NNT: 2024AIXM0000

THÈSE DE DOCTORAT

Soutenue à Aix-Marseille Université
le 11 décembre 2024 par

Emiliia MKRTYCHYAN

Linking linguistic and neural alignment of word meaning in the dyad

Discipline
Sciences du Langage

École doctorale
ED 356 – Cognition, Langage, Education

Laboratoire/Partenaires de recherche
Laboratoire Parole et Langage

Composition du jury

Martin PICKERING, PR University of Edinburgh	Rapporteur Président du jury
Anna K. KUHLEN, PR Aachen University	Rapporteuse
Véronique BOULENGER, HDR Université Lyon 2, CNRS	Examinatrice
Kristof STRIJKERS, HDR Aix Marseille Université, CNRS	Directeur de thèse
Benjamin MORILLON, HDR Aix Marseille Université, INSERM	Co-directeur de thèse

Affidavit

I, undersigned, Emilia Kerr, hereby declare that the work presented in this manuscript is my own work, carried out under the scientific supervision of Kristof Strijkers and Benjamin Morillon, in accordance with the principles of honesty, integrity and responsibility inherent to the research mission. The research work and the writing of this manuscript have been carried out in compliance with both the French national charter for Research Integrity and the Aix-Marseille University charter on the fight against plagiarism.

This work has not been submitted previously either in this country or in another country in the same or in a similar version to any other examination body.

Place Aix en Provence, date 1er octobre 2024

Liste de publications et/ou brevets et participation aux conférences

1) Liste des publications réalisées dans le cadre du projet de thèse:

1. Kerr, E., Morillon, B., & Strijkers, K. (2024, September 9). Predicting meaning in the dyad. https://doi.org/10.31234/osf.io/8cf2m

2. Kerr, E., Strijkers, K., & Morillon, B. (2024). Does prediction drive neural alignment in conversation? *In Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 46).

3. Kerr, E., Ivanova, B., & Strijkers, K. (2023). Lexical access in speech production: Psycho-and neurolinguistic perspectives on the spatiotemporal dynamics. *In Language Production* (pp. 32-65). Routledge.

4. Kerr, E., & Strijkers, K. (2022). The effects of prediction in conversational alignment. *In Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 44, No. 44).

3. Blache, P., Antoine, S., de Jong, D., Huttner, L.-M., Kerr, E., et al. (2022). The Badalona Corpus an Audio, Video and Neuro-Physiological Conversational Dataset. Language *Resources and Evaluation Conference,* 2022, Marseille, France.

2) Participation aux conférences et écoles d'été au cours de la période de thèse:

1. July 2024 – poster presentation at the CogSci conference (Rotterdam, the Netherlands).

2. June 2024 – poster presentation at the IWoLP conference (Marseille, France).

3. October 2023 – poster presentation at the joint COBRA workshop (Aix en Provence, France).

4. October 2023 - poster presentation at the SNL conference (Marseille, France).

5. July 2023 – short talk at the Waves Conference (Salerno, Italy).

6. 2020 – 2023 – participation (project presentations) in conferences and training events organized by the Cobra consortium (Bratislava (March 2021), Marseille (May 2021), Ferrara (June 2022), Louvain (2023).

7. May 2022 - IMPRS conference, poster presentation (Nijmegen, the Netherlands).

8. September 2021 – participation and project presentation in the European Researchers' Night event (Badalona, Spain).

Résumé

Cette thèse explore la relation entre l'alignement neuronal et linguistique dans la signification des mots, en examinant si la prédiction favorise cet alignement. L'alignement linguistique décrit la tendance des partenaires de conversation à synchroniser leurs états mentaux, mots et concepts (Clark, 1996 ; Pickering & Garrod, 2004), tandis que l'alignement neuronal se manifeste lorsque l'activité cérébrale des locuteurs et auditeurs devient corrélée (Stephens et al., 2010; Hasson et al., 2012). Malgré des preuves des deux types d'alignement, leur interrelation reste floue.

Le premier objectif de la thèse est de vérifier si l'alignement neuronal reflète l'alignement linguistique en étant sensible aux connaissances spécifiques, comme la signification des mots. Le second est d'examiner si la prédiction améliore les deux alignements, linguistique et neuronal, étant donné qu'elle est suggérée comme un mécanisme clé (Pickering & Garrod, 2013; Friston & Frith, 2015; Pickering & Gambi, 2018). Alors que le traitement prédictif a été bien étudié chez les individus, son rôle dans les interactions dyadiques est moins compris.

Un nouveau paradigme expérimental combinant hyperscanning EEG et jeux de langage interactifs a été conçu. Les participants ont effectué des tâches d'association sémantique simples, avec ou sans manipulation de prévisibilité. Cela a permis d'analyser (1) l'alignement neuronal pour le sens des mots et (2) l'effet de la prédiction sur cet alignement.

Les résultats montrent que l'alignement neuronal est sensible au sens des mots et que la prédiction facilite l'alignement neuronal précoce. Ces conclusions suggèrent que l'alignement neuronal pourrait refléter l'alignement linguistique et illustrent le rôle crucial de la prédiction dans le dialogue.

Mots-clés: alignement linguistique, alignement neuronal, traitement prédictif, hyperscanning

Abstract

This thesis explores the relationship between neural and linguistic alignment for word meaning, examining whether prediction drives alignment. Linguistic alignment refers to the tendency of conversational partners to synchronize mental states, words, and concepts (Clark, 1996; Pickering & Garrod, 2004). Neural alignment occurs when brain activity in speakers and listeners becomes correlated (Stephens et al., 2010; Hasson et al., 2012). Despite the evidence for both forms of alignment, whether they represent the same phenomenon remains unclear.

The first objective of this thesis is to investigate whether neural alignment reflects linguistic alignment by examining whether it is sensitive to specific linguistic knowledge, such as word meanings. The second objective is to explore whether prediction enhances both linguistic and neural alignment, since it has been proposed as a key mechanism of alignment (Pickering & Garrod, 2013; Friston & Frith, 2015; Pickering & Gambi, 2018). Although predictive processing has been well-studied in individuals, its role in dyadic interactions is less understood.

To address these goals, a novel experimental paradigm using EEG hyperscanning and interactive language games was developed. Participants engaged in simple semantic association tasks with or without predictability manipulation. This setup allowed investigation of (1) neural alignment for word meaning and (2) the effect of prediction on meaning-specific alignment.

Results demonstrated neural alignment sensitive to word meaning and showed that prediction facilitates earlier emergence of meaning-specific neural alignment. These findings suggest neural alignment could be the brain's signature of linguistic alignment, and offer a potential neural implementation of the interactive alignment model, highlighting the role of prediction in conversational behavior.

Keywords: linguistic alignment, neural alignment, predictive processing, hyperscanning

Remerciements

Since my first years at university, which was quite a long time ago, I have been dreaming of devoting my professional life to research. This dream was not fulfilled until I came here, to the Laboratoire Parole et Langage, where I was given the beautiful opportunity to learn so much from amazing people. And I am immensely grateful for this.

I feel incredibly lucky to have met Kristof, who offered me the chance to come here and be part of the Cobra project. Thank you, Kristof, for being such a great supervisor that you are. Learning from you has been an exciting and thoughtful experience. The way you pursue ideas and march through theories is inspiring and highly educating. Although a primary goal, not only teaching is important when it comes to supervision. You have been always so understanding and supportive, and this helped me get through difficult moments of my PhD years. Saying 'great work' after reading my drafts when I could surely have done better, or being forgiving for missing team meetings when my kids demanded my presence at home might seem like small details, but this is precisely the attitude that has put my mind at peace so many times. After these four years, I sometimes still can't believe that I have been so lucky to work with a supervisor who understands and helps me, without putting any unnecessary pressure. In today's world, and maybe academia especially, this is precious.

Not only was I lucky with my main supervisor, but also with the co-supervisor who agreed to get me through the statistical hurdles and challenges. Thank you, Benjamin, for all your help and understanding. I started knowing very little about the complex and exciting ways of exploring the data, but your guidance has taught me so much.

I also want to thank the people who worked on making the Cobra project possible. Noel, this is your creation, and it has become a great start for all the PhD students, including myself, who participated in Cobra. Some have already moved on; others are still finishing but we all have gained so much through the collaboration that you brought to life.

This thesis would not be finished without the help of those who assisted me with my experiments. Thank you, Thierry, Christelle and Daria. The dual EEG testing is not an easy task, but you made it so much easier for me.

Finally, I am immensely grateful to my family who have been with me throughout my studies. My parents and my sister, who let me pursue my interests and take pride in my achievements; my husband who has been so supportive and forgiving, and my children who are the source of my inspiration and motivation. Thank you, my beloved ones, – without you I would not have done this.

These past four years have been filled with challenges, as I believe, any PhD endeavour encompasses. But the challenges of the sort that allow one grow, both in their knowledge and confidence. Thank you all once again for giving me the opportunity to meet these challenges and enjoy them.

Table of Contents

AFFIDAVIT 2
LISTE DE PUBLICATIONS ET/OU BREVETS ET PARTICIPATION AUX CONFÉRENCES 3
ABSTRACT 5
REMERCIEMENTS 6
CHAPTER 1. GENERAL INTRODUCTION 10

1.1 On linguistic alignment 12
1.1.1. The Interactive Alignment Model 17
1.2 On neural alignment 19
1.3. On prediction 23
1.4 Interim Discussion 29

CHAPTER 2. LEXICAL ACCESS IN SPEECH PRODUCTION 32

Introduction 34
1. The sequential (serial) model of lexical access in word production. 34
2. Interactive models. 41
3. Dual-stream Feedback Models 47
4. The Parallel Assembly Model 52
Conclusion 57

CHAPTER 3. PREDICTING MEANING IN THE DYAD 62

Abstract 63
Introduction 64
Methods 65
Participants 65
Materials 66
Task and procedure 67
Analyses 67
Results 68
Discussion 70

CHAPTER 4. NEURAL ALIGNMENT OF WORD MEANING 72

Abstract 73
Introduction 74
Methods 76
Participants 76
Materials 76
Task and Procedure 76
EEG data collection 76
EEG Data Preprocessing 77
EEG Data Analyses 80
Statistical Analyses 80
Results 81
Behavioural Results 81
MVPA 82
Interbrain Synchrony 83
Discussion 85

CHAPTER 5. NEURAL ALIGNMENT AND PREDICTIVE PROCESSING 90

Abstract 91
Introduction 92
Methods 95

Participants 95
Materials 95
Task and Procedure 96
EEG data collection 97
EEG Data Analyses 99
Statistical Analyses 100
RESULTS 100
Behavioural Results 100
MVPA 102
Interbrain Synchrony 103
DISCUSSION 107
CHAPTER 6. GENERAL DISCUSSION 110
MEANING-SPECIFIC NEURAL ALIGNMENT 111
LINGUISTIC AND NEURAL ALIGNMENT AND PREDICTION 114
MODELS OF LEXICAL ACCESS AND MEANING-SPECIFIC NEURAL ALIGNMENT 116
LIMITATIONS AND FUTURE RESEARCH DIRECTIONS 118
SYNTHESIS 120
REFERENCES 121

The agency or source of sensory consequences is determined not by the hidden states per se but by fluctuations in sensory precision that underlie turn taking. In this sense, the expectations are without agency – they are neither yours nor mine, they are our expectations.

Friston & Frith

Chapter 1. General Introduction

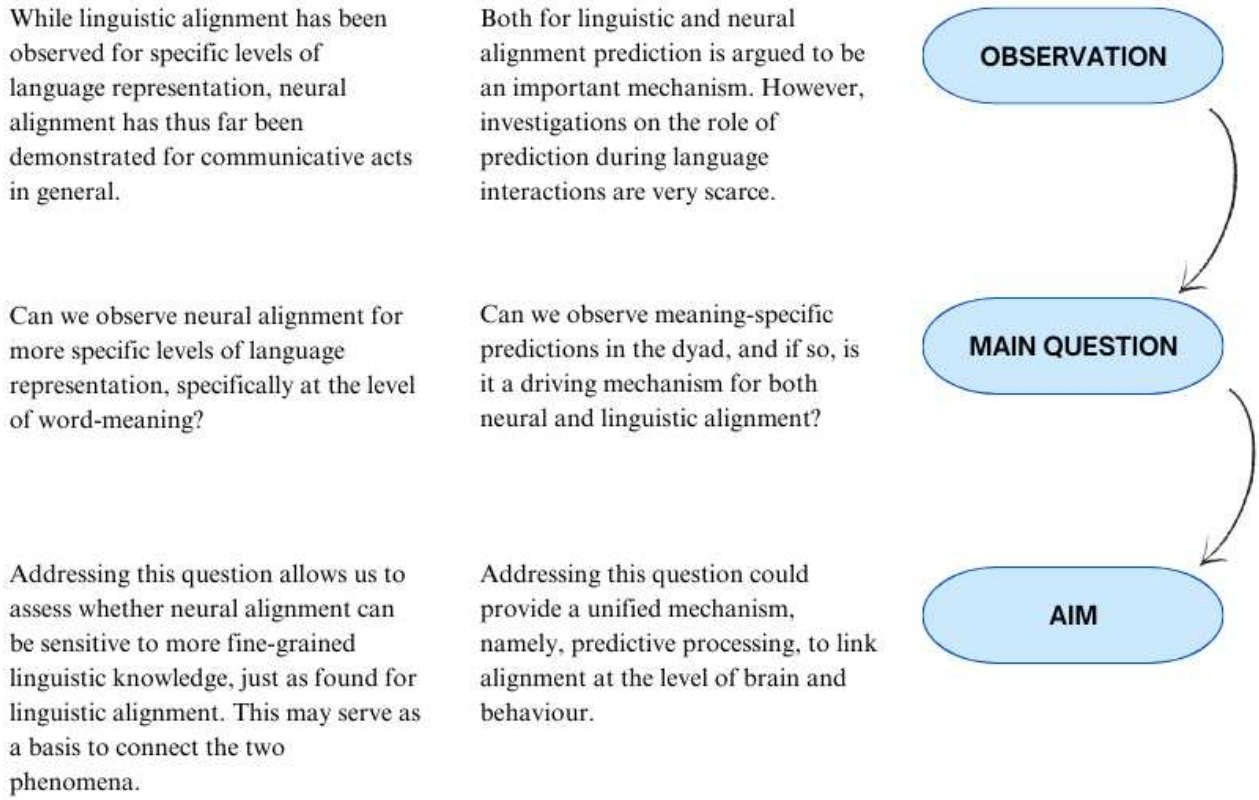
Research into the cognitive neuroscience of language processing during live human interactions has garnered significant interest in recent years, as it shifts the focus from isolated, individual cognition to the dynamic, interactive nature of communication. Hasson et al. (2012) propose the concept of brain-to-brain coupling, where the neural activity of individuals engaged in a conversation becomes synchronized, facilitating the creation and sharing of a social world. This synchronization allows for more efficient communication, as participants are not merely exchanging words but are also sharing mental states and intentions. This perspective underscores the importance of understanding language as a deeply social process, one that is embedded within and shaped by real-time interactions between individuals.

Complementing this, Schilbach et al. (2013) argue for a "second-person neuroscience," emphasizing the study of social interactions from a perspective that involves direct engagement between individuals. This approach challenges the traditional first-person (self-focused) and third-person (observer-focused) methodologies, suggesting that the nature of language processing can only be fully understood when considering the interactive, context-dependent dynamics of live communication. This relatively new approach to studying language and human communication can be related to Wittgenstein's concept of language games that proposes that the meaning of language arises from its use within specific social contexts, and these contexts inherently arise from communication. Indeed, as proposed in Philosophical Investigations (1953), the meaning of words and language in general is not fixed by reference to objects or facts found in the surrounding environment but instead it is derived from the various contextual uses that the words occur in. This view can be applied not only to the meaning of language but to the cognitive and neural processes that underlie language comprehension and production. Indeed, if we study language processing by analysing data extracted from individual people we might not be obtaining the full picture since the way we primarily use language is interactive and almost always involves more than one person.

Much of what we know about how language behaviour in humans is represented in the brain is based on task-based laboratory experiments that are limited to one individual speaker or listener. However, the complexity of this behaviour seems to be even greater when the perspective is shifted from one speaker to interactive speech acts between interlocutors. Making this shift to an interactive perspective is necessary given that dialogue is the most frequent form of language use (e.g., Clark, 1996; Pickering & Garrod, 2004; Levinson, 2016), especially since it seems that language (and cognition) as joint activity is not simply the individualistic model plus one (e.g., Sebanz et al., 2006; Kuhlen & Abdel Rahman, 2023). Even though people appear

to easily establish flowing conversations without much effort, how and why this is possible requires a dyadic, not solo, approach.

One interesting perspective to do so is alignment, which has been observed both in behavior and the brain and concerns one of the most prominent mechanisms to explain communication between interlocutors. In short, linguistic alignment concerns the phenomenon where interlocutors copy each other's language behavior, and neural alignment concerns the phenomenon where interlocutors' brain responses become correlated during interaction. In the present thesis I will focus on these forms of alignment from a dyadic perspective and will elaborate on the mechanisms that underlie them, with a special focus on trying to find a link between alignment in language behavior (linguistic alignment) with alignment of brain responses (neural alignment). The observations that motivate this thesis and the general questions and aims that it attempts to address are illustrated in the diagram below:



In **Chapter 1** I will detail theoretical background on the notions of *linguistic alignment*, *neural alignment*, and the role *prediction* is believed to play for both forms of alignment; **Chapter 2** will discuss different psycholinguistic and neurobiological models of lexical access which allows to derive predictions on which 'linguistic components' may align in a dyadic interaction and when (Kerr et al., 2023); **Chapter 3** will introduce the new experimental

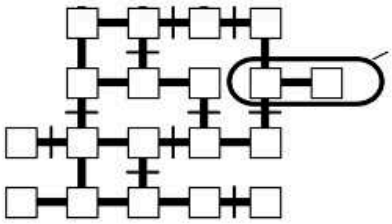
paradigm to explore simple dyadic interactions at the specific linguistic level of word meaning, and how prediction affects the behaviour of both interlocutors (Kerr et al., submitted); **Chapter 4** will present the EEG data recorded from both interlocutors (EEG hyper-scanning) performing simple dyadic interactions in the novel paradigm presented in Chapter 3, and demonstrate meaning-specific neural alignment (Kerr et al., in prep.); **Chapter 5** will discuss EEG hyper-scanning data on how prediction affects the meaning-specific neural alignment uncovered in Chapter 4; finally, in **Chapter 6** I will provide general conclusions on the potential link between linguistic and neural alignment, the role prediction may play in both forms of alignment and what the data presented in this thesis means for models of dialogue and neurobiological models of words in the brain.

1.1 On linguistic alignment

The first major theoretical pillar behind this project concerns linguistic alignment[1], a phenomenon known to occur when two or more interlocutors converge on mental representations during an interaction. These mental representations have been shown to include various linguistic levels. For example, people may converge on lexico-semantic choices, such as using the same word when two or more synonyms are available, even when the synonym is a less frequent word like 'couch' instead of 'sofa' (Brennan, 1996). Similarly, syntactical alignment takes place during a natural interaction, for instance, by repeating a prepositional construction instead of the one without (Reitter et al., 2006). Linguistic alignment can even occur at the level of phonology, for example, when two interlocutors converge on the same dialect (Coulston et al., 2002). Furthermore, it has been also demonstrated that people align not only on linguistic representations, but they also converge on co-occurring behaviours such as gestures and gaze (e.g., Dale et al. 2011; Schippers et al., 2010; Chartrand and Bargh, 1999; Rasenberg et al., 2020).

Experimental demonstrations of linguistic alignment date back to studies from the 1980s (e.g., Levelt & Kelter, 1982; Bock, 1986; Garrod & Anderson, 1987). Allow me to introduce below some of the seminal works that laid the ground in the field. In a series of experiments, Levelt and Kelter (1982) demonstrated that in question-answer interactions, people who gave answers tend to repeat, or converge on, the syntactic structure of the question given. For example, when asked *'(At) what time do you close?*', subjects' replies were directly influenced by the structure of the preceding question, that is, with or without the preposition depending on whether it was used in the question, *'(At) five o'clock*'. The authors claim that this convergence,

[1] Regarding the term *linguistic alignment*, it should be noted that various labels exist in the literature, including alignment, convergence, entrainment, coordination, etc. (e.g., Rasenberg et al., 2020). In this thesis I will mainly be using the term alignment for consistency.



A: I am one down from the upper right box
B: You mean the square just below the **right** corner?

A: Yes, the **square below** the upper **right corner**
B: So now you can move to the extreme **right square**

A: Yes, **I move to the extreme right square**

Figure 1. Example of alignment in the maze game between two interlocutors. *The left panel depicts an example of a maze where interlocutors have to move through by giving each other verbal instructions. The right panel gives an example transcript of a conversation between interlocutors A and B for the circled squares of the maze. In bold I highlighted when an interlocutor repeats a word, concept, or structure from the other interlocutor. As the conversation progresses, alignment increases. (Adapted with permission from Pickering & Garrod, 2006)*

or the 'correspondence effect' in their terms, is the retainment mechanism of preceding speech, which eases and accelerates the interaction by being more efficient than producing entirely new utterances.

Another example comes from Kathryn Bock (1986), using a priming design, the author showed that speakers align with the syntactic structure they were exposed to before. In this experiment, participants had to describe what they saw on a presented picture (e.g., the picture of *'a church and lightening'*). The syntactic structure they used to describe the picture was directly influenced by the structure of a sentence they had to listen to and produce just before (e.g., after an active prime, participants tend to say '*lightning hit the church*' vs. after a passive prime, they tend to say '*the church was hit by lightning*'). This demonstrates one's language choices, including abstract syntactic representations, are directly influenced by prior language input.

A final famous example of those early days of alignment research, and which in contrast to the syntactic priming study above measured alignment from both interacting interlocutors, is the maze-game developed by Garrod & Anderson (1987). In the maze-game two speakers had to co-ordinate their speech to find a way out of a virtual maze (Figure 1). When analyzing the produced speech, it became evident that the participants started to copy each other's lexical labels and semantic representations of space to successfully complete the maze.

Since these early demonstrations, linguistic alignment has been investigated rather extensively, demonstrating that interlocutors align on nearly all levels of linguistic representation. In fact, this sensitivity to specific linguistic information is a defining feature of linguistic alignment, and quite important in the current thesis since this aspect differs with neural alignment, which thus far has been demonstrated only for communication in general. Before going into more detail on neural alignment and how it may or may not relate to linguistic

alignment (see Section 1.2), I will describe some more examples of experimental studies showing alignment for these different linguistic representations, draw special attention to alignment of word meaning, which is the main component investigated in this thesis, and discuss the dominant cognitive model that can explain these different forms of alignment.

Alignment at the lexical level has been shown in numerous studies (e.g., Brennan & Clark 1996; Clark & Wilkes-Gibbs 1986; Wilkes-Gibbs & Clark, 1992; Garrod & Clark 1993; Garrod & Doherty 1994; Branigan et al., 2011; Nenkova et al, 2008; for an overview see Pickering & Garrod, 2021, Part 2). A good example of such lexical-specific alignment can be found in the work by Nenkova et al. (2008). The authors analyzed two corpora exploring the effects of high-frequency lexical items and how they relate to the dialogue's 'naturalness' and the task's success. The Naturalness parameter was annotated using the Switchboard Corpus, a collection of 6-minute telephone conversations, and was measured on a Likert scale (1 'very natural' – 5 'not natural at all'). The task's success was measured via the Columbia Games Corpus based on the game score (i.e., the higher the score, the higher the success). The results demonstrate that alignment on common words, that is, more frequent words overall and within a dialogue, predicts a more natural flow of the interaction and a more successful task completion.

Another level of linguistic representation that has received much attention in the alignment literature, as evident from the first studies discussed higher, is that of syntax (e.g.., Levelt & Kelter, 1982; Bock, 1986; Pickering & Branigan, 1998; Branigan, Pickering, Stewart, & McLean, 2000; Hartsuiker & Westenberg, 2000; Cleland & Pickering, 2003; Hartsuiker et al., 2004; for a review see e.g., Dell & Ferreira, 2016). Part of the appeal to study syntactic alignment concerns its abstract nature. That is, in the case of syntactic alignment what is copied are abstract, structural representations, but not (necessarily) the content or a mere repetition of the same words. For example, in Cleland and Pickering's (2003) study, participants were paired with a confederate, and they needed to describe images of objects to each other. Importantly, the task was conducted in turns, and after the trials when the confederate used a specific syntactic structure, namely, a pre-nominal adjective (e.g., ''*the green square*") or a relative clause (e.g., ''*the square that is green*''), the participant repeated that same structure when describing the next image. For example, if the confederate said, '*the square that is green*', the participant was likely to say '*the diamond that is red*' when it was their turn to describe the next image, that is, repeating the same grammatical structure. Similarly, Branigan et al. (2000) conducted a study on coordination of syntactic structure in dialogue and showed that a structure chosen by the experimenter's confederate (a prepositional object, e.g., *The pirate handing the cake to the sailor*, or a double object, e.g., *The pirate handing the sailor the cake*) affected the

syntactic structure of other speakers, that is, participants tended to choose the same syntactic structure as the one that was used by the confederate just before. For example, if the confederate described an image as '*The pirate handing the cake to the sailor'*, the participant was likely to describe the next image as '*The cowboy offering the gun to the robber'*.

While less studied and less robust than lexical and syntactic alignment, also at the level of phonology and phonetics alignment has been observed and where speakers tend to match their interlocutor' speech sound features (e.g., Bard et al. 2000; Babel & Bulatov, 2012; Babel, 2012; Nielsen, 2011; Fowler et al., 2003; Pardo, 2006; Pardo et al., 2017). For example, Pardo (2006) conducted an experiment where pairs of speakers repeated the same lexical items, and their production was then played to independent listeners. The listeners' task was to judge the degree of similarity between the spoken utterances. The results showed that the listeners perceived the utterances to be more similar between the two speakers when they were produced during an interaction compared to outside of the interactional setting. According to the authors, these data suggest that interlocutors tend to converge phonetically in an interactional environment.

Given that the objective of my thesis project is to test the link between linguistic and neural alignment at the lexico-semantic level (see Chapter 3 and 4), I would like to draw special attention to the studies that explore alignment of word meaning (Brennan & Clark, 1996; Branigan et al., 2011; Cirillo et al., 2022), in particular those investigating lexico-semantic[2] alignment in interaction rather than in the individual such as several of the studies mentioned above. Brennan and Clark (1996) explained such alignment as a result of a conceptual pact between two interlocutors, which they jointly construct as their interaction progresses. In their experiments, Brennan and Clark (1996) asked participants to sort identical sets of cards in the same order. Each participant had their card set in front of them and could not see the set of their partner. The recorded interactions that the participants had during the task show that interlocutors converge on lexical choices, and importantly, even infrequent ones. For example, after one interlocutor names a given card as *a loafer* instead of just *shoe*, their partner picks up on the same terms later in the discussion. This convergence is achieved, according to the authors, through establishing a shared conceptualization (a conceptual pact). Brennan and Clark note that these pacts are based on the frequency of past utterance and are partner- and interaction-specific. Branigan et al. (2011) report similar findings where speakers copy

[2] Please note that there is not always a clear distinction between lexical alignment (as described higher) and alignment of word meaning, and they may refer to the same phenomenon. I use the label 'lexico-semantic' here to refer to alignment which is not per se the repetition of the same word form (lexical label), but rather an alignment of lexical choices driven by conceptual representations.

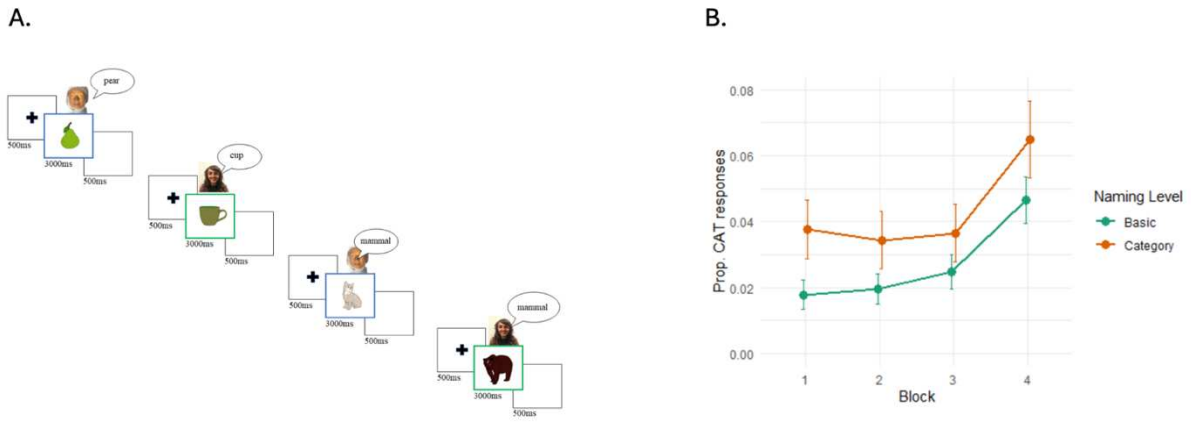


Figure 2. Panel A *illustrates an example procedure where the robot and the participant name pictures in turns.* ***Panel B*** *shows the main results with an increased proportion for the category (superordinate) name responses given by the participants after the robot produced category names (orange line) vs after the robot produced basic-level names (green line).*

infrequent lexical labels (e.g., saying *'coach'* instead of *'bus'*) to align with both a human partner or a computer (alignment was even stronger for the latter).

A particularly compelling demonstration of lexico-semantic (conceptual) alignment comes from recent work by Cirillo et al. (2022). The authors conducted a joint picture naming experiment where speakers were paired with a robot (see Figure 2 for the design and results representation). The main manipulation was that the robot was programmed to name a certain part of the pictures with a superordinate name (e.g., naming *'cat'* as *'mammal'*). Given the well-known basic level naming advantage (Rogers & Patterson, 2007; Alario et al., 2004; Almeida et al., 2007), where speakers tend to automatically utter the basic level name of an object, the authors thus created a situation where the robot was naming object in an a-typical manner. Interestingly, this manipulation significantly affected the naming behavior of the participants who adapted to the robot and started to increase their productions of superordinate names for pictures stemming from the same semantic category for which the robot gave a superordinate name. For example, when the robot named the picture of a *'cat'* as *'mammal'*, participants were more likely to name later on a picture of a *'bear'* as *'mammal'* as well (see Figure 2). Put differently, through lexico-semantic alignment the basic level naming advantage got on some occasions successfully overridden. Particularly neat here is that unlike in the work of Brennan and Clark (1996) and Branigan et al. (2011) where an infrequent word is repeated for the same input (e.g., picture of a *'bus'/'coach'*), the alignment in Cirillo et al., (2022) occurs for distinct input (e.g., picture of a *'cat'* versus *'bear'*). This leaves little doubt that the alignment emerged at the level of word meaning.

To conclude, it is important to mention that much research into linguistic alignment has been done via testing individual speakers as opposed to a dyadic interaction. For instance, and

as can be appreciated from several of the examples above, a common approach to study alignment concerned the priming of words, sentences and structures and assess how that affects the productions in an individual participant. Partially, this stems from a traditional single-person approach in psycholinguistics and neuroscience. Notably though, the shift towards a multi-person perspective has been perceived as an important step towards deepening our understanding about the mechanisms that underpin human communication (Schilbach et al., 2013). Indeed, when attempting to tease apart the processes that foster verbal interaction between two (or more) speakers, it appears to be logical to study this by taking the dyad as an essential and integral unit. This becomes even more important when considering that cognitive and neural processes studied in interaction are not necessarily similar to the individual model times two, and therefore an approach is required that addresses social and interactive behaviours from a joint perspective (e.g., Sebanz et al., 2006; Brown-Schmidt & Tanenhaus, 2008; Kuhlen & Abdel Rahman, 2023). This systemic approach is not entirely new; however, it has not been prevalent when tackling notions like linguistic alignment, hence, many issue remain to be explored further and in more detail through implementing dyadic and interactive designs. A few studies have adopted this view, some of which we discussed above (e.g., Garrod & Anderson, 1987; Brennan & Clark, 1996; Corps et al., 2018; Gambi et al., 2015; Kuhlen & Abdel Rahman, 2017; Cirillo et al., 2022), which is important for present purposes since neural alignment is a phenomenon that requires (at the least) two interacting persons. Before discussing the details of the dyadic paradigms and observations in neural alignment, I will first describe the dominant model in the literature to explain linguistic alignment.

1.1.1. The Interactive Alignment Model

A model that can accommodate a wide range of the linguistic alignment findings discussed above is the Interactive Alignment Model developed by Pickering & Garrod (2004) (Figure 3). This model posits that successful communication requires 'the development of aligned representations by the interlocutors' (p. 175). According to the model, the key mechanism that allows alignment of representations is ***priming***. This mechanism is believed to be (a) ***automatic***, which assures an effortless and largely problem-free interaction, and (b) ***bidirectional***, which means that if there is alignment in one direction, e.g., from production to perception, there must be alignment in the opposite direction, i.e., from perception to production. Allow me to illustrate how the model explains linguistic alignment taking Bock et al. (1986) as an example: when speakers produce a specific syntactic structure, for example, a sentence in the passive voice, they inherently prime their conversational partner's production system with that specific syntactic structure. Thus, when it is the partner's turn to produce a sentence, they are more likely to implement the same syntactic structure in their own speech.

This priming of the syntactic structure stems from perception of their partner's speech. Another important feature of the model is that alignment occurs at various linguistic levels so that alignment at a lower level percolates to alignment at higher levels, which altogether lead to an easier flow of a given verbal interaction. A well-known example here is the lexical boost, which shows that people align even more on syntactic representations if there is also alignment at a lower level, namely words (e.g., Pickering & Branigan, 1998; Traxler et al., 2014). A final central notion of the Interactive Alignment model is the ***parity*** between language production and perception. Dispensing with the classical notion where language production and perception rely on partially separated processing systems (e.g., Frazier, 1987; Levelt, 1989; for an overview see Meyer et al., 2016), the model proposes interactive activation at all levels of representations between the two modalities. Given that the activation of perception representations would lead to the activation of production representations and vice versa, the model can explain why linguistic alignment occurs at different levels of the language architecture.

While there are other (often mutually non-exclusive) mechanisms proposed to explain (some forms) of alignment, like audience design (e.g., Clark, 1996) which concerns a conscious adaptation strategy to our interlocutor rather than an automatic process as described by Pickering and Garrod (2004), nevertheless the Interactive Alignment model is likely the most studied model of alignment in psycholinguistics. Perhaps because of that, it is all the more surprising that a comprehensive neural implementation has never been materialized. The

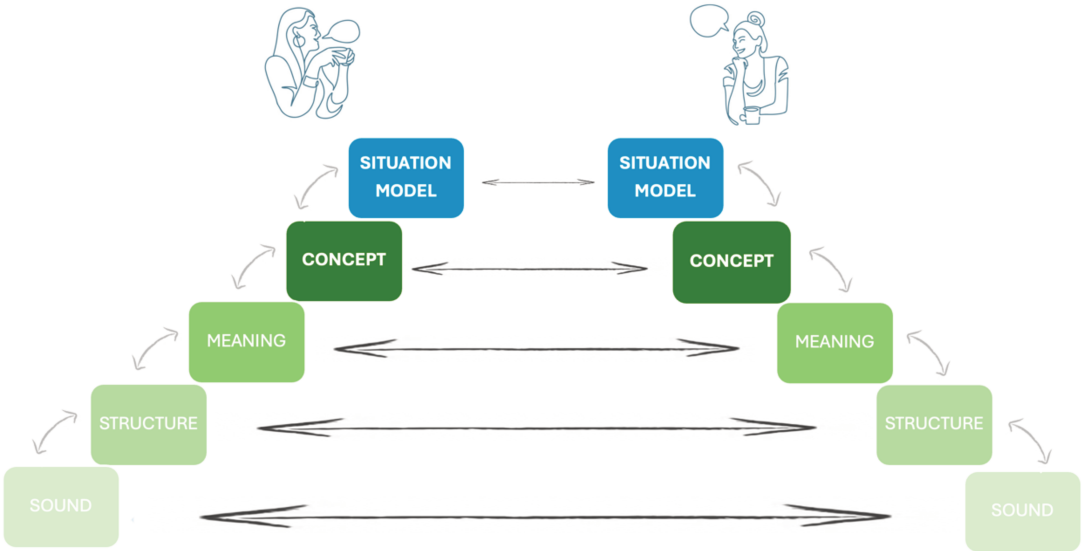


Figure 3. A schematic representation of the Interactive Alignment Model *that illustrates two interlocutors and their stages of language production and perception during an interaction. It illustrates the process of alignment, where grey arrows demonstrate the percolation of alignment at different linguistic levels and the black arrows represent different channels at which alignment occurs. Blue squares indicate higher level processing, while green squares indicate lower level processing. (Adapted from Pickering and Garrod, 2004).*

discoveries of neural alignment between speakers and listeners over the last decade may signify an important advance to start connecting the cognitive mechanisms of interactive alignment with neurobiological activity found for communication. Of course, to be able to do so, the first step is to assess whether both phenomena are (at least in part) related to each other.

1.2 On neural alignment

We have seen that at the behavioral level conversational partners align at various linguistic levels to mediate a successful flow of the interaction. A consecutive question then arises as to how in the neural system this alignment is achieved. Do the neural signatures of two or more brains also converge, or synchronize, as a result of the linguistic convergence? If so, can we capture this brain-to-brain synchronization with the existing experimental methods? Collaborative efforts in social neuroscience predict that the answers to these questions appear to be positive. Below I will cover the main advances in the study of brain-to-brain synchronization, how it can be done and what it means in the broader realm of research into social interaction.

Brain-to-brain synchronization, also termed as neural alignment, is the phenomenon when the cortical activity of two or more participants doing a joint task is correlated in the spatial and/or temporal domains and can inform us of the neurophysiological signatures that underpin mutual understanding, cooperation and communication in general between individuals. The measures of brain-to-brain synchronization include oscillatory activity from the cortex, that is, how the signal from two (or more) brains covariate in the temporal domain, and correlation of brain activity in specific brain regions, in other words, when an activity in region X of the interlocutor A corresponds to the activity in the same regions X of interlocutor B. Neural alignment was first observed in social tasks of joint activity, for example, playing card games (Montague et al, 2002; Babiloni et al., 2006), gestural activity (Schippers et al., 2010), imitation of hand movements (Dumas et al., 2010), processing music (Babiloni et al., 2012; Lindenberger et al., 2009), among many others.

Perhaps not surprisingly, neural alignment research was expanded to language processing since talking to other people is an example of joint action that most of us exhibit on a daily basis. One of the seminal studies that looked in neural alignment during a verbal interaction was the work by Stephens et al. (2010). The authors used functional magnetic resonance imaging (fMRI) and compared the brain activity between a speaker reading a story and listeners comprehending that same story. They observed a wide-spread correlated brain activity in Perisylvian language areas and beyond. Moreover, they observed that neural alignment was stronger in those listeners who had better comprehension of the story. The

authors concluded that there is much more neural overlap between the encoding and decoding of language than originally presumed, and this overlap underlies neural alignment which facilitates communication success. Later on, neural alignment between speakers and listeners has been extended to naturalistic narrative (e.g., Silbert et al., 2014, Liu et al., 2020), predictive sentence processing (e.g., Dikker et al., 2014), discourse processing (e.g., Heidlmayr et al., 2020), teacher-student interactions (Dikker et al., 2017) and second language processing (Perez et al., 2019).

Methodologically, research into neural alignment takes advantage of various methods available to neuroscience and recording the cortical activity from two (or more) brains (near) simultaneously is termed hyperscanning. Hemodynamic changes measured via fMRI allow to explore which brain regions in one participant correlate with the activity from the other participant (e.g., Stephens et al., 2010; Montague et al, 2002; King-Casas et al., 2005). Although fMRI provides high spatial resolution, it poses certain limitations for the hyperscanning research. First, it does not allow the detailed investigation of the temporal changes of the cortical activity, which appear to be important since the changes that form neural coupling between two brains happen rapidly. Secondly, fMRI setting allows very little mobility for participants and present a lot of noise due to its technical specifications, hence, the number of manipulations that can be performed with this technique is limited and does not appear to simulate an ecological and naturalistic environment (Koike et al., 2015; Kelsen et al., 2022). Thirdly, most of the time fMRI recordings cannot be done simultaneously. Instead, one participant goes in the scanner as a speaker and their brain activity is recorded, and at a later stage, the other participant goes in as a listener, which means that the recorded BOLD-responses did not occur at the same time (as was the case for instance in the study of Stephens et al. (2010) discussed above).

Another popular hyperscanning method concerns research into how oscillatory activity in one brain correlates with oscillatory activity in another brain during language processing has been conducted by means of electroencephalography (EEG)[3] (e.g., Kuhlen et al., 2012; Dikker et al., 2017; Perez et al., 2017; 2019). Thus far, using hyperscanning EEG seems to be a widely used solution in testing two or more participants engaging in a joint task. This can be explained by its cost efficiency, the relative mobility that this technique allows, virtually no noise coming from the machinery itself unlike in fMRI, ensuring the possibility of having a naturalistic verbal communication between participants, and, most importantly, what it allows to measure, namely,

[3] Apart from EEG, MEG and functional near-infrared spectroscopy (fNIRS) has also been used to investigate the interbrain synchrony with a more detailed temporal resolution (e.g., Cseszumski et al., 2020; Balconi et al., 2017; Cheng et al., 2015); although in my thesis I am focusing on EEG studies, where necessary I will mention studies that involve other techniques.

simultaneous rapid changes in cortical activity in both (or more) participants. Below I will discuss several of these studies to demonstrate the state-of-the-art and general conclusions we draw thus far from EEG hyper-scanning studies on language.

