NNT/NL : 0000AIXM0000/000ED000

THÈSE DE DOCTORAT

Soutenue à Aix-Marseille Université
le 9 décembre 2022 par

Giusy Cirillo

Behavioral and electrophysiological markers of partner-adaptation in joint language production

Discipline
Sciences du Langage

École doctorale
ED 356 – Cognition, Langage, Education

Laboratoire
Laboratoire Parole et Langage

Composition du jury

Anna KUHLEN RWTH Aachen University	Rapporteuse
Chiara GAMBI University of Warwick	Rapporteuse
François-Xavier ALARIO Aix-Marseille Université	Président du jury
Raphaël FARGIER Université Côte d'Azur	Examinateur
Noël NGUYEN Aix-Marseille Université	Directeur de thèse
Cristina BAUS Universitat de Barcelona	Co-directrice de thèse

Affidavit

I, undersigned, Giusy Cirillo, hereby declare that the work presented in this manuscript is my own work, carried out under the scientific direction of Noël Nguyen and Cristina Baus, 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 26 septembre 2022

Liste de publications et participation aux conférences

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

   1. 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
   2. Cirillo, G., Runnqvist, E., Strijkers, K., Nguyen, N., & Baus, C. (en soumission). Effects of shared attention on joint language production across processing levels. *Language, Cognition and Neuroscience*
   3.
   4. Cirillo, G., Runnqvist, E., Strijkers, K., Nguyen, N., & Baus, C. (en préparation). Electrophysiological evidence for prediction and prediction adaptation in human-robot interaction. A joint picture-naming study.

2) Participation aux conférences[2] et écoles d'été au cours de la période de thèse :
   1. The Virtual Social Interaction Conference, University of Glasgow, 30 juin 2021 (Communication orale)
   2. 7ème édition de la Journée Annuelle des Doctorants, LPL, Aix-Marseille Université, 13 décembre 2019 (Communication orale)
   3. 8th Joint Action Meeting, Instituto Italiano di Tecnologia, Genova, 10-13 Julliet 2019 (Poster)
   4. Ecole d'été en Bilinguisme et Multilinguisme, Universitat Pompeu Fabra, Barcelona, 15-25 Septembre 2019 (Poster)
   5. Colloque des Jeunes Chercheurs en Sciences Cognitives, École normale supérieure, Paris, 1-2 mars 2019 (Poster)
   6. 6ème édition de la Journée Annuelle des Doctorants, Laboratoire Parole et Langage, Aix-Marseille Université, 6 décembre 2018 (Poster)

---

[1] Cette liste comprend les articles publiés, les articles soumis à publication et les articles en préparation ainsi que les livres, chapitres de livre et/ou toutes formes de valorisation des résultats des travaux propres à la discipline du projet de thèse. La référence aux publications doit suivre les règles standards de bibliographie et doit être conforme à la charte des publications d'AMU.

[2] Le terme « conférence » est générique. Il désigne à la fois « conférence », « congrès », « workshop », « colloques », « rencontres nationales et/ou internationales » … etc.

Indiquer si vous avez fait une présentation orale ou sous forme de poster.

Abstract

Language interaction is a form of joint action in which individuals coordinate their verbal and non-verbal behavior to communicate successfully. To do so, interlocutors come to co-represent each other at multiple language dimensions, from conceptualization down to lexical selection and to phonological representation. They also come to copy each other's words and expressions. The present thesis aimed at exploring the social and cognitive dynamics that bring interlocutors to shift towards the perspective of their partner and discuss the effects that partner-adaptive behaviors have on memory (information encoding), prediction and production (word choices). In three empirical studies, we provided evidence for co-representation and adaptation processes underlying human-human and human-robot interaction.

In the first study, we employed a shared go/no-go task in which pairs of participants categorized objects displayed on a screen. Objects varied in whether they required a response of only one participant, both participants or neither. Each participant reacted either according to the animacy (semantic task) or to the first letter/phoneme of the object (phoneme-monitoring task). We tested whether shared attention over separate, selective linguistic properties affects language processing. We found that participants attending to the phonetic level were faster at responding in joint trials, suggesting that phonetics benefits from co-representing semantics. Results from a recall test revealed, in addition, a social memory advantage for both groups of participants when attending objects together with their partner. In the second and third study we tested prediction and adaptation to the lexico-semantic choices produced by a humanoid robot in a joint picture naming task using electrophysiological (study 2) and behavioral measures (study 3). The robot was programmed to give the superordinate, semantic category name (e.g., tool) instead of the more typical basic-level name (e.g., hammer) of objects belonging to specific semantic categories. Electrophysiological results revealed that 1) human partners are able to monitor and predict a robot's upcoming words, as signaled by the comparable activity when preparing to speak and when the robot prepares to speak and that 2) they can adapt to the idiosyncratic lexico-semantic patterns of the robot. From the behavioral analyses, we found that participants progressively produced the superordinate names in same semantic contexts of the robot, indicating conceptual alignment towards the robot's word choices.

The present thesis offers an exploratory overlook on the adaptation dynamics affecting language processing, combining relevant production stages (lexico-semantic, lexical, phonetic) with specific cognitive dimensions (prediction, production, memory) within controlled and ecologically-valid settings.

Keywords: Language production, Partner-adaptive behavior, Joint action, Shared attention, Co-representation, Linguistic alignment, Human-robot interaction, Picture-naming, EEG

Résumé

L'interaction linguistique est une forme d'action conjointe dans laquelle les individus coordonnent leur comportement verbal et non verbal pour communiquer avec succès. Pour ce faire, les interlocuteurs en viennent à se représenter mutuellement dans de multiples dimensions du langage, de la conceptualisation à la sélection lexicale et à la représentation phonologique. Ils en viennent également à copier les mots et les expressions des autres. La présente thèse vise à explorer les dynamiques sociales et cognitives qui amènent les interlocuteurs à se tourner vers la perspective de leur partenaire et à discuter des effets que les comportements d'adaptation au partenaire ont sur la mémoire (encodage de l'information), la prédiction et la production (choix des mots). Dans trois études empiriques, nous avons fourni des preuves des processus de co-représentation et d'adaptation qui sous-tendent l'interaction homme-homme et homme-robot.

Dans la première étude, nous avons utilisé une tâche partagée de type go/no-go dans laquelle des paires de participants classaient des objets affichés sur un écran. Les objets variaient selon qu'ils nécessitaient une réponse d'un seul participant, des deux participants ou d'aucun des deux. Chaque participant réagissait soit en fonction de l'animation (tâche sémantique), soit en fonction de la première lettre/phonème de l'objet (tâche de suivi du phonème). Nous avons testé si l'attention partagée sur des propriétés linguistiques distinctes et sélectives affecte le traitement du langage. Nous avons constaté que les participants qui portaient leur attention sur le niveau phonétique répondaient plus rapidement dans les essais conjoints. Cela suggère que la phonétique bénéficie de la coprésence de la sémantique. Les résultats d'un test de rappel ont révélé, en outre, un avantage de mémoire sociale pour les deux groupes de participants lorsqu'ils repondent aux objets ensemble.

Dans les deuxième et troisième études, nous avons testé la prédiction et l'adaptation aux choix lexico-sémantiques produits par un robot humanoïde dans une tâche conjointe de dénomination d'images en utilisant des mesures électrophysiologiques (étude 2) et comportementales (étude 3). Le robot était programmé pour donner le nom de la catégorie sémantique supérieure (p. ex., outil) au lieu du nom de niveau de base plus typique (p. ex., marteau) des objets appartenant à des catégories sémantiques spécifiques. Les résultats électrophysiologiques ont révélé que 1) les partenaires humains sont capables de surveiller et de prédire les mots à venir du robot, comme le signale l'activité comparable lorsqu'ils se préparent à parler et lorsque le robot se

prépare à parler, et que 2) ils peuvent s'adapter aux modèles lexico-sémantiques idiosyncrasiques du robot. D'après les analyses comportementales, nous avons constaté que les participants produisaient progressivement les noms superordonnés dans les mêmes contextes sémantiques que le robot, indiquant un alignement conceptuel sur les choix de mots du robot.

La présente thèse offre un regard exploratoire sur la dynamique d'adaptation affectant le traitement du langage, combinant des étapes de production pertinentes (lexico-sémantique, lexicale, phonétique) avec des dimensions cognitives spécifiques (prédiction, production, mémoire) dans des contextes contrôlés et écologiquement valides.

Dedication

A mia nonna, Lina Scalzo Teluso

Acknowledgements

What are you grateful for today? What are you grateful for in your life?

At the end of the day, it is good practice to look back and be grateful to the people that one way or another were there with (for) us.

To start with, I would like to thank my supervisors, Noel Nguyen and Cristina Baus. Thank you Noel for having believed in me as PhD candidate, for your warm welcome in Aix (and France). Thank you for supporting me during those very important years as early researcher and for giving me the possibility to be part of the Laboratore Parole et Langage (LPL) and of the ILCB. This was for me a great honor. Thank you Cristina for your human and scientific support, for having taught me how to write an article, how to analyze (EEG) data, how to be a kind-of-ok PhD fellow. You were, and still are, a polar star in my PhD. Thank you Elin and Kristof to be amazing members of this team. Thank you for your advices, and for your professional support. Thank you also for our informal chats, the aperos and the Auriol time.

I would also like to thank Anna Kuhlen and Chiara Gambi for accepting to be the reviewers of my thesis. Your work as leading researchers is for me source of inspiration. Thanks to the other members of the jury, Xavier Alario and Raphaël Fargier who were present in different moments of my PhD, always bringing brilliant, positive and encouraging advices. Raph, I was so lucky to have met you at the start of my PhD, when you were still a Post-doc, and I am proud of you, for the new position you got and that you deserve so much. You are a Thinker, but also a friend and a tutor. You are really the best examiner I could ask for my defense.

Thanks to all members of the LPL. You became a big family to me. A special thanks to Nadera, Cathy, Stephanie that accompanied all my administrative journeys with a big smile. Thank you Laurent and Serge for your extraordinary direction of the lab. Thank you Gilles to be right next door during the breaks.

A heartful *thank you* to the PhDs and the Postdocs of the building B. Thanks to the ones that left, the ones that came, the ones that are *toujours* there. Thank you Francesca Carbone, la mia amica geniale, ora volata a Canterbury. Le nostre discussioni sono sempre state fonte di inspirazione, ritorno alle origini da un lato e desiderio di rinnovamento dall'altro. Thank you Sneza. We met during the interview for the ILCB and were extremely happy to find each other afterwards. Thank you for sharing this journey with me as lego grads. Nanana Nana Nana. Grazie Chiaretta per la tua amicizia, per la tua risata, per le tue teorie creative e per il tuo aiuto con LaTeX. Elora, when I came I was a *little bunny* (your words) that wanted to go to folk festivals to enjoy her evenings and meet new people (guess what at the end I met my husband!). Now I am 4 years older, I am a mum, and I am a doctor! Just like you! Thank you my sweet Birgit, you became a real friend to me. We went through pregnancy together and now through ATER! You have a lightening soul, keep it! I wish you the best for your career.

Thank you to the people I met in Aix during those years. Milena, thank you for our adventures in the nature and the camping experience we shared. Grazie Eleonora per la tua contagiosa passione per la danza e la musica tradizionale della Calabria e del sud Italia che mi stai trasmettendo in queste magiche serate marsigliesi. Grazie ai membri di Aix Baleti e dell'ostau di Marsiglia, grazie per le serate di balli, sorrisi e buona musica che hanno ritmato il mio dottorato in questi anni.

Thank you to my Italian and international friends here and there, that I had the chance to meet during my university years between Pisa and Tubingen and that kept on being there for me. Everyone of you contributed in different ways to this achievement. Thank you Flavia (la migliore coinquilina di sempre), Erica, Laura, Teresa, Sarky (my goats need grass), my sweet Lizzie (I would like to find you again...and repair what has been stirred by my egoism), Chiara, Francesco, Elisa, Anwar, Noemi, Alex, Delia, Michael, Virgi...Thank you to the people that were close to me just before coming to Aix, when I was doing an intership in Maastricht. You were crucial for the decision to start over a PhD here in Aix. Thank you Olga and Paddy, my flatmates that made me embrace the house rule 'we never choose easy options' the days before the interview for the scholarship. Grazie Diego per avermi aiutato a preparare il colloquio, per avermi dato il coraggio di dare il mio best. Grazie per avermi insegnato a trovare quel Diamond in the rough. Thank you Katerina (sister seal), for being an inspiring person, and Xan for introducing me to Friends and bird-watching.

Grazie ai miei amici di sempre, legati alla mia terra di origine. Giovanna, Donatella, Simona...La vostra presenza nella mia vita mi ha probabilmente portato fino a questo giorno. Grazie a Nella Coletta, che a quindici anni mi iniettò come una medicina (o un incantesimo?) l'amore per il sapere, la cultura, la curiosità, la ricerca. Sono grata alla nostra corrispondenza speciale, iniziata 13 anni fa e non ancora finita.

Un grazie speciale alla mia famiglia. Al mio fratellino saggio che mi sta sempre accanto, a mia mamma Caterina e a mio papà Angelo che in questi anni hanno sempre accettato qualsiasi idea strampalata mi passasse per la testa. Grazie di essere lì per me e ora grazie di essere lì per Gabrièu. Un grazie speciale ai miei cuginOtti, ed in particolare a Carlo, il mio fratellone prossimo alla paternità. Un grazie immenso alle mie nonne Virginia e Lina, per me due figure mistiche, complete e complementari tra di loro.

E poi tu, Dani! Grazie non solo per l'amore che mi dai e che mi dimostri tutti i giorni nei piccoli gesti, ma anche per l'esempio di vita sana e autentica al quale hai introdotto questa ragazza selvaggia e 'poco chiara'. Grazie per questa famiglia che

abbiamo appena creato insieme e per la tanta speranza e tanta bellezza che si prospetta davanti a noi.

E per finire, Gabrièu. Mi hai fatto diventare Mamma, e questo è stato per me forse la metamorfosi più importante, bella e difficile della mia vita. I tuoi sorrisi in questi mesi di scrittura mi hanno aiutato a prendere le cose con più leggerezza. Ogni giorno con te è una scoperta ed una riscoperta del Bambino. Mentieni acceso il tuo *fuoco creativo*. Grazie anche a te.

Contents

Chapter 1

The theoretical context

1.1 General Introduction

Our ability to adapt our language behavior according to the interlocutor is one of the most remarkable cognitive skills we possess (Brennan & Hanna, 2009; Carrard, 2021; Luo, Robbins, Martin, & Demiray, 2019). Having a conversational partner affects the way we select and elaborate information, changes the way we refer to events and objects and, ultimately, shapes the way think and we speak. This multi-level process is complex and gradual, supported by mutual coordination in terms of mental representation and turn-taking, but also very automatic, as speakers adjust systematically and spontaneously to their interlocutors. A fundamentally adaptive component therefore characterizes any language exchange and is evident at different processing stages of language representation (e.g., phonetic, lexical, conceptual).

Imagine a couple contemplating the breathtaking view on the top of a 3000m mountain. After a moment of sacred and well deserved silence, they would start commenting what they *see*, and become at the same time speakers and listeners of a conversation. Each speaker contributes to the discussion by adding what he or she believes relevant to remark and use specific words to shape his/her own thoughts. Let's imagine that speaker 1 is impressed by the whiteness of the sky with its light shades of gray and starts to refer to it as the *pearly sky*. Speaker 2 would gradually shift towards the perspective of speaker 1 and start seeing the sky as *pearly*. In other words, he would form a mental representation of the sky which matches the one of his interlocutor. From a speaking point of view, he would most likely adopt the same vocabulary to describe the landscape (i.e., starting saying *pearly sky*), showing that he implicitly agrees to the perspective of his partner, and this way he optimizes the communication. Now imagine that speaker 2 starts focusing on the big trees with their large green leaves on the slopes of the mountains, and points to them along the conversation. He would draw the attention of his partner to another aspect of the landscape. He might call the trees with the label *tree* or, depending on his botanical knowledge, say *beech*, *olm* or *chestnut tree* instead. Speaker 1 would then get closer to either the generic or specific conceptualization of a tree and use the corresponding terms as established during the conversation. This is just an example of the way we adjust our representation and linguistic system to the interlocutor when sharing the same visual material. There are some agreements that are established implicitly over the course of a conversation between interlocutors. Brennan and Clark (1996) called them "conceptual pacts", which design a "temporary agreement" on how speakers conceptualize objects.

In the present PhD thesis, we monitored adaptation affecting diverse cognitive processes, including language prediction and production and memory, within joint language settings. In three experimental studies, we aimed at enlarging the theoretical and empirical knowledge on partner-adaptation in language processing and map its dynamics using experimental tasks targeting individual or multiple language dimensions. As for the example of the couple commenting the landscape, the presence of a language partner modulates our attentional focus, affecting basic processes such as categorization and learning (Böckler, Knoblich, & Sebanz, 2012). Once we mutually agree and direct our *mental gaze* to the same sensory information, we begin to co-represent the language actions of our partner over the course of the shared activity (Shintel & Keysar, 2009). Language production is itself a form of joint action in which interactive partners coordinate their verbal and non-verbal behavior across the different moments of a language exchange (Clark, 1996; Gambi & Pickering, 2011, 2013; Garrod & Pickering, 2009; van der Wel, Becchio, Curioni, & Wolf, 2021). The parity of representations between production and perception allows predicting a partner's upcoming speech (McEllin, Knoblich, & Sebanz, 2018; Pickering & Garrod, 2013b; Prinz, 1997). This mechanism is linked to neuroscientific accounts that interpret the brain as a proactive organ which does not passively perceive the sensory information present in the environment, but is actively involved in generating predictions about the upcoming sensory inputs (Bar, 2009; Brown & Brüne, 2012; Bubic, Von Cramon, & Schubotz, 2010). Importantly, those predictive mechanisms are adaptive: while we continually get new inputs from the environment, we use prediction errors to update internal models (Pickering & Garrod, 2013a, 2013b).

The effects of co-representation are directly evident at the production level, when interlocutors align with their conversational partners by adopting each other's verbal (e.g., words, sentences, syntactic structures, intonation) and non-verbal choices (e.g., posture and gestures, Bergmann, Branigan, & Kopp, 2015; Fusaroli et al., 2012; Reitter & Moore, 2014). Adaptation is also modulated by the type of interlocutor we engage our language exchange with (Brennan & Hanna, 2009). We behave differently whether we are in front of a close friend, family member or whether we interact with a complete stranger. Generally speaking, the more familiar we are with the interlocutor, the more we will be able to predict what to hear and plan what to say. However, it is also true that a less known interlocutor would require further monitoring from our part, as we would be less able to predict his language behavior. Belief about the interlocutor has therefore a central role in language processing, and cannot be omitted when investigating adaptation in joint production. Theories focusing on how adaptation emerges during dialogue and conversation claim that partner-specific information is integrated in memory as any other type of linguistic and contextual cue and used during dialogue and conversation to quickly make the appropriate adjustments in production (Brennan & Hanna, 2009; Hanna, Tanenhaus, & Trueswell, 2003; Metzing & Brennan, 2003).

With technology having increasingly taken the lead in our everyday society, sophisticated machines such as humanoid robots have started to operate as interactive partners in the same physical space as people in a number of industrial, educational and domestic settings (Cooper, Gow, Fensome, Dragone, & Kourtis, 2020; Wudarczyk, Kirtay, Kuhlen, et al., 2021). Human-robot interaction has therefore become more relevant in experimental investigations of joint action mechanisms, as it represents a key to understand to which extent humans are able to successfully carry out an activity with non-human agents and, ultimately, communicate with them (Branigan, Pickering, Pearson, & McLean, 2010; Branigan, Pickering, Pearson, McLean, & Brown, 2011; Chaminade & Cheng, 2009; Kim, Kwak, & Kim, 2013; Marge et al., 2022; Wykowska, Chaminade, & Cheng, 2016). Following this research current, we decided to explore the impact of having an artificial partner as interlocutor on language processing, to examine what are the prerequisites to reach coordination and mutual understanding in joint production. We examined how humans exchange in co-representation processes (planning, prediction, perception) with a social robot and eventually monitored adaptation at the language level in terms of alignment to its lexical choices.

The present chapter offers a review of the main theoretical concepts on which we based our three experimental works. First, we provide an overview of the joint action paradigm and its application on the language domain. This is followed by three sections respectively addressing 1) shared attention and co-representation, 2) adaptive prediction, as supported by the motor resonance account and the integrated theory of language production and comprehension, and 3) adaptive production, or else linguistic (lexical and conceptual) alignment. We reserve then a specific section to human-robot interaction in joint production. The last section of the chapter includes a brief introduction to our three empirical studies.

1.1.1 Before starting: Some terminology clarification

When scrolling through the research literature on the adaptive mechanisms in conversation and, more generally, in joint language production, we face the existence of a variegate terminology. Accommodation, alignment, convergence, entertainment, synchrony are terms that are often interchangeable, generating disagreement among researchers (Rasenberg, Özyürek, & Dingemanse, 2020, for an overview). The term 'alignment', widely present in the current work, has been used in psycholinguistic investigation to refer to interpersonal coordination at the mental representation level (Pickering & Garrod, 2004a), but also at the behavioral level via priming and repetition of salient linguistic features. In this case, *alignment* is preceded by an adjective to indicate the specific language dimension in which it takes place. We refer to lexical, syntactic or gestural alignment depending on whether we want to indicate adaptation of words, sentence structures or gestures respectively (Bergmann & Kopp, 2012; Branigan, Pickering, Pearson, Mclean, & Nass, 2003; Suffill, Kutasi, Pickering, & Branigan, 2021). Similar processes have been investigated using different labels, including entrainment, imitation, mimicry, repetition, and which can all be ultimately explained by priming effects. While they all share evident similarities, each label is associated with a specific framework (Rasenberg et al., 2020). There is no univocal perspective of the use of one term instead of another. Researchers adopt one framework and take all the relevant labels belonging to it. In the absence of agreement, we therefore believe it fundamental to specify our choice of terminology at the beginning of this work.

While we use the term co-representation -in line with the joint action paradigm- to refer to alignment between mental states, we use the term 'linguistic alignment' to denote interlocutors' mutual exchange of linguistic choices and patterns. Alignment is here used at multiple linguistic levels (lexical, semantic and syntactic). Relevant here, we refer to lexical alignment as the behavioral tendency to copy one's partner word choices, and to conceptual alignment to express the idea that interlocutors share the conceptual perspective of the interlocutor when making their linguistic choices. Both concepts will be developed in more details in Section 1.5 and will be object of experimental investigation in our third study (Chapter 4).

1.2 Joint action in language processing

1.2.1 Joint Action: A theory and an experimental paradigm

The capacity of sharing actions is a prerequisite of human beings, who have entrusted their survival on constituting groups rather than living isolated (Cacioppo & Patrick, 2008). Humans have evolved as ultrasocial animals (Tomasello, 2014), and developed sophisticated cognitive skills because of their ability to cooperate and exchange knowledge within and between groups (Dean, Kendal, Schapiro, Thierry, & Laland, 2012; Herrmann, Call, Hernàndez-Lloreda, Hare, & Tomasello, 2007). Most of what we humans do and achieve happens within a social context. While until two decades ago, traditional approaches in cognitive psychology privileged individual actions and thoughts, recent investigations have started treating cognitive processes as sprung by minds acting jointly (Brennan, Galati, & Kuhlen, 2010). In 2006, Sebanz together with a group of researchers introduced the *joint action paradigm* that invites to study human cognition, and its related processes (e.g., attention, decision making, language, memory) within a social setting (see, van der Wel et al., 2021 for an overview of the theoretical and empirical implications of joint action). Joint action has since then become a common methodological choice in various disciplines, bringing to the development of elaborated designs involving dyads or groups of people to target specific aspects of human social cognition.

The term 'joint action' refers to any kind of action in which "two or more people coordinate in space and time to bring about a change in the environment" (Sebanz, Bekkering, & Knoblich, 2006, p.70). In this sense, our everyday life is full of those joint actions, ranging from very simple and automatic actions such as exchanging a high five or shaking hands to more complex actions such as playing music in an orchestra (Knoblich, Butterfill, & Sebanz, 2011; Sebanz et al., 2006; Wenke et al., 2011). What those variegate actions have in common is that they require multiple coordination mechanisms between co-actors, who come to share mental representations as well as specific sensorimotor information (van der Wel, 2015). The ability to predict a partner's upcoming actions, in particular, has a decisive role in the accomplishment of more global joint actions as we will see in Section 1.4. A basic form of social interaction, widely adopted in experimental laboratories, consists into sharing a task, as it happens in a common double game with pre-defined rules in which two or more people take turns to play (Wenke et al., 2011). Task-sharing is used to explore numerous cognitive mechanisms under joint action in controlled settings. Task-sharing experiments are characterized by a go/no-go design: participants are instructed when to play/act (go trials) and when not (no-go trials) from the beginning (Obhi & Sebanz, 2011; Sebanz, Knoblich, & Prinz, 2003; see Figure 1.1). The presence of another person, which is perceived as *other* from the self (Novembre, Ticini, Schütz-Bosbach, & Keller, 2012; Pacherie, 2012), adds up another layer of turn-taking, corresponding to what has been often termed as the 'other-go' condition (Baus et al., 2014), to indicate the moment when the partner is *acting*, while the self is not acting. This other-go condition is particularly interesting when comparing behavioral perception or brain activation between partner's and no-one's trials, to highlight how attentional and selective processes are present when it is the partner's turn to act. In this section, we propose a schematic illustration of the task-sharing design, with all the possible combinations of turn-taking. We implemented the joint action paradigm via a task sharing design in all our three studies to monitor different aspects of adaptation in language and cognition under social presence.

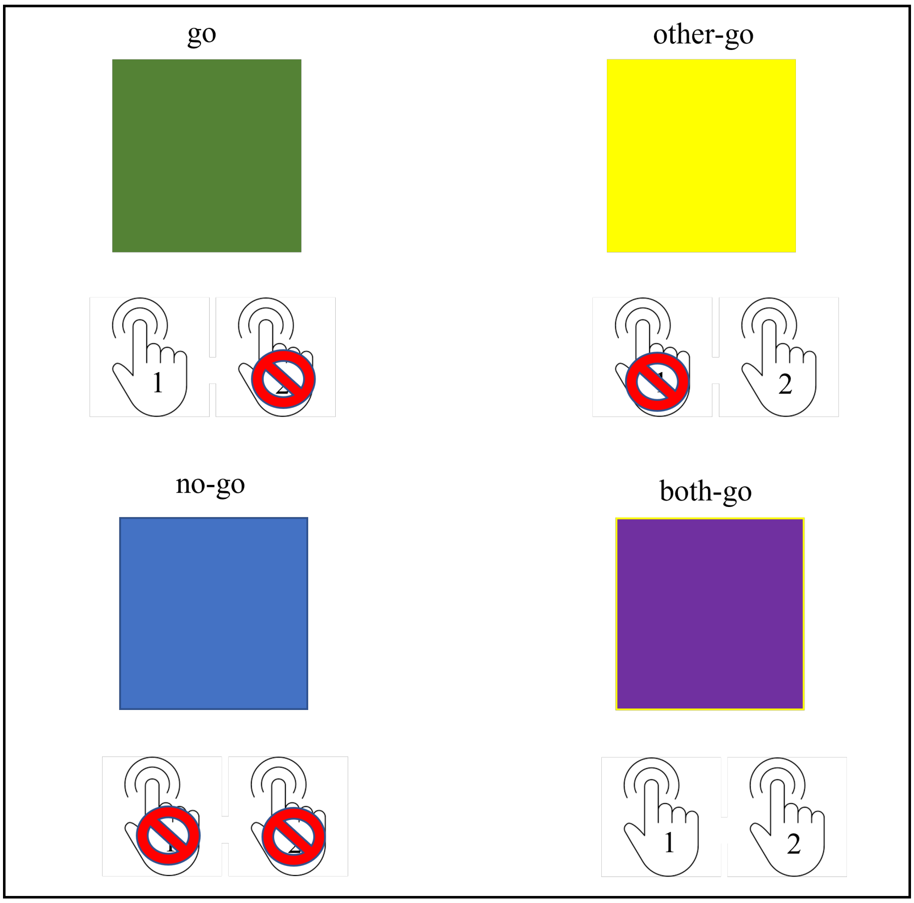

Figure 1.1: **Task sharing paradigm.**

Illustration of the four possible turn conditions within an example of a simple task sharing paradigm, where people are instructed to press a button according to the color presentation of the squares. The go condition, on the top left, corresponds to participant's playing/responding. The other-go condition on the top right corresponds to the participant's partner playing: the *other* compared to the self. The no-go condition on the bottom left means is no-one's turn to play. The go-together condition on the bottom right indicates that it is both participants' turn to play. Importantly, not all studies include all four turn conditions.

1.2.2 Language-as-action approach: joint language tasks

Verbal communication, and more specifically conversation, is perhaps the most paradigmatic expression of joint action. Generally speaking, there exist two main theoretical approaches on the study of language (see Brennan et al., 2010). The language-as-product approach considers language as evolved for thinking (e.g., Chomsky, 1965). This theory "treats language itself as object of study", de-contextualizing it from any social perspective and consequently privileging the use of individual language tasks, easier to design and to perform in controlled settings. The other tradition is the language-as-action approach, which interprets language as evolved for communication (Brennan & Clark, 1996; Clark, 1992; Hanna et al., 2003), and proposes to study language as well as any cognitive process associated to it within a social context and in real-time interaction, claiming for the necessity to use a second-person perspective (Schilbach et al., 2013) to account for language (Gambi & Pickering, 2013). Researchers following this approach embrace the use of joint language tasks rather than isolated settings, as they correspond more to the basic use we make of language: dialogue and conversation (Pickering & Garrod, 2004b).

Joint language tasks have been implemented to study language representation processes during both production (e.g., naming or describing pictures together) and perception (e.g., categorizing objects together). Those tasks allow to approximate the turn taking mechanisms typically found in dialogue and conversation, in which interlocutors alternate between speaking and listening or, in the joint action terminology, between 'go' and 'no-go' moments. While joint language designs are numerous and varied, for the experimental part of the present thesis we only used two types of joint task: a categorization task targeting separate linguistic features from the same visual input (Chapter 2), and a picture-naming task (Chapter 3 and 4). Language categorization is associated to identifying linguistic properties and eventually assigning a label to them. Distinguishing between semantic categories (Eskenazi, Doerrfeld, Logan, Knoblich, & Sebanz, 2013), between animate and inanimate (Nairne, VanArsdall, Pandeirada, Cogdill, & LeBreton, 2013) or among letters/phonemes (phoneme monitoring task; Howell and Ratner (2018)) can vary importantly in terms of speed and memorization. Performing a similar task with a partner can be used to study social effects of language processing at various production stages. Joint picture-naming tasks consist in having two participants alternating in naming pictures displayed on a screen. Participants are normally assigned to a shape or color cue indicating them when it is their turn to speak. In the individual version of the task, facilitation and disruption of lexical access have been attributed to numerous psycholinguistic factors, including lexical frequency, naming agreement, imageability, concept typicality and age of acquisition, impacting naming latencies as a consequence (Alario et al., 2004; Almeida, Knobel, Finkbeiner, & Caramazza, 2007; Barry, Morrison, & Ellis, 1997; Oldfield & Wingfield, 1965; Rossiter & Best, 2013; Strijkers, Costa, & Thierry, 2010). In addition to those features, the joint version of the task has highlighted how playing with a partner represents a social factor further influencing (e.g., via facilitation or interference) the performance. (Baus et al., 2014; Brehm, Taschenberger, & Meyer, 2019; Gambi, Van de Cavey, & Pickering, 2015; Gambi et al., 2015; Kuhlen & Abdel Rahman, 2017). In the present thesis, we adopted the language-as-action theoretical approach via the use of the joint action paradigm. We used a joint picture-naming experiment and a joint object categorization experiment to study relevant psycholinguistic processes such as lexical access and choice (Chapter 3 and 4), memory (Chapter 2), and decision making (Chapter 2, 3, 4) as shepherded by the (social) presence of a partner.

1.2.3 Event-related potentials for language adaptation

In typical behavioral language experiments, possible social effects of task-sharing are extracted from speakers' responses, at both levels of perception (e.g., via reaction times, decision making; see Chapter 2) and production (e.g., via priming/alignment, Pickering & Ferreira, 2008 and naming latency, Kuhlen & Abdel Rahman, 2017; see Chapter 4). However, those types of experiments lack the ability to capture the changes that the brain undergoes during verbal exchanges. Some language processing mechanisms are not automatically evident in behavior, as when accessing relevant linguistic features while the partner is speaking. Neurolinguistic studies are there to identify the neural and cognitive correlates underlying language processing and representation (Skeide & Friederici, 2017). In joint language production, electroencephalography (EEG) has been widely deployed to monitor interpersonal coordination (Kuhlen, Allefeld, & Haynes, 2012). EEG is a noninvasive neuroimaging technique that, given its high temporal resolution, has the advantage of coping with the conversational speed of alternating between speaking and listening (Kuhlen & Rahman, 2020), and show relevant processing during passive listening.

Event-related potentials (ERPs) are post-synaptic potentials corresponding to the measured brain activity response time-locked to specific sensory, cognitive or motor events (Luck, 2005, 2014). They are obtained from extracting and averaging the continuous EEG signal across trials and participants (Bradley & Keil, 2012). From a physiological point of view, they are positive or negative voltage deflections characterized by amplitude, onset latency and topography. Pertinent to joint naming (our experimental paradigm), two main neurolinguistic investigations involving ERPs allow to trace co-representation at the brain level. They are reflected into: 1) comparable ERP components before speaking and listening (as for the P2/P3 components when accessing word frequency in both go- and other-go trials in Baus et al., 2014), and 2) monitoring of ERPs related to prediction error and update for unexpected linguistic choices in regular contexts (Hodapp & Rabovsky, 2021).

Elicitation of specific components has been associated to prediction-error with consequent update of internal models. Among those components we find the feedback-related negativity (Chase, Swainson, Durham, Benham, & Cools, 2011; Hauser et al., 2014), the mismatch negativity (Pulvermüller & Shtyrov, 2006) and the error-related negativity (L. Wang, Gu, Zhao, & Chen, 2020). Within the language processing domain, the N400 component has shown to reflect prediction errors in semantic processing during language comprehension (see, Hodapp and Rabovsky (2021) for overview). The N400, in particular, is a centro-pariatal, negative-going wave observed around 300-500 ms and that peaks at 400 ms after stimulus onset, whose amplitude is larger for a novel or unexpected stimulus word (Chang et al., 2010; Hodapp & Rabovsky, 2021; Kutas & Federmeier, 2011). This component has been particularly investigated in comprehension of congruent/incongruent words within the context of a sentence, where it has been linked to the word's cloze probability (Pickering & Gambi, 2018). Other (positive) components have also been associated to prediction across language processing (e.g., P300, P600; Alday & Kretzschmar, 2019; Arbel, Spencer, & Donchin, 2011; Van Petten & Luka, 2012). For the present work, we measured prediction and prediction adaptation to less expected lexical choices that were made in a regular context while recording the continuous electrophysiological activity of participants during the whole experiment (Chapter 3).

1.3 From sharing attention to co-representing each other's wor(l)ds

A number of factors influence the way we perceive and represent the world and how we store it in memory. We do not access information about the physical environment passively; rather we pass-filter the surroundings through an active, evaluative process (Norman, 2002; Posner, 1980; Wimmer et al., 2015). Perception is highly subjective, as it is the perceiver's mind that defines and categorizes information as salient and less salient to process. This dynamics of attention can be mediated by previous experience, which gives rise to an "intention to perceive" (Gibson & Rader, 1979), but also by current needs or goals (Elekes & Király, 2021). For instance, the urge to find a small child in a big crowd during a concert would prevent us from noticing that a close friend is waiving right in front of us. Attention combines bottom-up and top-down approaches, as it brings together knowledge and physical input (Elekes & Király, 2021; Serences & Yantis, 2006). In language processing, this explains our ability to pick up a conversation in a loud room, or to disambiguate sentences using the context. Importantly, attention responds to social purposes, as the presence of another person attentionally ’shaping’ the environment has an impact on the way we conceptualize and process the surroundings (Freeth, Foulsham, & Kingstone, 2013; Gallup, Chong, & Couzin, 2012; Risko & Kingstone, 2011).

From our ancestors we have inherited what has been called an ‘altercentric cognition’ (Kampis & Southgate, 2020). In addition to the self-related perception, we possess an other-centered perception, corresponding to a way of processing the objects and events of the world which takes into consideration our fellows’ perspective (Bråten, 2007; Elekes & Király, 2021; Southgate, 2018). This cooperative behavior is proper of the human nature and is increasingly present in young children from the first year of life (Tomasello, 2009; Tomasello, Carpenter, Call, Behne, & Moll, 2005). The presence of another person and even simply the belief of the presence of other person has proven to have an impact on the way we perceive, represent and encode information, often resulting in a significant processing/learning boost (Shteynberg, Hirsh, Bentley, & Garthoff, 2020). This is because, above anything else, it orients our own attention, diverting it or intensifying it. In cognitive psychology, attention is a mechanism which guides and filters information in the environment (Lu, 2008; Posner, 1980). Attended information receives more processing as compared to elements of the environment falling outside attention. In its social dimension, the attentional field is enlarged by having multiple people attending together (Boothby, Clark, & Bargh, 2014; Mundy & Newell, 2007).

One of the theories exploring the social dynamics of attention is the shared attention theory (Shteynberg, 2010, 2015, 2018). The term ‘shared attention’ is used in the literature to indicate a psychological state in which personal and other perspectives overlap, and the self experiences the world from a social point of view. When tuned with the other’s perspective, experience is reinforced and co-attenders become sensitive to the perspective their partners take on the stimuli present in their environment. Something which is relevant to the partner becomes, consequently, relevant to the self to a certain extent. Crucially, under shared attention information receives deeper cognitive processing and further prioritization as compared not only to irrelevant information, but also to information attended alone. Shared attention has been associated to several empirical benefits, including perception and judgment (Tosi et al., 2020; Conson et al. 2019, Bockler 2011, 2013), motivation (Shteynberg & Galinsky, 2011), memory (Elekes, Brody, Halász, & Király, 2015; Elekes & Sebanz, 2020; Eskenazi et al., 2013), emotion and affection (Shteynberg et al., 2014) and, ultimately, action performance.

Following this, sharing attention can be interpreted as a building block to achieve joint action. Here we propose that the socio-cognitive state associated to it leads to spontaneous task co-representation, a crucial mechanism underlying joint action (Böckler et al., 2012; Tsai, Kuo, Jing, Hung, & Tzeng, 2006). Co-representation refers to the ability proper of co-actors to represent each other's actions, eventually supporting the formation and subsequent achievement of shared goals via mutual coordination (Sebanz et al., 2003). It has been suggested that partners co-represent each other's actions without particular effort and in a relatively automatic manner (for review, e.g., Sebanz & Knoblich, 2009; Wenke et al., 2011). Co-representation ranges from low cognitive levels such as motor movements to higher cognitive levels, including language processing, when monitoring partner's language behavior. Critically, co-representation does not only emerge as a consequence of action perception, but also as a result of the mere belief that those actions are to take place. In the present thesis, we tested empirically to what extent shared attention acts at a top-down level in social interaction, affecting the way relevant language features are processed. Our experimental designs allowed us to differentiate between the effects of shared attention when partners are assigned to the same (Chapter 3 and 4) and to different tasks (Chapter 2). We then explored co-representation (as established by different degrees of shared attention) shaping comprehension (Chapter 3) and production (Chapter 2 and 4).

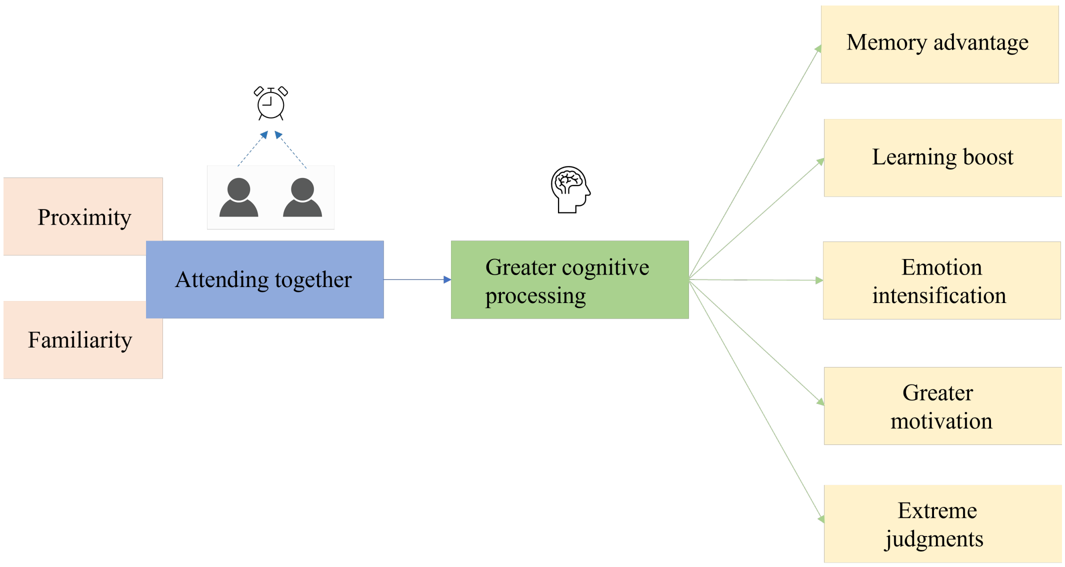

Figure 1.2: **Model and empirical benefits of shared attention.**
Adapted from Shteynberg (2015).

1.4 Adaptive prediction

1.4.1 The motor resonance in speech perception

Linked to co-representation is the notion of prediction (Pesquita, Whitwell, & Enns, 2018). Activity partners benefit from co-representing their co-actor's task in a way that they become able to predict his actions and integrate them in their own action planning and performance to facilitate coordination (Knoblich et al., 2011; Vesper et al., 2017). An explanation to this is linked to the "motor resonance" that takes place during action observation (Uithol, van Rooij, Bekkering, & Haselager, 2011), supporting crucial mechanisms such as learning-by-imitation (Paulus, Hunnius, Vissers, & Bekkering, 2011). Observing a partner's action activates the corresponding motor planning and execution of the targeted action, leading to an "action tendency" to perform the same action (Prinz, 1990; Sebanz, Knoblich, & Prinz, 2005). The engagement of the motor system during action perception has been traditionally established by the mirror neuron system (MNS), a neural *reality* in which production and perception are inherently linked (Rizzolatti & Craighero, 2004; Salo, Ferrari, & Fox, 2019). First shown in the brain of macaque monkeys (di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996), the MNS has rapidly found its counterpart in humans (Gallese, Keysers, & Rizzolatti, 2004). Physically, it corresponds to a group of neurons localized in the frontal lobe, and which modulate their activity both when an individual performs an action and when he observes the action being performed by other individuals (Salo et al., 2019). Crucially, this automatic motor resonance has a functional significance in joint action, supporting action understanding.

