



HAL
open science

The cortical dynamics of speaking: present shortcomings and future avenues

Kristof Strijkers

► **To cite this version:**

Kristof Strijkers. The cortical dynamics of speaking: present shortcomings and future avenues. *Language, Cognition and Neuroscience*, 2016, 31 (4), pp.484-503. 10.1080/23273798.2015.1120878 . hal-01731326

HAL Id: hal-01731326

<https://hal.science/hal-01731326>

Submitted on 14 Mar 2018

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 version)

**The Cortical Dynamics of Speaking:
Present Shortcomings and Future Avenues**

Kristof Strijkers^{1,2} & Albert Costa^{3,4}

¹Laboratoire Parole et Langage (LPL), UMR 7309, CNRS, Aix-Marseille Université, 13100, Aix-en-Provence, France.

²Laboratoire de Psychologie Cognitive (LPC), CNRS, Aix-Marseille Université, 13331, Marseille, France.

³Center for Brain and Cognition, Department of Technology, Pompeu Fabra University, 08018 Barcelona, Spain.

⁴ICREA (Institució Catalana de Recerca I Estudis Avançats), 08010 Barcelona, Spain.

Address for correspondence:

Kristof Strijkers

Centre National de la Recherche Scientifique (CNRS)

Laboratoire Parole et Langage (LPL) – Université d’Aix-Marseille

5, Av. Pasteur

13100 Aix-en-Provence, France

Kristof.Strijkers@gmail.com

Abstract

A central question in the cognitive neuroscience of language production concerns the cortical activation time-course by which different word production components become available to the speaker. According to the dominant neurocognitive model of word production this is achieved in a sequential fashion where each of the linguistic components involved in speaking has its specific time-course and dedicated processing center in the brain (Indefrey & Levelt, 2004). Language production would initiate with lexico-semantic processes reflected in posterior brain regions and slowly move ‘upwards’ through the representational formats until an articulatory program is activated in anterior brain regions. In this opinion piece we raise three critical issues aimed at questioning the strict feedforward and serial conceptualization of the spatiotemporal dynamics engendering word production. Concretely, we address the following three topics: 1) Are the data supporting the model or the model supporting the data? We conclude that, despite the evidence supporting the notion of some local sequentiality underpinning speech preparation in the brain, there is insufficient explicit spatiotemporal evidence to support the notion that such local sequentiality by itself is enough to explain how the brain computes language; 2) What is the role of top-down processing on the spatiotemporal dynamics? We argue that if top-down processes such as goal-directed behavior and attention can proactively modulate the linguistic system, the spatiotemporal correlates of word production components should likewise be regarded as context-depend and speaker-adaptive; 3) Can we map psycholinguistic stages onto the brain in a one-to-one fashion? We advocate to start considering, above and beyond the feedforward anatomical connectedness that underpin traditional sequential hierarchical models, other types of neural communication such as neural coherence, feedback projections and reverberatory brain activity in order to mechanistically explain how basic brain properties sustain language production.

Research into the time-course of brain activation (i.e., spatiotemporal dynamics) during speech production has flourished in the last decade, thanks in part to methodological developments and the possibility of recording neurophysiological data during overt speech production (e.g., Edwards et al., 2010; Eulitz, Hauk & Cohen, 2000; Ganushchak, Christoffels & Schiller, 2007; Graves, Grabowski, Mehta & Gordon, 2007; Llorens, Trebuchon, Liegeois-Chauvel & Alario, 2011; Price, 2012; Salmelin, Hari, Lounasmaa & Sams, 1994; Strijkers, Costa & Thierry, 2010; Strijkers & Costa, 2011; Wilson, Isenberg & Hickok, 2009). This, in turn, allowed researchers to start fleshing-out classical psycholinguistic questions (e.g., Caramazza, 1997; Dell, 1986; Levelt, Roelofs & Meyer, 1999), in a more fine-grained perspective. The most important contribution in relation to the spatiotemporal map of language production is the work by Indefrey and Levelt (I&L, 2004). Following the psycholinguistic model of Levelt and colleagues (1999), and after conducting an exhaustive meta-analysis, the authors provide a very precise framework of when and where the different processing stages occur. The goal of this opinion piece is to build upon this seminal work and critically assess some of its properties in order to stimulate research regarding the cortical dynamics of speech production.

According to the spatiotemporal framework proposed by I&L (2004), object naming displays the following step-wise progression (see Fig. 1): 1) Conceptual processing is achieved within the first 150 ms of processing in the occipito-temporal cortex (from visual input); 2) Lexical (lemma) access unfolds between 175-250 ms in the mid temporal gyrus (MTG); 3) The phonemes are encoded in the posterior MTG and the superior temporal gyrus (STG) between 250-350 ms and segmented into syllables between 350-450 ms in the left inferior frontal gyrus (LIFG); 4) Motor programming takes effect in the pre- and post-central gyri between 450-600 ms (see also: e.g., Indefrey, 2011; Levelt, Praamstra, Meyer, Helenius & Salmelin, 1998; Salmelin et al., 1994). On this view speech planning is thought of as a hierarchically organized process initiating with semantics and gradually 'moving upwards' through the required representational formats until a motor program is executed (e.g., Levelt et al., 1999). And although many researchers agree that the system need not be strictly serial and modular, allowing for cascading and interactivity (Caramazza, 1997; Dell, 1986; Dell & O'Sheadgha, 1992; Dell, Schwartz, Martin, Saffran & Gagnon, 1997; Rapp & Goldrick, 2000), the general notion that linguistic knowledge is engendered by specialized neural layers which are triggered

progressively in time, is widely held (2004)¹.

Insert Figure 1 around here

In the last decade, there have been many observations that are consistent with various aspects of this model (see Indefrey, 2011). In terms of timing, several electrophysiological studies of overt speech production have confirmed an onset of lexical access around 150-200 ms after stimulus presentation (e.g., Aristei, Melinger, Abdel Rahman, 2011; Christoffels, Firk & Schiller, 2007; Costa, Strijkers, Martin & Thierry, 2009; Strijkers et al., 2010; Strijkers & Costa, 2011; Strijkers, Holcomb & Costa, 2011a; Strijkers, Baus, Runnqvist, Fitzpatrick & Costa, 2013). Others have reported ERP evidence of lexeme retrieval and phonological encoding emerging (roughly) after 300 ms and 400 ms of processing, respectively (e.g., Christoffels et al., 2007; Koester & Schiller, 2008; Laganaro, Morand & Schnider, 2009a; Laganaro et al., 2009b; Laganaro, Python & Toepel, 2013). And most importantly, some studies (e.g., Laganaro et al., 2009a; 2009b; Sahin, Pinker, Cash, Schomer & Halgren, 2009; Schuhmann, Schiller, Goebel & Sack, 2012) support the sequential nature of the language production system. In terms of space, several studies have reliably linked MTG activation to lexico-semantic and lexico-syntactic operations (e.g., Graves et al., 2007; Menenti, Segaert & Hagoort, 2012; Schuhmann et al., 2012; Schwartz et al., 2009), posterior STG activation to word form encoding (e.g., Hickok, 2012; 2014a ; Graves et al., 2007; 2008; Wilson et al., 2009), and IFG activity to (motor) phonology and syllabification specifically (e.g., Papoutsi et al., 2009; Sahin et al., 2009; Schuhmann et al., 2009; 2012). Finally, a number of MEG studies have reported a similar spatiotemporal profile, with activation maxima slowly progressing bilaterally from occipital cortex, over left temporal and parietal brain regions, towards the left IFG (e.g., Hulten, Vilha, Laine & Salmelin, 2009; Liljestrom, Hulten, Parkkonen & Salmelin, 2009; Soros, Cornelissen, Laine & Salmelin, 2003; Vilha, Laine & Salmelin, 2006).

The evidence cited above may lead us to think that, at least when uttering words in isolation, I&L's spatiotemporal map (2004) provides a realistic (macroscopic) blueprint of the cortical dynamics

¹ Note that it is not our intention to flatten out the important theoretical and functional differences that exist between serial and cascaded/interactive models in explaining psycholinguistic phenomena, speech error patterns and speech pathologies (e.g., Caramazza, 1997; Dell, 1986; Dell et al., 1997; Levelt et al., 1991; 1999; Roelofs, 1992; 1997; Rapp & Goldrick, 2000; Dell et al., 2007). But in terms of a spatiotemporal blueprint of the core word production components in the brain, mist allowing some spatiotemporal flexibility (e.g., Dell & O'Sheagha, 1992), the neural sources linked to each mental operation and their relative time-course as specified by I&L (2004) are not necessarily incompatible with (slowly progressive) cascaded and interactive models (see also Indefrey, 2011).

underpinning word production components. If so, we should be able to start using this model as a roadmap to interpret both psycholinguistic and neurobiological research in language production. Following this rationale, and as long as single word production tasks are employed, psycholinguistic effects occurring roughly between 150-250 ms of processing can reliably be associated with lexical access and the MTG, and activation patterns in the left IFG can reliably be associated with phonological encoding occurring late in the processing stream (see Fig. 1). Needless to point out the immense power such interpretative tool would offer researchers. That said, herein lies exactly the biggest danger of I&L's map, both for past and future research. In this opinion piece we aim at pointing out that the use of this model for data interpretation is not always optimal and at present in many occasions premature. Instead, we will advocate that the data presented in I&L model (2004) are better approached empirically, namely as data on which we can *built*, not *consolidate*; a means to *predictions*, not *interpretations*; a framework which needs to be *accounted for*, but not *accounts for*.

Three Critical Questions for Present Research.

Concretely, we will bring forward 3 issues that we believe need to be addressed in order to advance on our understanding of how the cognitive skill of speaking is reflected in the brain². When doing so, we will advocate for stronger cross-talk between different domains of cognition (beyond the speech production tradition) and different domains of neuroscience (beyond the 'mapping' tradition). Ultimately we aim at showing that the current spatiotemporal map of language production is only a part, but not the whole story of how to integrate language, cognition and neuroscience. The three questions we will address are: 1) Are the data supporting the model or the model supporting the data? 2) What is the role of top-down processing on the spatiotemporal dynamics? 3) Can we map psycholinguistic stages onto the brain in a one-to-one fashion?

² There are obviously more questions to be addressed than the 3 we pose here. To name only a few, exciting novel work is being conducted about the integration of speech production and speech perception (e.g., Dell & Chang, 2014; Hickok, 2012; 2014a; Pickering & Garrod, 2013), to reveal the white matter fiber tracks connecting the important language regions in the brain (e.g., Catani & Mesulam, 2008; Glasser & Rilling, 2008; Petrides & Pandya, 2009) and their functional role in the production of speech (e.g., Dell et al., 2013; Hickok, 2012; Roelofs, 2014; Ueno et al., 2011), uncover the domain-general versus language-specific nodes in the language network (e.g., Fedorenko et al., 2012; Fedorenko & Thompson-Schill, 2014), dissect important language-related brain regions at the microscopic level (e.g., Amunts et al., 2010; Edwards et al., 2010; Sahin et al., 2009), etc. Knowledge on these vibrant areas of research will be vital to deepen our understanding of how the language network is organized in the brain and how that organization translates to our cognitive capability of speaking. However, for present purposes, we choose to focus on questions with a more direct link on how the temporal and spatial correlates of word production components were attributed in I&L's meta-analysis (2004).

The first question will cast some doubt on the neurobiological evidence typically cited to support of I&L's model and will emphasize the necessity of more theory-independent empirical data. The other two questions will consider additional and novel brain language mechanisms that are largely neglected by more traditional models. Although we will be critical with some of the tenets of I&L's model, we do want to underscore its importance in stimulating research into the spatiotemporal dynamics of language production. It provided researchers with a crucial tool to start linking psycholinguistic phenomena to the when and where in the brain, and it will (and should) remain to do so in the future. Moreover, based on the impressive meta-analysis and the available evidence (some of which mentioned above), it is clear that certain aspects of the model represent advancements in the right direction. Specifically, I&L's work (2004) and many studies inspired by it have convincingly demonstrated that (a) there is a sequential component to speech preparation, (b) that most (if not all) of the brain areas identified in I&L are (at least in part) involved in the specific linguistic operations assigned to them (e.g., lexical involvement of the MTG). While acknowledging the latter, we question here whether the proposed sequentiality and local neural selectivity are enough to explain how words in the brain are processed in order to produce speech.