Kuhlen et al. (2012) designed a study where a part of participants was recorded while narrating a story, and the other part of participants then listened and watched the recording of the narration. The main manipulation was that during the listening part of the experiment, participants were shown superimposed videos of two speakers, one female and one male, and they were instructed to only attend to one of them. During both the story telling and the listening the EEG activity was continuously recorded. The data obtained suggested that the listeners' EEG activity was more strongly correlated with the EEG activity of the speaker they were attending to. Perez et al. (2017) explored interbrain synchronization during a verbal interaction in dyads where participants exchanged short narratives on a given topic and their EEG activity was simultaneously recorded. The analyses included the computations of brain-to-speech synchrony (brain-speech envelope synchronization) and brain-to-brain synchrony. The study showed that the observed brain-to-brain synchrony was not limited to the auditorily induced speech-to-brain synchrony but that it also emerged as a result of the interactive processes. Put differently, the neural synchrony between interlocutors arises due to the verbal communicative act that they are engaged in. The authors emphasize that the mechanisms behind language processing during an interaction cannot be understood when studying interlocutors in isolation but instead, it is crucial to implement the two-person neuroscience framework. This framework has also been expanded to even higher-order language functions such as group-dynamics (e.g., Dikker et al., 2017) and turn-taking dynamics (e.g., Ahn et al., 2018).

While the first EEG hyperscanning studies explored neural alignment between speakers and listeners in general, more recent work stated to implement more specific manipulations. One example comes from a study by Perez et al. (2019). The authors recorded electrical activity over speakers and listeners while they were engaged in a simple information exchange covering common everyday topics (like Perez et al., 2017). The key manipulation was whether participants were speaking and listening in their native language (Spanish) or their second language (English). After seeing 5 guiding questions relating to one of the topics (sports, travel and food for the English condition, and movies, animals and music for the Spanish condition), one of the participants was assigned the role of the Speaker and had to elaborate on that topic while the participant in the Listener role had to attend to the speech of their partner. After each trial participants exchanged roles so that each person completed the task in both roles. The results showed that within a dyad, in both conditions participants' brain oscillatory activity synchronized in the alpha band, and this synchronization was significantly different between

the two conditions. More specifically, 5 electrode pairs showed larger synchronization for the native language condition, while 9 pairs showed larger synchronization for the foreign language condition. The fact that both conditions induced different patterns of interbrain synchronization indicates both interlocutors activated distributed language-specific networks (native vs. foreign), and those networks synchronized during the verbal interaction. Showing that in both the native and foreign conditions there is distributed interbrain synchronization is interesting and relevant in light of the current thesis because it demonstrates that the phenomenon of neural alignment can be sensitive to language manipulations.

Another example relevant for the current thesis of a hyperscanning study relying on a specific linguistic manipulation is Liu et al. (2019). Liu et al. compared the interbrain synchrony in pairs of participants who described pictures to each other while using either a syntactically consistent structure (always double-object (DO) or always prepositional-object (PO)) or a syntactically inconsistent structure (alternating between DO and PO). A significant increase in the interbrain synchrony in the right posterior superior temporal cortex was observed in the consistent conditions, that is, when both participants used the same syntactic structure. These findings propose that interlocutors share neural representations of syntax during a verbal interaction. To my best knowledge, this is the only study thus far which observed neural alignment of specific linguistic representations, namely syntax, in line with observations in linguistic alignment. However, this work must be confirmed and extended because a between-participants design was used, and they observed the effect only in one of the syntactic conditions (double-object). Therefore, it remains possible the effect was not syntax-specific alignment per se but indexed general communicative alignment in only one of the participant groups. Nevertheless, both the data from Perez et al. (2019) and Liu et al. (2019) suggest that neural alignment is sensitive to language manipulations and therefore might go beyond general communicative engagement.

Not surprisingly thus, the findings from neural alignment studies were met with enthusiasm since they can index the neural expression of linguistic alignment as observed in behavior (e.g., Pickering & Garrod, 2014; Kelsen et al., 2022). That said, at present such link has not been empirically established. As could be noted from the overviews given in the above sections, while linguistic alignment deals with specific language representations such as sounds, syntax, words and meaning, neural alignment predominately handles about communication in general and remains rather mute about the specific levels of language representation (see also Section 1.4 later). To find a possible relationship between linguistic and neural alignment, a first step is to assess whether neural alignment, in accordance with linguistic alignment, is sensitive to specific language representations rather than communicative interaction in general.

Exploring this for word meaning is one of the main objectives of this thesis (see also Section 1.4), and thanks to the method of EEG hyperscanning we have a method that allows us to explore the neural processes that form linguistic alignment. However, before detailing how the experimental part of this thesis attempts to contribute to this exciting topic and how we try to establish a more precise link between neural and linguistic alignment, I will dedicate some words on the topic of prediction first, since predictive processing is of importance for both forms of alignment and may be the link that connects neural and linguistic alignment.

1.3. On prediction

The human brain is not just a passive processor of incoming information but rather a complex system that predicts upcoming input by comparing it against generative models that it has learned through experience. This idea has been integrated in multiple fields of cognitive neuroscience, and empirical evidence suggests that one's behavior and their understanding of the others' behavior and the world around them is largely governed by prediction (e.g., Parr et al., 2022; Pezzulo et al., 2024; Friston, 2005; Friston & Frith, 2015). Consequently, the role prediction may play for producing and comprehending language, has been a central topic in the field over the last few decades.

Overall, prediction facilitates speech production and comprehension, aids in avoiding errors and disfluencies, and ensures successful communication in conversation (e.g., Kutas et al., 2011; Pickering & Garrod, 2013; Huettig, 2015; Kuperberg & Jaeger, 2016; Pickering & Gambi, 2018). Furthermore, many behavioral and neuroimaging studies have shown that prediction can occur at various levels of language processing, including semantics, syntax, and phonology (e.g., Altmann & Kamide, 1999; Federmeier & Kutas, 1999; DeLong et al., 2005; Van Berkum et al., 2005; Staub & Clifton, 2006; Grisoni et al., 2017; Ito et al., 2018; Strijkers et al., 2019).

Seminal findings on prediction of word meaning, the level of investigation in the current thesis, often concern the contrast of predictable versus unpredictable content following context sentences, both with behavioral measures such as eye movements and neurophysiological measures such as EEG (e.g., Altmann and Kamide, 1999; Federmeier & Kutas 1999; Kutas & Federmeier 2011). The paper by Altmann and Kamide (1999) explored how listeners incrementally interpret sentences by predicting upcoming information based on verbs. The authors relied on the measurements of eye-tracking in a visual world paradigm (e.g., Tanenhaus et al., 1995). In this paradigm participants were presented a visual scene that contained a few objects, for example, a toy train set, a toy car, a balloon, and a birthday cake. At the same time a prerecorded sentence was played. Crucially, for each scene the sentence could be either highly

constrained, that is, the verb of the sentence strictly dictated only one object from the scene, for example, '*the boy will eat the cake'*, or unconstrained, where the verb permitted all the objects from the scene, for example, '*the boy will move the cake'*. The eye-tracking data showed that when the scene was accompanied by a highly restrictive sentence participants were likely to fixate on the predictable object prior to the auditory completion of the sentence. The authors demonstrated that upon hearing a verb, listeners immediately use it to constrain their expectations about what objects are likely to be mentioned next. This rapid and anticipatory processing reflects how listeners use their understanding of the verb's meaning and its typical arguments to guide their interpretation of the sentence in real-time. The study provides evidence that language comprehension is an active, predictive process, rather than a passive one, with listeners continuously integrating linguistic input with contextual knowledge to anticipate future references.

Federmeier and Kutas (1999) investigated how long-term memory structures influence real-time sentence processing, focusing on how people anticipate and integrate upcoming words. Using event-related potentials (ERPs), the authors recorded participants' brain responses to sentences that ended with a highly predictable item, an unpredictable item from the same category, or an unpredictable item from a different category. For example, a sentence '*They wanted to make the hotel look more like a tropical resort. So along the driveway, they planted rows of...'* could be completed with *palms*, *pines* or *tulips*. The ERP component of interest was the N400, a negative-going potential that peaks at about 400 ms post-stimulus inset. This component has been especially robust as a response to semantically anomalous contexts (Kutas & Hillyard, 1980). The results showed that when the sentence was ended with a predictable item ('*palms'* for the example above), the N400 response was the smallest, increasing for the unpredictable item from the same semantic category ('*pines'*) and the biggest for the unexpected item from a different category ('*tulips'*). This indicates that readers generate expectations about forthcoming words. When these expectations are violated, the brain exhibits specific neural responses (N400), indicating that sentence processing is highly predictive and influenced by prior knowledge stored in long-term memory. The study suggests that comprehending language involves dynamically accessing and using stored information to predict and make sense of ongoing linguistic input. Meaning predictability has been demonstrated time and time again, both in behavior and for brain measurements (for reviews see, e.g., Kutas et al., 2011; Huettig, 2015; Pickering & Gambi, 2018). This leaves little doubt that prediction of meaning is an important part of language processing.

Effects of prediction are not exclusive to meaning and have been reported for syntax and phonology as well (van Berkum et al., 2005; Staub and Clifton, 2006; Delong et al., 2005;

Nieuwland et al., 2018). However, it does seem to be the case that those effects are less robust and systematic than at the meaning level (e.g., Pickering & Gambi, 2018). This also is one of the reasons why the current thesis focusses on word meaning, that is, the strongest variable, to start assessing possible relationships between linguistic and neural alignment and explore the role that prediction may play in such alignment.

Given the numerous demonstrations and importance of prediction for language processing, it was proposed that prediction may in fact be the mechanism that drives or at least enhances alignment between interlocutors (e.g., Pickering & Garrod, 2013; Pickering & Gambi, 2018). Because interlocutors in the dyad try to predict each other's utterances, the same language representations in both individuals become activated, allowing for mimicking behavior. Pickering and Gambi (2018) developed a theory of prediction in dialogue (Figure 4), where predictive processing is viewed as *general,* meaning that it occurs at all linguistic levels, phonological, semantic, and syntactic, and is production-based since this theory suggests that the same mechanism is implemented when interlocutors are comprehending the other's speech and when they are producing speech themselves. Overall, as can be seen in Figure 4, the comprehension system implements the mechanism of prediction-by-production. This mechanism involves covert imitation of attended speech, which leads to understanding the

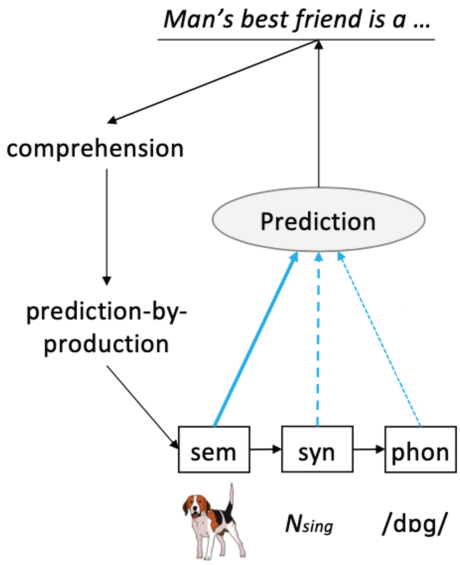


Figure 4. A schematic illustration of the prediction-by-production system during comprehension as proposed by Pickering and Gambi (2018). *Solid lines show essential comprehension processes, while dashed lines are optional. When hearing a sentence like 'Man's best friend is a…' comprehenders covertly imitate their interlocutor and then use their production system to estimate the most likely continuation (here: 'dog'). (Adapted from Pickering & Strijkers, in prep.)*

utterance[4]. Empirical evidence that the production system is actively involved in predictive processing comes from several studies (Martin et al., 2018; Lelonkiewicz et al., 2021). For example, in Martin et al. (2018) participants read highly predictable sentences that ended with either expected or unexpected word. Interestingly, the final words differed in gender (the experiments were conducted in Spanish), and when the sentence ended with an unexpected word, and, hence, the preceding article did not grammatically correspond with the expected word, the classic N400 effect (e.g., Kutas & Hillyard, 1984) was observed already when encountering the unexpected article. However, when participants were tasked to repeatedly pronounce syllables at the same time as comprehending the sentence, thus, taxing their production system, the N400 effect at the article level disappeared. This clearly suggests that when the production system is inhibited during comprehension it interferes with prediction. Pickering and Gambi's theory proposes that in order to understand their interlocutors, people first understand a given context (e.g., "*The boy went outside to fly a...*") and then transform that understanding into representations as if they were generating the speech themselves. These representations are used to form the intention that would naturally continue the interlocutor's sentence. The prediction process involves going through the stages of language production internally, without actual speech. The theory also suggests that since language production is demanding and resource-intensive, not all levels of representation are always computed, and the prediction process can be halted at any stage.

Interestingly, at the level of the brain, a similar mechanism is proposed which could explain how neural alignment comes about (e.g., Friston & Frith, 2015; Schoot et al., 2016). The dominant theory here is that of active inference which is based on the notion of predictive coding (Friston et al., 2005; Friston and Frith, 2015; Parr et al., 2022; Pezzulo et al., 2024).

Although this theory has been gaining scientific empirical attention rather recently, its motivation dates back to late 19th – mid 20th centuries. Several foundational theories lie behind the modern understanding of predictive coding and active inference. Below I will mention a few as explained in Pezzulo et al. (2024). For example, Craik (1943) proposed that the human brain develops a small-scale model of the world and before reacting to a stimulus, it uses its model to simulate the consequences of its actions. Another foundation is the idea of the Bayesian brain where during perception the brain uses an internal generative model to make an inference on a given state of affairs (Helmholtz, 1866; Doya et al., 2007). From the neurobiological perspective, this mechanism of perception-by-inference can be explained via the predictive coding (Pezzulo et al., 2024). One more historical basis for the development of

[4] It should be noted here that different models of lexical access in language production and comprehension, namely, serial and parallel accounts, have varying implications on how neural alignment and prediction operate. This will be discussed in more detail in Chapter 2.

the active inference theory is the notion that in order to understand the environment and act upon it (when necessary), an organism needs to be actively engaged with that environment. Put differently, perception builds upon anticipating one's own actions (or lack thereof) and their consequences on the outer world (Gibson, 1979).

As stated in Pezzulo et al. (2024), the historical roots of active inference are somewhat disconnected ideas that have been combined and developed into a unified perspective. This perspective's main idea is that the means by which living organisms perceive their environment and interact with it is the minimization of prediction errors (or free energy). This is achieved, according to the theory, by developing a generative model of the environment and testing it against a given sensory input. Active inference theory has been applied in studies of various domains of human cognition such as decision making (Parr et al., 2023; Rens et al., 2023), action-perception and higher cognition (Friston, et al., 2017; Pezzulo et al., 2015, 2018), and goal-directed behavior (Ahmadi & Tani, 2019). Not surprisingly, perhaps, active inference and predictive coding theory have also been extended to examine verbal communication (Friston and Frith, 2015; Schoot et al., 2016).

As stated in Friston and Frith (2015), when it comes to human interaction, the mechanism that we use to understand each other and then reply is basically the same as the one engaged in action and perception processing in general, and that mechanism is the minimization prediction error. As mentioned above, this major underlying principle is accomplished by constituting a generative model and testing it against the sensory input: while listening, one brain attends to the sensory input (sounds) which enacts the generative model in order to predict the speech of their interlocutor, whereas while speaking, the other brain (the interlocutor) uses their generative model (that is, predictions) to produce utterances. What here seems to be of major importance is the observation that in a communicative situation the interlocutors share the same generative model, or put differently, their initially separate generative models synchronize. In this context, successful communication occurs when the listener's generative model aligns with the speaker's, leading to generalized synchrony, where the neural states of both parties become coupled. Essentially, what can be inferred is that when the two systems are communicating, both engaging active inference system to predict each other's input, they establish 'generalized synchrony' (Friston and Frith, 2015, p.136).

In that regard, predictive coding during verbal interactions is amodal – as comprehenders, we predict what our interlocutor is about to say by implementing the same generative model as when we are planning our own speech. Generalized synchrony suggests that when a listener accurately predicts a speaker's utterance, their brain activity mirrors that of the speaker. This coupling occurs because both speaking and listening are driven by the same

underlying cognitive processes, where predictions are amodal and apply to both language production and comprehension. The extent of neural coupling, therefore, reflects the alignment of generative models between the speaker and listener, which varies across different levels of linguistic representation. Evidence from studies like Dikker et al. (2014) supports this hypothesis, showing stronger neural coupling when the listener successfully predicts the speaker's utterance.

Another important feature of active inference during communication is that it is of dynamical nature. As shown with birdsong simulations, Friston and Frith (2015) demonstrate that active inference updates as the interaction progresses, and the participants of a given interaction synchronize almost immediately, taking turns when appropriate and responding to each other's input (chirping in case of the songbird simulations, and speaking/responding in case of a human dialogue).

Despite the interesting correspondence between brain and behavior and the role that prediction may play in explaining both linguistic and neural alignment, scientific studies on how prediction may affect language processing in the dyad are scarce (e.g., Corps et al., 2018). Nevertheless, a few studies did examine the question and could give us useful hints on how to proceed from there. Some evidence suggests that interlocutors constantly and easily anticipate each other's utterances. This is evident from instances when one speaker interrupts or finishes the utterance of the other speaker without delay (Pickering & Garrod 2013; Dikker et al., 2014). In Stephens et al. (2010) study an intriguing finding was that in some areas of activation listeners preceded the speaker, which as the authors concluded is the result of predictive processing employed by the listeners. Interestingly, those listeners who showed certain activation patterns prior to the speaker also exhibited better comprehension of the narrative that was measured by a follow-up questionnaire. This finding can serve as a potential indication that prediction is the mechanism that drives a fluent and smooth interaction.

Prediction in interaction has also been investigated in how it affects turn-taking in listeners (e.g., Levinson, 2016). In this work it has been found that participants are faster responding to questions, either with button-presses (e.g., De Ruiter et al., 2006; Magyari et al., 2014; Ryskin et al., 2020) or verbal responses (e.g., Bogels et al., 2015; Corps et al., 2018; see also for written responses: Corps & Pickering, 2023), when the content of those questions is predictable. For instance, Corps and colleagues (2018) observed that people are faster in giving simple yes/no replies to questions when sentence meaning is predictable (e.g., *Are dogs your favorite animal*?) compared to when the content of the question is not predictable (e.g., *Do you enjoy going to the supermarket?*). These results are in line with the idea and supporting evidence that prediction aids language comprehension in individual speakers (e.g., Altmann & Kamide,

1999; Staub & Clifton, 2006; Ito et al., 2018), and they demonstrate that listeners' speech planning is facilitated via prediction. Nonetheless, whether and how prediction, and specifically prediction of word meaning, may affect the verbal behavior of the other interlocutor in the dyad is at present unclear. A second objective of this thesis lies in starting to understand better the role of prediction in the dyad and how it may affect neural and linguistic alignment between interlocutors.

1.4 Interim Discussion

In an interim summary of the above-mentioned theoretical accounts I would like to draw on certain common features that neural alignment studies share. First, all contribute to the shift towards a multi-person neuroscience where the neural correlates that underpin human interaction and communication are investigated by testing pairs, or dyads, of participants, or even groups, as opposed to isolated subjects. Second, so far, the hyperscanning literature successfully demonstrates that the natural flow that human interaction normally poses is largely established through significant interbrain synchronization, that is often (though not limited to) measured by phase covariance[5] in the two (or more) brain signals. The majority of these studies tend to create a (semi) naturalistic experimental set-up where the brain activity of the interlocutors is measured over the course of the entire interaction, for example, for the whole duration of a conversation, or a narrative. This approach allows to assess whether people's brain signals synchronize when engaging in interactions. Narrowing down to verbal interactions, this claim holds as long as higher-level linguistic processing is concerned. Put differently, thus far, interbrain synchronization has been evident from data that do not differentiate between the various linguistic levels of processing, like semantics, syntax, phonology, etc. This is different for linguistic alignment discussed in Chapter 1, that has been proven to happen automatically and bidirectionally at all linguistic levels. It is currently unclear which specific levels of language processing—whether lexical, semantic, phonological, or syntactic—are sensitive to neural alignment. Previous studies on neural alignment have generally taken a broad approach, often measuring neural alignment across entire interactions, leaving us unsure about which particular linguistic elements of the overall communicative signal become neurally aligned (e.g., Schoot et al., 2016; Kelsen et al., 2022). Although a key feature of linguistic alignment is that it occurs across most, if not all, linguistic representations (e.g., Pickering & Garrod, 2004), it is possible that neural alignment might be limited to higher levels of linguistic processing, such as shared understanding or situation models (e.g., Schoot et al., 2016). The present thesis

[5] Methods that measure phase covariance include phase-locking value (Lachaux et al., 1999), Circular Correlation Coefficient (Jammalamadaka & Sengupta, 2001), Kraskov mutual information (KMI) (Kraskov et al., 2004). Other measurements also exist, however. For example, coupling in the changes of EEG amplitude or power, or partial directed coherence (PDC) that allows to measure the direction of the information flow (Baccala & Sameshima, 2001).

will investigate this question for word meaning. More precisely, in Chapter 4 I will explore whether neural alignment occurs at lower levels of linguistic processing.

Another issue relates to how language is represented during production and perception (e.g., Menenti et al., 2012; Fairs et al., 2022). A key assumption underlying neural alignment is that language encoding and decoding share the same representation (e.g., Stephens et al., 2010; Hasson et al., 2012). However, models that explain linguistic alignment often suggest, directly or indirectly, that representations during production and perception are closely connected but still distinct (e.g., Pickering & Garrod, 2007; 2013; Hickok, 2012; Dell & Chang, 2014; Pickering & Gambi, 2018). Essentially, we make predictions about the meaning, grammar, and sounds of upcoming words based on mental representations in the opposite modality: when we speak, we predict what our spoken words will sound like using our comprehension system and compare those predictions with our speech planning in the production system. Similarly, when listening, we use our production system covertly to predict likely words we are about to hear and check those predictions against the actual auditory input. In other words, in a system where language input and output are compared, words require two representations, one in production and one in perception. This idea is in contrast to what is assumed in neural alignment studies, that, as I mentioned above, posit that brains neurally align because they essentially use the same, single, representation during encoding and decoding (see also Chapter 2 of the present thesis).

Given the widespread findings of neural alignment within the language network, intriguing questions arise about the relationship between linguistic and neural alignment. If we assume a correspondence between these types of alignment, does this imply that linguistic representations are shared across language modalities (e.g., Menenti et al., 2012; Strijkers & Costa, 2016; Kerr et al., 2023)? If so, what does this mean for the cognitive mechanisms underlying interactive alignment and prediction (e.g., Fairs et al., 2022)? Alternatively, could neural alignment be related to linguistic alignment only at higher levels of representation, such as shared understanding or situation models, where common systems across language modalities are typically assumed (e.g., Schoot et al., 2016)? Or perhaps the two forms of alignment are entirely distinct, representing separate cases of biological convergence that require different explanations? While the idea of linking linguistic and neural alignment is certainly exciting, the reality is that we currently lack the data and understanding to establish this connection, particularly when it comes to the fundamental components of language. Therefore, it is crucial to determine which specific word representations are responsive to neural alignment and whether this responsiveness translates into linguistic alignment. Furthermore, these research questions should be explored within the same interactive context for both types

of alignment: the dyad. Some of these questions will be discussed in greater detail in subsequent experimental chapters of the present thesis.

Chapter 2. Lexical access in speech production

Prior to delving into the experimental part of the present thesis, it is important to cover a question that appears to be of significance when trying to find a link between linguistic and neural alignment. The question is whether the linguistic components of words - such as meaning, grammar, and sounds – share neural representations (space) and temporal dynamics (time) between language production and perception. This question is pertinent because the general idea behind neural alignment is that it occurs when decoding in the speaker and encoding in the listener share neural representations and/or the time-course of processing (e.g., Hasson et al., 2012; Schoot et al., 2016). While most cognitive and neurobiological models of language agree by now that higher-levels of language processing like message- and discourse-levels strongly overlap (or even are identical) between production and perception, at more basic levels of language processing such as the lexical level much more debate exist. Given that the current thesis focusses on the lexico-semantic level of processing, this chapter wants to address the question about overlap across the modalities for lexical access.

To answer this question, I wish to rely on the book chapter we recently published and is added below, and which gives an overview of the main models of lexical access in psycho- and neurolinguistics. In this book chapter, we first discuss two dominant theoretical frameworks on lexical access in language production: (a) ***serial models*** where different components of words are active at different points in time and engage distinct layers of processing (e.g., Levelt et al., 1999; Indefrey & Levelt, 2004); and (b) ***interactive models*** where the lexical architecture still is hierarchical in nature, but overlap in time-course and cross-talk between (adjacent) representational layers is possible (e.g., Dell, 1986; Dell et al., 2013). From those classic frameworks, we turn to more recent neurobiological implementations, which, importantly also make comparisons to language perception: (c) ***dual-stream models*** which assume partially distinct processing pathways in time and space between the production and perception of words (e.g., Hickok & Poeppel, 2007; Hickok, 2012); and (d) ***parallel models*** which assume words become rapidly activated as a whole in overlapping distributed networks between production and perception (e.g., Strijkers, 2016; Strijkers & Costa, 2016; Pulvermuller, 2018). In the paragraphs below, published as a book chapter in *Language Production* (Kerr et al., 2023), each type of model will be described in more detail. After the book chapter, I will provide a shot overview about the implications of these models on if and when they predict neural alignment between interlocutors for lexico-semantic information.

Lexical access in speech production

Kerr Emilia[1]
Ivanova Bissera[1]
Strijkers Kristof[1,2]

[1]Aix-Marseille Université (AMU), Laboratoire de Parole et Langage (LPL), Aix-en-Provence, France
[2]CNRS, Provence-Corse, France

Introduction

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

1. The sequential (serial)[6] model of lexical access in word production.

The first model that we will cover is the serial model developed by Willem J.M. Levelt and colleagues (Levelt et al., 1999; see also Levelt, 1989) and later expanded into a neural

[6] The terms seriality and sequentiality do not denote the same concept: seriality (or discreteness as also often associated historically with these types of models) means that stage B can only be initiated after stage A has been completed; sequentiality, on the other hands, allows some degree of overlap between stage A and B (i.e., like 'cascading' where a representational layer lower in the hierarchy can become activated prior to selection of a representation higher in the processing hierarchy), though stage A would still initiate well before stage B. Historically, the Levelt et al. (1999) model is serial (discrete). However,

model (Indefrey and Levelt, 2004; see also Indefrey, 2011). As the name suggests, one of the main principles is the sequentiality underlying lexical access during word production (see footnote 1). It advocates progressive step-by-step processing of each linguistic level before advancing to the next level in the hierarchy (Levelt et al., 1999) and considers functional specialisation[7] of brain areas that are involved in the representation of word components (Indefrey & Levelt, 2004). According to this model, the mental lexicon concerns an independent processing layer within the speech production architecture preceded in time by conceptual processing and followed by form encoding, housed in the left mid temporal gyrus (MTG) and functionally active at a specific point in time (roughly between 150 and 250 ms of processing). Below we will outline the main principles and spatiotemporal dynamics behind this model.

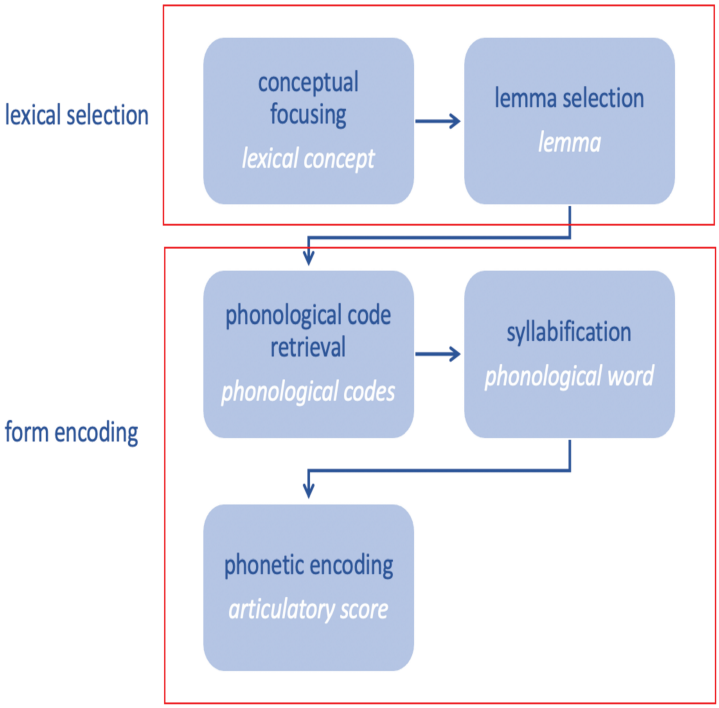


Figure 1. Schematic representation of the serial/sequential model language production model, *where lexical selection concerns the process of translating a lexical concept into a lemma representation, and form encoding concerns the retrieval of the phonological and phonetic information of the activated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

nowadays, most proponents of this model would agree with some degree of sequentially (cascading) instead of strict seriality. For the present chapter, we group the serial and sequential models together (and thus also the historic division between serial and cascaded processing), since their differences are less relevant for present purposes.

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

While most prior language production models were motivated by aphasic speech and speech error data (e.g., Dell, 1986; Garrett, 1975; Fromkin, 1971), the Levelt et al. model deviates from that tradition by being based on reaction time data of psycholinguistic experiments in healthy speakers (Glaser, 1992; Levelt, 1989; Levelt et al., 1991; Levelt & Kelter, 1982; Meyer, 1992; Schriefers et al., 1990). Figure 1 represents a schematic overview of the word production system as proposed by Levelt and colleagues (1999; see also Levelt, 2001). The first step is lexical selection, that is, when a speaker selects an item from their mental lexicon. The way this works is that a speaker activates an intended concept s/he wants to utter (i.e., conceptual focusing), which in turn will activate word candidates in the mental lexicon until an appropriate lemma, i.e., a morpho-syntactic representation of a lexical concept, is being singled out. During lexical selection there is thought to be competition between different lexical candidates due to spreading activation: That is, when activating an intended concept (e.g., CAT) related concepts (e.g., DOG) become (partially) activated as well because of their association with the target concept (spreading activation between strongly interconnected representations). This in turn will activate different lemma representations (e.g., cat and dog) which enter into a competitive process for selection, since after the lemma stage the system is thought to continue processing with a single representation, namely, the target lemma intended for articulation. The empirical input supporting these notions that lexical access is a competitive process where only a single lexical representation will be selected for further phonological processing stems from semantic interference experiments, namely the well-known picture-word interference (PWI) paradigm (Glaser & Düngelhoff, 1984; Levelt et al., 1991; Roelofs, 1992; Schriefers, Meyer & Levelt, 1990; Damian & Bowers, 2003; for review see: Abdel Rahman & Melinger, 2009; Piai et al., 2012) when a distractor word is either visually superimposed on the picture to be named or auditorily presented during stimulus (picture) presentation. The critical manipulation is that distractor words can either be semantically related to the target word or semantically unrelated. What is consistently shown is that semantically related distractors increase naming latency. That is, upon presentation of related distractors subjects are normally slower in naming than when an unrelated distractor is presented. This is because the speed of selecting a target item is proportional to the cumulative activation of all the lemmas that are competing for selection. Let us explain this with an example: when selecting the lemma ‘dog’, the lemma ‘cat’ is also activated for selection as a closely related conceptual item. If participants, when asked to name a picture of a dog, are presented with the word *cat* superimposed on that picture, the lemma ‘cat’ receives additional activation resulting in a longer naming latency. If, however, an unrelated word that is not part of the competitive selection process is presented over the picture, e.g., *hat,* it affects the response latency much less. These types of semantic interference effects

are amongst the most cited evidence favouring the notion that lexical access is a competitive selection process (but see: e.g., Mahon et al., 2007; Dhooge & Hartsuiker, 2010).

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

Furthermore, PWI experiments were also of importance to obtain initial chronometric evidence on the time course of lemma selection and form encoding. In their classic PWI experiment, Schriefers and colleagues (1990; see also Levelt et al., 1991) explored both semantic interference and phonological facilitation within the same study while manipulating the stimulus-onset asynchrony (SOA). SOA (here) refers to the time between the presentation of the (auditory) distractor and the target picture to name: the distractor could either be presented prior to the target (e.g., -150 ms SOA), at the same time (0 ms SOA), or after target presentation (e.g., +150 ms SOA). The authors showed that the semantic interference effect was maximal when the distractor was presented before the picture target (-150 ms SOA), while the

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

Given the enormous impact of the serial/sequential model developed by Levelt et al. (1999), and its very precise functional and temporal predictions to go from a concept to the utterance of that concept, it offered an ideal blueprint to implement at the neural level and link its different, sequential processing stages to their respective spatial and temporal brain correlates. In order to do so, Indefrey and Levelt (2004) performed a meta-analysis of most available neuroscientific data on picture naming at that point to identify those brain regions and their respective time course of cortical activation involved in the different word production components of the Levelt et al. model (1999). First, to identify the brain regions that are reliably active during word production they looked at 82 production experiments and 26 perception experiments from neuroimaging localisation studies (see Table 2 in Indefrey & Levelt, 2004). By contrasting the patterns of brain activity found in all 108 experiments, Indefrey and Levelt propose that the set of brain areas reliably found for both the picture naming and word generation tasks can be regarded as the core set of brain areas responsible for word production. These include 11 areas in the left hemisphere (posterior inferior frontal gyrus, ventral precentral gyrus, supplementary motor area, mid and posterior superior and middle temporal gyri, posterior temporal fusiform gyrus, anterior insula, thalamus, and medial cerebellum) and four in the right hemisphere (mid superior temporal gyrus, medial and lateral cerebellum, and the supplementary motor area). Next, the authors mapped these regions to the different word production components as proposed in the Levelt et al. (1999) model (see Figure 2): mid temporal regions linked to lexical access, superior temporal regions involved in phonological encoding, the inferior frontal gyrus for speech segmentation (syllabification), motor regions associated with phonetics and articulation, and finally the superior temporal cortex for speech monitoring (for more details on the neurobiological basis of speech monitoring see: Runnqvist, 2022 in this handbook). The result of the meta-analysis and its association with the sequential model of word production thus suggest a functional specialisation of brain areas (see footnote 2): each brain area is responsible for a particular kind of linguistic computation (i.e., conceptual, lexical, phonological, articulatory) during word production. In this view the linguistic

computations underpinning speech planning are neurally discrete and the different brain areas 'communicate' hierarchically: the output of one becomes the input of another. This hierarchical conceptualisation of word processing therefore copies the key property of the Levelt et al. (1999) model, namely the serial/sequential activation of brain regions linked to a specific word production component (see also later extensions: e.g., Indefrey, 2011; and also: WEAVER++/ARC model (Roelofs, 2014; 2018)).

Importantly, Indefrey and Levelt (2004) provided their model also with temporal estimates, hereby portraying not only the spatial components of word production, but also its temporal dynamics. This is an important addition in comparison to all previous brain language models of language production which solely focused on where in the brain language production processes may come about, but not when. In order to do so, Indefrey and Levelt adopted the following strategy: First, they estimated time-windows when each word production process would happen in the course of speech planning, and next they compared those temporal estimates with the results of magnetoencephalography (MEG) studies. With regard to the first step, the authors relied on ERP data of object categorization studies (e.g., Thorpe et al., 1996; Hauk et al., 2007; Johnson & Olshausen; 2005; Schmitt, Münte & Kutas, 2000) to estimate that conceptual processing (of an image) takes about 150-200 ms. With that as starting point, they subsequently added the chronometric estimates as assessed with the PWI studies mentioned above and their computational simulations (e.g., Roelofs, 1992; 1997). The following temporal

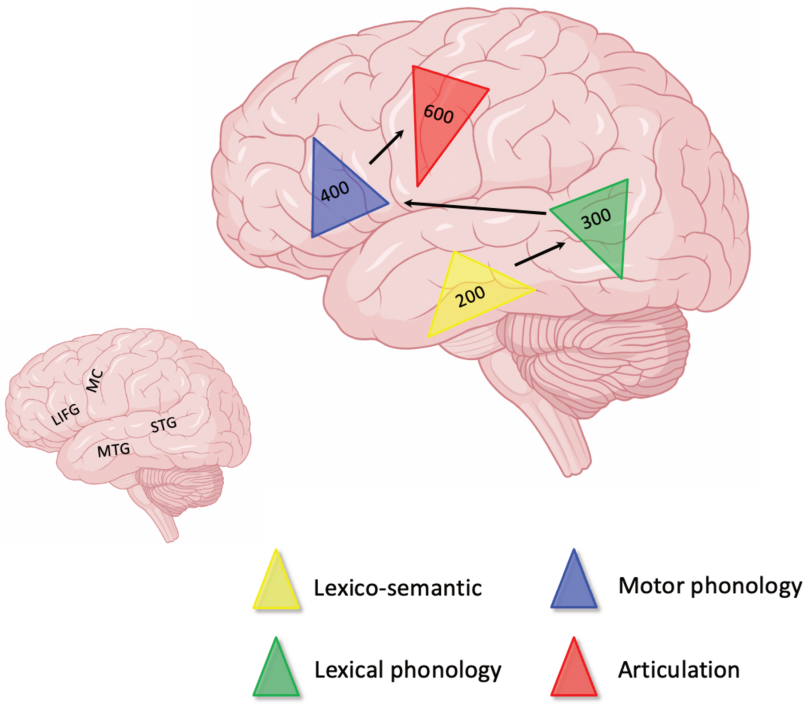


Figure 2. Schematic representation in time and space of the sequential brain language model of word production. *See text for explanations (based on Indefrey & Levelt, 2004; and adapted from Strijkers & Costa, 2016).*

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

In summary, the sequential model of lexical access in word production proposes that lexical access is a hierarchically organised process which begins with a concept and moves step-by step in a feedforward manner through linguistic levels of representation until a motor command is carried out to articulate the utterance. The core of this model postulates that the two parts of the lexical access system, i.e., lexical selection and form encoding, serving two different functions (selecting a target lexical item, and producing an articulatory score for that selected item, respectively), are segregated in time and space. That is, according to the model, information spreads in a feedforward manner, where each step is neurally realised at a well-defined time window by a distinct brain area, which specialises in one linguistic function only. This model is considered to be a detailed mapping of the neural dynamics supporting word production and to be a valuable tool and theoretical basis for producing precise predictions and hypotheses and testing them (e.g., Aristei et al., 2011; Christoffels et al., 2007; Habets et al., 2008; Koester & Schiller, 2008; Hulten et al., 2009; Hanulova et al., 2011; Dell'Acqua et al., 2010; Laganaro et al., 2009; 2012; Sahin et al., 2009; Piai et al., 2014; Fargier & Laganaro, 2017). Other researchers, however, have questioned the model's strictly sequential and highly localised properties (e.g., Strijkers & Costa, 2016; Munding, Dubarry & Alario, 2016), and alternative models of lexical access have been proposed; some of which we will discuss below. Nevertheless, and in spite that certain properties of the sequential model have led to much debate, most of the alternatives are directly built on Levelt's model and still have many aspects in common. The latter is remarkable in itself, especially when taken into account that there is

no model in language production that has been as extensively tested as the sequential model of lexical access.

2. Interactive models.

The next model we will discuss is the one proposed by Dell in 1986 (see also: Dell & O'Seaghdha, 1992; Dell et al., 1997; Dell et al., 2007; Dell et al., 2013), which introduced the notion of interactivity in lexical access (see also: e.g., Dell & Reich, 1977; Harley, 1984), and inspired many following models and much research in psycho- and neurolinguistics (e.g., Caramazza, 1997; Rapp & Goldrick, 2000; Vigliocco & Hartsuiker, 2002; Pickering & Garrod, 2004; 2013; Harley, 2008; Strijkers & Costa, 2011; Ueno et al., 2011; Strijkers, 2016; Walker & Hickok, 2016; Abdel Rahman & Melinger, 2019; Roelofs & Ferreira, 2019; Nozari & Pinet, 2020). The central characteristic of this model is that while just like sequential models it assumes linguistic representations are organized hierarchically in functionally dissociable processing layers, it allows for interactivity between those levels, meaning that information can flow bidirectionally between processing layers (see Figure 3). In this manner, and compared to the serial view on lexical access as discussed before, an interactive account of lexical access is somewhat more flexible in that there is some degree of processing overlap in both time and

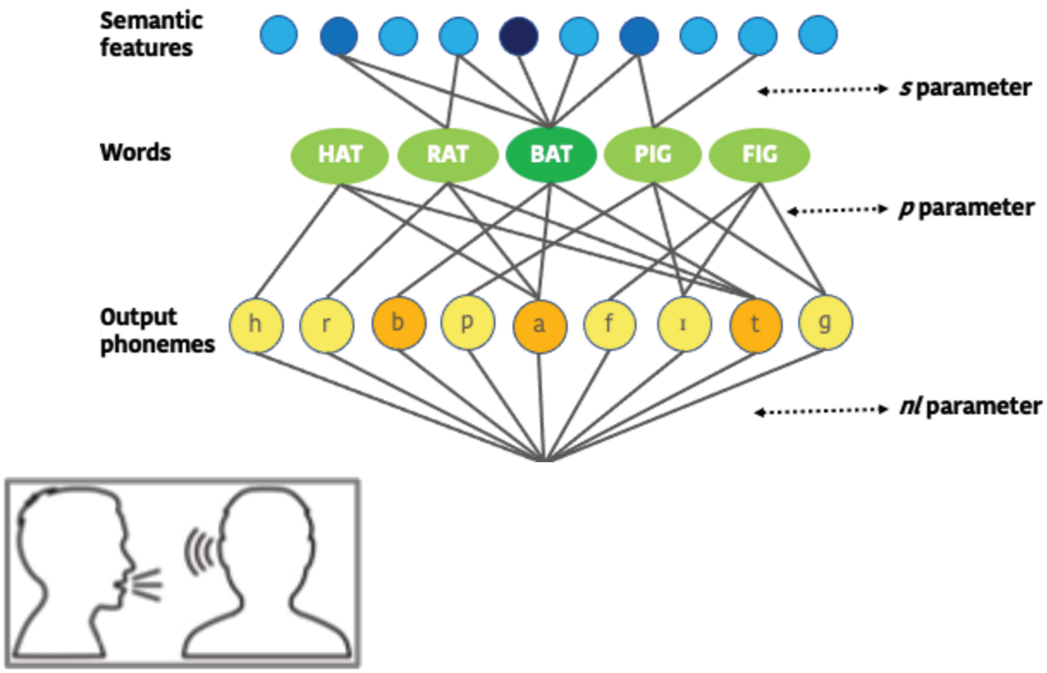


Figure 3. Schematic illustration of the two-step interactive activation model of Dell. *The two steps of the model during word production are from the semantic (features) to the lexical level (words), and from the lexical to the phonological level (output phonemes). Weights refer to the parameters s,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

space. Below we will first detail the computational principles of Dell's interactive lexical model further and subsequently give an overview of its proposed neural implementation.