A similar pattern is applicable to language and, in particular, to conversation, itself understood as a type of joint action in which conversational partners alternate between speaking (action) and listening (action perception; Clark, 1996; Garrod & Pickering, 2004). Accordingly, the motor circuits that are elicited in the articulation of speech sounds are also active during the perception of the same speech sounds (Michaelis, Miyakoshi, Norato, Medvedev, & Turkeltaub, 2021; Pulvermüller & Shtyrov, 2006). Researchers have shown that listening to speech generates motor resonance of the relevant articulatory gestures corresponding to the production of the speech sounds (Skipper, Devlin, & Lametti, 2017). In a series of transacranial magnetic stimulation (TMS) studies, it has been shown how somatosensory stimulation to areas of brain involved in specific articulatory movements affects language processing. Fadiga, Craighero, Buccino, and Rizzolatti (2002) revealed high motor-evoked potentials coming from listeners' tongue muscles when listening to syllables that involved tongue mobilization. Similar findings were replicated by Watkins, Strafella, and Paus (2003) for lip-based phonemes. More recently, D'Ausilio et al. (2009) revealed how cortical areas of the premotor cortex involved in the articulation of the lips/tongue can be controlled to generate faster or slower receptive brain responses to phonemes with congruent/incongruent places of articulation. In the present dissertation, we got inspiration from the motor resonance theory as the first, clear, empirical demonstration of the strict coupling between production and comprehension. In our studies, however, our focus was on higher stages of language production (lexical access, semantic and conceptual representation), and how those come to be co-represented by language partners.

1.4.2 Forward models mediate language comprehension

The theories as well as the empirical studies mentioned in the previous section highlight the involvement of production mechanisms during comprehension at the phonetic level, eliciting corresponding articulatory patterns, as suggested by the motor theory of speech perception (Hickok, 2012; Pulvermüller & Shtyrov, 2006). However, language is represented and, therefore, potentially predicted, at multiple levels of representation, including phonological, lexical, semantic, and syntactic. This aspect of prediction is presented by Pickering and Garrod (2013a) in their integrated theory of language production and comprehension, which draws on Wolpert (1997)'s computational model for motor control. Accordingly, when comprehending speech, listeners generate production intentions that draw them to covertly imitate their interlocutors by passing by the same dynamics they would go through if they were speaking themselves (Dell & Chang, 2014; Pickering & Garrod, 2013b). Comprehenders use forward models to predict upcoming speech, then compare the predicted percept formulated by the internal model to the actual percept (Pickering & Garrod, 2014b). In the case when the predicted percept corresponding to the expected language realization does not match the actual percept, a prediction error is sent back as efference copy to the action command, which will then integrate the error and, eventually, update the model. In sum, prediction is what supports robust and efficient perception (Kleinschmidt & Jaeger, 2015).

Contextual predictability influences numerous language processes, including sentence comprehension and lexico-semantic processing (Kuperberg & Jaeger, 2016; Ness & Meltzer-Asscher, 2021). Recent findings have highlighted how the predictability about an upcoming word is negatively correlated with 1) the processing time and

2) the neural activity associated to processing it. Behavioural studies such as naming and lexical decision show that reaction times are faster for predictable (or frequent) words compared to unpredictable (or unfrequent) words (Almeida et al., 2007; Navarrete, Basagni, Alario, & Costa, 2006; Oldfield & Wingfield, 1965). Eye tracking experiments indicate that fixations are shorter for predictable words (Altarriba, Kroll, Sholl, & Rayner, 1996; Rayner, 1998; Staub, 2015; Staub & Benatar, 2013). Similar, electrophysiological methods reveal how the amplitudes of ERP components related to prediction such as the N1 (linked to early auditory processing) and the P200, P300 and N400 (linked to lexico-semantic processing), are larger for surprise or unexpected words (Astheimer & Sanders, 2011; Dikker & Pylkkanen, 2011; Fjaellingsdal, Schwenke, Ruigendijk, Scherbaum, & Bleichner, 2020; Nieuwland, 2019; Nieuwland & Van Berkum, 2006). Habituation over repetition of the same word in the same context or the same lexico-semantic pattern can eventually adjust the behavioral and neural reactions via an update of the model which will in turn treat originally unpredictable words as predictable to ease and optimize comprehension (Ness & Meltzer-Asscher, 2021). In our second study (Chapter 3), we monitored prediction and prediction adaptation towards a robot's lexical choices produced in regular contexts measuring the electrophysiological activity of human participants.

1.5 Adaptive production: Linguistic alignment

One of the main characteristics of language is that it is enriched by our capacity to absorb relevant linguistic information from the environment and replicate it. During dialogue and conversation as well as any type of joint language production, interlocutors *align* on multiple linguistic levels, assuring a smooth and successful communication (Garrod & Pickering, 2004; Pickering & Garrod, 2006). The parity of representations between production and comprehension generates mutual priming between speakers and listeners, triggering alignment of relevant linguistic features (Menenti, Pickering, & Garrod, 2012; Pickering & Garrod, 2004a). While listening to our conversational partner, we come to establish common ground, and co-represent the different language dimensions (lexical, grammar, meaning...): This will not only update the way we perceive his speech but will also affect our own speaking. This mechanism occurs in many aspects of human social life more generally, in which the perception of an action influences future actions, leading to imitative and complementary behaviors (Dijksterhuis & Bargh, 2001). In this sense, alignment is what makes language an effective tool to reach social coordination (Clark, 1996; Tylén, Weed, Wallentin, Roepstorff, & Frith, 2010).

Linguistic alignment takes place when conversational partners *entrain* towards the linguistic features of their partner's utterance. In a study by Branigan, Pickering, and Cleland (2000), pairs of participants took turns to describe picture cards to each other. One speaker was a real participant, while the other was a lab confederate, who produced scripted descriptions, which varied in syntactic structure. They found that naive participants tended to adopt the same syntactic structures just used by the confederate, even in the absence of the same content words. This is only an example of linguistic alignment, affecting the syntactic level (see also, Kempen et al., 2011; Levelt and Kelter, 1982; Cleland and Pickering, 2003), which well depicts how adaptation is present in production. Beside syntax, interlocutors progressively converge in terms of phonetic categorization (Pardo, 2006), accent and speech rates (Giles, Coupland, & Coupland, 1991), and in terms of words (Branigan et al., 2011; Brennan & Clark, 1996), eventually associating their linguistic choices to specific contexts and conceptualizations (Garrod & Anderson, 1987). Additionally, alignment can percolate from one level of representation to another. For instance, alignment at the lexical level can elicit/increase alignment of sentence structures, up to situation models (Menenti et al., 2012; Zwaan & Radvansky, 1998). In the present thesis, we focused on two types of linguistic alignment, namely lexical and conceptual alignment, and monitored their emergence using the joint action paradigm exposed in the previous sections (Chapter 4).

1.5.1 Lexical and conceptual alignment in joint production

Lexical alignment consists in the adoption of the interlocutor's words to refer to a particular object or situation (Branigan et al., 2011; Brennan & Clark, 1996; Foltz, Gaspers, Thiele, Stenneken, & Cimiano, 2015; Pickering & Garrod, 2006; Reitter & Moore, 2014). For example, if during an exchange about a children party between Carlo and Lina, Lina uses the word *bonbon*, it is likely that Carlo would say *bonbon* instead of *candy* to refer to the sweet in his upcoming speech. Lexical alignment is supported by bottom-up priming as well as top-down cognitive processes such as perspect-taking and mentalizing (Branigan et al., 2011). When hearing the dispreferred name *bonbon*, bottom-up processes increase the activation of this less frequent lexical item to the level of the more frequent *candy*. Lexical alignment constitutes a basic level of alignment, as it consists into the copying of the words heard while conversing. However, interlocutors are also able to construct and align their respective mental models of the specific situation (i.e., *situational models*, Zwaan & Radvansky, 1998). For any type of joint task, individuals come to share their conceptual representations to benefit from social coordination (Tomasello et al., 2005; Fusaroli et al., 2012). In language, we refer to this mechanism as 'conceptual alignment': it implies a higher, more abstract cognitive level of shared representation as speakers come to progressively share the same conceptual knowledge and conversational schema (Garrod & Anderson, 1987). At the lexical level, conceptual alignment can be understood as a historical explanation to lexical alignment (Brennan, 1996). In a first stage, interlocutors come to achieve conceptual pacts, or shared conceptualizations. Afterwards, they start marking it by using the same names. In this sense, as Brennan and Clark (1996) point out, "while lexical variability is high between conversations, it is relatively low within a conversation" (Brennan, 1996). A characteristic of this type of alignment, however, is that speakers do not simply copy the same lexical items, but rather they acquire knowledge of the lexico-semantic pattern proper of the interlocutor and adapt to it.

1.5.2 Alignment and *belief*: the unmediated and mediated approach

The fact that alignment consists into a linguistic priming between interlocutors on multiple levels of language representation has therefore been well established in the literature (Costa, Pickering, & Sorace, 2008; Pickering & Garrod, 2006). However, why people align is still object of debate (Schober & Brennan, 2003). There exists two main theories, corresponding to the mediated and unmediated account of alignment (Branigan et al., 2011). According to the unmediated account, alignment is an automatic response of the language processing system to the interlocutor's linguistic behavior (see also Pickering & Branigan, 1998). As a consequence, it is not affected by non-linguistic factors, and is a simple result of priming, or previous exposure to related stimuli. The mediated account, on the other hand, suggests that alignment is affected by speaker's belief about the interlocutor. In this sense, alignment is a type of partner-adaptive behavior (Dubuisson Duplessis, Langlet, Clavel, & Landragin, 2021). Interlocutors adopt a specific word if they believe it is appropriate not only in the context but also with respect to the person they face (Bell, 1984; Brennan & Hanna, 2009). This account relates to the 'audience design' (Clark & Schaefer, 1987, p. 209) and to the 'communicative design' (Branigan et al., 2011), for which people tend to align more to interlocutors they believe with less communicative abilities, such as children, L2 speakers and, ultimately, non-human agents. Regarding non-human agents, Branigan et al. (2011) found that, while people reproduced more or less extensively the disfavored names used by their human and artificial interlocutors, they aligned more when this was a computer, and even more when they believed the computer was a less sophisticated machine.

The present dissertation combines the two accounts, treating them as complementary and not mutually exclusive theories. On one hand, alignment has an automatic component, represented by the fact that people are usually not aware to align and they align almost systematically. For instance, interlocutors copy atypical lexical responses such as less used synonyms (Brennan & Clark, 1996) and infrequent syntactic structures such as passives (Bock, 1986), thus overriding more frequent verbal behaviors. In addition, studies on 'atypical' populations suggest that it is so rooted in the communication dynamics that is one of the few social mechanisms which come to be spared in patients with psychiatric disorders such as schizophrenia, bipolar disorders (Sharpe et al., 2022) and autism (Branigan, Tosi, & Gillespie-Smith, 2016; Hopkins, Yuill, & Keller, 2015; Nadig, Seth, & Sasson, 2015; Vollmer, Rohlfing, Wrede, & Cangelosi, 2015). On the other hand, alignment can also be mediated by the type of interlocutor (audience design perspective). In particular, priming is modulated by higher-level beliefs, such as speaker's knowledge about the conversational partner (communicative design perspective; Suffill et al., 2021).

1.6 Social robots for language adaptation

The new challenges of modern society, in which technology has taken the lead in a number of public and private contexts and everyday activities, have brought researchers and engineers from different disciplines to equip computers and robots with cognitive skills that would allow them to socially interact with other humans (Breazel, 2004; Wiese et al., 2017; Yang et al., 2018; Kirtay, 2020; Wudarczyk et al., 2021; Belhassen, 2022). Human-robot interaction (HRI) has thus become a dominant topic in joint action research (Belhassein et al., 2022; Cooper et al., 2020), and different aspects of this interaction have started to be object of research (Kirtay et al., 2020). Co-representation in human-robot interaction has been confirmed at the motor (Cooper et al., 2020; Liepelt & Brass, 2010; Müller et al., 2011), and at the social level (Cross et al., 2019). A robot's physical presence, in particular, has proven to enhance *social presence* (i.e., the feeling of "being there" with a real person, Jung, Kwan, & Lee, 2004; Oh, Bailenson, & Welch, 2018), in a way that it positively affects human belief about the robot's capabilities and intentions, promoting and facilitating interaction (Bainbridge, Hart, Kim, & Scassellati, 2011). Humanoid robots, in particular, offer a perfect compromise between ecological validity and controlled settings. Those artificial agents allow for manipulation of diverse levels of behavior (motor, gesture, language), but they also bring along a human-like appearance and a real presence (Wykowska et al., 2016). Their implementation for shared experimental tasks responds to the requirements of second-person approach of social interaction (Schilbach et al., 2013).

Spoken language interaction is a powerful tool to investigate social cognition, as language constitutes the basic communicative channel promoting coordination among individuals (Fitch, Huber, & Bugnyar, 2010; Mercer, 2016; Tylén et al., 2010). Robots have been employed to monitor some important mechanisms supporting joint language production in typical human-human interaction, including alignment. Humans have been shown to align with their artificial partner in terms of their speech rates, speaking slower or faster according to whether listening to a slow or fast speaking machine (Bell et al., 2003), but also their prosody (Suzuki & Katagiri, 2007), their lexical choices (Brandstetter & Bartneck, 2017; Brennan, 1991; Iio et al., 2015) and synctactic structures (Branigan et al., 2003). Those studies suggest that alignment in HRI is not only preserved, but, possibly, reinforced (see Branigan et al., 2010 for overview). To mediate the extent to which people adapt to their artificial partner there is the belief they have about who they interact with (Branigan et al., 2011), leading them to reserve higher attention and monitoring to coordinate with someone (or something) never encountered before. Overall, those studies provide evidence that an artificial language partner can impact human verbal behavior and induce alignment of low-level representation features in human-robot interaction (e.g., lexical, syntactic, prosodic).

As introduced in previous sections, what is crucial to achieve joint actions is the ability to mutually share the mental states, coordinating at higher levels of representation. Similarly, what makes conversation successful is the speakers' ability to align at the semantic level, and ultimately, conceptually. At this stage, mere automatic copying behavior is discarded as an explanation of alignment which, in turn, must be supported by higher-up processes, such as perspect-taking or mentalization (Frith & Frith, 1999). The recent work by Wudarczyk, Kirtay, Pischedda, et al. (2021) found traces of co-representation in a human-robot task at the level of conceptualization. Co-representing the semantic level of words with a robot sped up the naming performance. In the present thesis, we explored this aspect in relation to lexico-semantic representation, possibly leading to conceptual alignment (Chapter 3 and 4).

1.6.1 Furhat Robot

While some theories on HRI claim that making human-like robots might hinder the interaction between a human and a robot and lead to a feeling of strangeness (the so called 'uncanny effect', S. Wang, Cheong, Dilks, & Rochat, 2020), the increasing interest in the interaction between humans and computers has brought researchers to adopt sophisticated machines to encounter the need for simulating more ecological settings. Social robots such as Pepper and Nao, both issued by the Softback Robotics, have become a successful tool in the interaction research. Those robots are equipped with social skills (from there the name) that allow them to communicate with people in an extremely natural way, alternating both verbal and non-verbal (gestural, emotional) signals (see Breazeal, Dautenhahn, & Kanda, 2016 for an overview on social robots).

Figure 1.3: **Furhat Robot.**

Photo taken from the official website of the Furhat Robotics company: www.furhatrobotics.com.

Furhat Robotics (see Figure 1.3) has been classed to pertain to this kind of machines. It is a product of the "conversational AI and social robotics" startup, founded in Stockholm in 2014, whose primary aim was that of giving birth to a proper 'ghost in the machine', a social robot with a personality and an ability to embody emotions in a conversational interaction. What makes Furhat a real conversational partner to its interlocutors lies in its sophisticated back-projection system with a 3D-printed mask which can resemble anyone's face. In particular, its set up was built to control its facial expressions (including eye brows, mouth, lips) to make its movements smooth and natural. The robot responds to two of the main facial gestures judged to be fundamental in interaction and conversation for the process of building prediction, namely lip-to-word synchronization and gaze direction (Moubayed, Skantze, & Beskow, 2013).

In the present work, we used Furhat to explore co-representation and adaptive prediction (Chapter 3) as well as alignment (Chapter 4) in language processing. Its use was crucial for two reasons. The first reason was methodological. A joint action setting with two human participants hardly permits to address specific brain underpinnings in adaptive prediction, as it is difficult to execute any control over timing and content of responses as well as over the acoustic structure of the speech sounds produced by the human confederate. By contrast, this was perfectly feasible with an artificial partner, which was programmed to produce pre-recorded stimuli at controlled time. The second reason was empirical and theoretical: by using a robot, we tried to establish to what extent predictive and imitative processes are elicited for what concerns specific linguistic patterns, as the lexico-semantic choices.

1.7 Summary of the present studies

The present thesis contains three main empirical investigations that were conducted via four experiments: three joint experiments and one individual control experiment (see Figure 1.4 for an illustration of the thesis's structure). All studies have in common the use of the joint action paradigm to explore different aspects of the adaptive dynamics characterizing language processing and memory in a social context. First, we characterized shared attention on processing relevant linguistic features at different levels of representation (semantics, phonetics; Study 1, see Chapter 2). We asked whether people are affected by the social presence of a partner in the first place, and whether this is eventually manifest in the way they perceive and memorize language. This allowed us to monitor how people come to co-represent a partner's task and his language behavior. In our second study (Chapter 3), we applied this mechanism to human-robot interaction and asked whether co-representation is also present when interacting with a robot (Baus et al., 2014). In the same study, we examined how people learn to predict idiosyncratic lexico-semantic choices, that in our case were made by a robot. Finally, in our third study (Chapter 4), we investigated lexical and conceptual alignment to those choices, therefore turning to the production level.

In the first study, we developed a joint language categorization task in which pairs of participants categorized pictures of objects according either to animacy (semantic task) or to the first letter/phoneme corresponding to the name of the object (phoneme monitoring task). The attentional split we performed in the design allowed us to explore the effects of **shared attention** on processing (via analyses of reaction times) and encoding (via a surprise recall test) of relevant linguistic features at separate representation levels. We modulated the focus of attention, having participants respond either alone, together, or not respond at all, and asked whether attending/responding together versus alone affects how language processing and memory are affected. In addition, we investigated whether this mechanism is mediated by the type of processing level (production stage) at stake. This study included, additionally, an individual control experiment. The reason for that was to isolate possible social factors (shared attention, co-representation) from the performance.

Our second and third study shared a similar experimental design, which consisted in a joint picture-naming task performed with a **social robot**. Robot's responses were manipulated to have him produce the semantic category name instead of the basic-level name of objects (e.g., *fruit* for the picture of a pear) for a regular subset of trials (e.g., all the fruits). In other words, we created an idiosyncratic lexico-semantic pattern in the robot's responses. In the EEG study, we tested 1) **co-representation** in terms of comparable lexicalization processes when preparing to speak and when the robot is expected to name the picture and 2) adaptation to unexpected and less frequent linguistic choices, indicated by a possible reduction of the amplitude of relevant ERPs linked to semantic **prediction** (e.g., P300, N400, P600) over the course of the task. In our third study, robot and participant shared the same categories. We recorded participant's responses and analyzed them to see whether they would start to increasingly say the category name instead of the basic-level name in the same context as the robot (**conceptual alignment**).

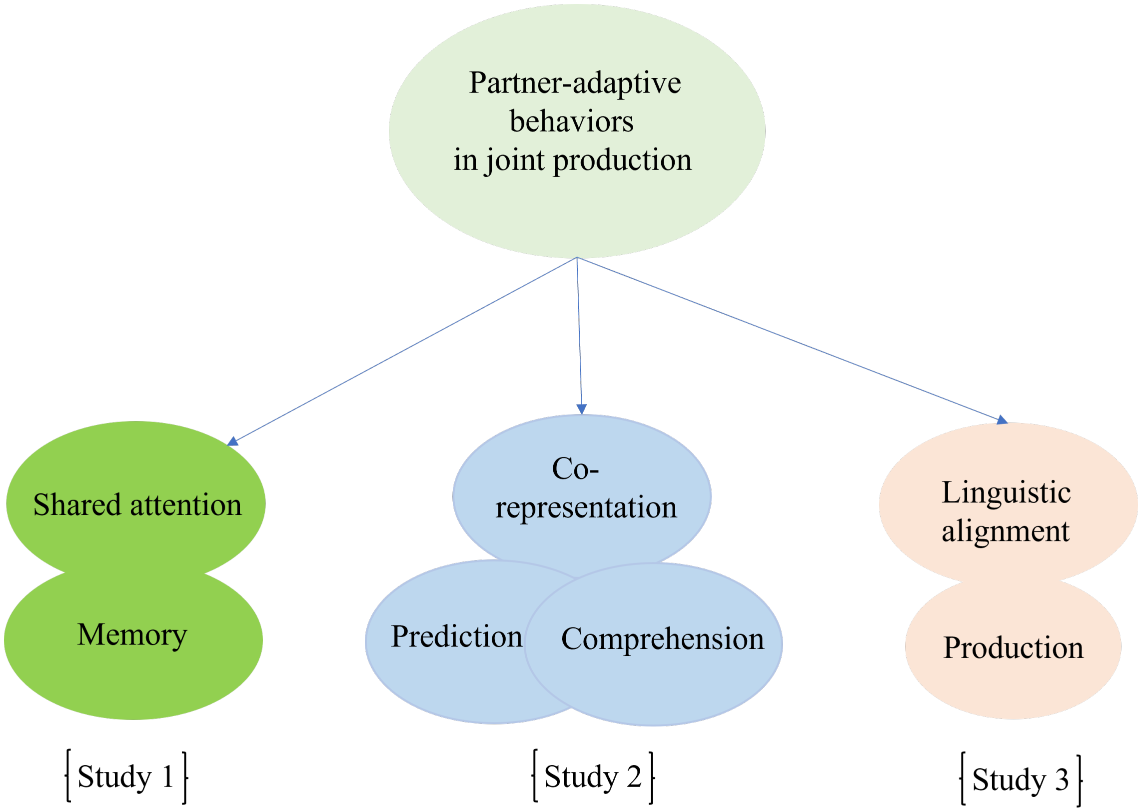

Figure 1.4: **Thesis structure for the three experimental investigations.**

The figure summarizes the three studies which all try to explore the partner-adaptive behaviors in joint language production. Each empirical investigation was built based on a theoretical concept: Shared attention (study 1), co-representation (Study 2) and linguistic alignment (study 3). Additionally, each experiment focused on a specific cognitive process, or else memory (study 1), prediction supporting comprehension (study 2), and production (study 3).

Chapter 2

Study 1: Shared attention in language processing

Article 1: Effects of shared attention on joint language production across processing levels

Cirillo, G., Strijkers, Kristof, Runnqvist. E., Nguyen, N., Baus, C.,
submission for *Language, Cognition, Neuroscience*

Abstract

Shared attention across individuals is a crucial component of joint activities, modulating how we perceive relevant information. In the present study, we explored shared attention in language production and memory across separate representation levels. In a shared go/no-go task, pairs of participants responded to objects displayed on a screen: One participant reacted according to the animacy of the object (semantic task), while her partner reacted to the first letter/phoneme (phoneme-monitoring task). Objects could require a response from either one participant, both participants, or no-one. Only participants assigned to the phoneme-monitoring task were faster at responding to joint than to alone trials. Results from a memory recall test showed that for both partners recall was more accurate for those items to which the partner responded (relative to those responded by no-one), but even more so for jointly responded items. Our findings suggest that shared attention boosts language processing and encoding, but also that shared attention in language is selective and depends on the linguistic feature a partner attends to.

Keywords: Shared attention; Co-representation; Joint Memory Effect; Language production

2.1 Introduction

Shared attention across individuals is a crucial component of any joint activity. Partners carrying out a joint activity (e.g., playing a tennis match, collaborating on a work project, conversing; Sebanz et al., 2006) automatically attend together to events and objects in the common environment. This shared attention has an impact on the way information is perceived, represented, and encoded. The mere fact of attending together has an impact at different cognitive levels, quantifiable in 1) better recall memory (Eskenazi et al., 2013), 2) higher affective intensity (emotions are amplified, Shteynberg et al., 2014), 3) strengthening of motivation and behavioral learning (Shteynberg, 2015). The 'we-mode' (Gallotti & Frith, 2013) or 'first-person plural perspective' (Hirst, Yamashiro, & Coman, 2018) is therefore responsible for a large range of actions and cognitive processes. The cognitive mechanisms of shared attention in joint actions have been investigated in experimental settings such as the task-sharing paradigm (Wenke et al., 2011). In general terms, this experimental manipulation consists in having dyads of participants carrying out complementary tasks, as in a common double game. One binary choice task (e.g., pressing 'E' when seeing an even number, pressing 'O' when seeing an odd number) is split between participants, with each responding to only one property of the stimulus (e.g., the odd or the even number). Studies comparing isolated and joint performance have shown how people acting together influence each other's task performance (for an overview, see Knoblich et al., 2011). This phenomenon has been explained by referring to a co-representation process, according to which actors constantly co-represent their partner's task and incorporate it as if it was their own (Atmaca, Sebanz, & Knoblich, 2011; Sebanz et al., 2005; Vlainic, Liepelt, Colzato, Prinz, & Hommel, 2010). As such, items falling under the partner's attention undergo a deeper cognitive processing than those falling outside.

Eskenazi et al. (2013) showed that shared attention in the joint action context impacts memory recall of partners. Pairs of participants performing a semantic categorization test were exposed to a stream of words belonging to three semantic categories (animals, plants, and objects). Each participant received specific instructions to press a button to only one out of the three categories. Memory performance was evaluated at the end of the experiment when participants were asked to complete a free recall test. Two main results were obtained: 1) higher recall rates for those words to which participants had responded to as compared to other trials (self-prioritization account; Turk et al., 2008). Second, as an indication of shared attention they showed higher rates of recall for the partner's targets (other-go trials) as compared to those words belonging to the category none of the two participants had to react to (no-go trials). Researchers called this effect the Joint Memory Effect (JME), indicating an advantage in recalling information which has been under the attention of one's partner as compared to irrelevant information (no-go trials). In a more recent study, Elekes and Sebanz (2020) elaborated a social epistemic account of the JME, suggesting that the JME is selective, depending on the type of feature the partner attends to (e.g., semantics but not physical color). Accordingly, partners tend to prioritize the most relevant information falling under the partner's attention, which can be used to establish common ground and is likely to be relevant in the future. In different experiments, Elekes and Sebanz (2020), showed a social memory enhancement when participants were asked to categorize items according to their meaning (semantic category, experiment 1), while no effect was found when they categorized items according to the color presentation (experiment 2).

In the language domain, several studies have used the task sharing paradigm to understand production. In an EEG study by Baus et al. (2014), two participants took turns in naming objects of high or low lexical frequency. The authors found comparable ERP modulations of the frequency effect independently of who was speaking and who was listening, demonstrating that participants predicted the lexical representations of their interlocutors while attending to the same materials. Another study by Kuhlen and Abdel Rahman (2017) found a cumulative semantic effect (i.e., a slowdown in naming latencies after a sequence of semantically-related pictures; Howard, Nickels, Coltheart, & Cole-Virtue, 2006) in the context where dyads of participants performed the task in an alternate way, suggesting once again that speakers use shared lexical representations. Those findings provide empirical evidence of lexical co-representation in shared naming tasks, during which language partners engage in a constant simulation of their co-actor's production plans. However, those studies do not directly put to test the notion of shared attention, as partners are asked to attend and process the same levels of processing.

In this study, we implemented a task-sharing paradigm to monitor the impact of shared attention on language processing. In particular, we asked whether shared attention in language is selective (as described for memory; Elekes & Sebanz, 2020) and depends on the linguistic feature a partner attends to. For this purpose, we developed an object categorization task as in Elekes & Sebanz (2020), where dyads were asked to respond to pictures according to different dimensions. While in their study participants responded to either semantics (language dimension) or color (physical dimension), we split participants' task into two language dimensions. Each participant reacted either to the semantic category (natural/animate vs artificial/inanimate; semantic task) or to the phonological category of the onset phoneme of the picture's name (vowel vs. consonant; phoneme monitoring task; Connine & Titone, 1996; Frauenfelder & Segui, 1989). In each dyad, one participant was assigned to the semantic task and the other to the phoneme monitoring task. They both had to press a button when facing their target items and do nothing for the rest of the trials. Our design was therefore characterized by the following: 1) the stimulus material was shared between participants, 2) each participant performed his own, well defined, independent task, and 3) items varied in whether they required a response of only one participant (go-alone trials), both participants (go-together trials; jointly relevant items) or neither (no-go trials). In this aspect, our design differed from previous research, as it put to test shared attention within different language processes, and considered its impact on jointly relevant items (go-together trials).

By comparing trials in which both participants responded to and trials for which only one participant was required to respond to, we were able to evaluate the influence of shared attention on the self-prioritization effect (Turk, Cunningham, & Macrae, 2008). We predicted that the fact of performing the task with a partner facilitates performance in those trials where participants shared the target stimuli (for instance an animal with a vowel-initial name for those participants that had animate/vowel as instruction) as compared to go-alone trials (see Table 1 and Table 2 for an illustration of the design). This could be explained by a more robust representation of the lexical item, enhanced by the two language features processed together. In addition, by having participants responding to different features of the same object, we were able to test whether the presence of a partner accessing one linguistic dimension which differs from one's own has an impact on language processing, and whether that depends on the type of language representation (semantic, phonetic).

From the language production literature, we know that production entails retrieval of multiple linguistic representations (e.g., semantic, lexical and phonological). While it is open to debate whether lexico-semantic features and phonology of a word are retrieved in sequence (Hickok, 2012; Indefrey & Levelt, 2004) or in parallel (Fairs, Michelas, Dufour, & Strijkers, 2021; Strijkers & Costa, 2016; Strijkers, Costa, & Pulvermüller, 2017), there is little doubt that linguistic decisions are faster for semantic than phonological features (Strijkers & Costa, 2011; van Turennout, Hagoort, & Brown, 1997). In the context of our study, we expected participants assigned to the phoneme-monitoring task to be slower compared to those assigned to the semantic task. This aspect could allow them to incorporate objects' semantic representations to ease the processing time to respond, while for those doing the semantic task, accessing the phonetic level might delay their performance. This could indicate that attending together to the same relevant object can affect language differently depending on the type of processing level targeted and to the different degrees of difficulty associated to each task/processing level. In sum, our study questions whether language co-representation as a consequence of shared attention is present (and traceable in timing performance and recall) when individuals are processing two distinct linguistic properties. We elaborated from Elekes and Sebanz (2020)'s findings to see whether people make a difference in terms of attentional focus from one linguistic feature to another (the selecting nature of the JME is here applied at the language level).

As in Eskenazi et al. (2013) and Elekes and Sebanz (2020), we faced participants with a recall test at the end of the experiment. From what concerns the impact of shared attention on memory, the aim of this design was: 1) to replicate previous findings showing better information retrieval for other-go stimuli as compared to no-go stimuli within a context where two language dimensions are separately evoked, and 2) to assess whether there would be a further advantage in the recall of items to which participants responded together compared to the items to which only one participant responded. We predicted a memory enhancement for words relevant to the self and the partner, as compared to words relevant only to the self, establishing a more fine-grained level to the self-prioritization account (Turk et al., 2008), and eventually show how this self-advantage can be improved by shared attention. Finally, to make our findings unequivocally linked to the social manipulation, we performed a control experiment, where we instructed participants to carry out one out of the two tasks without the presence of a partner, while being aware of the existence of the other task. Our aim was to rule out that any possible effect found in the joint condition was to be attributed to the bias of having the two language sources available rather than to the presence of a partner, and therefore to a social effect of shared attention.

2.2 Experiment 1: Joint language experiment

2.2.1 Methods

In accordance with Open Sciences Practices data, we made our script analyses and materials available in our OSF repository (https://osf.io/mvs8h/).

2.2.2 Participants

Sixty participants (30 pairs, 16 men; age: M = 22.5 years, SD = 3.2, range = 18–35 years) participated in the joint experiment. All participants were native speakers of French with normal or corrected-to-normal vision. None of the participants reported any neurological disorder, psychiatric disorder, or speech/language impairment. In accordance with theories of collective attention suggesting that shared attention is potentially stronger between people who are familiar with each other and are likely to interact again in the future (Shteynberg, 2015, 2018), participants knew each other's before. All participants indicated to be very familiar with their task partner on a 5-point likert scale (M = 4.2, SD = 1.0, range = 3–5). They gave informed consent and received 10 euros or course credits for their participation.

2.2.3 Design and materials

The set of visual stimuli consisted of 88 black and white pictures extracted from Snodgrass and Vanderwart (1980). The pool included 44 pictures depicting living items (e.g., a vegetable), out of which 22 had a name that started with a vowel (e.g., *ail* = garlic) and 22 had a name starting with a consonant (e.g., *carotte* = carrot). The remaining 44 images represented non-living items (e.g., a building), with 22 names starting with a vowel (e.g., igloo) and 22 starting with a consonant (e.g., *pont* = bridge). By combining the semantic conditions (living/non-living) with the two phonological (vowel onset/consonant onset) conditions, we ended up having 4 different task combinations (see Table 2.1).

Task	Semantic	Phoneme-monitoring
1	Living	Vowel onset
2	Living	Consonant onset
3	Non-living	Vowel onset
4	Non-living	Consonant onset

Table 2.1: **Turn conditions**. Four possible combinations of semantic and phoneme-monitoring task per pair of participants.

The joint language task followed a classic social go/no-go design, with three possible turn conditions. A go-alone condition, in which only one participant responded to the picture, i.e., either the participant doing the semantic task or her partner doing the phonological task. A go-together condition, in which both participants responded in the same trial, and a no-go condition for the trials where none of the participants had to respond. If we take the living-vowel onset combination, an example of go-alone trial for the semantic task would be the picture of a *poivron* (pepper), as it represents a living object starting with a consonant, while an example of go-alone trial for the phonological task would be *avion* (airplane), as it represents an artificial object starting with a vowel. In the same combination a go-together trial would be the picture of an *éléphant* (elephant), while a no-go trial would be a *porte* (door), as it is not a living entity nor does it start with a vowel (see Figure 2.1). We created 10 randomized lists of items to avoid any bias due to the effect of order of presentation. Each pair of participants was exposed to one list and carried out one out of the four task combinations. The experiment was divided into two blocks. In each block all the 88 pictures were presented but in a different order, with a total of 1 repetition per image (88 images x 2). We collected a total of 176 object points per participant per experiment or 352 (176 x 2) object points per pair.

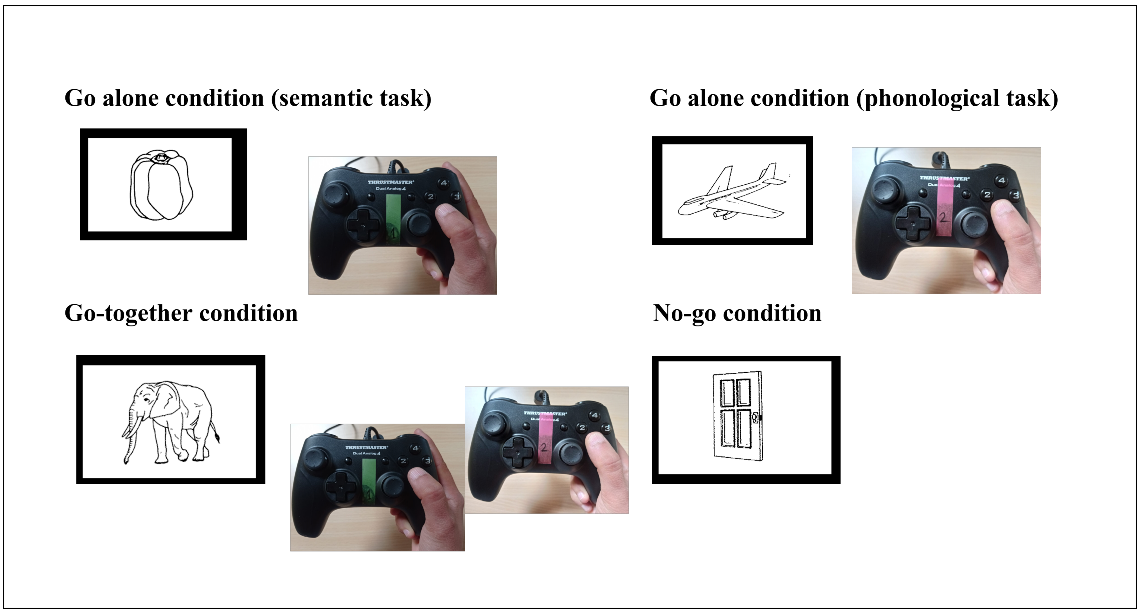

Figure 2.1: Example of turn condition in the living-vowel onset combination. The joystick with the green label was for participants carrying out the semantic task, while the joystick with the pink label was reserved to participants doing the phonological task.

2.2.4 Procedure

Participants were first informed about the study and gave informed consent. In Experiment 1 participants were tested in pairs. They sat side-by-side in front of a computer screen on which the visual stimuli were displayed by means of the Presentation® software (version 18.1, www.neurobs.com). The location of the participant assigned to each of the two tasks -semantic vs phonological - relative to the computer screen (left vs right) was counterbalanced across pairs. Participants were introduced to the experimental procedure by a written set of instructions on the screen. In the instructions, we explained that participant 1 (participant holding the joystick '1') had to respond according to the semantic category of the picture (semantic task), while participant 2 (participant holding the joystick '2') had to respond according to the first letter of the name associated with the picture (phonological task). They were asked to respond by pressing the button placed on the right side of the joystick in case the picture fulfilled their specific task.

Before starting the experiment, each pair went through a familiarization phase, consisting of a passive exposure to all the experimental images with their corresponding written names. This phase took three minutes and allowed us to ensure that the images were not ambiguous and that participants associated the expected name to each picture. Participants then moved to the experimental phase, where pictures were presented without their name. Each trial started with the presentation of an asterisk (for 500 ms), followed by a blank screen (for 500 ms) and by the experimental picture (for 2500 ms; see Fig. 2.2). Participants were asked to make their decision while the picture was on the screen and to respond as accurately and as quickly as possible. We collected their reaction times. The experiment was divided into two blocks of 7 minutes each. At the end of the first block, which comprised 88 trials, participants made a pause. At the end of the experiment, participants performed a free recall test, where they were asked to write down on a blank paper all the items they were able to recall after the categorization task.

2.2.5 Data analysis

One pair of participants was rejected due to the large number of error trials for the participant carrying out the phonological task (N = 58, i.e. 32% of the data points for the task). Thus, the final sample included 29 pairs. Reaction times were fitted to a linear mixed effects model using the lme4 package (Bates, Mächler, Bolker, &

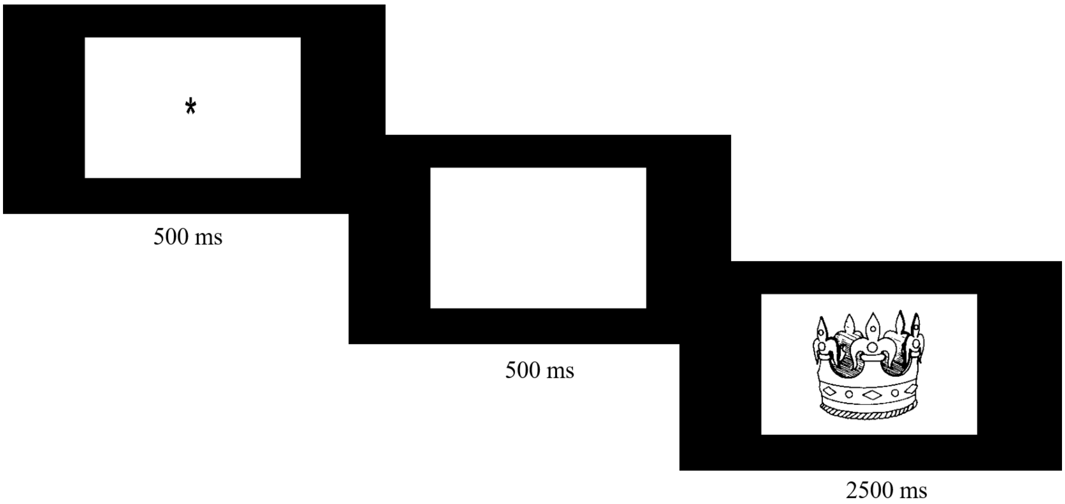

Figure 2.2: Timeline of the experiment (one trial).

Walker, 2015)) provided by the R environment to assess the impact of the fixed factors task (semantic vs. phonetic) and turn (go-alone vs. go-together), with pair and item taken as random factors. All the categorical variables (task and turn) were sum to zero coded. Reaction times were log transformed before analysis to reduce skewness. Missing responses and incorrect responses were excluded from the naming latencies analysis (N = 248). For the simplicity, figure and means reported in the article are not log-transformed. Errors were processed by fitting a logistic regression model, more suitable to binary data. The variable error (0 vs. 1) was taken as response variable, and task (semantic vs. phonological) and turn (go-alone vs. go-together) were the predictors. Pair and item constituted the random factors.

Finally, we performed three separate analyses on the number of words participants recalled in the memory test. First, we compared the number of items that required participants to respond (go trials: go alone and go together) to those items that did not require participants to respond (no-go trials: no-go and other-go) as a general measure. For this, we performed a linear mixed effects model with the fixed factors task (semantic vs. phonological) and turn (go vs. no-go), and random factor subject. Afterwards we replicated the model only to the items that did require participants' response during the categorization task, with the fixed factors task (semantic vs. phonetic) and turn (go-alone vs. go-together). To analyze recall for those items that did not require participants' response we conducted a further analysis with the fixed factors task (semantic vs. phonological) and turn (no-go vs. other-go).

2.3 Results

2.3.1 Reaction times and error rates

The analyses of reaction times revealed a significant effect of task (b = 1.900e-01, s.e. = 3.967e-03, t = 47.89, $p < 0.001$). Participants carrying out the phoneme-monitoring task responded slower than participants carrying out the semantic task (Semantic task: M = 579 ms, SD = 229; Phoneme-monitoring: M = 855 ms, SD = 351). An effect of the social partner was observed in the turn condition (b = 8.901e-03, s.e. = 3.969e-03, t = 2.242, p = 0.02), and, in particular, in the interaction between the two variables (b = 1.086e-02, s.e. = 3.965e-03, t = 2.73, p = 0.006), which revealed that the social effect of responding together was mediated by the type of task or by the type of language representation evoked. Items for the phoneme-monitoring task were named faster in the go-together turn (M = 827 ms, SD = 286) compared to the go-alone turn (M= 883 ms, SD = 401; b = 0.03952, 0.0113, t = 3.495, p = 0.0027), while little difference was spotted for the semantic task (go-alone: M = 580 ms, SD = 251; go-together: M = 577 ms, SD = 206; b = -0.00392, s.e. = 0.0111, t = -0.352, p = 0.9851; see Figure 2.3).