1. Are the data supporting the model or the model supporting the data?

Elsewhere (Strijkers & Costa, 2011) we have already argued that the empirical evidence supporting the temporal scope of the model is limited because much of the chronometric data does not stem from speech production behavior under normal conditions (i.e., immediate and overt language production) (e.g., Camen, Morand & Laganaro, 2010; Guo, Peng, Lu & Liu, 2005; Hanulova, Davidson & Indefrey, 2011; Laganaro et al., 2009a; 2009b; Sahin et al., 2009; Schiller, 2006; Van Turennout, Hagoort & Brown, 1997; 1998; Zhang & Damian, 2009). As a consequence one may question the relevance of temporal estimates derived from covert button-press tasks (e.g., does a picture name start with the letter /b/ or not) or delayed naming responses (provide the overt response after a cued interval in order to avoid motor contamination of the signal) to inform about the time course of overt and immediate speech production. Moreover, the additional cognitive processes necessary to perform these complex tasks (e.g., inhibition in the case of delayed object naming, or meta-linguistic decision making in the case of phoneme identification), could affect the time course with which the mental operations of speech preparation take effect (for a detailed discussion and concrete examples please consult Strijkers & Costa, 2011).

Another and more general problem concerns the difference in data that relies on the model versus data that tests the model; a difference which is often convoluted. Many researchers – including ourselves – have relied (more or less strongly) on the spatiotemporal correlates identified by I&L (2004) to establish which cognitive stage(s) of processing could be responsible (or not responsible) for an observed temporal or spatial finding (e.g., Acheson, Hamidi, Binder & Postle, 2011; Aristei et al., 2011; Cheng, Schafer & Akyurek, 2010; Costa et al., 2009; Christoffels et al., 2007; Dell'Acqua et al., 2010; Graves et al., 2007; Hulten et al., 2009; Koester & Schiller, 2008; 2011; Laganaro et al., 2009a; 2012; Laganaro & Perret, 2011; Perret, Bonin & Laganaro, 2014; Piai, Roelofs & Schriefers, 2011; Piai, Roelofs, Jensen, Schoffelen & Bonnefond, 2014a; Roelofs, 2014; Schuhmann et al., 2009; Strijkers et al., 2010; Vilha et al., 2006). For example, if a certain psycholinguistic effect appears at 200 ms (e.g., word frequency), then this effect is argued to occur at the lexical level since the model assumes that around 200 ms lexical processes are at play. While such theory-driven approach in order to hypothesize about the when and where of a psycholinguistic phenomenon is a common strategy in cognitive sciences, it does not serve to test the theory itself. Perhaps an obvious claim, but nonetheless worth mentioning explicitly since much data considered as evidence supporting the spatiotemporal map, did in fact rely on the models' assumptions for data interpretation rather than test the map itself.

In his review, Indefrey (2011) warns of this issue and gives some concrete examples. He argues that studies exploring the electrophysiological effects of lexical frequency and cognate status (e.g., Laganaro et al., 2009b; Strijkers et al., 2010), could not be used to test the time-course of word retrieval because the origin of those psycholinguistic phenomena can arise at multiple levels (e.g., Hanulova et al., 2011). That is, if a psycholinguistic phenomenon, such as lexical frequency, can arise at more than one specific component of word production, then one cannot use temporal data elicited by that phenomenon to inform the time-course of word retrieval. We agree with Indefrey's reasoning (2011) that testing time estimates by exploring the temporal (or spatial for that matter) effects of psycholinguistic variables is risky if no clear origin of that variable is known (see also Strijkers & Costa, 2011). Unfortunately, this same criticism also applies to the data usually cited in support of I&L's model (2004; Indefrey, 2011).

For example, Indefrey (2011) takes the early effects elicited by different types of semantic interference in object naming (e.g., Aristei et al., 2011; Costa et al., 2009; Maess, Friederici, Damian, Meyer & Levelt, 2002), as supporting the notion that lexical (lemma) selection occurs (roughly) in a time window

between 175 ms and 275 ms. However, not all researchers would agree with the assumption that semantic interference indexes lexical competition. Some argue that the cumulative semantic interference effect (e.g., Howard, Nickels, Coltheart & Cole-Virtue, 2006; Runnqvist, Strijkers, Alario & Costa, 2012), explored in the study of Costa et al. (2009), originates at the conceptual level, not the lexical one (e.g., Llorens, Trebuchon, Ries, Liegeois-Chauvel & Alario, 2014; Navarrete, Mahon & Caramazza, 2010). Likewise, whether the locus of semantic interference in picture-word interference and blocked cyclic naming paradigms, as explored in Aristei et al., (2011), is located during lexical selection is currently being questioned by several researchers (e.g., Caramazza & Costa, 2000; 2001; D'Hooge & Hartsuiker, 2011; Finkbeiner & Caramazza, 2006; Janssen, Hernandez-Cabrera, van der Meij & Barber, 2014; Mahon, Costa, Peterson, Vargas & Caramazza, 2007; but see, e.g., Abdel Rahman & Melinger, 2009; Piai, Roelofs & Schriefers, 2014b; Roelofs, Piai & Schriefers, 2013). Hence, consistent with the logic that psycholinguistic phenomena which have an ambiguous origin cannot be used to test the time-course of speech production, it follows that evidence from semantic interference paradigms cannot be considered (for now) to support I&L's spatiotemporal map (and similar objections can be made with respect to the time estimates of other word production components as well).

The pitfall of whether the data support the model or the model the data is often hard to disentangle. One of the more important sources of input for I&L's spatiotemporal model are MEG studies of object naming. This is because MEG provides both spatial and temporal resolution and can therefore map both dimensions of the model at the same time. As mentioned earlier, several studies observed that in the first 200 ms of processing source activations are maximal in occipital brain regions, between 200-400 ms in temporal and parietal regions, and between 400-600 ms in frontal regions (e.g., Hulten et al., 2009; Levelt et al., 1998; Liljestrom et al., 2009; Maess et al., 2002; Salmelin et al., 1994; Soros et al., 2003; Vilha et al., 2006; see also Llorens et al., 2011). These observations have served both as input to I&L's model (e.g. Levelt et al., 1998; Maess et al., 2002; Salmelin et al., 1994) and, in the case of the later studies (e.g., Hulten et al., 2009; Liljestrom et al., 2009; Soros et al., 2003; Vilha et al., 2006), as confirmation of the model (Indefrey, 2011). However, a few important remarks have to be taken under consideration. First, the MEG studies traced the neural generators behind the brain's magnetic response to object naming in general, without the presence of experimental contrasts which could isolate one (or more) of the linguistic components in word production (e.g., Miozzo, Pulvermuller & Hauk, 2014; Strijkers & Costa, 2011). Strictly speaking, to associate temporal

and parietal source activations to lexical and phonological processes, and the frontal source activations to phonetic and articulatory processes, those studies had to rely on a-priori assumptions about when the different representational codes of word production are thought to become activated in the course of speech planning (e.g., Levelt et al., 1991; 1999; Roelofs, 1992; 1997). The same serial time-estimates which serve as starting point to attribute the spatial and temporal correlates of word production in I&L's map. Secondly, to trace the cortical dynamics elicited by naming objects, a localization strategy was applied in which for a given time-frame the most active cortical source (activation maxima) was identified. Taking into account that the task of object naming is a visually driven motor act, it is perhaps not so surprising that the time-window closest to stimulus presentation shows the strongest occipito-temporal activation, and the time-window closest to articulation shows the strongest fronto-central activation (Broca and the motor cortex). Demonstrating that an object naming tasks starts with visual processes and ends with a motor act, does therefore not offer evidence that word activation in the brain follows a stepwise progression from posterior towards anterior areas. In fact, several MEG studies on object naming did find some activations in Broca's region and the motor cortex within 200 ms (e.g., Levelt et al., 1998; Liljestrom et al., 2009; Soros et al., 2003; see also Miozzo et al., 2014; Strijkers, Costa & Pulvermuller, under review), but just not with their peak activation in that time-window.

A potential way to circumvent or alleviate the problems presented above is to accumulate spatial and temporal data elicited by several psycholinguistic phenomena (e.g., Graves et al., 2007; Miozzo et al., 2014; Strijkers et al., 2010; 2013; Strijkers & Costa, 2011; Wilson et al., 2009). For instance, within the same study Strijkers and colleagues (2010; see also Strijkers et al., 2013) explored the electrophysiological correlates of the word frequency effect, the cognate effect (i.e., faster naming latencies for translation words which share phonology compared to those which do not; e.g., Costa, Caramazza & Sebastian-Galles, 2000) and the language effect (i.e., faster response times for naming pictures in one's first compared to second language; e.g., Ivanova & Costa, 2008). Given that word frequency is known to correlate with conceptual processes (e.g., Laganaro, 2014; Snodgrass & Vanderwart, 1980), based on the ERP effect of that variable alone it is unclear whether the observed time-course of 180 ms reflects the onset of lexical access or rather earlier correlated conceptual effects (e.g., Hanulova et al., 2011; Indefrey, 2011). However, since the other two variables tested in the study elicited identical ERP effects within a similar time frame, a conceptual origin of the observed word frequency effect could be ruled out. That is because for both the cognate effect (defined

by the amount of sound overlap between translations) and the language effect (defined by an identical concept linked to two different words) there is no obvious link with conceptual processes.

Another possibility to test the fitness of a model more explicitly is to rely on spatiotemporal 'markers' which are independently assessed from any given model (e.g., Costa et al., 2009; Edwards et al., 2010; Fedorenko, Duncan & Kanwisher, 2012; Graves et al., 2008; Heinks-Maldonado, Nagarajan & Houde, 2006; Peeva et al., 2010; Pulvermuller, 2013; Strijkers et al., 2011a; under review; Strijkers & Costa, 2011; Ylinen et al., 2014). For example, Peeva et al. (2010) relied on an imaging protocol called fMRI adaptation to distinguish brain regions involved in sublexical phonological processes during speech production. fMRI adaptation refers to the observation that the haemodynamic response to repeated items decreases. Importantly, the signal reduction following repetition is shown to be stimulus-specific and affect those brain regions which were most active during initial presentation, suggesting functional selectivity of the anatomical region where signal attenuation occurs (e.g., Grill-Spector, Henson & Martin, 2006). Based on this rationale, Peeva and colleagues (2010) were able to assess, in a theory-independent manner, which brain regions showed selective adaptation to phonemic repetitions (left SMA, pallidum, posterior STG and superior lateral cerebellum) and which to syllable repetitions (left ventral premotor cortex) (a result which is both in partial agreement and disagreement with the sources identified in I&L).

A third potential promising avenue is to design experimental settings where distinct spatial and temporal components of word production are explored within the same participants and experiment (e.g., Laganaro et al., 2009a; 2009b; Papoutsis et al., 2009; Sahin et al., 2009); ideally in a setting where immediate and overt naming is required (e.g., Acheson et al., 2011; Laganaro et al., 2013; Miozzo et al., 2014; Schuhmann et al., 2012; Strijkers et al., under review). For instance, in a repetitive TMS study of overt object naming, Schuhmann and colleagues (2012) stimulated 3 different 'nodes' (LIFG, left MTG and left STG) of I&L's spatiotemporal map at variable time frames. The chronometric TMS design allowed them to establish at which points in time neural activation in those three brain regions are functionally relevant for the naming behavior, hereby offering an explicit test of whether those brain regions indeed follow the proposed temporal progression. The data partially supported the model as well as partially disagreed with the model: In the MTG, stimulation around 225 ms after picture presentation significantly slowed down the naming latencies and then again around 400 ms; in the LIFG, the behavioral performance was disrupted by electrical stimulations applied around 300 ms after picture presentation; and in the STG, object naming went slower

when stimulated around 400 ms after stimulus onset. In other words, the study revealed that all three brain regions are relevant for language production and causally affect production at different time points, but suggest a different sequence of activation flow between those brain regions compared to I&L's model (2004).

