of the tongue. *The Third LACUS Forum*, 3, 448-455.
- Dell, G. S., Schwartz, M. F., Nozari, N., Faseyitan, O., & Branch Coslett, H. (2013). Voxel-based lesion-parameter mapping: Identifying the neural correlates of a computational model of word production. *Cognition*, 128(3), 380–396.
- Dell, G., Nazbanou, N., Oppenheim, G.M., Ferreira, V., Goldrick, M., & Miozzo, M. (2014). Word Production: Behavioral and Computational Considerations. In Goldrick, M., Ferreira, V.S. & Miozzo, M. (ed.), *The Oxford Handbook of Language Production*, Oxford University Press.
- Dell, G.S. and Reich, P.A., (1981). Stages in sentence production: An analysis of speech error data. *Journal of verbal learning and verbal behavior*, 20(6), 611-629.
- Dell, G.S., Lawler, E.N., Harris, H.D. and Gordon, J.K., (2004). Models of errors of omission in aphasic naming. *Cognitive Neuropsychology*, 21(2-4), 125-145.
- Dell, G.S., Martin, N. and Schwartz, M.F., (2007). A case-series test of the interactive two-step model of lexical access: Predicting word repetition from picture naming. *Journal of Memory and Language*, 56(4), 490-520.
- Dell, G.S., Schwartz, M.F., Martin, N., Saffran, E.M. and Gagnon, D.A., (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological review*, 104(4), 801.
- 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*.
- Duffau, H., Gatignol, P., Mandonnet, E., Capelle, L. and Taillandier, L., (2008). Intraoperative subcortical stimulation mapping of language pathways in a consecutive series of 115 patients with Grade II glioma in the left dominant hemisphere. *Journal of neurosurgery*, 109(3), 461-471.
- Edwards, E., Nagarajan, S. S., Dalal, S. S., Canolty, R. T., Kirsch, H., Barbaro, N. M., & Knight, R. T. (2010). Spatiotemporal imaging of cortical activation during verb generation and picture naming. *Neuroimage* 50, 291–301.

- Fairbanks, G. (1954). Systematic research in experimental phonetics: 1. A theory of the speech mechanism as a servosystem. *Journal of Speech and Hearing Disorders*, 19, 133-139.
- Fairs, A., Michelas, A., Dufour, S., Strijkers, K. (2021). The same ultra-rapid parallel brain dynamics underpin the production and perception of speech. *Cerebral Cortex Communications*, Volume 2, Issue 3.
- Fargier, R. Laganaro, M. (2017). Spatio-temporal dynamics of referential and inferential naming: different brain and cognitive operations to lexical selection. *Brain Topography*, 30(2), 182-197.
- Feng, C., Damian, M.F., Qu, Q. (2021). Parallel Processing of Semantics and Phonology in Spoken Production: Evidence from Blocked Cyclic Picture Naming and EEG. *J Cogn Neurosci*. 1-14.
- Ferreira, V. S., & Pashler, H. (2002). Central bottleneck influences on the processing stages of word production. *Journal of experimental psychology. Learning, memory, and cognition*, 28(6), 1187–1199.
- Ferreira, V.S. and Griffin, Z.M., (2003). Phonological influences on lexical (mis) selection. *Psychological Science*, 14(1), 86-90.
- Fridriksson, J., Yourganov, G., Bonilha, L., Basilakos, A., Den Ouden, D.B. and Rorden, C., (2016). Revealing the dual streams of speech processing. *Proceedings of the National Academy of Sciences*, 113(52), 15108-15113.
- Fries, P. (2005). A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence. *Trends in Cognitive Sciences*, 9(10), 474–480.
- Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language*, 47, 27–52.
- Fromkin, V. A. (1973). Slips of the tongue. *Scientific American*, 229(6), 110–117.
- Gallese, V., Fadiga, L., Fogassi, L., Luppino, G., & Murata, A. (1997). A parietal-frontal circuit for hand grasping movements in the monkey: evidence from reversible inactivation experiments. In P. Thier and H.-O. Karnath (Eds.) *Parietal lobe contributions to orientation in 3D-space* (pp. 255-270). Heidelberg: Springer-Verlag.
