UNIVERSITÉ DE LILLE

Laboratoire de Sciences Cognitives et Affectives (SCALab)

UMR CNRS 9193

École doctorale des Sciences de l'Homme et de la Société

The impact of dialogue on the organisation in memory of lexico-semantic representations

L'impact du dialogue sur l'organisation en mémoire des représentations lexico-sémantiques

---

THÈSE DE DOCTORAT

En vue de l'obtention du grade de Docteure en Psychologie

Présentée et soutenue publiquement le vendredi 29 Novembre 2024

Par

Alicia FASQUEL

Sous la direction du Pr. Angèle BRUNELLIÈRE

Et du Dr. Dominique KNUTSEN

Composition du jury :

Pr. Adrian BANGERTER	Université de Neuchâtel	Rapporteur, Président du jury
Pr. Gareth GASKELL	University of York	Rapporteur
Dr. Ana Luisa RAPOSO	Universidade de Lisboa	Examinatrice
Pr. Angèle BRUNELLIÈRE	Université de Lille	Directrice
Dr. Dominique KNUTSEN	Université de Lille	Co-Encadrante

Remerciements

Ces quatre années de thèse ont été pour moi la source de bien des challenges, mais la satisfaction et la fierté que j'en retire en valaient bien la peine. Heureusement pour moi, je n'ai jamais eu à surmonter ces épreuves seule, et c'est grâce à toutes les personnes qui m'ont entourées pendant quatre ans que j'ai pu arriver là où j'en suis aujourd'hui.

Les premières personnes que je souhaiterais remercier sont mes deux formidables directrices de thèse, pour la qualité de leur encadrement et pour m'avoir permis de tant évoluer, aussi bien professionnellement que personnellement. Merci à Angèle de m'avoir permis de travailler sur ce projet, malgré mon absence totale d'expertise préalable sur le sujet. Merci d'avoir toujours été présente, quelle que soit l'heure (souvent tardive) de la journée, ou l'état émotionnel dans lequel je me trouvais. Merci à Dominique également pour ton enthousiasme et ta positivité permanente qui illuminent ce laboratoire, et pour toutes les pauses café/réflexions que tu m'as accordée. Vous m'avez toutes deux permis, à votre façon, d'évoluer et de repousser mes limites, tout en m'offrant un cadre de travail bienveillant, et pour cela, je vous en serai éternellement reconnaissante.

J'aimerai ensuite te remercier, Maëlle, d'être restée et de m'avoir supportée pendant ces quatre années. J'espère que tu sais tout ce que tu représentes pour moi, et je tiens à te remercier de ne pas avoir fui après chacune de mes crises de nerfs ou d'angoisse, d'être restée à mes côtés même quand j'étais fatiguée ou quand je te parlais de mémoire sémantique et de représentations conceptuelles. J'aimerais pouvoir te dire que tout ça est enfin terminé pour moi, mais tu me connais, je ne peux pas encore te promettre de quitter définitivement la recherche pour me consacrer à une vie émotionnellement plus stable.

Merci également à mes parents et à ma sœur car sans vous, je n'aurai probablement pas eu ni les ressources, ni la motivation de me lancer dans des études supérieures et d'arriver là où j'en suis. Merci également de m'avoir toujours suivie et soutenue, malgré les décisions parfois hasardeuses (au sens propre) que j'ai pu prendre. Merci de m'avoir parfois remis les idées en place et poussée à ne pas baisser les bras, de m'avoir écouté parler de choses que vous ne compreniez probablement pas, mais avec toujours autant d'attention.

J'aimerai également remercier un ami qui m'aura indirectement beaucoup aidée pendant ma thèse, sans forcément toujours le savoir. Alors merci à toi Lilian, ma meilleure copine, pour ces moments de déconnection que tu m'as offert pendant ces 4 années, et bien avant cela déjà.

Et puis comme une thèse à SCALab ne se fait jamais seul dans un bureau lugubre plein de moquette, il me faut remercier mes collègues et amis de la salle des docs. Tout d'abord, je tiens à remercier Maria-Francesca sans qui cette thèse n'aurait pas été possible car c'est bien toi qui m'as conseillée de travailler avec Angèle, et c'est probablement le meilleur conseil que tu m'aies jamais donné. Ensuite, merci à toutes les personnes qui ont été de passage chez nous et qui m'ont aidées à avancer sur ce projet. Je pense évidemment à Wassila en premier lieu, ma compère préférée, à qui mon Chapitre 5 est entièrement dédié car il n'aurait pas existé sans toi. Merci également à Charlotte Stinkeste pour les coups de mains donnés sur le projet et pour ta bonne humeur. Et bien que je ne puisse pas tou.te.s les citer, merci à tou.te.s les étudiant.e.s ayant travaillé à plus ou moins long terme sur le projet, et notamment à Carla Ridel et Hélène Kwasigroch pour votre travail de toute une année sur le projet.

Et maintenant, il me faut remercier toutes les personnes qui ont fait de cette salle des doctorants un lieu où il est agréable de venir travailler. Mes premières pensées vont à la Team Langage, parce que vous êtes les meilleurs de ce labo (n'en doutez jamais). Merci à Junior et Ding d'être simplement qui vous êtes, de belles personnes avec un grand cœur. Sans vous, ces dernières années de thèse n'auraient pas été les mêmes. Merci à Matthieu pour les instants craquages et procrastination, et pour avoir été mon binôme de rédaction pendant une bonne partie de l'année.

Merci également à Fanny Grisetto et Marie Mathé, pour votre gentillesse et votre bonne humeur. Vous faites partie de ces personnes qui m'ont redonné le sourire à de nombreuses occasions, et ça, c'est tellement important quand on mène une thèse…

Je prends également un moment pour remercier tous ceux qui ne sont plus parmi nous en salle doc mais qui m'auront chacun à leur manière permis d'être la personne que je suis devenue : Yann-Romain, Amélie Menut, Fabrizia, Vasiliki, Dandan & Miao.

Et de manière générale, merci à toutes les personnes qui font de ce laboratoire ce qu'il est, tou.te.s les doctorant.e.s que je n'ai pas encore cité.e.s, Mélène, Isa, Aurélien, Angie, Emilie, Lucie, Maxime, Nahid, Lilas, Charlotte, Fanny, Louise, Clara, Nicolas, Raphaëlle, Constance…

Merci au personnel administratif sans qui nos thèses ne pourraient pas aussi bien se passer, Manu, Sabine et Ouria. Et merci au seul et unique Laurent Ott, sans qui ce laboratoire ne pourrait pas tourner aussi efficacement (et mes analyses du Chapitre 4 non plus).

Enfin, merci à l'équipe langage pour les séminaires, les retours faits au sujet de mes travaux de recherche, et pour votre bienveillance. Merci également à Isabelle Bonnotte d'avoir accepté de prendre de votre temps afin de partager votre expertise et d'échanger sur les projets qui ont rythmé ma thèse. Merci enfin à Solène Kalénine et Marion Fossard d'avoir accepté de participer à mon CSI et de m'avoir toujours conseillée avec bienveillance lors de ces rencontres.

Finally, I would also like to thank the members of my jury for agreeing to devote their time to evaluating this dissertation and for coming to Lille to attend my defense. I am very grateful for this and I look forward to discussing my research with you.