We believe that the above suggestions (data accumulation of psycholinguistic factors, independent spatiotemporal markers and multi-component research) may positively contribute to constrain and advance upon our knowledge about the spatiotemporal dynamics of speaking. At present, when considering the data coming from studies which follow such more independent approach, the evidence regarding the claims made by I&L's model are mixed: Some parts of the model have been reliably supported (e.g., Acheson et al., 2011; Costa et al., 2009; Graves et al., 2007; Laganaro et al., 2009a; 2009b; Papoutsis et al., 2009; Peeva et al., 2010; Sahin et al., 2009; Schuhmann et al., 2012; Strijkers et al., 2010; 2013; Wilson et al., 2009), but certainly not all (e.g., Costa et al., 2009; Edwards et al., 2010; Graves et al., 2007; Miozzo et al., 2014; Peeva et al., 2010; Sahin et al., 2009; Schuhmann et al., 2012; Strijkers et al., 2010; 2013; under review; Wilson et al., 2009) (see section 3 as well).

2. What is the role of top-down processing on the spatiotemporal dynamics?

It is well accepted that language production is an intentional act (Levelt, 1989). However, how this goal-directed behavior is implemented and how it may affect the underlying spatiotemporal dynamics of speech production has been largely neglected. This is not to say that the existing models negate the existence of top-down influences, but rather that they remain to a large extent silent of how such intention alters the dynamics of word production.

At present, most models assume that a word intended for production is accessed through spreading activation; a principle according to which information necessarily circulates between strongly connected representations (e.g., Collins & Loftus, 1975). This implies that an activated concept will automatically spread its activation to the lexical representation with which it is connected, regardless whether there is the intention to name that concept or not (e.g., Caramazza, 1997; Dell, 1986; Dell et al., 1997; Levelt et al., 1999; Navarrete & Costa, 2005; but see Bloem & La Heij, 2003). In this view, linguistic intentions, such as to engage in speech production or not, only affect language processing after the automatic access to words (referred to as reactive top-down processing; e.g., Roelofs, 2003; 2008). Note that also I&L's spatiotemporal map (2004) enters in this class of models. Although the authors remain silent about top-down processing, the

architecture is based on Levelt et al.'s model of lexical access (1999), which allows for reactive control in function of a speaker's linguistic intentions.

Evidence supporting the notion of automatic spreading activation in language production is abundant and mostly stems from Stroop-like tasks (Stroop, 1935) where distractor information which a speaker does not intend to verbalize nonetheless influences the speed of target naming (e.g., Bles & Jansma, 2008; Costa & Caramazza, 2002; Glaser & Glaser, 1989; Levelt et al., 1991; Morsella & Miozzo, 2002; Navarrete & Costa, 2005; Roelofs, 1992). These observations have been understood as revealing that activated concepts automatically trigger their corresponding lexical representations, regardless whether a speaker has the intention to verbalize those concepts (they may even be harmful for performance as in the case of Stroop-interference). It is only after the conceptual system has spread its activation to the language system that the speech-irrelevant information (distractor) is suppressed and the speech-relevant information (target) can be selected. But does it also follow from this evidence that goal-directed top-down influences in speech production are therefore restricted to late, reactive control only? Following the classical proposals concerning automaticity and control (e.g., Posner & Snyder, 1975), where the former is fast and occurs in the absence of top-down modulations and the latter is slow and consciously engaged, it would. However, over the last few decades, the rather dichotomic classification between automaticity and top-down control has come under strong scrutiny, in particular in the field of vision science. Today, there is ample evidence showing that even the earliest stages in the visual processing stream, classically defined as purely sensory-driven, are affected by attention (e.g., Bar, 2003; Corbetta & Shulman, 2002; Desimone & Duncan, 1995; Engel, Fries & Singer, 2001; Gilbert & Li, 2013; Hillyard, Vogel & Luck, 1998; Kastner, Pinsk, De Weerd, Desimone & Ungerleider, 1999). This has led to a radical change about the role of attention and goal-directed control in vision; one where the brain interprets input through the dynamical and immediate interplay between stimulus-driven bottom-up and attention-driven top-down processes. Importantly, this view can count by now on empirical demonstrations in other, higher-order domains of cognition as well, such as spoken word perception (e.g., Garagnani, Shtyrov & Pulvermuller, 2009; Heald & Nusbaum, 2014), visual word recognition (e.g., Chen, Davis, Pulvermuller & Hauk, 2013; Ruz & Nobre, 2008; Strijkers, Bertrand & Grainger, 2015), semantic memory (e.g., Kiefer, 2007; 2012) and consciousness (e.g., Kiefer, 2007; 2012; Naccache, Blandin & Dehaene, 2002).

Given the above, it seems reasonable to assume that also for the higher-order and conscious skill of

speaking, top-down processes will not be restricted to late, reactive modulations. Recently, Strijkers and colleagues (2011a) put this to the test. They compared the electrophysiological signature to word frequency, an established electrophysiological index for the onset of lexical access (e.g., Graves et al., 2007; Sahin et al., 2009; Strijkers et al., 2010), between a task where speakers had the conscious intention to verbalize a stimulus (overt object naming) and a task where no such intention was present (non-verbal object categorization). The spreading activation view on lexical access would thus predict an early word frequency effect in both tasks (i.e., automatic spreading activation between an activated concept and its corresponding word), followed afterwards by task-specific dissociations in the ERPs (i.e., reactive top-down control). We indeed observed a word frequency effect in the ERPs for both tasks. Crucially, however, while the ERP word frequency effect already occurred 150 ms after stimulus presentation (in the P2-range) during object naming, such effect was present much later in the object categorization task (350 ms after picture onset in the N400-range). The presence of a word frequency effect for a task where no speech is required confirmed the notion that activated concepts automatically trigger (some) word knowledge (the principle of spreading activation as demonstrated in Stroop-like production tasks), but the delay of 200 ms compared to the word frequency effect in the verbal task suggests that initial word activation is not free from top-down modulations. Instead, the data concur with the notion that top-down processes alter bottom-up processing in a proactive (i.e., prior to stimulus presentation) manner by enhancing the sensitivity of the (automatic) target-relevant pathways and decreasing the sensitivity of the (automatic) target-irrelevant pathways, just as advocated in other domains of cognition (e.g., Bar, 2003; Corbetta & Shulmann, 2002; Desimone & Duncan, 1995; Engel et al., 2001; Gilbert & Li, 2013; Heald & Nusbaum, 2014; Hillyard et al., 1998; Kastner et al., 1999; Kiefer, 2007; 2012; Naccache et al., 2002). Following this view, Strijkers and colleagues (2011a) concluded that the difference in the emergence of the word frequency effect between tasks revealed that the intention to speak strengthens the lexico-semantic pathway prior to stimulus presentation (resulting in speeded access = automatic spreading activation + top-down facilitation). This facilitation will not be present in the categorization task since such task does not require word retrieval and consequently cannot benefit from additional top-down enhancements (resulting in slower access = automatic spreading activation alone) (and for compatible evidence in speech production see also: Baus, Strijkers & Costa, 2013; Schotter, Ferreira & Rayner, 2013; Strijkers, Yum, Grainger & Holcomb, 2011b).

The observation that the speed of lexical access (and thereby the underlying neural structures which

support it) can fluctuate in function of top-down processes may pose a problem with respect to the attribution of time and space in I&L's model for two reasons. First, it points out that tasks that do not rely on overt naming might lose important temporal and spatial information that does form part of the signal during intentional language production. This is relevant since many of the spatiotemporal data used to inform the model came from metalinguistic tasks, where intention is confined to a specific level of representation, and non-verbal tasks, where there is no intention to speak. Secondly, one may question whether the spatiotemporal estimates of a strictly feedforward word production model will be accurate when considering early top-down modulations (see Fig. 2). The answer will depend on whether top-down biasing signals have a global effect on the language system or could modulate parts of the system specifically. So far, what the results show is that the intention to verbalize an activated concept affects the access to the lexical representations connected to that concept. It does not show whether the influence of top-down modulations goes beyond this global effect of 'placing the brain in a naming state'. If this is all what the top-down modulation does, then the core of I&L's spatiotemporal map (2004) can be very well maintained (i.e., when the speech intention is 'on', cortical activity can flow through the system in the manner described by I&L). However, if the role of preparatory top-down modulations are more specific and capable of enhancing particular words in function of our speech intentions, context and attentional resources, then a spatiotemporal map embedded in a feedforward framework might be limited (see Fig. 2).

Although at present there is no empirical evidence of the latter possibility, there are some good reasons to assume that this will likely be the case. In the field of vision it has been repeatedly demonstrated that top-down processes selectively guide stimulus activation to specific attributes. Goal-directed control and visual attention can enhance the neural gain of relevant spatial locations in a scene (e.g., Motter, 1993), pre-activate particular features of an object (e.g., Corbetta, Miezin, Dobmeyer, Shulman & Petersen, 1990; Giesbrecht, Woldorff, Song & Mangun, 2003; Kastner et al., 1999), bias towards task-relevant attributes of the same perceptual input (e.g., Cromer, Roy & Miller, 2010; Roelfsema, Lamme & Spekreijse, 1998), discriminate between complex object categories (e.g., O'Craven, Downing & Kanwisher, 1999; Peelen, Fei-Fei & Kastner, 2009) and even anticipate specific objects for recognition (e.g., Chelazzi, Miller, Duncan & Desimone, 1993; O'Craven & Kanwisher, 2000; Peelen et al., 2009; Ullman, 2007) (for reviews see: e.g., Corbetta & Shulman, 2002; Gilbert & Li, 2013; Kastner & Ungerleider, 2000; Kanwisher & Wojciulik, 2000). The diverse and rich nature of these predictive top-down modulations serve the common goal to

enhance the speed and efficiency of visual processing in accordance with the available attentional resources and goal-directed behavior at play (e.g., Desimone & Duncan, 1995); a property of the visual system which is crucial to rapidly adapt to a complex environment.

Like in vision, language can serve distinct intentions and occur in distinct contexts. For instance, a single concept (e.g., DOG) can be expressed with many different words (e.g., *animal, dog, dalmatian, 'spot'*) depending on which message we wish to convey (e.g., specific or general), a single word (e.g., *work*) can have different linguistic use depending on our grammatical intentions (e.g., noun or verb), the same concepts and words (e.g., *a proper name*) could sound differently depending on the valance we want to communicate (e.g., angry or happy) and different sounds (e.g., *dog or perro*) can have the same meaning depending on the language we choose to speak in (English or Spanish). In addition, these distinct linguistic intentions will often be shaped further by the different contexts in which language behavior can occur. Speaking with your boss is likely to rely on a somewhat different "lexicon" than talking to your friends in a bar and on a rainy day words such as *umbrella, clouds, wet* and *inside* are reasonable to anticipate in conversation, while on a sunny day it is more likely to talk about *tan, beaches, suncream* and *margarita's*. If top-down influences can bias the activation of semantic, lexical or phonological knowledge linked to specific word categories or even condense groups of words, then also their spatiotemporal dynamics will be context- and task-dependent. In language comprehension some empirical evidence has been reported which shows that word recognition is indeed subject to such context-dependent semantic, grammatical and phonological predictions (e.g., Delong, Urbach & Kutas, 2005; Federmeier & Kutas, 1999; Heald & Nusbaum, 2014; Kutas & Federmeier, 2000; 2011; Tian & Poeppel, 2010), and that the time-course of cortical area activations of linguistic knowledge can be task-specific (e.g., Chen et al., 2013; Garagnani et al., 2009; Ruz & Nobre, 2008; Strijkers et al., 2015). We predict that also in speech production context-dependent prediction and task-specific intention can affect the cortical dynamics of speaking. If this prediction turns out to be correct, then current models of speech production may provide an accurate spatiotemporal progression of word production components in some, but not most cases (see Fig. 2).

Insert Fig. 2 around here

In sum, by now there is compelling evidence that language production, specifically, and language

processing, in general, is subject to top-down modulations. Far less clear is how the inclusion of proactive and predictive top-down processes will affect the spatiotemporal correlates of word production (see Fig. 2). It is a challenge for future research to see whether top-down mechanisms can modulate the retrieval of words in a proactive and specific manner. If so, rather than conceiving word production in the brain as a general-purpose, context-invariant feedforward process in hierarchically organized neural architectures, it will need to be regarded as an adaptive and context-dependent process where 'bottom-up' (conceptual) driven word retrieval and top-down intentions, expectations and experiences dynamically interact. In this case, also the spatiotemporal correlates elicited by the mental operations of word production will need to be regarded as context-dependent and adaptive to a speaker's intentions.