- Garrett, M. F. (1975). The analysis of sentence production. In: *The psychology of learning and motivation: Vol. 9, ed. G. H. Bower. Academic Press.*
- Glaser, W. R. (1992). Picture naming. *Cognition* 42, 61–105.
- Glaser, W. R. & Glaser, M. O. (1989). Context effects in Stroop-like word and picture processing. *Journal of Experimental Psychology: General*, 118:13–42.
- Glaser, W. R., & Dünghoff, F.-J. (1984). The time-course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 640–654.
- Goldrick, M. (2011), Linking Speech Errors and Generative Phonological Theory. *Language and Linguistics Compass*, 5, 397-412.
- Goldrick, M., & Rapp, B. (2002). A restricted interaction account (RIA) of spoken word production: The best of both worlds. *Aphasiology*, 16(1-2), 20–55.

- Golfinoopoulos, E., Tourville, J.A., Guenther, F.H. (2010). The integration of largescale neural network modeling and functional brain imaging in speech motor control. *Neuroimage*, 52(3), 862–874.
- Goodale, M.A., Milner, A.D., Jakobson, L.S., & Carey, D.P. (1991). A neurological dissociation between perceiving objects and grasping them. *Nature*, 349, 154-156.
- Gracco, V. L., Lofqvist, A. (1994). Speech motor coordination and control: evidence from lip, jaw, and laryngeal movements. *Journal of Neuroscience*, 14(11 Pt 1), 6585–6597.
- Graves, W. W., Grabowski, T. J., Mehta, S., & Gordon, J. K. (2007). A neural signature of phonological access: distinguishing the effects of word frequency from familiarity and length in overt picture naming. *Journal of cognitive neuroscience*, 19(4), 617–631.
- Graves, W.W., Grabowski, T.J., Mehta, S. and Gupta, P., (2008). The left posterior superior temporal gyrus participates specifically in accessing lexical phonology. *Journal of cognitive neuroscience*, 20(9), 1698-1710.
- Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. *J Commun Disord*. 39(5), 350–365.
- Habets, B., Jansma, B.M. & Münte, T.F. (2008). Neurophysiological correlates of linearization in language production. *BMC Neurosci* 9, 77.
- Hanulová, J., Davidson, D. J., & Indefrey, P. (2011). Where does the delay in L2 picture naming come from? Neurocognitive evidence on second language word production. *Language and Cognitive Processes*, 26(7), 902–934.
- Harley, T. A. (1984). A critique of top-down independent levels models of speech production: Evidence from non-plan-internal speech errors. *Cognitive Science*, 8, 191-219.
- Harley, T. A. (1993). Phonological activation of semantic competitors during lexical access in speech production. *Language and Cognitive Processes*, 8, 291-309.
- Hartsuiker, R. J., Antón-Méndez, I., Roelstraete, B., & Costa, A. (2006). Spoonish spanerisms: A lexical bias effect in spanish. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(4), 949–953.
- Hartsuiker, R.J., Corley, M. and Martensen, H., (2005). The lexical bias effect is modulated by context, but the standard monitoring account doesn't fly: Related beply to Baars et al.(1975). *Journal of Memory and Language*, 52(1), 58-70.
- Hauk, O., Patterson, K., Woollams, A., Cooper-Pye, E., Pulvermuller, F. & Rogers, T. T. (2007). How the camel lost its hump: the impact of object typicality on event-related potential signals in object decision. *J. Cogn. Neurosci*. 19, 1338–1353.
- Hebb, D. O. (1949). The organization of behavior; a neuropsychological theory. *Psychology Press*.
- Heinks-Maldonado, T. H., Nagarajan, S. S., & Houde, J. F. (2006). Magnetoencephalographic evidence for a precise forward model in speech production. *Neuroreport*, 17(13), 1375–1379.
- Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, 13(2), 135–145.
- Hickok, G. (2014). The architecture of speech production and the role of the phoneme in speech processing. *Language and Cognitive Processes*, 29, 2–20.

- Hickok, G. and Poeppel, D., (2000). Towards a functional neuroanatomy of speech perception. *Trends in cognitive sciences*, 4(4), 131-138.
- Hickok, G. and Poeppel, D., (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, 92(1-2), 67-99.
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5), 393–402.
- Hickok, G., Okada, K., Serences, J.T. (2009). Area Spt in the human planum temporale supports sensory-motor integration for speech processing. *Journal of Neurophysiology*, 101(5), 2725–2732.