Interactive lexical access in the Dell-model is, as in the serial model (Levelt et al., 1999; but see Caramazza, 1997), regarded as a two-step process where activation flows through the lexical system in the following manner: The first step is initialised by a jolt of activation at the conceptual level ("Semantic features" in Fig. 3), continues as spreading activation through the network and activates a number of relevant units at the lexico-syntactic level ("Words"). Units at this level are called lemmas and are abstract symbols that unify the semantic-syntactic representation of a word. The first step is finalised when the lemma with the highest level of activation is selected, regulated via an 'insertion' rule that checks whether the lemma is semantically and syntactically suitable (e.g. "swim" if a verb is required and "swimming" if a noun), and activation of other lemmas is inhibited. The second step of word production begins with a jolt of activation spreading from the lemma to the phonological level ("Output phonemes"), activating the phonological units corresponding to the selected lemma representation. At this stage, beyond for example differences in compositionality (e.g., Roelofs, 1997b), the previously discussed serial model and Dell's interactive activation model are quite similar in their dynamics of lexical access. Crucially, however, interactive models propose that information travels bi-directionally in the lexical network. That is, while in the serial model (Levelt et al., 1999) lexical access is entirely feedforward, in Dell's model (1986) activation can flow back between the representational layers (see Figure 3). This means that even if linguistic levels are globally modular and initially become activated in a sequential manner, they are locally interactive, displaying temporal overlap in the activation of lexical representations: when level *x* is most active, there will be some activity too in levels *x+1* and *x-1* (Dell & O'Seaghdha, 1992). For example, spreading activation from the concept BAT will activate the lemmas "bat" and "pig" because they are categorically related. Then, spreading activation from the lemma "bat" will activate the phonological units "b", "a" and "t", and those phonological units will in their turn send activation back to those lemmas associated with the phonological features, hereby 'boosting' the activation of the target lemma "bat" and reducing the activation linked to the semantically related word "pig" (see Figure 3). Or put differently, within such interactive framework, the intended lexical entry for speech receives two sources of activation: the bottom-up driven feedforward activation from the concept a speaker wishes to utter and the feedback driven activation from the phonological level of processing allowing to "check" that the activated sounds indeed correspond to the word we wish to convey[8].

[8] While the Dell and colleagues' model does not include inhibitory connections, other interactive models do, suggesting this a biologically plausible mechanism for limiting

Importantly, beyond offering a mechanism on how our lexical system can avoid mis-selecting words, the interaction between levels allows for explaining common speech errors that people make. In fact, the original model of Dell (1986) was constructed exactly for that purpose. Indeed, while overall, we make surprisingly few speech errors when speaking, when we do make an error it seldomly is random, but instead follows linguistic constraints (e.g., Fromkin, 1973; Garret, 1975; Dell, 1986; Rapp & Goldrick, 2000). In this manner, speech errors have been an important source of information to understand the architecture underpinning language production (for more recent overviews: e.g., Goldrick, 2011; Dell et al., 2014; Runnqvist, 2022 this handbook). At the level of lexical access, three types of errors have been particularly informative: semantic, phonological and mixed errors. Semantic errors refer to saying for example “pig” instead of the intended “bat” and are explained in the model because of the spreading activation from related concepts (i.e., other animals) (see Figure 3). Phonological errors refer to saying for example “hat” instead of “bat” and come about in the second phase of activation in the model, namely when the selected lexical item activates overlapping phonological units. Finally, and most importantly here, the mixed error effect refers to errors which are both semantically and phonologically related, like saying “rat” instead of “bat”. Furthermore, these mixed errors occur more frequently than predicted in light of semantic or phonological errors in isolation (e.g., Dell & Reich, 1981; Harley, 1984; Martin et al., 1997; Goldrick & Rapp, 2002; Ferreira & Griffin, 2003); an observation which is assumed to be a consequence of interactivity. That is, keeping with the above example, it is more likely that the target "bat" will be substituted by "rat" than by "pig", because “rat” will receive activation from two sources, namely the spreading activation (from the target concept “bat”) from the semantic layer and the feedback activation (interactivity) from the phonological layer (because it has two phonological units that overlap with the phonological units of “bat”), while “pig” will only receive activation from one source, namely the spreading activation from the semantic layer (see Figure 3). In other words, due to the interactivity, mixed errors are readily explained in Dell’s model (1986; Dell et al., 2013). In contrast, in discrete and serial models like that of Levelt et al. (1999), such mixed error effects are particularly difficult to explain, since the holistic word forms do not affect the lexical representations, and thus mixed errors should not happen.

Another important behavioral observation that has typically been cited to support interactivity concerns the lexical bias effect (Baars et al., 1975; Dell & Reich, 1981; Humphreys, 2002; Hartsuiker, et al., 2005; 2006; Nooteboom, 2005; Nozari & Dell, 2009; Runnqvist et al., 2016; 2021). The lexical bias effect is the tendency of phonological errors to

interaction and cascading within the system. For more detail look at Harley (1993) and Schade and Berg (1992).

also be real words, for example people are more likely to substitute "bat" with "hat" than with "lat". Interactive models explain this because phonological units follow phonological rules in a language, they will favour the grouping together of legal over illegal phonological activations. Since in an interactive model activated phonological units feedback to the lexical system, it logically follows that a speech error will more likely be a word than a non-word for the simple fact that words are represented in our mental lexicon, while non-words are not. Hence, when making a phonological speech error like a phoneme substitution (e.g., changing the first phoneme of a word), within an interactive system it is predicted that the substitution will more likely result in an actual word ("hat" instead of "lat"), because it receives activation from the lexical layer, while a nonword does not. Similarly as for the mixed error effect, in a serial model (Levelt et al., 1999), this is not necessarily predicted (but see: e.g., Roelofs, 2004).

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

In terms of neural implementation, the notion of interactivity can fit different neuroanatomical architectures (e.g., Indefrey, 2011; Strijkers & Costa, 2011; Hickok, 2012). For example, it could be consistent with the previously described Indefrey and Levelt (2004) model if some more temporal and spatial flexibility between the different brain regions (linked to distinct functional word components) is allowed. However, while the Indefrey and Levelt (2004) model suggest a single feedforward processing pathway, going from mid temporal via superior temporal towards frontal brain regions, Dell and colleagues have linked their specific interactive model to a dual-route architecture (Dell et al., 2013; see also Nozari et al., 2010; Ueno et al., 2011), hereby integrating established neuroanatomical ideas from the perception literature (e.g., Hickok & Poeppel, 2000; 2004; 2007; Scott & Johnsrude, 2003; Rauschecker & Scott, 2009). In this manner, Dell et al. (2013) suggest that lexical access in an interactive model is realized by two partially distinct brain networks, a more ventral stream and a more dorsal stream (roughly corresponding to the red and blue patches in Figure 4). Previous research implicates the dorsal stream in phonological processing as there is evidence that phonological form retrieval takes place in the pSTG (e.g. de Zubicaray et al., 2002; Graves et al., 2008;

Edwards et al., 2010), phonological short-term memory is processed by temporo-parietal and inferior parietal regions (e.g. Buschsbaum et al., 2011), and phonologically-related errors stem from the dorsal pathway (e.g. Cloutman et al., 2009; Duffau et al., 2008; Schwartz et al., 2012). The ventral stream, on the contrary, has been implicated in more lexical-semantic processing as evidence suggests that lemma processing takes place in the MTG and pITG (e.g. Damasio et al., 1996; Graves et al., 2007; de Zubicaray & McMahon, 2009; de Zubicaray et al., 2015; Ries et al. 2017), objects and events are represented in the ATL and AG (e.g. Binder & Desai, 2011), and damage to the ventral stream, including the MTG, ITG, ATL, AG and IFG results in semantically-based word retrieval difficulties (e.g. Antonucci et al., 2008; DeLeon et al., 2007; Schnur et al., 2009).

Dell and colleagues (2013) further confirmed and extended this two-step neural architecture to brain damaged aphasic speakers via voxel-based lesion parameter mapping. The study used the interactive activation computational model to simulate the individual error pattern of 103 aphasic patients in three different tasks: word production, word repetition, and non-word repetition. Parameters corresponding to the connections between linguistic levels were adjusted so that the model can account for the variability in errors for all three tasks. The lesion status of a voxel was then used to predict the three critical parameters: *s* representing the ability to map from the conceptual to the lexical level, *p* the mapping from the lexical to the phonological level, and *nl* the mapping from auditory input to the phonological level (see Figures 3 and 4). While the *p* and *nl* parameters shared a significant proportion of their brain maps, for the *s* and *p* parameters overlap was much less pronounced (see Figure 4), suggesting that the brain areas responsible for realizing the lexical and phonological levels of word production are at least in part neurally distinct.

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

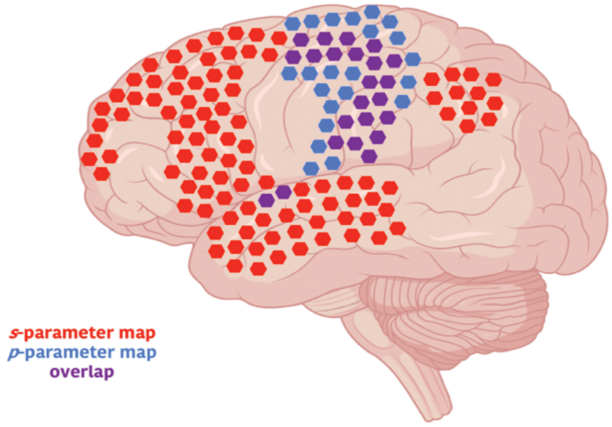


Figure 4. *The maps of two of the parameters from the interactive computational model suggest that semantic and lexical processing on the one hand and phonological and articulatory processing on the other have only a few brain areas in common (based on Dell et al., 2013).*

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

To summarize, lexical access in an interactive model (e.g., Dell, 1986; Dell et al., 2013) has some very similar assumptions to lexical access in a serial/sequential model (e.g., Levelt et al., 1999; Indefrey & Levelt, 2004), because it considers lexical access to involve sequential steps of processing which work with different kinds of representations, stored in different brain areas. Specifically, in this view, words in the brain have lexical and phonological representations, which are independent of each other and activation flows from the concept through the lemma and onto the phonological form in functionally distinct time steps. In contrast to the serial/sequential model of lexical access, however, these lexical and phonological layers are "globally modular, but locally interactive" (cf. Dell & O'Seaghdha, 1992), in that

they are much more distributed in our brain and can have a functional and temporal influence upon each other through feedback interactivity.

3. Dual-stream Feedback Models

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

The first model that we are going to discuss, the hierarchical state feedback control model (HSFC), has been developed by Hickok (2012; see also Hickok, 2014; Walker & Hickok, 2016) and is part of a larger integrated theory on speech processing, the dual-route model (see Figure 5; Hickok & Poeppel, 2000; 2004; 2007). The dual-route architecture, which we already mentioned in the previous section with regard to a specific neural implementation of Gary Dell's (1986) interactive activation model of lexical access, is borrowed from research in the visual domain where this processing mechanism, namely, the division of labour between the ventral and dorsal streams, has been well established and demonstrated empirically (e.g., Milner & Goodale, 1993; Ungerleider & Mishkin, 1982). The dual-route model of speech processing proposes two pathways, or streams, that underlie the neuroanatomy of language processing: the ventral stream, concentrated in the superior and middle parts of the temporal lobe, is responsible for the comprehension of speech, and the dorsal stream, involving structures in the posterior

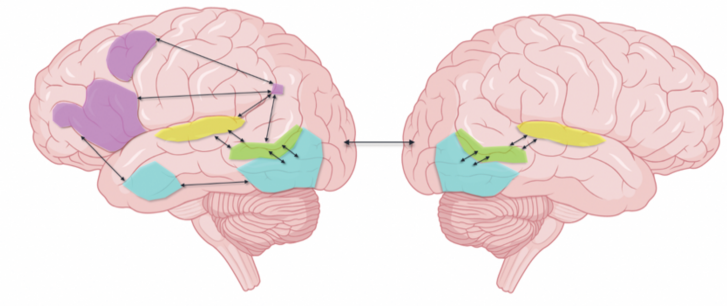


Figure 5. The dual-stream model of speech processing. *Bilateral auditory regions in the dorsal STG (yellow) and STS (green) are engaged in the early stages of speech processing, which later diverges into two streams: the ventral stream responsible for speech comprehension (blue) and a left dominant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

planum temporale and posterior frontal lobe, is engaged in speech production through sensory-motor integration (more on this below) (see Figure 5). While originally the dual-route architecture was developed to explain language comprehension (Hickok & Poeppel, 2000; 2004; 2007), here we will focus on the recent extension to language production (Hickok, 2012; 2014). This model serves as a neat bridge with the prior two sections in this chapter, since it borrows many properties of the sequential and interactive models discussed before but embeds it within the neuroanatomical principle of dual-stream processing and adds the concepts of feedback control and predictive processing (see for similar psycholinguistic feedback/predictive models of language production e.g., Pickering & Garrod, 2013; Dell & Chang, 2014).

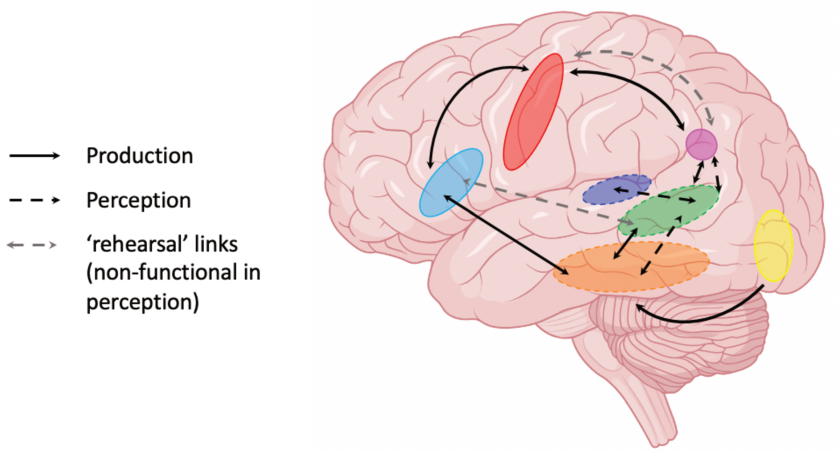


Figure 6. Simplified schematic representation of Hickok's dual-stream HSFC model, *where lexical representations (lemmas) are triggered (for example from a visual information -yellow area- we wish to utter) in the (posterior) MTG (orange) and activate two streams of processing: (1) towards inferior frontal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

The HSFC model attempts to bridge two traditions of speech processing research – psycholinguistics on the one hand (e.g., Dell, 1986; Levelt et al., 1999), and motor control theories on the other (e.g., Gracco & Lofqvist, 1994; Guenther, 2006; Golfinopoulos et al., 2010; Tourville & Guenther, 2011; Perkell, 2012). Drawing from the previous psycholinguistic theories (e.g., Dell, 1986; Levelt et al., 1999), the model agrees that the main stages of word production comprise conceptual and lexical selection, phonological encoding and articulatory preparation. These stages are considered to be implemented in a hierarchical manner where lexico-semantic processes give rise to phonological processes (Hickok, 2012; 2014; 2019). Neurally, this hierarchy is organized in the following regions: lexical selection occurs in the middle temporal regions and phonological processing takes place in the posterior STG (word forms), inferior parietal and inferior frontal and premotor cortex (motor phonology) (see Figure 6).

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

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

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

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

Other data, from healthy speakers, that finds an elegant explanation in the HSFC model concerns neuroimaging data related to (1) auditory suppression and to (2) speech monitoring. Concerning (1), auditory suppression refers to the phenomenon that the auditory cortex's response to one's own speech is reduced compared to the speech of others (e.g., Houde et al., 2002). The fact that this phenomenon seems to occur ultra-rapidly (already within 100 ms of processing; e.g., Heinks-Maldonado et al., 2006) fits well with the notion of forward prediction as implemented in Hickok's HSFC model: That is, because one can predict one's own speech better than that of others, the neural response in the auditory cortex is attenuated since there is a good match between the action (motor) and its consequence (auditory) and no further reprocessing is necessary. Concerning (2), and recent neuroimaging finding that goes well with the predictions of the HSFC, are studies where the speech monitoring system is investigated by using a slip task (e.g., Runnqvist et al., 2016; 2021). In a slip-task (e.g., Motley et al., 1982) researchers attempt to experimentally induce speech errors by priming people to make phoneme substitutions (e.g., barn door -> darn boor). Runnqvist and colleagues investigate this

phenomenon with TMS and fMRI for correct trials (that is, trials where the participant didn't make the error -the phoneme substitution- but the likelihood of making such error was high; put differently, conditions that tax the speech monitoring system). The authors observed the involvement of the right cerebellum and superior temporal brain regions (amongst a relevant set of other brain regions: e.g., Runnqvist et al., 2016; 2021). Given the cerebellum's role in forward modelling (e.g., Blakemore et al., 2001; Ito, 2008), the link of right cerebellar activity and left superior temporal regions to monitor phoneme substitutions fits neatly with the predictions of the HSFC (Hickok, 2012; 2014).

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

4. The Parallel Assembly Model

The last model we will describe proposes the notion of fast parallel processing where lexico-semantic and phonological-articulatory stages occur in the same temporal windows and simultaneously integrate distributed cortical activations. This model differs from previous models in that it does not assume a hierarchical structure underpinning the different linguistic components making up a word, but rather the full integration of all word components into a single functional whole; the word as the neural Gestalt of language in the brain. Another interesting difference this model has compared to the previous ones is that its basis is inspired by system neuroscience theory rather than psycholinguistics and neuropsychology. That is, while the previous models were mainly guided by reaction time data (Levelt et al., 1999), speech error patterns (Dell, 1986; Dell et al., 2013) or patient data (Hickok, 2012; 2014), the parallel assembly model took the reverse approach and is mainly driven by neurophysiological processing principles. More concretely, the driving principle behind the model is Hebbian-based learning, which in broad terms states that 'what fires together wires together' (Hebb,

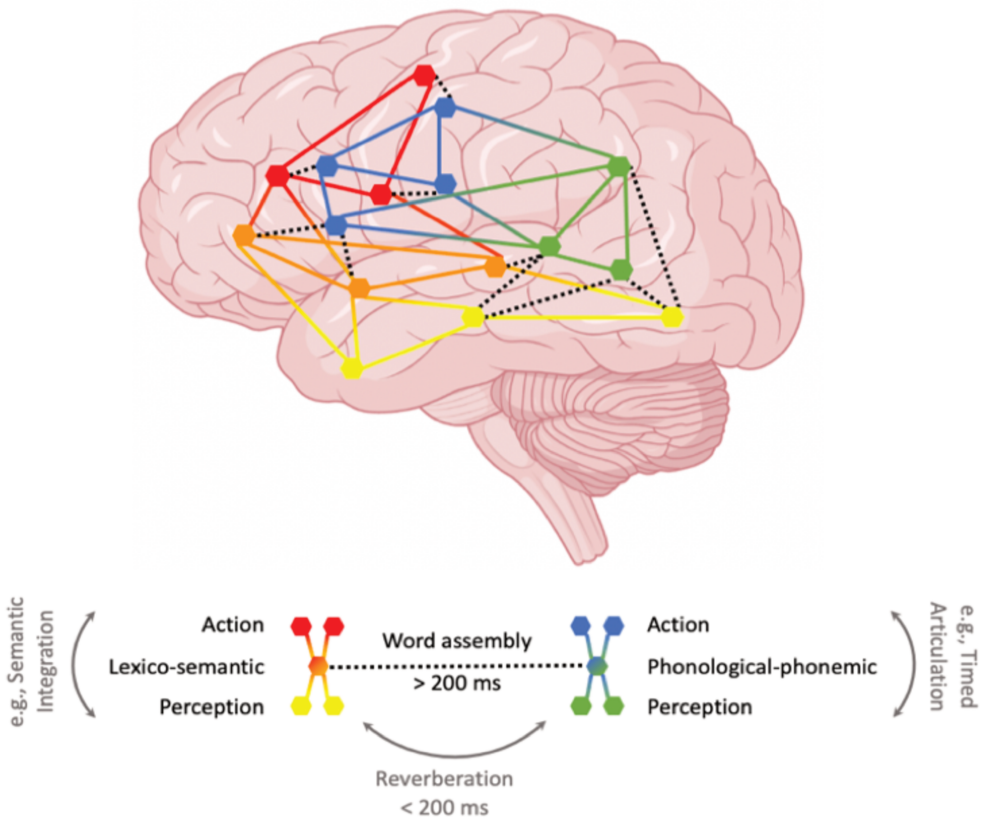


Figure 7. A schematic visualisation of the parallel assembly model. *A widely distributed lexico-semantic network embedded in action (red) – perception (yellow) circuits and a widely distributed phonological-phonemic network embedded action (blue) – perception (green) circuits form a word assembly which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

1949), and which Friedemann Pulvermüller used to develop a model of how our brain would represent words (Pulvermüller, 1999; and later: Pulvermüller, 2001; 2005; 2018; Pulvermüller & Fadiga, 2010). Recently this Hebbian-based assembly model of language was then adopted to the issue of lexical access in speech production (Strijkers, 2016; Strijkers & Costa, 2016).

In short, Hebbian-based learning (or assembly coding) means that those neural representations that are active at the same time will bind together into a single functional unit (e.g., Hebb, 1949; Braitenberg, 1978; Singer & Gray, 1995; Fries, 2005; Buzsaki, 2010; Singer, 2013). Or, put differently, if for a given event X there is temporal correlation (or coherence) between neural population A and neural population B, and this temporal correlation occurs often, then for that specific event X those two neural populations A and B will bind together in a novel, overarching neural assembly C that can reflect event X as a whole. Translating Hebb's postulate to words, the idea is that since the meaning and sounds of a given word always co-occur (e.g., when speaking about a 'ball', the semantic features and sound features of 'ball' will always be active at the same time), through Hebbian-like learning they form a single functional unit capable to reflect a word in its totality: a word assembly (e.g., Pulvermüller, 1999; 2005; 2018; Pulvermüller & Fadiga, 2010). Note that the starting point is not so different from the previously discussed models, namely a system dedicated to meaning-related processes and to sound-related processes that are (partially) separated and independent from one another. The difference emerges during development where, according to Pulvermüller's Hebbian model, the key to word learning lies in binding together meaning and sounds that form a coherent word. Viewing words in this way results in a substantially different spatiotemporal dynamic of lexical access when adopted to word production compared to the previously discussed serial and interactive theories (see Figure 7) (Strijkers, 2016; Strijkers & Costa, 2016): (a) the time course of lexical access would be parallel instead of sequential; (b) the spatial recruitment would involve distributed networks both for lexico-semantic and phonological knowledge, instead of localized brain regions or processing streams for the lexico-semantic (lemmas) representations on the one hand and phonological representations on the other (see Figure 7).

First, we will look at the temporal dynamics and the main differences between the parallel assembly and sequential or interactive models. Traditional language production theories suggest that the firing of the lower-level cells (input) spreads to the firing of the higher-level cells (output), thus constructing a hierarchy of activations that occur in a sequential manner, i.e., segregated in time with differences between lexico-semantic and phonological activation often estimated in the range of 100 ms (e.g., Dell & O'Sheagdha, 1992; Indefrey & Levelt, 2004; Indefrey, 2011). Note, as mentioned above for models incorporating interactive (and the same holds for cascading) properties (e.g., Hickok, 2012; 2014; Dell et al., 2013), that some overlapping (and in that sense "parallel") activation is predicted in such models. However,

and crucially, this overlapping activation occurs at a later point in time after initial sequential activation (for example, in an interactive model, first there is sequential activation of the hierarchical layers before they can start interacting with each other and overlap occurs). This is different from the parallel dynamics envisioned by Hebbian-based assembly models where the parallel activation of the meaning and sounds of a word occurs immediately during the first pass activation. Indeed, given that in a parallel assembly model words are reflected as integrated functional units, neuronal firing triggers the near-synchronous ignition of a word as a whole, because lexico-semantics and phonology are bound together in a single parallel distributed processing network.

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

Nevertheless, while such results pose serious problems for purely serial processing theories, it remains debated whether they fit word production models with a sequential time-course. That is, the type effects presented above where a lexico-semantic and phonological variable both emerge within (roughly) 250 ms of processing, may still reflect a sequential activation over hierarchically distinct processing layers (lexico-semanticàphonological), but simply where activation spreads from one layer onto the other in a faster manner than previously assumed, namely in 10s of ms instead of 100s of ms (e.g., Mahon & Navarrete, 2016; Strijkers

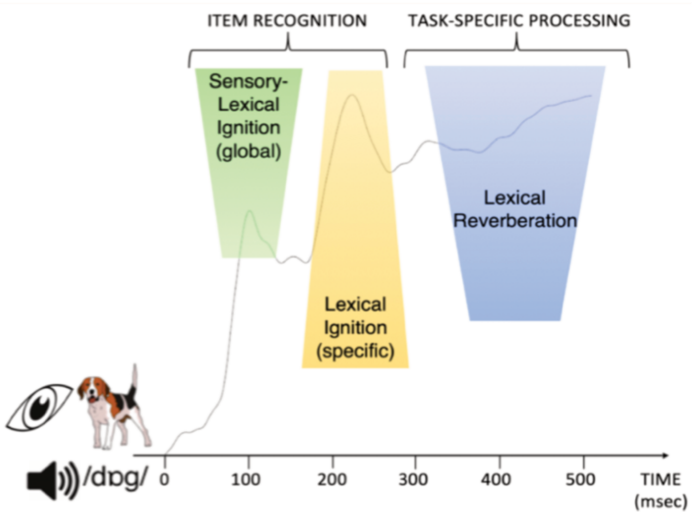


Figure 8. Schematic overview of word production processing according to the parallel assembly model. *First, up to about 300 ms lexical ignition takes place where the word assembly becomes active simultaneously. Second, task- and stimulus-specific reverberations occur after 300 ms of processing (taken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

et al., 2017). To assess this, Fairs et al. (2021) approached the issue differently and compared the time-course of lexico-semantic (targeted through lexical frequency) and phonological word processing (targeted through phonotactic frequency) during both production and perception. The logic here was the following: Given the hierarchical nature of sequential models, they predict the reverse time-course between production and perception. In production the lexical frequency effect, linked to lexico-semantic word knowledge, should emerge before the phonotactic frequency effect, linked to phonological word knowledge, while in perception the reverse should happen (regardless of whether this happens fast in 10s of ms or slow in 100s of ms). In contrast, according to a parallel assembly model, a word ignites as a whole both in production and perception and thus for both language modalities the lexical and phonotactic frequency effects should manifest simultaneously. The latter is indeed the result found by Fairs et al. (2021), where for both production and perception lexical frequency and phonotactic frequency variables modulated the ERP effects in the early time windows (74-145 ms and 186-287 ms). Differences between the language modalities occurred only in a later time- window (316-369 ms after stimulus onset).

The results by Fairs et al. (2021) are particularly intriguing because they demonstrate the parallel effects (early on) and the sequential effects (later on) within the same study. Indeed, their data showed that after 300 ms of processing lexical and phonotactic frequency only modulated word production not perception, hereby showing modality and task-specific reverberations at later stages of processing. In this manner, a parallel assembly model explains the temporal dynamics underpinning lexical access in the following manner (see Figure 8): After early initial sensorial activation in response to the input, (1) the first linguistic activation

emerges roughly between 75-150 ms after onset denoting the start of "globally" activating potential words fitting the initial sensory analyses (this ultra-rapid lexical access may be achieved via prediction); (2) next, this initial "global word space" activation is further refined and delineated within roughly 150-250 ms leading to the ignition and recognition of the specific lexical item associated with the input; (3) finally, roughly after 300 ms, slower, sequential-like reverberation upon the activated word assembly takes effect in order to embed the recognised target word into the proper linguistic and task context to be able to perform the intended behaviour, and modality-specific processing effects become visible. Note the last point is key, because it highlights that the parallel assembly model of language production does not invalidate all the studies reporting sequential activation time- courses observed in word production experiments, but rather offers an alternative interpretation to the same data. In particular, rather than reflecting the initial activation of word components in a sequential manner, the observed sequential effects reflect sensitivity to the later task- and language-specific processes upon a parallelly retrieved word representation (see e.g., Strijkers, 2016; Strijkers et al., 2017; Fairs et al., 2021). In a similar vein, dissociations between lexico-semantics and phonology in patients or as observed for speech errors (e.g., Brehm & Goldrick, 2016), would in this model be due to problems during reverberatory processing rather than initial word retrieval (for more details see: e.g., Strijkers, 2016; Strijkers & Costa, 2016b). In sum, this temporal dynamic of lexical access is markedly different from those described for the models in the previous sections in that there is not a single time frame (serial or interactive) where a given linguistic component is active, but (at the least) two functionally distinct time frames: a parallel linked to word activation (i.e., ignition) and a sequential one linked to task-specific operations upon that ignited word assembly (i.e., reverberation) (see Figures 7 and 8).

Another important difference between prior models and the parallel assembly model lies in the distinct language-to-brain mapping (e.g., Pulvermüller & Fadiga, 2010; Strijkers, 2016; Strijkers & Costa, 2016; Pulvermüller, 1999; 2018). While in the previous models that we discussed the link between brain localization and linguistic function was mainly driven by a one-to-one (for example, lexical representations in MTG, phonological representations in STG, etc.; e.g., Indefrey & Levelt, 2004), or one-to-many relationship (for example, lexical processing in a ventral stream, phonological processing in a dorsal stream; e.g., Hickok, 2012; 2014; Dell et al., 2013), in the parallel assembly model mapping is many-to-many. This means that the neural organization of linguistic components during word production (lexico- semantics and phonology) is reflected in distributed frontotemporal and parietal circuits that integrate sensorimotor networks (see Figure 7). The reason why different word components will map onto different overlapping and distributed neural networks in parallel assembly models is similar as to why time-course of lexical access manifests rapidly and simultaneously in these

models, namely because of the Hebbian neural binding between consistently co-activated input (perception) and output (production) (see Figure 7). This is the reason why an assembly model can explain 'embodied-like' responses during language perception (for example, the word 'kick' activating the leg-region in the motor cortex) (e.g., Watkins et al., 2003; Hauk et al., 2004; Pulvermuller et al., 2006; Carota et al., 2012; Dreyer et al., 2015), and why during object naming Strijkers and colleagues (2017) observed parallel distributed frontotemporal networks activated in response to both lexical and phonological properties (see also: e.g., Munding et al., 2016; Ries et al., 2017), with (just as in perception) feature-specific topographies in the sensorimotor cortex (e.g., bilabial phonemes linked to lip motor cortex and alveolar phonemes linked to tongue motor cortex; see also Fairs et al., 2021).

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

Conclusion

In this chapter we have outlined the main theories on lexical access during language production. Lexical access as a process of binding conceptual semantic information with its phonological form has been analyzed from different perspectives: as a serial/sequential hierarchical system (e.g., Levelt, 1999; Levelt & Indefrey, 2004; Indefrey, 2011), as an interactive system that allows more communication between its levels (e.g., Dell, 1986; Hickok, 2012; Dell et al., 2013), and as a parallel non-hierarchical system (e.g., Pulvermuller, 1999; 2018; Strijkers, 2016; Strijkers & Costa, 2016). The development of these theories spans a few decades of research, and the topic still remains a vital area of research with many exciting issues to resolve. Traditional serial and interactive accounts on word production have played an essential role in advancing the field by providing the first clear models where the main components of word production were identified. Recently, the advancements of research methods and imaging techniques have allowed to investigate word production in more fine-

grained detail and ecologically valid conditions. This, in turn, has triggered some reconsideration of past ideas and postulates that have been considered undeniable, especially the hierarchical and serial nature of speech production processing.

2.1 Summary of the book chapter in relation to the prerequisites for neural alignment

I have summarized the major theories on lexical access during language production and their spatial and temporal link to perception in the final two sections of the chapter. The process of connecting conceptual and semantic information with its grammatical and phonological forms has been examined from several angles: as a sequential hierarchical system (Levelt, 1999; Levelt & Indefrey, 2004; Indefrey, 2011), as an interactive system allowing more communication between its levels (Dell, 1986; Dell et al., 2013), within a dual-stream architecture proposing two (partially) distinct processing pathways between production and perception (e.g., Hickok & Poeppel, 2007; Hickok, 2012; 2014), and as a parallel non-hierarchical system where the retrieval of words is similar between production and perception (Pulvermuller, 1999; 2018; Strijkers, 2016; Strijkers & Costa, 2016).

Here I would like to speculate how these different approaches to word retrieval translate to the potential emergence of neural alignment during verbal communication. As mentioned in the Introduction of this chapter, a prerequisite for neural alignment to occur is that the representations under investigation have substantial spatiotemporal overlap across the language modalities (i.e., shared representations) (e.g., Hasson et al., 2012; Schoot et al., 2016), or at least that corresponding representations in production and perception are active at the same

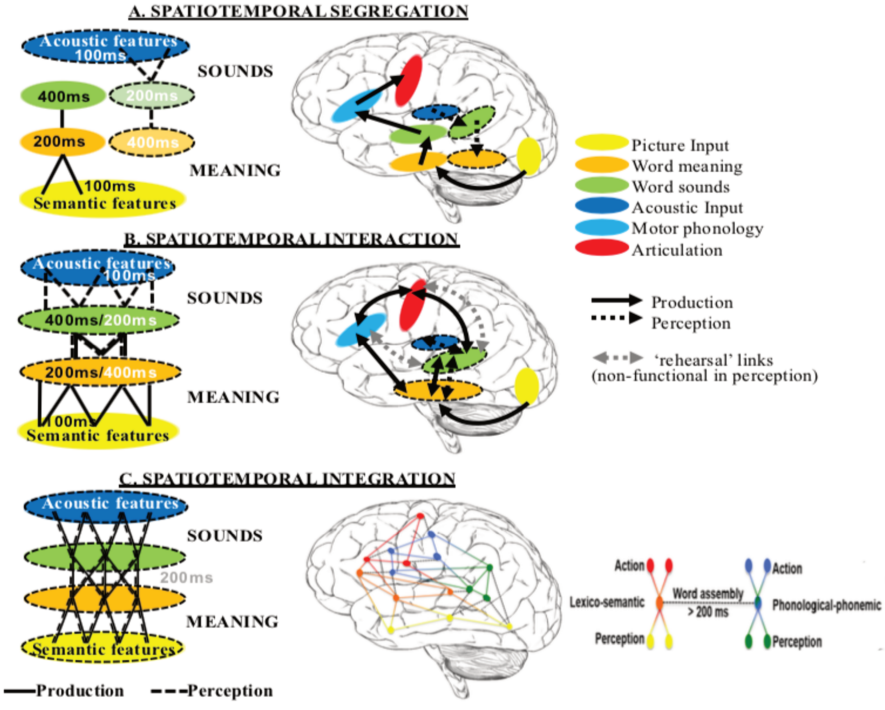


Figure 9. Schematic visualization of spatiotemporal dynamics of lexical access for three different classes of models. *(A) Spatiotemporal segregation where the production and perception of words rely on different brain regions and temporal dynamics. (B) Spatiotemporal interaction where there is a certain degree of spatial overlap but different temporal dynamics between the production and perception of words. (C) Spatiotemporal integration where the production and perception of words rely on the same spatiotemporal dynamics.*

time even when processing in a single modality (i.e., parity of activation of "copied" representations between the modalities) (e.g., Pickering & Garrod, 2004; Menenti et al., 2011; Pickering & Garrod, 2014). While for certain word components, in particular phonology, models strongly differ in assuming representational overlap or not across the modalities, it is clear from the overview given in this chapter that most models of lexical access do agree that for the level of investigation of this thesis, lexico-semantics, representations are shared (or at least strongly overlap) between production and perception. A fully segregated model between production and perception (see Figure 9 A), the dominant view for over a century (e.g., Geschwind, 1979), which would predict the absence of neural alignment at the word level, is no longer supported, and most of the models discussed in the chapter are compatible with an infrastructure that may sustain neural alignment at the word level, at least for lexico-semantic information.

Assuming overlap of lexico-semantic representations is an important requirement for neural alignment, but not (necessarily) a sufficient requirement. Representations must also be activated with a similar time-course across the modalities to neurally align, and here differences between models remain marked. The serial and interactive models, including dual-route models, discussed in the chapter assume a distinct time-course between the production and perception of words (e.g., Dell, 1986; Levelt et al., 1999; Indefrey & Levelt, 2004; Hickok & Poeppel, 2007; Indefrey, 2011; Hickok, 2012; Dell et al., 2013). That is given their hierarchical structure of word component activation they all propose (and notwithstanding important differences between those models as well), the lexico-semantic system is active earlier in production (around 200 ms) than in perception (around 400 ms) (see Figure 9 B). Or put differently, from the perspective of the dyad, even though lexico-semantic representations are shared between production and perception, they will not be active at the same time in the speaker as in the listener. This difference in temporal dynamics is most marked for the serial models described (e.g., Levelt et al., 1999; Indefrey & Levelt, 2004; Indefrey, 2011), since they only have feedforward connections. Interactive models can have some temporal overlap (between adjacent layers), but only through feedback projections (e.g., Dell, 1986; Hickok, 2012; Dell et al., 2013). That is, during initial processing there is no overlap, but near the end of processing, when interactivity between representational layers kicks in, lexico-semantics may be active in perception and production at the same time. Does this mean that for these types of models it is unlikely to observe neural alignment at the word level, at best near the end of processing, but not during retrieval? When testing words in isolation (see Chapter 4), the lack of strong temporal overlap indeed poses a problem to see strong neural alignment, but when testing words in context (see Chapter 5), the lack of temporal overlap may be compensated via predictive processing. That is, if both listeners and speakers pre-activate the same potential

word candidates in function of prior context, then both conditions to observe neural alignment would be fulfilled: the same (or similar) spatial representations activated at the same point in time.

The story is different for the last class of models we discussed in the chapter, namely parallel activation models. Given that these models dispense with the strict hierarchical format (at least at the word level) and assume words concern a single, integrated mental representation that becomes activated in parallel, the time-course of lexical access will be the same regardless of the language modality (e.g., Pulvermuller, 1999; 2018; Strijkers, 2016; Strijkers & Costa, 2016) (see Figure 9 C). This means that these types of models can predict neural alignment at the word level between speakers and listeners even in the absence of a prior context given words in parallel activation models share space and time between production and perception. Of course, just as for the other models, predictive contexts would also in these models likely enhance the timing and strength of neural alignment, but it may not be a necessary requirement compared to its hierarchical counterparts.

Chapter 3. Predicting meaning in the dyad

(under review)

Kerr Emilia[1]
Morillon Benjamin[2]
Strijkers Kristof[1,3]

---

[1]Aix-Marseille Université (AMU), Laboratoire de Parole et Langage (LPL), Aix-en-Provence, France
[2] Aix-Marseille Université (AMU) & Inserm, Institut de Neurosciences des Systemes (INS), Marseille, France
[3]CNRS, Provence-Corse, France

Abstract

This study investigated the role of meaning prediction in dyadic interactions. Participants engaged in a semantic association game where predictability of upcoming semantic categories was manipulated. Results showed that predictability sped up speech production in both interlocutors. Critically, our study reveals that meaning prediction not only aids language comprehension and subsequent speech production for the interlocutor completing a predictable sentence but facilitates the verbal behavior from the conversational partner even more. This result demonstrates an incremental dyadic predictability effect in interlocutors during language interaction; a finding which offers empirical support for a key property of theories on joint action and interactive dialogue models, namely that successful social interaction and language use involve predicting and representing linguistic representations of 'both' partners in a dyad.

Keywords: Psycholinguistics, prediction, interaction, conversation, language production.