3. Can we map psycholinguistic stages onto the brain in a one-to-one fashion?

One of the most appealing features of I&L's (2004) spatiotemporal map is its near perfect symmetry between psycholinguistic theory and the brain. We have already commended this symmetry because of its high falsifiability and its theory-driven approach. Concretely, the core mental operations of speaking, beginning with conceptual processing and ending with an articulatory motor program (e.g., Levelt et al., 1999), are plotted onto the brain from posterior towards anterior regions in an almost perfect one-to-one fashion. But does such exercise, as valuable and transparent as it may be, provide a realistic perspective on the neurobiological basis of language (and in general how the brain computes cognition)? Current advances in neuroscience suggest that such modular-like structure may be insufficient to fully capture the complexity between brain processing and language processing.

The approach presented above follows classical views in cognitive neuroscience where sensorial, motor and cognitive events are argued to come about in the brain through sequential bottom-up propagation of information along a hierarchy of dedicated processing centers (e.g., Barlow, 1972; Fodor, 1983; Hubel & Wiesel, 1965; Marr, 1982; Shallice, 1988). While the alignment of linguistic units to brain regions certainly represented significant progress (e.g., Price, 2010), the evidence is correlational, and does not offer mechanistic explanations (Poeppel, 2012; Pulvermuller, 2005; Pulvermuller & Fadiga, 2010). Furthermore, it suffers from a principled problem where linguistic concepts are mapped to brain functioning without clear linking hypotheses between the two distinct domains of abstraction (e.g., Poeppel, 2012; Poeppel, Emmorey,

Hickok & Pylkkanen, 2012)³. Here we would like to focus on one aspect of this problem, namely whether modern advances in neuroscience can indeed be aligned (mapped) with serial processing hierarchies of psycholinguistic function.

In a neurobiological context “serial” means that the activation time-course of a neuron (neural populations) sending a message (i.e., action potential) and the activation time-course of a receiving neuron (neural populations) are functionally separated in time. The term “hierarchical” means that the representational contents carried by individual neurons increases in complexity as one moves up the processing stream (e.g., Barlow, 1972; Shallice, 1988). This type of neural communication is thus defined by the structure of their connections, and the attribution of word production (or any cognitive) components should follow that anatomical connectedness from lower toward higher cortical regions of representation. While such forward anatomical connectedness likely plays an important role, this is not the only form of neural communication. Above and beyond the traditional stance of *fixed forward anatomical* communication, evidence has been brought forward that there are cortical mechanisms in place allowing for *flexible effective* communication (e.g., Fries, 2005); a tenet thought to be crucial in explaining the adaptive nature of the brain and its capacity to integrate many different types of information (and arguably of more importance to code for cognition in the brain compared to fixed forward neural communication; e.g., Tsodyks & Markram, 1997; Fries, 2005). For instance, neural communication is not restricted to *anatomical connectivity* between adjacent ‘input’ and ‘output’ neurons (or clusters), but can be achieved through *neural coherence*. That is, the synchronous activation of different neural populations, which can be located at distant sources in the brain and without necessarily direct connections between them, could reflect time-dependent neural communication and represent specific functional states during cognitive processes (e.g., Engel et al., 2001; Engel, Maye, Kurthen & Konig, 2013; Fries, 2005; Singer & Gray, 1995; Varela et al., 1991; 2001). Similarly, the brain is not solely equipped with *forward connections* between neurons and cortical areas, but also consists of a vast amount of *feedback projections* (e.g., Porrino, Crane & Goldman-Rakic, 1981; Rempel-Clower & Barbas, 2000). These are argued to be crucial for top-down processing and prediction. In addition, it has been shown that the firing of neurons is not *linear and monotone* (i.e., not time-constant as

³ Note that this *mapping problem* (cf. Poeppel, 2012) is not restricted to serial hierarchical models, but due to their reductionist nature between the elements of language (e.g., semantics versus phonology; or comprehension versus production) and the elements of the brain (e.g., neurons, brain areas or hemispheres), those models do represent an extreme example of the problem.

assumed in I&L), but rather have a strong and abrupt onset (i.e., *ignition*) followed by a longer period of sustained activity (i.e., *reverberation*) (e.g., Braitenberg, 1978; Hebb, 1949; Pulvermuller, 1999; Singer & Gray, 1995). Interestingly, these distinct firing patterns could relate to functionally different processing states where ignition may serve the recognition of a mental representation and reverberation important second-order processes such as reprocessing, memory consolidation, conscious processing and decision-making (e.g., Dehaene & Changeux, 2011; Dehaene, Changeux, Naccache, Sackur & Sergent, 2006; Engel & Singer, 2001; Friston, 2010; Fuster, 2003; Pulvermuller, 1999; 2002; Pulvermuller et al., 2009; Pulvermuller & Garagnani, 2014).

While the dominant neurobiological models of language production largely neglect physiological brain properties beyond forward anatomical connectedness, the implementation of more flexible neural communication is becoming increasingly popular in language comprehension models. Here we will discuss one such brain language model with as aim to see whether similar adaptations to the field of language production could be fruitful⁴. A popular current formalization of the language network is based on cell assembly theory (e.g., Braitenberg, 1978; Hebb, 1949; Mesulam, 1990; Singer & Gray, 1995), where words are represented as large-scale distributed neural circuits that emerge in a rapid and near-simultaneous fashion (e.g., Buszaki, 2010; Fuster, 2003; Mesulam, 1990; Pulvermuller, 1999; 2005; Pulvermuller & Fadiga, 2010; Pulvermuller & Garagnani, 2014). Concretely, through language development the different constituents of words (e.g., semantic, lexical and phonological knowledge) reflected in action, perception and domain-general (integrating) brain systems are thought to group together forming functional assemblies that represent a word as a whole (gestalt) (see Fig. 3). For example, words related to body-parts (e.g., ‘lip’ and ‘hip’) will involve assemblies linking visual properties (e.g., form, size, color, etc.) in temporal cortex with motor properties (e.g., neurons responsible for moving the lips or the hips) in fronto-central cortex. Assembly-overlap in visual and motor properties denotes generalizability between the words (e.g., common semantic category) and assembly-dissociations (e.g., more inferior face-somatotopy in precentral gyrus for

⁴ It is important to point out that also different (hierarchical) model implementations that go beyond simple serial forward connectivity could be equally fruitful to consider and may also capture some of the data we will discuss in this section. For instance, dual- or multi-stream models of language comprehension (e.g., Rauschecker & Tian, 2000 ; Hickok & Poeppel, 2007 ; Kotz & Schwarze, 2010 ; Friederici, 2012), and in particular its recent adaption to speech production (e.g., Hickok, 2012 ; 2014a; Pickering & Garrod, 2013) are promising alternative accounts to explore in language production. However, we decided to focus here on cell assembly models because they represent a more contrastive example with the traditional serial models, being non-hierarchical in nature (see e.g., Strijkers, 2014 ; Hickok, 2014b), and because they also make explicit temporal predictions, unlike the dual-stream models.

'lip' versus more superior leg-somatotopy for 'hip') denotes specificity between the words. Similarly, topographic dissociations in acoustics (auditory cortex) and place of articulation (motor cortex) ensure that sensorimotor assemblies differentiate words with phonological overlap (e.g., more dental-related motor recruitment for 'lip' compared to more glottal-related motor recruitment for 'hip'). This model has been especially fruitful in explaining the very rapid and simultaneous activation of phonology, syntax and meaning in spoken word recognition and reading (e.g., Chanceaux, Vitu, Bendahman, Thorpe & Grainger, 2012; MacGregor, Pulvermuller, van Casteren & Shtyrov, 2012; Näätänen, Paavilainen, Rinne & Alho, 2007; Pulvermuller et al., 2009), the parallel recruitment of frontal and temporal brain regions when comprehending language (e.g., Moseley, Pulvermuller, Shtyrov, 2013; Pulvermuller, 2013; Pulvermuller & Fadiga, 2010), the functional role of the motor cortex in speech perception (e.g., Boulenger, Hauk & Pulvermuller, 2009; D'Ausilio, 2009; Hauk et al., 2004; Pulvermuller & Fadiga, 2010; Pulvermuller et al., 2014; Wilson, Saygin, Sereno & Iacoboni, 2004) and language pathologies driven by sensorimotor lesions (e.g., Bak et al., 2001; Boulenger et al., 2008; Cotelli et al., 2006; Garagnani et al., 2008).

Although a neural assembly model explains several current neurophysiological findings in language comprehension that on itself is not necessarily a reason to start implementing such more flexible physiological principles in language production as well. However, in recent years, certain findings are difficult to account for with a local and sequential architecture as in I&L (2004), and may be better served by models that argue for more flexible spatiotemporal dynamics. In what follows we will highlight a few of such observations.

In section 1 we have already mentioned the chronometric TMS study on object naming of Schuhmann and colleagues (2012). In this study it was shown that stimulation of the LIFG affected the naming latencies prior to stimulation in the STG, and stimulation to the MTG influenced naming behavior at two different time-points, early (prior to LIFG) and late (after STG). The finding that MTG activation comes prior to STG and LIFG activation is consistent with I&L's predictions (2004). However, the fact that the LIFG is functionally relevant prior to the STG and that the MTG displays causal involvement at two different moments in time are not consistent with their predictions. Especially the latter result suggests that the time-course of cortical area activations is not restricted to a single time frame, but shows patterns consistent with the notions of feedback projections and reentrant activity. Another compelling example revealing that brain regions might not have a single time-course linked to a single linguistic function comes

from the work of Sahin and colleagues (2009; see also Edwards et al., 2010; Peeva et al., 2010). In this study a sequential brain response for lexical, grammatical and phonological encoding (which corresponded reasonably well with the temporal estimates provided by I&L) was observed. However, these sequential responses to different linguistic knowledge were observed within the same brain area, namely the LIFG (affecting different subparts of this structure). This finding is inconsistent with I&L (2004), who argue that lexical and grammatical processing should be restricted to temporal regions and involvement of the LIFG should be restricted to phonological processes occurring late in the course of speech preparation.

The above observations question the fixed spatial structure attributed to linguistic function in I&L's spatiotemporal map (2004). But also with respect to the fixed temporal structure of the model inconsistent findings have been reported. As discussed in Section 1, Strijkers and colleagues (2010; see also Strijkers et al., 2013) observed that the lexical frequency effect, the cognate effect and the language effect all produced the same early (within 200 ms) electrophysiological deviations (P2-range). Especially the latter two effects were surprisingly early given that the dimension which defines these variables is phonological in nature (the only difference between a cognate and non-cognate is whether they have overlapping sounds in different languages; e.g., the Spanish-English cognate pair *guitarra-guitar* vs. the non-cognate pair *tambor-drum*). This finding might index, analogous to findings in comprehension (e.g., Pulvermuller et al., 2009), the near-simultaneous retrieval of lexical and phonological knowledge in the course of speech preparation⁵. Compatible results including also spatial information have been reported by Miozzo and colleagues (2014) in a recent MEG study of overt object naming. The authors performed a multiple-linear regression analyses on the neuromagnetic brain response of variables related to lexico-semantic (specific semantic features and action features) and phonological processing (word form features). At around 150 ms after object presentation modulations elicited by lexico-semantic variables became manifest in the left fronto-temporal cortex. In the same latency-range modulations elicited by the phonological variable became manifest in the left posterior MTG (previously linked to word form processing; e.g., Graves et al., 2007; Wilson et al.,

⁵ Note that these results do not necessitate an explanation in terms of parallel retrieval dynamics. In fact, one of the explanations given by the authors is that all three variables occur during lexical processing (and therefore display a similar electrophysiological pattern). From this perspective, the cognate and language effects could become visible already during lexical instead of phonological retrieval if there is a high degree of (long-lasting) interactivity between the lexico-semantic (lemma) and phonological (lexeme) processing layers (for a detailed account please consult Strijkers et al., 2010; Strijkers & Costa, 2011). In any event, regardless whether the results are better explained through parallel or interactive processing, the important conclusion for present purposes is that they cannot be captured by the serial dynamics as portrayed in I&L (2004).