- Houde, J. F., & Nagarajan, S. S. (2011). Speech production as state feedback control. *Frontiers in human neuroscience*, 5, 82.
- Houde, J. F., Nagarajan, S. S., Sekihara, K., & Merzenich, M. M. (2002). Modulation of the auditory cortex during speech: an MEG study. *Journal of cognitive neuroscience*, 14(8), 1125–1138.
- Houde, J.F., Jordan, M.I. (1998). Sensorimotor adaptation in speech production. *Science*, 279, 1213–1216.
- Hultén, A., Vihla, M., Laine, M. and Salmelin, R. (2009), Accessing newly learned names and meanings in the native language. *Hum. Brain Mapp.*, 30, 976-989.
- Humphreys, K.R., (2002). Lexical bias in speech errors. University of Illinois at Urbana-Champaign.
- Indefrey, P. (2011). The spatial and temporal signatures of word production components: a critical update. *Front. Psychol.* 2, 255.
- Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, 92(1–2), 101–144.
- Ito, M. (2008). Control of mental activities by internal models in the cerebellum. *Nat Rev Neurosci* 9, 304–313.
- Janssen, N., Schirm, W., Mahon, B. Z. & Caramazza, A. (2008). The semantic interference effect in the picture-word interference paradigm: evidence for the response selection hypothesis. *J. Exp. Psychol. Learn. Mem. Cogn.* 34, 249–256.
- Jeannerod, M., Arbib, M. A., Rizzolatti, G., & Sakata, H. (1995). Grasping objects: the cortical mechanisms of visuomotor transformation. *Trends in neurosciences*, 18(7), 314–320.
- Jefferies, E., Crisp, J. and Ralph, M.A.L., (2006). The impact of phonological or semantic impairment on delayed auditory repetition: Evidence from stroke aphasia and semantic dementia. *Aphasiology*, 20(9), 963-992.
- Jescheniak, J. D. & Levelt, W. J. M. (1994). Word frequency effects in speech production: retrieval of syntactic information and of phonological form. *J. Exp. Psychol. Learn. Mem. Cogn.* 20, 824–843.
- Jescheniak, J. D. & Schriefers, H. (1998). Discrete serial versus cascaded processing in lexical access in speech production: Further evidence from the co-activation of near-synonyms. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24:1256–74.

- Jescheniak, J. D., Hahne, A., & Schriefers, H. (2003). Information flow in the mental lexicon during speech planning: evidence from event-related brain potentials. *Brain research. Cognitive brain research*, 15(3), 261–276.
- Johnson, J. S. & Olshausen, B. A. (2005). The earliest EEG signatures of object recognition in a cued-target task are postsensory. *J. Vis.* 5, 299–312.
- Kawashima, R., Naitoh, E., Matsumura, M., Itoh, H., Ono, S., Satoh, K., Gotoh, R., Koyama, M., Inoue, K., Yoshioka, S. & Fukuda, H. (1996). Topographic representation in human intraparietal sulcus of reaching and saccade. *Neuroreport*, 7, 1253-1256.
- Koester, D., & Schiller, N. O. (2008). Morphological priming in overt language production: electrophysiological evidence from Dutch. *NeuroImage*, 42(4), 1622–1630.
- Laganaro, M., Morand, S., Schwitter, V., Zimmermann, C., Camen, C., & Schnider, A. (2009). Electrophysiological correlates of different anomia patterns in comparison with normal word production. *Cortex*, 45(6), 697–707.
- Laganaro, M., Valente, A., & Perret, C. (2012). Time course of word production in fast and slow speakers: a high density ERP topographic study. *NeuroImage*, 59(4), 3881–3888.
- Levelt W. J. (2001). Spoken word production: a theory of lexical access. *Proceedings of the National Academy of Sciences of the United States of America*, 98(23), 13464–13471.
- Levelt, W. J. M. (1989). Speaking: From intention to articulation. *The MIT Press*.
- Levelt, W. J. M. & Kelter, S. (1982). Surface form and memory in question answering. *Cognitive Psychology*, 14, 78–106.
- Levelt, W. J. M., Praamstra, P., Meyer, A. S., Helenius, P., & Salmelin, R. (1998). An MEG study of picture naming. *Journal of Cognitive Neuroscience*, 10(5), 553–567.
- Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T. & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review*, 98, 122–42.
- Levelt, W.J., Roelofs, A. and Meyer, A.S., (1999). A theory of lexical access in speech production. *Behavioral and brain sciences*, 22(1), 1-38.
- Lichtheim, L. (1885). On aphasia. *Brain*, 7, 433–484.
- Lupker, S. J. (1982). The role of phonetic and orthographic similarity in picture-word interference. *Canadian Journal of Psychology*, 36, 349–367.
- 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.
- Mägiste, E. (1984). Stroop tasks and dichotic translation: The development of interference patterns in bilinguals. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 10, 304-315.
- Martin, N., (1996). Phonological facilitation of semantic errors in normal and aphasic speakers. *Language and Cognitive Processes*, 11(3), 257-282.
- Meyer, A. S. (1992). Investigation of phonological encoding through speech error analyses: Achievements, limitations, and alternatives. *Cognition*, 42, 181–211.

- Meyer, A. S., & Schriefers, H. (1991). Phonological facilitation in picture-word interference experiments: Effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(6), 1146–1160.
- Meyer, A.S., Roelofs, A. & Brehm, L. (2019). Thirty years of Speaking: An introduction to the Special Issue, *Language, Cognition and Neuroscience*, 34:9, 1073-1084.
- Milner, A. D. & Goodale, M. A. (1995). *The visual brain in action*. Oxford Univ. Press, Oxford.
- Milner, A.D. & Goodale, M.A. (1993). Visual pathways to perception and action. In T.P. Hicks, S. Molotchnikoff & T. Ono (Eds.) *Progress in Brain Research*, Vol. 95 (pp. 317-337). Amsterdam: Elsevier.
- 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. *Cereb Cortex*. 25, 3343-3355.
- Motley, M. T., Camden, C. T., & Baars, B. J. Covert formulation and editing of anomalies in speech production: Evidence from experimentally elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior*, 21(5), 578–594.
- 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.
- Navarrete, E., & Costa, A. (2005). Phonological activation of ignored pictures: Further evidence for a cascade model of lexical access. *Journal of Memory and Language*, 53(3), 359–377.
- Nooteboom, S. G. (2005). Lexical bias revisited: Detecting, rejecting and repairing speech errors in inner speech. *Speech Communication*, 47(1-2), 43–58.
- Nooteboom, S.G., (2005). Listening to oneself: Monitoring speech production. In *Phonological encoding and monitoring in normal and pathological speech* (pp. 179-198). Psychology Press.
- Nozari, N. and Dell, G.S., (2009). More on lexical bias: How efficient can a “lexical editor” be?. *Journal of Memory and language*, 60(2), 291-307.
- Nozari, N., & Pinet, S. (2020). A critical review of the behavioral, neuroimaging, and electrophysiological studies of co-activation of representations during word production. *Journal of Neurolinguistics*, 53, Article 100875.
- Nozari, N., Kittredge, A. K., Dell, G. S., & Schwartz, M. F. (2010). Naming and repetition in aphasia: Steps, routes, and frequency effects. *Journal of memory and language*, 63(4), 541–559.
- Nozari, N., Kittredge, A.K., Dell, G.S. and Schwartz, M.F., (2010). Naming and repetition in aphasia: Steps, routes, and frequency effects. *Journal of memory and language*, 63(4), 541-559.
- Perkell, J.S. (2012). Movement goals and feedback and feedforward control mechanisms in speech production. *Journal of Neurolinguistics*, 25(5), 382–407.
- Piai V., Roelofs A., Jensen O., Schoffelen J.M., & Bonnefond M. (2014). Distinct Patterns of Brain Activity Characterise Lexical Activation and Competition in Spoken Word Production. *PLOS ONE* 9(2): e88674.

- 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.
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *The Behavioral and brain sciences*, 36(4), 329–347.
- Pickering, M.J. and Garrod, S., (2004). Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences*, 27(2), 169-190.
- Pulvermüller F. (1999). Words in the brain's language. *The Behavioral and brain sciences*, 22(2), 253–336.
- Pulvermüller, F. (2002). The neuroscience of language. *Cambridge: Cambridge University Press*.
- Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, 6(7), 576–582.