Abstract – English

During dialogue, interlocutors must coordinate their linguistic behaviours to achieve mutual understanding (Clark, 1996; Garrod & Anderson, 1987). For instance, interlocutors adjust their linguistic behaviours to their partners such that they adapt their word usage to the context of the conversation, in order to match it with their partner (Brennan & Clark, 1996). At the representational level, the interactive alignment theory (Pickering and Garrod, 2004; 2021) suggests that linguistic representations stored in semantic memory increase their level of activation to become more similar between dialogue partners. However, it remains unclear whether or not dialogue can impact linguistic representations so that they remain changed after the dialogue. In the present thesis, we focus on the changes occurring at the level of lexico-semantic representations, given that dialogue promote changes in word usage to match new meanings in particular contexts. Crucially, studies outside the field of dialogue and working on language comprehension already showed that lexico-semantic representations can be adapted following exposure to new linguistic inputs (Rodd et al., 2012; 2013; 2016). However, such adaptations have not yet been evidenced after dialogue.

The present thesis thus aims to investigate the potential impact of dialogue on the access to pre-existing lexico-semantic representations and the extent to which their organisation within the lexico-semantic network remains changed after dialogue. Two types of changes were tested in this thesis: the first concerned the possibility to create a new relationship between pre-existing lexico-semantic representations, and the second focused on the possibility to change the organisation of semantic categories. To this aim, two sets of experiments were conducted in which participants were engaged in a collaborative dialogue task about pictures during which we manipulated these two types of changes. The organisation of the lexico-semantic representations was then assessed either immediately or after a night of sleep promoting offline consolidation, in order to track the time course of the changes. Crucially, the dialogue task for the first set of experiments required us to create the first database for standardised tangram pictures, as these pictures were discussed by the participants (Experiment 1). The first set of experiments which explored the possibility to create a new lexico-semantic relationships led participants to associate two semantically unrelated words with a common tangram picture (Experiments 2 to 4). In the second set of experiments, we examined the possibility of manipulating the typicality of words by making atypical words more typical of their semantic category (Experiments 5 to 6).

Neither the creation of new lexico-semantic relationships (Experiment 2 and 3) nor the changes in the organisation of semantic categories (Experiment 5 and 6) were evidenced when tested implicitly (i.e. lexical decision task). When participants were explicitly asked to focus on the semantic relationships between the newly semantically related words (i.e. semantic relatedness judgement task), a significant impact of dialogue on lexico-semantic representations was found after a night of sleep (Experiment 4). Overall, we interpreted these findings as evidence that dialogue impacts lexico-semantic representations even one day later, but that the changes in the connections within the lexico-semantic network are likely too weak to be detected without attentional focus on the semantic properties.

As the first empirical evidence of an impact of dialogue on the lexico-semantic representations, this thesis offers both theoretical and methodological contributions to the understanding of dialogue and lexico-semantic representations, while paving the way for future research in this area.

Keywords: Lexico-semantic representations, Dialogue, Adaptations

Abstract – French

Au cours d'un dialogue, les interlocuteurs coordonnent leurs comportements linguistiques afin d'atteindre une compréhension mutuelle (Clark, 1996 ; Garrod & Anderson, 1987). Ils peuvent par exemple adapter leur usage des mots dans le contexte de la conversation, afin de les faire correspondre à ceux de leur partenaire (Brennan & Clark, 1996). Au niveau représentationnel, la théorie de l'alignement interactif (Pickering et Garrod, 2004 ; 2021) suggère que les représentations linguistiques augmentent leur niveau d'activation pour devenir plus similaires entre les partenaires de dialogue. Cependant, il reste à déterminer si le dialogue impacte ou non les représentations linguistiques, de sorte qu'elles restent changées après l'interaction. Cette thèse se focalise plus particulièrement sur les changements qui se produisent au niveau des représentations lexico-sémantiques, étant donné que le dialogue favorise les modifications d'usages des mots. De plus, des études en dehors du domaine du dialogue portant sur la compréhension du langage ont déjà montré que les représentations lexico-sémantiques pouvaient changer après une exposition à de nouveaux inputs linguistiques (Rodd et al., 2012 ; 2013 ; 2016). Cependant, de telles adaptations n'ont pas encore été démontrées après un dialogue.

Cette thèse vise donc à examiner l'impact potentiel du dialogue sur l'accès aux représentations lexico-sémantiques préexistantes, et dans quelle mesure leur organisation au sein du réseau lexico-sémantique demeure changée après un dialogue. Deux types de changements étaient testés dans cette thèse : le premier concernait la possibilité de créer une nouvelle relation entre des représentations lexico-sémantiques préexistantes et le second la possibilité de modifier l'organisation de catégories sémantiques. Pour cela, deux séries d'expériences ont été menées au cours desquelles les participants prenaient part à une tâche de dialogue collaboratif autour d'images et durant laquelle ces deux types de changements étaient manipulés. L'organisation des représentations lexico-sémantiques était ensuite évaluée soit immédiatement après le dialogue, soit après une nuit de sommeil censée favoriser la consolidation en mémoire. La tâche de dialogue pour la première série d'expériences a nécessité la création de la première base de données d'images de tangram (Expérience 1). Ces images étaient en effet discutées par les participants des Expériences 2 à 4, qui devaient associer deux mots sémantiquement non-reliés à une même image afin de créer de nouveaux liens lexico-sémantiques. Dans les Expériences 5 et 6, la possibilité de manipuler la typicalité de mots était testée, en rendant des mots atypiques plus typiques de leur catégorie sémantique.

Ni la création de nouvelles relations lexico-sémantiques (Expériences 2-3) ni les changements dans l'organisation des catégories sémantiques (Expériences 5-6) n'ont été mis en évidence lorsqu'ils étaient testés de manière implicite (i.e. tâche de décision lexicale). Lorsque les participants étaient explicitement amenés à se concentrer sur les relations sémantiques entre les paires de mots (i.e. tâche de jugement de proximité sémantique), un impact significatif du dialogue sur les représentations lexico-sémantiques était observé après une nuit de sommeil (Expérience 4). Ces résultats sont interprétés comme une preuve de l'impact du dialogue sur les représentations lexico-sémantiques un jour après, bien que les changements de connexions au sein du réseau lexico-sémantique soient probablement trop faibles pour être détectés sans focalisation attentionnelle portée sur les propriétés sémantiques.

En tant que première preuve empirique d'un impact du dialogue sur les représentations lexico-sémantiques, cette thèse offre des contributions à la fois théoriques et méthodologiques à la compréhension du dialogue et de ses relations avec les représentations lexico-sémantiques, tout en ouvrant la voie à de futures recherches dans le domaine.

<u>Mots-clefs :</u> Représentations lexico-sémantiques, Dialogue, Adaptations

Table of Content

List of Figures

List of Tables

List of Appendices

Forewords

Dialogue is one of the most common activities in everyday life. Whether it serves to achieve a common work objective or to pass time by chatting in the lift, we rely on our language abilities to communicate with other people. However, being understood and understanding our interlocutor is not as simple as it seems. One of the most obvious challenges is to make successful use of the language function, as it represents our main communication medium. For instance, when talking with someone, our goal is to successfully convey our ideas by creating appropriate utterances. To do so, we need to select words according to what we know about their meanings, but also on what we know our partner can understand (e.g. depending on their knowledge about the topic of the conversation). However, while we already know that the meaning of words and how they are used with a given partner can be stored in memory (Barr & Keysar, 2002; Brown-Schmidt, 2012; Clark & Marshall, 1981; Horton & Gerrig, 2005a; Knutsen & Le Bigot, 2014; McKinley et al., 2017), there is to date no evidence that this knowledge (i.e. meaning and use of words) can in turn be impacted by a dialogue such that it remains changed after the interaction. To better understand the challenges raised by this question, we briefly introduce the notions of language, dialogue and memory in the next part of this foreword. This introduction will be followed by the presentation of the research questions and main objectives of this thesis, which will be the last part of this foreword before the introductory chapters of this manuscript.