The analysis on error rates revealed a significant effect of task (b = 0.91824, s.e. = 0.13130, z = 6.993, $p < 0.001$). Items from the phonetic-monitoring task were named incorrectly more often (N = 105) than items belonging to the semantic task (N = 18). We did not find any effect of turn (b = -0.17009, s.e. = 0.13221, z = -1.287, p = 0.19), nor of the interaction between the two variables (b = -0.04004, s.e. = 0.13114, z = -0.305, p= 0.76).

2.3.2 Memory Test

Our first analysis comparing overall names recalled with the go/no-go distinction showed that participants recalled more items belonging to the go condition (see Table 2.2. for a descriptive statistics of the recall data, and Figure 2.4 for its illustration) as compared to the no-go condition (b = 1.53, s.e = 0.15, t = 9.63, $p < 0.001$). In addition, participants carrying out the phoneme-monitoring task recalled more items than those doing the semantic task (b = 0.67, s.e. = 0.15, t = 4.25, $p < 0.001$). However, the model did not reveal a significant interaction between the two variables (b = -0.09, s.e. = 0.15, t = -0.6, p = 0.54). The second analysis focusing on the go data (go-alone vs. go-together) revealed an overall effect of turn (b = 0.9, s.e. = 0.2156, t = 4.175, <0.001): items that required a response from both participants were recalled more often than those items that required the response from only one participant. The model revealed

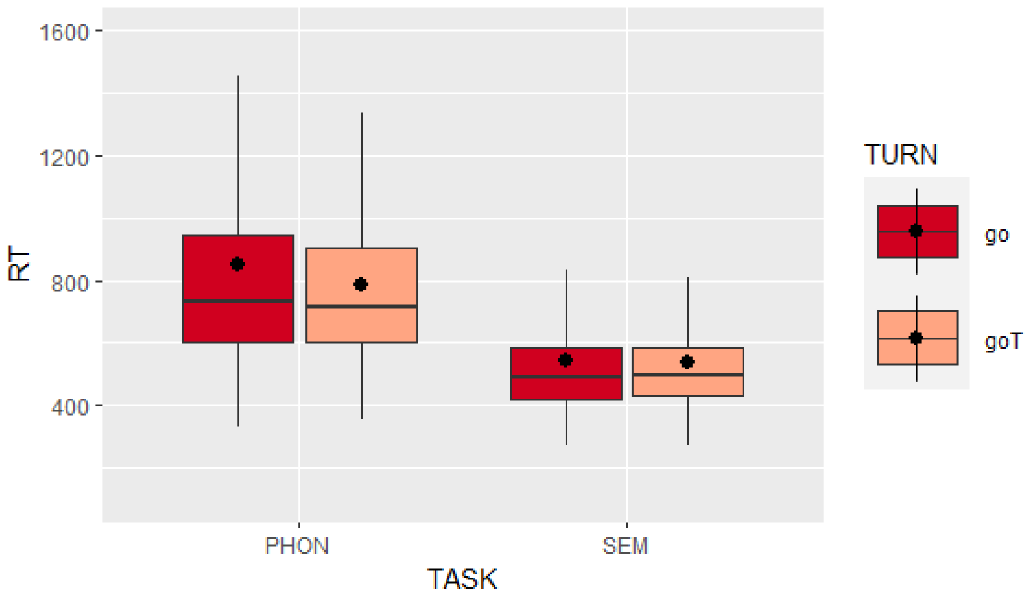

Figure 2.3: **Reaction times results for the joint experiment**. Mean and standard error for reaction times (in ms) plotted by task, where 'SEM' stands for 'semantic task' and 'PHON' for phoneme-monitoring task; and by turn condition. In the legend, the label 'go' corresponds to the individual condition where only one participant responds, while goT (go-together) corresponds to the condition where both participants react to the picture.

an effect of task (b = 0.58, s.e. = 0.21, t = 2.7, p = 0.008), indicating that participants doing the phoneme-monitoring task recalled more items than those doing the semantic task. The model did not find interaction between the variables (b = 0.3, s.e. = 0.21, t = 1.39, p = 0.16). Similarly, our third analysis on the no-go data (other-go vs. no-go) showed an effect of turn (b = 0.7, s.e. = 0.18, t = 3.761, p <0.001), indicating that participants recalled more items from other-go trials than no-go trials. It also showed a comparable effect of task as in the previous analyses (b = 0.77, s.e. = 0.18, t = 4.11, p <0.001) and no interaction (b = 0.07, s.e. = 0.18 t = 0.39, p = 0.69).

	alone	together	other	none
SEMANTIC	M = 7.13 SD = 2.3	M = 8.3 SD = 2.8	M = 5.1 SD = 2.2	M = 3.8 SD = 1.7
PHONEME-MONITORING	M = 7.7 SD = 2.6	M = 10.1 SD = 2.5	M = 6.8 SD = 3.1	M = 5.2 SD = 2.2

Table 2.2: **Descriptive statistics of the free recall test in the joint experiment.** The table summarizes the mean and standard deviation of the number of names correctly recalled by participants doing the semantic task (first row) and participants doing the phoneme-monitoring task (second row) for each condition (go: go-alone, go-together; no-go: other-go, no-go).

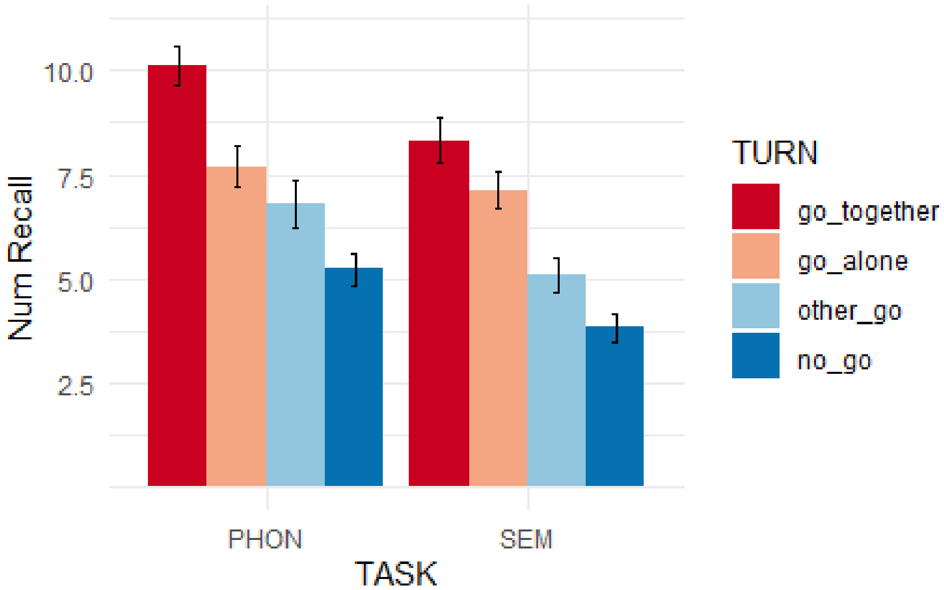

Figure 2.4: **Items recalled in the memory test in the joint experiment.** Number of items recalled by each pair considering the task (SEM= semantic task; PHON = phoneme-monitoring task) and turn condition. In particular, here we considered four levels corresponding to the variable turn, including the go-trials (go-together and go-alone) and no-go trials (other-go and no-go).

Overall, the results of the free recall test in the joint version of the experiment revealed that participants carrying out the phoneme-monitoring task were better at recalling items than participants doing the semantic test, which can be due to level of representation and the difficulty associated to it. In addition, we observed an effect of turn which remained stable and strong across the three analyses. First, participants recalled more items that belonged to the go trials (go-alone + go-together) than the no-go trials (other-go + no-go). Second, they recalled better those items that required two responses (go-together trials) as compared to those that required one response (go-alone trials). Finally, among the no-go trials, we found that participants recalled more items that belonged to the turn of their partner (other-go trials) rather than those requiring no one to respond (no-go trials). We can therefore assume that the presence of a partner resulted in a memory advantage, as items falling (also) under the attention of the partner were remembered more accurately.

Before discussing our results, in a second experiment we ruled out the possibility that - for both reaction time performance and recall - our results were not due to the social presence of a partner, but to the fact of having two cues (semantic and phonetic) to process. That is, the fact of explaining partners which two language features were relevant in the experiment could be driving the increased attention to those features. For that reason, we carried out a control individual experiment, maintaining the exact same structure and giving the same instructions to participants. Participants were therefore aware of the presence of the two language dimensions and were assigned to attend only one of them.

2.4 Experiment 2: Individual language experiment

As a control, we carried out an individual experiment, where we made a new sample of participants perform the same task (doing one of the two categorization tasks) without the presence of a task-partner. This way we were able to monitor participant's response and recall behavior in the absence of shared attention and across the two language channels. We carefully explained participants that the experiment was initially meant for two people and gave them the same instructions as in its joint version. We finally told them they would be assigned to only one task. Script analyses, and materials are available in the same project OSF repository (https://osf.io/mvs8h/).

2.4.1 Methods

2.4.2 Participants

Sixty participants (16 men; age:M = 25.8 years, SD = 5.04 , range = 18–38 years) took part in Experiment 2. None of them had taken part in the first experiment. All participants were native speakers of French with normal or corrected-to-normal vision.

None of the participants reported any neurological disorders, psychiatric disorders, or speech/language impairments. Participants gave informed consent and received 10 euros for their participation.

2.4.3 Design and materials

The experimental material was the same as in the joint experiment.

2.4.4 Procedure

The procedure was the same as in the joint version of the experiment, except that participants came alone and not in pairs.

2.4.5 Data analysis

As in the joint task, we carried out statistical data analyses on the reaction times and error rates of the participants, with the same predictors (task, turn) and random factors (subject, item). Three participants were excluded from the analysis. Two of them made more than 25% of errors, while for the third one, the experimental system crashed twice and could not be recovered. The final sample included 57 participants. Missing responses and incorrect responses were excluded from the naming latencies analysis (N = 243) and included in the error rate analysis. Similarly, we analyzed the results obtained in the memory test via three separate linear mixed effects models. We kept the same turn division, even though the control task had only two real conditions. However, the comparison with the conditions that in the joint experiment represented the presence of another person (other-go and go-together) was necessary for our main objective.

2.4.6 Results

2.4.7 Reaction times and error rates

The results on reaction times replicated those obtained in the joint experiment concerning the task (b = 2.914e-01, s.e. = 2.026e-02, t = 14.381, $p < 0.001$), with participants carrying out the phoneme-monitoring task reacting slower than participants carrying out the semantic task (Semantic task: M = 502.74 ms, SD = 193.14 ms; phoneme-monitoring task: M = 906.49 ms, SD = 348.69). However, the model revealed no significant difference in relation to the condition turn (b = -3.467e-04, s.e. = 3.828e-03, t = -0.091, p = .92), nor a significant interaction between the turn and task (b = 1.125e-03, s.e. = 3.805e-03, t = 0.29, p = 0.767; see Figure 2.5). The analysis on the error rates did not reveal any significant effect of task (b = -0.06, s.e. = 1.83, z = -0.03, p = 0.97), nor turn (b = 0.4, s.e. = 0.64, z = 0.63, p = 0.52) nor any interaction between the two variables (b = -0.56222, s.e. = 0.63610, z = -0.884, p = 0.37).

2.4.8 Memory test

Descriptive statistics for the recall data in the individual condition are summarized in Table 2.3. Our first analysis comparing go and no-go trials revealed a significant effect for turn (b = 0.87, s.e. = 0.14, t = 6.28, p <0.001), indicating that participants overall recalled their items better. It did not show a significant effect for task (b = 0.38, s.e. = 0.2, t = 1.890, p = 0.06) not for the interaction (b = -0.15, s.e. = 0.14 t = -1.1, p = 0.27). Similarly, the other two analyses revealed no significant difference neither between go-alone and go-together trials (b = -0.17, s.e. = 0.17, t = -1.008, p = 0.31), nor between other-go and no-go trials (b = 0.03, s.e. = 0.21, t = 0.16, p = 0.87). Those results highlight once more that playing with the partner does boost encoding, as in the individual version a corresponding memory advantage was not preserved.

	alone	together	other	none
SEMANTIC	M = 7.17 SD = 2.6	M = 7.06 SD = 2.5	M = 5.13 SD = 3.1	M = 4.96 SD = 1.9
PHONEME-MONITORING	M = 7.9 SD = 2.5	M = 7.3 SD = 2.2	M = 6.1 SD = 2.4	M = 6.1 SD = 2.1

Table 2.3: **Descriptive statistics of the free recall test in the individual experiment.** The table summarizes the mean and standard deviation of the number of names correctly recalled by participants doing the semantic task (first row) and participants doing the phoneme-monitoring task (second row) for each 'supposed' turn condition (go: go-alone, go-together; no-go: other-go, no-go).

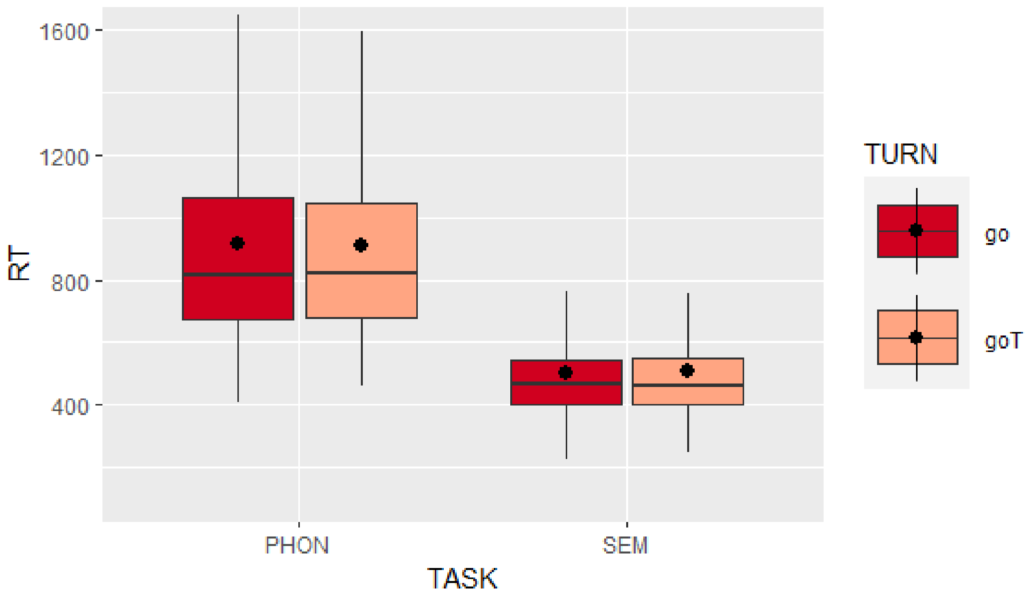

Figure 2.5: **Reaction times results for the individual experiment**. Mean and standard error for reaction times (in ms) plotted by task, where 'SEM' stands for 'semantic task' and 'PHON' for phoneme-monitoring task; and by turn condition. Here the labels 'go' (go-alone) and goT (go-together) in the legend are built based on the previous experiment and on the general instructions, while participants did the experiment individually.

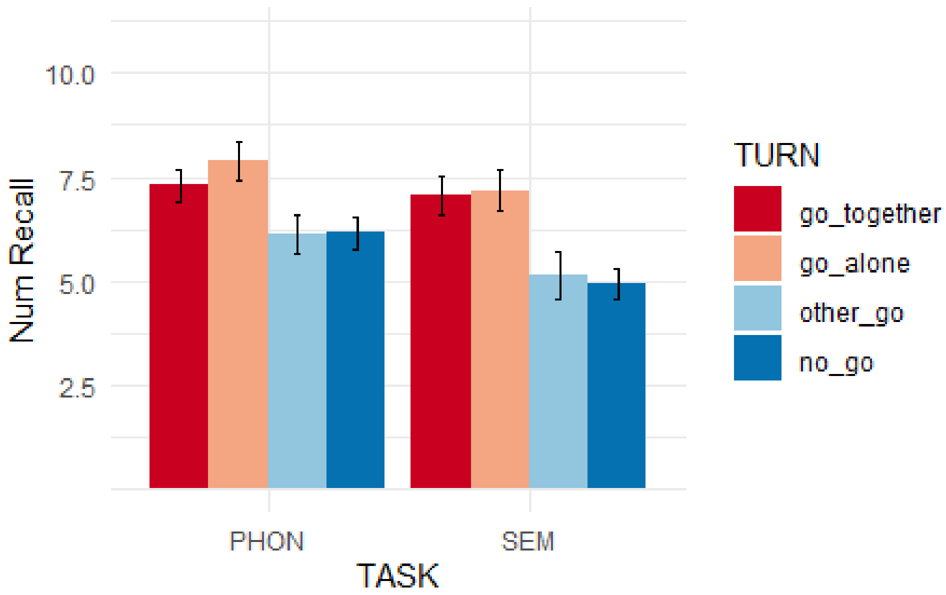

Figure 2.6: **Items recalled in the memory test in the individual experiment.** Number of items recalled by each pair considering the task (SEM= semantic task; PHON = phoneme-monitoring task) and turn condition. In particular, here we considered four levels corresponding to the variable turn, including the go-trials (go-together and go-alone) and no-go trials (other-go and no-go).

2.5 Discussion

Our work aimed to investigate how shared attention affects language processing and memory recall. We implemented a task-sharing experiment where pairs of relationally close people sat side-by-side and looked at objects displayed on a screen, while they carried out two separate tasks: One participant responded to the animacy of the object (semantic task), while her partner reacted to its corresponding first letter/phoneme (phoneme-monitoring task). In terms of design, our study was inspired by previous research works testing the impact of shared attention on memory (i.e., Elekes & Sebanz, 2020; Eskenazi et al., 2013). Importantly, our design differed from those because it 1) included jointly relevant items (i.e., items that required a response from both participants) in addition to items that required the response of only one participant, and items that required no one to respond (Bäss & Prinz, 2014; Pickering, McLean, & Gambi, 2022), and 2) covered two language dimensions that were accessed separately by participants. The choice of having pairs of people that knew each other was made in conformity with Shteynberg (2015, 2018), according to which proximity as well as being in close terms with one another are among the strongest factors to elicit shared attention.

We collected reaction times during the language task as well as written responses participants gave in a free recall test at the end of the experimental session to quantify the impact of shared attention in terms of online processing and memory recall of relevant linguistic properties. We asked whether people benefit from the presence of a co-actor, making faster responses and memorizing more items in trials that required a response from both participants. In addition, having two separate representation levels to process allowed us testing to which extent influences of shared attention are modulated by the specific language representation level at stake. We asked whether people benefit from the presence of a co-actor independently of whether accessing the semantic or phonetic/phonological level, or whether the fact of focusing on one specific level incentives co-representation of the joint action, possibly resulting in a boost in terms of behavioral response and memory encoding.

Our results revealed that participants assigned to the phoneme-monitoring task were significantly faster when responding to trials that required both participants to respond (go-together trials) compared to the trials where they responded alone (go-alone trials). Participants assigned to the semantic task, on the other hand, did not show the same behavioral pattern. Their reaction times were as fast in the go-alone trials as in the go-together trials. In contrast, results from the free recall test revealed that both groups of participants were better at recalling items previously assigned to their partner's task (other-go trials) than no-go trials, and better at recalling items from the go-together trials than go-alone trials. As part of our study, we also run a control, individual condition, which allowed us to rule out the possibility that our results were due to having participants accessing the two linguistic cues/features available from the instructions, rather than from the presence of a co-attending actor. In the present discussion, we try to interpret those findings integrating language processing and joint action through the lens of the shared attention theory.

According to Shteynberg (2010)'s social tuning effect, experiencing something together with relationally close others leads to better memory for the attended object/event. When two friends watch a movie together, for instance, they do not only assume their own perspective of the movie, but they also adopt the perspective of their partner, which, in turn, leads them to focus on details they would have not paid attention to being alone. This aspect is directly linked to the notion of relevance: anything falling under the attention of our partner becomes relevant to us, because we co-represent this information (Boothby et al., 2014; Eitam & Higgins, 2010). This explains why participants prioritized information that was relevant to their partner (items in other-go trials) as compared to irrelevant information (items in no-go trials). The fact that people invest higher cognitive resources to co-attended events (Shteynberg, 2015) was well represented by the higher recalling rate participants reached for the go-together trials, indicating an important memory advantage under social circumstances.

Replicating Elekes and Sebanz (2020), our results revealed that shared attention improves memory recall, so that information falling under the attention of the partner is prioritized. In addition, they show that shared attention is selective, depending on the cognitive process involved. As findings point out, this selective nature of shared attention is modulated by the cognitive task evaluated. Individuals do not automatically co-represent all features attended by a co-actor, but rather they prioritize information that is relevant to retain (i.e., linguistic information but not color), and likely to be brought up in the future (Elekes & Király, 2021; Gallup et al., 2012; Moorselaar & Slagter, 2020; Shteynberg et al., 2020). While we found evidence of shared attention in memory for both groups of participants (i.e., those reacting to the semantic and the phonetic level) in the recall test of the joint condition, only those doing the phoneme-monitoring task registered a social benefit in the object categorization task. A possible explanation to this observation must be searched for among theories on language production. In particular, we bring forward two interpretations, which do not mutually exclude each other, and which refer to 1) the temporal (sequential) dynamics characterizing language production (Levelt, Roelofs, & Meyer, 1999), and 2) the different degrees of difficulty associated to each task.

During verbal communication, people produce and comprehend language across multiple representation levels (Pickering & Garrod, 2013b). Each representation is defined by its temporal characteristics and is associated to a certain processing depth (Hauk, 2016; Indefrey & Levelt, 2004). Retrieving animacy (semantic dimension), for instance, is associated to a more superficial process as compared to retrieving the phonetic/phonological form. To start with, animacy is a universal feature, evolutionally anterior to the phonetic dimension which, in turn, is culturally based (Nairne et al., 2013; J. New, Cosmides, & Tooby, 2007; Rawlinson & Kelley, 2021). Theories of sequential processing on language production suggest that each type of language representation corresponds to a precise temporal dynamics in the brain (Sahin, Pinker, Cash, Schomer, & Halgren, 2009). Speaking or the simple act of generating a word is the result of a staged process, which goes from conceptual preparation to lexical selection, and then down to phonological and phonetic encoding till the final articulation of the word, which is described to occur within 600 msec after picture onset (Dell, 1986; Hickok, 2012; Indefrey & Levelt, 2004; Levelt et al., 1999; Roelofs, 1992). Following this, semantic access comes chronologically before phonological and phonetic access when retrieving a word. In the context of our work, this could mean that those participants reacting to the phonetic features passed through the semantic stage when elaborating the phonological form of the word. The social situation of responding at the same time with a partner attending to the semantic dimension might have brought them to co-represent this more superficial dimension and use it to speed up the performance.

A deeper process such as that of extracting the phonetic form is strictly associated to an increased difficulty in retrieving information as compared to the more superficial task of accessing semantics. This explains why decisions in the phonetic task were slower compared to those in the semantic task (Strijkers & Costa, 2011; van Turennout et al., 1997). In this context, increased task complexity resulted in higher social benefits. This could be that a difficult task, taking more time to be accomplished, leaves more room for effects to be observed. On the other hand, people accessing the semantic level did not need to process the cognitively higher up phonetic level to categorize an object as animate or inanimate. In other words, this group of participants were already the fastest, and could not have done better, as indicated by the results in the individual version of the experiment, in which reaction times for the semantic task were comparable to those obtained in the joint experiment.

In sum, our work outlines how the social effects of shared attention are mediated by the specific linguistic representation at stake. While long-term benefits seem to be reached when co-attending separate linguistic features (i.e., memory advantage), those benefits are not always integrated in response behavior, and are fully dependent on the processing depth associated with the linguistic representation. Deeper processing levels, in particular, benefits from co-representing more superficial dimensions under shared attention.

References

Atmaca, S., Sebanz, N., & Knoblich, G. (2011). The joint flanker effect: Sharing tasks with real and imagined co-actors. *Experimental Brain Research. Experimentelle Hirnforschung. Experimentation Cerebrale*, *211*(3–4), 371–385. https://doi.org/10.1007/s00221-011-2709-9

Bäss, P., & Prinz, W. (2014). My partner is also on my mind: Social context modulates the N1 response. *Experimental Brain Research*, *233*. https://doi.org/10.1007/s00221-014-4092-9

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), Article 1. https://doi.org/10.18637/jss.v067.i01

Baus, C., Sebanz, N., Fuente, V. de la, 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

Boothby, E. J., Clark, M. S., & Bargh, J. A. (2014). Shared Experiences Are Amplified. *Psychological Science*, *25*(12), 2209–2216. https://doi.org/10.1177/0956797614551162

Connine, C. M., & Titone, D. (1996). Phoneme monitoring. *Language and Cognitive Processes*, *11*(6), 635–645. https://doi.org/10.1080/016909696387042

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

Eitam, B., & Higgins, E. T. (2010). Motivation in mental accessibility: Relevance of a Representation (ROAR) as a new framework. *Social and Personality Psychology Compass*, *4*(10), 951–967. https://doi.org/10.1111/j.1751-9004.2010.00309.x

Elekes, F., & Sebanz, N. (2020). Effects of a partner's task on memory for content and source. *Cognition*, *198*. https://doi.org/10.1016/j.cognition.2020.104221

Eskenazi, T., Doerrfeld, A., Logan, G. D., Knoblich, G., & Sebanz, N. (2013). Your words are my words: Effects of acting together on encoding. *The Quarterly Journal of Experimental Psychology*, *66*(5), 1026–1034.

Frauenfelder, U. H., & Segui, J. (1989). Phoneme monitoring and lexical processing: Evidence for associative context effects. *Memory & Cognition*, *17*(2), 134–140. https://doi.org/10.3758/bf03197063

Gallotti, M., & Frith, C. D. (2013). Social cognition in the we-mode. *Trends in Cognitive Sciences*, *17*(4), 160–165. https://doi.org/10.1016/j.tics.2013.02.002

Gallup, A. C., Chong, A., & Couzin, I. D. (2012). The directional flow of visual information transfer between pedestrians. *Biology Letters*, *8*(4), 520–522. https://doi.org/10.1098/rsbl.2012.0160

Gambi, C., Van de Cavey, J., & Pickering, M. J. (2015). Interference in joint picture naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(1), 1–21. https://doi.org/10.1037/a0037438

Hauk, O. (2016). Only time will tell—Why temporal information is essential for our neuroscientific understanding of semantics. *Psychonomic Bulletin & Review*, *23*(4), 1072–1079. https://doi.org/10.3758/s13423-015-0873-9

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, *13*(2), 135–145. https://doi.org/10.1038/nrn3158

Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: Experimental and computational studies. *Cognition*, *100*(3), 464–482. https://doi.org/10.1016/j.cognition.2005.02.006

Indefrey, P. (2011). The spatial and temporal signatures of word production components: A critical update. *Frontiers in Psychology*, *2*, 255. https://doi.org/10.3389/fpsyg.2011.00255

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

Knoblich, G., Butterfill, S., & Sebanz, N. (2011). Psychological Research on Joint Action. In *Psychology of Learning and Motivation* (Vol. 54, pp. 59–101). Elsevier. https://doi.org/10.1016/B978-0-12-385527-5.00003-6

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

Levelt, W. J., Roelofs, A., & Meyer, A. (1999). A theory of lexical access in speech production. *The Behavioral and Brain Sciences*, *22*(1). https://doi.org/10.1017/s0140525x99001776

Moorselaar, D., & Slagter, H. (2020). Inhibition in selective attention. *Annals of the New York Academy of Sciences*, *1464*. https://doi.org/10.1111/nyas.14304

Nairne, J., VanArsdall, J., Pandeirada, J., Cogdill, M., & LeBreton, J. (2013). Adaptive Memory: The Mnemonic Value of Animacy. *Psychological Science*, *24*. https://doi.org/10.1177/0956797613480803

New, J., Cosmides, L., & Tooby, J. (2007). Category-specific attention for animals reflects ancestral priorities, not expertise. *Proceedings of the National Academy of Sciences*, *104*(42), 16598–16603. https://doi.org/10.1073/pnas.0703913104

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(04), Article 04. https://doi.org/10.1017/S0140525X12001495

Pickering, M. J., McLean, J. F., & Gambi, C. (2022). Interference in the shared-Stroop task: A comparison of self- and other-monitoring. *Royal Society Open Science*, *9*(4), 220107. https://doi.org/10.1098/rsos.220107

Rawlinson, H. C., & Kelley, C. M. (2021). In search of the proximal cause of the animacy effect on memory: Attentional resource allocation and semantic representations. *Memory & Cognition*, *49*(6), 1137–1152. https://doi.org/10.3758/s13421-021-01154-5

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*(1), 107–142. https://doi.org/10.1016/0010-0277(92)90041-F

Sahin, N., Pinker, S., Cash, S., Schomer, D., & Halgren, E. (2009). Sequential Processing of Lexical, Grammatical, and Phonological Information Within Broca's Area. *Science (New York, N.Y.)*, *326*, 445–449. https://doi.org/10.1126/science.1174481

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

Sebanz, N., Knoblich, G., & Prinz, W. (2005). How two share a task: Corepresenting stimulus-response mappings. *Journal of Experimental Psychology. Human Perception and Performance*, *31*(6), 1234–1246. https://doi.org/10.1037/0096-1523.31.6.1234

Shteynberg, G. (2010). A silent emergence of culture: The social tuning effect. *Journal of Personality and Social Psychology*, *99*(4), 683–689. https://doi.org/10.1037/a0019573

Shteynberg, G. (2015). Shared Attention. *Perspectives on Psychological Science*, *10*. https://doi.org/10.1177/1745691615589104

Shteynberg, G. (2018). A collective perspective: Shared attention and the mind. *Current Opinion in Psychology*, *23*, 93–97. https://doi.org/10.1016/j.copsyc.2017.12.007

Shteynberg, G., Hirsh, J. B., Apfelbaum, E. P., Larsen, J. T., Galinsky, A. D., & Roese, N. J. (2014). Feeling more together: Group attention intensifies emotion. *Emotion*, *14*(6), 1102–1114. https://doi.org/10.1037/a0037697

Shteynberg, G., Hirsh, J. B., Bentley, R. A., & Garthoff, J. (2020). Shared worlds and shared minds: A theory of collective learning and a psychology of common knowledge. *Psychological Review*, *127*(5), 918–931. https://doi.org/10.1037/rev0000200

Strijkers, K., & Costa, A. (2011). Riding the lexical speedway: A critical review on the time course of lexical selection in speech production. *Frontiers in Psychology*, *2*, 356. https://doi.org/10.3389/fpsyg.2011.00356

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

Turk, D. J., Cunningham, S. J., & Macrae, C. N. (2008). Self-memory biases in explicit and incidental encoding of trait adjectives. *Consciousness and Cognition*, *17*(3), 1040–1045. https://doi.org/10.1016/j.concog.2008.02.004

van Turennout, M., Hagoort, P., & Brown, C. M. (1997). Electrophysiological evidence on the time course of semantic and phonological processes in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(4), 787–806. https://doi.org/10.1037/0278-7393.23.4.787

Vlainic, E., Liepelt, R., Colzato, L., Prinz, W., & Hommel, B. (2010). The Virtual Co-Actor: The Social Simon Effect does not Rely on Online Feedback from the Other. *Frontiers in Psychology*, *1*, 208. https://doi.org/10.3389/fpsyg.2010.00208

Wenke, D., Atmaca, S., Holländer, A., Liepelt, R., Baess, P., & Prinz, W. (2011). What is Shared in Joint Action? Issues of Co-representation, Response Conflict, and Agent Identification. *Review of Philosophy and Psychology*, *2*(2), 147–172. https://doi.org/10.1007/s13164-011-0057-0

Appendix A: Statistical tables for reaction times

Fixed effects	Estimate	Std. Error	t value	Pr(> |t|)
(Intercept)	6.500e+00	3.053e-02	212.929	<0.001 ***
Task1 (Phon/Sem)	1.900e-01	3.967e-03	47.898	<0.001 ***
Turn1 (go/goT)	8.901e-03	3.969e-03	2.242	0.025 *
Task1/Turn1	1.086e-02	3.965e-03	2.738	0.006 **

Table 2.4: **Joint task: reactions times**. Results from the model on reaction times by task (semantic, phonetic) and turn (go, go-together) and their interaction.

Fixed effects	Estimate	Std. Error	t value	Pr(> |t|)
(Intercept)	6.470e+00	2.087e-02	309.937	<0.001 ***
Task1 (Phon/Sem)	2.914e-01	2.026e-02	14.381	<0.001 ***
Turn1 (go/goT)	-3.467e-04	3.828e-03	-0.091	0.928
Task1/Turn1	1.125e-03	3.805e-03	0.296	0.767

Table 2.5: **Individual task: reactions times**. Results from the model on reaction times by task (semantic, phonetic) and turn (go, go-together) and their interaction.

Appendix B: Statistical tables for memory recall test

2.5.1 Memory Recall test in the joint experiment

Fixed effects	Estimate	Std. Error	t value	Pr(> |t|)
(Intercept)	6.80915	0.28260	24.095	0.000117 ***
Task1 (Phon/Sem)	0.67917	0.15950	4.258	<0.001 ***
Turn1 (go/no-go)	1.53750	0.15950	9.639	<0.001 ***
Task1/Turn1	-0.09583	0.15950	-0.601	0.548608

Table 2.6: **Memory Recall Test for the joint experiment: GO and NO-GO trials**. Summary LMER for the memory recall rates according to Task (semantic, phonetic) and turn (go, no-go) and their interaction.

Fixed effects	Estimate	Std. Error	t value	Pr(> |t|)
(Intercept)	8.3167	0.2955	28.140	<0.001 ***
Task1 (Phon/Sem)	0.5833	0.2156	2.706	0.0082 **
Turn1 (go-together/go-alone)	0.9000	0.2156	4.175	<0.001 ***
Task1/Turn1	0.3000	0.2156	1.392	0.1676

Table 2.7: **Memory Recall Test for the joint experiment: GO-ALONE and GO-TOGETHER trials**. Summary LMER for the memory recall rates according to Task (semantic, phonetic) and turn (go, no-go) and their interaction.

Fixed effects	Estimate	Std. Error	t value	Pr(> |t|)
(Intercept)	5.2417	0.2799	18.723	<0.001 ***
Task1 (Phon/Sem)	0.7750	0.1883	4.115	<0.001 ***
Turn1 (other-go/no-go)	0.7083	0.1883	3.761	0.000306 ***
Task1/Turn1	0.0750	0.1883	0.398	0.691432

Table 2.8: **Memory Recall Test for the joint experiment: OTHER-GO and NO-GO trials**. Summary LMER for the memory recall rates according to Task (semantic, phonetic) and turn (go, no-go) and their interaction.

2.5.2 Memory Recall test in the individual experiment

Fixed effects	Estimate	Std. Error	t value	Pr(> $\|t\|$)
(Intercept)	6.4568	0.2062	31.319	<0.001***
Task1 (Phon/Sem)	0.3801	0.2011	1.890	0.0631 ***
Turn1 (go/no-go)	0.8797	0.1400	6.282	<0.001 ***
Task1/Turn1	-0.1547	0.1400	-1.105	0.2707

Table 2.9: **Memory Recall Test for the individual experiment: GO and NO-GO trials**. Summary LMER for the memory recall rates according to Task (semantic, phonetic) and turn (go, no-go) and their interaction.

Fixed effects	Estimate	Std. Error	t value	Pr(> $\|t\|$)
(Intercept)	7.3308	0.2733	26.827	<0.001 ***
Task1 (Phon/Sem)	0.1946	0.2652	0.734	0.466
Turn1 (go-together/go-alone)	-0.1759	0.1744	-1.008	0.317 ***
Task1/Turn1	-0.1241	0.1744	-0.712	0.479

Table 2.10: **Memory Recall Test for the joint task: GO-ALONE and GO-TOGETHER trials**. Summary LMER for the memory recall rates according to Task (semantic, phonetic) and turn (go, no-go) and their interaction.

Fixed effects	Estimate	Std. Error	t value	Pr(> $\|t\|$)
(Intercept)	5.59510	0.23025	24.301	<0.001 ***
Task1 (Phon/Sem)	0.55011	0.22935	2.399	0.0194 *
Turn1 (other-go/no-go)	0.03477	0.21635	0.161	0.8729
Task1/Turn1	-0.05144	0.21635	-0.238	0.8129

Table 2.11: **Memory Recall Test for the joint task: OTHER-GO and NO-GO trials**. Summary LMER for the memory recall rates according to Task (semantic, phonetic) and turn (go, no-go) and their interaction.

Chapter 3

Study 2: Co-representation and adaptive prediction using EEG

Article 2: Electrophysiological markers of language adaptation in joint production. Evidence from human-robot interaction.

Cirillo, G., Strijkers, Kristof, Runnqvist. E., Nguyen, N., Baus, C.,
in preparation

Abstract

The present study investigates whether people engage in lexico-semantic processing when performing a language task with a humanoid robot. A human participant and a robot alternated in naming pictures of objects belonging to 15 semantic categories, while participants' electrophysiological activity was recorded. We manipulated word frequency as a measure of lexical access having half of the pictures associated to high frequency names and half to low frequency names. In addition, the robot was programmed to give the semantic category name (e.g., tool for the picture of a hammer) instead of the more typical basic-level name (e.g., hammer for the picture of a hammer) for items belonging to 5 categories. Analyses on the picture-locked activity revealed a comparable ERP associated to frequency both when it was participant's turn and robot's turn to speak. Analyses on the response-locked activity show a different pattern for the category and basic-level responses in the first but not in the second part of the experiment, indicating acquisition and adaptation to the lexico-semantic pattern of the robot. Taken together, our findings provide empirical evidence for 1) the involvement of listeners' production system to predict robot's upcoming words, and for 2) partner-adaptive behavior supporting comprehension.

Keywords: Joint action; Speech production; Artificial partner; Lexical Frequency; Co-representation; Lexico-semantic processing

3.1 Introduction

According to the joint action theory (Sebanz et al., 2006; Sebanz & Knoblich, 2009), among the key elements allowing activity partners to coordinate and accomplish a joint goal is co-representation (Baus et al., 2014; Brehm et al., 2019; Gambi et al., 2015; Sahaï, Desantis, Grynszpan, Pacherie, & Berberian, 2019; Sebanz et al., 2003, 2005). Representing a co-actor's task enables individuals to make predictions about their partner's plans and upcoming actions (Baus et al., 2014; Vesper et al., 2017). This applies to a tennis match as much as to an everyday conversation. In a common exchange, conversational partners take turns between speaking and listening at extremely fast rates (at about 200 ms gaps), requiring them to process multiple language levels simultaneously (Levinson & Torreira, 2015). Predicting the partner's upcoming speech makes them able to plan and produce a response even before their interlocutor has actually ended up the utterance (Corps, Gambi, & Pickering, 2017; Garrod & Pickering, 2015). Those expectations are formed based on previous experience, and shaped by acquired phonological and grammatical constraints as well as meaning (Ramscar, Dye, & McCauley, 2013; Signoret et al., 2020). For instance, individuals use semantic information from the sentence to predict the upcoming verb (Freunberger & Roehm, 2016), or semantic information of verbs to predict the dependent noun (Maess, Mamashli, Obleser, Helle, & Friederici, 2016). Expectations are also supported by non-linguistic factors, including culture, social dynamics and familiarity with the interlocutor. The proactive nature of the human brain (Bar, 2009, 2021) finds therefore its path through the heterogeneity of verbal communication.

The integrated theory of language production and comprehension described by Pickering and Garrod (2013a) proposes that the ability to process different language dimensions rapidly and almost automatically comes from the strict coupling between language production and comprehension. The main idea behind this theory is that interlocutors use their production system to comprehend language via the formation of production intentions, retrieving all the relevant language representation levels, as if they were speaking themselves (Dell & Chang, 2014; Pickering & Garrod, 2007, 2013b). This account finds support from other studies showing how the quality of comprehension is highly dependent on the production skills (Huettig, 2015; Mani & Huettig, 2012). One key aspect of this prediction-by-production process supporting comprehension (Martin, Branzi, & Bar, 2018) is that it is subject to a continuous update, and interlocutors need to select relevant properties of the language environment and adapt to them. In this work, we implemented a joint picture-naming task to monitor the prediction dynamics to lexical and lexico-semantic properties of a robot's naming pattern using EEG.

Our design included two important manipulations. First, we manipulated word frequency having half of the pictures corresponding to high frequency words and half to low frequency words. Second, we manipulated robot's responses so that, for a regular subset of trials, the robot named the pictures with their corresponding semantic category label instead of the basic-level name. We asked whether 1) human participants engage in lexicalization processes when the robot prepares to speak and 2) whether they are able to adapt to the robot's lexico-semantic choices. Lexico-semantic patterns denote a language choice that, in addition to the lexical information, also make use of semantic information. Two different objects (e.g., a pen and a pencil), can share the same semantic category, such as 'tool' or 'stationery' (Katz & Fodor, 1963), while being associated to different names respectively (i.e., *pen* and *pencil*). Choosing to name an object with its category label would therefore guide the attention on this particular aspect of the object. In this sense, the generic perspective of the object seen as a piece of the stationery comes to be prioritized. This is in line with the shared attention theory (Shteynberg, 2015; Shteynberg et al., 2020).

In joint naming, researchers have highlighted how people co-represent their partner's task and speech, as comparable components associated to lexical access are present when both preparing to speak and preparing to listen to the partner, even though, given the task, no direct interaction is elicited between the participants. The most striking evidence comes from a joint picture naming study conducted by Baus et al. (2014), in which participants showed similar amplitudes of P200/P300 components related to easiness of processing associated to the lexical frequency of the words for both those trials that required their response and those that required the response of their partner as compared to no-go trials. Our study draws from Baus et al. (2014), and rises the question whether lexical co-representation is present also when having a robot as naming partner. Our second aim concerning adaptation relates to EEG studies reporting the elicitation of specific ERP components (e.g., P300, N400, P600) after processing deviant and unexpected linguistic events (Arbel et al., 2011; Fitz & Chang, 2019; Hodapp & Rabovsky, 2021; Kaan, Harris, Gibson, & Holcomb, 2000; Kutas & Federmeier, 2011; Van Petten & Luka, 2012). In our work, the use of a robot as language partner makes the quest for social effects even more challenging. Our first aim was to establish whether human participants predict what the robot is going to say, by monitoring lexical access via the frequency manipulation as in Baus et al. (2014). Secondly, and novel to the field, we aimed to capture adaptive prediction to the robot's responses over the course of the task, possibly evident in a gradual reduction of the amplitude of relevant ERP components (e.g., P300, N400) over the trials (Dikker & Pylkkanen, 2011; Fjaellingsdal et al., 2020; Hodapp & Rabovsky, 2021; Kutas & Federmeier, 2011; Nieuwland, 2019; Nieuwland & Van Berkum, 2006).