Introduction

Prediction is generally believed to play an important role in language processing. Prediction can speed up our speech production and understanding, helps in avoiding errors, and ensures successful communication in conversation (e.g., Kutas et al., 2011; Pickering & Garrod, 2013; Huettig, 2015; Kuperberg & Jaeger, 2016; Pickering & Gambi, 2018). And even though it remains debated whether prediction is a necessary mechanism to process language (e.g., Huettig & Mani, 2016; Nieuwland et al., 2018), many behavioral and neuroimaging studies have shown that when prediction is possible it aids the processing of semantics, syntax, and phonology (e.g., Altmann & Kamide, 1999; Federmeier & Kutas, 1999; DeLong et al., 2005; Van Berkum et al., 2005; Staub & Clifton, 2006; Grisoni et al., 2017; Ito et al., 2018; Strijkers et al., 2019; Giroud et al., 2023). Most of this work has focussed on the effect of prediction within the individual, where for instance the speed of processing for a final word is compared between a highly constraining versus unconstrained context. Much less investigations have been done exploring the role of prediction in the dyad (e.g., Corps et al., 2018). This is important though, not only because it concerns our most frequent form of language use (e.g., Clark, 1996), but also to investigate whether prediction affects communication across interlocutors, as hypothesized by interactive language models of dialogue (e.g. Pickering & Garrod, 2021). While we have plenty of evidence demonstrating that predictable contexts can facilitate the comprehension or production of upcoming language within an interlocutor, we currently lack evidence demonstrating that such predictions also facilitate the linguistic behavior of our conversational partner. With the current study we aim at addressing this open issue by introducing a novel paradigm where two interlocutors play word association games in either predictive or unpredictive contexts.

Psycholinguistic research within the dyad is not new, but remains scarce and is mainly focussed on alignment, where interlocutors tend to copy each other's linguistic behavior (e.g., Garrod & Anderson, 1987; Brennan & Clark, 1996; Branigan et al., 2000; Hartsuiker et al., 2004), or joint action, where speakers need to coordinate their language use to achieve a common goal (e.g., Brennan et al., 2010; Kuhlen & Abdel Rahman, 2023). Prediction in interaction has also been investigated, but mainly in how it affects turn-taking in listeners (e.g., Levinson, 2016). Participants are faster responding to questions, either with button-presses (e.g., De Ruiter et al., 2006; Magyari et al., 2014; Ryskin et al., 2020) or verbal responses (e.g., Bogels et al., 2015; Corps et al., 2018; see also for written responses: Corps & Pickering, 2023), when the content of those questions is predictable. For instance, Corps and colleagues (2018) observed that people are faster in giving simple yes/no replies to questions when sentence meaning is predictable (e.g., *Are dogs your favorite animal?*) compared to when the content of the question is not predictable (e.g., *Do you enjoy going to the supermarket?*). These findings

are compatible with studies demonstrating that prediction aids language comprehension (e.g., Altmann & Kamide, 1999; Staub & Clifton, 2006; Ito et al., 2018), and show that listeners can use those predictions to speed up their own speech planning.

In the present study we want to take these findings one step further and test whether meaning predictions affect dyadic interactions. That is, we will investigate the impact of prediction in two participants (rather than one) engaging in interaction with each other and explore whether meaning predictability not only speeds up the naming behaviour of an interlocutor completing a sentence, but also the replies given by the conversational partner.

To do so, we developed a paradigm where two participants play simple word association games with each other (see Figure 1). One interlocutor names an object presented on a computer screen (e.g., *dog*), and the conversational partner can freely reply, as long as the reply concerns a word from the same semantic category (e.g., *cat*). The paradigm thus concerns a simplified "mimicking" of conversational acts in that interlocutors can reply what they want as long as it is meaning relevant, but retains elements of a controlled design in that the experimenter dictates the input of the word association game. To manipulate predictability, the input (e.g., picture of a '*dog'*) could be preceded by an unpredictable sentence context (e.g., *'Outside my window I saw a...'*) or a predictable sentence context (e.g., *'Man's best friend is a...'*). Just as found in the work on turn-taking (e.g., Bogels et al., 2015; Corps et al., 2018), completing a predictable sentence should go faster than an unpredictable sentence. The key question here is whether it will also facilitate the naming behaviour of the conversational partner who formulates a reply. If dialogue models rooted in prediction are correct (e.g., Pickering & Garrod, 2013; Pickering & Gambi, 2018), the hypothesis is that prediction should facilitate language production of both interlocutors to a similar extent.

Methods

All materials, the code, analyses pipelines and raw data are made freely available on our online OSF repository (https://osf.io/ujpb8/).

Participants

40 French native speakers (34 female) were recruited to form 20 dyads. All participants were between 18 and 40 years old (mean age = 21.02 (SD = 2.18), had no neurological or speech disorders and had normal or corrected to normal vision. Each was paid 10 euro per session and received course credit if required.

Materials

36 French words were chosen, half of which corresponded to animal names and half to tool names, and which serve as picture input to start a dyadic interaction (items selected from the MultiPic database: Duñabeitia et al., 2018). We choose those two categories because they are well-established semantic categories with non-overlapping meaning representations (e.g., Caramazza & Mahon, 2003), are easily depictable, and categories with many lexical items allowing for a wide range of possible replies. We controlled the stimuli for their lexical frequency (animal words: M=5.65, SD=8.19; tool words: M=9.68, SD=12.9), number of letters (animal words: M=6.94, SD=1.80; tool words: M=6.61, SD=1.24) and number of syllables (animal words: M=2.22, SD=0.88; tool words: M=1.89, SD=0.58). We then constructed for each of the stimuli highly predictable (HP) and unpredictable (UN) sentences (see Table 1). To make sure that the sentences fulfilled the predictability requirement, a cloze-probably test (Bloom & Fischler, 1980; Block and Baldwin, 2010) was conducted online with 68 respondents (using the online FindingFive platform following criteria from previous work: Fairs & Strijkers, 2021). In this test participants were asked to complete the sentences with a word that first comes to their mind. We only kept those sentences for which at least 70% of the respondents gave the intended item. The mean cloze-probability result for the animal category was 86.4% (SD=8.7), and the mean cloze-probability result for the tool category was 87% (SD=8.9). Among the sentences in the UN condition, the highest probability score was 38%. Relying on this cut-off, we were left with 18 items and corresponding HP sentences in the animal category and 18 items and corresponding HP sentences in the tool lists (full stimuli set: https://osf.io/ujpb8/). Per item, the length (number of words) of the HP and UN sentences was identical.

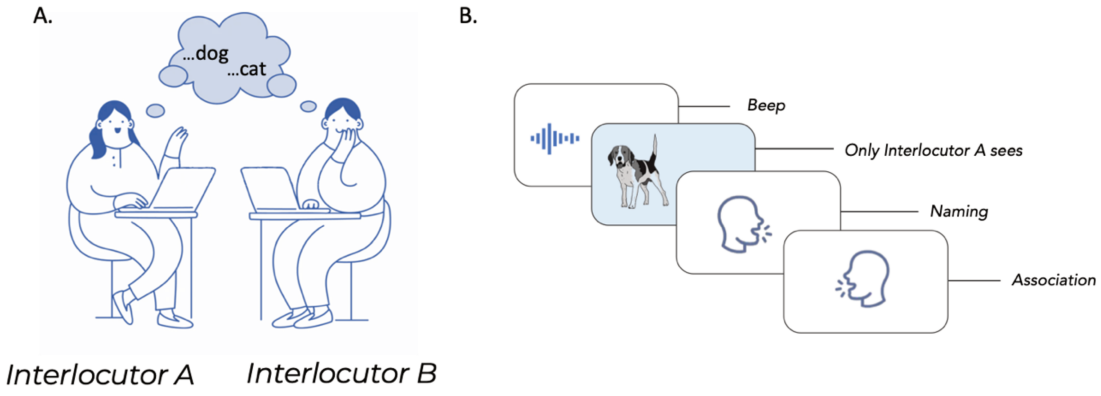


Figure 1. A visual representation of the experimental setting and design. *A. Participants are seated across each other, and both have presentation laptops in front of them. B. In all three conditions participants first hear an auditory input, a beep sound, or a sentence, proceeded by an image appearing on the Interlocutor A's screen. Interlocutor A names the image, and Interlocutor B replies with an association that belongs to the same semantic category, i.e., either an animal or a tool.*

Task and procedure

The experiment setup is shown in Figure 1. A trial had the following sequence: Interlocutor A saw an image and named the object aloud (i.e., standard object naming; e.g., Levelt, 1989), while Interlocutor B responded to the utterance of Interlocutor A with a semantically related word. They alternated roles in each trial. In the baseline condition (the Isolated Word condition; IW), participants heard a beep as attentional cue to signal the start of a trial. To explore the effect of predictability, two additional conditions were included: the highly predictable (HP) and unpredictable (UN) conditions. In these, both participants in the dyad heard auditorily presented sentences from the cloze-probability test (instead of the beep in the IW), after which an object was presented and Interlocutor A had to complete the sentence by naming the object (for similar sentence completion designs see e.g., Griffin & Bock, 1998; Ferreira & Griffin, 2003; Drake & Corley, 2015; Kleinman et al., 2015). Just as in the IW condition, Interlocutor B had to respond to Interlocutor A with a categorically related word. All participants cycled through all three conditions, which were administered randomly, for a total of 216 trials (72 per condition), with each stimulus repeated six times.

Category	*Word*	*Highly predictable sentence*	*Non-predictable sentence*
animal	**âne** *(donkey)*	Bourriquet, l'ami de winnie-l'ourson, était un… *(Eyore, Winnie-the-Pooh's friend, is a…).*	Au marché hier, il a pensé à acheter un… *(At the market yesterday, he bought a…).*
tool	**balai** *(broom)*	Elle nettoie le sol avec un… *(She is sweeping the floor with a…).*	Je ne trouve toujours pas un… *(I still can't find a…).*

Table 1. *Examples of the stimuli*

Analyses

Analyses included Linear Mixed Effects Model (LME) to analyze reaction times (naming latencies) with 'condition' (IW, HP and UN), 'repetition' (from 1 to 6) and 'category' ('animal' or 'tool') as fixed effects and 'dyads' and 'items' as crossed random effects (Baayen et al., 2008; Barr et al., 2013). The models were fitted with lmer (for LME) function from the lme4 package (Bates et al., 2015) in the R statistical computing environment (version 4.2.0, R Core Team, 2022) and the Pymer4 library in Python (Jolly, 2018). We used the control condition as the reference. Reaction times were log-transformed prior to analysis. We used the maximal random structure model that converged (Barr et al., 2013), and this included by-dyad and by-item random intercepts.

Results

To compute the main results, we included correct trials only. From the total 4320 trials 436 were removed due to either a wrong response or no response by either interlocutor, or when a word was preceded by an article, verbalised hesitation, or self-correction. Hence, 3884 correct trials (89.9%) entered the analyses. Using a combination of Python and Praat scripts (https://osf.io/ujpb8/), we calculated speech onset times for Interlocutor A as the picture onset to their speech onset, and for the Interlocutor B from Interlocutor A's speech offset to Interlocutor B's speech onset. Note, importantly, that by calculating Interlocutor B's naming latencies after Interlocutor A's response (rather than from stimulus onset or Interlocutor A's speech onset), the naming latencies of Interlocutor B are independent from the speed of naming of Interlocutor A. The mean reaction times are presented in Tables 2 and 3, and the response distributions per condition in Figure 2.

Condition	N	Interlocutor A RT	SD	SE	CI
Isolated Word	1282	1092	568.63	15.88	31.15
Predictable	1321	740	481.93	13.25	26.01
Unpredictable	1281	974	508.36	14.20	27.86

Table 2. *Mean reaction times (msec) per condition for Interlocutor A.*

Condition	N	Interlocutor B RT	SD	SE	CI
Isolated Word	1282	2309	1233.15	34.44	67.56
Predictable	1321	1875	1189.93	32.73	64.22
Unpredictable	1281	2276	1216.73	33.99	66.69

Table 3. *Mean reaction times (msec) per condition for Interlocutor B.*

Below we report regression coefficients (*b*), standard errors (SE), *t*-values, corresponding *p*-values and Cohen's *d*. For Interlocutor A, reaction times were significantly faster in the Predictable (HP) and Unpredictable (UN) conditions compared to the Isolated Word (IW) condition (HP condition: b = 0.39, SE = 0.058, t = 6.70, p < 0.001, d = 0.67; UN condition: b = 0.121, SE = 0.043, t = 2.85, p = 0.007, d = 0.22). A direct comparison between the HP and UN likewise resulted in significant differences, with faster responses in the HP condition (b = 0.264, SE = 0.053, t = 4.950, p = < 0.001, d = 0.47). Additionally, while Interlocutor A was progressively faster with each repetition of an item (b = 0.053, SE = 0.009, t = 5.865, p < 0.001), importantly, there was no significant interaction between the HP and Repetition or between the UN and Repetition (p = 0.218 and p = 0.745, respectively). Also, category did not affect the response times of the Interlocutor A (p = 0.447).

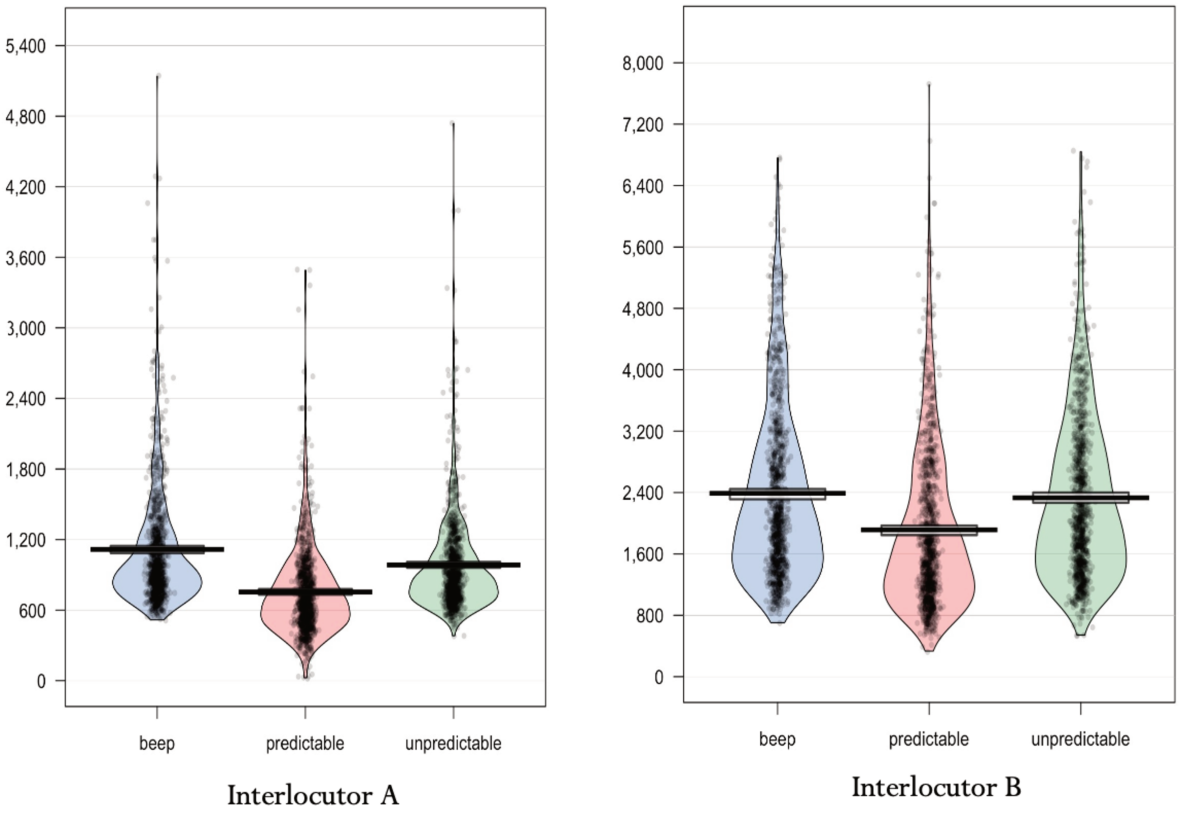


Figure 2. Pirate plots visualising the reaction time distributions for Interlocutor A (left) and Interlocutor B (right) per condition (blue – IW, red – HP, green – UN). *Grey dots reflect each datapoint per condition and black lines index the average reaction times (in msec) per condition.*

Importantly, also for Interlocutor B reaction times were significantly faster in the HP condition compared to IW (b = 0.219, SE = 0.05, t = 4.39, p < 0.001, d = 0.36) and compared to the UN (b = 0.240, SE = 0.06, t = 4.356, p = < 0.001, d = 0.35). There was no significant difference between IW and UN (p = 0.607). Interlocutor B's reaction times did not progressively differ with each Repetition (p = 0.112), and the interaction between Repetition and Predictability was not significant (p = 0.551 for the HP, and p = 0.375 for the UN). Also, Interlocutor B was faster replying to stimuli denoting animals compared to tools (b = 0.082, SE = 0.033, t = 2.505, p = 0.02), suggesting that it was easier to find animal than tool associates.

We also compared the roles of Interlocutor A and B directly by including 'Role' as a fixed effect in the model. We observed a significant effect of Role (b = 0.703, SE = 0.05, t = 14.33, p < 0.001, d = 1.30), with overall faster reaction times for interlocutor A than interlocutor B (see Tables 2 and 3). Interestingly, the interaction between Role and Predictability (HP – UN) was significant (b = 0.203, SE = 0.046, t = 4.439, p < 0.001, d = 1.05). This result highlights stronger prediction effects for interlocutor B than interlocutor A in a given dyad (see Figure 3).

Finally, we also explored psycholinguistic variables of the responses given by the Interlocutor B known to affect response time speed (e.g., Strijkers et al., 2010), to assess whether these variables would differ between the predictability conditions. The lexical

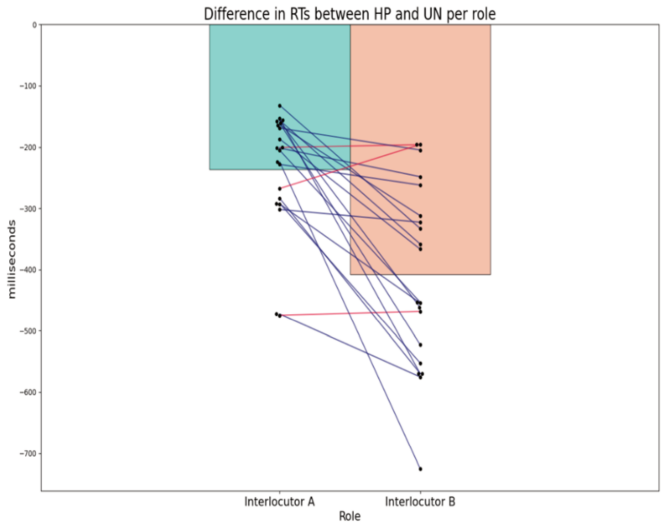


Figure 3. Representation of the prediction effect (expressed in 'reaction time gain' by subtracting HP from UN) per role in each dyad. *Black dots represent individual dyads, with blue lines connecting those interlocutors of the same dyad where the prediction effect (gain) was bigger for Interlocutor B and with red lines connecting those interlocutors of the same dyad where the prediction effect (gain) was bigger for Interlocutor A. The turquoise bar shows the average gain for Interlocutor A and the orange bar shows the average gain for Interlocutor B.*

frequency of the responses given by Interlocutor B did not differ in the HP condition ($p = 0.952$) or the UN condition ($p = 0.377$) compared to the control condition. The lexical frequency also did not differ with repetition ($p = 0.9$), or between tool and animal categories ($p = 0.537$). Likewise, the replies given by Interlocutor B did not differ in word length (phonemes or syllables) for the HP condition ($p = 0.227$), the UN condition ($p = 0.888$), repetition of a stimulus ($p = 0.136$), or category ($p = 0.932$) compared to the control condition. Of the replies given by Interlocutor B, broken down by the number of repetitions, 50.3% of responses were unique, 19.9% were repeated once, 10.8% were repeated twice, 8.2% were repeated three times, 5% were repeated four times, and 1,8% were repeated more than four times. In sum, neither differences in psycholinguistic properties like frequency or length of given replies nor excessive repetition of the same reply can account for the huge prediction effects observed in the experiment.

Discussion

In the present study we investigated whether meaning prediction occurs in a dyadic interaction, and whether it facilitates the communicative behaviour of both interlocutors. In the experiment dyads of participants engaged in a simple semantic association game where we manipulated whether the upcoming semantic category could be predicted or not. Results demonstrated that when the meaning of the upcoming utterance can be predicted, speech production is speed up by at least 200 ms in both interlocutors, and in an incremental fashion. Faster naming latencies for the interlocutor who completes the predictable sentence

(*interlocutor A*) replicates previous research that shows that meaning prediction aids language comprehension (e.g., Altmann & Kamide, 1999; Staub & Clifton, 2006; Ito et al., 2016; Grisoni et al., 2017), which in turn facilitates that listener's own productions to complete a sentence (e.g., Griffin & Bock, 1998; Drake & Corley, 2015; Kleinman et al., 2015; Staub et al., 2015), or to formulate an answer to the perceived sentence (e.g., Bogels et al., 2015; Corps et al., 2018). The important contribution here is that meaning prediction also facilitates the replies given by the conversational partner (*interlocutor B*), and this dyadic predictability effect is even 'larger' compared to the sentence completion response to which the original prediction applied (Figure 3).

This finding is in line with work on joint action which proposes that successful social interaction requires predicting and representing perceptual, motor and cognitive representations of our partner (e.g., Sebanz et al., 2006; Sebanz & Knoblich, 2009). Similarly, for language, intrinsically a social act (e.g., Tanenhaus & Brown-Schmidt, 2008), evidence has cumulated showing that we do not only activate language representations we intend to utter ourselves, but also those of our conversational partner in order to predict that partner's verbal behaviour (e.g., Brennan et al., 2010; Baus et al., 2014; Gambi et al., 2015; Kuhlen & Abdel Rahman, 2017; 2023; Brehm et al., 2019; Cirillo et al., 2022). Our results extend this work by highlighting that thanks to predicting from which semantic category our partner will utter a lexical item, we ourselves can prepare in advance a meaningful reply, which speeds up our proper productions and substantially facilitates the conversational interaction. This dynamic of simulating the linguistic representations of other interlocutors via predictive processing in order to enhance communication and alignment between language users is exactly the mechanism proposed by interactive dialogue models rooted in prediction (e.g., Pickering & Garrod, 2013; Pickering & Gambi, 2018). The present study offers empirical evidence for this role of prediction in the dyad and confirms a key property of interactive dialogue models, namely that predictive processing not only facilitates language processing of the interlocutor to whom the predictive content applies, but also that of the partner who participates in the conversation; in fact, if anything, the impact of prediction is stronger for the latter, which is highly suggestive that predictive processing could indeed be the key mechanism for alignment.

In conclusion, relying on a novel dyadic paradigm combining controlled picture naming with more free associative naming, we demonstrated that prediction strongly facilitates the verbal behavior of both interlocutors engaging in conversation. Such synergetic dyadic predictions can aid communication and alignment as suggested by interactive dialogue models.

Chapter 4. Neural alignment of word meaning

(in preparation)

Kerr Emilia[1]
Morillon Benjamin[2]
Strijkers Kristof[1,3]

---

[1]Aix-Marseille Université (AMU), Laboratoire de Parole et Langage (LPL), Aix-en-Provence, France
[2] Aix-Marseille Université (AMU) & Inserm, Institut de Neurosciences des Systemes (INS), Marseille, France
[3]CNRS, Provence-Corse, France

Abstract

This study investigates neural alignment between interlocutors as a marker of communicative engagement, focusing on both global interaction dynamics and finer, lexico-semantic level of processing. Using EEG hyperscanning, we examined pairs of participants playing a semantic association game, where one participant named objects from categories (animals or tools), and the other provided a semantically related response. Our results reveal neural alignment at both general communicative level and a more specific level, namely, lexico-semantic one. Multivariate pattern analysis (MVPA) showed meaning-specific classification effects, with neural signatures for animal and tool categories successfully distinguished in both participants. Importantly, inter-brain synchrony (IBS) results indicated task-general alignment in the alpha frequency band and meaning-specific alignment in the theta frequency band, suggesting that neural alignment reflects not only shared communicative goals but also semantic content. Our findings extend theories of neural and linguistic alignment by showing that neural alignment during verbal interaction operates at multiple levels, including specific lexico-semantic processing. This work bridges behavioural and neurophysiological measures, offering new insights into the neurobiological underpinnings of human communication.

Keywords: neural alignment, hyperscanning EEG, alpha and theta connectivity, verbal interaction

Introduction

Neural alignment, a phenomenon which reflects the synchronization of cortical activity between interacting individuals, is thought to be involved in social cognition (e.g., Hari & Kujala, 2009; Hasson et al., 2012; Schilbach et al., 2013). In language, neural alignment between conversational partners has been reported in a number of studies and for different neurophysiological responses (e.g., Stephens et al., 2010; Kuhlen et al., 2012; Dikker et al., 2014; Silbert et al., 2014; Perez et al., 2017; 2019; Ahn et al., 2018; Liu et al., 2019). For instance, in seminal work by Stephens and colleagues (2010) brain regions in the Perisylvian language cortex became coupled between a speaker telling a short story and a group of listeners hearing that same story. Moreover, this speaker-listener brain coupling was stronger for those who had better comprehension of the story. Based on this work and that of others, neural alignment during language processing is thought to facilitate effective communication by ensuring that the cortical processes underlying speech comprehension and production are closely coupled between interlocutors (Hasson et al., 2012; Schoot et al., 2016; Kelsen et al., 2022, Schwab-Mohamed et al., 2024). However, which linguistic information of the communicative signal drives this neural alignment is unclear.

Gaining more fine-grained insights on the language processing properties of neural alignment is important to uncover whether it reflects the brain's signature of a famous behavioral form of alignment, namely linguistic alignment. Linguistic alignment is a well-studied topic in psycholinguistics and concerns the tendency for interlocutors to copy each other's language use during conversation (e.g., Pickering & Garrod, 2004; 2021; Dell & Ferreira, 2016; Rasenberg et al., 2020). For instance, we often adjust our concepts and words, and how we structure those words, to mirror our conversational partner. Overall, the idea is that linguistic alignment enhances mutual understanding and facilitates successful communication (Pickering & Garrod, 2004; 2021). While it is tempting to draw a link between this linguistic alignment and the neural alignment of brain responses between speakers and listeners, empirical evidence that allows to establish such a link is missing (e.g., Pickering & Garrod, 2014; Schoot et al., 2016). For one, a defining property of linguistic alignment is that it occurs at multiple levels of language processing, including for instance semantic, syntactic and phonological representations (e.g., Bock, 1986; Garrod & Anderson, 1987; Brennan & Clark, 1996; Branigan et al., 2000, 2011; Pardo, 2006; Hartsuiker et al., 2004; Goldrick et al., 2014; Pardo et al., 2017; Cirillo et al., 2023). In contrast, the phenomenon of neural alignment has been predominantly investigated from a general communicative perspective, measuring interbrain synchrony for the discourse- and message-level of processing (e.g., Schoot et al., 2016; Kelsen et al., 2022), but without distinguishing between different linguistic representations. An important question that

arises is whether neural alignment concerns a domain-general marker of overall communicative engagement or whether it reflects more nuanced aspects of language processing and is the neural mechanism underlying linguistic alignment.

The present study aims to investigate whether neural alignment can be observed specifically for word meaning. To do so, participants grouped in dyads engaged in an interactive language processing task while the electrophysiological (EEG) activity from both participants in a given dyad was recorded simultaneously (see Figure 1). This method, known as EEG hyper-scanning, allows to measure the synchrony in scalp activity between interacting participants (e.g., Dumas et al., 2011; Babiloni & Astolfi, 2014). The task they performed concerns a semi-controlled semantic association game we recently developed to isolate the meaning of words during live interaction (Kerr et al., under review) (see Figure 1). Interlocutor A sees an object on a computer screen and has to utter aloud the object's name (i.e., standard object naming; e.g., Levelt, 1989). Next, Interlocutor B has to respond with a word belonging to the same semantic category. Even though we can't control what exactly interlocutor B will reply, we can control what objects are presented to interlocutor A. This way, if interlocutor A is presented with an animal to name (e.g., '*dog'*), then interlocutor B will also reply with an animal name (e.g., '*cat'*). And similarly, if interlocutor A names the picture of a tool (e.g., '*hammer'*), interlocutor B will also reply with the name of a tool (e.g., '*saw'*). This allows us to contrast animal versus tool word categories, which is a well-established meaning-related property engaging different neural networks in the brain (e.g., Caramazza & Mahon, 2003; Martin, 2007; Binder et al., 2009; Pulvermuller, 2013; Kemmerer, 2014).

The cortical activation differences between animal and tool words can be picked up at the scalp level with EEG and correlated between interlocutors to investigate whether we can observe meaning-specific neural alignment. Several outcomes are conceivable: If neural alignment in language solely reflects domain-general communicative engagement, then we should observe interbrain synchrony between interlocutors regardless of word meaning. Following previous work, such overall speaker-listener coupling will most likely be in the alpha band (e.g., Perez et al., 2017; 2019; Ahn et al., 2018). However, if neural alignment is sensitive to specific language processing components, then we should observe animal- and tool-related interbrain synchrony. If so, this synchrony may be qualitatively similar to the domain-general effects (e.g., alpha band for similar electrodes), suggesting previous speaker-listener coupling effects are at least in part driven by word meaning. Alternatively, meaning-specific neural alignment could be qualitatively different from domain-general neural alignment in which case we should observe that interbrain synchrony specific to animals and tools affects different frequency bands (and/or electrode sites) compared to general speaker-listener synchrony.

Methods

All materials, the code, analyses pipelines and raw data are made freely available on our online OSF repository (https://osf.io/rdmq6/).

Participants

40 French native speakers (32 female) were recruited to form 20 dyads. All participants were between 18 and 40 years old (mean age = 22.4 (SD = 3.5), right-handed, had no neurological or speech disorders and had normal or corrected to normal vision. The experiment was approved by the local ethics guidelines from Aix-Marseille University and were conducted following several guidelines of the Declaration of Helsinki. Each participant gave their written informed consent and was paid 20 euro per session and received course credit if required.

Materials

Stimuli included the same collection of words denoting animals and tools that were used in Kerr et al. (under review; Chapter 3; online OSF repository https://osf.io/ujpb8/). For this experiment we only used the Isolated Word condition (see Task and Procedure for details). As in Chapter 3, in each experimental session participants were exposed to 18 animal words and 18 tool words.

Task and Procedure

As in Kerr et al. (under review; Chapter 3), participants were comfortably seated across each other, each having a laptop in front of them. The difference from Kerr et al. (under review; Chapter 3) is that now both participants' EEG signal was recorded (Figure 1), as well as participants' speech. The task that participants needed to execute was the same as in Kerr et al. (under review; Chapter 3) with the main difference being that here we only tested the Isolated Word condition. Like Chapter 3 and as detailed in the Introduction of that Chapter, interlocutors engaged is a semantic association game, where Interlocutor A named aloud a picture of a tool or animal, and Interlocutor B had to reply with a categorically related word. In total, the experiment consisted of 216 trials, which means that each stimulus was repeated six times. The trial duration was fixed at 8 seconds. The order of presentation of the stimuli was completely randomised. During the mounting of the EEG caps, participants were familiarised with the stimuli by seeing each of the pictures with the intended written label below it.

EEG data collection

During the entire session, participants' electrical brain activity was continuously recorded from the scalp at a sampling rate of 2048 Hz using a BioSemi Active Electrodes

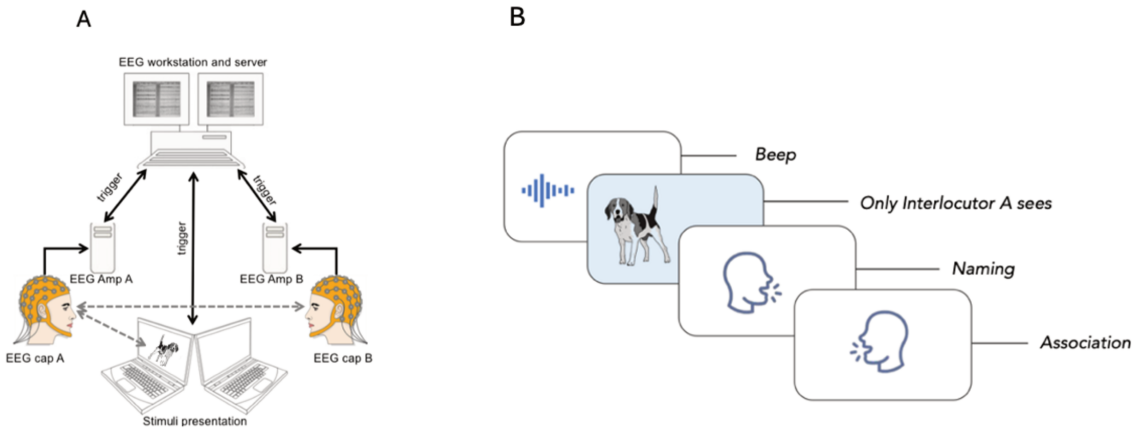


Figure 1. Schematic illustration of the dual-EEG set-up and trial sequence. ***(A)*** *Two participants wearing the EEG caps which are linked to two separate amplifiers and are connected via a Daisy-Bridge hardware to synchronise stimuli presentation and triggering.* ***(B)*** *A trial sequence started by an attentional cue (beep), after which Interlocutor A sees a picture and names it (naming), and Interlocutor B replies with a categorically related word (association).*

system and an online bandpass of 0.01-417Hz with two standard 64-electrode caps (the International 10/20 system (American Encephalographic Society; 1994) fitted to both participants and linked to two amplifiers which passed the amplified electrical scalp activity to the EEG server for data storage (see Figure 1). impedances were kept below 20 kΩ. Stimuli presentations and triggers were synchronized between the different systems using specialized hardware (BioSemi Daisy-Bridge). The triggers were time-locked to each picture onset time. Eye-movements were monitored by EOG electrodes positioned under the left eye of the participants and the prefrontal scalp electrodes, and 2 external electrodes were placed on the left and right mastoids of each of the participants.

EEG Data Preprocessing

Of the 20 dyads that were tested, 5 were excluded from the analyses due to too much noise in the EEG signal (mainly due to articulation and movement potentials), leaving too few remaining trials for analyses after ICA and artefact rejections (for details, see below). Note that even if only one of the participants in the dyad had too much noise, still both participants, that is the dyad as a whole, were removed from analyses. One additional dyad was removed due to missing audio recordings because of the microphone malfunction. Hence, 14 dyads (28 participants) in total entered the main analyses pipeline.

The EEG data were preprocessed in the MNE Python environment (Gramfort et al., 2013; v1.7.0). Preprocessing involved the following sequence: (1) removing bad trials; (2) running the Independent Component Analysis (ICA) on the continuous signal to remove

artefacts; (3) creating two differently filtered sets, one for the MVPA analysis and one for the interbrain synchrony analysis; (4) creating different epochs and time-windows of interest for analysing the interaction in an event-related like manner. Each of these steps will now be detailed more below.

The recorded signal was split into two recordings corresponding to each participant in the dyad. Trials that contained erroneous replies, no replies either from Interlocutor A or Interlocutor B, too much laughter or unrelated utterances between participants were annotated as 'bad' in the data and were excluded from further analyses, as well as trials where articulation onset was earlier than 500ms either for Interlocutor A or Interlocutor B (Mean of excluded trials = 22.8 (= 10.55% of all data), SD = 17.7, range = 2-66). The 500ms cut-off rate was chosen to have as little articulation artefacts in the data with as little data loss as possible (e.g., Strijkers et al., 2010). Also breaks (participants were given 3 breaks per session) were annotated and excluded from the analyses. Bad channels were defined (a) during data collection, (b) with an automatic bridged electrode detection algorithm (Tenke & Kayser, 2001), and (c) by visually inspecting the data (Mean of excluded electrodes = 5 (= 8% of all electrodes), SD = 3.9, range = 0-15). The following step was the application of the independent component analysis (ICA) technique to remove noise coming from the oculomotor activity and articulation. To do so the data were filtered at 1-40Hz which is optimal for ICA. The algorithm for implementation of the ICA was FastICA (Hyvarinen & Oja, 2000), and 15 components were computed to be passed to the ICA algorithm. On average, 4.4 (SD = 2.02, range = 2-8) components linked to eye movements, speech and movement artefacts were removed. Examples of excluded ICA components are presented in Figure 2. Next, the data were re-filtered at 0.3 – 8 Hz for the Mutli-Variate Pattern Analyses (MVPA) and at 0.3 – 30 Hz for the interbrain synchronisation (IBS) analysis (see section Analyses and Results below). The data were re-referenced using the average of all channels. Next, off-line triggers were created to time-lock the signal related to one trial in several relevant event-related epochs for analyses (i.e., Interlocutor A naming an

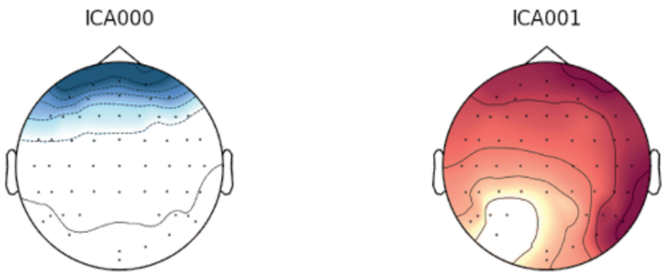


Figure 2. Examples of ICA component analysis from a sample subject. *The left part represents a sample eye blink component, and the right part illustrates a sample component that was attributed to articulation artefacts.*

object and Interlocutor B responding with a semantically related word). The following logic was implemented: each event (trigger) marked the speech onset and offset times on both interlocutors, except for the first trigger that marked the picture onset time. In total, 5 triggers were computed:

- Picture onset
- Speech onset of Interlocutor A
- Speech offset of Interlocutor A
- Speech onset of Interlocutor B
- Speech offset of Interlocutor B.

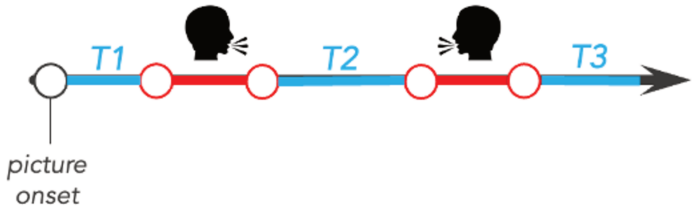


Figure 3. A schematic representation of how each trial was epoched in the analysis pipeline. *The three time windows of interest that entered the main analyses were the time window related to the speech planning by Interlocutor A to name the presented picture (T1), the time window related to the speech planning by Interlocutor B to reply with a semantically related word (T2), and the time window after the reply by Interlocutor B which indexes the end of the interaction (T3). The red areas contained most of the actual articulations (marked according to individual speech onsets and offsets) did not enter the analyses.*

Speech onset calculations were done in a similar fashion as in the Behavioural experiment, with the main difference being that here Interlocutor B's responses were measured from the picture onset to determine the actual trigger location, as opposed to Interlocutor A's speech onset. As shown in Figure 3, this results in three silent intervals, T1, T2 and T3, and two articulation intervals when participants uttered their responses. This was done for two reasons: First, to be able to analyse the verbal interaction between interlocutors in a more fine-grained manner and assess the electrical brain activity linked to (1) object naming (blue interval T1 in Figure 3), (2) associative naming (blue interval T2 in Figure 3), (3) end of the interaction (blue interval T3 in Figure 3); and second, in order to exclude the articulation intervals (the two red intervals in Figure 3) from further analyses hereby reducing further any remaining impact of potential articulation artefacts after the ICA.

The epochs for analyses were coded according to the experimental conditions (animal vs tool words) and their time-window of interest, resulting per trial in three time-windows of interests (T1, T2, T3) of 1000 ms long epochs each. These epochs were baseline-corrected using the 100 ms before the picture onset. Note that by epoching the EEG signal for each trial in this

manner, most of the interval containing articulations of Interlocutor A and B were excluded from analyses (see Figure 3). This final subset was resampled to 128 Hz. Before running the main analyses (see below) the bad channels were interpolated.

EEG Data Analyses

We performed two different analyses on the EEG data: Multivariate Pattern Analysis (MVPA) and Interbrain Synchrony (IBS). While the goal of the IBS analyses is to test the different neural alignment predictions of this study, we first perform time-resolved MVPA, which allows to assess whether a trained classifier can accurately dissociate the neural response pattern (the scalp topography) associated to animal and tool categories (e.g., Haxby et al., 2001; King & Dehaene, 2014). Specifically, it allows to determine (a) whether animal versus tool responses can be classified in each interlocutor of the dyad (i.e., verification whether the task and stimuli allows to assess this meaning-specific difference), (b) when in time these classifications occur in each interlocutor, and (c) whether there are time-windows when successful animal-tool classification overlaps between the interlocutors. One advantage of the MVPA approach is that it is time-resolved, while the IBS is not (calculated over 1 second epochs).

The MVPA was used under the Scikit-learn Python environment (Pedregosa et al., 2011). Each participants' EEG signal was analyzed separately to determine how well the machine learning algorithm can dissociate the neural data related to the animal and tool conditions. The algorithm of choice was the Logistic Regression classifier that uses one-vs-rest scheme for the binary data (Zhu et al., 1997).

To analyze IBS we used the HyPyP procedure (Ayrolles et al., 2021). We used the Circular correlation coefficient (Ccorr) to determine the covariance synchronisation between the two brain signals as it appears to be particularly well-suited for EEG data and is more robust against spurious hyper-connections than other connectivity measures (Burgess, 2013). We computed Ccorr between speakers and listeners, for all electrode pair combinations (64 X 64 electrodes) in the theta (4-7Hz), alpha (8-13Hz) and the beta (14-30Hz) frequency bands, which were extracted using functions developed in the HyPyP pipeline: FIR filtering and Hilbert transformation to perform frequency decomposition of the EEG signal. This was done for the two meaning-related conditions (animals and tools) to assess word-specific speaker-listener IBS, as well as grouped by the conditions to assess general speaker-listener IBS.