1. The study of language

1.1 Overall definition of language

Language is the system of words and symbols that is used to express and communicate thoughts and feelings (APA Dictionary of Psychology). It has been studied for decades in many research fields such as linguistics, philosophy or cognitive psychology, both in its expressive (speaking or writing) and receptive (listening or reading) forms. In cognitive psychology, researchers are specifically interested in understanding what underlies the ability to produce and comprehend language. Since the study of language is a vast field, researchers have divided it into several linguistic levels. One may study the way words are shaped (i.e. morphology), how they are ordered in a sentence (i.e. syntax), pronounced (i.e. phonology) or related to their meaning (i.e. semantics). In the present work, we focused on the latter domain (i.e. the semantics), which is the study of the meaning.

1.2. The study of meaning

Unlike syntax and morphology, the study of meaning is not specific to the field of language and can be found in every kind of human behaviour. However, in the present thesis we chose to focus on the study of linguistic meaning. Linguistic meaning is the field that studies the meaning of words, how the relation between word and meaning is built and updated (also called ‘lexical semantics’; Taylor, 2017), and how the meaning of a sentence depends on the meaning of the words that compose it. The study of meaning is restricted to the literal meaning of words and sentences, as found in dictionaries. In light of this definition, one question that has been explored is the way these meanings are stored as representations in the memory (Sachs, 1967). In other words, we can wonder how the external reality of the meaning of words is represented as an object or structure in the mind (Krcmar & Haberkorn, 2020). Most studies on the relationship between spoken language and the representation of meaning stored in memory rely on participants who are placed alone in a room and asked to produce either words or sentences (e.g. Barsalou, 1985), or to listen to pre-recorded words or sentences (e.g. Rodd et al., 2016). To date, the relation between the representation of meaning and dialogue situations has thus not been studied, although dialogue is one of the most frequent uses of language. To highlight the fact that dialogue is a specific setting for the study of representations of meaning compared to non-interactive settings, we present a brief overview of what dialogue represents in the field of cognitive psychology.

1.3 What is dialogue and why should we care about it?

Dialogue is a joint activity which involves that at least two partners interact using language in order to reach a common goal (e.g. choosing the hour of a meeting or the destination of a trip). This daily activity requires the partners to coordinate the timing and content of their conversation to ensure easy and efficient communication and to understand each other (Clark, 1996; Clark & Brennan, 1991; Pickering & Garrod, 2004, 2021; Sacks et al., 1974). According to the collaborative approach to dialogue, dialogue partners try to minimise the effort invested in dialogue by producing utterances designed to achieve mutual understanding (Clark, 1996; Clark & Brennan, 1991; Clark & Schaefer, 1989; Clark & Wilkes-Gibbs, 1986; Nault et al., 2023). They thus favour the production of words and utterances which are easily understandable by their partner (Clark & Murphy, 1982). The implications of this process, referred to as 'audience design', are twofold. First, the production of such words may result in the emergence of new uses and meanings for known words during dialogue. For example, when choosing the destination of a trip, interlocutor A may forget the name of one of the cities they planned to visit. In that case, this person can refer to the city using an expression such as "the city near the forest". If interlocutor B understands what this expression means, the expression gains a new meaning, as it does not anymore designate any city which can be near a forest but this particular city in which our characters are planning to go. The second implication of audience design is that partners tend to reuse the same words or expressions to refer repeatedly to the same objects throughout the conversation, to improve mutual understanding (Brennan & Clark, 1996; Clark, 1996). In our previous example, if B wants to be sure that A will understand the plan of the trip, they will be more likely to reuse the term “the city near the forest” than the real name of the city. These two phenomena are particularly interesting when it comes to understanding how the representations that are stored in memory are changed after a dialogue, as they constitute what makes dialogue an ideal setting for the integration of new uses and meanings for known words. To better understand this assumption, we will present some central notions in the literature on memory.

2. Memory systems

The aim of this section is to help the reader understand which memory representations will be studied in the present thesis and in which memory register they are stored. In particular, this section will be dedicated to the presentation of the early models of memory created in the field of cognitive psychology. This presentation should allow the reader to better understand which memory register we assume to be changed after a dialogue.

2.1 Organisation of memory systems

Memory is the system in which information is encoded and stored, and from which it is retrieved. This system can be divided into several sub-systems according to the function, type and format of information stored. One of the first divisions of memory into sub-systems was defined by the model of Atkinson and Schiffrin (1968). This model postulated the existence of three main registers of information: the sensory register, consisting in only few seconds of sensory-dependant storage of information; the short-term register, which was meant to maintain a limited amount of information for around 30 seconds; and the long-term register. The latter was defined as a relatively permanent storage of a relatively infinite amount of information, which represent at the same time a very intriguing and a very vague definition. As a consequence, subsequent models attempted to be more precise about the content of long-term memory. In the model of Cohen and Squire (1980), long-term memory is organised in two systems: declarative memory and procedural memory (Cohen & Squire, 1980). Procedural memory refers to the cognitive and motor abilities that can be stored and retrieved unconsciously, while declarative memory refers to the memories and knowledge that we are able to retrieve consciously. This declarative memory is where linguistic representations are thought to be stored, including the semantic representations which correspond to the representations of the meaning of words. Declarative memory can store representations under two different formats (Tulving, 1972). On the one hand, information that are spatio-temporally marked and organized are called episodic representations and are stored in episodic memory. On the other hand, information that have been extracted from the environment to give rise to representations that are independent of the context and of the source of learning are called abstract representations and are stored in semantic memory. While only a few psycholinguistic models consider that some linguistic representations can be stored under an episodic format (Goldinger, 1998), most agree on a more abstract format for linguistic representations (see Weber and Scharenborg, 2012 for a review in the field of word recognition). It is also important to note that in most psycholinguistic models, semantic memory is thought to contain only the semantic (abstract) representations (i.e. representations of the meaning of words), while models of cognitive psychology such as those presented above (e.g. Tulving, 1972) consider semantic memory as the storage for various types of abstract representations, including the semantic representations. Due to its ambiguous meaning, the term “semantic memory” will not be used much in the remainder of this manuscript and we will therefore refer mainly to the name of the representations (either abstract representation or semantic representation).

Finally, some models consider that semantic representations constitute only a subset of a more general and multimodal representation of knowledge, called conceptual representations (Andrew et al., 2009; Hoffman, 2018; Martinez-Manrique, 2010). In these models, semantic representations are the part of the conceptual representations accessible by the mean of a linguistic input. Therefore, studies that rely on linguistic material such as dialogue studies are limited to the examination of semantic representations. As the object of the present thesis is precisely to focus on dialogue, the following sections present the organisation of the semantic (rather than conceptual) representations stored in semantic memory.

2.2 Organisation of semantic representations

Several models have been created to conceptualize the organisation of semantic representations in memory. Semantic representations are often described as networks of highly interconnected nodes that can be activated to retrieve the meanings of words (Collins and Quillian, 1969; Collins & Loftus, 1975). These networks represent the degree of relatedness between words depending on the number of interconnections between nodes (Anderson, 1983; Collins and Loftus, 1975; Farah and McClelland, 1991; McRae et al., 1997; Plaut and Booth, 2000).

Interestingly, these interconnections have been shown to be flexible. Studies investigating semantic representations in relation to words (hereafter named lexico-semantic representations) have shown that learning new meanings in non-interactive language comprehension tasks may change pre-existing lexico-semantic representations and that these new meanings can be integrated into the lexico-semantic network (Rodd et al., 2012; 2013;

2016). However, these studies were conducted in non-interactive settings. Yet, as we already mentioned in previous sections, people may have the opportunity to create new mappings between words and meanings as they interact during dialogue. Therefore, we first investigated whether new lexico-semantic relationships which are created in the lexico-semantic network during dialogue persist after dialogue. In addition, as we know that updates in the lexico-semantic representations largely benefit from a consolidation period, especially during sleep, this adaptation was observed in both the short term (i.e. immediately after dialogue) and the long term (i.e. after one night of sleep; Gaskell et al., 2019).