Robots have proven to be good task partners in joint production studies (Brandstetter & Bartneck, 2017; Marge et al., 2022; Wudarczyk, Kirtay, Pischedda, et al., 2021). Typical coordinating signals, such as gaze cueing, facial expressions, hesitation sounds have been found in human-robot interaction as much as in human-human interaction (Kilner, Paulignan, & Blakemore, 2003; Loth, Jettka, Giuliani, & de Ruiter, 2015; Skantze, 2016). At the behavioral level, for instance, participants have shown to align with robots in terms of their lexical, syntactic or conceptual choices (Branigan et al., 2010; Cirillo, Runnqvist, Strijkers, Nguyen, & Baus, 2022; Iio et al., 2015) or to facilitate production (Wudarczyk, Kirtay, Pischedda, et al., 2021), suggesting that humans are able to co-represent a robot's task. On the other hand, little investigation is present in the joint production research to signal the electrophysiological responses to a robot's (verbal) action, while it is often limited to processing artificial movements (Cooper et al., 2020). Elicitation of ERP components associated to prediction typically found in human-human interaction might signal language adaptation in human-robot interaction.

3.2 Methods

3.2.1 Participants

Thirty-four participants (14 males; mean age 24.5 years; age range 18–33 years) took part in the EEG study. All participants were right-handed, had normal or corrected-to-normal vision, were native French speakers, had no known history of neurological or psychiatric disorders. Prior to the experiment, participants read the instructions, and gave written informed consent. All participants were paid for their participation to the experiment. Due to technical problems related to system synchronization between Furhat and Biosemi (the EEG system), we were not able to provide results for 6 participants. 4 participants were also excluded because the robot did not perform the category naming level condition. For 1 participant, the experimental computer crashed in the middle of the experiment and we were not able to recover the data. Our final pool included 23 participants.

3.2.2 Material and design

Our visual stimuli consisted of 450 pictures of objects belonging to 15 semantic categories (e.g., fruits, tools, musical instruments), with 30 items in each category (see Appendix A). 377 pictures were taken from the MultiPic database (Duñabeitia et al., 2018), while the remaining 73 were drawn by a local artist we hired for the purpose. We extracted the frequency value for each item from the online French word database 'Lexique' (B. New, Pallier, Ferrand, & Matos, 2001). We averaged the means of the frequency values for words contained in books and words contained in film subtitles, to account for both spoken and written usage. A median split of the stimuli resulted in a mean frequency value for the high frequency items of 72.5 occurrences per million (sd = 108.2, subjective frequency: mean = 3.3, sd= 0.8) and a mean frequency value for low frequency items of 5.7 occurrences per million (sd = 3.9; subjective frequency: mean 2.3, sd= 0.8). As further test, we run a survey on 30 students from Aix-Marseille University, who evaluated how frequently they encountered a list of words in a 5-point Likert scale, where 1 corresponded to 'very rarely' and 5 to 'very often'. This allowed us to have a subjective frequency value as additional anchor. The audio stimuli consisted of 465 word productions pre-recorded by the synthesized voice of Furhat, that we distributed across six sets of randomized naming latencies (M = 930 ms, SD = 250 ms, range = 500–1600 ms). The robot was programmed to produce the basic-level name (e.g., 'hammer') for items belonging to 10 semantic categories, and the semantic category name (e.g., 'tool') for items corresponding to the remaining five semantic categories.

Our design included three main turn conditions, corresponding to who had to name the pictures: go trials (participant naming), other-go trials (robot naming) and no-go trials (no one naming). The exact numbers of items/trials associated to each turn condition are summarized in Table 3.1. Overall, 100 (10 items x 10 semantic categories) were assigned to each turn condition. Those categories corresponded to the 10 semantic categories in which items were named with the basic-level name by Furhat. 75 additional items (15 items x 5 semantic categories) were named by the robot with the category name, and the remaining 75 items were included as no-go trials. We eventually assigned 75 filler items to participants, to have a more balanced naming alternation across trials. Pictures were presented within a square, of which the color (blue, green or orange) indicated whether it was participant's, robot's or nobody's turn to speak. Semantic categories, items and colors were counterbalanced across naming conditions and turn conditions.

Condition	go	other-go	no-go
Basic-level	10 items x 10 categories	10 items x 10 categories	10 items x 10 categories
Category-level		15 items x 5 categories	15 items x 5 categories

Table 3.1: Distribution of items across trials.

3.2.3 Procedure

Each participant was introduced to Furhat and initiated a free conversation (around 5 min) with him. Participant and robot sat alongside each other in front of a computer screen. They were given instructions together. The experimenter told them that their task was to name as quickly as possible the pictures presented inside the square of the color (either blue, green or orange) to which they were assigned and to remain silent for the rest of the pictures. Before getting started with the main experiment, participants were able to practice the task with the robot. They were presented with a total of 15 pictures to familiarize with the three main turn conditions (go, other-go, no-go). We run the experiment using the OpenSesame software (Mathôt, Schreij, & Theeuwes, 2012). Trials presented the following structure (see Figure 3.1 for a detailed illustration of the experimental timeline): they started with a fixation point (+) presented in the middle of the screen for 500 ms followed by the picture presentation (2500 ms) and finally by a white screen (500 ms). The experiment comprised 5 blocks of 125 trials each. After each block, participants were asked to take a 2-minutes pause.

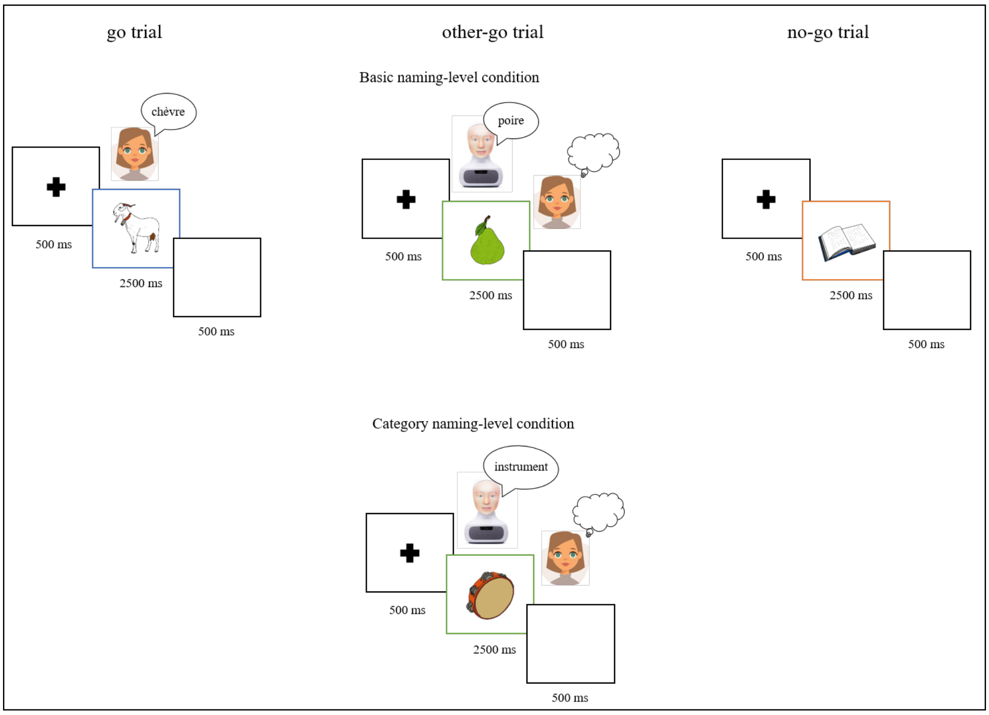

Figure 3.1: **Trials structure and timeline** with the three turn conditions (go, other-go and no-go) and the naming-level condition associated with the robot response. In the basic-level condition, the robot gave the basic-level name of the object (*pear*), while in the category condition, he gave the semantic category corresponding to the object (*instrument*). The thinking cloud icon corresponds to participant's tendency to co-represent robot's trials.

3.2.4 EEG data acquisition

Electroencephalographic (EEG) activity was recorded continuously by means of a 64 channel electrode cap (BioSemi Active). The impedance of the electrodes was kept below 20KΩ. Data was down-sampled offline to 512 Hz and a band-pass filter between 0.1 and 30 Hz was applied. The correction of ocular artifacts was conducted by Independent Component Analysis (ICA) in Brain Vision Analyzer 2.0 (Brain Products GmbH, Germany), for which segments containing artifacts and eye blinks were identified and manually rejected. Afterwards, data was re-referenced to electrodes closest to the mastoids (TP7 and TP8), segmented, artifact corrected (voltages above or below 100 microvolts) and time-locked to the onset of both picture and robot's response (-100 ms pre- and 700 ms post-stimulus). Cleaned epochs were baseline corrected to -100 ms pre-stimulus baseline, and were averaged separately according to each condition.

3.3 Data analysis

We analyzed behavioral responses (for go-trials) and ERP responses (for go, other-go and no-go trials) of 23 participants.

3.3.1 Behavioral analysis

We run an error rate and a naming latency analysis on participants' responses in go-trials, especially to replicate standard frequency effects as a sanity check for our design. We divided the experiment into two parts or blocks (block 1 and block 2). This methodological choice was adopted in order to understand how participants' responses changed over time, with a sufficient number of trials. It was used in the response-locked ERP analyses as well. Errors (N = 459) comprised no-responses, hesitations as well as responses that differed from the target name we aimed to elicit (e.g., 'boat' for the picture of a ship). They were fitted into a generalized linear mixed model (glmer) using the lmr4 package in R Bates et al. (2015), where frequency (low vs high) was taken as fixed factor while item and participant were taken as random factors. For naming latencies, the analysis was carried out by fitting a linear mixed effects model (lmer) with frequency (low vs high) as fixed factor, and item and participant as random factors. Naming latencies were log-transformed for a better fit to the model, and errors as well as 2.5 standard deviations above or below the mean response time for high and low frequency words respectively were excluded.

To monitor participant's adaptation to robot's conceptual patterns, we run an analysis of the category responses over time (across the two blocks; see also Chapter 4). We collected a total of 68 responses, indicating that participants did manifest a certain tendency to reproduce the behavioral pattern of the robot. We binary coded (0–1) participants' responses, indicating whether they used or not a category-related name during each trial. Category names (e.g., tool for hammer; N = 176) were fitted to a logits regression model with block (1 and 2; we divided the experiment in two parts) as fixed factor and participant and item as random factors.

3.3.2 ERP analysis

We conducted separate analyses to distinguish between brain activity locked to the picture presentation and brain activity locked to the robot's responses. For the first analysis we were interested to see the effect of the lexical frequency manipulation across the three turn conditions (go, no-go and other-go). Based on previous literature investigating the electrophysiological correlates of frequency, we focused on two time windows: 150-250 ms and 250-350 ms corresponding to the P200 and P300 respectively (Baus et al., 2014; Strijkers et al., 2010). In addition, we analyzed activity in the time window 350-450 for later effects. Go trials were analyzed separately from no-go and other-go trials. Analysis of go-trials served as a baseline to test the validity of the frequency manipulation. To do so, we run a linear mixed effects model for each time-window (Bates et al., 2015), with word frequency (high vs. low frequency) and region (Anterior, Central, Posterior) (Anterior: AF7, AF3, AF4, AFz, F1, F3, F5, FC3, FC5, Fz, FC1, FCz, FC2, F2, F4, F6, FC4, FC6; Central: C1, C3, C5, CP3, CP5, Cz, CP1, CPz, CP2, C2, C4, C6, CP4, CP6; Posterior: P3, P5, P7, PO3, PO7, P1, Pz, P2, POz, P4, P6, P8, PO4; PO8, O1, Iz, Oz, O2) as predictors and participants as random factor. Other-go and no-go trials were analyzed together as two different levels of no-go trials. The model included type of turn (other-go vs simple no-go) together with frequency (high vs. low frequency) and region (Anterior, Central, Posterior) as fixed factors and participant as random factor.

For the electrophysiological reactions to the robot's responses, we wanted to define how the category manipulation affected comprehension by contrasting the two naming levels associated with the robot's responses (saying the basic-level name vs. saying the category-level name). In addition, we compared activity between the two parts of the experiment (two blocks). For this, we fitted the continuous brain activity to a linear mixed effects model with naming level (basic vs. category), block (1 vs 2) and region (Anterior, Central, Posterior) as fixed factors and subject as random factor. We conducted the analysis on one time-window: 400 - 600 ms. This specific choice was made in order to capture the prediction error we expected to elicit when hearing the semantic category instead of the more predictable basic-level name. As the manipulation we performed concerned the semantic level, we based our approach on previous research suggesting that semantic and contextual patterns are normally integrated later in time compared to lexical patterns. A higher amplitude of components associated to stimuli deviating from contextually-induced expectations (e.g., P300, N400 or P600) at the category-level condition would evidence a processing difficulty when listening to the category name instead of the basic-level name. In addition, having block as a variable allowed us to monitor how the amplitudes of those components changed over the course of the experiment, and how different processing patterns were associated to the two naming conditions respectively.

3.4 Results

3.4.1 Behavioral results

The analysis on error rates revealed an effect of frequency. Low frequency words were named less accurately than high frequency words (M LF = 0.29, SD LF = 0.4; M HF = 0.13, SD HF = 0.3: b = 0.52167, s.e. = 0.05688, z = 9.172, $p < 0.001$). The analysis did not reveal any effect of block (b = -0.04, s.e. = 0.05, t = -0.71, p = 0.473) nor of the interaction between block and frequency (b = 0.06, s.e. = 0.056, z = 1.06, p = 0.28). The analysis on naming latencies showed a significant effect of frequency (b = -3.694e-02, s.e. = 7.750e-03, t = -4.766, $p < 0.001$) and block (b = 1.070e-02, s.e. = 4.613e-03, t = 2.32, p = 0.02), but not for the interaction between the variables (b = -3.210e-04, s.e. = 4.657e-03, t = -0.06, p = 0.94). Pictures associated with low frequency words were named slower (M = 1055.28 ms, SD = 292.79) than high frequency items (M = 995.2 ms, SD = 285.94). From the behavioral measures we were therefore able to replicate standard frequency effects characterizing lexical speed in picture-naming. The analysis on the category responses (N = 68) did not reveal a significant effect for block (b = 0.1371, s.e. = 0.2957, t = 0.464, p = 0.643), suggesting that lexical alignment does not systematically apply when partners do not share the same conceptual cues (i.e., the same categories).

3.4.2 ERP results

3.4.3 Picture-locked activity: Frequency effects

For the go-trials, the analysis performed in the 150-250 time-window (P200) revealed an effect of frequency, with low-frequency items eliciting more positive waveforms than high frequency items (b = 0.09589, s.e. = 0.04068, t = 2.357, p = 0.0185). No interaction with region was observed ($p > 0.5$). We found the same pattern in the 250-350 ms time-window, with a significant frequency effect (b = 1.285e-01, s.e. = 5.484e-02, t = 2.343, p = 0.0192), and no interaction between frequency and region ($p > 0.05$). The frequency effect, however, disappeared when testing it in the latest time-window (350-450 ms: b = 0.12755, s.e. 0.06824, t = 1.869, p = 0.0617). Full statistical results for go-trials are included in the Tables 3.2, 3.3 and 3.4.

For the no-go trials (robot's trials and no one's trials), the model on the 150-250 ms time-window did not reveal any effect of frequency (b = 2.231e-02, s.e. = 2.948e-02, t = 0.757, p = 0.44913), neither an interaction with type of turn (b = 5.256e-02, s.e. =

2.948e-02, t = 1.783, p = 0.07463) nor region (p >0.05). However, and relevant here, we observed an interaction between frequency and type of turn in the second time-window (250-350 ms: b = 0.07153, s.e. = 0.03517, t = 2.034, p = 0.042011). When performing the pair-wise comparison contrasting frequency and type of turn, we found that, for other-go trials, low frequency items elicited more positive waveforms compared to high frequency items (b = 0.2667, s.e. = 0.09, t = 2.6, p = 0.0370), while no significant frequency effect was reported for no-go trials (b = -0.0195, s.e. = 0.0995, t = -0.196, p = 0.9973). We also found a significant interaction in the time-window 350-450 ms (b = 7.024e-02, s.e. = 3.483e-02 , t = 2.017, p = 0.04380), confirming the elicitation of a P300 component associated to processing low-frequency compared to high frequency words for other-go trials (b = 0.4272, s.e. = 0.0985, t = 4.336, p = 0.0001) and not for no-go trials (b = 0.1462, s.e. = 0.0985, t = 1.484, p = 0.4472). Full statistical results are included in the Tables 3.5, 3.6 and 3.7. For full pair-wise comparison results, see Appendix C.

Fixed effects	Estimate	Std. Error	t value	Pr(> $\|t\|$)
(Intercept)	0.24798	0.17563	1.412	0.1719
Frequency1 (Low/High)	0.09589	0.04068	2.357	0.0185 *
Region1 (Posterior/Anterior)	0.51961	0.05627	9.235	<0.001 ***
Region2 (Posterior/Central)	-0.28126	0.05624	-5.001	<0.001 ***
Frequency1/Region1	0.01962	0.05627	0.349	0.7274
Frequency1/Region2	-0.08320	0.05623	-1.480	0.1391

Table 3.2: **Go trials - Time window 150-250 ms**. Summary LMER for the electrophysiological activity (measured in volt) according to Frequency (High vs. Low), Region (Anterior, Central, Posterior) and their interactions.

Fixed effects	Estimate	Std. Error	t value	Pr(> $\|t\|$)
(Intercept)	-2.695e-01	2.204e-01	-1.223	0.2343
Frequency1 (Low/High)	1.285e-01	5.484e-02	2.343	0.0192 *
Region1 (Posterior/Anterior)	1.446e+00	7.584e-02	19.071	<0.001 ***
Region2 (Posterior/Central)	-1.354e+00	7.580e-02	-17.863	<0.001 ***
Frequency1/Region1	-1.468e-03	7.584e-02	-0.019	0.9846
Frequency1/Region2	-8.781e-02	7.579e-02	-1.159	0.2468

Table 3.3: **Go trials - Time window 250-350 ms**. Summary LMER for the electrophysiological activity (measured in volt) according to Frequency (High vs. Low), Region (Anterior, Central, Posterior) and their interactions.

Fixed effects	Estimate	Std. Error	t value	Pr(> $\|t\|$)
(Intercept)	0.07699	0.27914	0.276	0.7853
Frequency1 (Low/High)	0.12755	0.06824	1.869	0.0617
Region1 (Posterior/Anterior)	0.96919	0.09438	10.269	<0.001 ***
Region2 (Posterior/Central)	-1.78779	0.09433	-18.953	<0.001***
Frequency1/Region1	-0.01658	0.09438	-0.176	0.8605
Frequency1/Region2	-0.04808	0.09432	-0.510	0.6103

Table 3.4: **Go trials - Time window 350-450 ms**. Summary LMER for the electrophysiological activity (measured in volt) according to Frequency (High vs. Low), Region (Anterior, Central, Posterior) and their interactions.

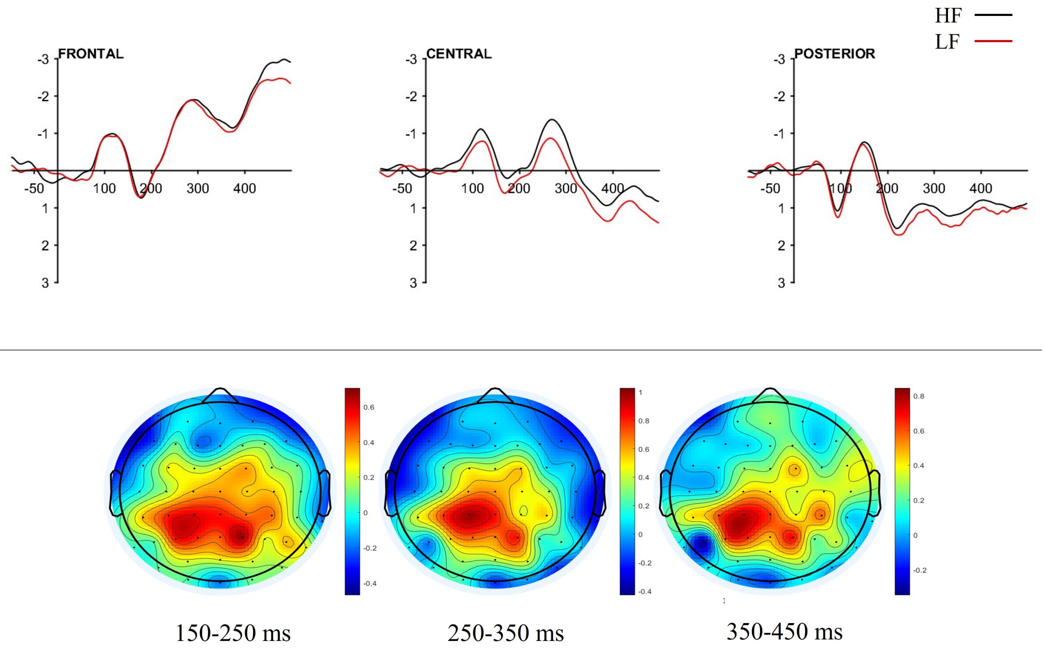

Figure 3.2: **EEG results for the GO trials time-locked to the presentation of the pictures**. The three figures in the upper panel represent the grand averages for go trials for anterior, central and posterior electrodes respectively. Electrodes were grouped according to the region (frontal, central, posterior). Red lines represent low-frequency words (LF) and black lines represent high-frequency words (HF). The lower panel illustrates the topographical maps representing the frequency effect in the 150-250, 250-350 and 350-450 time windows (low frequency words minus high frequency ones). Red colors indicate positive differences, corresponding to low frequency words bring more positive than high frequency words.

Fixed effects	Estimate	Std. Error	t value	Pr(> $\|t\|$)
(Intercept)	1.603e-01	1.234e-01	1.299	0.20746
Frequency1 (Low/High)	2.231e-02	2.948e-02	0.757	0.44913
Turn1 (Other/No)	3.488e-02	2.948e-02	1.183	0.23681
Region1 (Posterior/Anterior)	1.240e+00	4.077e-02	30.427	<0.001 ***
Region2 (Posterior/Central)	-8.706e-01	4.075e-02	-21.367	<0.001 ***
Frequency1/Turn1	5.256e-02	2.948e-02	1.783	0.07463
Frequency1/Region1	1.168e-02	4.077e-02	0.286	0.77457
Frequency1/Region2	-2.880e-02	4.074e-02	-0.707	0.47977
Turn1/Region1	1.272e-01	4.077e-02	3.120	0.00182 **
Turn1/Region2	-1.961e-01	4.074e-02	-4.812	<0.001 ***
Frequency1/Turn1/Region1	5.435e-02	4.077e-02	1.333	0.18253
Frequency1/Turn1/Region2	-5.258e-03	4.074e-02	-0.129	0.89732

Table 3.5: **No-go trials - Time window 150-250 ms**. Summary LMER for the electrophysiological activity (measured in volt) according to Frequency (High, Low), Turn (Other-go, No-go), Region (Anterior, Central, Posterior) and their interactions.

Fixed effects	Estimate	Std. Error	t value	Pr(> $\|t\|$)
(Intercept)	-0.44379	0.15681	-2.830	0.009735 **
Frequency1 (Low/High)	0.06180	0.03517	1.757	0.078938
Turn1 (Other/No)	0.10420	0.03517	2.963	0.003064 ***
Region1 (Posterior/Anterior)	1.60961	0.04864	33.094	<0.001 ***
Region2 (Posterior/Central)	-1.29616	0.04861	-26.664	<0.001 ***
Frequency1/Turn1	0.07153	0.03517	2.034	0.042011 *
Frequency1/Region1	0.02894	0.04864	0.595	0.551836
Frequency1/Region2	-0.07796	0.04861	-1.604	0.108806
Turn1/Region1	0.11945	0.04864	2.456	0.014092 *
Turn1/Region2	-0.16701	0.04861	-3.436	<0.001 ***
Frequency1/Turn1/Region1	0.04300	0.04864	0.884	0.376705
Frequency1/Turn1/Region2	0.01160	0.04861	0.239	0.811455

Table 3.6: **No-go trials - Time window 250-350 ms**. Summary LMER for the electrophysiological activity (measured in volt) according to Frequency (High, Low), Turn (Other-go, No-go), Region (Anterior, Central, Posterior) and their interactions.

Fixed effects	Estimate	Std. Error	t value	Pr(> $\|t\|$)
(Intercept)	-5.602e-01	1.737e-01	-3.224	0.00390 **
Frequency1 (Low/High)	1.433e-01	3.483e-02	4.115	<0.001 ***
Turn1 (Other/No)	1.288e-01	3.483e-02	3.698	0.00022 ***
Region1 (Posterior/Anterior)	1.008e+00	4.817e-02	20.925	<0.001 ***
Region2 (Posterior/Central)	-8.681e-01	4.814e-02	-18.031	<0.001 ***
Frequency1/Turn1	7.024e-02	3.483e-02	2.017	0.04380 *
Frequency1/Region1	-3.427e-02	4.817e-02	-0.711	0.47684
Frequency1/Region2	-2.842e-02	4.814e-02	-0.590	0.55505
Turn1/Region1	1.495e-01	4.817e-02	3.103	0.00193 **
Turn1/Region2	-2.346e-01	4.814e-02	-4.872	<0.001 ***
Frequency1/Turn1/Region1	-8.228e-04	4.817e-02	-0.017	0.98637
Frequency1/Turn1/Region2	6.914e-02	4.814e-02	1.436	0.15102

Table 3.7: No-go trials - **Time window 350-450 ms**. Summary LMER for the electrophysiological activity (measured in volt) according to Frequency (High, Low), Turn (Other-go, No-go), Region (Anterior, Central, Posterior) and their interactions.

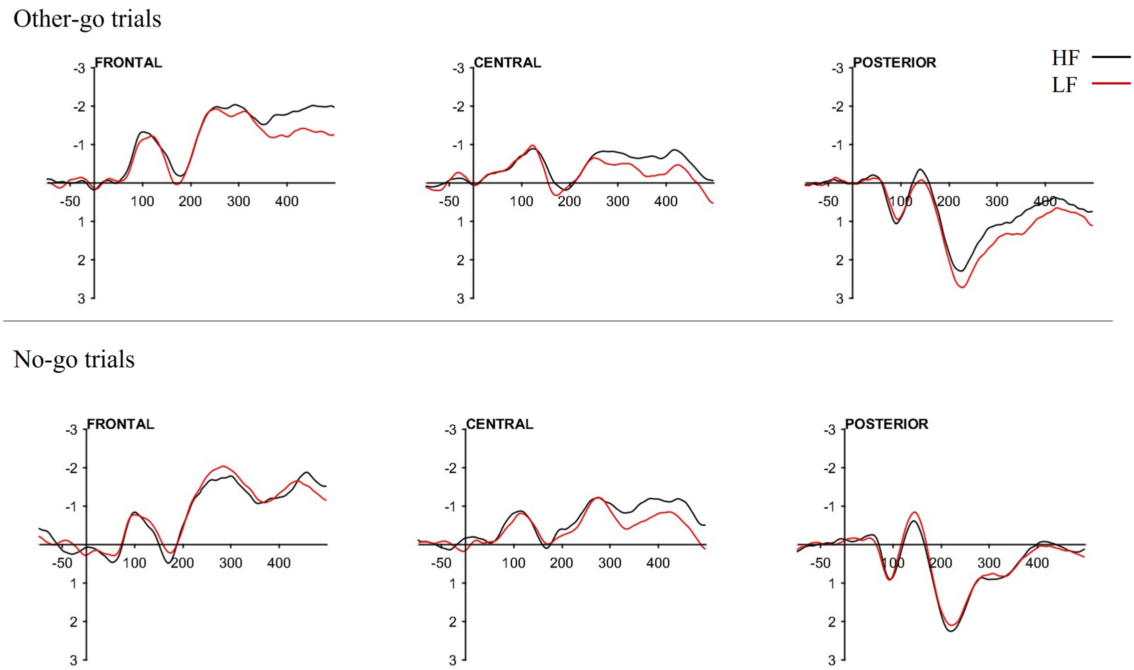

Figure 3.3: **EEG results for the NO-GO trials (other-go + no-go) time-locked to the presentation of the pictures**. The three figures in the upper panel represent the grand averages for other-go trials for anterior, central and posterior electrodes respectively. Red lines represent low-frequency words (LF) and black lines represent high-frequency words (HF). In the lower panel, the same grand averages are represented for simple no-go trials.

3.4.4 Response-locked activity: Adaptation to the lexico-semantic pattern

The statistical model on robot's responses on the time-window considered (400-600 ms) revealed a significant effect of naming level (b = 1.000e-01, s.e. = 2.744e-02, t = 3.646, p = 0.000269) and, relevant here, of the interaction between naming level and block (b = 1.347e-01, s.e. = 2.744e-02, t = 4.910, p <0.001). Category-level names elicited more positive amplitudes compared to basic-level names in the first block (b = 0.4695, s.e. = 0.0776, t = 6.050, p <0.0001). However, this difference disappeared in the second block (b = -0.0693, s.e. = 0.0776, t = -0.894, p = 0.8082). Those results signal 1) different processing ease associated to each naming condition respectively (first block) and, especially, 2) adaptation to the lexico-semantic choices of the robot, that became gradually more expected and accepted by participants (reduced difference in the second block). See Table 3.8, 3.9 and 3.10 for full statistical results. For full par-wise comparison results see Appendix C.

Fixed effects	Estimate	Std. Error	t value	Pr(> |t|)
(Intercept)	2.236e-01	1.520e-01	1.470	0.155587
NamingLevel1 (Category/Basic)	1.000e-01	2.744e-02	3.646	0.000269 ***
Block1 (1/2)	-5.339e-02	2.744e-02	-1.946	0.051721
Region1 (Posterior/Anterior)	-4.543e-01	3.810e-02	-11.925	¡ 2e-16 ***
Region2 (Posterior/Central)	2.880e-01	3.761e-02	7.658	2.28e-14 ***
NamingLevel1/Block1	1.347e-01	2.744e-02	4.910	9.42e-07 ***
NamingLevel1/Region1	-2.679e-02	3.810e-02	-0.703	0.481989
NamingLevel1/Region2	-1.304e-02	3.761e-02	-0.347	0.728732
Block1/Region1	4.751e-03	3.810e-02	0.125	0.900750
Block1/Region2	-8.292e-02	3.761e-02	-2.205	0.027516 *
NamingLevel1/Block1/Region1	3.026e-02	3.810e-02	0.794	0.427013
NamingLevel1/Block1/Region2	-1.874e-02	3.761e-02	-0.498	0.618371

Table 3.8: Other-go trials: response-locked activity - **Time window 400-600 ms**. Summary LMER for the electrophysiological activity (measured in volt) according to Naming Level (Basic, Category), Block (1,2), Region (Anterior, Central, Posterior) and their interactions.

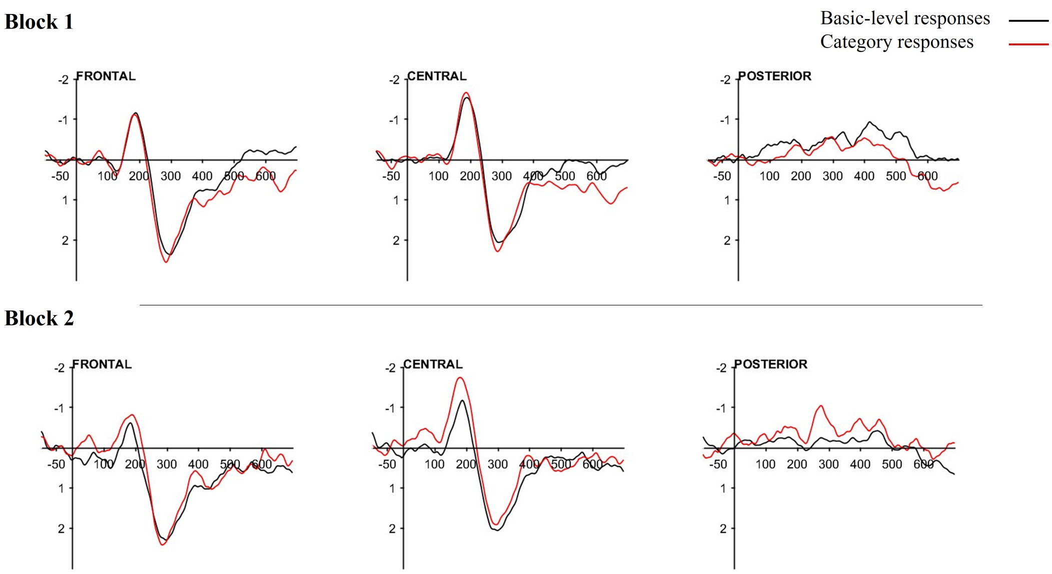

Figure 3.4: **EEG results for the OTHER-GO trials time-locked to robot's responses**. The figures show grand average of the brain activity elicited when hearing the robot responses across the two blocks. Electrodes were grouped according to the region (frontal, central, posterior). Black lines represent the basic-level naming, while red lines correspond to the category naming. In the upper panel, the grand average is related to the first block, while in the lower panel it is related to the second block.

3.5 Discussion

In the present study, we explored the electrophysiological correlates associated to co-representation and partner-adaptation in language processing when performing a joint picture-naming task with a social robot. Participants took turns in naming pictures with Furhat, a humanoid robot able to produce speech and embody physical emotions. The pictures displayed on the computer screen represented objects belonging to 15 semantic categories, and were distributed across three turn conditions: a go condition requiring the participant to name the pictures, an other-go condition in which Furhat named the pictures and a no-go condition in which none of them spoke. Half of the pictures corresponded to high frequency names, and half to low frequency names. Importantly, we manipulated the responses of the robot in a way to have the robot produce the semantic category name of the object (e.g., *fruit* for the picture of a pear) instead of the basic-level name (e.g., *pear* for the picture of a pear) for a regular subset of trials (5 out of the 15 semantic categories).

Our aim was twofold. First, we wanted to replicate with the robot previous studies in human-human interaction showing co-representation towards the partner's task, and find comparable electrophysiological patterns when preparing to speak and when preparing to listen to an interactive partner (Baus et al., 2014). We monitored lexical access via the frequency manipulation reflected in the different processing ease between low and high frequency words. Based on previous studies using lexical frequency, we expected that, if participants successfully co-represented robot's trials, we would find more positive amplitudes corresponding to the P200 and P300 associated to low frequency words both in go and other-go trials but not in the no-go trials. Our second objective was related to the manipulation we performed at the robot's naming level. We monitored participants' brain activity when listening to a less predictable name in comparison to when hearing the expected, basic-level name across the two main parts of the experiment. In accordance with research focusing on ERPs associated with processing of deviant and unexpected (lexical, semantic) linguistic events, we focused on the participants' brain activity starting around 300/400 ms, and hypothesized a reduced difference in voltage amplitude between the naming level condition from the first to the second block. This would be linked to semantic processing of less predictable words in the category condition but not in the basic-level condition (Fitz & Chang, 2019; Hodapp & Rabovsky, 2021).

The following main results were observed. At the behavioral level, naming latencies on go trials were faster and more accurate for high compared to low frequency words, replicating previous results of picture-naming studies using the frequency manipulation (Almeida et al., 2007; Baus et al., 2014; Navarrete et al., 2006; Strijkers et al., 2010). At the electrophysiological level, go trials elicited the expected ERP components associated to lexical frequency, that is the P200 and P300 components (Baus et al., 2014; Strijkers & Costa, 2011). Low-frequency words were associated to larger, more positive waveforms between 200 and 300 ms after stimulus onset. The analyses performed on no-go trials revealed elicitation of a comparable activity distinguishing low versus high frequency words peaking around 370 ms for other-go trials, while no difference was observed in simple no-go trials. Those results suggest that participants successfully **co-represented robot's task**, and they were actively monitoring his upcoming words.

When analyzing the electrophysiological activity time-locked to the robot's responses (for both the basic-level names and the category names), we observed that category names and basic-level names elicited different amplitudes in the first block. This difference disappeared over the course of the experiment. This result indicates **adaptation towards the robot's lexico-semantic choices**.

3.5.1 Task co-representation with robot: picture-locked activity

Our results show that participants co-represented the robot's task, and engaged in comparable lexicalization processing when preparing to name the pictures and when the robot was expected to name the pictures. This is indicated by the the fact that we found a frequency effect for other-go trials (trails in which the robot named the pictures) but not in simple no-go trials (trials in which no one named the pictures). Our study therefore confirms Baus et al. (2014)'s results for human-robot interaction, showing that humans are able to co-represent a robot's verbal actions and predict relevant lexical properties of his speech. More broadly, our work relates to the accounts of co-representation within the joint action theory, suggesting that, under shared attention, activity partners co-represent each other's action at different moments in the interaction (Atmaca et al., 2011; Sebanz et al., 2003, 2005) and use their production system to simulate what the partner is going to say (Pickering & Garrod, 2007, 2013b). Similarly with Baus et al. (2014), we also obtained a difference in latency between go and other-go trials for the frequency effect. Frequency effects appeared later when the other (human or robot) named the pictures (around 350 ms) than when the participant named the pictures (200-300 ms). The authors explained this aspect by referring to inhibitory processes participants engage with to refrain from responding to no-go trials. That is, the speed with which lexical access starts is modulated by whether a word will be finally uttered or not (Strijkers & Costa, 2011).

3.5.2 Adaptation to robot's lexico-semantic choices: response-locked activity

The more positive waveforms observed starting from 400 ms after response onset between category names and basic-level names indicate that participants found it harder to process words that did not correspond to the more typical and more expected basic-level names. Those electrophysiological responses are therefore an index of what is perceived as a deviant response in human-robot interaction. Importantly, this difference was not observed in the second block, indicating that participants acquired the robot's way of responding and adapted to it. From a linguistic point of view, it indicates that participants were able to adapt to the interlocutor at the conceptual level, and learned to predict the lexical choices of the robot.

Our work relates to research in linguistic alignment within human-robot interaction (Branigan et al., 2010, 2011; Pearson, Hu, Branigan, Pickering, & Nass, 2006; Wudarczyk, Kirtay, Pischedda, et al., 2021), in which humans have been shown to copy a robot's syntactic structures, and words. Here we were able to reproduce prediction adaptation towards a robot's word choices at the conceptual level (Wudarczyk, Kirtay, Pischedda, et al., 2021). Further investigations are needed to explore how this affects participants' speaking (naming). In particular if, when participants are exposed to name items from the same semantic categories of the robot (i.e., those whose robot's response was manipulated), they start replicating his lexico-semantic pattern (see Chapter 4).

3.6 Conclusion

The present study investigated the ERP components associated to co-representation and adaptation to lexical and lexico-semantic properties using a joint picture-naming task between a participant and a humanoid robot. Taken together, our results provide evidence for two important electrophysiological aspects emerging after picture presentation and after response production respectively. First, human participants were able to co-represent the robot's verbal actions. This was demonstrated by the parity of activity associated to processing low versus high frequency words when preparing to speak and preparing to listen compared to simple no-go trials. Our work therefore replicates that by Baus et al. (2014)'s in human-robot interaction, possibly suggesting that co-representation and prediction of one's partner upcoming speech are so eradicated in the joint language production dynamics that they are present even when the partner is a robot. Second, our results reveal adaptation to the response pattern of the robot, who produced the semantic category name instead of the basic-level name of objects for a regular subset of trials. The electrophysiological responses of the participants were, in fact, diverging between the basic-level condition and the category-level condition in the first part of the experiment, while this difference disappeared in the second part of the experiment. To our knowledge, our work offers the first empirical evidence on language adaptation in human-robot interaction at both the lexical and the conceptual level using EEG.

References

Almeida, J., Knobel, M., Finkbeiner, M., & Caramazza, A. (2007). The locus of the frequency effect in picture naming: When recognizing is not enough. *Psychonomic Bulletin & Review*, *14*(6), 1177–1182. https://doi.org/10.3758/BF03193109

Arbel, Y., Spencer, K. M., & Donchin, E. (2011). The N400 and the P300 are not all that independent. *Psychophysiology*, *48*(6), 861–875. https://doi.org/10.1111/j.1469-8986.2010.01151.x

Atmaca, S., Sebanz, N., & Knoblich, G. (2011). The joint flanker effect: Sharing tasks with real and imagined co-actors. *Experimental Brain Research. Experimentelle Hirnforschung. Experimentation Cerebrale*, *211*(3–4), 371–385. https://doi.org/10.1007/s00221-011-2709-9

Bar, M. (2009). The proactive brain: Memory for predictions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1235–1243. https://doi.org/10.1098/rstb.2008.0310

Bar, M. (2021). From Objects to Unified Minds. *Current Directions in Psychological Science*, *30*(2), 129–137. https://doi.org/10.1177/0963721420984403

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), Article 1. https://doi.org/10.18637/jss.v067.i01

Baus, C., Sebanz, N., Fuente, V. de la, 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

Brandstetter, J., & Bartneck, C. (2017). Robots will dominate the use of our language. *Adaptive Behavior*, *25*(6), 275–288. https://doi.org/10.1177/1059712317731606

Branigan, H. P., Pickering, M. J., Pearson, J., & McLean, J. F. (2010). Linguistic alignment between people and computers. *Journal of Pragmatics*, *42*(9), 2355–2368. https://doi.org/10.1016/j.pragma.2009.12.012

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

Branigan, H. P., Pickering, M. J., Pearson, J., Mclean, J., & Nass, C. (2003). Syntactic alignment between computers and people: The role of belief about mental states. *Cognitive Science - COGSCI*.