2009). From these results it was concluded that access to phonological information might become available in parallel to lexico-semantic information⁶.

Insert Fig. 3 around here

The results discussed in the previous two paragraphs are difficult to explain by a sequential and modular-like brain model of language production. Brain regions argued to be involved in a particular linguistic operation (such as the LIFG and MTG) were demonstrated to be sensitive to multiple levels of processing (e.g., Sahin et al., 2009; Schuhmann et al., 2012). Lexico-semantic processes were not restricted to the temporal cortex, but recruited a distributed fronto-temporal network simultaneously (e.g., Miozzo et al., 2014; Strijkers et al., under review). And, perhaps most problematic, well-established brain regions involved in lexico-semantic (e.g., MTG) and phonological/articulatory processing (e.g., pMTG, STG, motor cortex) did not show a sequential activation pattern, but ignited in parallel within 200 ms (e.g., Miozzo et al., 2014; Strijkers et al., under review). Although these findings need replication and more data on this topic in general is desirable, they suggest that I&L's spatiotemporal map is incomplete. In particular, it seems that the model might be too rigid both in terms of spatial and temporal functionality. That is, a reasonable modification could be that central (and highly interconnected) brain regions such as the LIFG and the MTG are allowed to have more than one function during language production (e.g., Fedorenko & Thompson-Schill, 2014; Hagoort, 2005; Hagoort & Levelt, 2009; Pulvermuller, 2013; Sahin et al., 2009; Schuhmann et al., 2012;). Likewise, in terms of timing, faster propagation dynamics seem at play. Allowing for (fast) interactivity and cascading between representational layers could already be helpful in this regard (e.g., Caramazza, 1997; Dell, 1986; Dell, Schwartz, Nozari, Faseyitan & Coslett, 2013).

That said, it is an open issue whether a 'dynamical upgrade' of the original model will be sufficient (especially for the data suggesting parallel retrieval of the different word production components in fronto-

⁶ Recently, these findings received additional strong support and got extended to later phonemic and articulatory processes as well in an anatomically constrained MEG study of overt object naming (Strijkers et al., under review). In that study it was demonstrated that a lexico-semantic variable (word frequency) activated within 200 ms a fronto-temporal network (MTG and LIFG), along with a simultaneous dependency of local brain responses in the motor cortex and the STG with respect to the particular articulatory movement (*lip vs. tongue*) required to utter different initial speech sounds of a word (e.g., *Monkey vs. Donkey*). This specific nature of the spatiotemporal pattern responding to a word's frequency and initial phoneme provided compelling evidence that in the course of speech planning lexico-semantic and phonological-phonemic processes emerge rapidly together, drawing in parallel on temporal and frontal cortex.

temporal circuits as discussed above). If not, it could be fruitful to take into consideration neurobiological accounts that integrate both parallel distributed and local sequential processing. Indeed, the observations that lexico-semantic and phonological properties of to-be-uttered words are supported by fronto-temporal brain circuits which become activated rapidly in a near-simultaneous fashion, does not only find an elegant explanation in neural assembly models of language, it is in fact a necessary prediction of those models. Furthermore, in such models, after parallel ignition of the ‘word web’, activity can reverberate in local parts of the assembly producing well-timed spatiotemporal sequences. These could be responsible for second-order language processes such as semantic and syntactic integration or timed articulation (e.g., Buszaki, 2010; Pulvermuller, 1999; 2002; Pulvermuller et al., 2009; Pulvermuller & Garagnani, 2014; Strijkers et al., under review) (see Fig. 3). This is important since, as mentioned in the Introduction, there is much evidence demonstrating that there is a sequential and local component to language production: The MTG likely plays a role in lexico-semantic retrieval (e.g., Indefrey, 2011; I&L, 2004; Graves et al., 2007; Schwartz et al., 2009; Schuhmann et al., 2012), the LIFG seems indeed to be involved in phonology and syllabification (e.g., Indefrey, 2011; I&L, 2004; Papoutsis et al., 2009; Sahin et al., 2009; Schuhmann et al., 2009; 2012), sequential brain responses to different linguistic knowledge have been repeatedly observed (e.g., Indefrey, 2011; I&L, 2004; Laganaro et al., 2009a, 2009b; Sahin et al., 2009; Schuhmann et al., 2012) and activation maxima slowly progressing from posterior towards anterior brain regions are frequently reported in response to object naming (e.g., Hulstén et al., 2009; Indefrey, 2011; I&L, 2004; Liljestrom et al., 2009; Salmelin et al., 1994; Soros et al., 2003; Vilha et al., 2006). However, the linguistic functionality of their time-course may vary considerably depending on the neurobiological conceptualization of the language production architecture.

For instance, those models solely relying on sequential hierarchical brain dynamics argue that activation in the motor cortex is only functionally involved near the end of speech preparation in order to trigger a motor program to the vocal folds (e.g., Indefrey, 2011; Indefrey & Levelt, 2004). In contrast, a brain language theory built upon Hebb’s concept of cell assemblies, argues that neurons in the motor cortex will have (at least) two functional contributions to the preparation of a speech act: In an early phase of speech preparation (ignition) as representational part of the word assembly (action semantics and motor phonology), and in a later phase (reverberation) to trigger a sequential articulatory plan (e.g., Pulvermuller, 1999; 2005; 2013; Pulvermuller & Fadiga, 2010). While assembly ignition could explain the fronto-temporal parallel

dynamics observed in response to different word production components, reverberation could explain the sequential and local brain responses observed in neurolinguistic research and serve as a key neurobiological mechanism for sequential behavior. For example, if a word is represented in the brain as a distributed cell assembly that ignites in parallel, then how are we capable of correctly uttering an intended word as “BAL” instead of its undesired mirror image “LAB”? One possible answer provided by neural assembly theory is that after the synchronous ignition, structured reverberatory activation embedded in the circuit (drawing on motor cortex) ensures the timed articulation of the correct sequence /b/-/a/-/l/ instead of the incorrect sequence /l/-/a/-/b/ (which has a different internal assembly structure linked to it and thus a different reverberatory sequence) (e.g., Pulvermuller, 2002; Pulvermuller et al., 2014; Strijkers et al., under rev.). Future research investigating such linking hypotheses between brain mechanics and linguistic mechanics, specifically, and investigating which type of neural communication (e.g., sequential or parallel) and neuroanatomical organization (e.g., hierarchical or non-hierarchical) underpin speech preparation in the brain, more generally, will help us to substantially advance in our understanding of how basic brain properties sustain language production.

Conclusion.

According to the dominant neurocognitive model of speech production, uttering words is a sequential and feedforward process where each linguistic component is represented by a specialized neural system and a specific time-course of activation (I&L, 2004). While this work constitutes an impressive progress concerning our understanding of the cortical dynamics underlying word production, in this opinion piece we have questioned whether such modular-like dynamic is sufficient to capture the complexity between brain processing and language processing. We pointed out that (1) despite a considerable amount of evidence being consistent with I&L’s meta-analysis (2004), studies which explicitly test the serial and local hypotheses of the model remain all in all scarce. We asked (2) how the inclusion of top-down processing and prediction during speech preparation might modulate the spatiotemporal correlates of word production components. And finally, we highlighted that (3) advances in neuroscience suggest that above and beyond fixed forward anatomical communication, other neurophysiological processing principles of the brain, such as neural coherence, feedback connectivity and reverberatory activation, may serve a role in the production of speech. We concluded that still several aspects of I&L’s spatiotemporal map (2004), in particular its

uniquely sequential time-course and region-specific linguistic functionality, remain open issues. Future research endeavors dedicated to unraveling the topics addressed here will hopefully help to elucidate further the cortical dynamics underpinning speech production.

Acknowledgments

K.S. was supported by the Intra-European Fellowship (FP7-PEOPLE-2012-IEF) of the Marie Curie Actions (grant number: 302807) awarded by the European Commission. A.C. was supported by two grants from the Spanish Government, PSI2011-23033 and and a grant from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013 Cooperation grant agreement n° 613465 - AThEME).

References

- Abdel Rahman, R., & Melinger, A. (2009). Semantic context effects in language production: A swinging lexical network proposal and a review. *Language and Cognitive Processes*, 24, 713-734.
DOI:10.1080/01690960802597250
- Acheson, D.J., Hamidi, M., Binder, J.R. & Postle, B.R. (2011). A common neural substrate for language production and verbal working memory. *Journal of Cognitive Neuroscience*, 23, 1358–1367.
doi:10.1162/jocn.2010.21519
- Amunts, K., et al. (2010). Broca's region: novel organizational principles and multiple receptor mapping. *PLoS Biology*, 8:e1000489. DOI: 10.1371/journal.pbio.1000489
- Aristei, S., Melinger, A., & Abdel Rahman, R. (2011). Electrophysiological chronometry of semantic context effects in language production. *Journal of Cognitive Neuroscience*, 23, 1567–1586.
doi:10.1162/jocn.2010.21474
- Bak, T.H. et al. (2001). Selective impairment of verb processing associated with pathological changes in Brodmann areas 44 and 45 in the motor neurone disease–dementia–aphasia syndrome. *Brain*, 124, 103–120. DOI: <http://dx.doi.org/10.1093/brain/124.1.103>
- Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience*, 15, 600–609. doi:10.1162/089892903321662976

- Barlow, H. B. (1972). Single units and sensation: a neuron doctrine for perceptual psychology? *Perception* 1, 371–394.
- Baus, C., Strijkers, K., & Costa, A. (2013). When does word frequency influence written production?. *Frontiers in psychology*, 4. doi: 10.3389/fpsyg.2013.00963
- Bles, M., & Jansma, B.M. (2008). Phonological processing of ignored distractor pictures: an fMRI investigation. *BMC Neuroscience*, 9, 20. doi:10.1186/1471-2202-9-20
- Bloem, I. & La Heij, W. (2003). Semantic facilitation and semantic interference in word translation: Implications for models of lexical access in language production. *Journal of Memory and Language*, 48, 468–488. doi:10.1016/S0749-596X(02)00503-X
- Boulenger, V. et al. (2008). Word processing in Parkinson's disease is impaired for action verbs but not for concrete nouns. *Neuropsychologia*, 46, 743–756. doi:10.1016/j.neuropsychologia.2007.10.007
- Boulenger, V., Hauk, O., & Pulvermüller, F. (2009). Grasping ideas with the motor system: semantic somatotopy in idiom comprehension. *Cerebral Cortex*, 19, 1905-1914. doi: 10.1093/cercor/bhn217
- Braitenberg, V. (1978). Cell assemblies in the cerebral cortex. In: *Theoretical approaches to complex systems. (Lecture notes in mathematics, vol. 21)*, ed. R. Heim & G. Palm. Springer.
- Camen, C., Morand, S., & Laganaro, M. (2010). Re-evaluating the time course of gender and phonological encoding during silent monitoring tasks estimated by ERP: serial or parallel processing? *Journal of Psycholinguistic Research*, 39, 35–49.
- Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, 14, 177–208. DOI:10.1080/026432997381664
- Caramazza, A., & Costa, A. (2000). The semantic interference effect in the picture–word interference paradigm: Does the response set matter? *Cognition*, 75, B51–B64. doi:10.1016/S0010-0277(99)00082-7
- Catani, M., & Mesulam, M. (2008). The arcuate fasciculus and the disconnection theme in language and aphasia: history and current state. *Cortex*, 44, 953-961. doi:10.1016/j.cortex.2008.04.002
- Chanceaux, M., Vitu, F., Bendahman, L., Thorpe, S & Grainger, J. (2012). Word processing speed in peripheral vision measured with a saccadic choice task. *Vision Research*, 56, 10–19. doi:10.1016/j.visres.2012.01.014
- Chelazzi, L., Miller, E.K., Duncan, J. & Desimone, R. (1993). A neural basis for visual search in inferior