Another way of characterizing the organisation of lexico-semantic representations is to think about their organisation in the lexico-semantic network in terms of categories (Rosch, 1975). Clinical studies have shown that lexico-semantic representations follow such category organisation and that not all the members of a given category are equal in terms of accessibility to the representation (Rossiter & Best, 2013). This typicality effect arises from the fact that some members of a category are judged as more representative of their category (i.e. more typical) than others, as they share more properties that are representative of their belonging semantic category. However, one may wonder whether this organisation can be updated thanks to a dialogue setting in which the representative properties of a semantic category are repeatedly associated with less typical members. This question is the object of our second research objective. In addition, as for the first objective, the question of the adaptation will be observed in both the short and long term (Gaskell et al., 2019).

3. Objectives and organisation of the chapters

3.1 Objectives

As mentioned in the previous sections, the impact of dialogue on the organisation of lexico-semantic representations has not been evaluated to date. To investigate this issue, we examined the impact of dialogue on two aspects of the organisation of lexico-semantic representations. The first one was the possibility of creating new lexico-semantic relations between already existing words. This was examined by associating two semantically unrelated words with a common visual representation. To do so, we used tangram pictures which are often used in dialogue literature to create collaborative settings for dialogue tasks (Bangerter et al., 2020; Clark et Wilkes-Gibbs, 1986). Thus, the achievement of this objective required in the first place to collect possible names for many tangram pictures that we gathered into a database of 332 tangram pictures (Fasquel et al., 2023). The second aspect of semantic representations organisation that we explored was the typicality of lexico-semantic representations. We investigated the possibility that an atypical member of a semantic category could become more typical by means of a dialogue that focused on the shared properties of the representations with its superordinate semantic category. Finally, we examined these two types of adaptation in both the short and long term to follow their time-course. The measure of the adaptations was thus performed either immediately after the interaction or after one night of sleep.

3.2 Organisation of the chapters

In order to address these objectives, we begin by presenting the literature on dialogue in Chapter 1. In this chapter, we focus on the two main approaches to the way people coordinate their linguistic behaviours in dialogue settings. Then, Chapter 2 focuses on the literature on lexico-semantic representations stored in semantic memory. This chapter begins with the presentation of the key notions of the literature and is followed by an overview of the research on adaptation and consolidation of lexico-semantic representations. These first two chapters allow us to introduce Chapter 3 in which the research questions, main objectives and general methodology are detailed. The three following chapters present the experiments created to answer our research questions. These chapters include the methodology used in each experiment, the detailed results and the discussion of those results. Chapter 4 describes the creation of the tangram pictures database. The creation of this database will be particularly interesting to obtain standardised stimuli for the second line of research. Chapter 5 presents the three experiments that used the tangram pictures as stimuli and investigated the creation of new lexico-semantic relationships between already existing representations induced by dialogue. Chapter 6 presents two experiments carried out to assess the ability of dialogue to make an atypical member of a semantic category more typical of its superordinate semantic category. Chapter 7 consists in the general discussion of the results obtained in chapters 5 and 6, in light of the literature presented in the first two chapters. Overall, the aim of this thesis is to determine whether the organisation of lexico-semantic representations can be impacted by dialogue such that it remains subsequently changed.

Chapter 1 – Dialogue

The aim of this chapter is to present an overview of the main concepts and theories that are frequently discussed in the field of dialogue. Their presentation is important to better understand the crucial role of memory representations in dialogue. The following chapter follows a historical organisation in which each theory is first presented with its key concepts, followed by a discussion of its implications in terms of memory representations.

As defined in the foreword, dialogue is a fundamental activity that corresponds to a joint action involving at least two partners who coordinate their use and knowledge of language to reach a common goal (Clark, 1996). This definition reflects the main difference between dialogue settings and non-interactive settings of language use. While the research in psycholinguistics often consider speakers and listeners in isolation in terms of cognitive processes (Allopenna et al., 1998; Cleland & Pickering, 2003; Dell, 1986; Frazier & Rayner, 1982; Gaskell & Marslen-Wilson, 1997; Levelt et al., 1999), the heart of dialogue studies lies in the comprehension of how interlocutors coordinate their linguistic behaviours and respective knowledge in a given context (Garrod & Anderson, 1987). Such coordination is crucial to ensure mutual understanding between both partners (Clark, 1996; Clark & Brennan, 1991; Pickering & Garrod, 2004, 2021). However, the processes involved constitute a highly debated question in the field. For instance, it has been shown that people tend to adapt their choices of words to their interlocutor or to reuse the same terms to coordinate their perspectives on a particular object or event (Garrod & Anderson, 1987; Brennan & Clark, 1996). Yet, little is known as to the type of memory representation involved in these processes and the impact such coordination to the interlocutor can have on these representations. Indeed, we could imagine that the changes in linguistic behaviours involved in such need for coordination could affect long-term linguistic representations, including the semantic ones, as we know that they are sensible to our everyday life experiences (Rodd et al., 2016). In the following sections, we consequently review the two main approaches that propose an interpretation of the coordination processes in dialogue. Although these theories are presented in separate sections, we believe they are not fundamentally incompatible. Instead, they convey functionally distinct yet complementary points of view regarding the mechanisms underlying coordination in dialogue, which will lead us to the question of the involvement of memory representations. The first approach presented in this chapter is the collaborative approach, which proposes behavioural evidences of the role of memory in the coordination between dialogue partners. The second approach is the interactive alignment theory which proposes a new theoretical framework on the way long-term memory representations (both episodic and abstract) end up coordinated between dialogue partners.

1. The collaborative approach to dialogue

1.1. How common ground is used for the coordination in dialogue?

In 1981, Clark and Marshall made a major proposition to account for the coordination between the linguistic behaviours of dialogue partners. Their approach was in stark contrast with the literature on psycholinguistics as it placed the dyad as the main system instead of the individual comprehender or producer. In this pioneering article, the authors posited that speakers choose their words to ensure their addressee’s understanding is guided by the knowledge that they are both aware of sharing. This statement contains two of the main ideas that paved the way for the rest of the literature on the collaborative approach: the notion of mutual knowledge and the concept of audience design, which are presented in the following sections.

1.1.1. Mutual knowledge

When Clark and Marshall (1981) proposed the notion of awareness of shared knowledge, the idea was not new per se. Before them, the notion of shared knowledge had already been studied in at least as many ways as there have been authors working on the subject (e.g. Lewis, 1969, cited by Clark & Marshall, 1981; Karttunen & Peters, 1975; Stalnaker, 1978). However, Clark & Marshall (1981) laid the foundations for the mechanisms underlying what they called 'mutual knowledge' and would later be renamed 'common ground' (Clark & Murphy, 1982). At this stage, mutual knowledge or common ground was defined as the knowledge that two dialogue partners are aware of sharing. Its content was thought to be inferred on the basis of three main components: (a) the past shared linguistic experience in conversations, (b) shared community membership (e.g. dialogue partners living in the same country or going to the same school) and (c) the physical co-presence of both partners in the same environment. This definition constituted a first break with the literature which originally stated that mutual knowledge was inferred by an iterative process. In other words, partner A was thought to infer what partner B knew about A's inferences, including inferences about B's inferences, and so on. However, such iterative model seemed irrelevant in real-life conversation, given the infinite number of iterations that may occur over an infinite amount of time. Each iteration would also have to take the previous ones into account, which would be cognitively consuming. Clark and Marshall (1981) thus proposed a more realistic theory in terms of memory processes, based on only one inference about what the other could have in mind. Such inferences were supposed to be easier and faster to produce than iterations, thus making it possible to follow the sustained rhythm of dialogue in a more realistic way. Therefore, they posited that coordination in dialogue was achieved through a process of inferences about mutual knowledge, instead of inferring what each partner has in mind individually. However, at this stage the mutual knowledge was thought to be stored in specialised memory storage and was not meant to interact with ordinary memory representations (e.g. episodic or abstract representations).