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

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

Cooper, S., Gow, S., Fensome, S., Dragone, M., & Kourtis, D. (2020). *An EEG investigation on planning human-robot handover tasks*. TechRxiv. https://doi.org/10.36227/techrxiv.12950342.v2

Corps, R., Gambi, C., & Pickering, M. (2017). Coordinating Utterances During Turn-Taking: The Role of Prediction, Response Preparation, and Articulation. *Discourse Processes*, *55*, 1–11. https://doi.org/10.1080/0163853X.2017.1330031

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). https://doi.org/10.1098/rstb.2012.0394

Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly*

Journal of Experimental Psychology, *71*(4), 808–816. https://doi.org/10.1080/17470218.2017.1310261

Fitz, H., & Chang, F. (2019). Language ERPs reflect learning through prediction error propagation. *Cognitive Psychology*, *111*, 15–52. https://doi.org/10.1016/j.cogpsych.2019.03.002

Freunberger, D., & Roehm, D. (2016). Semantic prediction in language comprehension: Evidence from brain potentials. *Language, Cognition and Neuroscience*, *31*(9), 1193–1205. https://doi.org/10.1080/23273798.2016.1205202

Gambi, C., Van de Cavey, J., & Pickering, M. J. (2015). Interference in joint picture naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(1), 1–21. https://doi.org/10.1037/a0037438

Garrod, S., & Pickering, M. J. (2015). The use of content and timing to predict turn transitions. *Frontiers in Psychology*, *6*, 751. https://doi.org/10.3389/fpsyg.2015.00751

Hodapp, A., & Rabovsky, M. (2021). The N400 ERP component reflects an error-based implicit learning signal during language comprehension. *European Journal of Neuroscience*, *54*(9), 7125–7140. https://doi.org/10.1111/ejn.15462

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

Iio, T., Shiomi, M., Shinozawa, K., Shimohara, K., Miki, M., & Hagita, N. (2015). Lexical Entrainment in Human Robot Interaction. *International Journal of Social Robotics*, *7*. https://doi.org/10.1007/s12369-014-0255-x

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, *15*(2), 159–201. https://doi.org/10.1080/016909600386084

Kilner, J. M., Paulignan, Y., & Blakemore, S. J. (2003). An Interference Effect of Observed Biological Movement on Action. *Current Biology*, *13*(6), 522–525. https://doi.org/10.1016/S0960-9822(03)00165-9

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*, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, *6*. https://www.frontiersin.org/articles/10.3389/fpsyg.2015.00731

Loth, S., Jettka, K., Giuliani, M., & de Ruiter, J. P. (2015). Ghost-in-the-Machine reveals human social signals for human–robot interaction. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01641

Maess, B., Mamashli, F., Obleser, J., Helle, L., & Friederici, A. D. (2016). Prediction Signatures in the Brain: Semantic Pre-Activation during Language Comprehension. *Frontiers in Human Neuroscience*, *10*. https://www.frontiersin.org/articles/10.3389/fnhum.2016.00591

Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake—But only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(4), 843–847. https://doi.org/10.1037/a0029284

Marge, M., Espy-Wilson, C., Ward, N. G., Alwan, A., Artzi, Y., Bansal, M., Blankenship, G., Chai, J., Daumé, H., Dey, D., Harper, M., Howard, T., Kennington, C., Kruijff-Korbayová, I., Manocha, D., Matuszek, C., Mead, R., Mooney, R., Moore, R. K., … Yu, Z. (2022). Spoken language interaction with robots: Recommendations for future research. *Computer Speech & Language*, *71*, 101255. https://doi.org/10.1016/j.csl.2021.101255

Martin, C. D., Branzi, F. M., & Bar, M. (2018). Prediction is Production: The missing link between language production and comprehension. *Scientific Reports*, *8*(1), 1–9. https://doi.org/10.1038/s41598-018-19499-4

Navarrete, E., Basagni, B., Alario, F.-X., & Costa, A. (2006). Does word frequency affect lexical selection in speech production? *Quarterly Journal of Experimental Psychology*, *59*(10), 1681–1690. https://doi.org/10.1080/17470210600750558

New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). Une base de données lexicales du français contemporain sur internet: LEXIQUE™//A lexical database for contemporary french : LEXIQUE™. *Annee Psychologique - ANNEE PSYCHOL*, *101*, 447–462. https://doi.org/10.3406/psy.2001.1341

Pearson, J., Hu, J., Branigan, H., Pickering, M. J., & Nass, C. (2006). *Adaptive language behavior in HCI: How expectations and beliefs about a system affect users' word choice*. *2*, 1177–1180. https://doi.org/10.1145/1124772.1124948

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, *11*(3), 105–110. https://doi.org/10.1016/j.tics.2006.12.002

Pickering, M. J., & Garrod, S. (2013a). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(04), Article 04. https://doi.org/10.1017/S0140525X12001495

Pickering, M. J., & Garrod, S. (2013b). Forward models and their implications for production, comprehension, and dialogue. *Behavioral and Brain Sciences*, *36*(4), 377–392. https://doi.org/10.1017/S0140525X12003238

Ramscar, M., Dye, M., & McCauley, S. (2013). Error and expectation in language learning: The curious absence of 'mouses' in adult speech. *Language*, *89*, 760–793. https://doi.org/10.1353/lan.2013.0068

Sahaï, A., Desantis, A., Grynszpan, O., Pacherie, E., & Berberian, B. (2019). Action co-representation and the sense of agency during a joint Simon task: Comparing human and machine co-agents. *Consciousness and Cognition*, *67*, 44–55. https://doi.org/10.1016/j.concog.2018.11.008

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

Sebanz, N., & Knoblich, G. (2009). Prediction in joint action: What, when, and where. *Topics in Cognitive Science*, *1*(2), 353–367. https://doi.org/10.1111/j.1756-8765.2009.01024.x

Sebanz, N., Knoblich, G., & Prinz, W. (2003). Representing others' actions: Just like one's own? *Cognition*, *88*(3), B11–B21. https://doi.org/10.1016/S0010-0277(03)00043-X

Sebanz, N., Knoblich, G., & Prinz, W. (2005). How two share a task: Corepresenting stimulus-response mappings. *Journal of Experimental Psychology. Human Perception and Performance*, *31*(6), 1234–1246. https://doi.org/10.1037/0096-1523.31.6.1234

Signoret, C., Andersen, L. M., Dahlström, Ö., Blomberg, R., Lundqvist, D., Rudner, M., & Rönnberg, J. (2020). The Influence of Form- and Meaning-Based Predictions on Cortical Speech Processing Under Challenging Listening Conditions: A MEG Study. *Frontiers in Neuroscience*, *14*. https://www.frontiersin.org/articles/10.3389/fnins.2020.573254

Skantze, G. (2016). Real-Time Coordination in Human-Robot Interaction Using Face and Voice. *AI Magazine*, *37*(4), Article 4. https://doi.org/10.1609/aimag.v37i4.2686

Strijkers, K., & Costa, A. (2011). Riding the lexical speedway: A critical review on the time course of lexical selection in speech production. *Frontiers in Psychology*, *2*, 356. https://doi.org/10.3389/fpsyg.2011.00356

Strijkers, K., Costa, A., & Thierry, G. (2010). Tracking lexical access in speech production: Electrophysiological correlates of word frequency and cognate effects. *Cerebral Cortex (New York, N.Y.: 1991)*, *20*(4), 912–928. https://doi.org/10.1093/cercor/bhp153

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*(2), 176–190. https://doi.org/10.1016/j.ijpsycho.2011.09.015

Vesper, C., Abramova, E., Bütepage, J., Ciardo, F., Crossey, B., Effenberg, A., Hristova, D., Karlinsky, A., McEllin, L., Nijssen, S. R. R., Schmitz, L., & Wahn, B. (2017). Joint Action: Mental Representations, Shared Information and General Mechanisms for Coordinating with Others. *Frontiers in Psychology*, *7*. https://www.frontiersin.org/articles/10.3389/fpsyg.2016.02039

Wudarczyk, O., Kirtay, M., Kuhlen, A. K., Abdel Rahman, R., Haynes, J.-D., Hafner, V. V., & Pischedda, D. (2021). Bringing Together Robotics, Neuroscience, and Psychology: Lessons Learned From an Interdisciplinary Project. *Frontiers in Human Neuroscience*, *15*. https://www.frontiersin.org/article/10.3389/fnhum.2021.630789

Appendix A: List of words per semantic category

The words were selected after having 30 students of Aix-Marseille University evaluate a list of 500 words for concept typicality (Morrow & Duffy, 2005; Woollams, 2012). They rated how much each word was representative for the semantic category indicated in brackets in a 5-point Likert scale, where 1 corresponded to ‘not at all (representative)’ and 5 to ‘very much (representative)’. Items that had a mean of at least 3.0 were selected for the main experiment.

In a second phase, we checked that those words corresponded to the preferred name of the picture. For this, we had the MultiPic’s naming agreement to cover 377 words. The rest 73 words were normalized by having 30 students name the new pictures.

Accessoires	*Appareilles électronique*	*Constructions*	*Fruits*	*Instruments de musique*
bague	antenne	balcon	abricot	accordéon
barrette	aspirateur	barrière	ananas	banjo
bouclier	clavier	cathédrale	avocat	batterie
bouton	distributeur	chalet	banane	castagnettes
bracelet	écouteurs	chapiteau	cacahouète	clarinette
broche	écran	château	cerise	cloche
brosse	four	cheminée	châtaigne	cor français
canne	frigo	colonne	citron	cornemuse
chapelet	hélice	église	figue	cymbales
collier	imprimante	escalier	fraise	flûte
couronne	interrupteur	fontaine	framboise	gong
étiquette	magnétophone	labyrinthe	gland	grelot
éventail	micro	maison	grenade	guitare
lacet	micro-ondes	marche	kaki	harmonica
lunettes	ordinateur	moulin	kiwi	harpe
masque	parabole	mur	mandarine	lyre
médaille	pile	parking	melon	mandoline
montre	radiateur	pont	mûre	maracas
parapluie	radio	porte	myrtille	orgue
parfum	robot	puits	noix	piano
peigne	scanner	pyramide	olive	saxophone
perruque	sonnette	route	orange	sitar
pipe	souris	serre	papaye	tambour
poche	stéréo	stade	pastèque	tambourin
portefeuille	télécommande	statue	pêche	triangle
ruban	téléphone	toit	pistache	trombone
sac	télévision	tour	poire	trompette
talon	torche	trottoir	pomme	violon
tétine	ventilateur	tunnel	raisin	violoncelle
valise	prise	usine	tomate	xylophone

Mammifères	*Nature*	*Nourriture*	*Outils*	*Papeterie*
âne	arbre	artichaut	aiguille	agrafeuse
cerf	blé	asperge	aimant	cahier
chat	branche	beurre	ampoule	calculatrice
cheval	cactus	bonbon	balai	calendrier
chèvre	champignon	brocoli	cadenas	carte
chien	désert	café	chaîne	carton
cochon	éclair	carotte	clé	ciseaux
écureuil	feu	chocolat	clou	classeur
éléphant	feuille	chou-fleur	corde	compas
girafe	fleur	crevette	écrou	craie
gorille	iceberg	fromage	enclume	crayon
hippopotame	île	glace	faucille	enveloppe
kangourou	lac	hamburger	fil	équerre
koala	lune	jambon	filet	feutre
lama	marguerite	maïs	hache	gomme
lapin	montagne	miel	marteau	journal
lion	nid	moule	pelle	lettre
loup	nuage	muffin	perceuse	livre
mouton	planète	œuf	pinceau	mappemonde
ours	pluie	oignon	pioche	pochette
panthère	racine	paella	râteau	punaise
putois	rivière	pain	rouleau	règle
rat	rose	pizza	scie	scotch
renard	sapin	poulet	sécateur	stylo
rhinocéros	soleil	salade	serrure	surligneur
singe	tournesol	saucisse	tondeuse	tableau
taureau	trèfle	saucisson	tournevis	tampon
tigre	tronc	soupe	tronçonneuse	timbre
vache	vague	steak	tuyau	trombone
zèbre	volcan	viande	vis	trousse

Parties du corps	*Professions*	*Ustensiles de cuisine*	*Véhicules*	*Vêtements*
bouche	astronaute	assiette	ambulance	bavoir
bras	berger	balance	avion	béret
cerveau	boucher	bocal	barque	bretelles
cheveux	boxeur	bol	bateau	cape
cheville	bûcheron	bouilloire	bus	casquette
cou	chanteur	bouteille	camion	ceinture
coude	chasseur	cafetière	canoë	chapeau
dent	coiffeur	carafe	caravane	chaussette
doigt	danseur	casserole	charrette	chaussure
dos	dentiste	cocotte	dirigeable	chemise
épaule	facteur	couteau	fusée	collant
fesses	infirmière	couvercle	hélicoptère	costume
genou	jardinier	cuillère	limousine	cravate
jambe	juge	entonnoir	montgolfière	culotte
langue	livreur	fouet	pelleteuse	écharpe
main	marin	fourchette	radeau	gant
menton	médecin	louche	scooter	gilet
moustache	militaire	manche	skate	jupe
nez	mineur	mixeur	sous-marin	manteau
œil	peintre	mortier	tank	moufle
oreille	photographe	paille	taxi	pantalon
os	pilote	passoire	téléphérique	peignoir
pectoraux	plombier	planche à découper	tracteur	pull
pied	policier	plateau	train	robe
pouce	pompier	râpe	tricycle	short
poumon	prêtre	spatule	trottinette	soutien-gorge
sein	professeur	tasse	vélo	sweat-shirt
squelette	sculpteur	théière	voilier	tablier
veine	serveur	thermos	voiture	tee-shirt
visage	serveuse	verre	wagon	veste

Appendix B: Statistical tables for go-trials (behavioral)

Fixed effects	Estimate	Std. Error	t value	Pr(> |t|)
(Intercept)	6.908e+00	2.343e-02	294.8	<0.001 ***
Frequency1 (Low/High)	-3.694e-02	7.750e-03	-4.76	<0.001 ***
Block1 (1/2)	-1.449e-02	3.032e-02	-0.478	0.63277
Frequency1/Block1	-3.210e-04	4.657e-03	-0.06	0.94

Table 3.9: **Go trials: Naming Latencies analysis**. Summary LMER for the naming latencies (log-transformed) according to Frequency (Low, High), Block (1,2), and their interaction.

Fixed effects	Estimate	Std. Error	t value	Pr(> |t|)
(Intercept)	-1.43215	0.10374	-13.806	<0.001 ***
Frequency1 (Low/High)	0.52167	0.05688	9.172	<0.001 ***
Block1 (1/2)	-0.04045	0.05631	-0.718	0.473
Frequency1/Block1	0.06009	0.05642	1.065	0.287

Table 3.10: **Go trials: Error rates analysis**. Summary GMER for the error rates (binary coded) according to Frequency (Low, High), Block (1,2), and their interaction.

Appendix C : Statistical tables on pair-wise comparisons (ERP)

Contrast	Estimate	Std. Error	t value	Pr(> $\|t\|$)
(Other-go LF) - (No-go LF)	0.3515	0.0995	3.533	0.0023 **
(Other-go LF) - (Other-go HF)	0.2667	0.0995	2.681	0.0370 *
(Other-go LF) - (No-go HF)	0.3320	0.0995	3.338	0.0047 **
(No-go LF) - (Other-go HF)	-0.0848	0.0995	-0.852	0.8292
(No-go LF) - (No-go HF)	-0.0195	0.0995	-0.196	0.9973
(Other-go HF) - (No-go HF)	0.0653	0.0995	0.657	0.9132

Table 3.11: **No-go trials - Picture-locked. Time window 250-350 ms**. Summary for the pair-wise comparison (tukey method) on the electrophysiological activity when contrasting Frequency (Low, High) to Type of Turn (Other-go, No-go). Results are averaged over the levels of Region (Anterior, Central, Posterior).

Contrast	Estimate	Std. Error	t value	Pr(> $\|t\|$)
(Other-go LF) - (No-go LF)	0.3981	0.0985	4.041	0.0003
(Other-go LF) - (Other-go HF)	0.4272	0.0985	4.336	0.0001
(Other-go LF) - (No-go HF)	0.5443	0.0985	5.525	<0001
(No-go LF) - (Other-go HF)	0.0291	0.0985	0.295	0.9910
(No-go LF) - (No-go HF)	0.1462	0.0985	1.484	0.4472
(Other-go HF) - (No-go HF)	0.1171	0.0985	1.189	0.6340

Table 3.12: **No-go trials - Picture-locked. Time window 350-450 ms**. Summary for the pair-wise comparison (tukey method) on the electrophysiological activity when contrasting Frequency (Low, High) to Type of Turn (Other-go, No-go). Results are averaged over the levels of Region (Anterior, Central, Posterior).

Contrast	Estimate	Std. Error	t value	Pr(> $\|t\|$)
Category Block1 - Basic Block1	0.4695	0.0776	6.050	<0.0001
Category Block1 - Category Block2	0.1627	0.0776	2.096	0.1545
Category Block1 - Basic Block2	0.0933	0.0776	1.202	0.6254
Basic Block1 - Category Block2	-0.3069	0.0776	-3.954	0.0005
Basic Block1 - Basic Block2	-0.3762	0.0776	-4.848	<0.0001
Category Block2 - Basic Block2	-0.0693	0.0776	-0.894	0.8082

Table 3.13: **Other-go trials - Response-locked. Time window 400-600 ms**. Summary for the pair-wise comparison (tukey method) on the electrophysiological activity when contrasting Naming Level (Basic, Category) to Block (1,2). Results are averaged over the levels of Region (Anterior, Central, Posterior).

Appendix D

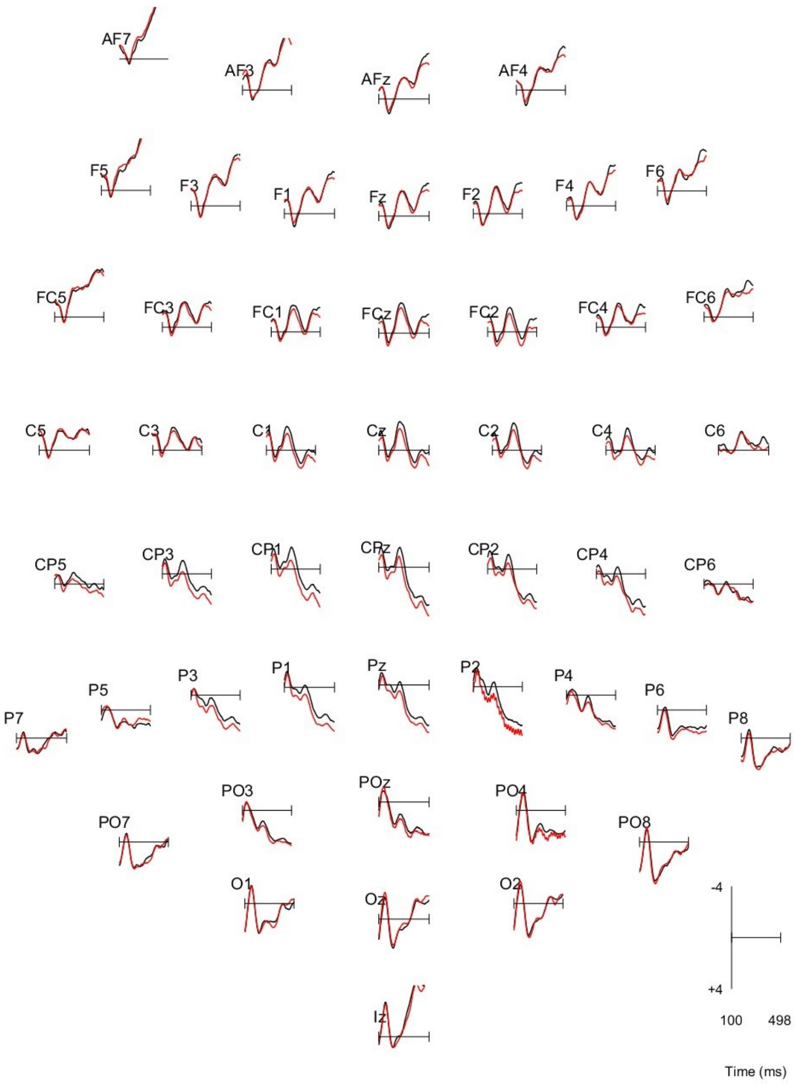

Figure 3.5: **EEG activity GO trials - Frequency effect**. The figure show grand average for each electrode included in the analysis. Red lines represent low-frequency words (LF) and black lines represent high-frequency words (HF).

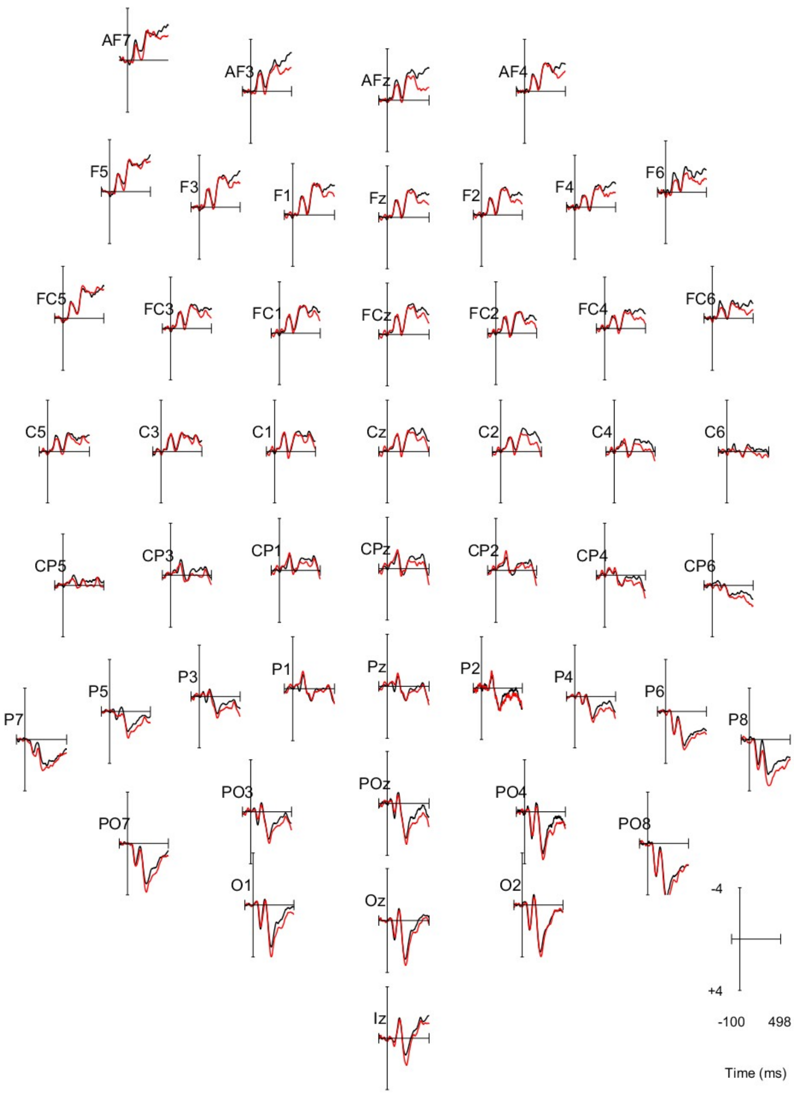

Figure 3.6: **EEG activity OTHER-GO trials - Frequency effect**. The figure show grand average for each electrode included in the analysis. Red lines represent low-frequency words (LF) and black lines represent high-frequency words (HF).

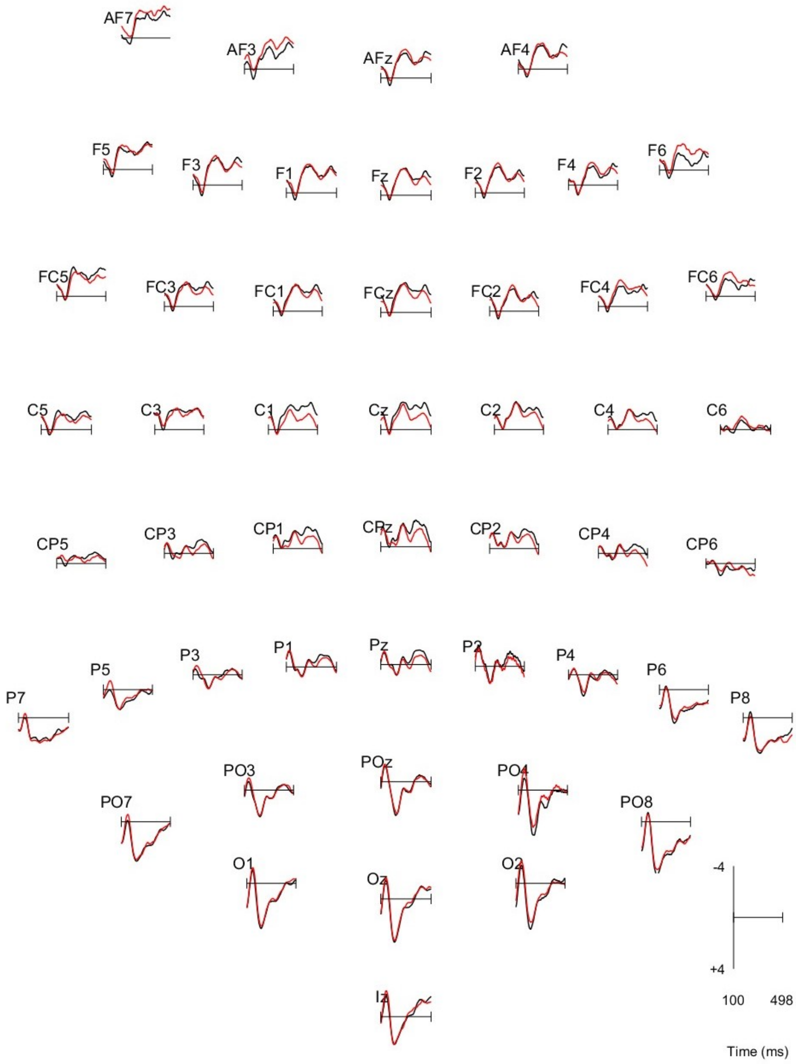

Figure 3.7: **EEG activity NO-GO trials - Frequency effect**. The figure shows grand average for each electrode included in the analysis. Red lines represent low-frequency words (LF) and black lines represent high-frequency words (HF).

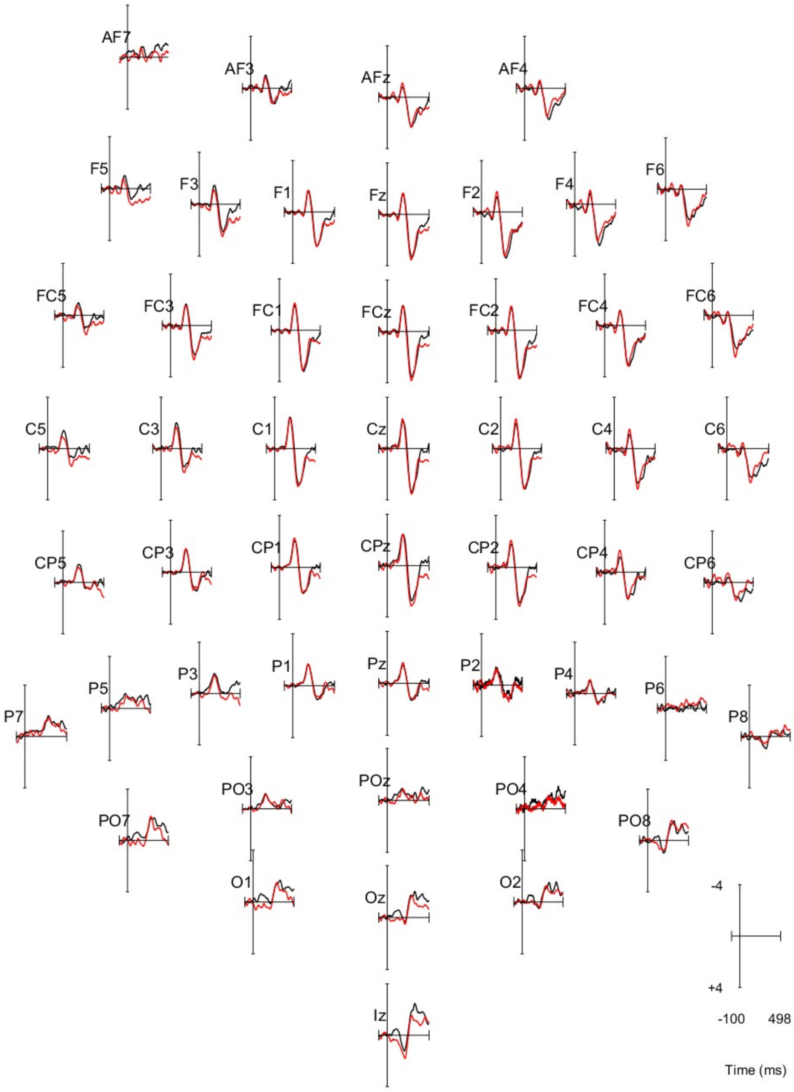

Figure 3.8: **EEG activity OTHER-GO trials - Adaptation effect: Block1**. The figure shows grand average for each electrode included in the analysis. Black lines represent the basic-level naming, while red lines correspond to the category naming. The activity (locked to the response of the robot) refers to the first part of the experiment (block 1).

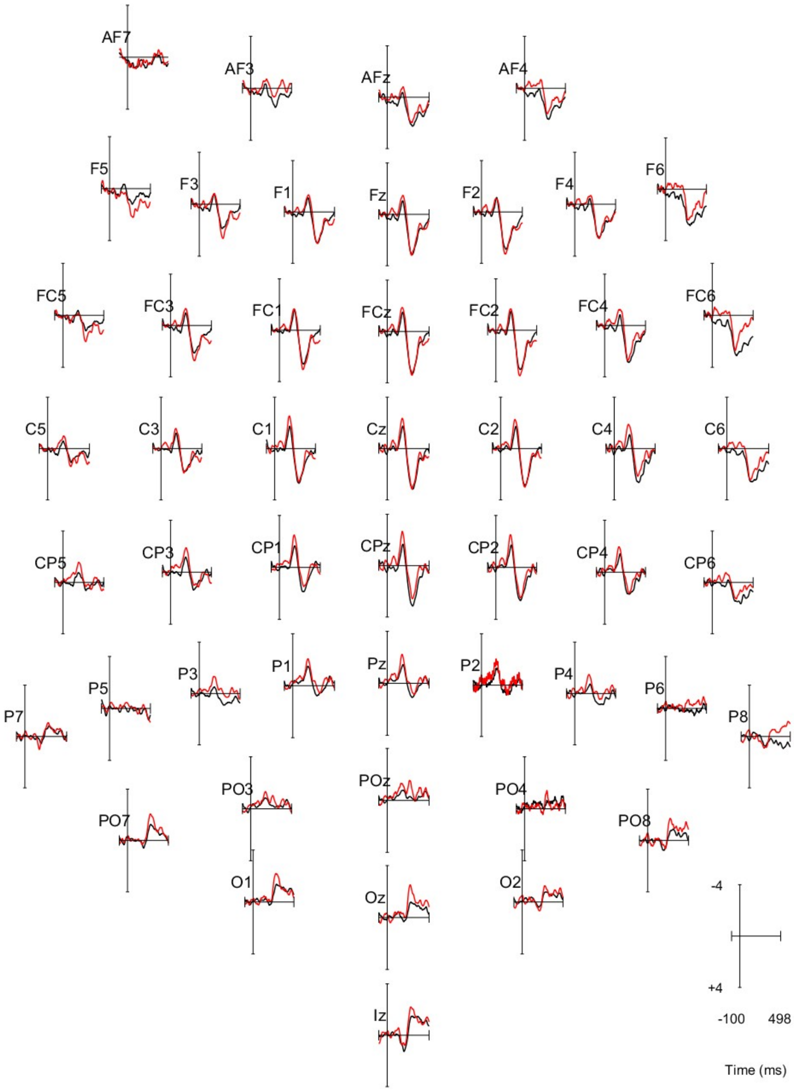

Figure 3.9: **EEG activity OTHER-GO trials - Adaptation effect: Block2**. The figure shows grand average for each electrode included in the analysis. Black lines represent the basic-level naming, while red lines correspond to the category naming. The activity (locked to the response of the robot) refers to the second part of the experiment (block 2).

Chapter 4

Study 3: Conceptual alignment with social robot

4.1 Article 2: Conceptual alignment in a joint picture-naming task performed with a social robot

Cirillo, G., Runnqvist. E., Strijkers, Kristof, Nguyen, N., Baus, C., *Cognition*, 2022

Cognition 227 (2022) 105213

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Conceptual alignment in a joint picture-naming task performed with a social robot

Giusy Cirillo [a,b,*], Elin Runnqvist [a,b], Kristof Strijkers [a,b], Noël Nguyen [a,b], Cristina Baus [c]

[a] Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France
[b] Institute for Language, Communication and the Brain (ILCB), Marseille, France
[c] Department of Cognition, Development and Educational Psychology, University of Barcelona, Spain

ARTICLE INFO

Keywords:
Spoken word production
Conceptual alignment
Lexical alignment
Joint action
Artificial partner
Picture naming

ABSTRACT

In this study we investigated whether people conceptually align when performing a language task together with a robot. In a joint picture-naming task, 24 French native speakers took turns with a robot in naming images of objects belonging to fifteen different semantic categories. For a subset of those semantic categories, the robot was programmed to produce the superordinate, semantic category name (e.g., fruit) instead of the more typical basic-level name associated with an object (e.g., pear). Importantly, while semantic categories were shared between the participant and the robot (e.g., fruits), different objects were assigned to each of them (e.g., the object of 'a pear' for the robot and of 'an apple' for the participant). Logistic regression models on participants' responses revealed that they aligned with the conceptual choices of the robot, producing over the course of the experiment more superordinate names (e.g., saying 'fruit' to the picture of an 'apple') for those objects belonging to the same semantic category as where the robot produced a superordinate name (e.g., saying 'fruit' to the picture of a 'pear'). These results provide evidence for conceptual alignment affecting speakers' word choices as a result of adaptation to the partner, even when the partner is a robot.

1. Introduction

People engage in joint actions daily. When playing or dancing together and when talking to each other, individuals transmit and react to relevant information from their partners with the aim of making their performance smoother and faster. These joint actions are a fundamental part of social cognition, as they explain not only how humans' social bonds are established, but also how they mutate depending on the situation and the partner. An intrinsic characteristic of any joint action is alignment (Pickering & Garrod, 2007). Activity partners align their action representations through automatic imitation at different levels of representation (e.g., motor, cognitive).

In the context of language, alignment has been described as crucial for successful communication (Pickering & Garrod, 2006; Pickering & Garrod, 2013). Aligned interlocutors achieve similar mental representations of the communicative situation, which improves their mutual understanding. Numerous studies show that speakers mimic each other in a number of non-verbal behaviors, including facial expressions (Dimberg, Thunberg, & Elmehed, 2000) and gestures (Bergmann & Kopp, 2012; Louwerse, Dale, Bard, & Jeuniaux, 2012), as well as verbal behaviors at several levels of representation. For example, speakers align to each other in terms of articulatory patterns, such as accent, speech rate and other phonetic dimensions (Giles, Coupland, & Coupland, 1991; Pardo, 2006). They also adopt each other's referring expressions, such as word choices (Brennan & Clark, 1996; Garrod & Pickering, 2004) and sentence structures (Branigan, Pickering, & Cleland, 2000). Particularly convincing for the notion of alignment, is that speakers even copy atypical lexical responses such as rarely used synonyms (e.g., Brennan & Clark, 1996) and infrequent syntactic structures such as passives (e.g., Bock, 1986). Findings like these highlight that linguistic alignment is a powerful communicative mechanism, capable even of overriding more frequent verbal behaviors.

In the present study we explored linguistic alignment beyond the copying of verbal utterances and asked whether speaker's lexical choices are affected by alignment with the partner at the conceptual level. That is, whether speakers adopt their partner's conceptual patterns, leading to the production of infrequent, yet meaningful lexico-semantic choices to all lexical items belonging to that conceptual category. To do so, we adopted a joint production design.

Joint production settings in which two people share a linguistic task

* Corresponding author.
E-mail address: giusy.cirillo@univ-amu.fr (G. Cirillo).

https://doi.org/10.1016/j.cognition.2022.105213
Received 5 May 2021; Received in revised form 31 May 2022; Accepted 27 June 2022
Available online 5 July 2022

Fig. 1. Experimental setting.
Participant and robot are positioned alongside and face the same computer screen, where the pictures are shown.

have been used to study parity of lexical representations between speakers and listeners (language co-representation). In particular, joint picture-naming tasks have been employed to explore whether lexical representations are shared across speakers (e.g., Baus et al., 2014; Gambi, Van de Cavey, & Pickering, 2015; Kuhlen & Abdel Rahman, 2017). Most of those studies have focused on the interference and facilitation effects observed in participants' performance because of naming objects with a partner.

In the current study, the joint-production setting constituted the method of investigation, while linguistic (lexical and conceptual) alignment was the object of investigation. We asked participants to perform a picture-naming task together with a social robot — Furhat Robot, a 3D humanoid talking head which ensures a more socially immersive experience compared to a computer or a virtual agent (see Fig. 1). We made this methodological choice to accurately control the dynamics of the joint performance, and to easily manipulate the robot's lexical choices. We manipulated the robot's response for a regular subset of trials, in which the robot did not give the basic-level name of the item but its semantic category name instead (e.g., *fruit* instead of pear*)*. As people naturally tend to name objects using their basic-level name, a phenomenon referred as the basic-level advantage (Rogers & Patterson, 2007), we predicted that, if alignment at the conceptual level was to take place, it would result in a progressive adaptation to the behavioral patterns of the robot, evident by a copy of the robot's conceptual choices.

We aimed to show a naming pattern going beyond word repetition and address speakers' capacity to adapt to the conceptual language space of the robot. Importantly, in our experiment robot and participants named different items for the same semantic category (e.g., for the category 'fruits' the robot named a pear, while participant an apple). Consequently, the use of an atypical name (category name) for items belonging to the same category for which the robot employs a category name, would constitute strong evidence for adaptation at the conceptual level. In short, this would mean that speakers align conceptually rather than to simple lexical choices.

2. Methods

2.1. Participants

Twenty-four participants (5 men; age: $M = 22.25$ years, $SD = 2.9$, range = 18–30 years) participated in the study. Our sample size was based on previous studies of alignment for infrequent names and syntactic structures (e.g., Branigan, Pickering, Pearson, McLean, & Brown, 2011; Suffill, Kutasi, Pickering, & Branigan, 2021). All participants were native speakers of French with normal or corrected-to-normal vision. None of the participants reported any neurological disorders, psychiatric disorders, or speech/language impairments. The experiment was conducted in line with the ethical guidelines laid down in the 6th (2008) Declaration of Helsinki. Participants gave informed consent and received 10 euros for their participation.

2.2. Furhat robot

In our experiment we used Furhat (https://www.furhatrobotics.com/; (Al Moubayed, Beskow, Skantze, & Granström, 2012), a humanoid robot equipped with a sophisticated back-projection system with a 3D-printed mask which can resemble anyone's face (see Fig. 1). The felicitous use of Furhat for joint-activity settings has been demonstrated in numerous experiments, where the robot played as partner in both perception and production tasks (Birgit et al., 2019; Moubayed, Skantze, & Beskow, 2013; Skantze, 2016).

2.3. Materials

The set of visual stimuli consisted of 450 pictures. 377 of them were taken from the dataset MultiPic (Duñabeitia et al., 2018). The remaining 73 pictures were drawn by a professional designer who used MultiPic as model. The images represented objects belonging to 15 semantic categories (e.g., fruits, mammals, clothes), with 30 items in each category (see Appendix A).

Half of the items of each semantic category (15 items of 15 semantic categories = 225 items) were assigned to the participant's trials (go trials) and the other half to the robot's trials (other-go trials). The robot's responses were pre-recorded using the synthesized voice of Furhat Robotics and played via the robot's loudspeakers. They consisted of 465 productions (450 basic-level names and 15 category names) of unique monosyllabic, disyllabic and trisyllabic words of between 200 ms and 1000 ms of duration each. We created six sets of randomized naming latencies ($M = 930$ ms, $SD = 250$ ms, range = 500–1600 ms) to mimic the intra-individual variability in naming latency observed in human speakers. The robot was programmed to produce the basic-level name (e.g., *châtaigne* = chestnut) for items belonging to 10 semantic categories, and the semantic category name (e.g., *outil* = tool) for items corresponding to the remaining five semantic categories. Pictures were presented within a square, of which the color (blue or green) indicated whether it was the participant's or the robot's turn to speak. Rotation of semantic categories, items and color cue resulted in 12 experimental lists which were counterbalanced across participants using a Latin-square design. This ensured that all items equally appeared in the basic-level and superordinate-level conditions, and as robot and participant responses.

Finally, we evaluated the effect of lexical frequency on response times and accuracy (extracted from the Lexique database, New, Pallier, Ferrand, & Matos, 2001) to ensure that participants displayed typical language production effects as in isolated object naming, validating the current joint set-up with a robot partner. Detailed analyses considering lexical frequency can be found in Supplementary Materials.

2.4. Procedure

Before the task, participants were introduced to Furhat, and engaged in a 5-min conversation with it. Afterwards, they were instructed to complete a joint picture-naming task with it, and name as quickly as possible the objects within the square of the color they were assigned to.

Participant and robot were positioned alongside facing the same computer screen, on which the visual stimuli were displayed (see Fig. 1). Each trial started with a fixation cross (500 ms), followed by the picture with the color cue assigned to either participant or robot (3000 ms) and a white screen (500 ms) to separate the trials (see Fig. 2). Every 90 trials, participants took a 2-min pause. Verbal responses were recorded for each participant.