Statistical Analyses

The results from both methods mentioned above were tested for significance using permutation clustering approach that includes correction for multiple comparisons. For the

MVPA analysis we used a 1-sample permutation cluster test (Maris & Oostenveld, 2007), which is a non-parametric cluster-level t-test that was implemented through the MNE Python package (Gramfort et al., 2013). For the IBS a cluster-level permutation test from the HyPyP package (Ayrolles et al., 2021) was used, where the synchronisation data from the real dyads was compared to simulated data created via random dyad shuffling (i.e. surrogate dyads; e.g., Bilek et al., 2015).

Results

Behavioural Results

Figure 4 represents the reaction time (speech onsets) distribution for Interlocutor A and B per condition. For the behavioural analysis, speech onsets for Interlocutor A were calculated from the picture onset, and for Interlocutor B from Interlocutor A's onsets, as was done in Kerr et al. (under review; Chapter 3). Mean reaction times for Interlocutor A (i.e., object naming) in the animal condition were 1.03 seconds (SD = 0.45), and for the tool condition 1.03 seconds (SD = 0.41). Interlocutor B's mean reaction times (i.e., associative naming) for the animal condition were 2.09 seconds (SD = 1.16) and for the tool condition 2.16 seconds (SD= 1.15). To explore the behavioural effects, we ran Linear Mixed Effects Model (LME) analysis similar to how it was done in Kerr et al. (under review, Chapter 3). The model included 'role' (Interlocutor A or B) and 'condition' (animal or tool) as fixed effects and 'dyads' as crossed random effects (Baayen et al., 2008; Barr et al., 2013). The models were fitted with lmer (for LME) function from the Pymer4 library in Python (Jolly, 2018). Reaction times were log-transformed prior to analysis. As expected, Interlocutor A (object naming) was faster than Interlocutor B (associative naming) and this difference was significant (b = 0.86, SE = 0.07, t = 12.66, p = < 0.001) (see Figure 4). Critically, reaction times did not differ between conditions

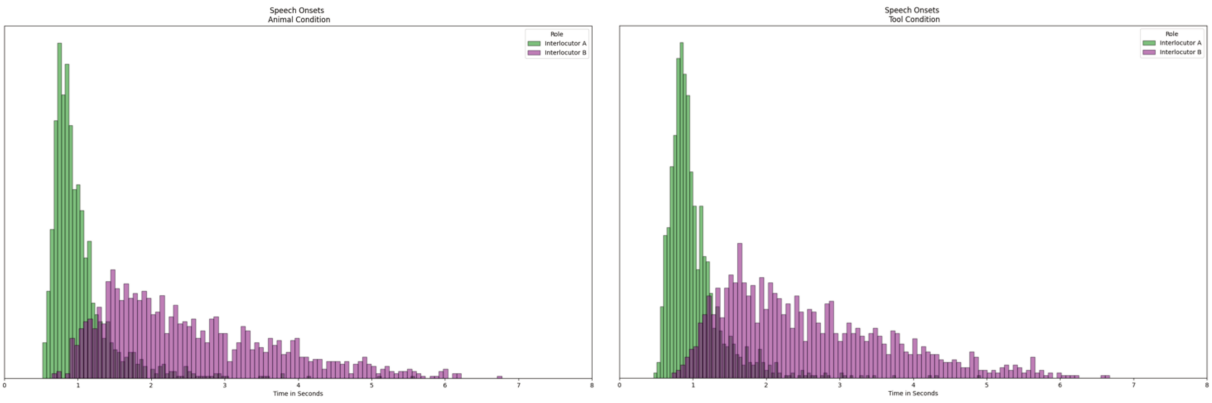


Figure 4. ***Reaction time distribution of the speech onsets for the animal condition (left) and the tool condition (right).*** The green reaction time distributions represent Interlocutor A and the purple reaction time distributions represents interlocutor B.

(p = 0.174), which indicates that the neural time courses can be compared between conditions. This behavioral pattern replicated that of the behavioral study (Chapter 3).

MVPA

The results of the MVPA after correcting for multiple comparisons (see above) are represented in Figure 5. For Interlocutor A (panel A in Figure 5), statistically significant clusters ($p<0.05$) where the algorithm successfully differentiated between the animal and tool conditions are present in T1, from roughly 200 ms after picture onset till the end of the 1 s epoch. There was also significant classification of animals and tools in T2 from Interlocutor's A speech offset (0 ms) until roughly 500 ms after speech offset, and then again from roughly 600ms after speech offset until the end of the epoch; that is, before Interlocutor B's reply. During T3 no significant clusters that survived corrections for multiple comparisons were present.

For Interlocutor B (panel B in Figure 5), there were no significant classification clusters in T1. Crucially, there was a significant classification cluster in T2, when Interlocutor B

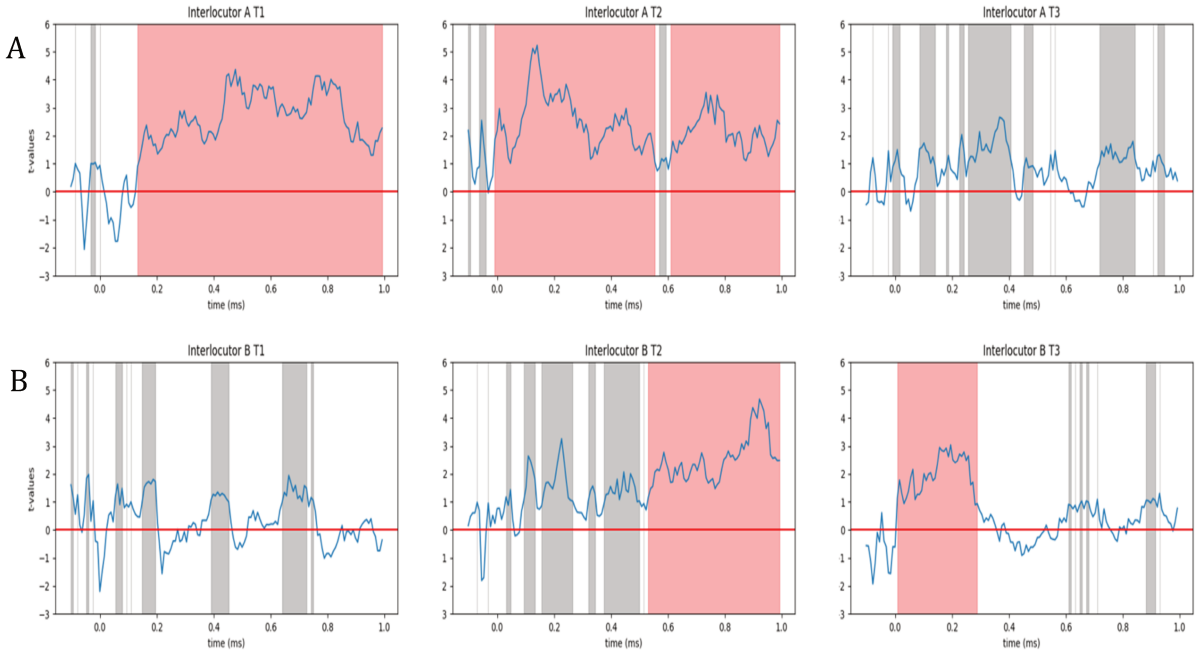


Figure 5. ***MVPA time-resolved results for classifying animals and tools in Interlocutor A (panel A) and Interlocutor B (panel B) per time window (T1, T2, T3).*** Each panel depicts the statistics (t-values) of the decoding at the group-level (N=14), estimated for each time-point of the epochs (from -200ms to +1 s). The red line represents chance level (t=0). The pink areas show significant clusters of successful classification between animals and tools, whereas the grey areas show formed clusters that did not survive correction for multiple comparison and white areas the absence of any successful classification clusters. As can be appreciated, in T2 from roughly 600ms after Interlocutor A named the picture until the end of the epoch, the MVPA successfully classified the EEG data in animals and tools categories in both interlocutors (i.e., classification overlap between interlocutors).

prepares a semantically related response, from 600 ms after the speech offset of Interlocutor A until the end of the epoch. We observed also a significant classification cluster in T3 from Interlocutor B's speech offset (0 ms) and which lasted about 300 ms.

In sum, the time-window when there was classification overlap and thus both interlocutors were processing animal and tool words simultaneously, was in T2 from 600 ms after Interlocutor A's speech offset until Interlocutor B's speech onset (see Figure 5).

Interbrain Synchrony

Figures 6 and 7 show the significant speaker-listener IBS (Ccor) when contrasted to the surrogate datasets and after permutation cluster tests run for correcting multiple comparisons ($p<0.05$). IBS was estimated for the three time-windows (T1, T2, T3) and for three neural frequency bands (theta, alpha, beta). There were no significant effects in T1 for any condition in any of the frequency bands. In T2 there was significant synchronisation when grouped by conditions, that is, for communication in general (irrespectively of the meaning variable), in the theta band (Figure 6). Interestingly and crucially for present purposes, we also observed significant meaning-specific interbrain synchrony for both the animal and tool conditions in T2 for the theta band (Figure 6). Additionally, also in T2 there was significant meaning-specific IBS in the alpha band only in the tool condition (Figure available on the OSF repository). In T3 in the theta band, significant synchrony was observed when grouped by the semantic categories and also for the animal condition (Figures available on the OSF repository, https://osf.io/rdmq6/). We also tested IBS grouped by the semantic categories and all the time-windows. Here we observed significant synchrony in the theta and alpha (but not beta) bands (Figure 7).

T2
Theta Band

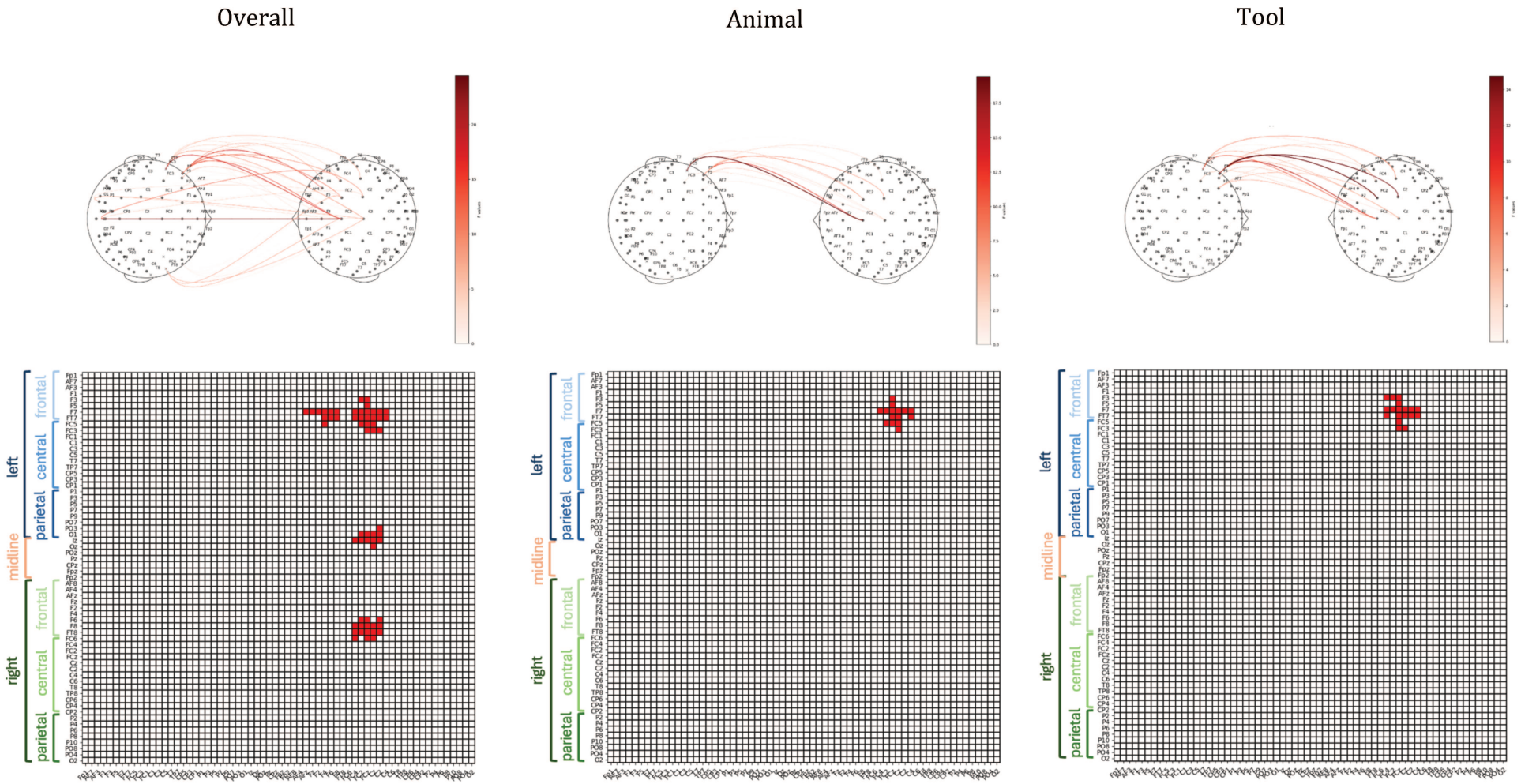


Figure 6. Matrix plots and dual head plots illustrating the interbrain synchrony (IBS) per condition in T2 in the theta band. *The left panel shows the results of speaker-listener synchrony across the semantic categories, the middle panel shows the results of speaker-listener synchrony in the animal condition, and the right panel shows the results of speaker-listener synchrony in the tool condition. Lines linking electrodes between head plots represent electrode pairs where significant synchrony was found (the darker the lines, the higher the significance).*

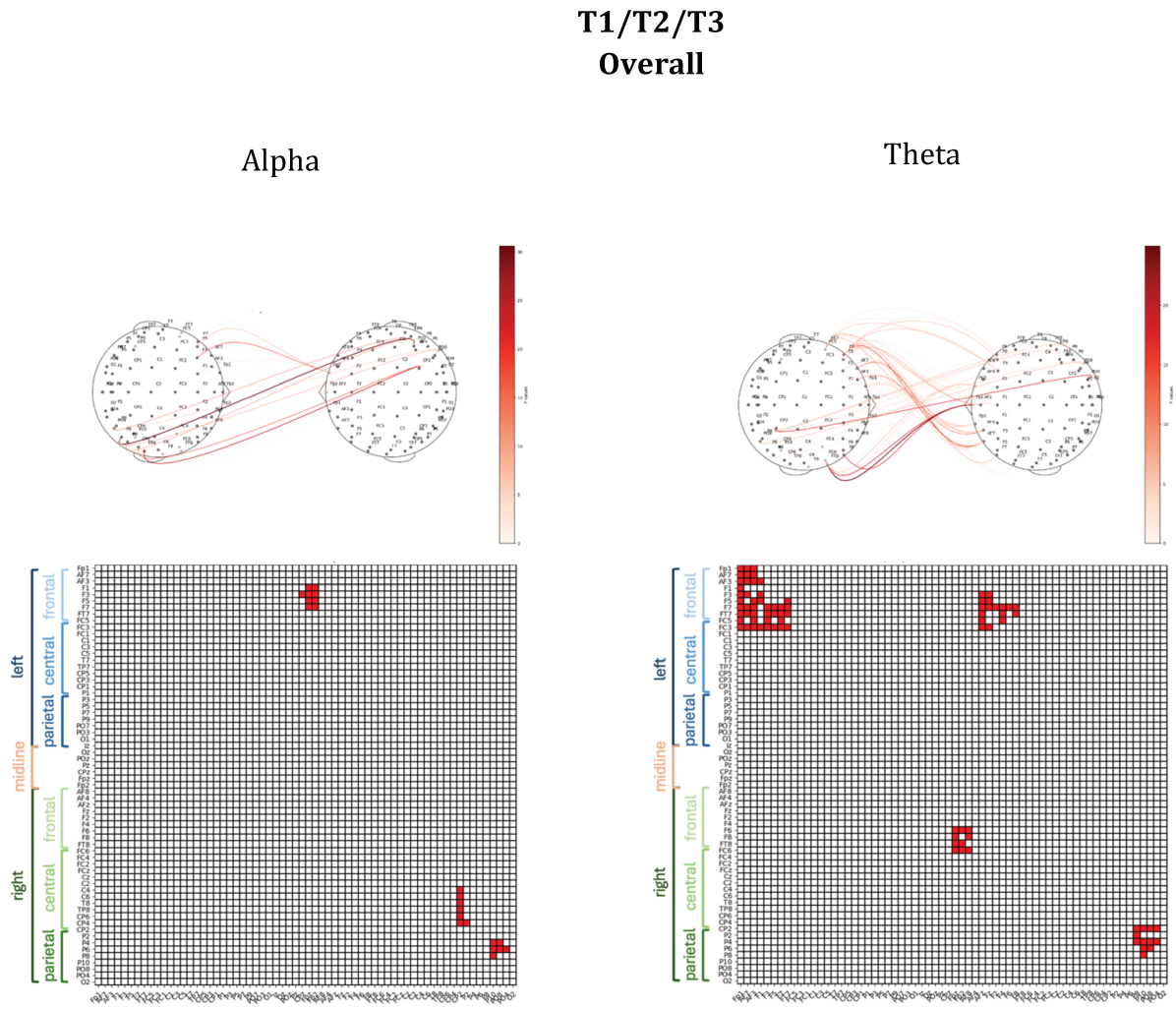


Figure 7. Matrix plots and dual head plots of the interbrain synchrony (IBS) analysis, performed across the semantic categories (animal/tool) and across all time-windows (T1/T2/T3). *The left panel shows the results from the alpha band (8-13 Hz) and the right panel shows the results from the theta band (4-7 Hz). Same convention as Figure 6.*

Discussion

In this study we explored whether neural alignment between interlocutors is a marker of domain-general communicative engagement. This was done by utilising a paradigm where participants were engaged in a game with a semantic manipulation. Therefore, this manipulation allowed us to explore neural alignment at the overall, general, level of the interaction and at a more fine-grained, lower level of processing, namely, semantics. Addressing this question is crucial to mechanistically link alignment between interlocutors at the level of the brain with alignment of language use at the level of behaviour. Using an EEG hyperscanning set-up, pairs of participants played a semantic association game where one participant named aloud objects of animals or tools, and the other participant had to reply with words from the same semantic category, a set-up which allowed us to test whether neural

alignment can be measured at the level of a given semantic category (animal/tool). Both our MVPA and IBS results indicated meaning-specific classification and synchrony between the interlocutors, which could also be distinguished from the task-general synchrony observed. This is the main contribution of our study and provides strong evidence that neural alignment between interlocutors does not only concern a general communicative phenomenon, but also has specific sensitivity to more fine-grained language representations. This finding makes the idea that neural alignment is the brain's implementation of linguistic alignment all the more likely.

The MVPA results revealed that in the first time-window (T1), in the period where Interlocutor A sees the picture and performs a standard object naming procedure (e.g., Levelt, 1989), tool and animal words were successfully classified in the speaker naming the object (Interlocutor A), but not in the listener (Interlocutor B); a logical result since the listener did not have access to the picture and could not yet know which object the speaker was about to utter. Note also that in T1, the time-course of the animal-tool classification, from 200 ms after visual object presentation, fits with previous studies exploring the time-course of semantic categorisation using animal and tool words (e.g., Kiefer & Pulvermuller, 2012), and the time-course of lexico-semantic processing in object naming (e.g., Indefrey & Levelt, 2004; Strijkers & Costa, 2016). The pattern of results found in T1, while not surprising, is relevant since it generalises robust neurophysiological findings observed in individual paradigms, to dyadic paradigms. In the time window corresponding to the speech planning of a semantically related word (T2), we observed significant neural classification of tool and animal words for both participants (A and B). In the participant who had just named the object (Interlocutor A), there were significant effects in the first 500 ms after speech onset, possibly related to the external monitoring of their own speech (e.g., Indefrey & Levelt, 2004; Runnqvist, 2023). More interestingly, from around 600 ms after speech offset until the onset of the reply, the classifier successfully differentiated the EEG responses to animals and tools in both participants. This moment perfectly aligns with Interlocutor B having processed the object-naming utterance (e.g., Indefrey & Levelt, 2004; Costa et al., 2009), and starting to plan a reply related to its meaning, while Interlocutor A starts to expect a semantically related response. In the final time-window after a reply has been formulated (T3), only Interlocutor B showed classification effects for the first 300 ms. Similar as suggested for the early classification response for Interlocutor A in T2, it likely reflects the monitoring of Interlocutor B's own speech.

The IBS results showed no significant effects in T1 across any frequency band (theta 4-7 Hz, alpha 8-13 Hz, or beta 14-30 Hz) or condition (grouped, animal, or tool), which seems logical since during this time-window only Interlocutor A is actively engaged in the task and Interlocutor B has not been exposed to any visual input and has not heard Interlocutor A's

response yet. Hence, at this stage of the experiment, the two participants are not 'in sync' yet. In contrast to T1, in T2 significant overall synchrony (i.e., across the semantic categories of interest) and meaning-specific synchrony (i.e., for both the animal and tool conditions separately) was observed in the theta band (see Figure 6). These results show that during the interaction, when both interlocutors are actively engaged in the processing of a given input, their brain signals start to synchronise. The meaning-specific IBS results are especially of interest because they indicate that the brain-to-brain synchronisation that is observed during a verbal interaction (e.g., Perez et al., 2019; Ahn et al, 2018), can be measured not only in function of higher-level or domain-general processing (e.g., shared situation models or communicative goal) but also at lower levels, specifically the lexico-semantic processing level. The IBS results evident in T2 in the theta band seem to be the most robust and reliable in our experiment, and they also neatly correspond with the time-period of the MVPA results where classification of the animal/tool conditions was significant in both interlocutors simultaneously. Beyond this meaning-specific neural alignment in T2, we also found neural alignment in the theta and alpha bands when looking at the whole trial (i.e., collapsing across the three time-windows (T1/T2/T3), and across the semantic categories (animal/tool)). This is interesting because the alpha connectivity between speaker and listener replicates previous work exploring neural alignment from a global communicative perspective (e.g., Perez et al., 2017; 2019; Ahn et al., 2018). Finally, in T3, we also observed some meaning-specific theta connectivity. However, the pattern of results was less consistent and convincing than in T2 since it was only present for animal words (not tool words), and unlike in T2, showed no correspondence with the complementary MVPA analysis.

Our main findings thus show meaning-specific neural alignment within the theta frequency band, which furthermore is qualitatively different from the more task-general neural alignment we and others observed in the alpha band. These findings represent a major contribution to prior research: While we replicate neural alignment in the alpha band as predominantly found in previous work exploring communicative interactions (e.g., Perez et al., 2017; 2019; Ahn et al., 2018), we manage to demonstrate that neural alignment durig language interactions is also sensitive to more fine-grained linguistic knowledge, namely word category, in the theta band. Oscillations in the theta band have been associated with memory retrieval (e.g., Buzsaki, 2002; Solomon et al., 2017; Herweg et al., 2020), and interestingly, in language, with the encoding of lexico-semantic representations, both in language comprehension (e.g., Bastiaansen & Hagoort, 2006; Meyer, 2018) and production (e.g., Piai & Zheng, 2019). Such interpretation perfectly fits our meaning-specific neural alignment data: The task concerns semantic association and the point in time we observe theta connectivity between listeners and speakers is when one interlocutor is planning a semantically related word to reply to the speech

of the partner and that partner is in anticipation of hearing a semantically related word. In contrast, for the alpha band, we only observed neural alignment when grouping across animal and tool words over the three time-windows of interest. Such more general task-related pattern highlights that alpha connectivity is not related to specific semantic content per se, but rather confirms previous interpretations linking alpha connectivity to verbal working memory and attentional demands to maintain overall communicative coherence (e.g., Klimesh, 1999; Schoot et al., 2016; Perez et al., 2017; 2019; Kelsen et al., 2022).

Our findings provide compelling evidence that neural alignment during dialogue can occur at a more granular linguistic level than previously recognized, specifically at the level of word meaning. This extends existing theories of neural and linguistic alignment by demonstrating that such IBS can encompass specific lexico-semantic components. The observed meaning-specific theta band synchrony suggests that interlocutors achieve a form of neural coupling that correlates with the linguistic alignment at the word meaning level, potentially reflecting aligned mental representations. This provides an empirical basis for the integration of dialogue models, such as the famous Interactive Alignment Model (Pickering and Garrod, 2004), with neural data, thereby bridging the gap between behavioral and neurophysiological measures of human communication. To our knowledge, there is only one other study that reports neural alignment for a specific linguistic component, namely, syntax (Liu et al., 2019). In their study, the authors recorded the brain activity of two interlocutors during an interaction using the fNIRS method. The main manipulation was the difference in the syntactic structures that were used during the interaction (double-object or prepositional-object structures). Their results showed increased IBS for the syntactic manipulation in the right posterior superior temporal cortex (pSTC). However, given that study used a between-participants design and observed the effect only in one of the syntactic conditions (double-object), it remained possible the effect was not syntax-specific alignment per se but indexed general communicative alignment in only one of the participant groups. A similar alternative account for our results is not possible since the meaning-specific neural alignment was observed to both animal and tool words, which furthermore was qualitatively distinct (different connections) from the task-general communicative alignment. In other words, when interlocutors linguistically align at the level of word meaning, as induced by our semantic association task, their brain signals also align, or synchronise, not only due to the interactive environment they are in, but also in function of word-meaning processing. This lays an empirical basis to develop a neurobiological implementation of dialogue models and investigate the mechanisms behind our capacity to align with our conversational partners including neurophysiological measurements.

In conclusion, this study highlights the importance of considering both the frequency, spatial connections and timing of neural alignment during verbal interactions in understanding the mechanisms underlying effective human communication. The distinct roles of theta and alpha band synchrony in facilitating different aspects of dialogue underscore the complexity of neural processes involved in dialogue. Future research should aim to further delineate these mechanisms and explore the conditions under which different types of neural synchrony emerge during interpersonal communication.

Chapter 5. Neural alignment and predictive processing

Kerr Emilia[1]
Morillon Benjamin[2]
Strijkers Kristof[1,3]

[1]Aix-Marseille Université (AMU), Laboratoire de Parole et Langage (LPL), Aix-en-Provence, France
[2] Aix-Marseille Université (AMU) & Inserm, Institut de Neurosciences des Systemes (INS), Marseille, France
[3]CNRS, Provence-Corse, France

Abstract

Using EEG hyperscanning, this study explores the role of predictive processing in facilitating efficient communication by examining whether enhanced prediction increases neural alignment between interlocutors engaged in a simple communicative task. While behavioural results confirmed previous findings of stronger prediction effects in both interlocutors, neural data provided only tentative support for the role of predictability in shaping neural alignment. This was caused by a decrease in power due to more noisy EEG data because prediction resulted in earlier articulation onsets. Despite this issue, the results did indicate that predictable contexts promote stronger neural alignment compared to unpredictable contexts. These findings highlight the potential of prediction to modulate neural synchrony during conversation and underscore the need for further investigation into how predictability influences specific aspects of neural alignment. Despite current limitations, our results lay a foundation for future work exploring predictive mechanisms in dyadic communication.

Keywords: neural alignment, predictive processing, hyperscanning EEG

Introduction

When individuals converse, their language behaviour undergoes a certain degree of convergence, or alignment. Lingusitic alignment is an integral part of human verbal communication. Linguistic alignment refers to the phenomenon where conversational partners synchronize various aspects of their language use, such as word choice, sentence structure, and speech patterns, during communication (e.g., Clark, 1996; Pickering & Garrod, 2004; 2021). This process is thought to facilitate mutual understanding and effective communication by reducing cognitive load, allowing interlocutors to predict each other's speech and intentions more easily (Pickering & Garrod, 2004). Linguistic alignment occurs at multiple linguistic levels, including phonological, lexical and syntactic levels (e.g., Brennan, 1996; Coulston et al., 2002; Reitter et al., 2006; Levelt & Kelter, 1982; Bock, 1986; Garrod & Anderson, 1987; Branigan et al., 2011; Cirillo et al., 2022; Hartsuiker et al., 2004; Corps et al., 2018; Gambi et al., 2015; Kuhlen & Abdel Rahman, 2017; for an overview see Pickering & Garrod, 2021, Part 2). This natural tendency toward alignment reflects a broader set of cognitive and social mechanisms that enable people to interact fluently in real-time conversations.

The Interactive Alignment Model (IAM), proposed by Pickering and Garrod (2004), provides a theoretical framework for understanding how linguistic alignment operates in dialogue. According to this model, alignment between speakers occurs automatically through a series of "priming" processes. As one person speaks, they activate certain linguistic representations in the listener's mind, and these representations become more likely to be used by the listener when they respond. IAM suggests that language comprehension and production are closely linked and that alignment is a foundational aspect of successful dialogue, fostering mutual understanding with minimal effort. The model also emphasizes that linguistic alignment is not merely a byproduct of language use but a fundamental mechanism for achieving shared meaning in interaction. Interestingly, more recently, the model has been extended to include prediction as a driving mechanism to enhance alignment between interlocutors (e.g., Pickering & Garrod, 2013; Pickering & Gambi, 2018).

Predictive processing has emerged as a central concept in cognitive science (e.g., Friston, 2010; Clark, 2013), and for the production and comprehension of language, investigations into the role of prediction have been amongst the most vibrant areas of research in the field (e.g., Kutas et al., 2011; Huettig, 2015; Kuperberg & Jaeger, 2016; Pickering & Gambi, 2018). At its core, predictive processing refers to the brain's capacity to anticipate sensory and cognitive outcomes of a process (top-down), before that process has been fully implemented (bottom-up), allowing for more efficient processing of information (Pezzulo et al., 2024; Friston, 2010). In the realm of language production and comprehension, predictive

processing allows for the proactive activation of upcoming sounds, grammar, words, phrases, and meanings based on prior context (e.g., Altmann & Kamide, 1999; Federmeier et al., 1999; Wicha et al., 2004; Delong et al., 2005; Van Berkum et al., 2005; Otten & Van Berkum, 2008; Strijkers et al., 2011; 2019; Dikker & Pylkkanen, 2013; Grisoni et al., 2017; Ito et al., 2018; but see for exceptions on predictive processing in language: e.g., Huettig & Mani, 2016; Nieuwland et al., 2018). This anticipatory function streamlines communication, making language processing faster, more fluid, and less prone to errors.

Concerning dialogue and broader human communication, research is increasingly focusing on the implications of predictive processing (e.g., Pickering & Garrod, 2013; Friston & Frith, 2015; Schoot et al., 2016; Pickering & Gambi, 2018). Dialogue, as a dynamic and collaborative process, demands a high level of coordination between speakers and listeners. In order to sustain effective communication, interlocutors must anticipate what their conversational partner will say next and prepare their responses accordingly. A particularly strong demonstration of the latter concerns turn-taking, where interlocutors are too fast (within 200 ms) to achieve such a feat without prediction (e.g., Levinson, 2016). Given this prominent role of prediction for language processing, as mentioned higher, the IAM has been extended to include predictive processing as a key mechanism ensuring the 'copying' of linguistic representations between speakers and listeners in dialogue (e.g., Pickering & Garrod, 2013; Pickering & Gambi, 2018)[9]. In short, the idea is that because we try to predict what our interlocutor is about to say, the same linguistic representations can become active in the speaker's and listener's brain resulting in alignment of their language behavior. In this manner, the capacity to predict during dialogue ensures that communication is optimal between the interlocutors and can proceed without constant disruptions, even in the presence of ambiguous or incomplete information.

While the role of prediction during language interactions has thus far been investigated scarcely and mainly focused on turn-taking (e.g., Corps et al., 2018), we recently provided compelling evidence that prediction affects both interlocutors in a dyad. In an experiment (Kerr et al., under review, Chapter 3) that involved a dyadic interaction, participants played a simple semantic association game in pairs. We measured the reaction times (speech onsets) for both interlocutors, one in the role of the 'speaker', and the other in the role of the 'listener'. The main manipulation was the predictability factor. More precisely, on some of the trials, the target word was preceded by a highly predictable sentence while on other trials it was preceded by an

[9] Note that different, non-mutually exclusive, mechanisms for prediction in dialogue have been proposed, like forward models (e.g., Pickering & Garrod, 2013), or prediction-by-association and prediction-by-production (e.g., Pickering & Gambi, 2018). It is beyond the scope of the current study to differentiate between these different mechanisms.

unpredictable sentence. The results showed that prediction directly affected the response speed of both interlocutors, where the responses from the interlocutor in the 'listener' role were modulated by prediction to a greater extent than the ones from the 'speaker'.

Apart from linguistic alignment and predictive processing that are part and parcel of human verbal interaction, there is one more aspect that is believed to be an integral part of the communicative process. That aspect is neural alignment, a phenomenon where the brain signals of two (or more) interloctuors align, or synchronise (interbrain synchrony), when people are engaged in an interaction (e.g., Silbert et al., 2014, Liu et al., 2020; Dikker et al., 2014; Perez et al., 2017, 2019; Stephens et al., 2010; Kuhlen et al., 2012). Studies on interbrain synchrony (IBS) have shown that the cognitive and neural processes that underlie verbal communication are coordinated to a certain extent in the individuals involved. Importantly, this synchronization is not only due to the shared input, e.g., speech, but is promoted by the communicative behaviour itself (e.g., Perez et al., 2017). As was shown in previous work (Kerr et al., in prep.; Chapter 4), IBS can be measured not only at higher levels of language processing, e.g., over an entire conversation, but at lower levels as well. In that study, IBS was measured in pairs of participants who were engaged in a semantic association game. The crucial finding was that we succeeded in capturing IBS both in a task-general manner (grouped over semantic categories showing alpha-connectivity) and for the meaning-specific component (per each semantic category showing theta-connectivity). This suggests that neural alignment follows similar patterns as linguistic alignment. Namely, it occurs at more general levels of the interaction, as demonstrated by prior research, and additionally, at more fine-grained levels of linguistic processing. For example, at the level of word meaning, interlocutors seem to exhibit neural synchrony in function of the semantic category they are exposed to (see also Liu et al., 2019 for syntax-specific IBS).

The goal of the present study is to explore the role of predictive processing in dialogue from a neurophysiological perspective. More specifically, below we will attempt to bridge the prediction in linguistic alignment and neural alignment by investigating whether prediction affects interbrain synchrony in a similar way that it does for linguistic alignment. That is, given that prediction modulates linguistic alignment in a simple interaction that involves single word processing (Kerr et al., under review) and that IBS can be measured at the same meaning-specific level (Chapter 4 of the present thesis), the question that we want to explore here is whether the prediction manipulation as implemented in Kerr et al. (under review; Chapter 3) will in any way affect neural alignment between interlocutors.

To do this, I conducted a hyperscanning EEG experiment, similar to the one described in Kerr et al. (in prep.; Chapter 4), with the main difference being that now the manipulation

involves the prediction modulation, i.e., prior to giving their own responses participants are exposed to either highly predictable or unpredictable sentences (the procedure is described in more detail in the Methods section below). Based on the previous work (Kerr et al., under review; Kerr et al., in prep.), we hypothesize that prediction will speed up the naming behaviour in both interlocutors, with an even stronger dyadic prediction effect in the interlocutor replying with a semantic associate (Interlocutor B) compared to the interlocutor completing a predictable sentence (Interlocutor A). Consequently, if neural alignment is linked to linguistic alignment as our results from Kerr et al., (in prep.) suggest, we expect that prediction should have a similar impact on meaning-specific IBS. That is, the neural alignment specific to tools and animals between the interlocutors should emerge sooner (already from the first time-window; see Methods below) and be stronger for the predictable compared to unpredictable contexts. Moreover, we expect that the prediction effect upon neural alignment will also be stronger for the interlocutor replying (Interlocutor B) compared to the conversational partner completing the predictable sentence (Interlocutor A).

Methods

All materials, the code, analyses pipelines and raw data will be made freely available on our online OSF repository[10].

Participants

40 French native speakers (31 female) were recruited to form 20 dyads. All participants were between 18 and 40 years old (mean age = 23.9 (SD = 3.8), right-handed, had no neurological or speech disorders and had normal or corrected to normal vision. The experiment was approved by the local ethics guidelines from Aix-Marseille University and were conducted in accordance with the Declaration of Helsinki. Each participant gave their written informed consent and was paid 20 euro per session and received course credit if required.

Materials

Stimuli included the same collection of words denoting animals and tools and sentences that were used in the first study (see Chapter 3). For this experiment we used the Highly Predictable and Unpredictable conditions (see Task and Procedure for details), hence, in each experimental session participants were exposed to 18 highly predictable sentences and 18 unpredictable sentences with the animal words, and 18 highly predictable sentences and 18 unpredictable sentences with the tool words.

[10] Given a noticeable drop in power in this experiment, more data needs to be collected and all materials and data will be placed in the online repository after the experiment is finished.

Task and Procedure

Participants were comfortably seated across each other, each having a laptop in front of them. Both participants' electroencephalographic signal was recorded (Figure 1), as well as participants' speech. The task that participants needed to execute was the same as Kerr et al., (in prep., Chapter 4) but now including the prediction manipulation as was done in Kerr et al., (under review, Chapter 3). That is, prior to picture onset, both participants heard either a highly predictable or an unpredictable sentence up to the final word, then Interlocutor A saw an image on the screen (either an animal or a tool) that completed that sentence and needed to name aloud the presented object (i.e., typical object naming, e.g., Levelt, 1989). Interlocutor B heard the word uttered by Interlocutor A and needed to reply with a word that belonged to the same semantic category. On the next trial participants changed roles and Interlocutor B then became the participant who named the presented object, whilst Interlocutor A became the one who replied with an association. In total, the experiment consisted of 216 trials, which means that each stimulus was repeated six times. The trial duration was fixed at 8 seconds. The order of

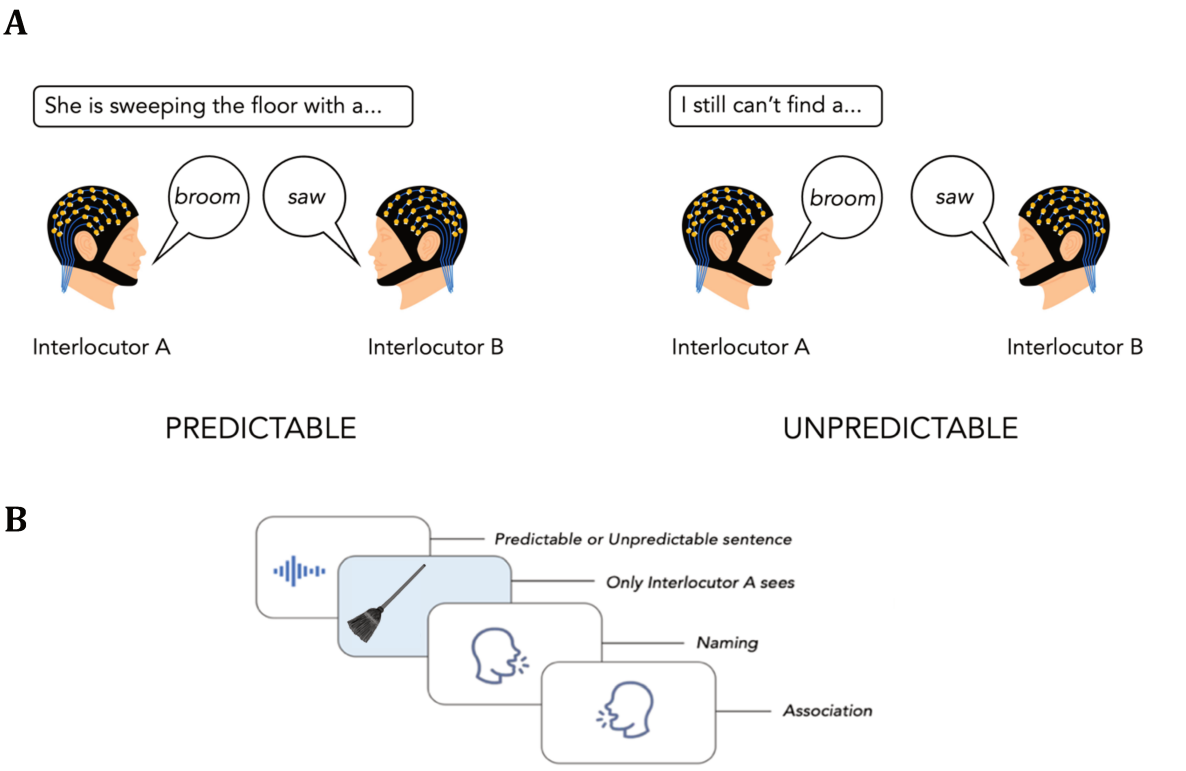


Figure 1. Schematic illustration of the dual-EEG set-up and trial sequence. ***(A)*** *Two participants are seated across each other wearing the EEG caps to record their EEG simultaneously. They are playing the semantic association game where at first both hear either a predictable or an unpredictable sentence, then Interlocutor A sees an image of a word that completes that sentence and names it, and Interlocutor B replies with a semantic association.* ***(B)*** *A trial sequence started by a sentence (predictable or unpredictable), after which Interlocutor A sees a picture and names it (naming), and Interlocutor B replies with a categorically related word (association).*

presentation of the stimuli was completely randomised. During the mounting of the EEG caps, participants were familiarised with the stimuli by seeing each of the pictures with the intended written label below it.

EEG data collection

During the entire session, participants' electrical brain activity was continuously recorded from the scalp at a sampling rate of 2048 Hz using a BioSemi Active Electrodes system and an online bandpass of 0.01-417Hz with two standard 64-electrode caps (the International 10/20 system - American Encephalographic Society; 1994) fitted to both participants and linked to two amplifiers which passed the amplified electrical scalp activity to the EEG server for data storage (see Figure 1). Impedances were kept below 20 kΩ. Stimuli presentations and triggers were synchronized between the different systems using specialized hardware (BioSemi Daisy-Bridge). The triggers were time-locked to each picture onset time. Eye-movements were monitored by EOG electrodes positioned under the left eye of the participants and the prefrontal scalp electrodes, and 2 external electrodes were placed on the left and right mastoids of each of the participants.