1.1.2. Audience design

In the same paper, Clark and Marshall (1981) also suggested that mutual knowledge was used to shape both the production and comprehension of a given sentence. This idea was later developed by Clark and Murphy (1982) and called 'audience design'. Audience design is a feature of an utterance that makes it understandable by the interlocutor for which the speaker constructed it, thanks to their mutual knowledge. For instance, when speakers choose a particular word or expression to refer to a particular object, they consider that this reference allow their addressee to uniquely identify the object referred to. This example was, besides, not taken as random as the choices of references (i.e. word or expression used to refer to an object) constitutes the most studied phenomenon in dialogue and is largely discussed in the following sections (Bovet et al., 2023; Brennan & Clark, 1996; Clark & Marshall, 1981; Clark & Wilkes-Gibbs, 1986; Hanna et al., 2003). Yet, choosing a reference according to the addressee is not specific to dialogue and can also be observed in a monologue given for a specific audience. But the concept of audience design is only part of a more general framework developed by Clark & Wilkes-Gibbs (1986) about the way dialogue partners coordinate with each other, based on their common ground.

In their collaborative model, Clark & Wilkes-Gibbs (1986) suggested that dialogue partners tend to coordinate their contributions in a collaborative way to ensure mutual understanding. This coordination is believed to have an impact on both the form of the discourse and on the resulting mental processes. On form, the coordination impacts the way speech turns are organized when new information is provided by a speaker. More specifically, interlocutors discuss the information until they both possess enough knowledge about it for the current purpose and can agree that it has become part of their common ground. This process is called 'acceptance cycle'. It does not mean that the information is accepted as true, only that both partners have established the mutual belief that the information is part of their common ground. On the other hand, coordination between dialogue partners also impacts the mental process according to which the information is supposed to be mutually accepted and added to the common ground, which has been termed 'grounding'. Generally speaking, some conversations may lead to the grounding of information such as the time for a meeting after a few negotiation speech turns. However, the most studied case of grounding in dialogue, as already mentioned, is the way conversational partners negotiate the choice of references (see Bovet et al., 2023 for a review). For instance, a new reference may be grounded because interlocutors refer to a new object in a conversation, or because the original reference to that object lacks precision in the context of the conversation. In such cases, the grounding of the new association between the reference and the object is referred to as a 'conceptual pact'. More precisely, the conceptual pact constitutes a flexible and temporary agreement on the perspective to adopt on an object, represented by a reference that has been negotiated between partners (Brennan & Clark, 1996).

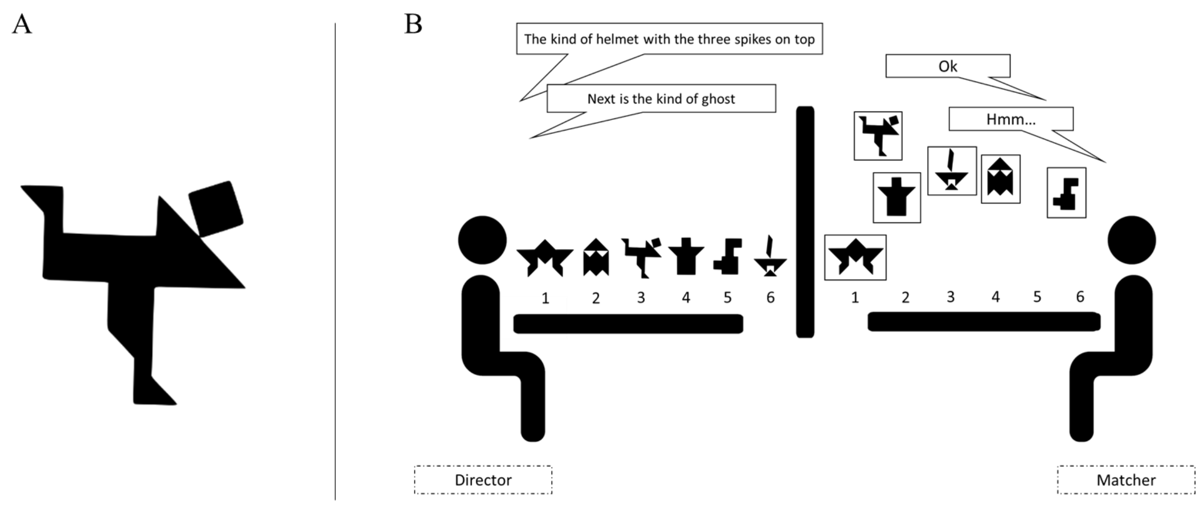
The fact that conceptual pacts are meant to be temporary refers to the many situations in which a reference is changed because there is a need for clarification. For example, imagine a conversation between a man and his wife, talking about their neighbour's new dog. If the man wants to speak about the dog, he will probably simply call it "the neighbour's dog". If his wife accepts this term by showing that she understood the referent, the conceptual pact is established. As a consequence, both partners will tend to reuse the same term repeatedly to refer to the dog in their following conversations, to maximise their mutual understanding. This phenomenon has been coined 'lexical entrainment' (Brennan & Clark, 1996; Garrod & Anderson, 1987). Now imagine that their other neighbour also adopts a dog. The man will not be able to use "the neighbour's dog" as a reference anymore to talk with his wife, as she may no longer be able to uniquely identify the reference precisely. The man will have to be more specific in his reference if he wants to specify, for example, which of the two dogs is barking. In this situation, the first conceptual pact established with his wife has to be broken and a new one is needed to refer uniquely to the first dog arrived in the neighbourhood. A more specific phrase such as "the neighbour's grey dog" is then more likely to be used. This illustrates why conceptual pacts are temporary and how coordination in conversation relies on the update of the common ground (Clark & Brennan, 1991; Garrod and Anderson, 1987).

This conclusion may have major implications on the relation between dialogue and memory representations, as it shows how flexible representations used in dialogue have to be to match what is observed in everyday life conversations. It consequently enlightens us on way representation can be mobilised during dialogue. However, before going into more detail on the relation between dialogue and memory representations, it is necessary to present other concepts from the collaborative approach that anchor dialogue into a process of coordination and may have implications for memory representations.

In their article, Clark and Wilkes-Gibbs (1986) referred to the principle of 'minimization of the collaborative effort' to explain how coordination can emerge in dialogue. The novelty of this principle lies in the notion of collaborative effort, which corresponds to the effort put into the coordination at the level of the dyad instead of at the individual level. Historically, the theory of the least effort considered that speakers tended to minimize the effort put into the production of their utterances (i.e. creating shorter utterances) by giving only the minimum of information needed by the addressee (Brown, 1958; Olson, 1970). However, Clark and Wilkes-Gibbs (1986) showed that this theory did not match the reality of a conversation. To do so, they derived a paradigm first created by Krauss and Weinheimer (1964, 1966), the referential communication task, and asked two participants to discuss abstract tangram figures. Tangram figures are abstract pictures which are usually made from multiple geometrical figures (squares, triangles, or parallelograms; see Figure 1a for an example). The main advantage of these figures is that their abstract nature forces people to find an agreement on how to name them, owing to the absence of pre-existing labels. In the study by Clark and Wilkes-Gibbs, participants were presented 12 tangram figures in random order and were asked to perform a matching task (see Figure 1b). One of the participants (identified as the director) had to give instructions to the other participant (identified as the matcher) on the way the figures had to be arranged to match his/her own order of presentation. Participants thus had to find an agreement on how to name the figures in order to communicate efficiently. They also had to rearrange the figures in six consecutive trials, always using the same figures but in different orders. This allowed to examine the way conceptual pacts created to name the pictures evolved throughout the conversation.