- Pulvermüller, F. (2018), Neurobiological Mechanisms for Semantic Feature Extraction and Conceptual Flexibility. *Top Cogn Sci*, 10, 590-620.
- Pulvermüller, F., & Fadiga, L. (2010). Active perception: Sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, 11(5), 351–360.
- Rapp, B. and Goldrick, M., (2000). Discreteness and interactivity in spoken word production. *Psychological review*, 107(3), 460.
- Rapp, B. and Goldrick, M., (2004). Feedback by any other name is still interactivity: a reply to Roelofs (2004).
- Rauschecker, J.P. and Scott, S.K., (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature neuroscience*, 12(6), pp.718-724.
- Riès, S. K., 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. *Proc Natl Acad Sci*. 114 : E4530-E4538.
- Rizzolatti, G., Fogassi, L. & Gallese, V. (1997). Parietal cortex: from sight to action. *Curr. Opin. Neurobiol.* 7, 562–567.
- Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, 42, 107–142.
- Roelofs, A. (1997a). The WEAVER model of word-form encoding in speech production. *Cognition*, 64, 249–284.
- Roelofs, A. (1997b). A case for non-decomposition in conceptually driven word retrieval. *Journal of Psycholinguistic Research* 26, 33–67.
- Roelofs, A. (2004). Error biases in spoken word planning and monitoring by aphasic and nonaphasic speakers: comment on Rapp and Goldrick (2000). *Psychological Review*, 111, 579-80.
- Roelofs, A. (2014). A dorsal-pathway account of aphasic language production: The WEAVER++/ARC model. *Cortex*, 59, 33–48.
- Roelofs, A. (2018). A unified computational account of cumulative semantic, semantic

- blocking, and semantic distractor effects in picture naming. *Cognition*, 172, 59-72.
- Roelofs, A., & Ferreira, V. (2019). The architecture of speaking. In Hagoort, P. (ed.), *Human language: From genes and brains to behavior*, 35 - 50. Cambridge, MA: MIT Press.
- Runnqvist, E., Bonnard, M., Gauvin, H.S., Attarian, S., Trébuchon, A., Hartsuiker, R.J. & Alario, F.-X. (2016). Internal modeling of upcoming speech: A causal role of the right posterior cerebellum in non-motor aspects of language production, *Cortex*, 81, 203-214.
- Runnqvist, E., Chanoine, V., Strijkers, K., Pattamadilok, C., Bonnard, M., Nazarian, B., Sein, J., Anton, J.-L., Dorokhova, L., Belin, P. & Alario, F.-X. (2021). Cerebellar and Cortical Correlates of Internal and External Speech Error Monitoring. *Cerebral Cortex Communications*. 2 (2).
- Sahin, N. T., Pinker, S., Cash, S. S., Schomer, D., & Halgren, E. (2009). Sequential processing of lexical, grammatical, and phonological information within Broca's area. *Science (New York, N.Y.)*, 326(5951), 445–449.
- Salmelin, R., Hari, R., Lounasmaa, O. V., & Sams, M. (1994). Dynamics of brain activation during picture naming. *Nature*, 368, 463–465.
- Schade, U., & Berg, T. (1992). The role of inhibition in a spreading-activation model of language production. II. The simulational perspective. *Journal of Psycholinguistic Research*, 21, 435-462.
- Schmitt, B. M., Münte, T. F. & Kutas, M. (2000). Electrophysiological estimates of the time course of semantic and phonological encoding during implicit picture naming. *Psychophysiology* 37, 473–484.
- Schnur, T.T., Schwartz, M.F., Kimberg, D.Y., Hirshorn, E., Coslett, H.B. and 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), pp.322-327.
- Schriefers, H., Meyer, A. S. & Levelt, W. J. M. (1990). Exploring the time course of lexical access in speech production: Picture-word interference studies. *Journal of Memory and Language*, 29, 86–102.
- Schwartz, M.F., Dell, G.S., Martin, N., Gahl, S. and Sobel, P., (2006). A case-series test of the interactive two-step model of lexical access: Evidence from picture naming. *Journal of Memory and language*, 54(2), 228-264.
- Schwartz, M.F., Faseyitan, O., Kim, J. and Coslett, H.B., (2012). The dorsal stream contribution to phonological retrieval in object naming. *Brain*, 135(12), 3799-3814.