EEG Data Preprocessing

Of the 20 dyads that were tested, 2 were excluded from the analyses due to too much noise in the EEG signal (mainly due to articulation and movement potentials) and 1 dyad due to too many mistakes or missing replies, leaving too few remaining trials for analyses after ICA and artefact rejections (details see below). Note that even if only one of the participants in the dyad had too much noise, still both participants, that is the dyad as a whole, were removed from analyses. Hence, 17 dyads (34 participants) in total entered the main analyses pipeline.

The EEG data were preprocessed in the MNE Python environment (Gramfort et al., 2013; v1.7.0). Preprocessing involved the following sequence: (1) removing bad trials; (2) running the Independent Component Analysis (ICA) on the continuous signal to remove artefacts; (3) creating two differently filtered sets, one for the MVPA analysis and one for the interbrain synchrony analysis; (4) creating different epochs and time-windows of interest for analysing the interaction in an event-related like manner. Each of these steps will now be detailed more below.

The recorded signal was split into two recordings corresponding to each participant in the dyad. Trials that contained erroneous replies, no replies either from Interlocutor A or Interlocutor B, too much laughter or unrelated utterances between participants were annotated as 'bad' in the data and were excluded from further analyses, as well as trials where articulation onset was earlier than 200ms either for Interlocutor A or Interlocutor B (Mean of excluded trials = 22.65 (=10.5% of all data) , SD = 18.79, range = 1-61). The 200ms cut-off rate was chosen

to have as little articulation artefacts in the data with as little data loss as possible (Strijkers et al., 2010). This rate is different in the present study than in Kerr et al. (in prep., Chapter 4) where it was 500ms due to faster replies provoked by the prediction manipulation in the current experiment. In a similar way breaks (participants were given 3 breaks per session) were annotated and excluded from the analyses as well. Bad channels were defined (a) during data collection, (b) with an automatic bridged electrode detection algorithm (Tenke & Kayser, 2001), and (c) by visually inspecting the data (Mean of excluded electrodes = 3.9, SD = 3, range = 0-12). The following step was the application of the independent component analysis (ICA) technique for data denoising with the aim to remove noise coming from the oculomotor activity and articulation. To do so the data were filtered at 1-40Hz which is optimal for ICA. The algorithm for implementation of the ICA was FastICA (Hyvarinen & Oja, 2000), and 15 components were computed to be passed to the ICA algorithm. On average, 6.18 components linked to eye movements, speech and movement artefacts were removed (SD = 2.19, range = 2-10. Next, the data were re-filtered at 0.3 – 8 Hz for the Multivariate Pattern Analyses (MVPA) and at 0.3 – 30 Hz for the interbrain synchronisation (IBS) analysis (see section Analyses and Results below). The data were re-referenced using the average of all channels. Next, off-line triggers were created to time-lock the signal and divide the signal related to one trial (i.e., Interlocutor A naming an object and Interlocutor B responding with a semantically related word) in several relevant event-related epochs for analyses. The following logic was implemented: each event (trigger) marked the speech onset and offset times on both interlocutors, except for the first trigger that marked the picture onset time. In total, 5 triggers were computed:

- Picture onset
- Speech onset of Interlocutor A
- Speech offset of Interlocutor A
- Speech onset of Interlocutor B
- Speech offset of Interlocutor B.

Speech onset calculations were done in a similar fashion as in the Behavioural experiment, with the main difference being that here Interlocutor B's responses were measured from the picture onset to determine the actual trigger location, as opposed to Interlocutor A's speech onset. As shown in Fig. 2, this results in three silent intervals, T1, T2 and T3, and two articulation intervals when participants uttered their responses. This was done for two reasons: First, to be able to analyse the verbal interaction between interlocutors in a more fine-grained manner and assess the electrical brain activity linked to (1) object naming (blue interval T1 in Figure 2), (2) associative naming (blue interval T2 in Figure 2), (3) end of the interaction (blue

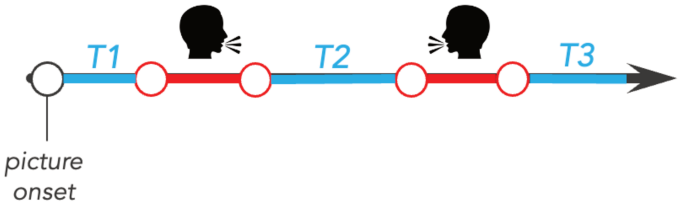


Figure 2. A schematic representation of how each trial was epoched in the analysis pipeline. *The three time windows of interest that entered the main analyses were the time window related to the speech planning by Interlocutor A to name the presented picture (T1), the time window related to the speech planning by Interlocutor B to reply with a semantically related word (T2), and the time window after the reply by Interlocutor B which indexes the end of the interaction (T3). The red areas contained most of the actual articulations (marked according to individual speech onsets and offsets) did not enter the analyses.*

interval T3 in Figure 2); and second, in order to exclude the articulation intervals (the red intervals in Figure 2) from further analyses hereby reducing further any remaining impact of potential speech artefacts after the ICA.

The epochs for analyses were coded according to the experimental conditions (animal vs tool words and predictable vs unpredictable) and their time-window of interest, resulting per trial in three time-windows of interests (T1, T2, T3) of 1000 ms long epochs each which were baseline-corrected using 100ms before the picture onset. This final subset was resampled to 128 Hz. Before running the main analyses (see below) the bad channels were interpolated.

EEG Data Analyses

The first analysis method that was used was Multivariate Pattern Analysis (MVPA) under the Scikit-learn Python environment (Pedregosa et al., 2011). Each participants' EEG signal was processed separately to determine how well the machine learning algorithm can differentiate between the animal and tool categories. The algorithm of choice was the Logistic Regression classifier that uses one-vs-rest scheme for the binary data.

The second method was the HyPyP procedure (Ayrolles et al., 2021) and it investigated the IBS directly. We used the circular correlation coefficient (Ccorr) to determine the covariance synchronisation between the two brain signals as it appears to be particularly well-suited for EEG data and is more robust against spurious hyper-connections than other connectivity measures (Burgess, 2013). We computed Ccorr between speakers and listeners, for all electrode pair combinations (64 X 64 electrodes) in the theta (4-7Hz), alpha (8-13Hz) and the beta (14-30Hz) frequency bands, which were extracted using functions developed in the HyPyP pipeline: FIR filtering and Hilbert transformation to perform frequency decomposition of the EEG signal. This was done for the two meaning-related categories

(animals and tools) and the predictable and unpredictable conditions, as well as grouped by the conditions to assess general speaker-listener IBS.

Statistical Analyses

Behavioural data were analyzed using the Linear Mixed Effects Model (LME) 'predictability' ('predictable' or 'unpredictable'), 'repetition' (from 1 to 6) and 'condition ('animal' or 'tool') as fixed effects and 'dyads' and 'items' as crossed random effects (Baayen et al., 2008; Barr et al., 2013). The models were fitted with lmer (for LME) function from the lme4 package (Bates et al., 2015) in the R statistical computing environment (version 4.2.0, R Core Team, 2022) and the Pymer4 library in Python (Jolly, 2018). Reaction times were log-transformed prior to analysis. We used the maximal random structure model that converged (Barr et al., 2013), and this included by-dyad and by-item random intercepts.

The results from the MVPA and IBS methods mentioned above were tested for significance using permutation clustering approach that includes correction for multiple comparisons. For the MVPA analysis we used a 1-sample permutation cluster test (Maris & Oostenveld, 2007), a non-parametric cluster-level t-test that was implemented through the MNE Python package (Gramfort et al., 2013). For the IBS a cluster-level permutation test from the HyPyP package (Ayrolles et al., 2021) was used, where the synchronisation data from the real dyads was compared to simulated data created via random dyad shuffling (i.e. surrogate dyads; e.g., Bilek et al., 2015).

Results

Prior to describing the results, it is necessary to state that this study is significantly underpowered due to (a) fewer trials per condition compared to Kerr et al. (in prep., Chapter 4); (b) more articulation noise in the EEG signal compared because of faster reaction times (i.e., speech onsets) since at the behavioural level we observe the facilitating prediction effect as in Kerr et al. (under review). Hence, we plan to test more dyads and double the number of participants. Therefore, the results presented below should be taken with caution and treated as preliminary. Below we will mainly focus on general patterns rather than specific data interpretation.

Behavioural Results

Mean reaction times for Interlocutors A and B split by condition (animal and tool) and predictability (predictable and unpredictable) are presented in Tables 1 and 2, and in Figure 3. The main effect of predictability in this study largely replicated the results obtained in Kerr et al. (under review, Chapter 3). We ran two LME models per role to test how predictability

Condition	animal	tool
Predictable	743 (SD = 392.38)	726 (SD = 367.03)
Unpredictable	939 (SD = 454.37)	908 (SD = 399.44)

Table 1. Mean reaction times (and SD) in milliseconds for Interlocutor A per semantic category and predictability factor.

Condition	animal	tool
Predictable	1729 (SD = 984.75)	1857 (SD = 1008.76)
Unpredictable	1956 (SD = 986.45)	2189 (SD = 1023.67)

Table 2. Mean reaction times (and SD) in milliseconds for Interlocutor B per semantic category and predictability factor.

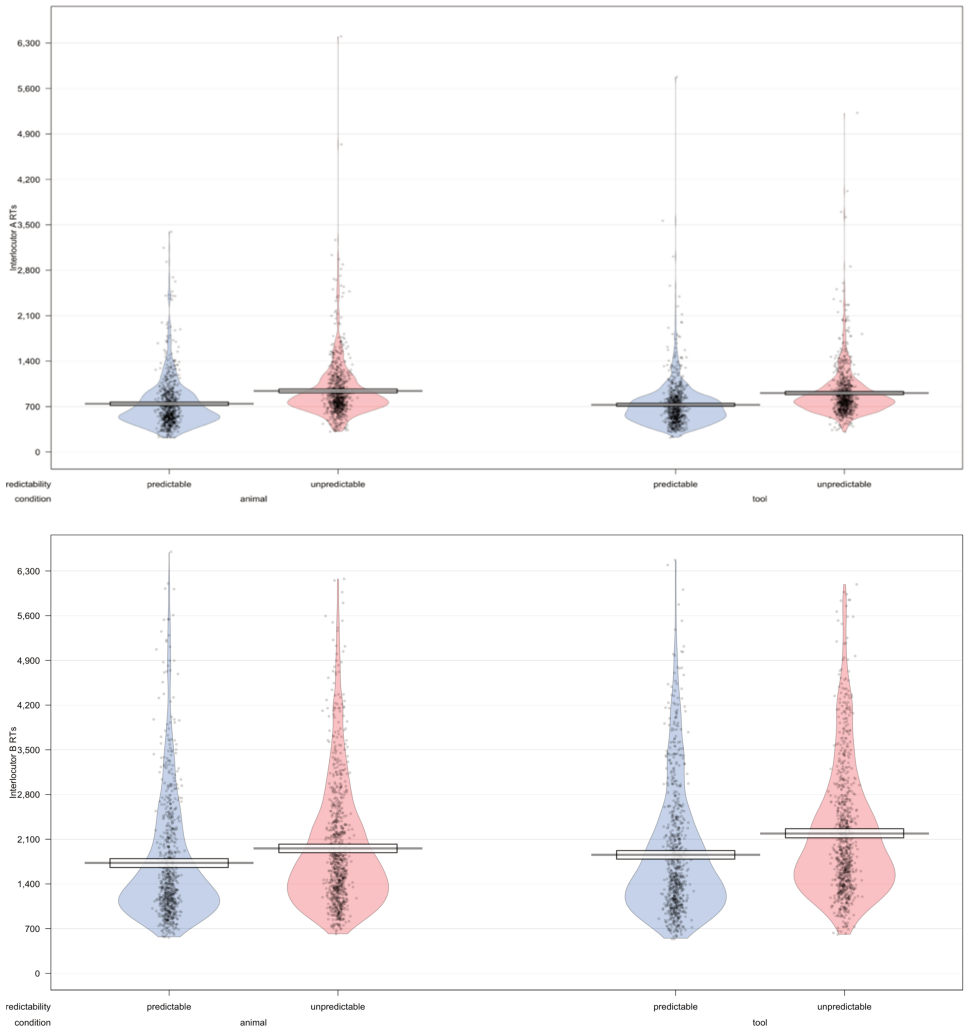


Figure 3. Pirate plots visualising the reaction time distributions *for Interlocutor A (top) and Interlocutor B (bottom) per condition (blue – predictable, red – unpredictable).*

affected each interlocutor separately and one model comparing the reaction times in function of the role. The model for Interlocutor A yielded significant effect of predictability (b = 0.23, SE = 0.06, t = 3.97, p < 0.001). The repetition and condition factors were not significant (p = 0.16 and p = 0.69 respectively). Similar results were obtained for Interlocutor B, where the predictability effect was significant (b = 0.18, SE = 0.05, t = 3.69, p = 0.001). The repetition factor was not significant (p = 0.66), but similarly to Kerr et al. (under review, Chapter 3), the semantic category affected the reaction times (b = 0.09, SE = 0.03, t = 3.1, p < 0.005), where participants replied faster in the animal trials compared to the tool trials.

The comparison of the reaction times in function of role showed that Interlocutor A replied significantly faster than Interlocutor B (b = 1.039, SE = 0.07, t = 16.09, p < 0.001). As in Kerr et al. (under review, Chapter 3), the interaction between Role and Predictability was significant (b = 0.09, SE = 0.02, t = 3.82, p < 0.001), with stronger prediction effects for Interlocutor B (on average: 280 ms) than for Interlocutor A (on average: 189 ms).

MVPA

The results of the MVPA after correcting for multiple comparisons are represented in Figure 4. The pattern observed here is different from what was obtained in previous work (Chapter 4), most probably to the lack of power in the present experiment.

For Interlocutor A in both predictable and unpredictable conditions statistically significant clusters (p<0.05) where the algorithm successfully differentiated between the animal and tool conditions are present in T1, that is during the picture naming phase. The pattern looks more extended in the predictable than in the unpredictable conditions (roughly from 100ms till the end of the epoch for the predictable condition, while only up to 750 ms for the unpredictable condition). However, no significant clusters were formed in T2 and T3 in either of the conditions. Similarly, for Interlocutor B no clusters survived the multiple comparison correction.

Nevertheless, if we look at the clusters in grey in Figure 4 (these are clusters which are significant prior to permutations to correct for multiple comparisons; these results are therefore not reliable or robust at present, but we use them as rough indicative trends as to the possible direction the data may go with more power), we can notice that some of them resemble the patterns from Kerr et al. (in prep., Chapter 4), especially the clusters formed in T2 for Interlocutor B in the predictable condition (from 600ms until the end of the epoch). Also noticeable, is the cluster for Interlocutor B in T1 (between 300 and 500 ms), but only in the predictable condition. And finally, when comparing predictable and unpredictable across all three Ts, it is evident that there are more cluster for the former compared to the latter condition.

Hence, be it very tentative, these patterns might be a sign that prediction indeed affects the meaning-related classification in both interlocutors.

Interbrain Synchrony

When testing the IBS compared to simulated data, and after permutation cluster tests to correct for multiple comparisons ($p<0.05$), we observed significant alpha-connectivity between interlocutors to the task in general (regardless of the meaning manipulation) in T2, both for the predictable and unpredictable conditions (see Figure 5). We also observed significant overall theta-connectivity in T2 (see Figure 6). These patterns thus followed the results from Kerr et al. (in prep., Chapter 4), but the remaining patterns we found were clearly distinct. In T2 for the predictable condition, we observed significant alpha-connectivity for animal- and tool-words, and which seems to be somewhat distinct in its connectivity profile from the alpha-connectivity

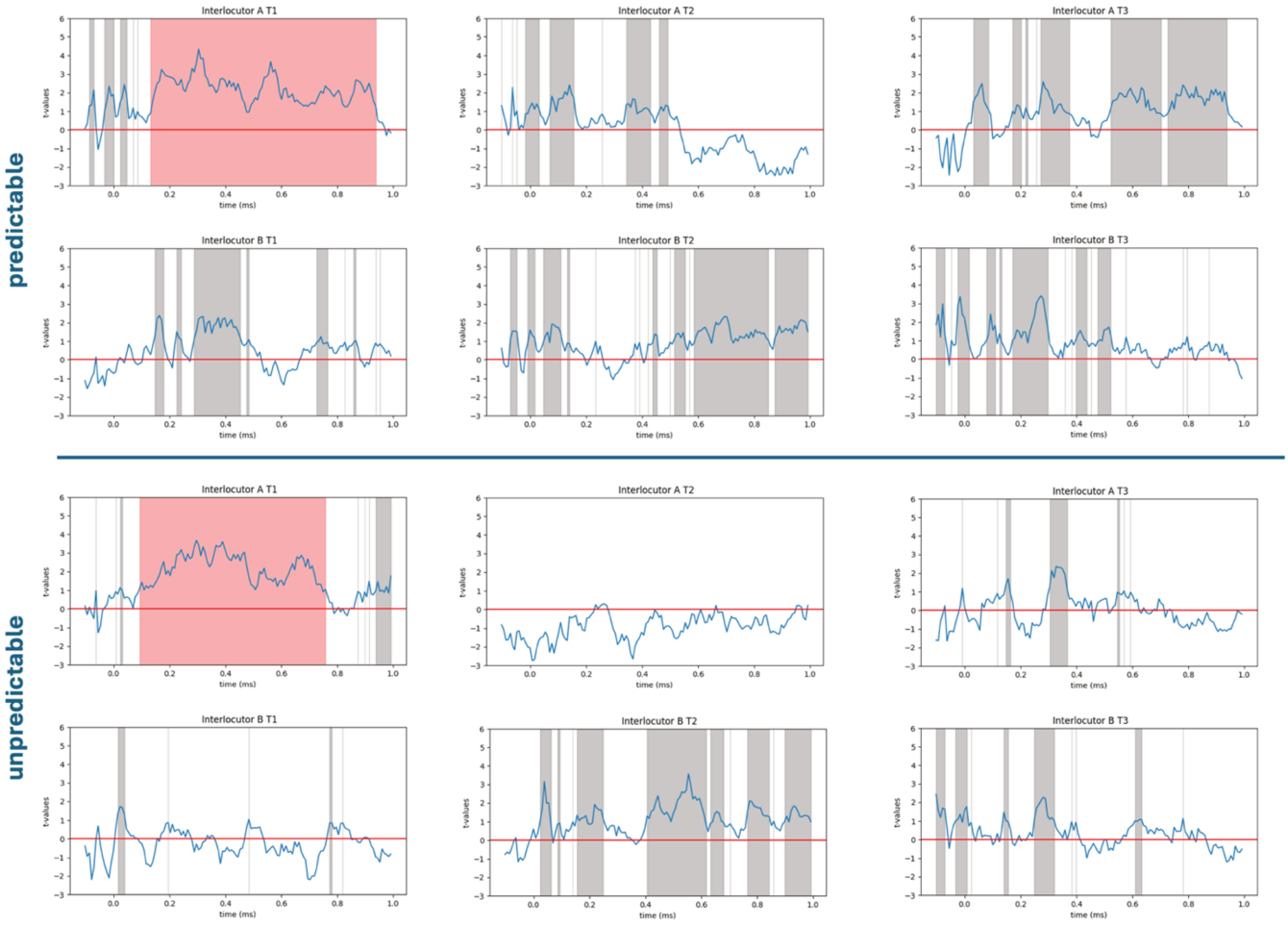


Figure 4. MVPA results for classifying animals and tools in Interlocutor A and Interlocutor B per time window (T1, T2, T3) for the predictable condition (top) and the unpredictable condition (bottom). *The red line represents chance level. The red areas show significant clusters of successful classification between animals and tools, whereas the grey areas show formed clusters that did not survive correction for multiple comparison and white areas the absence of any successful classification clusters.*

to the task in general (see Figure 5). Still for the predictable condition, we also observed significant theta-connectivity to the task in general in T1, and tool-specific theta-connectivity in T3. Finally, in the unpredictable condition we observed animal-specific connectivity in the alpha-band in T2, though that does not seem distinct from the overall alpha-connectivity in T2.

However, as for the MVPA data, particular care and caution must be taken with these IBS results given the low power at present in this study. From a general perspective, at present, the most reliable IBS findings seem to be the overall alpha-connectivity in T2 (which replicated the observations of Kerr et al. (in prep., Chapter 4)), and, similarly to the MVPA results, there are clearly more significant connectivity effects for the predictable compared to the unpredictable conditions. However, the more specific patterns thus far will need to be confirmed and extended by running more dyads to increase power.

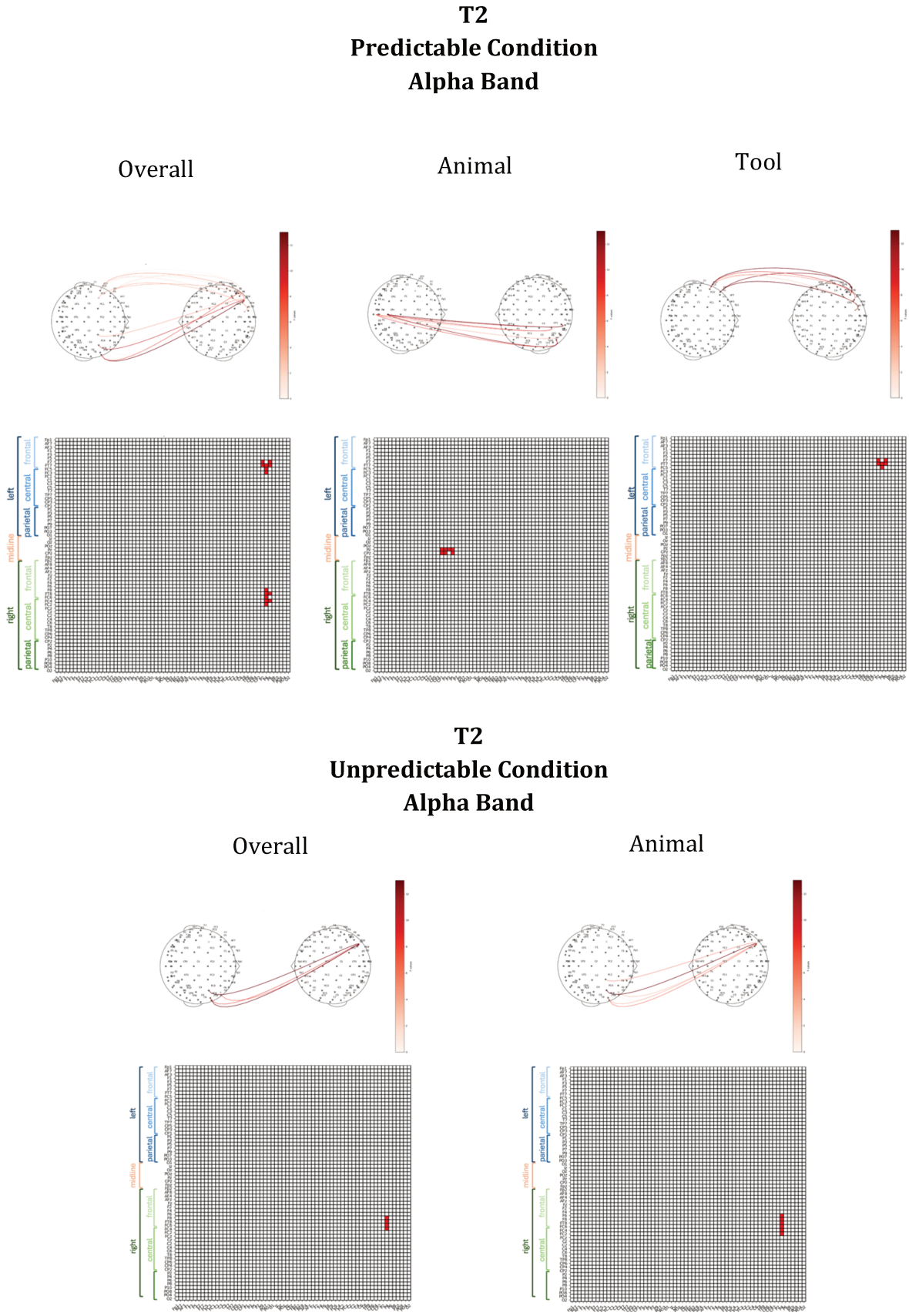


Figure 5. Dual head plots and matrix plots illustrating significant IBS results in T2 for the predictable and unpredictable conditions in the alpha band. *The top left panel represents significant IBS across the semantic categories; the top middle panel shows the results for the animal category; the top right panel shows the results for the tool category. The bottom left shows the results across the semantic categories and the bottom right – the results for the animal category.*

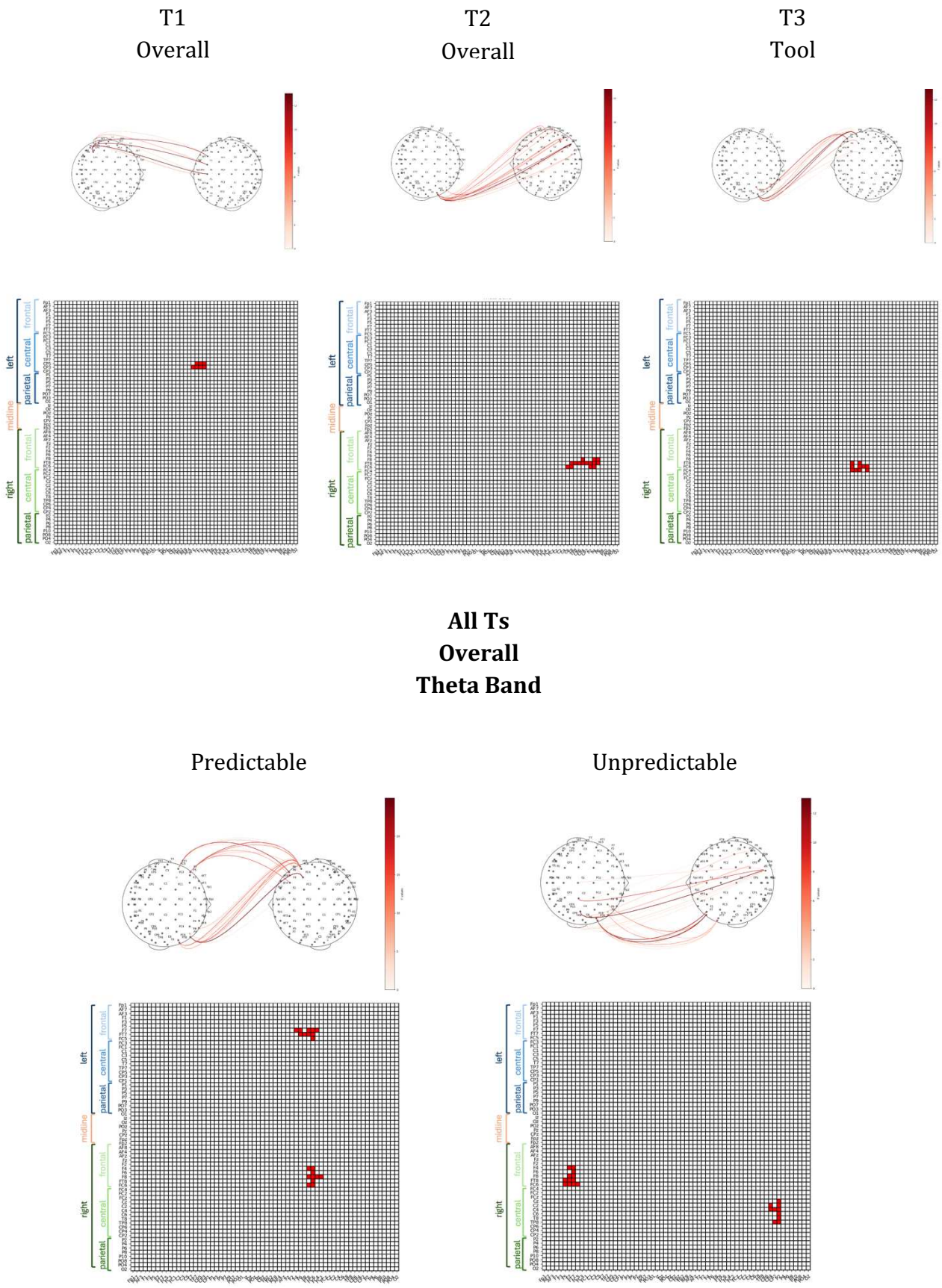


Figure 6. Dual head plots and matrix plots illustrating significant IBS results for the theta band. *The top panel represents significant IBS found in the predictable condition in T1 and T2 grouped by the semantic category, and in T3 for the tool category; the bottom panel shows the results for the unpredictable condition grouped for all time windows and semantic categories in the predictable (left) and unpredictable (right) conditions.*

Discussion

This study was devoted to the investigation whether predictive processing is a major factor in communication through facilitating efficient and fluid conversation. More specifically, our aim was to examine whether enhanced prediction will increase IBS in participants who are engaged in a simple communicative act, and if so, which component: task-general communicative alignment and/or meaning-specific alignment.

In the current study, we replicated the dyadic prediction effect at the behavioural level reported in Kerr et al. (under review, Chapter 3), where both interlocutors in the dyad show strong prediction effects, with the larger effect for the person replying to a conversational partner compared to the person completing the actual predictable sentence. This result confirms again that during verbal interaction we do not only present our own linguistic representations, but also predict those of our partner (e.g., Pickering & Garrod, 2004; 2013; Pickering & Gambi, 2018). Moreover, the fact that the prediction effect is stronger in the interlocutor replying to its partner compared to the interlocutor who simply completes the predictable sentence, demonstrates that language processing in the dyad is a synergetic process (e.g., Brown-Schmidt & Tanenhaus, 2008; Sebanz et al., 2009; Kuhlen & Abdel-Rahman, 2023). This demonstration underscores the need to explore how language, specifically, and cognition, in general, functions in interactive and social contexts, since simply relying on a model of the individualistic research tradition will fall short.

For the MVPA and IBS results however, while some patterns were replicated (specifically, successful animal-tool classification to picture naming and significant alpha-connectivity between speakers and listeners to the task in general during the time-window corresponding to the preparation of Interlocutor B's reply), at present this study lacks sufficient power to reliably assess the meaning-specific neural alignment and the possible impact prediction may have upon it. Rather, at this stage, the most reliable pattern we can observe, is that there are noticeably more significant clusters (MVPA) and connectivity (IBS) between interlocutors when contexts are predictable compared to unpredictable. This observation is promising in its potential to track the effect of prediction on neural alignment in this paradigm, but beyond that our conclusions cannot go any further than tentative suggestions until a reliable amount of power is obtained.

Regarding the IBS, the most reliable finding thus far seems to be the overall alpha-connectivity in T2 for both the predictable and unpredictable conditions (see Figures 5). Not surprisingly, and as in Kerr et al. (in prep., Chapter 4), this is the time window that corresponds to when both participants are most actively involved in the interactional task. Following our interpretation in Kerr et al. (in prep., Chapter 4), as well as previous IBS results (e.g., Klimesh,

1999; Schoot et al., 2016; Perez et al., 2017; 2019; Kelsen et al., 2022), this alpha-connectivity might be reflecting domain-general processes such as attention and verbal working memory, necessary to keep track of the overall communicative goals and context during the interaction. Another, more overall effect which was also present in Kerr et al. (in prep., Chapter 4) concerns the theta-connectivity grouped over all time-windows and variables (see Figure 6). However, while in Kerr et al. (in prep., Chapter 4) we did additionally find animal- and tool-specific theta-connectivity in T2 (and which was distinct from the overall connectivity), in the current experiment we failed to observe such meaning-specific connectivity patterns thus far. The overall theta-connectivity in response to all conditions and time-windows -and which is thus involves more power- might indicate lexico-semantic sensitivity just as in Kerr et al. (in prep., Chapter 4) (e.g., Bastiaansen et al., 2008; Meyer, 2018; Piai & Zheng, 2019). However, given the lack of animal- and tool-specific effects (for now), an alternative possibility is that this general theta-connectivity reflects processing between the speaker and listener (e.g., Luo & Poeppel, 2007; Morillon et al., 2012; Doelling et al., 2014; Meyer, 2018; Boulenger, 2024). It will be very interesting to see when doubling the dyads in this study whether meaning-specific theta-connectivity can be observed again between the interlocutors, and how predictability may shape the effect.

Unlike in our previous work (Kerr et al. (in prep., Chapter 4)), it is difficult to draw any analogy between the MVPA and IBS results obtained here since the MVPA patterns do not represent many significant clusters. However, a few patterns do coincide between the two experiments. Namely, the significant cluster sin T1 for Interlocutor A during the object naming pertain in both studies. This is an expected result since it is driven by the sensory input and therefore much more sensible to the classifier (also in Kerr et al. (in prep., Chapter 4) the object naming response of Interlocutor A gave the strongest classification scores). Nevertheless, it is important to replicate because it suggests that also the predictive version of the paradigm is sensitive enough to pick up neural activity differences and similarities between interlocutors. Therefore, by increasing power we are confident that the MVPA analyses will be able to inform us on the precise temporal effects prediction has on the classification of word meaning between interlocutors in the dyad.

To conclude, I wish to dedicate a few words on the cause of the lack of power since it constitutes a relevant lesson (especially in light of a doctoral thesis) when developing novel paradigms: While trying to maintain all three experiments identical in the number of trials and repetitions per condition to ensure a reasonable duration of the experiment for the participants (note also that the set-up is lengthier in an EEG hyperscanning experiment compared to a traditional EEG experiment), we did not fully consider the impact of our new manipulation, namely, predictability, on data quality. That is, while we were aware that we would have less

trials per meaning condition compared to the experiment from Kerr et al. (in prep., Chapter 4) (given that the semantic condition was now divided by the predictability condition), we did not foresee that we would have more noise in our data. That is, due to the behavioural facilitatory prediction effect, both Interlocutors' A and B responses are faster in this study compared to the previous experiment, which inevitably leads to more articulation-related noise in the EEG signal. Consequently, this factor affects the amount of rejected trials and overall data quality. In hindsight, it makes sense that faster reaction times lead to more articulation artefacts in the data, and we should have a-priori increased the number of trials. Of course, because we want prediction effects in this paradigm, it is perhaps understandable that one only sees the positive side (namely, the observation that our paradigm is highly sensitive for picking up prediction effects as already demonstrated in Kerr et al. (under review, Chapter 3)) and does not reflect upon the possible negative consequences of finding an intended effect. That said, and more importantly, it is not a detrimental issue to our study, and can be remediated by increasing the number of dyads to include in the analyses. From that perspective, these tentative results are a promising preliminary step in examining how prediction affects neural synchrony in pairs of interlocutors.

Chapter 6. General Discussion

In this concluding chapter, I will discuss the key findings of the research presented in this thesis, which investigated the intricate relationship between linguistic alignment, neural alignment, and predictive processes in conversation, in a more integrated manner.

The main work builds upon the well-established theory of interactive alignment, which posits that interlocutors adapt their language at multiple levels, from syntax, phonology and semantics to situation models and communicative goal, in order to facilitate effective communication (Pickering & Garrod, 2004). However, while linguistic alignment has been extensively explored at the behavioural level, the underlying neural mechanisms and their sensitivity to predictive processes remain less understood. This thesis addressed this gap by examining whether interbrain synchrony (IBS) reflects alignment not only at the communicative level but also at more specific linguistic levels, such as word meaning. Put differently, I attempted to uncover whether the alignment of neural activity between interlocutors mirrors linguistic alignment, by unfolding in function of specific linguistic information, that is, distinct levels of processing like semantics, syntax and phonology (Chapter 4).

Furthermore, I explore how predictive processing - the brain's ability to anticipate upcoming linguistic input - modulates both behavioural and neural alignment during verbal interactions. In the context of communication, prediction plays a crucial role in assisting speakers and listeners process and generate language more efficiently. Theories of predictive coding suggest that individuals constantly generate predictions about upcoming language based on prior knowledge and the unfolding context (Friston & Frith, 2015; Schoot et al., 2016). The brain's capacity to predict linguistic structures, meaning, and conversational flow has been argued to support smoother and more efficient dialogue (Pickering & Gambi, 2018). As such, prediction has been shown to be a key factor in successful communication (Chapter 3) and to potentially enhance neural alignment when interlocutors predict each other's utterances (Chapter 5).

Additionally, I outlined specific prerequisites that seem to be necessary for both linguistic and neural alignment to function (Chapter 2). These prerequisites concern the rapid and unified activation of linguistic representations for both speakers and listeners.

In summary, this discussion chapter will synthesize the findings across the experiments reported in this thesis, highlighting the theoretical and empirical contributions to our understanding of how prediction, linguistic alignment, and neural alignment intersect in conversational contexts. I will also address the limitations that were encountered during the

work presented in the previous chapters and outline directions for future research that could further disentangle the neural and cognitive mechanisms underlying dialogue.

Meaning-specific neural alignment

In Chapter 4 of this thesis, I investigated whether neural alignment between interlocutors serves as a general indicator of communicative engagement or if it is sensitive to specific linguistic representations, such as word meaning. This question is essential for linking neural alignment between speakers with linguistic alignment observed in behaviour. Using EEG hyperscanning, pairs of participants engaged in a semantic association game: one participant named objects, animals or tools, while the other responded with words from the same category. This setup enabled us to test whether we can measure neural alignment not only reflecting the task in general (communicative interaction) but also in function of the specific meaning-related component. Our MVPA and IBS analyses demonstrated both general and meaning-specific neural synchrony between interlocutors, highlighting that neural alignment is not just a general phenomenon but that it is also sensitive to more fine-grained linguistic information.

The time-window where the most interesting results regarding meaning-specific neural alignment were observed in this thesis was after the first interlocutor made a naming response and before the second interlocutor uttered a reply (T2). For the MVPA, both interlocutors showed significant classification of animal and tool words, specifically from 600ms after the speech offset of the first interlocutor until right before speech onset of the second interlocutor. For the IBS, it was only in this time-window that meaning-specific theta synchrony was observed between the interlocutors. That specifically this time-window highlights the clearest sensitivity to neural alignment at a more fine-grained linguistic level makes perfect sense since it is the most active and interactive moment of the conversational interaction between the interlocutors. The tool- and animal-specific classification overlap and the meaning-specific synchrony observed at that particular moment in time is a compelling demonstration that neural alignment during verbal interactions does not only reflect general communicative goals or alignment of situation models, but can also display more fine-grained linguistic information, like lexico-semantic information.

In language processing, the theta band activity has been attributed to memory formation and retrieval processes (e.g., Mai & Wang, 2023; Lam et al., 2016; Hermes et al., 2014; Solomon et al., 2017; Buzsaki, 2002; Herweg et al., 2020). For language specifically, theta has been linked to lexico-semantic processing in comprehension (Bastiaansen & Hagoort, 2006; Meyer, 2018) and production (e.g., Piai & Zheng, 2019). Given the manipulation in Chapter 4 concerned a lexico-semantic variable and the main task concerned semantic word association,

interpreting the tool- and animal-specific theta connectivity between speakers and listeners as retrieval of lexico-semantic information seems highly plausible. Nevertheless, the objective of this thesis was not to assess the functional interpretation of different connectivity-profiles, but rather if we could observe 'any' connectivity-profile for word-specific information during language use in interaction. Hence, other interpretations of the theta connectivity we found are possible.

While parsimony favours for now the lexico-semantic interpretation, in terms of future research it may be interesting to briefly assess some alternatives: Oscillations in the theta band have also been associated with increased cognitive load (e.g., Puma et al., 2018) and cognitive control (e.g., Adamczyk & Wyczesany, 2023; Snipes et al., 2022; Ness et al., 2024). That cognitive control plays an essential part in language is evident (e.g., Schnur et al., 2009; Fedorenko & Thompson-Schill, 2014), especially for language production, since we need to continuously inhibit irrelevant information (such as semantic competitors) and facilitate the intended words we wish to utter. Given that in a semantic association task one must retrieve related items and inhibit unrelated information, linking theta activity in our task to cognitive control over lexico-semantic representations is a plausible alternative. Another interesting candidate for theta band activity in language concerns syllable processing (e.g., Luo & Poeppel, 2007; Peelle et al., 2013; Doelling et al., 2014; Meyer, 2018; Boulenger, 2024). But how could that fit with meaning-specific neural alignment observed in this thesis? In this case, one must assume that syllables, which by means of articulation and speech input naturally cause two brains to synchronize, interact or overlap with the meaning features connected to those syllables. That is, by assuming a sequence of syllables interacts with their connected meaning, or even are fully integrated with meaning features as argued by parallel models (e.g., Strijkers, 2016; Strijkers & Costa, 2016; see also later), then the speech syllables would be the carrier signal allowing the brains to synchronize, and that carrier signal is sensitive to lexico-semantic modulations. It will be very interesting for future work to assess the functional interpretation of the meaning-specific neural alignment found in this thesis.

Beyond the key finding of this thesis of meaning-specific neural alignment observed between interlocutors, we also found task-general synchrony between the listeners and speakers, both in the theta and alpha band frequencies. The task-general alpha-connectivity is interesting because it replicates prior studies exploring neural synchrony to interactive communication in general (e.g., Perez et al., 2017; 2019; Ahn et al., 2018). In that work alpha was linked to attentional processing (e.g., Klimesch, 1999; Meyer, 2018), and could reflect the attentional demands to keep the task goals and overall context of the dialogue online to ensure communicative coherence (e.g., Perez et al., 2017; 2019). Given that this alpha-connectivity was present only when considering the entire interaction (grouped over the three time-windows)

and no alpha-connectivity was observed for tools and animals specifically, one possibility is that these different frequencies index qualitatively different alignment stages during language interactions (see also Figure 1 below). While an intriguing and possibly impactful idea, some caution is warranted as well. This idea of having qualitatively different oscillatory dynamics between different levels of alignment concerns an exploratory observation from this thesis's results, but not a direct test that this is indeed the case. Perhaps if conditions to observe neural alignment are enhanced such as with predictive contexts, alpha may be sensitive to more fine-grained levels of linguistic representation as well (see Chapter 5 and the section below). Similarly, while the task-general theta we observed may be a remanent of the meaning-specific neural alignment, another possibility is that the task-general theta-connectivity is (somewhat) functionally distinct from the meaning-specific theta-connectivity (e.g., syllable-rate driven oscillations between articulation (output) and speech perception (input) without being sensitive or modulated by word-level meaning). Clearly, these are very interesting future research flowing out the present thesis.