At the end of the experiment, participants filled out a questionnaire where they rated on 5-point Likert scales: a) their subjective view of and

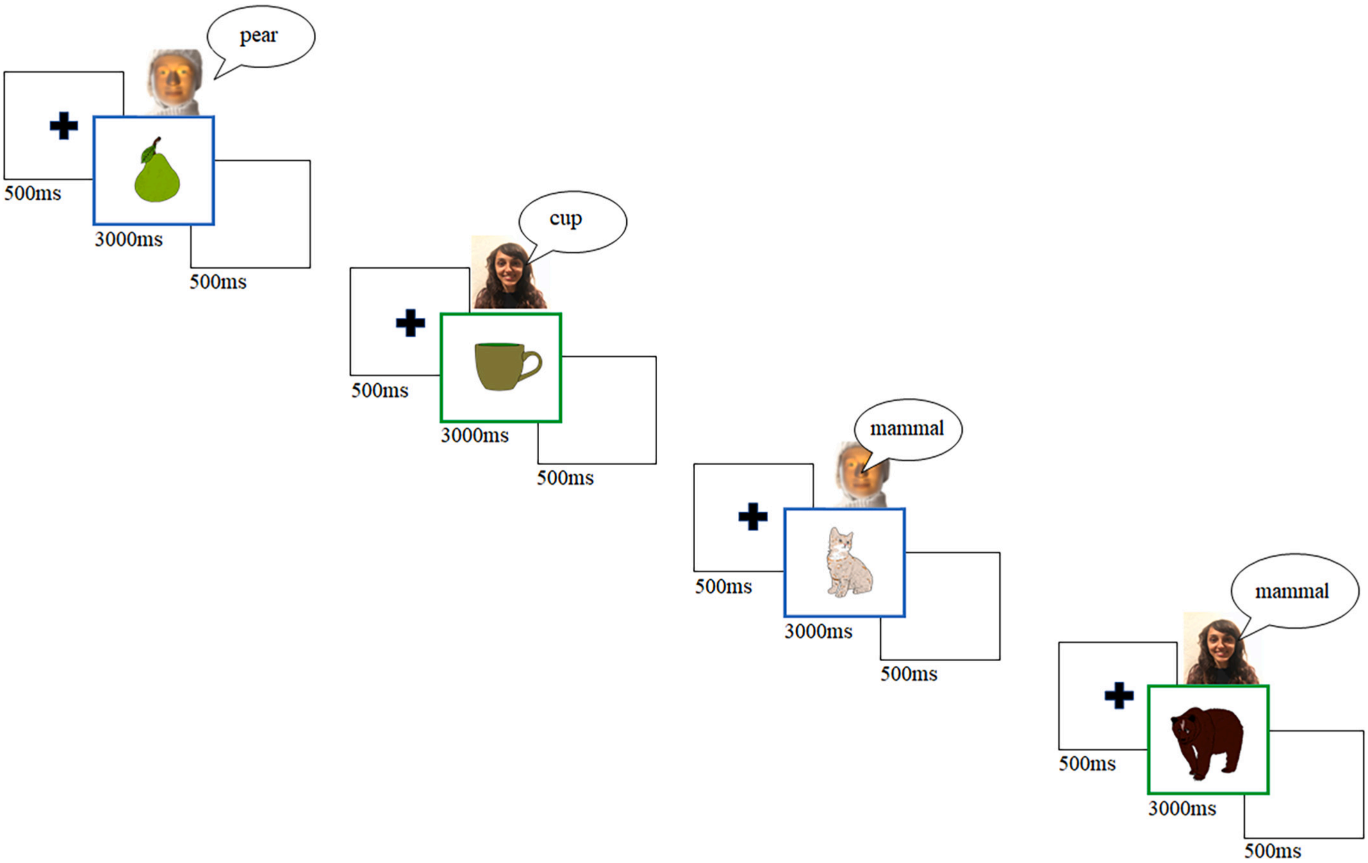

Fig. 2. Example of the trial structure of the experiment. In the first trial (on the top left), the robot produces the basic-level name of the object (basic condition), and the participant in the second trial does the same. In the third trial the robot produces the category name (category condition). In the fourth trial the participant uses the category name as well, as the item belongs to the same category. This is an example of response suggesting conceptual alignment within category.

familiarity with (humanoid) robots, b) experiment's difficulty, c) physical and social characteristics of Furhat (facial expression, voice, behavior, and sociability).

2.5. Data analysis

All data and scripts analyses are available in our project's OSF link (https://osf.io/f6gu3/). No participant was excluded from the analyses (i.e., all 24 participants were used). Responses were binary coded (0–1), representing whether participants used or not a category-related name during each trial. Category names (e.g., *tool* for *hammer*; $N = 176$) were fitted to a logits regression model (Bates, Mächler, Bolker, & Walker, 2015) with naming-level condition (basic vs. category; indicating whether the robot named the items of a given category with the basic-level or the category name) and block (1 to 4) as fixed factors and participant and item as random factors. We run a post-hoc power analysis on the mixed model data using the simr package in R (Green & MacLeod, 2016) with 200 iterations and reached a power of 65% (95% CI: 57.95, 71.59) for the effect of the naming-level condition.

Moreover, we used the quantitative variable 'order of trials' within an additional analysis of the responses on a trial-by-trial basis. All variables were contrast-coded using the Helmert contrast method, in which each level of a factor is contrasted to the mean of the previous ones. This method was implemented for the variable 'block', as we were able to compare each block to the average of the previous ones. Conceptual alignment would be indexed by significant differences in the number of category names employed by the participants for those items belonging to the same semantic category as when the robot used a category name versus when the robot used a basic-level name.

We also performed a series of Spearman's rank correlations between groups of Likert-scale questions in the final questionnaire and number of category responses. We then took the most representative questions per group to see whether conceptual alignment would correlate with certain beliefs about robots or whether category responses would be corroborated by stated awareness of one's adaptive/predictive behavior.

3. Results

The model exploring category responses across naming-level (basic vs. category) and block (1:4) revealed a significant effect of naming-level, showing that category names were produced more often in the category condition ($b = 4.257 \times 10^{-1}$, s.e. $= 1.059 \times 10^{-1}$, $z = 4.019$, $p < .001$; See Fig. 3), and showed a significant effect of block, indicating that this category naming effect was present throughout all four blocks, but was enhanced in the last block ($b = 1.729 \times 10^{-1}$, s.e. $= 6.225 \times 10^{-2}$, $z = 2.779$, $p = .005$; see Table 1).

Similarly, the analysis including 'order of trials' as a continuous variable revealed a significant effect of category responses by order (b =

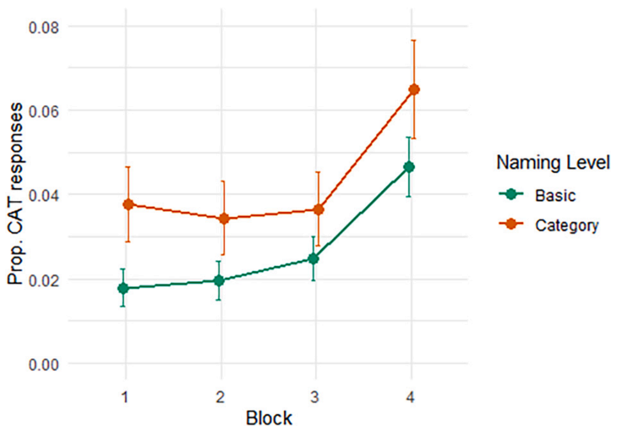

Fig. 3. Proportion of category responses given by the participants across block and naming level. The mean and standard error for all category responses made by the participants are plotted across the four blocks. The graph demonstrates that participants made more category responses for items belonging to the same semantic category where the robot also made a category response (plotted in orange) compared to category responses for semantic categories where the robot made basic-level responses (plotted in green). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

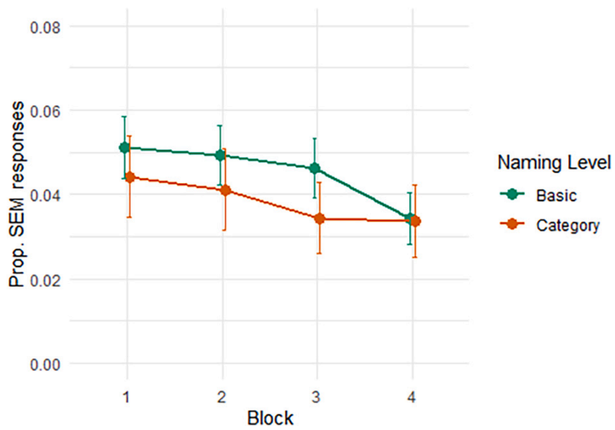

Fig. 4. Proportion of semantic-related errors given by the participants across block and naming level. The mean and standard error for the production of semantically related errors made by the participants are plotted across the four blocks. Participants made more semantically related errors in the beginning of the experiment than near the end of the experiment, regardless of whether the robot gave category (orange line) or basic-level responses (green line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
The table summarizes the model for participants' category responses explained by the variable block (1:4) and naming level condition (basic vs category). Significant results are marked in bold.

	b	s.e.	z value	p
(Intercept)	**−6.884**	**1.179**	**−5.835**	**<0.001**
BLOCK 1 (2 vs1)	−0.065	0.184	−0.354	0.722
BLOCK 2 (3 vs2 1)	0.143	0.099	1.449	0.147
BLOCK 3 (4 vs 3 2 1)	**0.172**	**0.062**	**2.775**	**0.005**
NAMING_LEVEL	**0.425**	**0.105**	**4.019**	**<0.001**
BLOCK 1: NAMING_LEVEL	−0.164	0.173	−0.949	0.342
BLOCK 2: NAMING_LEVEL	−0.102	0.099	−1.028	0.303
BLOCK 3: NAMING_LEVEL	−0.019	0.06	−0.323	0.746

Table 2
The table summarizes the model for participants' semantic-related responses explained by the variable block (1:4) and naming level condition (basic vs category). Significant results are marked in bold.

	b	s.e.	z value	p
(Intercept)	**−5.313**	**0.401**	**−13.219**	**<0.001**
BLOCK 1 (2–1)	−0.112	0.134	−0.835	0.403
BLOCK 2 (3−21)	−0.113	0.078	−1.441	0.149
BLOCK 3 (4–3 2 1)	**−0.128**	**0.058**	**−2.193**	**0.028**
NAMING_LEVEL	−0.104	0.084	−1.245	0.213
BLOCK 1: NAMING_LEVEL	0.041	0.128	0.321	0.748
BLOCK 2: NAMING_LEVEL	−0.019	0.083	−0.232	0.816
BLOCK 3: NAMING_LEVEL	−0.019	0.058	0.337	0.736

5.634×10^{-3}, s.e. = 1.832×10^{-3}, z = 3.076, *p* = .002), indicating that participants were more likely to produce category responses towards the end of the task.

In addition, we run a similar analysis on a second type of alternative responses, consisting in names belonging to the same semantic category of the target name, but that differed from it (semantically-related basic-level names**,** e.g., *bracelet* for *necklace; N* = 232) to ensure that the use of conceptual names was not strategically-driven (that is, participants giving a category name because they do not know or are uncertain about the correct basic-level response). If conceptual alignment was not determined by participants lacking knowledge of the pictures' names, we predicted a similar distribution of errors across conditions.

Analyses of the distribution of semantically-related errors revealed the inverse effect compared to the category naming effect, showing participants made more semantically related errors in the beginning of the experiment than at the end of the experiment (by block: b = − 0.128, s.e. = 0.058, z = −2.913, *p* = .028); by trial order (b = − 4.319×10^{-3}, s.e. = 1.474×10^{-3}, z = −2.93, *p* = .003; see Fig. 4 and Table 2). The number of semantically-related names was not different between naming-level conditions (*p* = .21).

Finally, results correlating category responses and questionnaire ratings are summarized in Appendix B. We found that category productions were negatively correlated with familiarity with robots (*r* = −0.445, *p* = .029) and imageability (*r* = −0.5, *p* < .012), and positively correlated with statement of adaptive prediction (*r* = 0.57, p = .003).

4. Discussion

In this study, we explored whether people align to the conceptual choice patterns of an artificial partner in a joint naming task. The robot was programmed to produce the semantic category name of objects belonging to 5 semantic categories (category condition), while for the rest of the trials it produced the basic-level name (basic-level condition). By doing this we were able to create an atypical, yet meaningful lexico-semantic choice pattern.

Our results show that participants named objects with the corresponding basic-level name most of the time, replicating standard picture-naming studies (Alario et al., 2004; Almeida, Knobel, Finkbeiner, & Caramazza, 2007). Importantly, they also reveal a significant alignment effect when comparing responses in the category-level condition to the basic-level condition. They produced more frequently the atypical category names for those items belonging to the same semantic category where the robot gave a category instead of basic-level naming response. This data pattern fits previous demonstrations where through alignment speakers copy infrequent lexical names or structures uttered by a human or computer partner (Bock, 1986; Brennan & Clark, 1996; Ivanova, Pickering, McLean, Costa, & Branigan, 2012). Our results go beyond the direct emulation of an atypical lexical label and show a conceptual alignment modifying an interlocutor's verbal behavior for previously unseen pictures. We observed that the robot's category response (e.g., naming the picture of a hammer as 'tool') caused the human interlocutor to utter a category name for different items of the same category (e.g., naming the previously not seen picture of a 'saw' as 'tool'). In addition, by performing fine-grained analyses over the course of the experiment for participants' category responses, we observed that this form of conceptual alignment is subject to two temporal dynamics: (1) It emerged from the first block, which we hypothesize to demonstrate very rapid, automatic adaptation to the robot's concepts, for example by enhancing the resting level activation of those category names where the robot gave a category response; (2) there was a significant increase in the participants' alignment in the final block, indicating an additive effect of possibly more conscious alignment with the robot's concepts, meaning that participants were aware for which categories the robot gave a category instead of basic-level response. The questionnaire ratings where participants reported they were trying to predict Furhat's behavior also fits this explanation of the additive effect registered in the final block. Future research should dig further into this option of having potentially two different linguistic adaptation mechanisms as suggested by our data.

The current observations of conceptual alignment cannot be accounted for by assuming that participants simply strategically updated their type of response nor that they were confused by the robot's replies of what the actual task was. Indeed, an explanation for this behavior would be that the category responses given by the participants were not an index of proper linguistic alignment, but rather used such category naming when participants were uncertain or did not know the basic-level name. However, several observations make this alternative account improbable. First, and this is the key observation in our study, conceptual alignment was most pronounced for those categories where the robot gave a category response (Fig. 3). Such contextual effect, where the frequency of a category response is directly associated with the robot's behavior, is a strong argument in favor of conceptual alignment. Second, the promptness with which this conceptual alignment emerged suggests it contains an automatic component and is not simply a strategic effect. Third, if participants were merely giving category responses when uncertainty about the basic-level name was high, then such effect should be particularly pronounced when error behavior is high as well (which is an index of such uncertainty). An additional analysis looking at the distribution of semantically-related errors over the experiment clearly argues against that option since most errors were made in the beginning of the experiment (Fig. 4), while the category naming effect was stronger at the end of the experiment (Fig. 3). Taken together, our data strongly support an interpretation in terms of conceptual alignment.

We also observed that participants gave more category names for items belonging to basic-level categories in the final block. Although this was significantly less compared to those items belonging to the same semantic category where the robot gave a category label, this generalization behavior most likely reflects spreading activation of conceptual choices. That is, participants' adaptation to category-level lexical concepts might have spilled over to other items (and the more conceptual adaption, the higher probability for these spill-over effects, as evidenced in the final block). More importantly for our purposes, this generalization behavior confirms the interpretation that the current data originates at the conceptual level of processing, rather than any other form of alignment or task-dependent strategies. In sum, the results are best understood following linguistic alignment theory and the adaptive link between language production and perception (Marsh, Richardson, & Schmidt, 2009; Pickering & Garrod, 2013).

Our findings support the notion that listeners update their productions in line with their partner's response patterns, and that they do so not simply by repeating the same lexical names, but by adopting the same conceptual word knowledge of the interlocutor. The behavioral shift of our participants who started imitating the response pattern of their artificial interlocutor can be understood as a way to connect to it, to bring about the same "change in the environment" as their partner (Sebanz, Bekkering, & Knoblich, 2006). Our results contribute and extend the theory of alignment in that this behavioral shift was characterized by a conceptual adaptation of the lexico-semantic patterns of the interlocutor.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2022.105213.

Author contributions

All authors developed the study concept and contributed to the study design. Testing, data collection and data analysis were performed by G. C. under the supervision of all authors. G.C. drafted the manuscript, and E.R., K.S., N·N and C.B. provided critical revisions. All authors approved the final version of the manuscript for submission.

Declaration of Competing Interest

The authors have no competing interests to declare.

Acknowledgements

This study has received funding from the "Investissements d'Avenir" French Government program managed by the French National Research Agency (reference: ANR-16-CONV-000× / ANR -17-EURE-00XX) and from Excellence Initiative of Aix-Marseille University - A*MIDEX" through the Institute of Language, Communication and the Brain. G.C. was supported by the Ecole Doctorale *Cognition*, Langage, Education / ED 356 of Aix-Marseille University. C.B. was supported by the Ramon y Cajal research program (RYC2018-026174-I). E.R. has benefited from support from the French government, managed by the French National Agency for Research (ANR) through a research grant (ANR-18- CE28-0013). K.S. was supported by a research grant of the ANR (ANR-16-CE28-0007-01).

We are grateful to Laurent Prévot for useful discussions and the technical help with the robot Furhat.

Appendix A. Word stimuli list grouped by semantic category

The words were selected after having 30 students of Aix-Marseille University evaluate a list of 500 words for concept typicality (Morrow & Duffy, 2005*; Woollams, 2012**). They rated how much each word was representative for the semantic category indicated in brackets in a 5-point Likert scale, where 1 corresponded to 'not at all (representative)' and 5 to 'very much (representative)'. Items that had a mean of at least 3.0 were selected for the main experiment.

In a second phase, we checked that those words corresponded to the preferred name of the picture. For this, we had the MultiPic's naming agreement to cover 377 words. The rest 73 words were normalized by having 30 students name the new pictures.

Accessoires	*Appareilles électronique*	*Constructions*	*Fruits*	*Instruments de musique*
bague	antenne	balcon	abricot	accordéon
barrette	aspirateur	barrière	ananas	banjo
bouclier	clavier	cathédrale	avocat	batterie
bouton	distributeur	chalet	banane	castagnettes
bracelet	écouteurs	chapiteau	cacahouète	clarinette
broche	écran	château	cerise	cloche
brosse	four	cheminée	châtaigne	cor français
canne	frigo	colonne	citron	cornemuse
chapelet	hélice	église	figue	cymbales
collier	imprimante	escalier	fraise	flûte
couronne	interrupteur	fontaine	framboise	gong
étiquette	magnétophone	labyrinthe	gland	grelot
éventail	micro	maison	grenade	guitare
lacet	micro-ondes	marche	kaki	harmonica
lunettes	ordinateur	moulin	kiwi	harpe
masque	parabole	mur	mandarine	lyre
médaille	pile	parking	melon	mandoline
montre	radiateur	pont	mûre	maracas
parapluie	radio	porte	myrtille	orgue
parfum	robot	puits	noix	piano
peigne	scanner	pyramide	olive	saxophone
perruque	sonnette	route	orange	sitar
pipe	souris	serre	papaye	tambour
poche	stéréo	stade	pastèque	tambourin
portefeuille	télécommande	statue	pêche	triangle
ruban	téléphone	toit	pistache	trombone
sac	télévision	tour	poire	trompette
talon	torche	trottoir	pomme	violon
tétine	ventilateur	tunnel	raisin	violoncelle
valise	prise	usine	tomate	xylophone

Mammifères	*Nature*	*Nourriture*	*Outils*	*Papeterie*
âne	arbre	artichaut	aiguille	agrafeuse
cerf	blé	asperge	aimant	cahier
chat	branche	beurre	ampoule	calculatrice
cheval	cactus	bonbon	balai	calendrier
chèvre	champignon	brocoli	cadenas	carte
chien	désert	café	chaîne	carton
cochon	éclair	carotte	clé	ciseaux
écureuil	feu	chocolat	clou	classeur
éléphant	feuille	chou-fleur	corde	compas
girafe	fleur	crevette	écrou	craie
gorille	iceberg	fromage	enclume	crayon
hippopotame	île	glace	faucille	enveloppe
kangourou	lac	hamburger	fil	équerre
koala	lune	jambon	filet	feutre
lama	marguerite	maïs	hache	gomme
lapin	montagne	miel	marteau	journal
lion	nid	moule	pelle	lettre
loup	nuage	muffin	perceuse	livre
mouton	planète	œuf	pinceau	mappemonde
ours	pluie	oignon	pioche	pochette
panthère	racine	paella	râteau	punaise
putois	rivière	pain	rouleau	règle
rat	rose	pizza	scie	scotch
renard	sapin	poulet	sécateur	stylo
rhinocéros	soleil	salade	serrure	surligneur
singe	tournesol	saucisse	tondeuse	tableau
taureau	trèfle	saucisson	tournevis	tampon
tigre	tronc	soupe	tronçonneuse	timbre
vache	vague	steak	tuyau	trombone
zèbre	volcan	viande	vis	trousse

Parties du corps	*Professions*	*Ustensiles de cuisine*	*Véhicules*	*Vêtements*
bouche	astronaute	assiette	ambulance	bavoir
bras	berger	balance	avion	béret
cerveau	boucher	bocal	barque	bretelles
cheveux	boxeur	bol	bateau	cape
cheville	bûcheron	bouilloire	bus	casquette
cou	chanteur	bouteille	camion	ceinture
coude	chasseur	cafetière	canoë	chapeau
dent	coiffeur	carafe	caravane	chaussette
doigt	danseur	casserole	charrette	chaussure
dos	dentiste	cocotte	dirigeable	chemise

(*continued on next page*)

(*continued*)

épaule	facteur	couteau	fusée	collant
fesses	infirmière	couvercle	hélicoptère	costume
genou	jardinier	cuillère	limousine	cravate
jambe	juge	entonnoir	montgolfière	culotte
langue	livreur	fouet	pelleteuse	écharpe
main	marin	fourchette	radeau	gant
menton	médecin	louche	scooter	gilet
moustache	militaire	manche	skate	jupe
nez	mineur	mixeur	sous-marin	manteau
œil	peintre	mortier	tank	moufle
oreille	photographe	paille	taxi	pantalon
os	pilote	passoire	téléphérique	peignoir
pectoraux	plombier	planche à découper	tracteur	pull
pied	policier	plateau	train	robe
pouce	pompier	râpe	tricycle	short
poumon	prêtre	spatule	trottinette	soutien-gorge
sein	professeur	tasse	vélo	sweat-shirt
squelette	sculpteur	théière	voilier	tablier
veine	serveur	thermos	voiture	tee-shirt
visage	serveuse	verre	wagon	veste

* Morrow, L. I., & Duffy, M. F. (2005). The representation of ontological category concepts as affected by healthy aging: Normative data and theoretical implications. *Behavior Research Methods*, *37*(4), 608–625. https://doi.org/10.3758/BF03192731

** Woollams, A. M. (2012). Apples are not the only fruit: The effects of concept typicality on semantic representation in the anterior temporal lobe. Frontiers in Human Neuroscience, 6. https://doi.org/10.3389/fnhum.2012.00085

References

Al Moubayed, S., Beskow, J., Skantze, G., & Granström, B. (2012). Furhat: a back-projected human-like robot head for multiparty human-machine interaction. In A. Esposito, A. M. Esposito, A. Vinciarelli, R. Hoffmann, & V. C. Müller (Eds.), *Cognitive Behavioural Systems (pp. 114–130)*. Springer. https://doi.org/10.1007/978-3-642-34584-5_9.

Alario, F.-X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (2004). Predictors of picture naming speed. *Behavior Research Methods, Instruments, & Computers, 36*(1), 140–155. https://doi.org/10.3758/BF03195559

Almeida, J., Knobel, M., Finkbeiner, M., & Caramazza, A. (2007). The locus of the frequency effect in picture naming: When recognizing is not enough. *Psychonomic Bulletin & Review, 14*(6), 1177–1182. https://doi.org/10.3758/BF03193109

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

Bergmann, K., & Kopp, S. (2012). Gestural alignment in natural dialogue. *Proceedings of the Annual Meeting of the Cognitive Science Society, 34*(34). https://escholarship.org/uc/item/73z0q063.

Birgit, R., Bruno, N., Morgane, B., Magalie, O., Laurent, P., & Thierry, C. (2019). Brain activity during reciprocal social interaction investigated using conversational robots as control condition. *Philosophical Transactions of the Royal Society, B: Biological Sciences, 374*(1771), 20180033. https://doi.org/10.1098/rstb.2018.0033

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

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

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

Dimberg, U., Thunberg, M., & Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychological Science, 11*(1), 86–89. https://doi.org/10.1111/1467-9280.00221

Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: a standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology, 71*(4), 808–816. https://doi.org/10.1080/17470218.2017.1310261

Gambi, C., Van de Cavey, J., & Pickering, M. J. (2015). Interference in joint picture naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(1), 1–21. https://doi.org/10.1037/a0037438

Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences, 8*(1), 8–11. https://doi.org/10.1016/j.tics.2003.10.016

Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: communication, context, and consequence. In *Contexts of accommodation: Developments in applied sociolinguistics* (pp. 1–68). Editions de la Maison des Sciences de l'Homme. https://doi.org/10.1017/CBO9780511663673.001.

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution, 7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Ivanova, I., Pickering, M. J., McLean, J. F., Costa, A., & Branigan, H. P. (2012). How do people produce ungrammatical utterances? *Journal of Memory and Language, 67*(3), 355–370. https://doi.org/10.1016/j.jml.2012.06.003

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

Louwerse, M. M., Dale, R., Bard, E. G., & Jeuniaux, P. (2012). Behavior matching in multimodal communication is synchronized. *Cognitive Science, 36*(8), 1404–1426. https://doi.org/10.1111/j.1551-6709.2012.01269.x

Marsh, K. L., Richardson, M. J., & Schmidt, R. C. (2009). Social connection through joint action and interpersonal coordination. *Topics in Cognitive Science, 1*(2), 320–339. https://doi.org/10.1111/j.1756-8765.2009.01022.x

Moubayed, S. A., Skantze, G., & Beskow, J. (2013). The furhat back-projected humanoid head–lip reading, gaze and multi-party interaction. *International Journal of Humanoid Robotics, 10*(01), 1350005. https://doi.org/10.1142/S0219843613500059

New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). Une base de données lexicales du français contemporain sur internet: LEXIQUETM//A lexical database for contemporary french : LEXIQUETM. *Annee Psychologique - ANNEE PSYCHOL, 101*, 447–462. https://doi.org/10.3406/psy.2001.1341

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

Pickering, M. J., & Garrod, S. (2006). Alignment as the basis for successful communication. *Research on Language and Computation, 4*(2), 203–228. https://doi.org/10.1007/s11168-006-9004-0

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences, 11*(3), 105–110. https://doi.org/10.1016/j.tics.2006.12.002

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences, 36*(04), 329–347. https://doi.org/10.1017/S0140525X12001495

Rogers, T. T., & Patterson, K. (2007). Object categorization: Reversals and explanations of the basic-level advantage. *Journal of Experimental Psychology: General, 136*(3), 451–469. https://doi.org/10.1037/0096-3445.136.3.451

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

Skantze, G. (2016). Real-time coordination in human-robot interaction using face and voice. *AI Magazine, 37*(4), 19–31. https://doi.org/10.1609/aimag.v37i4.2686

Suffill, E., Kutasi, T., Pickering, M. J., & Branigan, H. P. (2021). Lexical alignment is affected by addressee but not speaker nativeness. *Bilingualism: Language and Cognition, 24*(4), 746–757. https://doi.org/10.1017/S1366728921000092

Supplementary Materials

Data analysis of Naming latencies and error rates considering lexical frequency

Naming latencies and error rates analyses were carried out by fitting Generalized Linear Mixed Effects models using the lmr4 package in R (Bates et al., 2015), where frequency (low vs high), naming level condition (basic vs category) and block (1 to 4) were taken as fixed factors while item and participant were taken as random factors. Missing responses, hesitations and verbal responses which differed from the target name were excluded from the naming latencies analysis (N = 1000). In addition, values that were 2.5 standard deviations above or below the mean response time for high and low frequency words respectively were excluded (N= 117). Naming latencies were log-transformed for a better fit of the model.

Results of naming latencies analysis and error rates

The naming latency results revealed that low frequency words were produced significantly slower than high frequency words (Mean HF = 1148 ms and SD = 299 ms; Mean LF = 1241 ms and SD = 328 ms; b = 0.05144, s.e. = 0.006935, t = 7.418, $p < .001$). Moreover, participants responded faster towards the end of the experiment, and in particular in the fourth block as compared to the previous blocks (b = -0.03287, s.e. = 0.002135, t = -15.398 $p < .001$). However, there was no effect of naming level condition (p = .361) nor of any interaction ($p < 0.05$), showing that the frequency effect was overall distributed in the same way across both conditions.

The error rates showed a similar frequency effect, with low frequency words eliciting errors more often as compared to high frequency words (Prop. HF = 0.1 and SD = 0.3; Prop LF = 0.26 and SD = 0.4; b = 0.71, s.e. = 0.09, z = 7.36, $p < .001$). However, the model did not reveal any effect of the robot's response condition (p = .715), nor block ($p < .05$) nor an interaction between these ($p < .05$).

4.1.1 Descriptive statistics for category and semantic-related responses

Block	Naming Level	Mean	SD
1	Basic	0.017	0.13
2	Basic	0.019	0.13
3	Basic	0.024	0.15
4	Basic	0.046	0.21
1	Category	0.037	0.19
2	Category	0.034	0.18
3	Category	0.036	0.18
4	Category	0.065	0.24

Table 4.1: The table summarizes the mean and standard deviation values of category responses treated as a continuous variable across block and naming level.

Block	Naming Level	Mean	SD
1	Basic	0.051	0.22
2	Basic	0.049	0.21
3	Basic	0.046	0.21
4	Basic	0.034	0.18
1	Category	0.044	0.20
2	Category	0.041	0.19
3	Category	0.034	0.18
4	Category	0.033	0.18

Table 4.2: The table summarizes the mean and standard deviation of the semantic errors treated as a continuous variable across block and naming level.

Results from the final questionnaire

Topic	Representative question	r	p
Familiarity	*I am familiar with humanoid robots*	**-0.445**	**.029**
Imageability	*I can easily imagine interacting with robots*	**-0.5**	**.012**
Adaptive prediction	*My way of predicting changed during the experiment*	**0.57**	**.003**
Stated difficulty	*I found the experiment difficult*	-0.15	.341
Human-like appearance	*Rate how human-like the facial expressions are*	0.147	.492

Table 4.3: The results of the Spearman's rank correlations are shown for the most relevant questions of the final questionnaire as classified by topic.

Final Questionnaire

Avez-vous de l'expérience dans les jeux vidéo? ☐ OUI ☐ NON

Enfant, combien de temps avez-vous passé à jouer aux jeux vidéo par semaine ?

☐ Moins d'1h

☐ Entre 2-3h

☐ Entre 3-5 h

☐ Entre 5-8h

☐ Plus de 8h

Adolescent, combien de temps avez-vous passé à jouer aux jeux vidéo par semaine ?

☐ Moins d'1h

☐ Entre 2-3h

☐ Entre 3-5 h

☐ Entre 5-8h

☐ Plus de 8h

A quel genre de jeux vidéo avez-vous joué?

☐ Plate-forme

☐ Action-aventure

☐ FPS (First-Person Shooter)

☐ RPG (Role Playing Game)

☐ Stratégie

☐ Sport

☐ Course

☐ Racing

☐ Autre (veuillez préciser)

En moyenne, combien de temps passez-vous à jouer aux jeux vidéo par semaine ?

☐ Moins d'1h

☐ Entre 2-3h

☐ Entre 3-5 h

☐ Entre 5-8h

☐ Plus de 8h

Jeux vidéo pratiqués régulièrement :

☐ Plate-forme

☐ Action-aventure

☐ FPS (First-Person Shooter)

☐ RPG (Role Playing Game)

☐ Stratégie

☐ Sport

☐ Course

☐ Racing

☐ Autre (veuillez préciser)

Aimez-vous lire des livres de science-fiction ?

☐ Oui

☐ Non

Aimez-vous lire des livres de fantasy ?

☐ Oui

☐ Non

Avez-vous déjà interagit avec des robots humanoïdes (physiques ou virtuels) ? ☐ OUI ☐ NON

Si oui, merci de préciser :

____

____

____

____

Avez-vous déjà participé à des expériences avec des robots humanoïdes (physiques ou virtuels)?

 OUI NON

Veuillez indiquer votre ressenti à propos des robots humanoïdes (physiques ou virtuels):

Très négatif	Plutôt négatif	Neutre	Plutôt positif	Très positif
1	2	3	4	5

Êtes -vous d'accord avec les propositions suivantes ?

Je suis familier avec les des robots humanoïdes (physiques ou virtuels).

Fortement en désaccord	En désaccord	Neutre	D'accord	Fortement d'accord
1	2	3	4	5

Je peux facilement m'imaginer interagir avec des robots.

Fortement en désaccord	En désaccord	Neutre	D'accord	Fortement d'accord
1	2	3	4	5

Dans le future, nous interagirons quotidiennement avec les robots.

Fortement en désaccord	En désaccord	Neutre	D'accord	Fortement d'accord
1	2	3	4	5

Les robots me font peur

Fortement en désaccord	En désaccord	Neutre	D'accord	Fortement d'accord

Je ne peux pas m’imaginer interagir avec les robots

Fortement en désaccord	En désaccord	Neutre	D'accord	Fortement d'accord
1	2	3	4	5

Les robots n’ont pas d’intentionnalité

Fortement en désaccord	En désaccord	Neutre	D'accord	Fortement d'accord
1	2	3	4	5

Veuillez indiquer votre degré d'accord avec les phrases suivantes

J'ai trouvé l'expérience difficile

Fortement en désaccord	En désaccord	Neutre	D'accord	Fortement d'accord
1	2	3	4	5

J'étais bon à la tâche

Fortement en désaccord	En désaccord	Neutre	D'accord	Fortement d'accord
1	2	3	4	5

Furhat était bon à la tâche

Fortement en désaccord	En désaccord	Neutre	D'accord	Fortement d'accord
1	2	3	4	5

Je sentais qu'il y avait une coopération entre moi et Furhat (Nous avons tous les deux travaillé pour le succès du jeu)

Fortement en désaccord	En désaccord	Neutre	D'accord	Fortement d'accord
1	2	3	4	5

Le succès du jeu est principalement dû à moi

Fortement en désaccord	En désaccord	Neutre	D'accord	Fortement d'accord
1	2	3	4	5

Pendant toute la durée de l'expérience, j'ai essayé de comprendre les intentions de Furhat pendant la tâche

Fortement en désaccord	En désaccord	Neutre	D'accord	Fortement d'accord
1	2	3	4	5

Pendant toute la durée de l'expérience, j'étais toujours conscient(e) de jouer avec un robot

Fortement en désaccord	En désaccord	Neutre	D'accord	Fortement d'accord
1	2	3	4	5

Il y avait des moments où j'oubliais que je jouais avec le robot, et c'était comme si je jouais seul(e)

Fortement en désaccord	En désaccord	Neutre	D'accord	Fortement d'accord
1	2	3	4	5

Il y avait des moments où j'oubliais que je jouais avec un robot, et c'était comme jouer avec un humain

Fortement en désaccord	En désaccord	Neutre	D'accord	Fortement d'accord
1	2	3	4	5

Le robot rendait la tâche plus difficile que de jouer avec un humain

Fortement en désaccord	En désaccord	Neutre	D'accord	Fortement d'accord
1	2	3	4	5

Je pense que j'aurais mieux réussi si j'avais joué avec un humain

Fortement en désaccord	En désaccord	Neutre	D'accord	Fortement d'accord
1	2	3	4	5

La présence du robot m'a perturbé(e)

Fortement en désaccord	En désaccord	Neutre	D'accord	Fortement d'accord
1	2	3	4	5

J’ai remarqué qu’il y avait quelque chose de particulier dans les réponses de Furhat

Fortement en désaccord	En désaccord	Neutre	D'accord	Fortement d'accord
1	2	3	4	5

Je pense qu'il y avait une personne qui contrôlait le robot

Fortement en désaccord	En désaccord	Neutre	D'accord	Fortement d'accord
1	2	3	4	5

Je pense que le robot a commis des erreurs

Fortement en désaccord	En désaccord	Neutre	D'accord	Fortement d'accord
1	2	3	4	5

Je pense que le robot a sa propre façon de jouer et je l'ai acceptée

Fortement en désaccord	En désaccord	Neutre	D'accord	Fortement d'accord
1	2	3	4	5

J'ai eu du mal à comprendre ce que faisait le robot

Fortement en désaccord	En désaccord	Neutre	D'accord	Fortement d'accord
1	2	3	4	5

Je me suis adapté(e) au comportement de Furhat

Fortement en désaccord	En désaccord	Neutre	D'accord	Fortement d'accord
1	2	3	4	5

Ma performance a été influencée négativement par son comportement

Fortement en désaccord	En désaccord	Neutre	D'accord	Fortement d'accord
1	2	3	4	5

Ma performance a été influencée positivement par son comportement

Fortement en désaccord	En désaccord	Neutre	D'accord	Fortement d'accord
1	2	3	4	5

Avez-vous perçu un comportement anormal de la part de Furhat ? □ OUI □ NON

Si oui, veuillez indiquer le type de comportement que vous avez perçu comme anormal:

____

____

____

Les caractéristiques suivantes vous paraissent-elles humaines **:**

L'apparence physique de Furhat

Pas du tout humain	Plutôt inhumain	Neutre	Plutôt humain	Très humain
1	2	3	4	5

Les expressions faciales de Furhat

Pas du tout	Plutôt inhumain	Neutre	Plutôt humain	Très proche
1	2	3	4	5

La voix de Furhat

Pas du tout	Plutôt inhumain	Neutre	Plutôt humain	Très proche
1	2	3	4	5

Le comportement de Furhat

Pas du tout	Plutôt inhumain	Neutre	Plutôt humain	Très proche
1	2	3	4	5

La sociabilité de Furhat

Pas du tout	Plutôt inhumain	Neutre	Plutôt humain	Très proche
1	2	3	4	5

Chapter 5

General Discussion

Current neuro-linguistic and psycho-linguistic research proposes to investigate the cognitive and neural underpinnings associated to language processing through the lens of the joint action theory (Brennan et al., 2010; Clark, 1996; Fusaroli et al., 2012; Fusaroli & Tylén, 2013; Gambi & Pickering, 2011, 2013; Garrod & Anderson, 1987; Hoedemaker & Meyer, 2019). The go/no-go design typical of task-sharing experiments approximates the alternation between speaking and listening which characterizes any kind of language exchange (Gambi & Pickering, 2011; Garrod & Pickering, 2009). The application of those designs has shown how partner-adaptive behaviors play a central role in language, supporting coordination and communicative intention (Kuhlen et al., 2012; Pearson et al., 2006). Language is highly complex and variable, and yet interlocutors understand each other in extremely various situations. They are able to extract cues and regularities during interaction, and use them to adapt their behavior accordingly (Shintel & Keysar, 2009). In the present dissertation, we examined interlocutors' adaptation across three relevant cognitive dimensions: memory, prediction and production. First, conversational partners come to share attention, or else they come to attend the world in a 'we' mode, adding to their own perspective the perspective of their partner, this way reformulating the common environment from an individual to a collective point of view (Shteynberg, 2015, 2018). This brings them to encode information differently, resulting in a significant processing boost. Secondly, they adapt their way of comprehending language according to the interlocutor via a constant update of the predictive processing mediating comprehension (Fitz & Chang, 2019; Pickering & Garrod, 2013b; Vujović, Ramscar, & Wonnacott, 2021). Third, adaptation is present at the production level via linguistic alignment, which corresponds to the adoption of each other's language choices and patterns at multiple levels of representation (semantic, phonological, syntactic, prosodic, Branigan et al., 2010; Pickering & Garrod, 2004b). All those steps are fundamental because they optimize the language interaction and explain how language processing is essentially effortless. In three empirical chapters, this thesis explored partner-adaptive mechanisms characterizing joint language production via task-sharing designs in human-human and human-robot interaction across different production stages (conceptualization, lexical selection, phonetic realization).

In Chapter 2, we investigated shared attention via a joint language task in which each participant was assigned a specific language dimension to respond to, namely semantics (animate/inanimate; semantic task) or phonetics (consonant/vowel; phoneme-monitoring task). Previous studies on joint language production have shown how sharing a task together with a conversational partner affects language processing, as people come to share their relevant representations and simulate their production plans (Baus et al., 2014; Gambi et al., 2015; Howard et al., 2006; Kuhlen et al., 2012). Beside language access, shared attention has been pointed out to affect memory and recall (Shteynberg et al., 2020). In particular, it is responsible of the Joint Memory Effect, an advantage in encoding and subsequent recall for objects and events that have been under the attention of one's partner as compered to information falling outside of it (Eskenazi et al., 2013; Shteynberg, 2010). This JME has proven to be selective: it is present for some features instead of others, depending on the interest people have to retain them for future exchanges (Elekes & Király, 2021; Elekes & Sebanz, 2020). Our approach aimed to explore 1) how shared attention modulates language processing and recall in the case when people are explicitly asked to attend to different dimensions/details of the same language item and 2) how this is further affected by whether they react to the objects alone (go trials), together with their partner (go-together trials), whether their partner responds alone (other-go trials) or whether no one attends to the items (no-go trials). Our results revealed that only participants accessing the phonetic level improved their performance when responding together with their partner, while the reaction times of those reacting to the animacy of the object did not change across turn conditions. Memory recall, on the other hand, was better for other-go trials compared to no-go trials (Elekes & Sebanz, 2020; Eskenazi et al., 2013), and, crucially, better for go-together trials compared to go-alone trials. This was true for both levels of language representation, but those participants reacting to the phonetic level recalled more items overall. Our findings outline a new dynamics of shared attention in joint language production, in which language processing is mediated by the specific level of representation attended, and in which social benefits are extracted from deeper language production stages (semantics, conceptualization).

In Chapter 3, we set out an EEG experiment to investigate 1) co-representation through lexical access and 2) the neural sensitivity to variations from expected words to less frequent lexico-semantic patterns in human-robot interaction. We implemented a joint task in which a human participant alternated with a humanoid robot in naming pictures of objects belonging to different semantic categories. We manipulated word frequency having half of the pictures corresponding to high frequency names and half to low frequency names to have a measure of lexical access. In addition, the robot was programmed to produce, for a regular subset of items, the category name (e.g., *tool*) instead of the typical basic-level name (*hammer*). We recorded participants' electrophysiological activity during the whole duration of the experiment. Related to the frequency manipulation, we found comparable ERP components when the participants prepared to speak (go trials) and when the robot prepared to speak (other-go trials) but not for no-go trials (no one naming). We interpreted the present finding by referring to the fact that participants were able to co-represent robot's task at the level of lexical selection. In this sense, we replicated Baus et al. (2014)'s study with an artificial partner. Related to the robot's response manipulation, the study revealed that category names elicited a different (more positive) activity after response onset compared to basic-level names. This was true in the first block but not in the second block, suggesting that participants successfully adapted to the interlocutor's atypical lexico-sematic pattern, and learned to predict it. Main theories of neural processing propose that the brain is a proactive organ, formulating predictions about the upcoming events and this way mediating and facilitating comprehension (Pickering & Gambi, 2018). Importantly, they implicate that the production system is actively involved in comprehension (Martin et al., 2018). During interaction, we form expectations of the upcoming speech and compare our expectation to what has been actually said/done (Pickering & Garrod, 2014b). Less frequent or surprising responses are integrated to update perception for future exchanges (Den Ouden, Kok, & De Lange, 2012). The present investigation aligns to those accounts, providing evidence for adaptive prediction at the lexico-semantic stage of language production.