- Scott, S. and Johnsrude, I., (2003). The organisation and functional organisation of speech perception. *Trends in Neurosciences*, 26(2), 100-107.
- Seidenberg, M.S. and McClelland, J.L., (1989). A distributed, developmental model of word recognition and naming. *Psychological review*, 96(4), 523
- Shadmehr, R., Krakauer, J.W. (2008). A computational neuroanatomy for motor control. *Exp. Brain. Res.*, 185(3), 359–381.
- Shadmehr, R., Smith, M.A., Krakauer, J.W. (2010). Error correction, sensory prediction, and adaptation in motor control. *Annual Review of Neuroscience*, 33, 89–108.
- Singer, W. (2013). Cortical dynamics revisited. *Trends Cogn Sci.*, 17, 616-626.

- Singer, W., Gray, C.M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annu Rev Neurosci.*, 18, 555-586.
- Starreveld, P. A., & La Heij, W. (1996). Time-course analysis of semantic and orthographic context effects in picture naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(4), 896–918.
- Strijkers, K. (2016), A Neural Assembly–Based View on Word Production: The Bilingual Test Case. *Language Learning*, 66, 92-131.
- Strijkers, K. & Costa, A. (2016) The cortical dynamics of speaking: present shortcomings and future avenues. *Language, Cognition and Neuroscience*, 31:4, 484-503.
- Strijkers, K., & Costa, A. (2016b). On words and brains: Linking psycholinguistics with neural dynamics in speech production. *Language, Cognition and Neuroscience*, 31(4), 524–535.
- Strijkers, K., & Costa, A. (2011). Riding the Lexical Speedway: A Critical Review on the Time Course of Lexical Selection in Speech Production. *Frontiers in Psychology*, 2(December), 1–16.
- Strijkers, K., Costa, A., & Pulvermüller, F. (2017). The cortical dynamics of speaking: Lexical and phonological knowledge simultaneously recruit the frontal and temporal cortex within 200 ms. *NeuroImage*, 163, 206–219.
- 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.
- Stuart, A., Kalinowski, J., Rastatter, M.P., Lynch, K. (2002). Effect of delayed auditory feedback on normal speakers at two speech rates. *J. Acoust. Soc. Am.*, 111(5 Pt 1), 2237–2241.
- Thorpe, S., Fize, D. & Marlot, C. (1996). Speed of processing in the human visual system. *Nature* 381, 520–522.
- Tian, X., & Poeppel, D. (2010). Mental imagery of speech and movement implicates the dynamics of internal forward models. *Frontiers in psychology*, 1, 166.
- Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*, 26(7), 952–981.
- Ueno, T., Saito, S., Rogers, T.T. and Ralph, M.A.L., (2011). Lichtheim 2: synthesizing aphasia and the neural basis of language in a neurocomputational model of the dual dorsal-ventral language pathways. *Neuron*, 72(2), 385-396.
- Ungerleider, L.G. & Mishkin, M. (1982). Two cortical visual systems. In D.J. Ingle, M.A. Goodale & R.J.W. Mansfield (Eds.). *Analysis of visual behavior* (pp. 549-586). Cambridge, MA: MIT Press
- Van Turenout, M., Hagoort, P. & Brown, C. M. (1998) Brain activity during speaking: From syntax to phonology in 40 msec. *Science*, 280, 572–74.
- VanRullen, R. & Thorpe, S. J. (2002). Surfing a spike wave down the ventral stream. *Vision Res.* 42, 2593–2615.
- Vigliocco, G., & Hartsuiker, R. J. (2002). The interplay of meaning, sound, and syntax in sentence production. *Psychological Bulletin*, 128(3), 442–472.

- Walker, G. M., & Hickok, G. (2016). Bridging computational approaches to speech production: The semantic-lexical-auditory-motor model (SLAM). *Psychonomic bulletin & review*, 23(2), 339–352.
- Wernicke, C. (1874/1969). The symptom complex of aphasia: A psychological study on an anatomical basis. In: Cohen, R.S., Wartofsky, M.W., editors. *Boston studies in the philosophy of science*. Dordrecht: D. Reidel Publishing Company, 34-97.
- Wolpert, D.M., Ghahramani, Z., Jordan, M.I. (1995). An internal model for sensorimotor integration. *Science*, 269(5232), 1880–1882.