For now, our results demonstrated that neural alignment is not restricted to the highest-levels of language processing and occurs at a more detailed linguistic level as well, particularly at the word meaning level. This finding highlights the possibility to directly link the phenomenon of linguistic alignment with that of neural alignment, and supports the integration

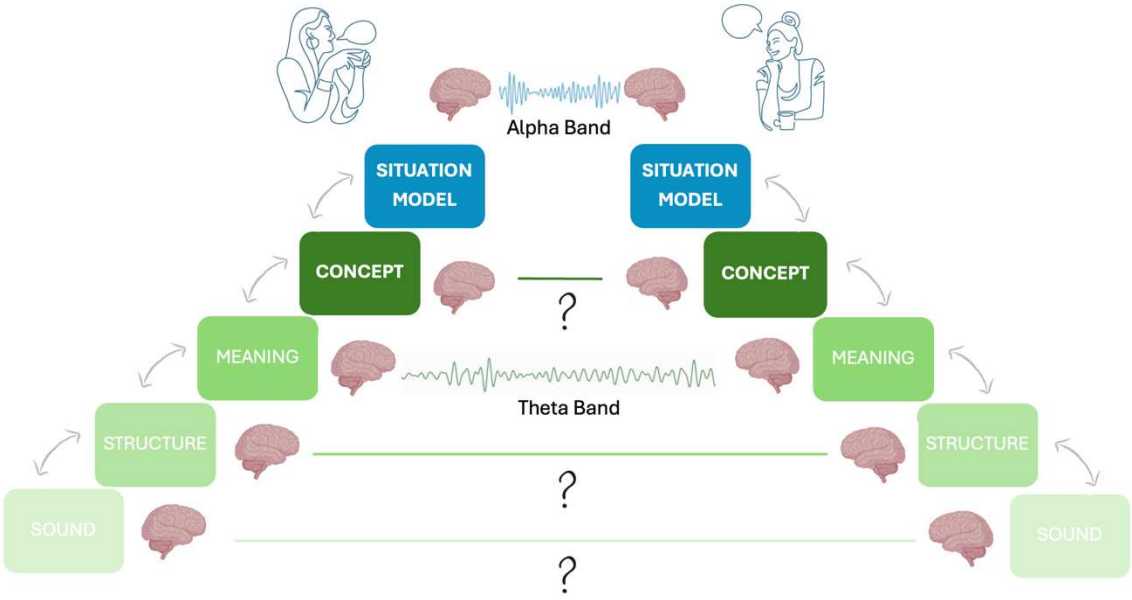


Figure 1. A schematic representation of the roles for the alpha and theta band proposed for the Interactive Alignment Model and neural alignment. *Alpha band connectivity might reflect the overall communicative coherence at higher levels of processing (the blue squares), while theta band connectivity could be involved in the retrieval of specific linguistic components, like lexico-semantics. The connecting lines with question marks concern the levels of linguistic representation not addressed in this thesis, and which therefore represent an interesting area for future research.*

of behavioural models like the Interactive Alignment Model of Pickering & Garrod (2004; 2021) with neurophysiological data (Figure 1). Such integration could advance strongly our understanding of both the cognitive mechanisms underlying human communication and dialogue, as well as their neural implementations to sustain those mechanism. Moreover, if directly linked, any advance in one domain (e.g., behaviour) will signify an advance in the other domain as well (e.g., neuroscience). This thesis thus emphasizes the importance of both the timing and frequency of neural alignment in understanding linguistic alignment. Based on the findings of this thesis, Pickering and Garrod's (2004; 2021) IAM could be extended with neurophysiological implementations as tentatively proposed in Figure 1, where alignment at the level of word meaning may be sustained by theta oscillations and higher-order alignment at the level of situation models by alpha oscillations between conversational partners. Whether this initial and tentative classification holds as well as filling in the other levels of representation (Figure 1) are in my opinion intriguing and important future questions to explore. In this manner, the current thesis may have provided a vibrant new area of research and highlight a way to address the complex interaction of both behavioural and neurobiological factors that can sustain and facilitate human conversation.

Linguistic and neural alignment and prediction

A substantial part of this thesis was devoted to the question of predictive processing during verbal interaction, and how prediction affects human's communicative behaviour and neural alignment between interlocutors.

At the level of behaviour, we observed a strong prediction effect in dyadic interaction (Chapter 3). Specifically, participants engaged in a semantic association game, where the predictability of the upcoming semantic category was manipulated. The findings revealed that when participants could anticipate the meaning of the next utterance, both interlocutors showed significantly faster speech production. The interlocutor who completed the predictable sentence, showed faster naming times, consistent with previous research suggesting that meaning prediction facilitates language comprehension and production (e.g., Altmann & Kamide, 1999; Staub & Clifton, 2006; Ito et al., 2016; Grisoni et al., 2017). More importantly, meaning prediction significantly sped up the replies of the other interlocutor, who benefited even more than the speaker generating the predictable content. This dyadic predictability effect aligns with joint action research, where successful social interactions depend on anticipating the partner's perceptual, motor, and cognitive states (Sebanz et al., 2006; Tanenhaus & Brown-Schmidt, 2008). Similar to other joint actions, language is an inherently social activity (e.g., Tanenhaus & Brown-Schmidt, 2008), and growing evidence indicates that we do not merely activate linguistic representations for our own speech but also simulate those of our

conversation partners to predict their verbal behaviour (e.g., Brennan et al., 2010; Baus et al., 2014; Gambi et al., 2015; Kuhlen & Abdel Rahman, 2017, 2023; Brehm et al., 2019; Cirillo et al., 2022).

Our results extend this body of work by demonstrating that when we predict the semantic category of our partner's upcoming utterance, we can prepare our own responses in advance, resulting in faster speech production and smoother conversational flow. This process, where we engage in predictive simulation of our interlocutor's linguistic representations, resonates with interactive dialogue models that emphasize prediction (e.g., Pickering & Garrod, 2013, 2021; Pickering & Gambi, 2018). In this way, our study provides empirical support for the critical role of prediction in dyadic communication, confirming a key tenet of interactive dialogue models: predictive processing not only facilitates the interlocutor producing the predictable content but also enhances the response of the partner, potentially playing a pivotal role in achieving linguistic alignment between speakers.

At the level of the brain, we aimed aimed to extend these behavioral findings by investigating how enhanced prediction influences neural alignment during simple communicative acts (Chapter 5). The study replicated the behavioural prediction effect from the behavioural experiment, reinforcing the idea that the interlocutor who replies to a conversational partner experiences stronger prediction effects than interlocutor completing the predictable utterance. While the behavioural results strongly support the role of predictive processing in conversation, the neural findings are inconclusive at present due to too low power (generated by an increase in noisy data). Nevertheless, the tentative results we have so far do show that those significant effects which are present predominantly arise from the predictable condition compared to the unpredictable condition. This general observation does suggest that prediction may indeed enhance neural alignment between interlocutors, also possibly at specific linguistic levels of processing. The latter finds support from the observation that there was significant animal- and tool-specific synchrony in the alpha band during the most interactive time-window (T2), but only under predictable circumstances. This observation is both surprising and interesting, since it differs from Chapter 4, where alpha was only present at a task-general level of processing (see also Figure 1). As mentioned in the previous section, this could mean that under circumstances that enhance the possibility to align between interlocutors, such as prediction, alpha band frequencies also fulfil a role in meaning-specific neural alignment. In this case, there may be a more domain-general form of alpha-connectivity allocating attention to maintain the communicative context online, and a more specific form of alpha-connectivity allocating attention to the relevant word categories of the task to successfully engage in semantic associations with the conversational partner.

Of course, before adopting this as a real possibility and incorporating this hypothesis into Figure 1, more clean data is required. As mentioned, due to increased noise levels (see Chapter 5), the reliability of the data set at present is low and several key aspects are left unaddressed: For now we fail to see meaning-specific theta-connectivity, which was the clearest finding in Chapter 4, we did not only expect that prediction would enhance neural synchrony (the only prediction which seems to be fulfilled in the study so far), but also that neural synchronization between interlocutors would emerge much earlier in time, and that both the strength and time-course of neural alignment would be directly related to the strength of the behavioural prediction effects in a given dyad (i.e., dyads with stronger prediction effects in behaviour would show stronger and earlier meaning-specific neural alignment). At present, we simply lack the power to assess whether or not these predictions will bear out, and based on the data we have, I estimate we will need to double the number of dyads to be able to make reliable conclusions.

In sum, the behavioural and neurophysiological results of this thesis clearly highlight that prediction has a strong impact on both linguistic and neural alignment, but to understand the specific role of prediction and its intrinsic relationship to both behaviour and brain during communicative interactions more work is needed. These are interesting future objects which will allow us to extend the model presented in Figure 1 with predictive processes and will help to increase our understanding on the role of prediction in dialogue.

Models of lexical access and meaning-specific neural alignment

Chapter 2 of the present thesis outlined the key theories on lexical access during language production and perception in order to establish which of the theories applies best to the notions of linguistic and neural alignment.

The process of linking conceptual and semantic information with its grammatical and phonological forms has been studied from several perspectives: as a sequential hierarchical system (Levelt, 1999; Levelt & Indefrey, 2004; Indefrey, 2011), as an interactive system with communication between its levels (Dell, 1986; Hickok, 2012; Dell et al., 2013), and as a parallel, non-hierarchical system (Pulvermuller, 1999; 2018; Strijkers, 2016; Strijkers & Costa, 2016).

Pickering and Garrod (2004), in their Interactive Alignment Model, proposed that for dialogue to function properly, there must be a parity in linguistic representations (see also Menenti et al., 2012; Pickering & Garrod, 2014). This means that similar linguistic representations are active during both production and comprehension. That is, there are either "copied" linguistic representations between production and perception (e.g., perceptual phonemes and motor phonemes; e.g., Hickok, 2012), which closely interact and can prime each

other (e.g., Pickering & Garrod, 2014), or there are identical linguistic representations between production and perception (e.g., Menenti et al., 2012; Strijkers, 2016). This is critical for dialogue because it enables quick linguistic alignment, ensuring smooth communication. While classic language theories based on neuropsychological data of brain damaged individuals argued that the mechanisms underlying language production and perception operate independently (e.g., Geschwind, 1979; Price, 2012), as we saw in Chapter 2 of this thesis current theories assume at least some degree of overlap between the language modalities (e.g., Kerr et al., 2023). From the perspective of dialogue and conversation this is important, because language production and comprehension need to be intertwined. A speaker's formulation of an utterance is influenced by what they have just heard, and the listener's interpretation is shaped by the speaker's previous statement. Dialogue, by its nature, is a collaborative process, with mental and neural processes of both participants closely linked. Moreover, since linguistic alignment occurs at various linguistic levels and critically, alignment in production influences alignment in comprehension, it is problematic to consider word retrieval in speakers and listeners as separate, independent systems in dialogue.

Despite most modern theories of lexical access being compatible with this notion that production and perception need to be able to communicate with each other to ensure dialogue, in general, and linguistic alignment, specifically, how much overlap varies strongly between theories. In Chapter 2 I reviewed such more current theories of lexical access and compared them in their degree of spatial and temporal overlap to hypothesize what each model would predict with regard to meaning-specific neural alignment: For serial models (e.g., Levelt et al., 1999; Indefrey & Levelt, 2004), given their distinct and independent temporal dynamics between production and perception, it was proposed that predictive processing would be a necessary factor to observe lexico-semantic neural alignment between listeners and speakers. For interactive models (e.g., Dell, 1986; Hickok, 2012; Dell et al., 2013), some lexico-semantic neural alignment for isolated words may be observed, though restricted to the final steps of processing when interactivity between representational layers occurs. For parallel models (e.g., Pulvermuller, 1999; 2018; Strijkers, 2016; Strijkers & Costa, 2016), strong lexico-semantic neural alignment between speakers and listeners should be observable for words in isolation already since these models assume temporal and spatial overlap of words between production and perception.

Given the data we observed, the parallel model of language production and perception seems best suited to explain the results. We found meaning-specific neural alignment between listeners and speakers without the necessity of having a predictive context (Chapter 4). In a parallel model this happens because word representations are identical in time and space between production (speaker) and perception (listener), meaning that perceiving a word and

producing a word can perfectly overlap; ideal conditions to observe the phenomenon of neural alignment (e.g., Hasson et al., 2012; Schoot et al., 2016). Also, interactive models may explain the meaning-specific neural alignment observed in Chapter 4, though the time-course of the effects does not make it the most parsimonious candidate per se. That is, given that in these models the first activation flow is sequential and interactivity between layers only kicks in later on, strictly speaking T3, the time-window after a reply has been uttered, would have been an expected moment in time to observe overlap between the production and perception due to interactive processing. But the meaning-specific neural alignment was restricted to T2. And also in that time-window, when adding the MVPA results, the time-course of classification overlap seems too early to be readily explained by interactivity (at least in current conceptualizations of the time-course in interactive lexical models). In any event, while interactive models may still capture (or be adapted to capture) the neural alignment patterns found in this thesis, the same cannot be said for serial models. In fact, to make a strong claim, one could argue that serial models of lexical access in general are not optimal to bridge neural alignment with linguistic alignment at a fine-grained level such as words.

In sum, parsimony seems to favour a parallel model of lexical access to explain neural alignment at the word level. Indeed, parallel models seem intrinsically perhaps the best suited to provide the necessary mechanisms for successful alignment and communication in dialogue. This model proposes significant overlap in both time and space for key linguistic components across production and perception, laying the groundwork to investigate the connection between neural alignment and linguistic alignment at the word level, as we have done here, and beyond.

Limitations and future research directions

The work presented in this thesis, while innovative, informative and, I believe, successful in its main objective to start linking the behaviour of alignment with the neurophysiology of alignment, also have limitations of course, and I will address some of them here.

A primary concern, as already highlighted, is the lack of statistical power, which constrains the generalizability and reliability of the findings. In EEG research, having sufficient participants is critical to reliably detect subtle neural effects, especially when investigating cognitive processes in pairs of participants. Both experiments were conducted with the same number of dyads (n = 20) in each experiment. While in the first experiment, and despite having to remove 6 dyads, we did obtain a reliable amount of data (roughly around 60 observations on average per critical condition) thanks to the high number of trials. However, for the second experiment, even though we had a higher number of dyads (17), the amount of data was much lower (roughly around 20 observations per critical condition) and the data we could retain was

less clean as in the first EEG experiment. As discussed in Chapter 5, in the second experiment we tested the same number of trials, but now split by predictability, and moreover the participants' speeded responses due to prediction created more articulation-related noise in the EEG signal. This dictates the necessity to treat the results of the second experiment with caution and as tentative only.

Another point worth mentioning here is the fact that the experiments we conducted are still rather controlled. Although this was partially the goal in order to ensure that the IBS we measured reflected specific levels of language processing (namely, lexico-semantics), it leaves the open question whether the same effects could be observed in a more naturalistic environment. Hence, an important direction for future research would be to create experimental designs that would allow to bridge the controlled findings for low-level processing with more general high-level work on neural alignment.

In our work we focused on the lexico-semantic level of processing, but future work should explore other levels of representations, such as syntax and phonology (see Figure 1). Given that linguistic alignment encompasses all levels, it is logical to assume that they are all reflected in IBS, as least when assuming as concluded in this thesis that neural and linguistic alignment are two sides of the same coin. It is interesting to see in the future whether, for example, the IBS measured at the level of syntax or phonology will replicate the same patterns we observed for lexico-semantics, or whether they are governed by different types of connectivity.

Lastly, in considering the social aspects of neural alignment, several factors beyond mere dyadic interaction could influence the outcomes observed in interactional tasks. While we kept dyad-grouping unconstrained in our study, future research should explore how variables such as gender, social status, and familiarity between participants may affect neural synchrony for language processing. For instance, existing literature suggests that social hierarchies or perceived status can modulate attention and behaviour in group settings, potentially altering how individuals align cognitively and behaviourally during communication (e.g., Sievers et al., 2024). Gender dynamics, too, might influence communication styles and the extent to which participants engage in predictive processing (e.g., Baker et al., 2016). Moreover, the degree of familiarity between participants—whether they are strangers or not—could impact the ease with which they predict each other's utterances and align their neural activity (Bizzego et al., 2020; Long et al., 2021). Prior familiarity often leads to smoother and more efficient exchanges, suggesting stronger IBS, whereas unfamiliarity might slow down alignment due to the cognitive effort required to adjust to a new partner's communicative style. Investigating these social factors would provide a richer understanding of how context and interpersonal dynamics shape neural and behavioural alignment in communication tasks.

To conclude, the research framework presented in the current thesis can be extended beyond neurotypical population to explore whether specific linguistic manipulations and predictability modulations affect linguistic and neural alignment in individuals with such conditions as language disorders and autism spectrum disorders.

Synthesis

This thesis is devoted to the relationship between linguistic alignment, neural alignment, and predictive processes in conversation, emphasizing how these factors contribute to smooth and effective communication.

Equipped with the ability to rapidly activate key linguistic components necessary for language production and perception, interlocutors employ specific mechanisms to engage in successful dialogues.

One such mechanism is neural alignment between interlocutors. It appears that it can reflect more fine-grained linguistic information, such as word meaning, particularly in the theta frequency band. By employing EEG hyperscanning during a semantic association task, we found that brain activity between dyads synchronized not only at a general communicative level but also in response to specific lexical content, suggesting that neural alignment goes beyond broad communicative engagement and can capture detailed, word-level linguistic alignment.

Another mechanism is predictive coding that plays a crucial role in shaping the communicative behaviour of both interlocutors in a dyad by enabling each participant to anticipate and simulate the linguistic actions of the other. The shared mental model created through prediction enhances linguistic alignment between partners, making communication more coordinated and cohesive. Thus, predictive coding is essential for optimizing the dynamic interaction between individuals in dialogue.

Linking these two mechanisms is a tempting endeavour since it seems logical to propose that predictive coding acts as a facilitating factor for neural alignment similarly to what we observe at the behavioural level. Although we presented tentative findings that predictability enhances neural alignment, particularly in the alpha band, these results so far lack in power and stability urging more scrupulous investigation.

In conclusion, this work explores the intricate link prediction, linguistic alignment, and neural alignment in communication. It has the potential to serve as a groundwork for future research into specific neural and cognitive processes that govern our ability to communicate with one another in effective dialogues.

References

Abdel Rahman, R., & Melinger, A. (2009). Semantic context effects in language production: A swinging lexical network proposal and a review. *Language and Cognitive Processes*, 24(5), 713–734.

Adamczyk, A. K., & Wyczesany, M. (2023). Theta-band Connectivity within Cognitive Control Brain Networks Suggests Common Neural Mechanisms for Cognitive and Implicit Emotional Control. *Journal of cognitive neuroscience*, 35(10), 1656–1669. https://doi.org/10.1162/jocn_a_02034

Ahmadi, A., & Tani, J. (2019). A novel predictive-coding-inspired variational RNN model for online prediction and recognition. *Neural computation*, 31(11), 2025-2074.

Ahn, S., Cho, H., Kwon, M., Kim, K., Kwon, H., Kim, B. S., Chang, W. S., Chang, J. W., & Jun, S. C. (2018). Interbrain phase synchronization during turn-taking verbal interaction-a hyperscanning study using simultaneous EEG/MEG. *Human brain mapping*, *39*(1), 171–188. https://doi.org/10.1002/hbm.23834

Alario, F.-X., Segui,J. & Ferrand, L. (2000). Semantic and associative priming in picture naming. *Q. J. Exp. Psychol*. A 53A, 741–764.

Alario, F.X., Ferrand, L., Laganaro, M. et al. (2004). Predictors of picture naming speed. *Behavior Research Methods, Instruments, & Computers* 36, 140–155. https://doi.org/10.3758/BF03195559

Almeida, J., Knobel, M., Finkbeiner, M., and Caramazza, A. (2007). The locus of the frequency effect in picture naming: when recognizing is not enough. *Psychon. Bull. Rev*. 14, 177–1182.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*(3), 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1

Antonucci, S.M., Beeson, P.M., Labiner, D.M. and Rapcsak, S.Z., (2008). Lexical retrieval and semantic knowledge in patients with left inferior temporal lobe lesions. *Aphasiology*, 22(3), 281-304.

Aristei, S., Melinger, A., & Abdel Rahman, R. (2011). Electrophysiological chronometry of semantic context effects in language production. *Journal of Cognitive Neuroscience*, 23(7), 1567–1586.

Ayrolles, A., Brun, F., Chen, P., Djalovski, A., Beauxis, Y., Delorme, R., Bourgeron, T., Dikker, S., & Dumas, G. (2021). HyPyP: a Hyperscanning Python Pipeline for inter-brain connectivity analysis. *Social cognitive and affective neuroscience*, 16(1-2), 72–83. https://doi.org/10.1093/scan/nsaa141

Baars, B. J., Motley, M. T., & MacKay, D. G. (1975). Output editing for lexical status in artificially elicited slips of the tongue. *Journal of Verbal Learning & Verbal Behavior*, 14(4), 382–391.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390-412. https://doi.org/10.1016/j.jml.2007.12.005

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics, 40*(1), 177–189. https://doi.org/10.1016/j.wocn.2011.09.001

Babel, M., & Bulatov, D. (2012). The Role of Fundamental Frequency in Phonetic Accommodation. *Language and Speech*, 55(2), 231-248. https://doi.org/10.1177/0023830911417695

Babiloni, C., Buffo, P., Vecchio, F., Marzano, N., Del Percio, C., Spada, D., Rossi, S., Bruni, I., Rossini, P. M., & Perani, D. (2012). Brains "in concert": frontal oscillatory alpha rhythms and empathy in professional musicians. *NeuroImage*, 60(1), 105–116. https://doi.org/10.1016/j.neuroimage.2011.12.008

Babiloni, F., & Astolfi, L. (2014). Social neuroscience and hyperscanning techniques: past, present and future. *Neuroscience and biobehavioral reviews*, 44, 76–93. https://doi.org/10.1016/j.neubiorev.2012.07.006

Babiloni, F., Cincotti, F., Mattia, D., Mattiocco, M., Fallani, F. D. V., Tocci, A., ... & Astolfi, L. (2006, August). Hypermethods for EEG hyperscanning. *In 2006 International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 3666-3669). IEEE.

Baker, J., Liu, N., Cui, X. et al. Sex differences in neural and behavioral signatures of cooperation revealed by fNIRS hyperscanning. *Scientific Reports* 6, 26492 (2016). https://doi.org/10.1038/srep26492

Baldo, J.V., Klostermann, E.C., Dronkers, N.F. (2008). It's either a cook or a baker: patients with conduction aphasia get the gist but lose the trace. *Brain and Language*, 105(2), 134–140.

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, 42(1), 1–22. https://doi.org/10.1006/jmla.1999.2667

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, 68(3), 10.1016/j.jml.2012.11.001. https://doi.org/10.1016/j.jml.2012.11.001

Bastiaansen, M. C., Oostenveld, R., Jensen, O., & Hagoort, P. (2008). I see what you mean: theta power increases are involved in the retrieval of lexical semantic information. *Brain and language*, 106(1), 15–28. https://doi.org/10.1016/j.bandl.2007.10.006

Bastiaansen, M., & Hagoort, P. (2006). Oscillatory neuronal dynamics during language comprehension. *Progress in brain research*, 159, 179–196. https://doi.org/10.1016/S0079-6123(06)59012-0

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48. https://doi.org/10.18637/jss.v067.i01

Baus, C., Sebanz, N., de la Fuente, V., Branzi, F. M., Martin, C., & Costa, A. (2014). On predicting others' words: electrophysiological evidence of prediction in speech production. *Cognition*, 133(2), 395–407. https://doi.org/10.1016/j.cognition.2014.07.006

Bilek, E., Ruf, M., Schäfer, A., Akdeniz, C., Calhoun, V. D., Schmahl, C., ... & Meyer-Lindenberg, A. (2015). Information flow between interacting human brains: Identification, validation, and relationship to social expertise. *Proceedings of the National Academy of Sciences, 112*(16), 5207-5212. https://doi.org/10.1073/pnas.1421831112

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex, 19*(12), 2767-2796. https://doi.org/10.1093/cercor/bhp055

Binder, J.R. and Desai, R.H., (2011). The neurobiology of semantic memory. *Trends in cognitive sciences*, 15(11), 527-536.

Bizzego, A., Azhari, A., Campostrini, N., Truzzi, A., Ng, L. Y., Gabrieli, G., Bornstein, M. H., Setoh, P., & Esposito, G. (2019). Strangers, Friends, and Lovers Show Different Physiological Synchrony in Different Emotional States. *Behavioral sciences* (Basel, Switzerland), 10(1), 11. https://doi.org/10.3390/bs10010011

Blakemore, S.-J., Frith, C. D. Wolpert, D. M. (2001) The cerebellum is involved in predicting the sensory consequences of action, *Neuroreport*,12(9), 1879-1884.

Bloem, I. & LaHeij, W. (2003). Semantic facilitation and semantic interference in word translation: implications for models of lexical access in language production. *J. Mem. Lang.* 48, 468–488.

Bock, K. (1986). Syntactic persistence in language production. *Cognitive Psychology, 18*(3), 355–387. https://doi.org/10.1016/0010-0285(86)90004-6

Bogels, S., Magyari, L., & Levinson, S. C. (2015). Neural signatures of response planning occur midway through an incoming question in conversation. *Scientific Reports, 5*, 12881. https://doi.org/10.1038/srep12881

Boulenger, V. (2024). Evaluating neural tracking of rhythmic information in speech: Some caveats and challenges. *Rhythms of Speech and Language: Culture, Cognition, and the Brain*.

Braitenberg, V. (1978). Cell assemblies in the cerebral cortex. *In Theoretical approaches to complex systems* (pp. 171-188). Springer, Berlin, Heidelberg.

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition*, 75(2), B13–B25. https://doi.org/10.1016/s0010-0277(99)00081-5

Branigan, H. P., Pickering, M. J., Pearson, J., McLean, J. F., & Brown, A. (2011). The role of beliefs in lexical alignment: evidence from dialogs with humans and computers. *Cognition*, 121(1), 41–57. https://doi.org/10.1016/j.cognition.2011.05.011

Brehm, L., & Goldrick, M. (2016). Empirical and conceptual challenges for neurocognitive theories of language production. *Language, Cognition and Neuroscience*, 31(4), 504-507.

Brehm, L., Taschenberger, L., & Meyer, A. (2019). Mental representations of partner task cause interference in picture naming. *Acta Psychologica*, 199, Article 102888. https://doi.org/10.1016/j.actpsy.2019.102888

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(6), 1482–1493. https://doi.org/10.1037/0278-7393.22.6.1482

Brennan, S. E., Galati, A., & Kuhlen, A. K. (2010). Two minds, one dialog: Coordinating speaking and understanding. *In B. H. Ross (Ed.), The psychology of learning and motivation: Advances in research and theory* (pp. 301–344). Elsevier Academic Press. https://doi.org/10.1016/S0079-7421(10)53008-1

Brown-Schmidt, S., & Tanenhaus, M. K. (2008). Real-time investigation of referential domains in unscripted conversation: A targeted language game approach. *Cognitive Science, 32*(4), 643–684. https://doi.org/10.1080/03640210802066816

Buchsbaum, B., Hickok, G., Humphries, C. (2001). Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cognitive Science*, 25, 663–678.

Buchsbaum, B.R., Baldo, J., Okada, K., Berman, K.F., Dronkers, N., D'Esposito, M. and Hickok, G., (2011). Conduction aphasia, sensory-motor integration, and phonological short-term memory–an aggregate analysis of lesion and fMRI data. *Brain and language*, 119(3), 119-128.

Burgess, A. P. (2013). On the interpretation of synchronization in EEG hyperscanning studies: A cautionary note. *Frontiers in Human Neuroscience, 7*, 881. https://doi.org/10.3389/fnhum.2013.00881

Burnett, T.A., Freedland, M.B., Larson, C.R., Hain, T.C. (1998). Voice F0 responses to manipulations in pitch feedback. *J. Acoust. Soc. Am*., 103(6), 3153–3161.

Buzsaki, G. (2002). Theta oscillations in the hippocampus. *Neuron, 33*(3), 325-340. https://doi.org/10.1016/S0896-6273(02)00586-X

Buzsáki, G. (2010). Neural syntax: cell assemblies, synapsembles, and readers. *Neuron*, 68 : 362-385.

Caramazza, A., (1997). How many levels of processing are there in lexical access?. *Cognitive neuropsychology*, 14(1), 177-208.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology, 76*(6), 893–910. https://doi.org/10.1037/0022-3514.76.6.893

Christoffels, I. K., Firk, C., & Schiller, N.O. (2007). Bilingual language control: An event-related brain potential study. *Brain Research*, 1147, 192–208.

Cirillo, G., Runnqvist, E., Strijkers, K., Nguyen, N., & Baus, C. (2022). Conceptual alignment in a joint picture-naming task performed with a social robot. *Cognition*, 227, 105213. https://doi.org/10.1016/j.cognition.2022.105213

Clark, H. H. (1996). Using language. *Cambridge University Press.* https://doi.org/10.1017/CBO9780511620539

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*(1), 1–39. https://doi.org/10.1016/0010-0277(86)90010-7

Cleland, A.A., & Pickering, M.J. (2003). The use of lexical and syntactic information in language production: Evidence from the priming of noun-phrase structure. *Journal of Memory and Language*, 49, 214-230.

Cloutman, L., Gottesman, R., Chaudhry, P., Davis, C., Kleinman, J.T., Pawlak, M., Herskovits, E.H., Kannan, V., Lee, A., Newhart, M. and Heidler-Gary, J., (2009). Where (in the brain) do semantic errors come from?. *Cortex*, 45(5), 641-649.

Coltheart, M., Rastle, K., Perry, C., Langdon, R. and Ziegler, J., (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological review*, 108(1), 204.

Corps, R. E., & Pickering, M. J. (2023). The role of prediction in dialogue. *Trends in Cognitive Sciences, 27*(6), 468–479. https://doi.org/10.1016/j.tics.2023.03.005

Corps, R. E., Crossley, A., Gambi, C., & Pickering, M. J. (2018). Early preparation during turn-taking: Listeners use content predictions to determine what to say but not when to say it. *Cognition*, 175, 77-95. doi:10.1016/j.cognition.2018.01.015.

Costa, A., Alario, F.-X. & Caramazza, A. (2005). On the categorical nature of the semantic interference effect in the picture–word interference paradigm. *Psychon. Bull. Rev*. 12, 125–131.

Coulston, R., Oviatt, S., & Darves, C. (2002). Amplitude convergence in children's conversational speech with animated personas. *In 7th International Conference on Spoken Language Processing*, ICSLP 2002: Denver; United States; 16 September 2002 through 20 September 2002 (pp. 2689-2692)

Craik, K. J. W. (1943). The nature of explanation. *Cambridge University Press.*

Cutting, J. C., & Ferreira, V. S. (1999). Semantic and phonological information flow in the production lexicon. Journal of experimental psychology. *Learning, memory, and cognition*, 25(2), 318–344.

Dale, R., Kirkham, N. Z., & Richardson, D. C. (2011). The dynamics of reference and shared visual attention. *Frontiers in psychology*, 2, 355. https://doi.org/10.3389/fpsyg.2011.00355

Damasio, H., Grabowski, T.J., Tranel, D., Hichwa, R.D. and Damasio, A.R., (1996). A neural basis for lexical retrieval. *Nature*, 380(6574), 499-505.

Damian, M., & Bowers, J. (2003). Locus of semantic interference in picture-word interference tasks. *Psychonomic Bulletin & Review*, 10, 111-117.

De Ruiter, J. P., Mitterer, H., & Enfield, N. J. (2006). Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language, 82*(3), 515–535. https://doi.org/10.1353/lan.2006.0130

de Zubicaray, G. I., McMahon, K. L., Eastburn, M. M., & Wilson, S. J. (2002). Orthographic/phonological facilitation of naming responses in the picture-word task: an event-related fMRI study using overt vocal responding. *NeuroImage*, 16(4), 1084–1093.

de Zubicaray, G.I., McMahon K.L. & Howard D. (2015). Perfusion fMRI evidence for priming of shared feature-to-lexical connections during cumulative semantic interference in spoken word production. *Language, Cognition and Neuroscience*, 30:3, 261-272.

de Zubicaray, G.I., McMahon, K.L. (2009). Auditory context effects in picture naming investigated with event-related fMRI. *Cognitive, Affective, & Behavioral Neuroscience* 9, 260–269.

DeLeon, J., Gottesman, R.F., Kleinman, J.T., Newhart, M., Davis, C., Heidler-Gary, J., Lee, A. and Hillis, A.E., (2007). Neural regions essential for distinct cognitive processes underlying picture naming. *Brain*, 130(5), 1408-1422.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93(3), 283-321. https://doi.org/10.1037/0033-295X.93.3.283

Dell, G. S., & 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*(1634), 20120394. https://doi.org/10.1098/rstb.2012.0394

Dell, G. S., & Ferreira, V. S. (2016). Thirty years of structural priming: An introduction to the special issue. *Journal of Memory and Language, 91*, 1-4. https://doi.org/10.1016/j.jml.2016.05.005

Dell, G. S., & O'Seaghdha, P. G. (1991). Mediated and Convergent Lexical Priming in Language Production: A Comment on Levelt et al. (1991*). Psychological Review*, 98(4), 604–614.

Dell, G. S., & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition*, 42(1–3), 287–314.

Dell, G. S., & Reich, P. A. (1977). A model of slips of the tongue. *The Third LACUS Forum*, 3, 448-455.

Dell, G. S., Nozari, N., & Oppenheim, G. M. (2014). Word production: Behavioral and computational considerations. *The Oxford handbook of language production*, 88-104.

Dell, G. S., Schwartz, M. F., Nozari, N., Faseyitan, O., & Branch Coslett, H. (2013). Voxel-based lesion-parameter mapping: Identifying the neural correlates of a computational model of word production. *Cognition*, 128(3), 380–396.

Dell, G.S. and Reich, P.A., (1981). Stages in sentence production: An analysis of speech error data. *Journal of verbal learning and verbal behavior*, 20(6), 611-629.

Dell, G.S., Lawler, E.N., Harris, H.D. and Gordon, J.K., (2004). Models of errors of omission in aphasic naming. *Cognitive Neuropsychology*, 21(2-4), 125-145.

Dell, G.S., Martin, N. and Schwartz, M.F., (2007). A case-series test of the interactive two-step model of lexical access: Predicting word repetition from picture naming. *Journal of Memory and Language*, 56(4), 490-520.

Dell, G.S., Schwartz, M.F., Martin, N., Saffran, E.M. and Gagnon, D.A., (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological review*, 104(4), 801.

Dell'Acqua, R., Sessa, P., Peressotti, F., Mulatti, C., Navarrete, E., & Grainger, J. (2010). ERP evidence for ultra-fast semantic processing in the picture-word interference paradigm. *Frontiers in Psychology.*

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

Dikker, S., & Pylkkänen, L. (2013). Predicting language: MEG evidence for lexical preactivation. *Brain and language*, 127(1), 55–64. https://doi.org/10.1016/j.bandl.2012.08.004

Dikker, S., Silbert, L. J., Hasson, U., & Zevin, J. D. (2014). On the same wavelength: Predictable language enhances speaker-listener brain-to-brain synchrony in posterior superior temporal gyrus. *Journal of Neuroscience, 34*(18), 6267–6272. https://doi.org/10.1523/JNEUROSCI.3796-13.2014

Dikker, S., Wan, L., Davidesco, I., Kaggen, L., Oostrik, M., McClintock, J., ... & Poeppel, D. (2017). Brain-to-brain synchrony tracks real-world dynamic group interactions in the classroom. *Current Biology, 27*(9), 1375–1380. https://doi.org/10.1016/j.cub.2017.04.002

Doelling, K. B., Arnal, L. H., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage*, *85*(1), 761-768. https://doi.org/10.1016/j.neuroimage.2013.06.035

Doya, K., Ishii, S., Pouget, A., & Rao, R. P. N. (Eds.). (2007). *Bayesian brain: Probabilistic approaches to neural coding*. MIT Press.

Duffau, H., Gatignol, P., Mandonnet, E., Capelle, L. and Taillandier, L., (2008). Intraoperative subcortical stimulation mapping of language pathways in a consecutive series of 115 patients with Grade II glioma in the left dominant hemisphere. *Journal of neurosurgery*, 109(3), 461-471.

Dumas, G., Nadel, J., Soussignan, R., Martinerie, J., & Garnero, L. (2010). Inter-brain synchronization during social interaction. *PLoS ONE, 5*(8), e12166. https://doi.org/10.1371/journal.pone.0012166

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

Fairbanks, G. (1954). Systematic research in experimental phonetics: 1. A theory of the speech mechanism as a servosystem. *Journal of Speech and Hearing Disorders*, 19, 133-139.

Fairs, A., Fargier, R., & Strijkers, K. (2022). Shared or different: How linked are word production and comprehension? *TIPA. Travaux interdisciplinaires sur la parole et le langage, 38*, https://doi.org/10.4000/tipa.4879

Fairs, A., Michelas, A., Dufour, S., Strijkers, K. (2021). The same ultra-rapid parallel brain dynamics underpin the production and perception of speech. *Cerebral Cortex Communications*, Volume 2, Issue 3.

Fargier, R. Laganaro, M. (2017). Spatio-temporal dynamics of referential and inferential naming: different brain and cognitive operations to lexical selection. *Brain Topography*, 30(2), 182-197.

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*(4), 469–495. https://doi.org/10.1006/jmla.1999.2660

Federmeier, K. D., Wlotko, E. W., Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, 1146, 75–84. https://doi.org/10.1016/j.brainres.2006.06.101

Feng, C., Damian, M.F., Qu, Q. (2021). Parallel Processing of Semantics and Phonology in Spoken Production: Evidence from Blocked Cyclic Picture Naming and EEG. *J Cogn Neurosci.* 1-14.

Ferreira, V. S., & Pashler, H. (2002). Central bottleneck influences on the processing stages of word production. Journal of experimental psychology. *Learning, memory, and cognition*, 28(6), 1187–1199.

Ferreira, V.S. and Griffin, Z.M., (2003). Phonological influences on lexical (mis) selection. *Psychological Science*, 14(1), 86-90.

Fowler, C. A., Brown, J. M., Sabadini, L., & Weihing, J. (2003). Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory and Language, 49*(3), 396-413. https://doi.org/10.1016/S0749-596X(03)00072-X

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 559–586). Lawrence Erlbaum Associates, Inc.

Fridriksson, J., Yourganov, G., Bonilha, L., Basilakos, A., Den Ouden, D.B. and Rorden, C., (2016). Revealing the dual streams of speech processing. *Proceedings of the National Academy of Sciences*, 113(52), 15108-15113.

Fries, P. (2005). A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence. *Trends in Cognitive Sciences*, 9(10), 474–480.

Friston K. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 360(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Friston, K. (2010). The free-energy principle: A unified brain theory? Nature Reviews *Neuroscience*, 11(2), 127–138. https://doi.org/10.1038/nrn2787

Friston, K. J., & Frith, C. D. (2015). Active inference, communication and hermeneutics. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, 68, 129–143. https://doi.org/10.1016/j.cortex.2015.03.025

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017). Active Inference: A Process Theory. *Neural computation*, 29(1), 1–49. https://doi.org/10.1162/NECO_a_00912

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language*, 47, 27–52.

Fromkin, V. A. (1973). Slips of the tongue. *Scientific American*, 229(6), 110–117.

Gallese, V., Fadiga, L., Fogassi, L., Luppino, G., & Murata, A. (1997). A parietal-frontal circuit for hand grasping movements in the monkey: evidence from reversible inactivation experiments. *In P. Thier and H.-O. Karnath (Eds.) Parietal lobe contributions to orientation in 3D-space* (pp. 255-270). Heidelberg: Springer-Verlag.

Gambi, C., Cop, U., & Pickering, M. J. (2015). How do speakers coordinate? Evidence for prediction in a joint word-replacement task. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, 68, 111–128. https://doi.org/10.1016/j.cortex.2014.09.009

Garrett, M. F. (1975). The analysis of sentence production. *In: The psychology of learning and motivation:* Vol. 9, ed. G. H. Bower. Academic Press.

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition, 27*(2), 181–218. https://doi.org/10.1016/0010-0277(87)90018-7

Garrod, S., & Clark, A. (1993). The development of dialogue co-ordination skills in schoolchildren. *Language and Cognitive Processes*, 8(1), 101–126. https://doi.org/10.1080/01690969308406950

Garrod, S., & Doherty, G. (1994). Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition*, 53(3), 181–215. https://doi.org/10.1016/0010-0277(94)90048-5

Gibson, J. J. (1979). The ecological approach to visual perception. *Houghton Mifflin.*

Giroud, J., Lerousseau, J. P., Pellegrino, F., & Morillon, B. (2023). The channel capacity of multilevel linguistic features constrains speech comprehension. *Cognition*, 232, 105345. https://doi.org/10.1016/j.cognition.2022.105345

Glaser, W. R. (1992). Picture naming. *Cognition* 42, 61–105.