In Chapter 4, we implemented a similar experiment as in Chapter 3 to explore adaptation at the production level in terms of conceptual alignment to less frequent lexico-semantic choices produced by the robot. Differently to the experiment in Chapter 3, here participants shared all categories with the robot. Previous psycho-linguistic investigations have shown how linguistic alignment is a strong phenomenon, as people can override automatic and more frequent language patterns and adopt atypical words (Brennan & Clark, 1996) and syntactic structures (Bock, 1986) in order to adapt to the partner. Studies focusing on joint production in human-robot interaction reveal that people's degree of alignment depends on what they believe about their interlocutor, suggesting that there exists a positive tendency to align to artificial partners (Branigan et al., 2011; Pearson et al., 2006). In the present study, our aim was to explore adaptation at the conceptual level, beyond the mere copying of words. For that, we monitored participant's responses over the course of the experiment. We found that participants aligned conceptually with the robot, producing increasingly more category names (e.g., saying 'tool' to the picture of an 'nail') for those items belonging to the same semantic category for which the robot produced the category name. Those results cast a new light on the alignment literature, providing evidence for conceptual adaptation of the lexico-semantic patterns of the interlocutor. In the following sections, we discuss around three main findings: selective attention in language processing, co-representation in human-human and human-robot interaction, and conceptual alignment.

5.1 The selective nature of shared attention in language processing

The first key observation we can extract from our experimental work is that shared attention in language processing is selective. Our findings demonstrate that the social epistemic account proposed by Elekes and Sebanz (2020) for the Joint Memory Effect (Eskenazi et al., 2013) is even more pertinent for language, in which different representations are associated to different processing stages (Hauk, 2016; Indefrey & Levelt, 2004; Levelt et al., 1999; Sahin et al., 2009). We showed that, within language, shared attention is modulated by 1) the specific linguistic feature attended (phonetic, lexical, semantic), and by 2) the type of action/performance it brings to (i.e., responding together or alone). Also, we found that shared attention affects cognition differently. While it enhances memory recall and incidental learning (Eagle & Leiter, 1964; Hulstijn, 2012; Tresselt & Mayzner, 1960), it is not always evident in production and decision making.

Shared attention in language is, therefore, additionally fine-grained. From a social perspective, the attentional focus is enlarged and boosted when humans attend/process something at the same time (Boothby et al., 2014; Mundy & Newell, 2007), even though not exactly the same thing. From a language processing perspective, attention strictly depends on the specific representation level addressed. Semantics and conceptualization, in particular, come to be prioritized as relevant information to retain compared to phonetics and phonology. We propose two possible explications to this. One explanation is that the conceptual and semantic levels of language production respond to more universal (sometimes even visual) cues (e.g., the concept of animacy is the same in all cultures), and people are able to identify them beyond the specific language system (i.e., beyond phonology and phonetics). As a consequence, semantics is accessed faster and more easily than phonology (Strijkers & Costa, 2011; van Turennout et al., 1997). The second explanation we propose is that semantic and conceptual information are used by people to reach common ground (Clark & Marshall, 1981; Elekes & Sebanz, 2020). Within the joint production framework, successful monitoring of conceptual states allows for establishing the situation models for comprehension and memory (Zwaan & Radvansky, 1998), leading for co-representation of one's partner task, intentions, and goals.

Selective attention is here intended as a prioritization mechanism (van Moorselaar & Slagter, 2020), and our work suggests that typical adults are able to use the other's perspective when this has a positive influence on behavior, indirectly improving learning (memory recall). This investigation was relevant to the general quest for adaptation in language processing as 1) it confirms the parity between production and comprehension (Gambi & Pickering, 2017), impacting all levels of language processing but in a fine-grained fashion, 2) stipulates linguistic coordination already at the level of attention, and 3) clarifies that those patterns are not automatically included in behavior, but rather they are calibrated according to the aim or the situation. Those findings are in line with Pickering and Garrod (2004a) when defining the role and limits of adaptation in interactive alignment. They argue that interlocutors monitor each other's production stages moderately as it would be too costly to integrate all features of the common ground, and would slow down the joint language action. Following this, interlocutors must select and prioritize information that can socially benefit the exchange.

5.2 Co-representation from human-human to human-robot interaction

A second characteristic of joint language production is linked to the ability, proper of interlocutors, to co-represent each other's task and language behavior (Gambi & Pickering, 2011; Garrod & Pickering, 2009). The selective account of shared attention that we formulated for language processing suggests that co-representation does not simply entail the intended state and final goal of the action (Schmitz, Vesper, Sebanz, & Knoblich, 2018), but, rather, is sequential, and partners co-represent also the detailed structure of their respective actions. Our work suggests that similar co-representation dynamics are potentially achieved with a (humanoid) robot as much as with a human, opening up future avenues testing to which extent co-representation is systematically required to follow a robot's speech. A robot is, in fact, the extreme case of what is perceived other from the self, falling into the category of 'stranger' or 'foreigner' (if not alien...). A robot is an artificial device, something we are not used to interact with, especially at this moment of society in which, while machines are increasingly present in our everyday activities, their functionalities remain widely unknown to most people. An intentional and more robust attention to monitor a robot's language behavior would therefore be needed to reach successful communication. This might be taken as a measure of the will to accept the partner and try to interact with him. It can also be linked to the novelty or the surprise of the situation, suggesting that more cognitive effort is reserved for novel and extreme situations, like that of having a talking head right beside us. In both cases, our investigation enriches both the joint action and the language processing frameworks, testing to which extent humans can attribute social intentions (of speaking, doing a joint task) to a robot (Kim et al., 2013; Kirtay et al., 2020; Marge et al., 2022). Regarding this aspect, our questionnaire at the end of the third experiment, where participants rated their belief and imageability about Furhat and robots in general as well as their (adaptive) behavior during the experiment, revealed a negative correlation between experience and imageability with robots and participants' (adaptive) performance. The less the experience with robots, the larger the adaptation (alignment) effect. In our opinion, this merits further investigations in future experiments.

We were able to find traces of co-representation in human-robot interaction at two production stages: lexical and lexico-semantic (conceptual). Lexical access is at the core of comprehension, as people create and model their thoughts via words. The ability to predict the upcoming word, along with its relevant features, is therefore essential during interaction (Baus et al., 2014). Words, however, are also often chosen on the basis of a broader context, based on knowledge, previous experience, situational constraints, and adapting to a person's conceptual plans represents a way of establishing common ground between interlocutors (Brennan & Hanna, 2009; Elekes & Sebanz, 2020; Hanna et al., 2003; Kobayashi, Yasuda, Igarashi, & Suzuki, 2020; Shintel & Keysar, 2009). We saw it in our first study, where participants co-represented semantics but not phonetics, and we saw it in our experiments on human-robot interaction, where the conceptual stage of word production affected participants' (behavioral and electrophysiological) response.

5.3 Conceptual alignment in human-robot interaction

Our last investigation concerned alignment in human-robot interaction. We showed that, when partners were actually sharing the same task and same categories (including those to which the robot was programmed to name the items with the semantic category name), they tended to *gradually* and *contextually* shift towards the same lexico-semantic pattern.

5.3.1 Conceptual alignment is gradual

We used the term 'gradually' in relation to the timing characterizing alignment: starting early during the interaction, but becoming more prominent across the task. We can hypothesize two different mechanisms, linked to the two different timings: (1) A rapid adaptation where the brain automatically adjusted to the new statistical realities, namely shifting activation levels towards the category-level labels, for instance enhancing for relevant categories their resting level of activation, and lowering activation levels of the basic-level labels. This adaptation effect was immediate (only a few category responses are required to adjust activation levels) and therefore it was already present at the beginning of the experiment (block 1). (2) A slower adaptation where the partner became conscious of the systematic robot's category-level naming for some of the categories, and strategically (more consciously) adjusted in light of this awareness (e.g., holding in working memory the rule that categories X and Y lead to a superordinate response). This slower, strategic adaptation required time, and was reflected in the boost of the effect we saw in the fourth block. It was also reflected by the different brain activity elicited by basic-level and category level names we registered in the first part but not in the second part of our EEG experiment. In sum, the potential explanation relies on a difference between automatic (but likely unconscious) adaptation versus strategic (and thus conscious) adaptation, which highlights the interest for future research.

5.3.2 Conceptual alignment is contextual

Conceptual alignment is also 'contextual', and participants showed the tendency to adopt the category names especially for those items belonging to the manipulated categories of the robot. Overall, we observed that the majority of responses participants gave were basic-level names. When we compare our work to previous research in alignment, we can observe that (1) the proportion of non-aligned responses are always larger than the proportion of aligned responses, and (2) when alignment effects do not concern repetition priming (lexical alignment), effect sizes of alignment are comparable to what we observed in the current study. Take for example Katherine Bock's seminal work on syntactic alignment (Bock, 1986). In her experiments, the alignment-effects were between 2% and 10% throughout the different experiments for different conditions, and similar numbers can be observed in other syntactic alignment work (e.g., Griffin & Bock, 2000). And while with lexical alignment/priming higher percentages are typically observed, this is not surprising since it concerns repetition behavior, given that the response consists in the same lexical label for the same stimulus (Pickering & Ferreira, 2008). Therefore, it is more appropriate to compare the size of our alignment effects with syntactic priming experiments, since our study does not concern repetition of the same response for the same input, but generalization of an infrequent conceptual label to an entirely new input. Here we were evaluating production of conceptual alignment on non-repeated items (i.e., items named by the robot were not subsequently presented to the participant). Hence, the relevant comparison is not the proportion of naming type A vs B (e.g., actives vs passives, or in our experiment: basic vs category names), but the proportion of increase in naming type A after priming A (e.g., how much more actives after an active prime: roughly 3-10%) and naming type B after priming B (e.g., how much more passives after a passive prime: roughly 2-9%). In a similar vein, for our experiment, the meaningful comparison was the proportion of category-responses in the basic-level vs category-level 'priming' conditions (alignment effect around 3%).

The fact that conceptual alignment is contextual explains also why our results cannot be reduced to a simple strategic effect (i.e., category naming when not knowing the basic-level name), as alignment was significantly enhanced for those categories where the robot also gave a category name. In addition, if this strategic effect was the explanation, then those category names should also have been more common when a lot of semantic errors were made (index of uncertainty). However, we found the exact reverse pattern: In our study, participants gave more category responses in the last block, but made more errors in the first blocks. In sum, conceptualization determining lexical choice is a fundamental stage of language production, which is used to guide partners across an interaction. Aligned responses are not errors, rather, they denote partner-adaptive behavior, and support and facilitate comprehension and production.

5.4 Limitations and future perspectives

This dissertation aimed to provide empirical evidence of the central role that adaptive mechanisms play in joint language production. In addition, it tried to illustrate its features across human-human and human-robot interaction. However, our work lacked a direct comparison between those two paradigms. In other words, we spoke of 'partner-adaptive behavior' without going into the details about the characteristics of the partner. This choice was mainly methodological: our first experiment on shared attention was hardly realizable with a robot. On the other hand, our last two studies which, for some aspects (i.e., lexical co-representation), found a counterpart in the literature (Baus et al., 2014), were even easier to perform with an artificial partner. The use of Furhat allowed us a perfect control in terms of timing and stimuli, possibly giving the impression to participants that this lexico-semantic pattern of naming some items with the semantic category label was proper of the robot as language partner. This way we simulated an idiosyncratic use of words and monitored progressive adjustment to it. In addition, we know that humans co-represent each other at both lexical and conceptual level (Garrod & Anderson, 1987; Pickering & Garrod, 2013b), while, to us, this had not been clearly demonstrated in human-robot interaction, and evidence about it is mixed (Branigan et al., 2010; Wudarczyk, Kirtay, Pischedda, et al., 2021). For future investigations, we aim to elaborate on the factor 'partner' when investigating language adaptation, possibly monitoring its dynamics when modulating the degrees of belief about the interlocutor, or even when playing with different categories of partner. In particular, we want to test whether we can find a common denominator among different types of partner that are all equally believed less competent and more naive than 'typical adults', and that, according to the mediated account of alignment (Branigan et al., 2011), would therefore draw to an increased co-representation/alignment. We would categorize those partners as 'simple systems' (e.g.,robots, children, 2L speakers) compared to the more 'complex systems' with whom we daily interact (e.g., 'typical' adults).

5.5 Conclusion

As any human behavior, language is a dynamic and multi-model system, shaped by the social and cultural environment (Dale, Fusaroli, Duran, & Richardson, 2013; Fusaroli & Tylén, 2013; Raczaszek-Leonardi & Kelso, 2008). **Adaptation** towards the interlocutor's linguistic choices is one important evidence for that, facilitating comprehension via a continuous update of knowledge- and experience-based expectations from the upcoming sensory input (Cohen & Kassis-Henderson, 2012). Language is co-represented at various levels, including phonetic, lexical and semantic (Pickering & Garrod, 2013a), thanks to the common representation codes shared between action (production) and action perception (comprehension) supporting speech dynamics (Pickering & Garrod, 2014a; Silbert, Honey, Simony, Poeppel, & Hasson, 2014). In production, this often results in alignment of word choice, syntactic construction, conceptual perspective, pronunciation and other forms of language behavior by both interlocutors (Garrod & Anderson, 1987; Giles, 2008; Pardo, 2006; Pickering & Garrod, 2004b, 2006). Our first aim was to provide empirical evidence of how language coordination emerge early during interaction, already at the attention level, and in particular how **shared attention** and **co-representation** influence production and improve memory and learning (Shteynberg, 2015; Shteynberg et al., 2020). Adaptive mechanisms are also affected by external, social factors such as **partner-specific** properties (Brennan & Hanna, 2009), including familiarity with the interlocutor (e.g., whether interacting with a close friend or a stranger), and the belief we have about his/her (language) knowledge, impacting the trajectory and the degree of **alignment** (Branigan et al., 2011). This aspect is of particular evidence when comparing human-human to **human-robot interaction**, or else a typical case of interaction to a paradigmatic case in which individuals find themselves in front of an artificial system, which they have little experience with (Kobayashi et al., 2020; Pearson et al., 2006). Thus, understanding how adaptive features emerge and are processed when sharing a language task with a robot can provide further insight into the limits and the advantages of adaptation in joint production.

In three empirical studies, this dissertation provided evidence for crucial partner-adaptive behaviors characterizing joint language production. We were able to illustrate relevant behavioral and electrophysiological markers related to adaptation, from shared attention (i.e., people attending the same objects and events) to co-representation (having people monitoring each other's task and language choices) and down to lexical alignment (having people copying the word choices). We also explored its impact on a number of cognitive processes, including memory, prediction and production, supporting comprehension as well as learning. Our research approach combined multiple language dimensions (phonetic, lexical, semantic), modulating the degree of social benefit in interaction. Our focus in human-robot interaction, in particular, allowed us to pinpoint to what extent those mechanisms are robust, that are present even (more) when interacting with a robot. Ultimately, the present work opens up new avenues on the boundaries between what it is considered natural and artificial when it comes to processing language.

References

Alario, F. X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (2004, February). Predictors of picture naming speed. *Behavior Research Methods, Instruments, & Computers*, *36*(1), 140–155. Retrieved 2019-10-30, from https://doi.org/10.3758/BF03195559 doi: 10.3758/BF03195559

Alday, P. M., & Kretzschmar, F. (2019, August). Speed-Accuracy Tradeoffs in Brain and Behavior: Testing the Independence of P300 and N400 Related Processes in Behavioral Responses to Sentence Categorization. *Frontiers in Human Neuroscience*, *13*, 285. Retrieved 2020-10-17, from https://www.frontiersin.org/article/10.3389/fnhum.2019.00285/full doi: 10.3389/fnhum.2019.00285

Almeida, J., Knobel, M., Finkbeiner, M., & Caramazza, A. (2007, December). The locus of the frequency effect in picture naming: When recognizing is not enough. *Psychonomic Bulletin & Review*, *14*(6), 1177–1182. Retrieved 2019-03-13, from http://www.springerlink.com/index/10.3758/BF03193109 doi: 10.3758/BF03193109

Altarriba, J., Kroll, J. F., Sholl, A., & Rayner, K. (1996, July). The influence of lexical and conceptual constraints on reading mixed-language sentences: evidence from eye fixations and naming times. *Memory & Cognition*, *24*(4), 477–492. doi: 10.3758/bf03200936

Arbel, Y., Spencer, K. M., & Donchin, E. (2011). The N400 and the P300 are not all that independent. *Psychophysiology*, *48*(6), 861–875. Retrieved 2022-09-08, from http://onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-8986.2010.01151.x (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1469-8986.2010.01151.x) doi: 10.1111/j.1469-8986.2010.01151.x

Astheimer, L., & Sanders, L. (2011, August). Predictability affects early perceptual processing of word onsets in continuous speech. *Neuropsychologia*, *49*, 3512–6. doi: 10.1016/j.neuropsychologia.2011.08.014

Atmaca, S., Sebanz, N., & Knoblich, G. (2011, June). The joint flanker effect: sharing tasks with real and imagined co-actors. *Experimental Brain Research. Experimentelle Hirnforschung. Experimentation Cerebrale*, *211*(3-4), 371–385. Retrieved 2020-09-24, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3102196/ doi: 10.1007/s00221-011-2709-9

Bainbridge, W., Hart, J., Kim, E., & Scassellati, B. (2011, October). The Benefits of Interactions with Physically Present Robots over Video-Displayed Agents. *I. J. Social Robotics*, *3*, 41–52. doi: 10.1007/s12369-010-0082-7

Bar, M. (2009, May). The proactive brain: memory for predictions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1235–1243. Retrieved 2019-09-30, from https://royalsocietypublishing.org/doi/full/10.1098/rstb.2008.0310 doi: 10.1098/rstb.2008.0310

Bar, M. (2021, April). From Objects to Unified Minds. *Current Directions in Psychological Science*, *30*(2), 129–137. Retrieved 2022-09-20, from https://doi.org/10.1177/0963721420984403 (Publisher: SAGE Publications Inc) doi: 10.1177/0963721420984403

Barry, C., Morrison, C. M., & Ellis, A. W. (1997, August). Naming the Snodgrass and Vanderwart Pictures: Effects of Age of Acquisition, Frequency, and Name Agreement. *Quarterly Journal of Experimental Psychology: Section A*, *50*(3), 560–585. Retrieved 2019-10-30, from http://search.ebscohost.com/login.aspx?direct=truedb=bthAN=7615692lang=frsite=ehost doi: 10.1080/027249897392026

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015, October). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. Retrieved 2021-04-18, from https://www.jstatsoft.org/index.php/jss/article/view/v067i01 (Number: 1) doi: 10.18637/jss.v067.i01

Baus, C., Sebanz, N., Fuente, V. d. l., Branzi, F. M., Martin, C., & Costa, A. (2014, November). On predicting others' words: Electrophysiological evidence of prediction in speech production. *Cognition*, *133*(2), 395–407. Retrieved 2019-03-13, from https://linkinghub.elsevier.com/retrieve/pii/S0010027714001437 doi: 10.1016/j.cognition.2014.07.006

Belhassein, K., Fernández-Castro, V., Mayima, A., Clodic, A., Pacherie, E., Guidetti, M., . . . Cochet, H. (2022, February). Addressing joint action challenges in HRI: Insights from psychology and philosophy. *Acta Psychologica*, *222*, 103476. Retrieved 2022-09-16, from https://www.sciencedirect.com/science/article/pii/S0001691821002262 doi: 10.1016/j.actpsy.2021.103476

Bell, A. (1984). Language Style as Audience Design. *Language in Society*, *13*(2), 145–204. Retrieved 2022-06-13, from https://www.jstor.org/stable/4167516 (Publisher: Cambridge University Press)

Bergmann, K., Branigan, H. P., & Kopp, S. (2015). Exploring the Alignment Space – Lexical and Gestural Alignment with Real and Virtual Humans. *Frontiers in ICT*, *2*. Retrieved 2021-04-12, from https://www.frontiersin.org/articles/10.3389/fict.2015.00007/full (Publisher: Frontiers) doi: 10.3389/fict.2015.00007

Bergmann, K., & Kopp, S. (2012). Gestural Alignment in Natural Dialogue. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *34*(34). Retrieved 2021-01-21, from https://escholarship.org/uc/item/73z0q063

Bock, J. K. (1986, July). Syntactic persistence in language production. *Cognitive Psychology*, *18*(3), 355–387. Retrieved 2021-04-13, from https://www.sciencedirect.com/science/article/pii/0010028586900046 doi: 10.1016/0010-0285(86)90004-6

Boothby, E. J., Clark, M. S., & Bargh, J. A. (2014, December). Shared Experiences Are Amplified. *Psychological Science*, *25*(12), 2209–2216. Retrieved 2022-03-17, from https://doi.org/10.1177/0956797614551162 (Publisher: SAGE Publications Inc) doi: 10.1177/0956797614551162

Bradley, M., & Keil, A. (2012, January). Event-Related Potentials (ERPs). In *Encyclopedia of Human Behavior: Second Edition* (pp. 79–85). doi: 10.1016/B978-0-12-375000-6.00154-3

Brandstetter, J., & Bartneck, C. (2017, December). Robots will dominate the use of our language. *Adaptive Behavior*, *25*(6), 275–288. Retrieved 2022-08-28, from https://doi.org/10.1177/1059712317731606 (Publisher: SAGE Publications Ltd STM) doi: 10.1177/1059712317731606

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000, May). Syntactic coordination in dialogue. *Cognition*, *75*(2), B13–B25. Retrieved 2021-04-13, from https://www.sciencedirect.com/science/article/pii/S0010027799000815 doi: 10.1016/S0010-0277(99)00081-5

Branigan, H. P., Pickering, M. J., Pearson, J., Mclean, J., & Nass, C. (2003, January). Syntactic alignment between computers and people: The role of belief about mental states. *Cognitive Science - COGSCI*.

Branigan, H. P., Pickering, M. J., Pearson, J., & McLean, J. F. (2010, September). Linguistic alignment between people and computers.

Journal of Pragmatics, *42*(9), 2355–2368. Retrieved 2021-04-13, from https://www.sciencedirect.com/science/article/pii/S0378216609003282 doi: 10.1016/j.pragma.2009.12.012

Branigan, H. P., Pickering, M. J., Pearson, J., McLean, J. F., & Brown, A. (2011, October). The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers. *Cognition*, *121*(1), 41–57. Retrieved 2022-05-09, from https://www.sciencedirect.com/science/article/pii/S0010027711001363 doi: 10.1016/j.cognition.2011.05.011

Branigan, H. P., Tosi, A., & Gillespie-Smith, K. (2016, November). Spontaneous lexical alignment in children with an autistic spectrum disorder and their typically developing peers. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *42*(11), 1821–1831. doi: 10.1037/xlm0000272

Breazeal, C., Dautenhahn, K., & Kanda, T. (2016, June). Social Robotics. In *Springer Handbook of Robotics* (pp. 1935–1972). (Journal Abbreviation: Springer Handbook of Robotics) doi: 10.1007/978-3-319-32552-1$_7$2

Brehm, L., Taschenberger, L., & Meyer, A. (2019, August). Mental representations of partner task cause interference in picture naming. *Acta Psychologica*, *199*, 102888. Retrieved 2022-08-04, from https://www.sciencedirect.com/science/article/pii/S0001691819300630 doi: 10.1016/j.actpsy.2019.102888

Brennan, S. E. (1991). LEXICAL ENTRAINMENT IN SPONTANEOUS DIALOG. , 4.

Brennan, S. E., & Clark, H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of experimental psychology. Learning, memory, and cognition*. doi: 10.1037/0278-7393.22.6.1482

Brennan, S. E., Galati, A., & Kuhlen, A. K. (2010). Two Minds, One Dialog. In *Psychology of Learning and Motivation* (Vol. 53, pp. 301–344). Elsevier. Retrieved 2021-08-06, from https://linkinghub.elsevier.com/retrieve/pii/S0079742110530081 doi: 10.1016/S0079-7421(10)53008-1

Brennan, S. E., & Hanna, J. E. (2009). Partner-Specific Adaptation in Dialog. *Topics in Cognitive Science*, *1*(2), 274–291. Retrieved 2022-06-13, from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1756-8765.2009.01019.x (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1756-8765.2009.01019.x) doi: 10.1111/j.1756-8765.2009.01019.x

Brown, E. C., & Brüne, M. (2012). The role of prediction in social neuroscience. *Frontiers in Human Neuroscience*, *6*. Retrieved 2019-03-13, from https://www.frontiersin.org/articles/10.3389/fnhum.2012.00147/full doi: 10.3389/fnhum.2012.00147

Bråten, S. (2007). Altercentric infants and adults: On the origins and manifestations of participant perception of others' acts and utterances. In *On being moved: From mirror neurons to empathy* (pp. 111–135). Amsterdam, Netherlands: John Benjamins Publishing Company. doi: 10.1075/aicr.68.11bra

Bubic, A., Von Cramon, D. Y., & Schubotz, R. I. (2010). Prediction, cognition and the brain. *Frontiers in Human Neuroscience*, *4*. Retrieved 2019-03-11, from `https://www.frontiersin.org/articles/10.3389/fnhum.2010.00025/full` doi: 10.3389/fnhum.2010.00025

Bäss, P., & Prinz, W. (2014, September). My partner is also on my mind: social context modulates the N1 response. *Experimental brain research*, *233*. doi: 10.1007/s00221-014-4092-9

Böckler, A., Knoblich, G., & Sebanz, N. (2012, December). Effects of a coactor's focus of attention on task performance. *Journal of Experimental Psychology. Human Perception and Performance*, *38*(6), 1404–1415. doi: 10.1037/a0027523

Cacioppo, J. T., & Patrick, W. (2008). *Loneliness: Human nature and the need for social connection*. New York, NY, US: W W Norton & Co. (Pages: xiv, 317)

Carrard, V. (2021). Non-verbal Adaptation to the Interlocutors' Inner Characteristics: Relevance, Challenges, and Future Directions. *Frontiers in Psychology*, *12*. Retrieved 2022-08-09, from `https://www.frontiersin.org/articles/10.3389/fpsyg.2021.612664`

Chaminade, T., & Cheng, G. (2009, May). Social cognitive neuroscience and humanoid robotics. *Journal of Physiology-Paris*, *103*(3), 286–295. Retrieved 2022-09-16, from `https://www.sciencedirect.com/science/article/pii/S0928425709000527` doi: 10.1016/j.jphysparis.2009.08.011

Chang, E. F., Rieger, J. W., Johnson, K., Berger, M. S., Barbaro, N. M., & Knight, R. T. (2010, November). Categorical Speech Representation in Human Superior Temporal Gyrus. *Nature neuroscience*, *13*(11), 1428–1432. Retrieved 2019-08-02, from `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2967728/` doi: 10.1038/nn.2641

Chase, H. W., Swainson, R., Durham, L., Benham, L., & Cools, R. (2011, April). Feedback-related negativity codes prediction error but not behavioral adjustment during probabilistic reversal learning. *Journal of Cognitive Neuroscience*, *23*(4), 936–946. doi: 10.1162/jocn.2010.21456

Chomsky, N. (1965). *Aspects of the Theory of Syntax* (50th ed.). The MIT Press. Retrieved 2022-10-12, from `https://www.jstor.org/stable/j.ctt17kk81z`

Cirillo, G., Runnqvist, E., Strijkers, K., Nguyen, N., & Baus, C. (2022, October). Conceptual alignment in a joint picture-naming task performed with a social robot. *Cognition*, *227*, 105213. Retrieved 2022-07-14, from https://www.sciencedirect.com/science/article/pii/S0010027722002013 doi: 10.1016/j.cognition.2022.105213

Clark, H. H. (1992). *Arenas of language use*. Chicago, IL, US: University of Chicago Press. (Pages: xviii, 419)

Clark, H. H. (1996). *Using Language*. Cambridge University Press. (Google-Books-ID: b8bLCgAAQBAJ)

Clark, H. H., & Marshall, C. R. (1981). Definite Knowledge and Mutual Knowledge. In A. K. Joshi, B. L. Webber, & I. A. Sag (Eds.), *Elements of Discourse Understanding* (pp. 10–63). Cambridge, UK: Cambridge University Press.

Clark, H. H., & Schaefer, E. F. (1987, January). Collaborating on contributions to conversations. *Language and Cognitive Processes*, *2*(1), 19–41. Retrieved 2022-10-15, from https://doi.org/10.1080/01690968708406350 (Publisher: Routledge _eprint: https://doi.org/10.1080/01690968708406350) doi: 10.1080/01690968708406350

Cohen, L., & Kassis-Henderson, J. (2012, October). Language use in establishing rapport and building relations : implications for international teams and management education. *Management Avenir*, *n° 55*(5), 185–207. Retrieved 2021-08-16, from https://www.cairn.info/revue-management-et-avenir-2012-5-page-185.htm (Bibliographie_available: 1 Cairndomain: www.cairn.info Cite Par_available: 1 Publisher: Management Prospective Ed.)

Connine, C. M., & Titone, D. (1996). Phoneme monitoring. *Language and Cognitive Processes*, *11*(6), 635–645. (Place: United Kingdom Publisher: Taylor & Francis) doi: 10.1080/016909696387042

Cooper, S., Gow, S., Fensome, S., Dragone, M., & Kourtis, D. (2020, October). *An EEG investigation on planning human-robot handover tasks.* TechRxiv. Retrieved 2022-08-01, from https://www.techrxiv.org/articles/preprint/An$_E EG_i nvestigation_o n_p lanning_h uman-robot_h andover_t asks$/12950342/2 doi: 10.36227/techrxiv.12950342.v2

Corps, R., Gambi, C., & Pickering, M. (2017, June). Coordinating Utterances During Turn-Taking: The Role of Prediction, Response Preparation, and Articulation. *Discourse Processes*, *55*, 1–11. doi: 10.1080/0163853X.2017.1330031

Costa, A., Pickering, M. J., & Sorace, A. (2008, June). Alignment in second language dialogue. *Language and Cognitive Processes*, *23*(4), 528–556. Retrieved 2021-04-13, from https://doi.org/10.1080/01690960801920545 (Publisher: Routledge _eprint: https://doi.org/10.1080/01690960801920545) doi: 10.1080/01690960801920545

Cross, E. S., Riddoch, K. A., Pratts, J., Titone, S., Chaudhury, B., & Hortensius, R. (2019, April). A neurocognitive investigation of the impact of socializing with a robot on empathy for pain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1771), 20180034. Retrieved 2022-09-16, from `https://royalsocietypublishing.org/doi/10.1098/rstb.2018.0034` (Publisher: Royal Society) doi: 10.1098/rstb.2018.0034

Dale, R., Fusaroli, R., Duran, N. D., & Richardson, D. C. (2013, January). Chapter Two - The Self-Organization of Human Interaction. In B. H. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. 59, pp. 43–95). Academic Press. Retrieved 2022-09-14, from `https://www.sciencedirect.com/science/article/pii/B9780124071872000022` doi: 10.1016/B978-0-12-407187-2.00002-2

D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009, March). The Motor Somatotopy of Speech Perception. *Current Biology*, *19*(5), 381–385. Retrieved 2019-09-30, from `http://www.sciencedirect.com/science/article/pii/S0960982209005569` doi: 10.1016/j.cub.2009.01.017

Dean, L. G., Kendal, R. L., Schapiro, S. J., Thierry, B., & Laland, K. N. (2012, March). Identification of the social and cognitive processes underlying human cumulative culture. *Science.*, *335*(6072), 1114–1118. Retrieved 2022-03-14, from `http://dx.doi.org/10.1126/science.1213969` (Number: 6072 Publisher: American Association for the Advancement of Science) doi: 10.1126/science.1213969

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*, 283–321. (Place: US Publisher: American Psychological Association) doi: 10.1037/0033-295X.93.3.283

Dell, G. S., & Chang, F. (2014, January). The P-chain: relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1634). Retrieved 2022-07-31, from `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3866424/` (Publisher: The Royal Society) doi: 10.1098/rstb.2012.0394

Den Ouden, H., Kok, P., & De Lange, F. (2012). How Prediction Errors Shape Perception, Attention, and Motivation. *Frontiers in Psychology*, *3*. Retrieved 2022-08-10, from `https://www.frontiersin.org/articles/10.3389/fpsyg.2012.00548`

Dijksterhuis, A., & Bargh, J. A. (2001). The perception–behavior expressway: Automatic effects of social perception on social behavior. In *Advances in experimental social psychology, Vol. 33* (pp. 1–40). San Diego, CA, US: Academic Press.

Dikker, S., & Pylkkanen, L. (2011, July). Before the N400: Effects of lexical–semantic violations in visual cortex. *Brain and Language*, *118*(1-2), 23–28. Retrieved 2019-06-28, from `https://linkinghub.elsevier.com/retrieve/pii/S0093934X11000381` doi: 10.1016/j.bandl.2011.02.006

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992, October). Understanding motor events: a neurophysiological study. *Experimental Brain Research*, *91*(1), 176–180. Retrieved 2022-07-21, from `https://doi.org/10.1007/BF00230027` doi: 10.1007/BF00230027

Dubuisson Duplessis, G., Langlet, C., Clavel, C., & Landragin, F. (2021, June). Towards alignment strategies in human-agent interactions based on measures of lexical repetitions. *Language Resources and Evaluation*, *55*(2), 353–388. Retrieved 2022-08-12, from `https://doi.org/10.1007/s10579-021-09532-w` doi: 10.1007/s10579-021-09532-w

Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018, April). MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology*, *71*(4), 808–816. Retrieved 2019-09-30, from `http://journals.sagepub.com/doi/10.1080/17470218.2017.1310261` doi: 10.1080/17470218.2017.1310261

Eagle, M., & Leiter, E. (1964). Recall and recognition in intentional and incidental learning. *Journal of Experimental Psychology*, *68*(1), 58–63. (Place: US Publisher: American Psychological Association) doi: 10.1037/h0044655

Eitam, B., & Higgins, E. T. (2010). Motivation in mental accessibility: Relevance of a Representation (ROAR) as a new framework. *Social and Personality Psychology Compass*, *4*(10), 951–967. (Place: United Kingdom Publisher: Wiley-Blackwell Publishing Ltd.) doi: 10.1111/j.1751-9004.2010.00309.x

Elekes, F., Brody, G., Halász, E., & Király, I. (2015, December). Enhanced encoding of the co-actor's target stimuli during a shared non-motor task. *Quarterly journal of experimental psychology (2006)*, *69*, 1–33. doi: 10.1080/17470218.2015.1120332

Elekes, F., & Király, I. (2021, January). Attention in naïve psychology. *Cognition*, *206*, 104480. Retrieved 2022-03-18, from `https://www.sciencedirect.com/science/article/pii/S0010027720302997` doi: 10.1016/j.cognition.2020.104480

Elekes, F., & Sebanz, N. (2020, February). Effects of a partner's task on memory for content and source. *Cognition*, *198*. doi: 10.1016/j.cognition.2020.104221

Eskenazi, T., Doerrfeld, A., Logan, G. D., Knoblich, G., & Sebanz, N. (2013). Your words are my words: Effects of acting together on encoding. *The Quarterly Journal of Experimental Psychology*, *66*(5), 1026–1034. Retrieved 2021-08-19, from https://www.academia.edu/19467297/Your$_{w}ords_{a}re_{m}y_{w}ords_{E}ffects_{o}f_{a}cting_{t}ogether_{o}n_{e}ncoding$

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *European Journal of Neuroscience*, *15*(2), 399–402. Retrieved 2019-09-30, from http://onlinelibrary.wiley.com/doi/abs/10.1046/j.0953-816x.2001.01874.x doi: 10.1046/j.0953-816x.2001.01874.x

Fairs, A., Michelas, A., Dufour, S., & Strijkers, K. (2021, January). The Same Ultra-Rapid Parallel Brain Dynamics Underpin the Production and Perception of Speech. *Cerebral Cortex Communications*, *2*(3), tgab040. Retrieved 2022-03-25, from https://doi.org/10.1093/texcom/tgab040 doi: 10.1093/texcom/tgab040

Fitch, W. T., Huber, L., & Bugnyar, T. (2010, March). Social Cognition and the Evolution of Language: Constructing Cognitive Phylogenies. *Neuron*, *65*(6), 795–814. Retrieved 2022-09-16, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4415479/ doi: 10.1016/j.neuron.2010.03.011

Fitz, H., & Chang, F. (2019, June). Language ERPs reflect learning through prediction error propagation. *Cognitive Psychology*, *111*, 15–52. doi: 10.1016/j.cogpsych.2019.03.002

Fjaellingsdal, T. G., Schwenke, D., Ruigendijk, E., Scherbaum, S., & Bleichner, M. G. (2020, March). Studying brain activity during word-by-word interactions using wireless EEG. *PLOS ONE*, *15*(3), e0230280. Retrieved 2022-08-20, from https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0230280 (Publisher: Public Library of Science) doi: 10.1371/journal.pone.0230280

Foltz, A., Gaspers, J., Thiele, K., Stenneken, P., & Cimiano, P. (2015, February). Lexical alignment in triadic communication. *Frontiers in Psychology*, *6*. Retrieved 2020-12-11, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4327576/ doi: 10.3389/fpsyg.2015.00127

Frauenfelder, U. H., & Segui, J. (1989, March). Phoneme monitoring and lexical processing: evidence for associative context effects. *Memory & Cognition*, *17*(2), 134–140. doi: 10.3758/bf03197063

Freeth, M., Foulsham, T., & Kingstone, A. (2013, January). What Affects Social Attention? Social Presence, Eye Contact and Autistic Traits. *PLoS ONE*, *8*(1), e53286. Retrieved 2022-09-15, from `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3541232/` doi: 10.1371/journal.pone.0053286

Freunberger, D., & Roehm, D. (2016, October). Semantic prediction in language comprehension: evidence from brain potentials. *Language, Cognition and Neuroscience*, *31*(9), 1193–1205. Retrieved 2022-09-25, from `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5080973/` doi: 10.1080/23273798.2016.1205202

Frith, C. D., & Frith, U. (1999, November). Interacting minds–a biological basis. *Science (New York, N.Y.)*, *286*(5445), 1692–1695. doi: 10.1126/science.286.5445.1692

Fusaroli, R., Bahrami, B., Olsen, K., Roepstorff, A., Rees, G., Frith, C., & Tylén, K. (2012, August). Coming to Terms: Quantifying the Benefits of Linguistic Coordination. *Psychological Science*, *23*(8), 931–939. Retrieved 2022-07-14, from `https://doi.org/10.1177/0956797612436816` (Publisher: SAGE Publications Inc) doi: 10.1177/0956797612436816

Fusaroli, R., & Tylén, K. (2013, January). Fusaroli, R, & Tylén, K. (2013) Linguistic Coordination: Models, Dynamics and Effects, New Ideas in Psychology. *New Ideas in Psychology*, *32*. doi: 10.1016/j.newideapsych.2013.03.002

Gallese, V., Keysers, C., & Rizzolatti, G. (2004, September). A unifying view of the basis of social cognition. *Trends in Cognitive Sciences*, *8*(9), 396–403. doi: 10.1016/j.tics.2004.07.002

Gallotti, M., & Frith, C. D. (2013, April). Social cognition in the we-mode. *Trends in Cognitive Sciences*, *17*(4), 160–165. Retrieved 2022-03-14, from `https://www.sciencedirect.com/science/article/pii/S1364661313000417` doi: 10.1016/j.tics.2013.02.002

Gallup, A. C., Chong, A., & Couzin, I. D. (2012, August). The directional flow of visual information transfer between pedestrians. *Biology Letters*, *8*(4), 520–522. Retrieved 2022-09-15, from `https://royalsocietypublishing.org/doi/abs/10.1098/rsbl.2012.0160` (Publisher: Royal Society) doi: 10.1098/rsbl.2012.0160

Gambi, C., & Pickering, M. J. (2011). A Cognitive Architecture for the Coordination of Utterances. *Frontiers in Psychology*, *2*. Retrieved 2020-12-07, from `https://www.frontiersin.org/articles/10.3389/fpsyg.2011.00275/full` (Publisher: Frontiers) doi: 10.3389/fpsyg.2011.00275

Gambi, C., & Pickering, M. J. (2013, August). Talking to each other and talking together: Joint language tasks and degrees of interactivity. *The Behavioral and brain* *sciences*, *36*, 423–4. doi: 10.1017/S0140525X12001926

Gambi, C., & Pickering, M. J. (2017). Models Linking Production and Comprehension. In *The Handbook of Psycholinguistics* (pp. 157–181). John Wiley & Sons, Ltd. Retrieved 2022-09-15, from https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118829516.ch7 (Section: 7 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118829516.ch7) doi: 10.1002/9781118829516.ch7

Gambi, C., Van de Cavey, J., & Pickering, M. J. (2015). Interference in joint picture naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(1), 1–21. (Place: US Publisher: American Psychological Association) doi: 10.1037/a0037438

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

Garrod, S., & Pickering, M. J. (2004, January). Why is conversation so easy? *Trends in Cognitive Sciences*, *8*(1), 8–11. Retrieved 2021-04-30, from https://www.sciencedirect.com/science/article/pii/S136466130300295X doi: 10.1016/j.tics.2003.10.016

Garrod, S., & Pickering, M. J. (2009). Joint Action, Interactive Alignment, and Dialog. *Topics in Cognitive Science*, *1*(2), 292–304. Retrieved 2020-12-03, from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1756-8765.2009.01020.x (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1756-8765.2009.01020.x) doi: https://doi.org/10.1111/j.1756-8765.2009.01020.x

Garrod, S., & Pickering, M. J. (2015, June). The use of content and timing to predict turn transitions. *Frontiers in Psychology*, *6*, 751. Retrieved 2022-09-20, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4463931/ doi: 10.3389/fpsyg.2015.00751

Gibson, E., & Rader, N. (1979, January). Attention: The Perceiver as Performer. In (pp. 1–21). doi: 10.1007/978-1-4613-2985-5$_1$

Giles, H. (2008). Communication accommodation theory. In *Engaging theories in interpersonal communication: Multiple perspectives* (pp. 161–173). Thousand Oaks, CA, US: Sage Publications, Inc. doi: 10.4135/9781483329529.n12

Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. In *Contexts of accommodation: Developments in applied sociolinguistics* (pp. 1–68). Paris, France: Editions de la Maison des Sciences de l'Homme. doi: 10.1017/CBO9780511663673.001

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological*

Science, *11*, 274–279. (Place: United Kingdom Publisher: Blackwell Publishing) doi: 10.1111/1467-9280.00255

Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*, *49*(1), 43–61. (Place: Netherlands Publisher: Elsevier Science) doi: 10.1016/S0749-596X(03)00022-6

Hauk, O. (2016, August). Only time will tell - why temporal information is essential for our neuroscientific understanding of semantics. *Psychonomic Bulletin & Review*, *23*(4), 1072–1079. doi: 10.3758/s13423-015-0873-9

Hauser, T. U., Iannaccone, R., Stämpfli, P., Drechsler, R., Brandeis, D., Walitza, S., & Brem, S. (2014, January). The feedback-related negativity (FRN) revisited: new insights into the localization, meaning and network organization. *NeuroImage*, *84*, 159–168. doi: 10.1016/j.neuroimage.2013.08.028