- temporal cortex. *Nature*, 363, 345–347. doi:10.1038/363345a0
- Chen, Y., Davis, M. H., Pulvermuller, F., & Hauk, O. (2013). Task modulations of brain responses in visual word recognition as studied using EEG/MEG and fMRI. *Frontiers in Human Neuroscience*, 7:376. doi: 10.3389/fnhum.2013.00376
- Cheng, X.R., Schafer, G., & Akyurek, E.G. (2010). Name agreement in picture naming: an ERP study. *International Journal of Psychophysiology*, 76, 130–141. doi:10.1016/j.ijpsycho.2010.03.003
- Christoffels, I.K., Firk, C., & Schiller, N.O. (2007). Bilingual language control: an event-related brain potential study. *Brain Research*, 1147, 192–208. doi:10.1016/j.brainres.2007.01.137
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic priming. *Psychological Review*, 82, 407–428.
- Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L., & Petersen, S. E. (1990). Attentional modulation of neural processing of shape, color, and velocity in humans. *Science*, 248, 1556–1559.
- Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, 3, 201–215. doi:10.1038/nrn755
- Costa, A., & Caramazza, A. (2002). The production of noun phrases in English and Spanish: Implications for the scope of phonological encoding in speech production. *Journal of Memory and Language*, 46, 178–198. doi:10.1006/jmla.2001.2804
- Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect: implications for models of lexical access. *Journal of Experimental Psychology: Learning Memory and Cognition*, 26, 1283–1296. <http://dx.doi.org/10.1037/0278-7393.26.5.1283>
- Costa, A., Strijkers, K., Martin, C., & Thierry, G. (2009). The time course of word retrieval revealed by event-related brain potentials during overt speech. *Proceedings of the National Academy of Science U.S.A.*, 106, 21442–21446. doi: 10.1073/pnas.0908921106
- Cotelli, M., et al. (2006). Action and object naming in frontotemporal dementia, progressive supranuclear palsy, and corticobasal degeneration. *Neuropsychology*, 20, 558. <http://dx.doi.org/10.1037/0894-4105.20.5.558>
- Cromer, J. A., Roy, J. E., & Miller, E. K. (2010). Representation of multiple, independent categories in the primate prefrontal cortex. *Neuron*, 66(5), 796–807. doi:10.1016/j.neuron.2010.05.005
- D’Ausilio, A., et al. (2009). The motor somatotopy of speech perception. *Current Biology*, 19, 381–385.

doi:10.1016/j.cub.2009.01.017

Dehaene, S., & Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing.

Neuron, 70, 200-227. doi:10.1016/j.neuron.2011.03.018

Dehaene, S., Changeux, J. P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: A testable taxonomy. *Trends in Cognitive Sciences*, 10(5), 204-211.

doi:10.1016/j.tics.2006.03.007

Dell, G.S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93, 283-321. <http://dx.doi.org/10.1037/0033-295X.93.3.283>

Dell, G. S., & Chang, F. (2014). The P-chain: relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369, 20120394. DOI: 10.1098/rstb.2012.0394

Dell, G. S., Martin, N., & Schwartz, M. F. (2007). A case-series test of the interactive two-step model of lexical access: Predicting word repetition from picture naming. *Journal of Memory and Language*, 56, 490-520. doi:10.1016/j.jml.2006.05.007

Dell, G. S., & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition*, 42, 287-314. doi:10.1016/0010-0277(92)90046-K

Dell, G.S., Schwartz, M.F., Martin, N., Saffran, E.M., & Gagnon, D.A. (1997). Lexical access in aphasic and non-aphasic speakers. *Psychological Review*, 104, 801-838. <http://dx.doi.org/10.1037/0033-295X.104.4.801>

Dell, G. S., Schwartz, M. F., Nozari, N., Faseyitan, O., & Coslett, H. B. (2013). Voxel-based lesion-parameter mapping: Identifying the neural correlates of a computational model of word production. *Cognition*, 128, 380-396. doi:10.1016/j.cognition.2013.05.007

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*. <http://dx.doi.org/10.3389/fpsyg.2010.00177>

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word preactivation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8, 1117-1121. doi:10.1038/nn1504

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of*

Neuroscience, 18, 193–222. DOI: 10.1146/annurev.ne.18.030195.001205

D'hooge, E., & Hartsuiker, R. J. (2011). How do speakers resist distraction? Evidence from a taboo picture–word interference task. *Psychological Science*, 22, 855–859. doi: 10.1177/0956797611410984

Edwards, E., Nagarajan, S.S., Dalal, S.S., Canolty, R.T., Kirsch, H.E., Barbaro, N.M., & Knight, R.T. (2010). Spatiotemporal imaging of cortical activation during verb generation and picture naming. *Neuroimage*, 50, 291–301. doi:10.1016/j.neuroimage.2009.12.035

Engel, A.K. & Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. *Trends in Cognitive Sciences*, 5, 16–25. doi:10.1016/S1364-6613(00)01568-0

Engel, A.K., Maye, A., Kurthen, M., König, P. (2013). Where's the action? The pragmatic turn in cognitive science. *Trends in Cognitive Sciences*, 17, 202-209. doi:10.1016/j.tics.2013.03.006

Engel, A. K., Fries, P., & Singer, W. (2001). Dynamic predictions: oscillations and synchrony in top-down processing. *Nature Reviews Neuroscience*, 2, 704-716. doi:10.1038/35094565

Eulitz, C., Hauk, O., & Cohen, R. (2000). Electroencephalographic activity over temporal brain areas during phonological encoding in picture naming. *Clinical Neurophysiology*, 111, 2088–2097. doi:10.1016/S1388-2457(00)00441-7

Fedorenko, E., Duncan, J. & Kanwisher, N. (2012). Language-selective and domain-general regions lie side by side within Broca's area. *Current Biology*, 22, 2059-2062. doi:10.1016/j.cub.2012.09.011

Fedorenko, E., & Thompson-Schill, S. L. (2014). Reworking the language network. *Trends in cognitive sciences*, 18, 120-126. doi:10.1016/j.tics.2013.12.006

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long term memory structure and sentence processing. *Journal of Memory and Language*, 41, 469–495. doi:10.1006/jmla.1999.2660

Finkbeiner, M., & Caramazza, A. (2006). Now you see it, now you don't: On turning semantic interference into facilitation in a Stroop-like task. *Cortex*, 42, 790–796. doi:10.1016/S0010-9452(08)70419-2

Fodor, J. A. (1983). *Modularity of mind: An essay on faculty psychology*. Cambridge: MIT Press.

Friederici, A.D. (2012). The cortical language circuit: from auditory perception to sentence comprehension. *Trends in Cognitive Sciences*, 16, 262–268. doi:10.1016/j.tics.2012.04.001

Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in cognitive sciences*, 9(10), 474-480. doi:10.1016/j.tics.2005.08.011

- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127-138. doi:10.1038/nrn2787
- Fuster, J.M. (2003). *Cortex and Mind: Unifying Cognition* (Oxford, UK: Oxford Univ. Press).
- Ganushchak L. Y., Christoffels I. K. & Schiller N. O. (2011). The use of electroencephalography in language production research: a review. *Frontiers in Psychology*, 2:208. doi: 10.3389/fpsyg.2011.00208
- Garagnani, M., Wennekers, T., & Pulvermüller, F. A. (2008). Neuroanatomically-grounded Hebbian learning model of attention-language interactions in the human brain. *European Journal of Neuroscience*, 27, 492–513. DOI: 10.1111/j.1460-9568.2008.06015.x
- Garagnani, M., Shtyrov, Y. & Pulvermüller F. (2009). Effects of attention on what is known and what is not: MEG evidence for functionally discrete memory circuits. *Frontiers in human neuroscience*, 3(10). doi: 10.3389/neuro.09.010.2009
- Giesbrecht, B., Woldorff, M.G., Song, A.W. & Mangun, G.R. (2003). Neural mechanisms of top-down control during spatial and feature attention. *Neuroimage*, 19, 496–512. doi:10.1016/S1053-8119(03)00162-9
- Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews Neuroscience*, 14, 350–363. doi:10.1038/nrn3476
- Glaser, W. R., & Glaser, M. O. (1989). Context effects on Stroop-like word and picture processing. *Journal of Experimental Psychology: General*, 118, 13–42. <http://dx.doi.org/10.1037/0096-3445.118.1.13>
- Glaser, M. F., & Rilling, J. K. (2008). DTI tractography of the human brain's language pathways. *Cerebral Cortex*, 18, 2471-2482. doi: 10.1093/cercor/bhn011
- 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, 617–631. doi:10.1162/jocn.2007.19.4.617
- Graves, W.W., Grabowski, T.J., Mehta, S., & Gupta, P. (2008). The left posterior superior temporal gyrus participates specifically in accessing lexical phonology. *Journal of Cognitive Neuroscience*, 20, 1698–1710. DOI:10.1162/jocn.2008.20113
- Guo, T.M., Peng, D.L., Lu, C.M., & Liu, H.Y. (2005). The temporal course of semantic and phonological activation in Chinese word production: an ERP study. *Acta Psychologica Sinica*, 37, 569–574.
- Hagoort, P. (2005). On Broca, brain, and binding: a new framework. *Trends in Cognitive Sciences*, 9, 416–

423. doi:10.1016/j.tics.2005.07.004

Hagoort, P., & Levelt, W.J.M. (2009). The speaking brain. *Science*, 326, 372–373.

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, 902–934. DOI:10.1080/01690965.2010.509946

Hauk, O. et al. (2004). Somatotopic representation of action words in the motor and premotor cortex. *Neuron* 41, 301–307. doi:10.1016/S0896-6273(03)00838-9

Heald, S. L., & Nusbaum, H. C. (2014). Speech perception as an active cognitive process. *Frontiers in systems neuroscience*, 8. doi: 10.3389/fnsys.2014.00035

Hebb, D. O. (1949). *The organization of behavior. A neurophysiological theory*. Wiley.

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. doi: 10.1097/01.wnr.0000233102.43526.e9

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, 13, 135-145. doi:10.1038/nrn3158

Hickok, G. (2014a). The architecture of speech production and the role of the phoneme in speech processing. *Language and Cognitive Processes*, 29, 2-20. DOI:10.1080/01690965.2013.834370

Hickok, G. (2014b). Towards an integrated psycholinguistic, neurolinguistic, sensorimotor framework for speech production. *Language, Cognition and Neuroscience*, 29, 52-59. DOI:10.1080/01690965.2013.852907

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8, 393–402. doi:10.1038/nrn2113

Hillyard, S. A., Vogel, E. K., & Luck, S. J. (1998). Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence. *Philosophical Transactions of the Royal Society Biological Sciences*, 353, 1257–1267. DOI: 10.1098/rstb.1998.0281

Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: Experimental and computational studies. *Cognition*, 100, 464-482. doi:10.1016/j.cognition.2005.02.006

- Hubel, D. H. & Wiesel, T. N. (1965). Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *Journal of Neurophysiology*, 28, 229–289.
- Hulten, A., Vihla, M., Laine, M., & Salmelin, R. (2009). Accessing newly learned names and meanings in the native language. *Human Brain Mapping*, 30, 976–989. DOI: 10.1002/hbm.20561
- Indefrey, P. (2011). The spatial and temporal signatures of word production components: A critical update. *Frontiers in Psychology*, 2, 255. doi: 10.3389/fpsyg.2011.00255
- Indefrey, P., & Levelt, W.J.M. (2004). The spatial and temporal signatures of word production components. *Cognition*, 92, 101–144. doi:10.1016/j.cognition.2002.06.001
- Ivanova, I., & Costa A. (2008). Does bilingualism hamper lexical access in speech production? *Acta Psychologica*, 127, 277–288. doi:10.1016/j.actpsy.2007.06.003
- Janssen, N., Hernández-Cabrera, J. A., van der Meij, M., & Barber, H. A. (2014). Tracking the Time Course of Competition During Word Production: Evidence for a Post-Retrieval Mechanism of Conflict Resolution. *Cerebral Cortex*, 25, 2960-2969. doi: 10.1093/cercor/bhu092
- Kanwisher, N., & Wojciulik, E. (2000). Visual attention: insights from brain imaging. *Nature Reviews Neuroscience*, 1, 91-100. doi:10.1038/35039043
- Kastner, S., & Ungerleider, L. G. (2000). Mechanisms of visual attention in the human cortex. *Annual Review of Neuroscience*, 23, 315–341. DOI: 10.1146/annurev.neuro.23.1.315
- Kastner, S., Pinsk, M., De Weerd, P., Desimone, R., & Ungerleider, L. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron*, 22, 751–761. doi:10.1016/S0896-6273(00)80734-5
- Kiefer, M. (2007). Top–down modulation of unconscious “automatic” processes: A gating framework. *Advances in Cognitive Psychology*, 3, 289–306. doi: 10.2478/v10053-008-0032-2
- Kiefer, M. (2012). Executive control over unconscious cognition: Attentional sensitization of unconscious information processing. *Frontiers in Human Neuroscience*, 6, 61. doi: 10.3389/fnhum.2012.00061
- Koester, D., & Schiller, N.O. (2008). Morphological priming in overt language production: electrophysiological evidence from Dutch. *Neuroimage*, 42, 1622–1630. doi:10.1016/j.neuroimage.2008.06.043
- Koester, D., & Schiller, N. O. (2011). The functional neuroanatomy of morphology in language production. *NeuroImage*, 55(2), 732-741. doi:10.1016/j.neuroimage.2010.11.044