Thanks to this task, Clark and Wilkes-Gibbs (1986) demonstrated that both the speaker and the addressee tended to put more effort into the first utterances (e.g. over specification of the reference by the speaker; more feedbacks produced by the addressee) in order to minimize

Figure 1. *A: Example of a tangram picture. B: Example of a matching task with tangram pictures*



the number of speech turns needed to reach acceptance. This result was thus more in line with the principle of least *collaborative* effort than the principle of least effort, as the effort was minimized at the dyad level (i.e. less speech turns), rather than the individual level. Interestingly, the notion of *collaborative* effort confirms the idea that speakers tend to take their addressees into consideration in their production of utterances, a mechanism that is discussed in more detail in the next section.

1.1.3 Partner specificity of the common ground

The literature about common ground assumed that the mutuality of the shared knowledge was specific to the partner with whom the information was shared, and this was especially true for the conceptual pact. In their study, Brennan and Clark (1996) used a matching task to demonstrate that once a conceptual pact is established between two interlocutors, speakers tend to reuse this pact in subsequent references even though they may use simpler references in later trials. In comparison, when speakers are asked to change addressee for the last trials, the conceptual pact was immediately abandoned and the simplest reference was preferred. In addition, similar results were found from the addressee's side by Metzing and Brennan (2003). In their study, they used an eye-tracking measure to demonstrate that addressees are slower to look at a target object if the speaker with whom a conceptual pact has been established uses a new reference, than if a new speaker uses a new reference. Therefore, breaking the conceptual pact affects early processing of the target object from the addressee's side, showing again the partner-specificity of the common ground. However, while the partner specificity of the common ground was proven to have an impact on both the production and comprehension of references in matching tasks, common ground was suggested not to be always necessary to comprehension and dialogue not to be a collaborative process at all times. In the following section, we present literature that questioned the use of representations that have been stored in the common ground and in which circumstances interlocutors may not rely on them during dialogue.

1.1.4 Questioning the use of common ground in dialogue

The idea that conversational partners use their common ground to tailor utterances that are easy to understand for their addressee has been the basis of a big part of the literature on partner's coordination in dialogue. However, the fully collaborative aspect of this approach has been questioned by some authors, who consider that creating utterances that are specifically designed for the addressee is too costly (Barr & Keysar, 2002; Keysar et al., 1998; Keysar et al., 2000). These authors favour an interpretation of coordination in dialogue in terms of low-level egocentric processes such as frequency and recency effects. In other words, instead of being the result of an audience-designed lexical choice, references are chosen because they are available in memory representations, either because they have been the most recently used reference or because they are highly frequent in the given language. The tenants of the egocentric approach thus challenge the collaborative view of dialogue by proposing a 'two-stage model' in which the first processes used to produce or comprehend sentences are egocentric and may later, if needed, be supported by a more effortful adjustment specific to the partner (Brown & Dell, 1987; Horton & Keysar, 1996; Barr & Keysar, 2002).

However, the two-stage model has not been the only model to challenge the concept of partner specificity and audience design in dialogue comprehension. A second alternative approach suggests that the conversational partner and its associated common ground act as a contextual cue that helps retrieve the most appropriate linguistic behaviour, including the lexical choices (Hanna & Tanenhaus, 2004; Hanna et al., 2003). It has long been known that the context of a sentence can sustain its understanding by constraining the possible productions or interpretations (McRae, Spivey-Knowlton, & Tanenhaus, 1998; Spivey & Tanenhaus, 1998;

Van Berkum et al., 2005). In line with this idea, the constraint-based model suggests that common ground acts as any other contextual cue stored in episodic memory to constrain the possible interpretations for the listener during dialogue (Brennan & Hanna, 2009; Hanna & Tanenhaus, 2004; Hanna et al., 2003). The constraint-based model may therefore be a compromise between the egocentric and the collaborative approaches, as it does not rule out that conversational partners can sometimes adopt an egocentric perspective when the common ground is not available to constrain the interpretation of the sentence. It is also more in line with the idea that common ground relies on existing (episodic) representations, instead of being a separate mechanism dedicated to the storage of dialogue-specific information. This idea is developed in more detail in the memory-based account of language production.

In sum, the collaborative approach proposes an interesting way of thinking production and comprehension in dialogue. It highlights the coordination of linguistic behaviours by showing that conversational partners coordinate speech turns to create conceptual pacts that are then stored in their common ground, together with all the information delivered during the interaction. Conversely, the content of the common ground has been shown to impact the coordination of subsequent utterances. For example, conceptual pacts constrain the process of creating and understanding the references in relation to the partner with which they have been created. Therefore, the notion of common ground is essential to the idea that information used in dialogue is stored in a memory register to be reused in subsequent utterances or interactions. However, the mechanisms by which the common ground is supposed to be used during dialogue have been subject to debate which is still ongoing. In addition, even if the collaborative approach relies on the idea that mutual knowledge is somehow stored in memory, the original studies did not propose explanations in terms of memory representations from a cognitive point of view.

In the following section, we review two models that were proposed to address these limitations. First, the memory-based model (Horton & Gerrig, 2005a) was mainly an attempt to bypass the second limitation, although the answer of the authors regarding the question of memory processes constitutes, by extension, an answer to the debate on the working mechanisms of the common ground. Second, the multiple perspective theory (Heller & Brown-Schmidt, 2023) primarily attempted to bypass the first limitation but consequently proposed a framework for the memory processes that may be at work. In both cases, these models provide different, although not necessarily compatible, explanations of how memory processes and memory representations impact and are impacted by dialogue.

1.2 New approaches to the collaborative view of coordination in dialogue

1.2.1 The memory-based model

The conceptualisation of the common ground has changed significantly since the first proposal by Clark and Marshall in 1981, in which they suggested that the knowledge used in conversation is stored in specific memory structures. This idea has however been challenged in the article by Metzing and Brennan (2003), who first introduced the notion of episodic memory. Specifically, they posited that a conceptual pact consists in the mapping of a referent with its reference, plus the context in which the mapping has been done. Such mapping would account for the partner specificity of the conceptual pact without the need to create and conceptualise new specific memory structures. Horton and Gerrig (2005b) then developed the idea according to which domain-general structures may be the basis for audience-designed utterances in dialogue. Specifically, episodic memory could work the same way in dialogue as in any other type of activity requiring memory, by keeping traces of past conversations and events, including their context. This contextual information, such as the dialogue partner, then serves as cues to retrieve the information that could help produce tailored utterances. The common ground is thus thought to be the construct in which each episodic trace associated with the presence of the partner is stored to be used during dialogue. This is consistent with the findings related to partner-specific conceptual pacts. In a second paper published the same year, the same authors developed this idea and created the memory-based model of language production (Horton & Gerrig, 2005a). The aim of this model was to provide an explanation for each mechanism previously described in terms of ordinary memory functioning instead of postulating the existence of diaries (as in Clark and Marshall, 1981). Although the model does not rule out the existence of common ground, Horton and Gerrig (2005a) divided the concept of audience design into two interdependent processes: commonality assessment and message formation. Commonality assessment evaluates the likelihood of specific information being shared with a particular addressee, to determine the percentage of chance that a knowledge is part of the common ground. The automatic process underlying commonality assessment is thought to rely on a cue-based mechanism called resonance, which is supposed to search automatically into a network of information stored in long-term memory if there is any association between the addressee and the information. The more a piece of information is consistently associated with the addressee, the more it is retrievable to constrain language production. Conversely, message formation is thought to constrain the construction of a sentence, as for instance the choice of reference. As such, message formation is the process of how to refer to a particular information that is believed to be shared with the partner (thanks to the commonality assessment). If we take the example of a matching task, while the commonality assessment evaluates whether the picture of the ice skater itself is part of the common ground, message formation consists in finding the best possible reference to ensure that one's addressee can uniquely find the picture one is referring to. This is where the commonality assessment and the message formation differ. However, they also converge in that they both constrain sentence production. This model thus brings an interesting view on the format under which information provided during a dialogue is stored in the common ground (i.e. episodic representations) to constrain subsequent production of utterances and sustain coordination between dialogue partners. In addition, the fact that information stored in the common ground is meant to rely on ordinary memory representations and can be automatically retrieved to constrain, as any other type of information, the formation of utterances, seems to be a more realistic proposition from a cognitive point of view in comparison with the original proposal made by Clark and colleagues (Clark and Marshall, 1981; Clark and Murphy, 1982).