Glaser, W. R. & Glaser, M. O. (1989). Context effects in Stroop-like word and picture processing. *Journal of Experimental Psychology: General*, 118:13–42.

Glaser, W. R., & Düngelhoff, F.-J. (1984). The time-course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 640–654.

Goldrick, M. (2011), Linking Speech Errors and Generative Phonological Theory. *Language and Linguistics Compass*, 5, 397-412.

Goldrick, M., & Rapp, B. (2002). A restricted interaction account (RIA) of spoken word production: The best of both worlds. *Aphasiology*, 16(1-2), 20–55.

Goldrick, M., Runnqvist, E., & Costa, A. (2014). Language switching makes pronunciation less nativelike. *Psychological science*, 25(4), 1031-1036.

Golfinopoulos, E., Tourville, J.A., Guenther, F.H. (2010). The integration of largescale neural network modeling and functional brain imaging in speech motor control. *Neuroimage*. 52(3), 862–874.

Goodale, M.A., Milner, A.D., Jakobson, L.S., & Carey, D.P. (1991). A neurological dissociation between perceiving objects and grasping them. *Nature*, 349, 154-156.

Gracco, V. L., Lofqvist, A. (1994). Speech motor coordination and control: evidence from lip, jaw, and laryngeal movements. *Journal of Neuroscience*. 14(11 Pt 1), 6585–6597.

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., ... & Hämäläinen, M. S. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience, 7*, 267. https://doi.org/10.3389/fnins.2013.00267

Graves, W. W., Grabowski, T. J., Mehta, S., & Gordon, J. K. (2007). A neural signature of phonological access: distinguishing the effects of word frequency from familiarity and length in overt picture naming. *Journal of cognitive neuroscience*, 19(4), 617–631.

Graves, W.W., Grabowski, T.J., Mehta, S. and Gupta, P., (2008). The left posterior superior temporal gyrus participates specifically in accessing lexical phonology. *Journal of cognitive neuroscience*, 20(9), 1698-1710.

Grisoni, L., Miller, T. M., & Pulvermüller, F. (2017). Neural correlates of semantic prediction and resolution in sentence processing. *Journal of Neuroscience, 37*(18), 4848–4858. https://doi.org/10.1523/JNEUROSCI.2800-16.2017

Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. *J Commun Disord*. 39(5), 350–365.

Habets, B., Jansma, B.M. & Münte, T.F. (2008). Neurophysiological correlates of linearization in language production. *BMC Neurosci* 9, 77.

Hanulová, J., Davidson, D. J., & Indefrey, P. (2011). Where does the delay in L2 picture naming come from? Neurocognitive evidence on second language word production. *Language and Cognitive Processes*, 26(7), 902–934.

Hari, R., & Kujala, M. V. (2009). Brain basis of human social interaction: From concepts to brain imaging. *Physiological Reviews, 89*(2), 453-479. https://doi.org/10.1152/physrev.00041.2007

Harley, T. A. (1984). A critique of top-down independent levels models of speech production: Evidence from non-plan-internal speech errors. *Cognitive Science*, 8, 191-219.

Harley, T. A. (1993). Phonological activation of semantic competitors during lexical access in speech production. *Language and Cognitive Processes*, 8, 291-309.

Hartsuiker, R. J., & Westenberg, C. (2000). Word order priming in written and spoken sentence production. *Cognition, 75*(1), B27–B39. https://doi.org/10.1016/S0010-0277(99)00080-3

Hartsuiker, R. J., Antón-Méndez, I., Roelstraete, B., & Costa, A. (2006). Spoonish spanerisms: A lexical bias effect in spanish. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(4), 949–953.

Hartsuiker, R. J., Pickering, M. J., & Veltkamp, E. (2004). Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish–English bilinguals. Psychological Science, 15(6), 409–414. https://doi.org/10.1111/j.0956-7976.2004.00693.x

Hartsuiker, R.J., Corley, M. and Martensen, H., (2005). The lexical bias effect is modulated by context, but the standard monitoring account doesn't fly: Related beply to Baars et al.(1975). *Journal of Memory and Language*, 52(1), 58-70.

Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S., & Keysers, C. (2012). Brain-to-brain coupling: A mechanism for creating and sharing a social world. *Trends in Cognitive Sciences, 16*(2), 114–121. https://doi.org/10.1016/j.tics.2011.12.007

Hauk, O., Patterson, K., Woollams, A., Cooper-Pye, E., Pulvermuller, F. & Rogers, T. T. (2007). How the camel lost its hump: the impact of object typicality on event-related potential signals in object decision. *J. Cogn. Neurosci*. 19, 1338–1353.

Hebb, D. O. (1949). The organization of behavior; a neuropsychological theory. *Psychology Press.*

Heidlmayr, K., Weber, K., Takashima, A., & Hagoort, P. (2020). No title, no theme: The joined neural space between speakers and listeners during production and comprehension of multi-sentence discourse. *Cortex*, 130, 111-126. doi:10.1016/j.cortex.2020.04.035.

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.

Helmholtz von, H. (1962). Helmholtz's treatise on physiological optics (J. P. C. Southall, Trans. and Ed.). New York, NY: *Dover Publication* (Original work published 1866).

Hermes, D., Miller, K. J., Vansteensel, M. J., Edwards, E., Ferrier, C. H., Bleichner, M. G., ... & Ramsey, N. F. (2014). Cortical theta wanes for language. *Neuroimage*, *85*, 738-748.

Herweg, N. A., Solomon, E. A., & Kahana, M. J. (2020). Theta oscillations in human memory. *Trends in Cognitive Sciences, 24*(3), 208-227. https://doi.org/10.1016/j.tics.2019.12.006

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

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, 13(2), 135–145.

Hickok, G. (2014). The architecture of speech production and the role of the phoneme in speech processing. *Language and Cognitive Processes*, 29, 2–20.

Hickok, G. and Poeppel, D., (2000). Towards a functional neuroanatomy of speech perception. *Trends in cognitive* sciences, 4(4), 131-138.

Hickok, G. and Poeppel, D., (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, 92(1-2), 67-99.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5), 393–402.

Hickok, G., Okada, K., Serences, J.T. (2009). Area Spt in the human planum temporale supports sensory-motor integration for speech processing. *Journal of Neurophysiology*, 101(5), 2725–2732.

Houde, J. F., & Nagarajan, S. S. (2011). Speech production as state feedback control. *Frontiers in human neuroscience*, 5, 82.

Houde, J. F., Nagarajan, S. S., Sekihara, K., & Merzenich, M. M. (2002). Modulation of the auditory cortex during speech: an MEG study. *Journal of cognitive neuroscience*, 14(8), 1125–1138.

Houde, J.F., Jordan, M.I. (1998). Sensorimotor adaptation in speech production. *Science*, 279, 1213–1216.

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research, 1626*, 118–135. https://doi.org/10.1016/j.brainres.2015.02.014

Huettig, F., & Mani, N. (2016). Is prediction necessary to understand language? Probably not. Language, Cognition and Neuroscience, 31(1), 19–31. https://doi.org/10.1080/23273798.2015.1072223

Hultén, A., Vihla, M., Laine, M. and Salmelin, R. (2009), Accessing newly learned names and meanings in the native language. *Hum. Brain Mapp*., 30, 976-989.

Humphreys, K.R., (2002). Lexical bias in speech errors. *University of Illinois at Urbana-Champaign.*

Hyvärinen, A., & Oja, E. (2000). Independent component analysis: Algorithms and applications. *Neural Networks, 13*(4-5), 411-430. https://doi.org/10.1016/S0893-6080(00)00026-5

Indefrey, P. (2011). The spatial and temporal signatures of word production components: a critical update. *Front. Psychol*. 2, 255.

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition, 92*(1-2), 101-144. https://doi.org/10.1016/j.cognition.2002.06.001

Ito, A., Corley, M., Pickering, M. J., Martin, A. E., & Nieuwland, M. S. (2016). Predicting form and meaning: Evidence from brain potentials. *Journal of Memory and Language*, 86, 157–171. https://doi.org/10.1016/j.jml.2015.10.007

Ito, A., Pickering, M. J., & Corley, M. (2018). Investigating the time-course of phonological prediction in native and non-native speakers of English: A visual world eye-tracking study. Journal of Memory and Language, 98, 1-11. https://doi.org/10.1016/j.jml.2017.09.002

Ito, M. (2008). Control of mental activities by internal models in the cerebellum. *Nat Rev Neurosci* 9, 304–313.

Janssen, N., Schirm, W., Mahon, B. Z. & Caramazza, A. (2008). The semantic interference effect in the picture-word interference paradigm: evidence for the response selection hypothesis. *J. Exp. Psychol. Learn. Mem. Cogn*. 34, 249–256.

Jeannerod, M., Arbib, M. A., Rizzolatti, G., & Sakata, H. (1995). Grasping objects: the cortical mechanisms of visuomotor transformation. *Trends in neurosciences*, 18(7), 314–320.

Jefferies, E., Crisp, J. and Ralph, M.A.L., (2006). The impact of phonological or semantic impairment on delayed auditory repetition: Evidence from stroke aphasia and semantic dementia. *Aphasiology*, 20(9), 963-992.

Jescheniak, J. D. & Levelt, W. J. M. (1994). Word frequency effects in speech production: retrieval of syntactic information and of phonological form. *J. Exp. Psychol. Learn. Mem. Cogn. 20*, 824–843.

Jescheniak, J. D. & Schriefers, H. (1998). Discrete serial versus cascaded processing in lexical access in speech production: Further evidence from the co-activation of near-synonyms. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24:1256–74.

Jescheniak, J. D., Hahne, A., & Schriefers, H. (2003). Information flow in the mental lexicon during speech planning: evidence from event-related brain potentials. *Brain research. Cognitive brain research*, 15(3), 261–276.

Johnson, J. S. & Olshausen, B. A. (2005). The earliest EEG signatures of object recognition in a cued-target task are postsensory. *J. Vis*. 5, 299–312.

Jolly, E. (2018). Pymer4: Connecting R and Python for Linear Mixed Modeling. *Journal of Open Source Software, 3*(31), 862. https://doi.org/10.21105/joss.00862

Kawashima, R., Naitoh, E., Matsumura, M., Itoh, H., Ono, S., Satoh, K., Gotoh, R., Koyama, M., Inoue, K., Yoshioka, S. & Fukuda, H. (1996). Topographic representation in human intraparietal sulcus of reaching and saccade. *Neuroreport*, 7, 1253-1256.

Kelsen, B. A., Sumich, A., Kasabov, N., Liang, S. H. Y., & Wang, G. Y. (2022). What has social neuroscience learned from hyperscanning studies of spoken communication? A systematic review. Neuroscience and biobehavioral reviews, 132, 1249–1262. https://doi.org/10.1016/j.neubiorev.2020.09.008

Kemmerer D. (2014). Word classes in the brain: implications of linguistic typology for cognitive neuroscience. *Cortex; a journal devoted to the study of the nervous system and behavior*, 58, 27–51. https://doi.org/10.1016/j.cortex.2014.05.004

Kerr, E., Ivanova, B., & Strijkers, K. (2023). Lexical access in speech production: Psycho-and neurolinguistic perspectives on the spatiotemporal dynamics. *In Language Production* (pp. 32-65). Routledge.

Kiefer, M., & Pulvermüller, F. (2012). Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. *Cortex*, 48(7), 805-825. https://doi.org/10.1016/j.cortex.2011.04.006

King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P. R. (2005). Getting to know you: Reputation and trust in a two-person economic exchange. *Science, 308*(5718), 78–83. https://doi.org/10.1126/science.1108062

King, J.-R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: The temporal generalization method. *Trends in Cognitive Sciences, 18*(4), 203-210. https://doi.org/10.1016/j.tics.2014.01.002

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. *Brain Research Reviews, 29*(2-3), 169-195. https://doi.org/10.1016/S0165-0173(98)00056-3

Koester, D., & Schiller, N. O. (2008). Morphological priming in overt language production: electrophysiological evidence from Dutch. *NeuroImage*, 42(4), 1622–1630.

Koike, T., Tanabe, H. C., & Sadato, N. (2015). Hyperscanning neuroimaging technique to reveal the “two-in-one” system in social interactions. *Neuroscience Research, 90*, 25–32. https://doi.org/10.1016/j.neures.2014.11.006

Kuhlen, A. K., & Abdel Rahman, R. (2017). Having a task partner affects lexical retrieval: Spoken word production in shared task settings. *Cognition*, 166, 94–106. https://doi.org/10.1016/j.cognition.2017.05.024

Kuhlen, A. K., & Abdel Rahman, R. (2023). Beyond speaking: neurocognitive perspectives on language production in social interaction. *Philosophical Transactions of the Royal Society* B, 378(1875), 20210483.

Kuhlen, A. K., Allefeld, C., & Haynes, J. D. (2012). Content-specific coordination of listeners' to speakers' EEG during communication. *Frontiers in human neuroscience*, 6, 266.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience, 31*(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62*(1), 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*(4427), 203–205. https://doi.org/10.1126/science.7350657

Kutas, M., DeLong, K. A., & Smith, N. J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. *In M. Bar (Ed.), Predictions in the brain: Using our past to generate a future* (pp. 190–207). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195395518.003.0065

Laganaro, M., Morand, S., Schwitter, V., Zimmermann, C., Camen, C., & Schnider, A. (2009). Electrophysiological correlates of different anomic patterns in comparison with normal word production. *Cortex*, 45(6), 697–707.

Laganaro, M., Valente, A., & Perret, C. (2012). Time course of word production in fast and slow speakers: a high density ERP topographic study. *NeuroImage*, 59(4), 3881–3888.

Lane, L. W., & Ferreira, V. S. (2008). Speaker-external versus speaker-internal forces on utterance form: do cognitive demands override threats to referential success?. *Journal of experimental psychology. Learning, memory, and cognition,* 34(6), 1466–1481. https://doi.org/10.1037/a0013353

Lane, L. W., Groisman, M., & Ferreira, V. S. (2006). Don't talk about pink elephants! Speaker's control over leaking private information during language production. *Psychological science*, 17(4), 273–277. https://doi.org/10.1111/j.1467-9280.2006.01697.x

Lelonkiewicz, J. R., Rabagliati, H., & Pickering, M. J. (2021). The role of language production in making predictions during comprehension. *Quarterly journal of experimental psychology* (2006), 74(12), 2193–2209. https://doi.org/10.1177/17470218211028438

Levelt W. J. (2001). Spoken word production: a theory of lexical access. *Proceedings of the National Academy of Sciences of the United States of America,* 98(23), 13464–13471.

Levelt, W. J. M. (1989). Speaking: From intention to articulation. *The MIT Press*.

Levelt, W. J. M. (1999). Producing spoken language: A blueprint of the speaker. *In C. M. Brown, & P. Hagoort (Eds.), The neurocognition of language (pp. 83-122)*. Oxford University Press.

Levelt, W. J. M., & Kelter, S. (1982). Surface form and memory in question answering. *Cognitive Psychology, 14*(1), 78–106. https://doi.org/10.1016/0010-0285(82)90005-6

Levelt, W. J. M., Praamstra, P., Meyer, A. S., Helenius, P., & Salmelin, R. (1998). An MEG study of picture naming. *Journal of Cognitive Neuroscience*, 10(5), 553–567.

Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T. & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review*, 98, 122–42.

Levelt, W.J., Roelofs, A. and Meyer, A.S., (1999). A theory of lexical access in speech production. *Behavioral and brain sciences*, 22(1), 1-38.

Levinson, S. C. (2016). Turn-taking in human communication—Origins and implications for language processing. *Trends in Cognitive Sciences, 20*(1), 6–14. https://doi.org/10.1016/j.tics.2015.10.010

Lichtheim, L. (1885). On aphasia. *Brain*, 7, 433–484.

Lindenberger, U., Li, S.-C., Gruber, W., & Müller, V. (2009). Brains swinging in concert: Cortical phase synchronization while playing guitar. *BMC Neuroscience, 10*, 22. https://doi.org/10.1186/1471-2202-10-22

Liu, L., Zhang, Y., Zhou, Q., Garrett, D. D., Lu, C., Chen, A., ... & Ding, G. (2020). Auditory–articulatory neural alignment between listener and speaker during verbal communication. *Cerebral Cortex*, 30(3), 942- 951.

Liu, W., Branigan, H. P., Zheng, L., Long, Y., Bai, X., Li, K., Zhao, H., Zhou, S., Pickering, M. J., & Lu, C. (2019). Shared neural representations of syntax during online dyadic communication. *NeuroImage*, 198, 63–72. https://doi.org/10.1016/j.neuroimage.2019.05.035

Long, Y., Zheng, L., Zhao, H., Zhou, S., Zhai, Y., & Lu, C. (2021). Interpersonal Neural Synchronization during Interpersonal Touch Underlies Affiliative Pair Bonding between Romantic Couples. Cerebral cortex (New York, N.Y. : 1991), 31(3), 1647–1659. https://doi.org/10.1093/cercor/bhaa316

Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, *54*(6), 1001-1010. https://doi.org/10.1016/j.neuron.2007.06.004

Lupker, S. J. (1982). The role of phonetic and orthographic similarity in picture-word interference. *Canadian Journal of Psychology*, 36, 349–367.

Maess, B., Friederici, A. D., Damian, M., Meyer, A. S., & Levelt, W. J. M. (2002). Semantic category interference in overt picture naming: Sharpening current density localization by PCA. *Journal of Cognitive Neuroscience*, 14(3), 455–462.

Mägiste, E. (1984). Stroop tasks and dichotic translation: The development of interference patterns in bilinguals. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 10, 304-315.

Magyari, L., Bastiaansen, M., de Ruiter, J. P., & Levinson, S. C. (2014). Early anticipation lies behind the speed of response in conversation. *Journal of Cognitive Neuroscience, 26*(11), 2530–2539. https://doi.org/10.1162/jocn_a_00673

Mai, G., & Wang, W. S.-Y. (2023). Distinct roles of delta- and theta-band neural tracking for sharpening and predictive coding of multi-level speech features during spoken language processing. *Human Brain Mapping*, 44(17), 6149–6172. https://doi.org/10.1002/hbm.26503

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164*(1), 177-190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology, 58*, 25-45. https://doi.org/10.1146/annurev.psych.57.102904.190143

Martin, C.D., Branzi, F.M. & Bar, M. (2018). Prediction is Production: The missing link between language production and comprehension. *Scientific Reports* 8, 1079. https://doi.org/10.1038/s41598-018-19499-4

Martin, N., (1996). Phonological facilitation of semantic errors in normal and aphasic speakers. *Language and Cognitive Processes*, 11(3), 257-282.

Menenti, L., Gierhan, S. M. E., Segaert, K., & Hagoort, P. (2012). Shared language: Overlap and segregation of the neuronal infrastructure for speaking and listening revealed by functional MRI. *Psychological Science, 23*(3), 269–276. https://doi.org/10.1177/0956797611418347

Meyer, A. S. (1992). Investigation of phonological encoding through speech error analyses: Achievements, limitations, and alternatives. *Cognition*, 42, 181–211.

Meyer, A. S., & Schriefers, H. (1991). Phonological facilitation in picture-word interference experiments: Effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(6), 1146–1160.

Meyer, A.S., Roelofs, A. & Brehm, L. (2019). Thirty years of Speaking: An introduction to the Special Issue, *Language, Cognition and Neuroscience*, 34:9, 1073-1084.

Meyer, L. (2018). The neural oscillations of speech processing and language comprehension: State of the art and emerging mechanisms. *European Journal of Neuroscience*, 48(7), 2609-2621. https://doi.org/10.1111/ejn.13748

Meyer, L., & Friederici, A. D. (2016). Neural systems underlying the processing of complex sentences. *Neurobiology of language*, 597-606. https://doi.org/10.1016/B978-0-12-407794-2.00048-1

Milner, A. D. & Goodale, M. A. (1995). The visual brain in action. *Oxford Univ. Press,* Oxford.

Milner, A.D. & Goodale, M.A. (1993). Visual pathways to perception and action. *In T.P. Hicks, S. Molotchnikoff & T. Ono (Eds.) Progress in Brain Research*, Vol. 95 (pp. 317-337). Amsterdam: Elsevier.

Miozzo, M., Pulvermüller, F., Hauk, O. (2015). Early parallel activation of semantics and phonology in picture naming: Evidence from a multiple linear regression MEG study. Cereb Cortex. 25, 3343-3355.

Montague, P. R., Berns, G. S., Cohen, J. D., McClure, S. M., Pagnoni, G., Dhamala, M., ... & Fisher, R. E. (2002). Hyperscanning: Simultaneous fMRI during linked social interactions. *NeuroImage, 16*(4), 1159–1164. https://doi.org/10.1006/nimg.2002.1150

Morillon, B., Liégeois-Chauvel, C., Arnal, L. H., Bénar, C. G., & Giraud, A. L. (2012). Asymmetric function of theta and gamma activity in syllable processing: an intra-cortical study. *Frontiers in psychology*, 3, 248. https://doi.org/10.3389/fpsyg.2012.00248

Motley, M. T., Camden, C. T., & Baars, B. J. Covert formulation and editing of anomalies in speech production: Evidence from experimentally elicited slips of the tongue. Journal of Verbal Learning and Verbal Behavior, 21(5), 578–594.

Munding, D., Dubarry, A.-S., & Alario, F.-X. (2016). On the cortical dynamics of word production: A review of the MEG evidence. Language, Cognition and Neuroscience, 31(4), 441–462.

Navarrete, E., & Costa, A. (2005). Phonological activation of ignored pictures: Further evidence for a cascade model of lexical access. Journal of Memory and Language, 53(3), 359–377.

Nenkova, A., Gravano, A., & Hirschberg, J. (2008). High frequency word entrainment in spoken dialogue. *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies*, 169-172.

Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics*, 39, 132-142.

Nieuwland, M. S., et al. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *eLife*, 7, e33468. https://doi.org/10.7554/eLife.33468

Nooteboom, S. G. (2005). Lexical bias revisited: Detecting, rejecting and repairing speech errors in inner speech. *Speech Communication*, 47(1-2), 43–58.

Nozari, N. and Dell, G.S., (2009). More on lexical bias: How efficient can a "lexical editor" be*?*. *Journal of Memory and language*, 60(2), 291-307.

Nozari, N., & Pinet, S. (2020). A critical review of the behavioral, neuroimaging, and electrophysiological studies of co-activation of representations during word production. *Journal of Neurolinguistics*, 53, Article 100875.

Nozari, N., Kittredge, A. K., Dell, G. S., & Schwartz, M. F. (2010). Naming and repetition in aphasia: Steps, routes, and frequency effects. *Journal of memory and language*, 63(4), 541–559.

Otten, M., & Van Berkum, J. J. A. (2008). Discourse-based word anticipation during language processing: Prediction or priming? *Discourse Processes*, 45(6), 464–496. https://doi.org/10.1080/01638530802356463

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America, 119*(4), 2382-2393. https://doi.org/10.1121/1.2178720

Pardo, J.S., Urmanche, A., Wilman, S. et al. Phonetic convergence across multiple measures and model talkers. *Attention, Perception, & Psychophysics* 79, 637–659 (2017). https://doi.org/10.3758/s13414-016-1226-0

Parr, T., Pezzulo, G., & Friston, K. J. (2022). Active Inference: The Free Energy Principle. *In Mind, Brain, and Behavior*. MIT Press.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825-2830. https://jmlr.org/papers/v12/pedregosa11a.html

Perez, A., Carreiras, M., & Duñabeitia, J. A. (2017). Brain-to-brain entrainment: EEG interbrain synchronization while speaking and listening. *Scientific Reports, 7*, 4190. https://doi.org/10.1038/s41598-017-04464-4

Pérez, A., Dumas, G., Karadag, M., & Duñabeitia, J. A. (2019). Differential brain-to-brain entrainment while speaking and listening in native and foreign languages. *Cortex; a journal devoted to the study of the nervous system and behavior,* 111, 303–315. https://doi.org/10.1016/j.cortex.2018.11.026

Perkell, J.S. (2012). Movement goals and feedback and feedforward control mechanisms in speech production. Journal of Neurolinguistics, 25(5), 382–407.

Pezzulo, G., Parr, T., & Friston, K. (2024). Active inference as a theory of sentient behavior. *Biological Psychology*, 186, 1–9. https://doi.org/10.1016/j.biopsycho.2023.108741

Piai V., Roelofs A., Jensen O., Schoffelen J.M., & Bonnefond M. (2014). Distinct Patterns of Brain Activity Characterise Lexical Activation and Competition in Spoken Word Production. PLOS ONE 9(2): e88674.

Piai, V., & Zheng, X. (2019). Speaking waves: Neuronal oscillations in language production. *In K. D. Federmeier (Ed.), The psychology of learning and motivation* (pp. 265–302). Elsevier Academic Press. https://doi.org/10.1016/bs.plm.2019.07.002

Piai, V., Roelofs, A., & van der Meij, R. (2012). Event-related potentials and oscillatory brain responses associated with semantic and Stroop-like interference effects in overt naming. Brain research, 1450, 87–101.

Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language*, 39(4), 633–651. https://doi.org/10.1006/jmla.1998.2592

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. Psychological Bulletin, 144(10), 1002–1044. https://doi.org/10.1037/bul0000158

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences,* 27(2), 169–226. https://doi.org/10.1017/S0140525X04000056

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences, 36*(4), 329–347. https://doi.org/10.1017/S0140525X12001495

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *The Behavioral and brain sciences*, 36(4), 329–347.

Pickering, M. J., & Garrod, S. (2021). Understanding dialogue: Language use and social interaction*. Cambridge University Press.*

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(2), 816-847. https://doi.org/10.1016/j.neuroimage.2012.04.062

Pulvermüller F. (1999). Words in the brain's language. *The Behavioral and brain sciences*, 22(2), 253–336.

Pulvermüller, F. (2002). The neuroscience of language. *Cambridge: Cambridge University Press.*

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience,* 6(7), 576–582.

Pulvermüller, F. (2013). How neurons make meaning: Brain mechanisms for embodied and abstract-symbolic semantics. *Trends in Cognitive Sciences, 17*(9), 458-470. https://doi.org/10.1016/j.tics.2013.06.004

Pulvermüller, F. (2018). Neural reuse of action perception circuits for language, concepts and communication. *Progress in Neurobiology*, 160, 1-44. https://doi.org/10.1016/j.pneurobio.2017.07.001

Pulvermüller, F., & Fadiga, L. (2010). Active perception: Sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, 11(5), 351-360. https://doi.org/10.1038/nrn2811

Rapp, B. and Goldrick, M., (2000). Discreteness and interactivity in spoken word production. *Psychological review*, 107(3), 460.

Rapp, B. and Goldrick, M., (2004). Feedback by any other name is still interactivity: a reply to Roelofs (2004).

Rasenberg, M., Özyürek, A., & Dingemanse, M. (2020). Alignment in Multimodal Interaction: An Integrative Framework. *Cognitive science*, 44(11), e12911. https://doi.org/10.1111/cogs.12911

Rauschecker, J.P. and Scott, S.K., (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature neuroscience*, 12(6), pp.718-724.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372-422. https://doi.org/10.1037/0033-2909.124.3.372

Redcay, E., Kleiner, M., & Saxe, R. (2012). Look at this: The neural correlates of initiating and responding to bids for joint attention. *Frontiers in Human Neuroscience, 6*, 169. https://doi.org/10.3389/fnhum.2012.00169

Reitter, D., & Moore, J. D. (2006). Priming of syntactic rules in task-oriented dialogue and spontaneous conversation. *In Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 28, No. 28).

Rens, N., Lancia, G. L., Eluchans, M., Schwartenbeck, P., Cunnington, R., & Pezzulo, G. (2023). Evidence for entropy maximisation in human free choice behaviour. *Cognition*, 232, 105328. https://doi.org/10.1016/j.cognition.2022.105328

Riès, S. K., Dhillon, R.K., Clarke, A., King-Stephens, D., Laxer, K.D., Weber, P.B., ... Knight, R.T. (2017). Spatiotemporal dynamics of word retrieval in speech production revealed by cortical high-frequency band activity. *Proc Natl Acad Sci*. 114 : E4530-E4538.

Rilling, J. K., & Sanfey, A. G. (2011). The neuroscience of social decision-making. *Annual Review of Psychology, 62*(1), 23–48. https://doi.org/10.1146/annurev.psych.121208.131647

Rizzolatti, G., & Sinigaglia, C. (2016). The mirror mechanism: A basic principle of brain function. *Nature Reviews Neuroscience, 17*(12), 757-765. https://doi.org/10.1038/nrn.2016.135.

Rizzolatti, G., Fogassi, L. & Gallese, V. (1997). Parietal cortex: from sight to action. *Curr. Opin. Neurobiol*. 7, 562–567.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, 42, 107–142.

Roelofs, A. (1997a). The WEAVER model of word-form encoding in speech production. *Cognition*, 64, 249–284.

Roelofs, A. (1997b). A case for non-decomposition in conceptually driven word retrieval. *Journal of Psycholinguistic Research* 26, 33–67.

Roelofs, A. (2004). Error biases in spoken word planning and monitoring by aphasic and nonaphasic speakers: comment on Rapp and Goldrick (2000*). Psychological Review*, 111, 579-80.

Roelofs, A. (2014). A dorsal-pathway account of aphasic language production: The WEAVER++/ARC model. *Cortex*, 59, 33–48.

Roelofs, A., & Ferreira, V. (2019). The architecture of speaking. In Hagoort, P. (ed.), Human language: From genes and brains to behavior, 35 - 50. *Cambridge*, MA: MIT Press.

Rogers, T. T., & Patterson, K. (2007). Object categorization: reversals and explanations of the basic-level advantage. *Journal of Experimental Psychology*: General, 136(3), 451.

Runnqvist, E. (2023) Self-Monitoring: The Neurocognitive Basis of Error Monitoring in Language Production. *Language Production, 1, Routledge*, pp.168-190, 2023, doi:10.4324/9781003145790-7.

Runnqvist, E., Bonnard, M., Gauvin, H.S., Attarian, S., Trébuchon, A., Hartsuiker, R.J. & Alario, F.-X. (2016). Internal modeling of upcoming speech: A causal role of the right posterior cerebellum in non-motor aspects of language production, *Cortex*, 81, 203-214.

Runnqvist, E., Chanoine, V., Strijkers, K., Pattamadilok, C., Bonnard, M., Nazarian, B., Sein, J., Anton, J.-L., Dorokhova, L., Belin, P. & Alario, F–X. (2021). Cerebellar and Cortical Correlates of Internal and External Speech Error Monitoring. *Cerebral Cortex Communications.* 2 (2).

Sahin, N. T., Pinker, S., Cash, S. S., Schomer, D., & Halgren, E. (2009). Sequential processing of lexical, grammatical, and phonological information within Broca's area. *Science* (New York, N.Y.), 326(5951), 445–449.

Salmelin, R., Hari, R., Lounasmaa, O. V., & Sams, M. (1994). Dynamics of brain activation during picture naming. *Nature*, 368, 463–465.

Schade, U., & Berg, T. (1992). The role of inhibition in a spreading-activation model of language production. II. The simulational perspective. *Journal of Psycholinguistic Research*, 21, 435-462.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *The Behavioral and brain sciences*, 36(4), 393–414. https://doi.org/10.1017/S0140525X12000660.

Schippers, M. B., Roebroeck, A., Renken, R., Nanetti, L., & Keysers, C. (2010). Mapping the information flow from one brain to another during gestural communication. *Proceedings of the National Academy of Sciences*, 107(20), 9388-9393.

Schmitt, B. M., Münte, T. F. & Kutas, M. (2000). Electrophysiological estimates of the time course of semantic and phonological encoding during implicit picture naming. *Psychophysiology* 37, 473–484.

Schnur, T.T., Schwartz, M.F., Kimberg, D.Y., Hirshorn, E., Coslett, H.B. and Thompson-Schill, S.L., (2009). Localizing interference during naming: Convergent neuroimaging and neuropsychological evidence for the function of Broca's area. *Proceedings of the National Academy of Sciences*, 106(1), pp.322-327.

Schoot, L., Hagoort, P., & Segaert, K. (2016). What can we learn from a two-brain approach to verbal interaction? *Neuroscience & Biobehavioral Reviews*, 68, 454-459.

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

Schwartz, M.F., Dell, G.S., Martin, N., Gahl, S. and Sobel, P., (2006). A case-series test of the interactive two-step model of lexical access: Evidence from picture naming. *Journal of Memory and language*, 54(2), 228-264.

Schwartz, M.F., Faseyitan, O., Kim, J. and Coslett, H.B., (2012). The dorsal stream contribution to phonological retrieval in object naming. *Brain*, 135(12), 3799-3814.

Scott, S. and Johnsrude, I., (2003). The organisation and functional organisation of speech perception. *Trends in Neurosciences*, 26(2), 100-107.

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, *10*(2), 70-76. https://doi.org/10.1016/j.tics.2005.12.009

Segaert, K., Menenti, L., Weber, K., Petersson, K. M., & Hagoort, P. (2012). Shared syntax in language production and language comprehension—An fMRI study. *Cerebral Cortex*, 22(7), 1662-1670. https://doi.org/10.1093/cercor/bhr249

Seidenberg, M.S. and McClelland, J.L., (1989). A distributed, developmental model of word recognition and naming. *Psychological review,* 96(4), 523

Shadmehr, R., Krakauer, J.W. (2008). A computational neuroanatomy for motor control. *Exp. Brain. Res*., 185(3), 359–381.

Shadmehr, R., Smith, M.A., Krakauer, J.W. (2010). Error correction, sensory prediction, and adaptation in motor control. *Annual Review of Neuroscience*, 33, 89–108.

Sievers, B., Welker, C., Hasson, U., Kleinbaum, A. M., & Wheatley, T. (2024). Consensus-building conversation leads to neural alignment. Nature communications, 15(1), 3936. https://doi.org/10.1038/s41467-023-43253-8

Silbert, L. J., Honey, C. J., Simony, E., Poeppel, D., & Hasson, U. (2014). Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proceedings of the National Academy of Sciences*, 111(43), E4687-E4696.

Singer, W. (2013). Cortical dynamics revisited. *Trends Cogn Sci.*, 17, 616-626.

Singer, W., Gray, C.M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annu Rev Neurosci*., 18, 555-586.

Skipper, J. I., Goldin-Meadow, S., Nusbaum, H. C., & Small, S. L. (2007). Speech-associated gestures, Broca's area, and the human mirror system. *Brain and Language, 101*(3), 260–277. https://doi.org/10.1016/j.bandl.2007.02.008

Solomon, E.A., Kragel, J.E., Sperling, M.R. et al. Widespread theta synchrony and high-frequency desynchronization underlies enhanced cognition. *Nature Communications* 8, 1704 (2017). https://doi.org/10.1038/s41467-017-01763-2

Starreveld, P. A., & La Heij, W. (1996). Time-course analysis of semantic and orthographic context effects in picture naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 22(4), 896–918.

Staub, A., & Clifton, C., Jr (2006). Syntactic prediction in language comprehension: evidence from either...or. *Journal of experimental psychology. Learning, memory, and cognition*, 32(2), 425–436. https://doi.org/10.1037/0278-7393.32.2.425

Stephens, G. J., Silbert, L. J., & Hasson, U. (2010). Speaker-listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences, 107*(32), 14425–14430. https://doi.org/10.1073/pnas.1008662107

Strijkers, K. (2016), A Neural Assembly–Based View on Word Production: The Bilingual Test Case. *Language Learning*, 66, 92-131.

Strijkers, K., & Costa, A. (2011). Riding the Lexical Speedway: A Critical Review on the Time Course of Lexical Selection in Speech Production. *Frontiers in Psychology*, 2(December), 1–16.

Strijkers, K., & Costa, A. (2016a). The cortical dynamics of speaking: Present shortcomings and future avenues. *Language, Cognition and Neuroscience*, 31(4), 484–503. https://doi.org/10.1080/23273798.2015.1120878

Strijkers, K., & Costa, A. (2016b). On words and brains: Linking psycholinguistics with neural dynamics in speech production. *Language, Cognition and Neuroscience*, 31(4), 524–535.

Strijkers, K., Costa, A., & Pulvermüller, F. (2017). The cortical dynamics of speaking: Lexical and phonological knowledge simultaneously recruit the frontal and temporal cortex within 200 ms. *NeuroImage*, 163, 206–219.

Strijkers, K., Costa, A., & Thierry, G. (2010). Tracking lexical access in speech production: Electrophysiological correlates of word frequency and cognate effects. *Cerebral Cortex*, 20(4), 912–928. https://doi.org/10.1093/cercor/bhp153

Strijkers, K., Costa, A., & Thierry, G. (2010). Tracking lexical access in speech production: electrophysiological correlates of word frequency and cognate effects. *Cerebral cortex*, 20(4), 912–928.

Stuart, A., Kalinowski, J., Rastatter, M.P., Lynch, K. (2002). Effect of delayed auditory feedback on normal speakers at two speech rates. *J. Acoust. Soc. Am*., 111(5 Pt 1), 2237–2241.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*(5217), 1632–1634. https://doi.org/10.1126/science.7777863

Tannen, D. (2005). *Conversational style: Analyzing talk among friends* (2nd ed.). Oxford University Press.

Tenke, C. E., & Kayser, J. (2001). A convenient method for detecting electrolyte bridges in multichannel electroencephalogram and event-related potential recordings. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, 112(3), 545–550. https://doi.org/10.1016/s1388-2457(00)00553-8

Thorpe, S., Fize, D. & Marlot, C. (1996). Speed of processing in the human visual system. *Nature* 381, 520–522.

Tian, X., & Poeppel, D. (2010). Mental imagery of speech and movement implicates the dynamics of internal forward models. *Frontiers in psychology*, 1, 166.

Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*, 26(7), 952–981.

Traxler, M. J., Tooley, K. M., & Pickering, M. J. (2014). Syntactic priming during sentence comprehension: Evidence for the lexical boost. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(4), 905–918. https://doi.org/10.1037/a0036377

Tylén, K., Weed, E., Wallentin, M., Roepstorff, A., & Frith, C. D. (2010). Language as a tool for interacting minds. *Mind & Language, 25*(1), 3-29. https://doi.org/10.1111/j.1468-0017.2009.01379.x

Ueno, T., Saito, S., Rogers, T.T. and Ralph, M.A.L., (2011). Lichtheim 2: synthesizing aphasia and the neural basis of language in a neurocomputational model of the dual dorsal-ventral language pathways. *Neuron*, 72(2), 385-396.

Ungerleider, L.G. & Mishkin, M. (1982). Two cortical visual systems. *In D.J. Ingle, M.A. Goodale & R.J.W. Mansfield (Eds.). Analysis of visual behavior* (pp. 549-586). Cambridge, MA: MIT Press

Van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. *Journal of experimental psychology. Learning, memory, and cognition*, 31(3), 443–467. https://doi.org/10.1037/0278-7393.31.3.443

Van Berkum, J. J., van den Brink, D., Tesink, C. M., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of cognitive neuroscience*, 20(4), 580–591. https://doi.org/10.1162/jocn.2008.20054

Van Turennout, M., Hagoort, P. & Brown, C. M. (1998) Brain activity during speaking: From syntax to phonology in 40 msec. *Science*, 280, 572–74.

VanRullen, R. & Thorpe, S. J. (2002). Surfing a spike wave down the ventral stream. *Vision Res*. 42, 2593–2615.

Varela, F. J., Lachaux, J. P., Rodriguez, E., & Martinerie, J. (2001). The brainweb: Phase synchronization and large-scale integration. *Nature Reviews Neuroscience*, 2(4), 229-239. https://doi.org/10.1038/35067550

Vesper, C., Butterfill, S., Knoblich, G., & Sebanz, N. (2010). A minimal architecture for joint action. *Neural Networks, 23*(8–9), 998–1003. https://doi.org/10.1016/j.neunet.2010.06.002

Vigliocco, G., & Hartsuiker, R. J. (2002). The interplay of meaning, sound, and syntax in sentence production. *Psychological Bulletin*, 128(3), 442–472.

Walker, G. M., & Hickok, G. (2016). Bridging computational approaches to speech production: The semantic-lexical-auditory-motor model (SLAM*). Psychonomic bulletin & review*, 23(2), 339–352.

Wernicke, C. (1874/1969). The symptom complex of aphasia: A psychological study on an anatomical basis. *In: Cohen, R.S., Wartofsky, M.W., editors. Boston studies in the philosophy of science. Dordrecht: D. Reidel Publishing Company*, 34-97.

Wheatley, T., Kang, O., Parkinson, C., & Looser, C. E. (2012). From mind perception to mental connection: Synchrony as a mechanism for social understanding. *Social and Personality Psychology Compass, 6*(8), 589-606. https://doi.org/10.1111/j.1751-9004.2012.00450.x

Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. Journal of Cognitive Neuroscience, 16(7), 1272–1288. https://doi.org/10.1162/0898929041920487

Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Language, 31*(2), 183–194. https://doi.org/10.1016/0749-596X(92)90010-U

Willems, R. M., & Hagoort, P. (2007). Neural evidence for the interplay between language, gesture, and action: A review. *Brain and Language*, 101(3), 278-289. https://doi.org/10.1016/j.bandl.2007.03.004

Wittgenstein, L. (1953). Philosophical investigations (G. E. M. Anscombe, Trans.). *Macmillan*.

Wolpert, D.M., Ghahramani, Z., Jordan, M.I. (1995). An internal model for sensorimotor integration. *Science*, 269(5232), 1880–1882.

Zhu, C., Byrd, R. H., and Nocedal, J. (1997) L-BFGS-B: Algorithm 778: L-BFGS-B, FORTRAN routines for large scale bound constrained optimization. *ACM Transactions on Mathematical Software*, Vol 23, Num. 4, pp. 550 - 560.