- Kotz, S.A. & Schwartz, M. (2010). Cortical speech processing unplugged: a timely subcortico-cortical framework. *Trends in Cognitive Sciences*, 14, 392–399. doi:10.1016/j.tics.2010.06.005
- Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, 4, 463–470. doi:10.1016/S1364-6613(00)01560-6
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62(1), 621–647. doi: 10.1146/annurev.psych.093008.131123
- Laganaro, M. (2014). ERP topographic analyses from concept to articulation in word production studies. *Frontiers in Psychology*, 5:493. doi: 10.3389/fpsyg.2014.00493
- Laganaro, M., Morand, S., & Schnider, A. (2009a). Time course of evoked-potential changes in different forms of anomia in aphasia. *Journal of Cognitive Neuroscience*, 21, 1499–1510. doi:10.1162/jocn.2009.21117
- Laganaro, M., et al. (2009b). Electrophysiological correlates of different anomic patterns in comparison with normal word production. *Cortex*, 45, 697–707. doi:10.1016/j.cortex.2008.09.007
- Laganaro, M., & Perret, C. (2011) Comparing electrophysiological correlates of word production in immediate and delayed naming through the analysis of word age of acquisition effects. *Brain Topography*, 24, 19–29.
- Laganaro M., Python G. & Toepel U. (2013). Dynamics of phonological-phonetic encoding in word production: evidence from diverging ERPs between stroke patients and controls. *Brain and Language*, 126, 123–132. doi:10.1016/j.bandl.2013.03.004
- 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, 3381–3388. doi:10.1016/j.neuroimage.2011.10.082
- Levelt, W.J.M. (1989). *Speaking: From Intention To Articulation*. Cambridge, MA: MIT Press.
- 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, 553–567. DOI:10.1162/089892998562960
- Levelt, W.J.M., Roelofs, A., & Meyer, A.S. (1999). A theory of lexical access in speech production. *Behavioral Brain Sciences*, 22, 1–38.
- 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, 122–142. <http://dx.doi.org/10.1037/0033-295X.98.1.122>
- Liljestrom, M., Hulten, A., Parkkonen, L., & Salmelin, R. (2009). Comparing MEG and fMRI views to naming actions and objects. *Human Brain Mapping*, 30, 1845–1856. DOI: 10.1002/hbm.20785
- Llorens, A., Trébuchon, A., Liégeois-Chauvel, C., & Alario, F. X. (2011). Intra-cranial recordings of brain activity during language production. *Frontiers in psychology*, 2. doi: 10.3389/fpsyg.2011.00375
- Llorens, A., Trébuchon, A., Riès, S., Liégeois-Chauvel, C., & Alario, F. (2014). How familiarization and repetition modulate the picture naming network. *Brain and language*, 133, 47-58.
doi:10.1016/j.bandl.2014.03.010
- MacGregor, L.J., Pulvermüller, F., van Casteren, M. & Shtyrov, Y. (2012). Ultra-rapid access to words in the brain. *Nature Communications*, 3, 7-11. doi:10.1038/ncomms1715
- 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, 455–462. doi:10.1162/089892902317361967
- Mahon, B.Z., Costa, A., Peterson, R., Vargas, K.A., & Caramazza, A. (2007). Lexical Selection is Not by Competition: A Reinterpretation of Semantic Interference and facilitation Effects in the Picture-Word Interference Paradigm. *Journal of Experimental Psychology: Learning Memory and Cognition*, 33, 503-535. <http://dx.doi.org/10.1037/0278-7393.33.3.503>
- Marr, D. (1982). *Vision: A computational approach*. San Francisco: Freeman & Co.
- Menenti, L., Segaert, K., & Hagoort, P. (2012). The neuronal infrastructure of speaking. *Brain and language*, 122, 71-80. doi:10.1016/j.bandl.2012.04.012
- Mesulam, M. M. (1990). Large-scale neurocognitive networks and distributed processing for attention, language, and memory. *Annals of Neurology*, 28, 597–613.
- Miozzo, M., Pulvermüller, F., & Hauk, O. (2014). Early parallel activation of semantics and phonology in picture naming: Evidence from a multiple linear regression MEG study. *Cerebral Cortex*, 25, 3343-3355. doi: 10.1093/cercor/bhu137
- Morsella, E., & Miozzo, M. (2002). Evidence for a cascade model of lexical access in speech production. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 28, 555–563.

<http://dx.doi.org/10.1037/0278-7393.28.3.555>

- Moseley, R. L., Pulvermüller, F., & Shtyrov, Y. (2013). Sensorimotor semantics on the spot: brain activity dissociates between conceptual categories within 150 ms. *Scientific reports*, 3.
doi:10.1038/srep01928
- Motter, B.C. (1993). Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *Journal of Neurophysiology*, 70, 909–919.
- Näätänen, R., Paavilainen, P., Rinne, T. & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clinical Neurophysiology*, 118, 2544-2590.
doi:10.1016/j.clinph.2007.04.026
- Naccache, L., Blandin, E., & Dehaene, S. (2002). Unconscious masked priming depends on temporal attention. *Psychological Science*, 13, 416-424. doi: 10.1111/1467-9280.00474
- 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, 359–377.
doi:10.1016/j.jml.2005.05.001
- Navarrete, E., Mahon, B. Z. & Caramazza, A. (2010). The cumulative semantic cost does not reflect lexical selection by competition. *Acta Psychologica*, 134, 279-289. doi:10.1016/j.actpsy.2010.02.009
- O’Craven, K., & Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *Journal of Cognitive Neuroscience*, 12, 1013-1023.
doi:10.1162/08989290051137549
- O’Craven, K. M., Downing, P. E., & Kanwisher, N. (1999). FMRI evidence for objects as the units of attentional selection. *Nature*, 401, 584–587. doi:10.1038/44134
- Papoutsis, M., de Zwart, J.A., Jansma, J.M., Pickering, M.J., Bednar, J. A., & Horwitz, B. (2009). From phonemes to articulatory codes: an fmri study of the role of Broca’s area in speech production. *Cerebral Cortex*, 19, 2156–2165. doi: 10.1093/cercor/bhn239
- Peelen, M. V., Fei-Fei, L., & Kastner, S. (2009). Neural mechanisms of rapid natural scene categorization in human visual cortex. *Nature*, 460, 94-97. doi:10.1038/nature08103
- Peeva, M.G., Guenther, F.H., Tourville, J.A., Nieto-Castanon, A., Anton, J.L., Nazarian, B., & Alario, F. X. (2010). Distinct representations of phonemes, syllables, and supra-syllabic sequences in the speech production network. *Neuroimage*, 50, 626–638. doi:10.1016/j.neuroimage.2009.12.065

- Perret C., Bonin P. & Laganaro M. (2014). Exploring the multiple-loci hypothesis of AoA effects in spoken and written object naming using a topographic ERP analysis. *Brain and Language*, 135, 20-31.
doi:10.1016/j.bandl.2014.04.006
- Petrides, M. & Pandya, D.N. (2009). Distinct parietal and temporal pathways to the homologues of Broca's area in the monkey. *PLoS Biology*, 7:e1000170. DOI: 10.1371/journal.pbio.1000170
- Piai, V., Roelofs, A., Jensen, O., Schoffelen, J.-M., & Bonnefond, M. (2014a). Distinct patterns of brain activity characterise lexical activation and competition in spoken word production. *PLoS One*, 9, e88674. DOI: 10.1371/journal.pone.0088674
- Piai, V., Roelofs, A. & Schriefers, H. (2011). Semantic interference in immediate and delayed naming and reading: Attention and task decisions. *Journal of Memory and Language*, 64, 404–423.
doi:10.1016/j.jml.2011.01.004
- Piai, V., Roelofs, A., & Schriefers, H. (2014b). Locus of semantic interference in picture naming: Evidence from dual-task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40, 147–165. <http://dx.doi.org/10.1037/a0033745>
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36, 329-347. DOI: <http://dx.doi.org/10.1017/S0140525X12001495>
- Poeppl, D. (2012). The maps problem and the mapping problem: two challenges for a cognitive neuroscience of speech and language. *Cognitive Neuropsychology*, 29, 34-55.
DOI:10.1080/02643294.2012.710600
- Poeppl, D., Emmorey, K., Hickok, G. & Pylkkanen, L. (2012). Towards a new neurobiology of language. *Journal of Neuroscience*, 32, 14125–14131. doi: 10.1523/JNEUROSCI.3244-12.2012
- Porrino, L. J., Crane, A. M., & Goldman-Rakic, P. S. (1981). Direct and indirect pathways from the amygdala to the frontal lobe in rhesus monkeys. *Journal of Comparative Neurology*, 198, 121–136.
DOI: 10.1002/cne.901980111
- Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 55–85). Hillsdale, NJ: Erlbaum.
- Price, C. J. (2010). The anatomy of language: a review of 100 fMRI studies published in 2009. *Annals of the New York Academy of Sciences*, 1191, 62-88. DOI: 10.1111/j.1749-6632.2010.05444.x

- Price, C.J. (2012). A review and synthesis of the first 20 years of PET and fmri studies of heard speech, spoken language and reading. *Neuroimage*, 62, 816-847. doi:10.1016/j.neuroimage.2012.04.062
- Pulvermuller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences*, 22, 253-336. DOI: <http://dx.doi.org/10.1017/S0140525X9900182X>
- Pulvermuller F. (2002). *The neuroscience of language*. Cambridge University Press, Cambridge.
- Pulvermuller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience* 6, 576-582. doi:10.1038/nrn1706
- Pulvermuller, F. (2013). How neurons make meaning: brain mechanisms for embodied and abstract-symbolic semantics. *Trends in Cognitive Sciences*, 17, 458-470. doi:10.1016/j.tics.2013.06.004
- Pulvermuller, F. & Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, 11, 351–360. doi:10.1038/nrn2811
- Pulvermuller F, Garagnani M. (2014). From sensorimotor learning to memory cells in prefrontal and temporal association cortex: a neurocomputational study of disembodiment. *Cortex*, 75, 1–21. doi:10.1016/j.cortex.2014.02.015
- Pulvermuller, F., Moseley, R., Egorova, N. et al. (2014). Motor cognition - motor semantics: action-perception theory of cognitive and communicative cortical function. *Neuropsychologia*, 55, 71–84. doi:10.1016/j.neuropsychologia.2013.12.002
- Pulvermuller, F., Shtyrov, Y., & Hauk, O. (2009). Understanding in an instant: Neurophysiological evidence for mechanistic language circuits in the brain. *Brain and Language*, 110, 81–94. doi:10.1016/j.bandl.2008.12.001
- Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological review*, 107(3), 460. <http://dx.doi.org/10.1037/0033-295X.107.3.460>
- Rauschecker, J. P., & Tian, B. (2000). Mechanisms and streams for processing of “what” and “where” in auditory cortex. *Proceedings of the National Academy of Sciences USA*, 97, 11800-11806. doi: 10.1073/pnas.97.22.11800
- Rempel-Clower, N. L., & Barbas, H. (2000). The laminar pattern of connections between prefrontal and anterior temporal cortices in the rhesus monkey is related to cortical structure and function. *Cerebral Cortex*, 10, 851–865.