1.2.2 Multiple Perspectives Theory of mental states in communication

Another alternative view to the original collaborative approach was the model by Heller and Brown-Schmidt called the Multiple Perspectives Theory (MPT; 2023). The latter went beyond the notion of common ground by elaborating a cognitive architecture based on findings from cognitive studies that account for the coordination between dialogue partners' representations. According to the MPT, conversational partners are able to communicate successfully thanks to the constant comparison between the *representations of self* and the *representations of other*. Therefore, what distinguishes the MPT from the collaborative approach is its ability to represent both the similarities and the differences between each other's representations, whereas the collaborative approach only relies on the notion of representation of mutual knowledge.

To do so, the MPT is composed of three components. The first one, named the 'representation of self', corresponds to the activation of one's own representations relevant to the discussion. Such representations are derived from the egocentric approach to dialogue (Barr & Keysar, 2002; Keysar et al., 2000; Knutsen & Le Bigot, 2014), according to which our own representations are taken into account independently and prior to any mutual knowledge. However, in the MPT, they are considered together with the second component of the model, the 'representation of other', which refers to the inference of what representation a conversational partner has probably activated due to their relevance for the current conversation. Interestingly, the authors suggest that this component may be related to the well-known mechanism of Theory of Mind, which refers to the ability of people to represent the mental state of others in order to predict and understand their behaviours (Achim et al., 2015; Apperly, 2012; Premack & Woodruff, 1978). Finally, the third component of the MPT is a process that works continuously throughout the conversation to compare the content of the first two components. In other words, this component is meant to continuously compare the representations of self with the inferred representations of the other, and to give an output in terms of similarities and differences between them. As to what kind of representations are supposed to be compared, the authors posit that only representations that are relevant to the discussion are compared to limit the cognitive cost of such a process. Thus, this process is supposedly what allows conversational partners to coordinate their perspective on a given object. The subsequent output is expected to support the update of both representations of self and other which then guides the subsequent utterance. This may serve to overcome comprehension issues or to keep track of differences in perspectives as in the case when debate with opposite opinions.

As previously mentioned, this approach relies on evidence from cognitive studies to answer the question of the kinds of representations (i.e. episodic or abstract representations) that are activated during dialogue to ensure the coordination between conversational partners. It provides an interesting alternative to the notion of common ground by including both shared and non-shared knowledge in the representations used during dialogue. However, detail about the cognitive functions involved in each of the three components is lacking, with only the second component having been defined as relying on the Theory of Mind. In addition, while the model heavily relies on memory representations, it does not specify under which representational format (episodic or abstract representations) the ‘representations of self’ and ‘representations of others’ are stored. The only information provided in this model is that, contrary to what had already been proposed (Horton & Gerrig, 2005a; Metzing & Brennan, 2003), episodic traces cannot be the only representations used during dialogue coordination and other memory representations are probably also involved.

In the following section dedicated to the relation between dialogue and memory, we take a closer look at Heller and Brown-Schmidt’s arguments (2023) against episodic representations necessarily supporting dialogue processes. However, before presenting these arguments, we first review how memory representations in dialogue have been conceptualised since the beginning of the collaborative approach. By making some extrapolations on the original proposals, we attempt to explain why episodic representations have been presented as good candidates for dialogue coordination, before examining the counter arguments made by Heller and Brown-Schmidt in their model.

1.3 The role of memory representations in dialogue

In his first theory about common ground (Clark and Marshall, 1981; Clark and Murphy, 1982), Clark made several assumptions regarding the type of memory representations used in dialogue. In his proposal, the memory registers used are two-fold and dialogue-specific (see Table 1). The first one, called ‘encyclopaedia’, is the register of all the generic and particular knowledge shared between all the members of a given community. This could be for example the knowledge that an American citizen has about the history of his country. Interestingly, this notion of general knowledge resembles the definition of semantic memory proposed by Tulving (1972). The second register proposed by Clark is a diary for personal and shared experiences, which is close to Tulving’s notion of episodic memory (Brown-Schmidt, 2012). Although Clark’s work did not focus on cognitive processes, his pragmatic proposal on dialogue

Table 1. *Summary of the models presented in the first section of the introduction. NA means that the information is not specified in the model.*

Name	Main authors	Section (page)	Is common ground involved	Memory registers involved
Collaborative approach	Brennan & Clark (1996); Clark & Marshall (1981); Clark & Murphy (1982); Clark & Wilkes-Gibbs (1986); Garrod & Anderson (1987)	1.1.1 - 1.1.3 (p.10-17)	Common ground is involved at every stage during the dialogue	Dialogue-specific memory registers: 'encyclopaedia' and 'diary'
Two-stage model	Barr & Keysar (2002); Keysar et al. (1998; 2000)	1.1.4 (p.18)	Common ground is only optionally involved in later adjustments	Not specified by the model
Constraint-based model	Hanna & Tanenhaus (2004) Hanna et al. (2003)	1.1.4 (p.18-19)	Common ground act as any other contextual cue to constraint comprehension	Not specified by the model
Memory-based account	Horton & Gerrig (2005a)	1.2.1 (p.21-22)	Common ground act as any other contextual cue to constraint production	Episodic memory
Multiple Perspective Theory	Heller & Brown-Schmidt (2023)	1.2.2 (p.23-24)	No common ground involved	Episodic memory + other unspecified registers
Interactive alignment theory	Pickering & Garrod (2004); Pickering & Garrod (2021)	2.1 – 2.2 (p.29-35)	Implicit common ground	Episodic memory + semantic memory

coordination could, in a way, be related to memory processes. However, in later contributions to the collaborative approach, the relation between dialogue and memory evolved towards domain-general memory representations (Metzing & Brennan, 2003; Brennan & Hanna, 2009; Horton & Gerrig, 2005a; 2005b). In particular, Horton and Gerrig (2005a; 2005b) created a framework in which the coordination in dialogue was meant to rely on episodic representations of past utterances. This claim was essentially based on the findings about the partner-specificity of the conceptual pact, which was considered as an evidence for the existence of an episodic trace of the mapping between the conceptual pact and its context of creation (e.g. the partner). This model also specified that the activation of episodic traces during dialogue (either automatic or strategic), constrain the retrieval of long-term representations.