Herrmann, E., Call, J., Hernàndez-Lloreda, M. V., Hare, B., & Tomasello, M. (2007, September). Humans Have Evolved Specialized Skills of Social Cognition: The Cultural Intelligence Hypothesis. *Science*, *317*(5843), 1360–1366. Retrieved 2022-03-17, from `https://www.science.org/doi/10.1126/science.1146282` (Publisher: American Association for the Advancement of Science) doi: 10.1126/science.1146282

Hickok, G. (2012, February). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, *13*(2), 135–145. Retrieved 2019-03-13, from `http://www.nature.com/articles/nrn3158` doi: 10.1038/nrn3158

Hirst, W., Yamashiro, J. K., & Coman, A. (2018, May). Collective Memory from a Psychological Perspective. *Trends in Cognitive Sciences*, *22*(5), 438–451. Retrieved 2022-03-14, from `https://www.sciencedirect.com/science/article/pii/S1364661318300469` doi: 10.1016/j.tics.2018.02.010

Hodapp, A., & Rabovsky, M. (2021). The N400 ERP component reflects an error-based implicit learning signal during language comprehension. *European Journal of Neuroscience*, *54*(9), 7125–7140. Retrieved 2022-07-18, from `https://onlinelibrary.wiley.com/doi/abs/10.1111/ejn.15462` (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ejn.15462) doi: 10.1111/ejn.15462

Hoedemaker, R. S., & Meyer, A. S. (2019, April). Planning and coordination of utterances in a joint naming task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(4), 732–752. Retrieved 2019-04-17, from `http://doi.apa.org/getdoi.cfm?doi=10.1037/xlm0000603` doi:

10.1037/xlm0000603

Hopkins, Z., Yuill, N., & Keller, B. (2015, January). Children with autism align syntax in natural conversation. *Applied Psycholinguistics*, *37*, 1–24. doi: 10.1017/S0142716414000599

Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006, July). Cumulative semantic inhibition in picture naming: experimental and computational studies. *Cognition*, *100*(3), 464–482. Retrieved 2020-12-03, from http://www.sciencedirect.com/science/article/pii/S0010027705001393 doi: 10.1016/j.cognition.2005.02.006

Howell, T. A., & Ratner, N. B. (2018, September). Use of a phoneme monitoring task to examine lexical access in adults who do and do not stutter. *Journal of fluency disorders*, *57*, 65–73. Retrieved 2022-08-28, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8965821/ doi: 10.1016/j.jfludis.2018.01.001

Huettig, F. (2015, November). Four central questions about prediction in language processing. *Brain Research*, *1626*, 118–135. Retrieved 2019-10-14, from http://www.sciencedirect.com/science/article/pii/S0006899315001146 doi: 10.1016/j.brainres.2015.02.014

Hulstijn, J. (2012, November). Incidental Learning in Second Language Acquisition.. doi: 10.1002/9781405198431.wbeal0530

Iio, T., Shiomi, M., Shinozawa, K., Shimohara, K., Miki, M., & Hagita, N. (2015, April). Lexical Entrainment in Human Robot Interaction. *International Journal of Social Robotics*, *7*. doi: 10.1007/s12369-014-0255-x

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

Jung, Y., Kwan, M., & Lee, K. (2004, January). Effects of physical embodiment on social presence of social robots. *Proceedings of Presence*.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, *15*(2), 159–201. (Place: United Kingdom Publisher: Taylor & Francis) doi: 10.1080/016909600386084

Kampis, D., & Southgate, V. (2020, November). Altercentric Cognition: How Others Influence Our Cognitive Processing. *Trends in Cognitive Sciences*, *24*(11), 945–959. Retrieved 2022-03-07, from https://www.sciencedirect.com/science/article/pii/S1364661320302175 doi: 10.1016/j.tics.2020.09.003

Katz, J. J., & Fodor, J. A. (1963). The Structure of a Semantic Theory. *Language*, *39*(2), 170–210. Retrieved 2022-09-14, from https://www.jstor.org/stable/411200 (Publisher: Linguistic Society of America) doi: 10.2307/411200

Kilner, J. M., Paulignan, Y., & Blakemore, S. J. (2003, March). An Interference Effect of Observed Biological Movement on Action. *Current Biology*, *13*(6), 522–525. Retrieved 2021-01-21, from http://www.sciencedirect.com/science/article/pii/S0960982203001659 doi: 10.1016/S0960-9822(03)00165-9

Kim, Y., Kwak, s., & Kim, M. (2013, May). Am I acceptable to you? Effect of a robot's verbal language forms on people's social distance from robots. *Computers in Human Behavior*, *29*, 1091–1101. doi: 10.1016/j.chb.2012.10.001

Kirtay, M., Wudarczyk, O., Pischedda, D., Kuhlen, A., Abdel Rahman, R., Haynes, J.-D., & Hafner, V. (2020, December). Modeling robot co-representation: state-of-the-art, open issues, and predictive learning as a possible framework.. doi: 10.1109/ICDL-EpiRob48136.2020.9278031

Kleinschmidt, D. F., & Jaeger, T. F. (2015, April). Robust speech perception: recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, *122*(2), 148–203. doi: 10.1037/a0038695

Knoblich, G., Butterfill, S., & Sebanz, N. (2011). Psychological Research on Joint Action. In *Psychology of Learning and Motivation* (Vol. 54, pp. 59–101). Elsevier. Retrieved 2019-03-13, from https://linkinghub.elsevier.com/retrieve/pii/B9780123855275000036 doi: 10.1016/B978-0-12-385527-5.00003-6

Kobayashi, H., Yasuda, T., Igarashi, H., & Suzuki, S. (2020, November). Language Use in Joint Action: The Means of Referring Expressions. *International Journal of Social Robotics*, *12*(5), 1021–1029. Retrieved 2022-09-09, from https://doi.org/10.1007/s12369-017-0462-3 doi: 10.1007/s12369-017-0462-3

Kuhlen, A. K., & Abdel Rahman, R. (2017, September). Having a task partner affects lexical retrieval: Spoken word production in shared task settings. *Cognition*, *166*, 94–106. Retrieved 2019-04-27, from http://www.sciencedirect.com/science/article/pii/S0010027717301415 doi: 10.1016/j.cognition.2017.05.024

Kuhlen, A. K., Allefeld, C., & Haynes, J.-D. (2012). Content-specific coordination of listeners' to speakers' EEG during communication. *Frontiers in Human Neuroscience*, *6*. Retrieved 2022-06-15, from https://www.frontiersin.org/article/10.3389/fnhum.2012.00266

Kuhlen, A. K., & Rahman, R. A. (2020, September). Joint language production: an electrophysiological investigation of simulated lexical access on behalf of task partner. *bioRxiv*, 2020.09.18.303099. Retrieved 2021-02-24, from https://www.biorxiv.org/content/10.1101/2020.09.18.303099v1 (Publisher: Cold Spring Harbor Laboratory Section: New Results) doi: 10.1101/2020.09.18.303099

Kuperberg, G. R., & Jaeger, T. F. (2016, January). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59. Retrieved 2022-08-19, from https://doi.org/10.1080/23273798.2015.1102299 (Publisher: Routledge _eprint: https://doi.org/10.1080/23273798.2015.1102299) doi: 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*, 621–647. Retrieved 2019-09-30, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4052444/ doi: 10.1146/annurev.psych.093008.131123

Levelt, W. J., Roelofs, A., & Meyer, A. (1999, February). A theory of lexical access in speech production. *The Behavioral and brain sciences*, *22*(1). Retrieved 2022-03-25, from https://pubmed.ncbi.nlm.nih.gov/11301520/ (Publisher: Behav Brain Sci) doi: 10.1017/s0140525x99001776

Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, *6*. Retrieved 2022-07-31, from https://www.frontiersin.org/articles/10.3389/fpsyg.2015.00731

Liepelt, R., & Brass, M. (2010). Top-down modulation of motor priming by belief about animacy. *Experimental Psychology*, *57*(3), 221–227. doi: 10.1027/1618-3169/a000028

Loth, S., Jettka, K., Giuliani, M., & de Ruiter, J. P. (2015). Ghost-in-the-Machine reveals human social signals for human–robot interaction. *Frontiers in Psychology*, *6*. Retrieved 2021-04-13, from https://www.frontiersin.org/articles/10.3389/fpsyg.2015.01641/full (Publisher: Frontiers) doi: 10.3389/fpsyg.2015.01641

Lu, Z.-L. (2008, January). Mechanisms of attention. *Kiso shinrigaku kenkyu*, *27*(1), 38–45. Retrieved 2022-08-10, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2879667/

Luo, M., Robbins, M. L., Martin, M., & Demiray, B. (2019). Real-Life Language Use Across Different Interlocutors: A Naturalistic Observation Study of Adults Varying in Age. *Frontiers in Psychology*, *10*. Retrieved 2022-08-09, from https://www.frontiersin.org/articles/10.3389/fpsyg.2019.01412

Maess, B., Mamashli, F., Obleser, J., Helle, L., & Friederici, A. D. (2016). Prediction Signatures in the Brain: Semantic Pre-Activation during Language Comprehension. *Frontiers in Human Neuroscience*, *10*. Retrieved 2022-09-25, from https://www.frontiersin.org/articles/10.3389/fnhum.2016.00591

Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake—But only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(4), 843–847. Retrieved 2019-10-14, from http://doi.apa.org/getdoi.cfm?doi=10.1037/a0029284 doi: 10.1037/a0029284

Marge, M., Espy-Wilson, C., Ward, N. G., Alwan, A., Artzi, Y., Bansal, M., ... Yu, Z. (2022, January). Spoken language interaction with robots: Recommendations for future research. *Computer Speech & Language*, *71*, 101255. Retrieved 2022-09-15, from https://www.sciencedirect.com/science/article/pii/S0885230821000620 doi: 10.1016/j.csl.2021.101255

Martin, C. D., Branzi, F. M., & Bar, M. (2018, January). Prediction is Production: The missing link between language production and comprehension. *Scientific Reports*, *8*(1), 1–9. Retrieved 2019-10-14, from https://www.nature.com/articles/s41598-018-19499-4 doi: 10.1038/s41598-018-19499-4

Mathôt, S., Schreij, D., & Theeuwes, J. (2012, June). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. Retrieved 2022-07-19, from https://doi.org/10.3758/s13428-011-0168-7 doi: 10.3758/s13428-011-0168-7

McEllin, L., Knoblich, G., & Sebanz, N. (2018). Imitation from a joint action perspective. *Mind & Language*, *33*(4), 342–354. Retrieved 2022-06-09, from https://onlinelibrary.wiley.com/doi/abs/10.1111/mila.12188 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/mila.12188) doi: 10.1111/mila.12188

Menenti, L., Pickering, M. J., & Garrod, S. C. (2012, June). Toward a neural basis of interactive alignment in conversation. *Frontiers in Human Neuroscience*, *6*, 185. Retrieved 2022-10-08, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3384290/ doi: 10.3389/fnhum.2012.00185

Mercer, N. (2016, July). Education and the social brain: linking language, thinking, teaching and learning. *Éducation et didactique*, *10*(2), 9–23. Retrieved 2022-

09-16, from https://journals.openedition.org/educationdidactique/2523 (ISBN: 9782753551787 Number: 2 Publisher: Presses universitaires de Rennes) doi: 10.4000/educationdidactique.2523

Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language*, *49*(2), 201–213. (Place: Netherlands Publisher: Elsevier Science) doi: 10.1016/S0749-596X(03)00028-7

Michaelis, K., Miyakoshi, M., Norato, G., Medvedev, A. V., & Turkeltaub, P. E. (2021, January). Motor engagement relates to accurate perception of phonemes and audiovisual words, but not auditory words. *Communications Biology*, *4*(1), 1–12. Retrieved 2022-07-21, from https://www.nature.com/articles/s42003-020-01634-5 (Number: 1 Publisher: Nature Publishing Group) doi: 10.1038/s42003-020-01634-5

Moorselaar, D., & Slagter, H. (2020, January). Inhibition in selective attention. *Annals of the New York Academy of Sciences*, *1464*. doi: 10.1111/nyas.14304

Morrow, L. I., & Duffy, M. F. (2005, November). The representation of ontological category concepts as affected by healthy aging: Normative data and theoretical implications. *Behavior Research Methods*, *37*(4), 608–625. Retrieved 2020-12-07, from https://doi.org/10.3758/BF03192731 doi: 10.3758/BF03192731

Moubayed, S. A., Skantze, G., & Beskow, J. (2013, March). The furhat back-projected humanoid head–lip reading, gaze and multi-party interaction. *International Journal of Humanoid Robotics*, *10*(01), 1350005. Retrieved 2020-12-11, from https://www.worldscientific.com/doi/abs/10.1142/S0219843613500059 (Publisher: World Scientific Publishing Co.) doi: 10.1142/S0219843613500059

Mundy, P., & Newell, L. (2007, October). Attention, Joint Attention, and Social Cognition. *Current Directions in Psychological Science*, *16*(5), 269–274. Retrieved 2022-03-14, from https://doi.org/10.1111/j.1467-8721.2007.00518.x (Publisher: SAGE Publications Inc) doi: 10.1111/j.1467-8721.2007.00518.x

Müller, B. C. N., Brass, M., Kühn, S., Tsai, C.-C., Nieuwboer, W., Dijksterhuis, A., & van Baaren, R. B. (2011, April). When Pinocchio acts like a human, a wooden hand becomes embodied. Action co-representation for non-biological agents. *Neuropsychologia*, *49*(5), 1373–1377. doi: 10.1016/j.neuropsychologia.2011.01.022

Nadig, A., Seth, S., & Sasson, M. (2015, December). Global Similarities and Multifaceted Differences in the Production of Partner-Specific Referential Pacts by Adults with Autism Spectrum Disorders. *Frontiers in Psychology*, *6*, 1888. Retrieved 2022-10-15, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4681838/ doi:

10.3389/fpsyg.2015.01888

Nairne, J., VanArsdall, J., Pandeirada, J., Cogdill, M., & LeBreton, J. (2013, August). Adaptive Memory: The Mnemonic Value of Animacy. *Psychological science*, *24*. doi: 10.1177/0956797613480803

Navarrete, E., Basagni, B., Alario, F.-X., & Costa, A. (2006, October). Does word frequency affect lexical selection in speech production? *Quarterly Journal of Experimental Psychology*, *59*(10), 1681–1690. Retrieved 2019-03-13, from `https://doi.org/10.1080/17470210600750558` doi: 10.1080/17470210600750558

Ness, T., & Meltzer-Asscher, A. (2021). Rational Adaptation in Lexical Prediction: The Influence of Prediction Strength. *Frontiers in Psychology*, *12*. Retrieved 2022-08-20, from `https://www.frontiersin.org/articles/10.3389/fpsyg.2021.622873`

New, B., Pallier, C., Ferrand, L., & Matos, R. (2001, January). Une base de données lexicales du français contemporain sur internet : LEXIQUE™//A lexical database for contemporary french : LEXIQUE™. *Annee Psychologique - ANNEE PSYCHOL*, *101*, 447–462. doi: 10.3406/psy.2001.1341

New, J., Cosmides, L., & Tooby, J. (2007, October). Category-specific attention for animals reflects ancestral priorities, not expertise. *Proceedings of the National Academy of Sciences*, *104*(42), 16598–16603. Retrieved 2022-03-25, from `https://www.pnas.org/doi/10.1073/pnas.0703913104` (Publisher: Proceedings of the National Academy of Sciences) doi: 10.1073/pnas.0703913104

Nieuwland, M. S. (2019, January). Do 'early' brain responses reveal word form prediction during language comprehension? A critical review. *Neuroscience & Biobehavioral Reviews*, *96*, 367–400. Retrieved 2019-03-11, from `https://linkinghub.elsevier.com/retrieve/pii/S0149763418303609` doi: 10.1016/j.neubiorev.2018.11.019

Nieuwland, M. S., & Van Berkum, J. J. A. (2006, July). When Peanuts Fall in Love: N400 Evidence for the Power of Discourse. *Journal of Cognitive Neuroscience*, *18*(7), 1098–1111. Retrieved 2022-03-25, from `https://direct.mit.edu/jocn/article/18/7/1098/4187/When-Peanuts-Fall-in-Love-N400-E` doi: 10.1162/jocn.2006.18.7.1098

Norman, J. (2002). Two visual systems and two theories of perception: An attempt to reconcile the constructivist and ecological approaches. *Behavioral and Brain Sciences*, *25*, 73–144. (Place: United Kingdom Publisher: Cambridge University Press) doi: 10.1017/S0140525X0200002X

Novembre, G., Ticini, F., Luca, Schütz-Bosbach, S., & Keller, P. (2012, January). Distinguishing Self and Other in Joint Action. Evidence from a Musical Paradigm. *Cerebral cortex (New York, N.Y. : 1991)*, *22*. doi: 10.1093/cercor/bhr364

Obhi, S., & Sebanz, N. (2011, June). Moving together: Toward understanding the mechanisms of joint action. *Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale*, *211*, 329–36. doi: 10.1007/s00221-011-2721-0

Oh, C. S., Bailenson, J. N., & Welch, G. F. (2018). A Systematic Review of Social Presence: Definition, Antecedents, and Implications. *Frontiers in Robotics and AI*, *5*. Retrieved 2022-09-16, from https://www.frontiersin.org/articles/10.3389/frobt.2018.00114

Oldfield, R. C., & Wingfield, A. (1965, December). Response Latencies in Naming Objects. *Quarterly Journal of Experimental Psychology*, *17*(4), 273–281. Retrieved 2019-03-13, from http://journals.sagepub.com/doi/10.1080/17470216508416445 doi: 10.1080/17470216508416445

Pacherie, E. (2012, January). The Phenomenology of Joint Action: Self-Agency vs. Joint-Agency. *Joint Attention: New Developments*.

Pardo, J. S. (2006, March). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, *119*(4), 2382–2393. Retrieved 2021-01-21, from https://asa.scitation.org/doi/abs/10.1121/1.2178720 (Publisher: Acoustical Society of America) doi: 10.1121/1.2178720

Paulus, M., Hunnius, S., Vissers, M., & Bekkering, H. (2011). Imitation in Infancy: Rational or Motor Resonance? *Child Development*, *82*(4), 1047–1057. Retrieved 2022-07-21, from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-8624.2011.01610.x (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-8624.2011.01610.x) doi: 10.1111/j.1467-8624.2011.01610.x

Pearson, J., Hu, J., Branigan, H., Pickering, M. J., & Nass, C. (2006, April). Adaptive language behavior in HCI: how expectations and beliefs about a system affect users' word choice. In (Vol. 2, pp. 1177–1180). doi: 10.1145/1124772.1124948

Pesquita, A., Whitwell, R. L., & Enns, J. T. (2018, October). Predictive joint-action model: A hierarchical predictive approach to human cooperation. *Psychonomic Bulletin & Review*, *25*(5), 1751–1769. Retrieved 2019-02-20, from https://doi.org/10.3758/s13423-017-1393-6 doi: 10.3758/s13423-017-1393-6

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

Pickering, M. J., & Ferreira, V. S. (2008, May). Structural Priming: A Critical Review. *Psychological bulletin*, *134*(3), 427–459. Retrieved 2021-01-21, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2657366/ doi: 10.1037/0033-2909.134.3.427

Pickering, M. J., & Gambi, C. (2018, April). Predicting While Comprehending Language: A Theory and Review. *Psychological Bulletin*, *144*. doi: 10.1037/bul0000158

Pickering, M. J., & Garrod, S. (2004a, April). The interactive-alignment model: Developments and refinements. *Behavioral and Brain Sciences*, *27*(2), 212–225. Retrieved 2022-09-15, from https://www.cambridge.org/core/journals/behavioral-and-brain-sciences/article/abs/i (Publisher: Cambridge University Press) doi: 10.1017/S0140525X04450055

Pickering, M. J., & Garrod, S. (2004b, May). Toward a Mechanistic Psychology of Dialogue. *The Behavioral and brain sciences*, *27*, 169–90; discussion 190. doi: 10.1017/S0140525X04000056

Pickering, M. J., & Garrod, S. (2006, October). Alignment as the Basis for Successful Communication. *Research on Language and Computation*, *4*(2), 203–228. Retrieved 2020-12-03, from https://doi.org/10.1007/s11168-006-9004-0 doi: 10.1007/s11168-006-9004-0

Pickering, M. J., & Garrod, S. (2007, March). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, *11*(3), 105–110. doi: 10.1016/j.tics.2006.12.002

Pickering, M. J., & Garrod, S. (2013a, August). Forward models and their implications for production, comprehension, and dialogue. *Behavioral and Brain Sciences*, *36*(4), 377–392. Retrieved 2019-02-20, from https://www.cambridge.org/core/journals/behavioral-and-brain-sciences/article/forwa doi: 10.1017/S0140525X12003238

Pickering, M. J., & Garrod, S. (2013b, August). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(04), 329–347. Retrieved 2020-12-07, from https://eprints.gla.ac.uk/82858/ (Number: 04) doi: 10.1017/S0140525X12001495

Pickering, M. J., & Garrod, S. (2014a, October). Neural integration of language production and comprehension. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(43), 15291–15292. Retrieved 2019-10-14, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4217443/ doi: 10.1073/pnas.1417917111

Pickering, M. J., & Garrod, S. (2014b). Self-, other-, and joint monitoring using forward models. *Frontiers in Human Neuroscience*, *8*. Retrieved 2022-08-13, from https://www.frontiersin.org/articles/10.3389/fnhum.2014.00132

Pickering, M. J., McLean, J. F., & Gambi, C. (2022). Interference in the shared-Stroop task: a comparison of self- and other-monitoring. *Royal Society Open Science*, *9*(4), 220107. Retrieved 2022-09-16, from https://royalsocietypublishing.org/doi/full/10.1098/rsos.220107 (Publisher: Royal Society) doi: 10.1098/rsos.220107

Posner, M. (1980, March). Orienting of Attention. *The Quarterly journal of experimental psychology*, *32*, 3–25. doi: 10.1080/00335558008248231

Prinz, W. (1990). A Common Coding Approach to Perception and Action. In O. Neumann & W. Prinz (Eds.), *Relationships Between Perception and Action: Current Approaches* (pp. 167–201). Berlin, Heidelberg: Springer. Retrieved 2020-12-07, from https://doi.org/10.1007/978-3-642-75348-0$_7$ doi: 10.1007/978-3-642-75348-0$_7$

Prinz, W. (1997). Perception and action planning. *European Journal of Cognitive Psychology*, *9*(2), 129–154. (Place: United Kingdom Publisher: Taylor & Francis) doi: 10.1080/713752551

Pulvermüller, F., & Shtyrov, Y. (2006, May). Language outside the focus of attention: the mismatch negativity as a tool for studying higher cognitive processes. *Progress in Neurobiology*, *79*(1), 49–71. doi: 10.1016/j.pneurobio.2006.04.004

Ramscar, M., Dye, M., & McCauley, S. (2013, December). Error and expectation in language learning: The curious absence of 'mouses' in adult speech. *Language*, *89*, 760–793. doi: 10.1353/lan.2013.0068

Rasenberg, M., Özyürek, A., & Dingemanse, M. (2020, November). Alignment in Multimodal Interaction: An Integrative Framework. *Cognitive Science*, *44*, 12911. doi: 10.1111/cogs.12911

Rawlinson, H. C., & Kelley, C. M. (2021, August). In search of the proximal cause of the animacy effect on memory: Attentional resource allocation and semantic representations. *Memory & Cognition*, *49*(6), 1137–1152. Retrieved 2022-03-25, from https://search.ebscohost.com/login.aspx?direct=truedb=bthAN=151585031lang=frsite=eh (Publisher: Springer Nature) doi: 10.3758/s13421-021-01154-5

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

Reitter, D., & Moore, J. D. (2014). Alignment and task success in spoken dialogue. *Journal of Memory and Language*, *76*, 29–46. (Place: Netherlands Publisher: Elsevier Science) doi: 10.1016/j.jml.2014.05.008

Risko, E. F., & Kingstone, A. (2011, February). Eyes wide shut: implied social presence, eye tracking and attention. *Attention, Perception & Psychophysics*, *73*(2), 291–296. doi: 10.3758/s13414-010-0042-1

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*, 169–192. doi: 10.1146/annurev.neuro.27.070203.144230

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996, March). Premotor cortex and the recognition of motor actions. *Brain Research. Cognitive Brain Research*, *3*(2), 131–141. doi: 10.1016/0926-6410(95)00038-0

Roelofs, A. (1992, January). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*(1), 107–142. Retrieved 2022-09-16, from https://www.sciencedirect.com/science/article/pii/001002779290041F doi: 10.1016/0010-0277(92)90041-F

Rossiter, C., & Best, W. (2013, July). "Penguins don't fly": An investigation into the effect of typicality on picture naming in people with aphasia. *Aphasiology*, *27*(7), 784–798. Retrieved 2019-01-17, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3935221/ doi: 10.1080/02687038.2012.751579

Raczaszek-Leonardi, J., & Kelso, S. (2008, September). Reconciling symbolic and dynamic aspects of language: Toward a dynamic psycholinguistics. *New ideas in psychology*, *26*, 193–207. doi: 10.1016/j.newideapsych.2007.07.003

Sahaï, A., Desantis, A., Grynszpan, O., Pacherie, E., & Berberian, B. (2019, January). Action co-representation and the sense of agency during a joint Simon task: Comparing human and machine co-agents. *Consciousness and Cognition*, *67*, 44–55. Retrieved 2022-09-25, from https://www.sciencedirect.com/science/article/pii/S1053810018301156 doi: 10.1016/j.concog.2018.11.008

Sahin, N., Pinker, S., Cash, S., Schomer, D., & Halgren, E. (2009, October). Sequential Processing of Lexical, Grammatical, and Phonological Information Within Broca's Area. *Science (New York, N.Y.)*, *326*, 445–9. doi: 10.1126/science.1174481

Salo, V. C., Ferrari, P. F., & Fox, N. A. (2019, April). The role of the motor system in action understanding and communication: Evidence from human infants and non-human primates. *Developmental psychobiology*, *61*(3), 390–401. Retrieved 2022-07-21, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6447431/ doi: 10.1002/dev.21779

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

Schmitz, L., Vesper, C., Sebanz, N., & Knoblich, G. (2018, December). Co-actors represent the order of each other's actions. *Cognition*, *181*, 65–79. Retrieved 2022-10-14, from `https://www.sciencedirect.com/science/article/pii/S0010027718302154` doi: 10.1016/j.cognition.2018.08.008

Schober, M. F., & Brennan, S. E. (2003). Processes of interactive spoken discourse: The role of the partner. In *Handbook of discourse processes* (pp. 123–164). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Sebanz, N., Bekkering, H., & Knoblich, G. (2006, February). Joint action: bodies and minds moving together. *Trends in Cognitive Sciences*, *10*(2), 70–76. Retrieved 2019-02-20, from `https://linkinghub.elsevier.com/retrieve/pii/S1364661305003566` doi: 10.1016/j.tics.2005.12.009

Sebanz, N., & Knoblich, G. (2009, April). Prediction in joint action: what, when, and where. *Topics in Cognitive Science*, *1*(2), 353–367. doi: 10.1111/j.1756-8765.2009.01024.x

Sebanz, N., Knoblich, G., & Prinz, W. (2003, July). Representing others' actions: just like one's own? *Cognition*, *88*(3), B11–B21. Retrieved 2021-08-19, from `https://linkinghub.elsevier.com/retrieve/pii/S001002770300043X` doi: 10.1016/S0010-0277(03)00043-X

Sebanz, N., Knoblich, G., & Prinz, W. (2005, December). How two share a task: corepresenting stimulus-response mappings. *Journal of Experimental Psychology. Human Perception and Performance*, *31*(6), 1234–1246. doi: 10.1037/0096-1523.31.6.1234

Serences, J. T., & Yantis, S. (2006, January). Selective visual attention and perceptual coherence. *Trends in Cognitive Sciences*, *10*(1), 38–45. doi: 10.1016/j.tics.2005.11.008

Sharpe, V., Schoot, L., Lewandowski, K. E., Öngür, D., Türközer, H. B., Hasoğlu, T., & Kuperberg, G. R. (2022, May). We both say tomato: Intact lexical alignment in schizophrenia and bipolar disorder. *Schizophrenia Research*, *243*, 138–146. Retrieved 2022-07-14, from `https://www.sciencedirect.com/science/article/pii/S0920996422001104` doi: 10.1016/j.schres.2022.02.032

Shintel, H., & Keysar, B. (2009). Less Is More: A Minimalist Account of Joint Action in Communication. *Topics in Cognitive Science*, *1*(2), 260–273. Retrieved 2021-02-11, from `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1756-8765.2009.01018.x` (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1756-8765.2009.01018.x) doi: https://doi.org/10.1111/j.1756-8765.2009.01018.x

Shteynberg, G. (2010, October). A silent emergence of culture: The social tuning effect. *Journal of Personality and Social Psychology*, *99*(4), 683–689. Retrieved 2022-03-14, from https://search.ebscohost.com/login.aspx?direct=truedb=pdhAN=2010-16629-001lang=frsi (Publisher: American Psychological Association) doi: 10.1037/a0019573

Shteynberg, G. (2015, January). Shared Attention. *Perspectives on Psychological Science*, *10*. doi: 10.1177/1745691615589104

Shteynberg, G. (2018, October). A collective perspective: shared attention and the mind. *Current Opinion in Psychology*, *23*, 93–97. Retrieved 2022-03-14, from https://www.sciencedirect.com/science/article/pii/S2352250X17302506 doi: 10.1016/j.copsyc.2017.12.007

Shteynberg, G., & Galinsky, A. D. (2011). Implicit coordination: Sharing goals with similar others intensifies goal pursuit. *Journal of Experimental Social Psychology*, *47*, 1291–1294. (Place: Netherlands Publisher: Elsevier Science) doi: 10.1016/j.jesp.2011.04.012

Shteynberg, G., Hirsh, J. B., Apfelbaum, E. P., Larsen, J. T., Galinsky, A. D., & Roese, N. J. (2014, December). Feeling more together: Group attention intensifies emotion. *Emotion*, *14*(6), 1102–1114. Retrieved 2022-03-17, from https://search.ebscohost.com/login.aspx?direct=truedb=pdhAN=2014-34829-001lang=frsi (Publisher: American Psychological Association) doi: 10.1037/a0037697

Shteynberg, G., Hirsh, J. B., Bentley, R. A., & Garthoff, J. (2020). Shared worlds and shared minds: A theory of collective learning and a psychology of common knowledge. *Psychological Review*, *127*(5), 918–931. Retrieved 2022-03-07, from https://search.ebscohost.com/login.aspx?direct=truedb=pdhAN=2020-27106-001lang=frsi (Publisher: American Psychological Association) doi: 10.1037/rev0000200

Signoret, C., Andersen, L. M., Dahlström, , Blomberg, R., Lundqvist, D., Rudner, M., & Rönnberg, J. (2020). The Influence of Form- and Meaning-Based Predictions on Cortical Speech Processing Under Challenging Listening Conditions: A MEG Study. *Frontiers in Neuroscience*, *14*. Retrieved 2022-08-01, from https://www.frontiersin.org/articles/10.3389/fnins.2020.573254

Silbert, L. J., Honey, C. J., Simony, E., Poeppel, D., & Hasson, U. (2014, October). Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(43), E4687–E4696. Retrieved 2019-10-14, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4217461/ doi: 10.1073/pnas.1323812111

Skantze, G. (2016). Real-Time Coordination in Human-Robot Interaction Using Face and Voice. *AI Magazine*, *37*(4), 19–31. Retrieved 2020-12-11, from https://ojs.aaai.org/index.php/aimagazine/article/view/2686 (Number: 4) doi: 10.1609/aimag.v37i4.2686

Skeide, M. A., & Friederici, A. D. (2017). Neurolinguistic Studies of Sentence Comprehension. In *The Handbook of Psycholinguistics* (pp. 438–456). John Wiley & Sons, Ltd. Retrieved 2022-07-18, from https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118829516.ch19 (Section: 19 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118829516.ch19) doi: 10.1002/9781118829516.ch19

Skipper, J. I., Devlin, J. T., & Lametti, D. R. (2017, January). The hearing ear is always found close to the speaking tongue: Review of the role of the motor system in speech perception. *Brain and Language*, *164*, 77–105. Retrieved 2022-07-21, from https://www.sciencedirect.com/science/article/pii/S0093934X16301420 doi: 10.1016/j.bandl.2016.10.004

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 174–215. (Place: US Publisher: American Psychological Association) doi: 10.1037/0278-7393.6.2.174

Southgate, V. (2018, March). *Are Infants Altercentric? The Other and the Self in Early Social Cognition.* PsyArXiv. Retrieved 2022-08-03, from https://psyarxiv.com/y2ghq/ doi: 10.31234/osf.io/y2ghq

Staub, A. (2015). The Effect of Lexical Predictability on Eye Movements in Reading: Critical Review and Theoretical Interpretation. *Language and Linguistics Compass*, *9*(8), 311–327. Retrieved 2022-08-20, from https://onlinelibrary.wiley.com/doi/abs/10.1111/lnc3.12151 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/lnc3.12151) doi: 10.1111/lnc3.12151

Staub, A., & Benatar, A. (2013, December). Individual differences in fixation duration distributions in reading. *Psychonomic Bulletin & Review*, *20*(6), 1304–1311. Retrieved 2022-08-20, from https://doi.org/10.3758/s13423-013-0444-x doi: 10.3758/s13423-013-0444-x

Strijkers, K., & Costa, A. (2011). Riding the lexical speedway: a critical review on the time course of lexical selection in speech production. *Frontiers in Psychology*, *2*, 356. doi: 10.3389/fpsyg.2011.00356

Strijkers, K., & Costa, A. (2016, January). The cortical dynamics of speaking: present shortcomings and future avenues. *Language, Cognition and Neuroscience*, *31*, 1–20.

doi: 10.1080/23273798.2015.1120878

Strijkers, K., Costa, A., & Pulvermüller, F. (2017, December). The cortical dynamics of speaking: Lexical and phonological knowledge simultaneously recruit the frontal and temporal cortex within 200 ms. *NeuroImage*, *163*, 206–219. Retrieved 2019-08-02, from `http://www.sciencedirect.com/science/article/pii/S1053811917307899` doi: 10.1016/j.neuroimage.2017.09.041

Strijkers, K., Costa, A., & Thierry, G. (2010, April). Tracking lexical access in speech production: electrophysiological correlates of word frequency and cognate effects. *Cerebral Cortex (New York, N.Y.: 1991)*, *20*(4), 912–928. doi: 10.1093/cercor/bhp153

Suffill, E., Kutasi, T., Pickering, M. J., & Branigan, H. P. (2021, August). Lexical alignment is affected by addressee but not speaker nativeness. *Bilingualism: Language and Cognition*, *24*(4), 746–757. Retrieved 2022-05-16, from `https://www.cambridge.org/core/journals/bilingualism-language-and-cognition/article` (Publisher: Cambridge University Press) doi: 10.1017/S1366728921000092

Suzuki, N., & Katagiri, Y. (2007). Prosodic alignment in human-computer interaction. *Connection Science*, *19*, 131–141. (Place: United Kingdom Publisher: Taylor & Francis) doi: 10.1080/09540090701369125

Tomasello, M. (2009). *Why We Cooperate* (D. Chasman, Ed.). Cambridge, MA, USA: MIT Press.

Tomasello, M. (2014, April). The ultra-social animal. *European Journal of Social Psychology*, *44*(3), 187–194. Retrieved 2022-03-17, from `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4302252/` doi: 10.1002/ejsp.2015

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005, October). Understanding and sharing intentions: the origins of cultural cognition. *The Behavioral and Brain Sciences*, *28*(5), 675–691; discussion 691–735. doi: 10.1017/S0140525X05000129

Tresselt, M., & Mayzner, M. (1960, October). A Study of Incidental Learning. *Journal of Psychology - J PSYCHOL*, *50*, 339–347. doi: 10.1080/00223980.1960.9916451

Tsai, C.-C., Kuo, W.-J., Jing, J.-T., Hung, D. L., & Tzeng, O. J.-L. (2006, November). A common coding framework in self-other interaction: evidence from joint action task. *Experimental Brain Research*, *175*(2), 353–362. doi: 10.1007/s00221-006-0557-9

Turk, D. J., Cunningham, S. J., & Macrae, C. N. (2008, September). Self-memory biases in explicit and incidental encoding of trait adjectives. *Consciousness and Cogni-*

tion, *17*(3), 1040–1045. doi: 10.1016/j.concog.2008.02.004

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. Retrieved 2022-10-12, from `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-0017.2009.01379.x` (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-0017.2009.01379.x) doi: 10.1111/j.1468-0017.2009.01379.x

Uithol, S., van Rooij, I., Bekkering, H., & Haselager, P. (2011). Understanding motor resonance. *Social Neuroscience*, *6*(4), 388–397. doi: 10.1080/17470919.2011.559129

van der Wel, R. P. (2015, July). Me and we: Metacognition and performance evaluation of joint actions. *Cognition*, *140*, 49–59. Retrieved 2019-03-13, from `https://linkinghub.elsevier.com/retrieve/pii/S001002771500061X` doi: 10.1016/j.cognition.2015.03.011

van der Wel, R. P., Becchio, C., Curioni, A., & Wolf, T. (2021, April). Understanding joint action: Current theoretical and empirical approaches. *Acta Psychologica*, *215*, 103285. Retrieved 2022-05-20, from `https://www.sciencedirect.com/science/article/pii/S0001691821000354` doi: 10.1016/j.actpsy.2021.103285

van Moorselaar, D., & Slagter, H. A. (2020). Inhibition in selective attention. *Annals of the New York Academy of Sciences*, *1464*(1), 204–221. Retrieved 2022-10-16, from `https://onlinelibrary.wiley.com/doi/abs/10.1111/nyas.14304` (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/nyas.14304) doi: 10.1111/nyas.14304

Van Petten, C., & Luka, B. J. (2012, February). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*(2), 176–190. Retrieved 2022-07-31, from `https://www.sciencedirect.com/science/article/pii/S0167876011002819` doi: 10.1016/j.ijpsycho.2011.09.015

van Turennout, M., Hagoort, P., & Brown, C. M. (1997). Electrophysiological evidence on the time course of semantic and phonological processes in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(4), 787–806. (Place: US Publisher: American Psychological Association) doi: 10.1037/0278-7393.23.4.787

Vesper, C., Abramova, E., Bütepage, J., Ciardo, F., Crossey, B., Effenberg, A., … Wahn, B. (2017). Joint Action: Mental Representations, Shared Information and General Mechanisms for Coordinating with Others. *Frontiers in Psychology*, *7*. Retrieved 2022-07-21, from https://www.frontiersin.org/articles/10.3389/fpsyg.2016.02039

Vlainic, E., Liepelt, R., Colzato, L., Prinz, W., & Hommel, B. (2010, December). The Virtual Co-Actor: The Social Simon Effect does not Rely on Online Feedback from the Other. *Frontiers in psychology*, *1*, 208. doi: 10.3389/fpsyg.2010.00208

Vollmer, A.-L., Rohlfing, K. J., Wrede, B., & Cangelosi, A. (2015, April). Alignment to the Actions of a Robot. *International Journal of Social Robotics*, *7*(2), 241–252. Retrieved 2022-10-15, from https://doi.org/10.1007/s12369-014-0252-0 doi: 10.1007/s12369-014-0252-0

Vujović, M., Ramscar, M., & Wonnacott, E. (2021, August). Language learning as uncertainty reduction: The role of prediction error in linguistic generalization and item-learning. *Journal of Memory and Language*, *119*, 104231. Retrieved 2022-08-04, from https://www.sciencedirect.com/science/article/pii/S0749596X21000140 doi: 10.1016/j.jml.2021.104231

Wang, L., Gu, Y., Zhao, G., & Chen, A. (2020, March). Error-related negativity and error awareness in a Go/No-go task. *Scientific Reports*, *10*(1), 4026. Retrieved 2022-07-18, from https://www.nature.com/articles/s41598-020-60693-0 (Number: 1 Publisher: Nature Publishing Group) doi: 10.1038/s41598-020-60693-0

Wang, S., Cheong, Y. F., Dilks, D. D., & Rochat, P. (2020, October). The Uncanny Valley Phenomenon and the Temporal Dynamics of Face Animacy Perception. *Perception*, *49*(10), 1069–1089. Retrieved 2022-08-12, from https://doi.org/10.1177/0301006620952611 (Publisher: SAGE Publications Ltd STM) doi: 10.1177/0301006620952611

Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, *41*(8), 989–994. doi: 10.1016/s0028-3932(02)00316-0

Wenke, D., Atmaca, S., Holländer, A., Liepelt, R., Baess, P., & Prinz, W. (2011, June). What is Shared in Joint Action? Issues of Co-representation, Response Conflict, and Agent Identification. *Review of Philosophy and Psychology*, *2*(2), 147–172. Retrieved 2021-08-17, from http://link.springer.com/10.1007/s13164-011-0057-0 doi: 10.1007/s13164-011-0057-0

Wimmer, R. D., Schmitt, L. I., Davidson, T. J., Nakajima, M., Deisseroth, K., & Halassa, M. M. (2015, October). Thalamic control of sensory selection in divided attention. *Nature*, *526*(7575), 705–709. Retrieved 2022-09-15, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4626291/ doi: 10.1038/nature15398

Wolpert, D. M. (1997, September). Computational approaches to motor control. *Trends in Cognitive Sciences*, *1*(6), 209–216. Retrieved 2022-07-22, from https://www.sciencedirect.com/science/article/pii/S136466139701070X doi: 10.1016/S1364-6613(97)01070-X

Woollams, A. M. (2012, April). Apples are not the only fruit: the effects of concept typicality on semantic representation in the anterior temporal lobe. *Frontiers in Human Neuroscience*, *6*. Retrieved 2019-01-17, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3328795/ doi: 10.3389/fnhum.2012.00085

Wudarczyk, O., Kirtay, M., Kuhlen, A. K., Abdel Rahman, R., Haynes, J.-D., Hafner, V. V., & Pischedda, D. (2021). Bringing Together Robotics, Neuroscience, and Psychology: Lessons Learned From an Interdisciplinary Project. *Frontiers in Human Neuroscience*, *15*. Retrieved 2022-06-15, from https://www.frontiersin.org/article/10.3389/fnhum.2021.630789

Wudarczyk, O., Kirtay, M., Pischedda, D., Hafner, V., Haynes, J.-D., Kuhlen, A. K., & Abdel Rahman, R. (2021, August). Robots facilitate human language production. *Scientific Reports*, *11*. doi: 10.1038/s41598-021-95645-9

Wykowska, A., Chaminade, T., & Cheng, G. (2016, May). Embodied artificial agents for understanding human social cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1693), 20150375. Retrieved 2022-09-16, from https://royalsocietypublishing.org/doi/10.1098/rstb.2015.0375 (Publisher: Royal Society) doi: 10.1098/rstb.2015.0375

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*, 162–185. (Place: US Publisher: American Psychological Association) doi: 10.1037/0033-2909.123.2.162