- Roelfsema, P.R., Lamme, V.A. & Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature*, 395, 376–381. doi:10.1038/26475
- Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, 42, 107–142. doi:10.1016/0010-0277(92)90041-F
- Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition*, 64, 249–284. doi:10.1016/S0010-0277(97)00027-9
- Roelofs, A. (2003). Goal-referenced selection of verbal action: Modelling attentional control in the Stroop task. *Psychological Review*, 110, 88–125. <http://dx.doi.org/10.1037/0033-295X.110.1.88>
- Roelofs, A. (2008). Attention to spoken word planning: Chronometric and neuroimaging evidence. *Language and Linguistic Compass*, 2, 389–405. DOI: 10.1111/j.1749-818X.2008.00060.x
- Roelofs, A. (2014). A dorsal-pathway account of aphasic language production: The WEAVER++/ARC model. *Cortex*, 59, 33-48. doi:10.1016/j.cortex.2014.07.001
- Roelofs, A., Piai, V. & Schriefers, H. (2013). Context effects and selective attention in picture naming and word reading: Competition versus response exclusion. *Language and Cognitive Processes*, 28, 655–671. DOI:10.1080/01690965.2011.615663
- Runnqvist, E., Strijkers, K., Alario, F.-X., & Costa, A. (2012). Cumulative semantic interference is blind to language: Implications for models of bilingual speech production. *Journal of Memory and Language*, 66, 850-869. doi:10.1016/j.jml.2012.02.007
- Ruz, M., & Nobre, A. C. (2008). Attention modulates initial stages of visual word processing. *Journal of Cognitive Neuroscience*, 20, 1727–1736. DOI: 10.1162/jocn.2008.20119
- 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*, 326, 445–449. DOI: 10.1126/science.1174481
- Salmelin, R., Hari, R., Lounasmaa, O.V., & Sams, M. (1994). Dynamics of brain activation during picture naming. *Nature*, 368, 463–465.
- Schiller, N.O. (2006). Lexical stress encoding in single word production estimated by event-related brain potentials. *Brain Research*, 1112, 201–212. doi:10.1016/j.brainres.2006.07.027
- Schotter, E. R., Ferreira, V. S., & Rayner, K. (2013). Parallel object activation and attentional gating of information: Evidence from eye movements in the multiple object naming paradigm. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, 39, 365.

<http://dx.doi.org/10.1037/a0028646>

Schuhmann, T., Schiller, N.O., Goebel, R., & Sack, A.T. (2009). The temporal characteristics of functional activation in Broca's area during overt picture naming. *Cortex*, 45, 1111–1116.

doi:10.1016/j.cortex.2008.10.013

Schuhmann, T., Schiller, N.O., Goebel, R. & Sack, A.T. (2012). Speaking of which: dissecting the neurocognitive network of language production in picture naming. *Cerebral Cortex*, 22, 701-709.

doi: 10.1093/cercor/bhr155

Schwartz, M.F., Kimberg, D.Y., Walker, G. M., Faseyitan, O., Brecher, A., Dell, G.S., & Coslett, H.B. (2009). Anterior temporal involvement in semantic word retrieval: voxel- based lesion symptom mapping evidence from aphasia. *Brain*, 132, 3411–3427. DOI: <http://dx.doi.org/10.1093/brain/awp284>

Shallice, T. (1988). *From Neuropsychology to Mental Structure*. Cambridge University Press

Singer, W. & Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annual Review Neuroscience*, 18, 555–586. DOI: 10.1146/annurev.ne.18.030195.003011

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norm for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 174-215.

Soros, P., Cornelissen, K., Laine, M., & Salmelin, R. (2003). Naming actions and objects: cortical dynamics in healthy adults and in an amnic patient with a dissociation in action/object naming. *Neuroimage*, 19, 1787–1801. doi:10.1016/S1053-8119(03)00217-9

Strijkers, K. (2014). How does Hebb both 'divide' and 'conquer' speech perception and production?. *Language, Cognition and Neuroscience*, 29(1), 49-51. DOI:10.1080/01690965.2013.852231

Strijkers K., Baus C., Runnqvist E., Fitzpatrick I., Costa A. (2013). The temporal dynamics of first versus second language production. *Brain and Language*, 127, 6–11. doi:10.1016/j.bandl.2013.07.008

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, 356. doi: 10.3389/fpsyg.2011.00356

Strijkers, K., Costa, A., & Pulvermuller, F. (2017). The cortical dynamics of speaking: Lexical and phonological knowledge simultaneously recruit the frontal and temporal cortex within 200 ms.

<https://doi.org/10.1016/j.neuroimage.2017.09.041>

- Strijkers, K., Costa, A., & Thierry, G. (2010). Tracking lexical access in speech production: electrophysiological correlates of word frequency and cognate effects. *Cerebral Cortex*, 20, 912–928. doi: 10.1093/cercor/bhp153
- Strijkers, K., Holcomb, P., & Costa, A. (2011a). Conscious intention to speak facilitates lexical access during overt object naming. *Journal of Memory and Language*, 65, 345–362. doi:10.1016/j.jml.2011.06.002
- Strijkers K., Yum Y. N., Grainger J., & Holcomb P. J. (2011b). Early goal-directed top-down influences in the production of speech. *Frontiers in Psychology*, 2:371. doi: 10.3389/fpsyg.2011.00371
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643–662.
- Tian, X., & Poeppel, D. (2010). Mental imagery of speech and movement implicates the dynamics of internal forward models. *Frontiers in Psychology*, 1. <http://dx.doi.org/10.3389/fpsyg.2010.00166>
- Ueno, T., Saito, S., Rogers, T.T. & Lambon Ralph, M.A. (2011). Lichtheim 2: synthesizing aphasia and the neural basis of language in a neurocomputational model of the dual dorsal-ventral language pathways. *Neuron*, 72, 385–396. doi:10.1016/j.neuron.2011.09.013
- Ullman S. (2007). Object recognition and segmentation by a fragment-based hierarchy. *Trends in Cognitive Sciences*, 11, 58–64. doi:10.1016/j.tics.2006.11.009
- van Turenout, M., Hagoort, P., & Brown, C.M. (1997). Electrophysiological evidence on the time course of semantic and phonological processes in speech production. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 23, 787–806. <http://dx.doi.org/10.1037/0278-7393.23.4.787>
- van Turenout, M., Hagoort, P., & Brown, C.M. (1998). Brain activity during speaking: from syntax to phonology in 40 milliseconds. *Science*, 280, 572–574. DOI: 10.1126/science.280.5363.572
- Varela, F.J. et al. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. MIT Press.
- Varela, F. et al. (2001) The brainweb: phase synchronization and large-scale integration. *Nature Reviews Neuroscience*, 2, 229–239. doi:10.1038/35067550
- Vihla, M., Laine, M., & Salmelin, R. (2006). Cortical dynamics of visual/semantic vs. phonological analysis in picture confrontation. *Neuroimage*, 33, 732–738. doi:10.1016/j.neuroimage.2006.06.040
- Wilson, S.M., Isenberg, A.L., & Hickok, G. (2009). Neural correlates of word production stages delineated by parametric modulation of psycholinguistic variables. *Human Brain Mapping*, 30, 3596–3608. DOI: 10.1002/hbm.20782

Wilson, S.M., Saygin, A.P., Sereno, M.I. & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, 7, 701-702. doi:10.1038/nn1263

Ylinen, S., Nora, A., Leminen, A., Hakala, T., Huotilainen, M., Shtyrov, Y., & Mäkelä, J. P. (2014). Two Distinct Auditory-Motor Circuits for Monitoring Speech Production as Revealed by Content-Specific Suppression of Auditory Cortex. *Cerebral Cortex*, 25, 1576-1586. doi: 10.1093/cercor/bht351

Zhang, Q.F., & Damian, M.F. (2009). The time course of semantic and orthographic encoding in Chinese word production: an event-related potential study. *Brain Research*, 1273, 92–105. doi:10.1016/j.brainres.2009.03.049

Figure Captions

Fig. 1. Schematic (and simplified) visualization of the dominant spatiotemporal model of word production. Brain activity sequentially progresses from the left mid temporal gyrus (MTG) linked to lexico-semantic processes, towards the left superior temporal gyrus (STG) involved in lexical phonology, over the left inferior frontal gyrus (LIFG) associated with syllabification and ending in the motor cortex (MC) for sending an articulation program. Numbers in the different processing hierarchies represent approximates in ms of when brain activity is thought to be maximal in that particular region. (*Based on e.g. I&L, 2004; Indefrey, 2011*)

Fig. 2. Schematic example of how context- and task-dependent top-down modulations may affect the spatiotemporal correlates of word production components. (A) Example of the time-course of brain activations recruited in speech production without the top-down intention to speak. (B) Example of the time-course of brain activations recruited in speech production with the top-down intention to speak. Numbers represent activation estimates in ms and are hypothetical (purely for illustrative purposes). The point is that without top-down prediction the time-course of brain activations involved in word production will likely go slower than in I&L (2004), while under conditions of top-down prediction they will likely go faster (*based on e.g. Strijkers, Holcomb & Costa, 2011*). The context- and task-dependent top-down modulations (which may affect the whole language production network as well as specific parts of it) are hypothesized to be generated

in structures of the frontal lobe such as the prefrontal cortex (PFC) and the frontal eye field (FEF), and structures in the parietal lobe such as the inferior parietal cortex (IP) and the temporo-parietal junction (TPJ) (note that arrows from these regions to the language areas are for visual purposes only and do not represent a structural claim about the connectivity). (Based on e.g. Corbetta & Shulman, 2002).

Fig. 3. Schematic (and simplified) visualization of a Hebbian-based cell assembly network of word production. A widely distributed lexico-semantic network embedded in local-specific action (red) – perception (yellow) circuits and a widely distributed phonological-phonemic network embedded in specific 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 (e.g., reverberation in the lexico-semantic sensorimotor circuit for semantic integration and reverberation in the phonological-phonemic sensorimotor circuit for timed articulation). (Based on e.g. Pulvermuller, 1999; 2001; 2005; Pulvermuller & Fadiga, 2010; Strijkers et al., under rev.).

Figures.

Fig. 1

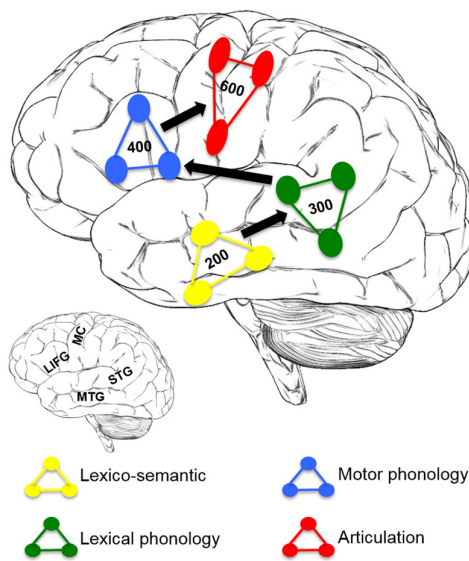


Fig. 2

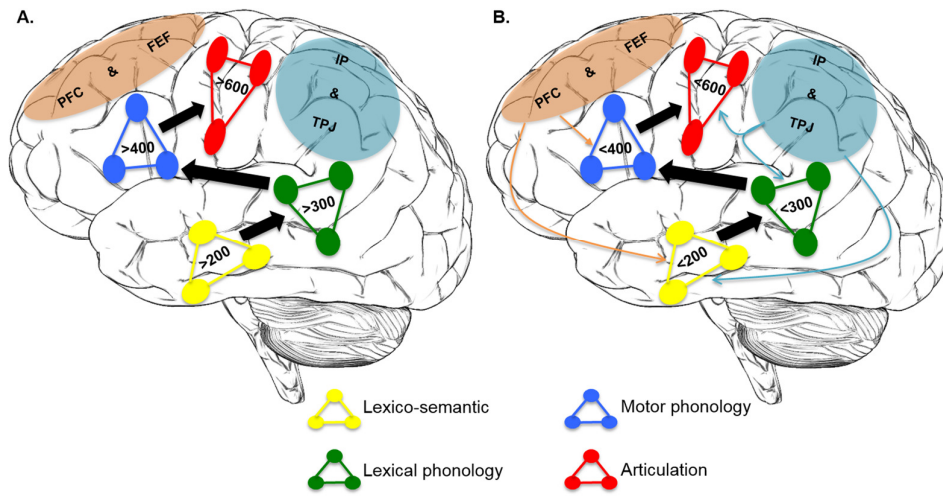


Fig. 3