At this point, episodic representations were thus the only representations to be mentioned in the literature on the collaborative approach to dialogue (as shown in Table 1). However, later neuropsychological studies suggested that other memory representations are also likely to be used during dialogue, since patients with hippocampal amnesia, who are thought to present a deficit in episodic memory, are still able to manage collaborative referencing (Duff et al., 2006; 2011). In a matching task with tangram pictures, Duff and collaborators (2006; 2011) demonstrated that these patients were still able to learn the labels of each picture that they negotiated with their confederate. More importantly, the partner specificity of such learning was later shown to remain intact, with amnesic participants still able to switch from their conceptual pact to less specific references when their partner changed from one trial to another (Yoon et al., 2017). Interestingly, these patients were still impaired in their ability to use definite reference correctly to mark their use of common ground. These findings were interpreted as the fact that patients with amnesia, instead of recalling the conversation itself, developed a shared perspective with their partner on the way the picture has to be seen. From these results, it was suggested that coordination in dialogue do not rely only on episodic traces but other memory representations also (Duff et al., 2006; Heller & Brown-Schmidt, 2023). Such findings also highlight the need for the literature on dialogue to go beyond the study of episodic representations and to test the involvement of other types of memory representations such as the abstract ones.

In sum, although the collaborative approach emerged from the pragmatic literature, it evolved towards more cognitive studies in order to explain the mechanisms of dialogue coordination (see Table 1 for a summary of these models). However, while these models have attempted to test the memory representations likely to be involved during a dialogue, they have mainly focused on episodic representations. In the second part of this chapter, we review a more mechanistic approach regarding the way cognitive processes may underpin dialogue, and which explicitly proposes the involvement of both episodic and abstract representations in dialogue coordination i.e. the interactive alignment model by Pickering and Garrod (2004; 2021).

2. A system-based approach to coordination in dialogue

The central idea of the interactive alignment theory lies in the word 'alignment', which refers to the fact that, during a dialogue, conversational partners develop mutual understanding by increasing the degree of similarity between their mental representations. In their first proposition of interactive alignment, Pickering & Garrod (2004) presented the mechanisms underlying alignment and their impact on the coordination between conversational partners. This proposition was later enriched in their book published in 2021, which offered to bring a more complete answer to the question of collaborative joint-action in dialogue settings. In this section, we begin with the presentation of the bases of the interactive alignment model as they have been proposed in 2004, and then highlight some of the key enrichments provided by the model of 2021. The choice of presenting both models sequentially allow us to present the key concepts of this theory with a gradual level of complexity, as the model of 2021 covers the one of 2004 while providing several additional components. Our final section examines the implications of such an approach in terms of memory representations in dialogue.

2.1 The interactive alignment theory: a first version of the model

2.1.1 Alignment of the situation model

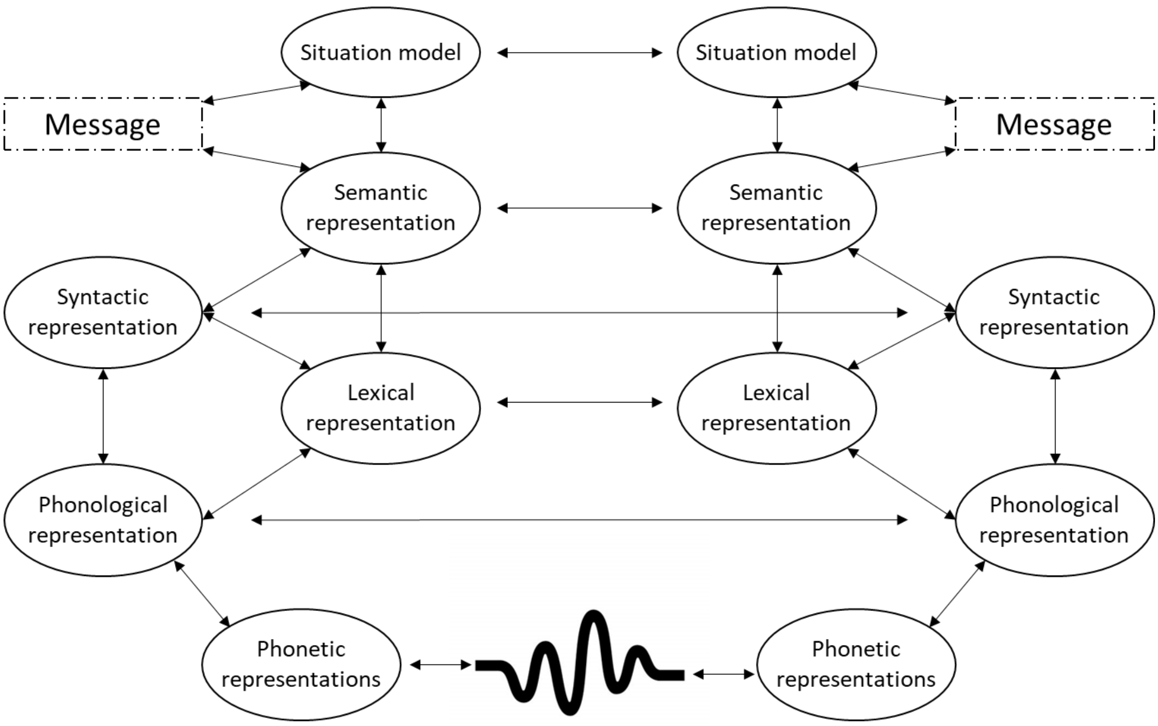
The interactive alignment model is based on the postulate that conversational partners develop an aligned situation model of their dialogue instead of developing mutual knowledge. To better understand this definition and the implications that underlay such assumption, some concepts need to be clarified. To begin with, the situation model can be defined as the representation of all the key elements that are needed to represent a situation in a unified network of information. Since its first mention by Van Dijk and Kintsch (1983), the situation model has dominated the literature on text comprehension as a way to represent how readers organise the information extracted from a story to create a network of relations between characters and events (in terms of spatio-temporal relations, causality and intentionality; see also Zwaan & Radvansky, 1998 for an overview). For instance, someone reading a novel will be able to keep track of the goals of the characters and to relate them to their movements through space and time. In dialogue settings, the notion of situation model has been reused to explain how people creates on the fly a representation of all the key elements that are needed to understand the discussion. Therefore, the situation model is a form of "implicit common ground" (Pickering & Garrod, 2004). However, contrary to the original version of the common ground, the interactive alignment theory posits that the conversation can be successful only if both interlocutors are aligned in terms of implicit common ground (characters, spatio-temporal relations and causality). This means that both dialogue partners have to share the same state of representations, instead of A representing B's state of representations and vice versa. In the next section, we describe how the model suggests that dialogue partners end up aligned in their implicit common ground.

2.1.2 Automaticity of the alignment: the role of priming

According to Pickering and Garrod (2004), the alignment process comes from a priming mechanism between dialogue partners' representations. This priming mechanism corresponds to the activation of a particular linguistic representation that emerges from two different sources. The first source of activation in the interactive alignment model concerns the activation of one level of linguistic representation (e.g. lexical level) by another level of linguistic representation (e.g. syntactic level), within an individual. According to the model, every level of linguistic representation is taken into account and associated with the other levels in a network of mental representations. By network, we mean that every representation at a given level (or strata) is represented by a node which is connected to its representation at every other level and to other nodes within the same level. In other words, the alignment at one level, let's say the lexical one, is meant to lead to alignment at other linguistics levels as the syntactic one (see vertical and diagonal arrows in Figure 2). As soon as all levels are aligned, conversational partners develop 'routines' that are fixed expressions in regards to all the linguistics levels, that can be reused throughout the entire conversation. Such routines are by definition easy to produce and understand because of the reduced need to select the appropriate reference, meaning or articulation. They are created through a process called 'routinisation'.

The second source of activation concerns the alignment between conversational partners' representations. This means that the production of a linguistic input by a speaker at a particular linguistic level of representation (lexical, semantic, syntactic, etc...) primes or activates the related representation in the listener's linguistic representations (see the horizontal arrows in Figure 2). In addition, the model specifies that both production and comprehension systems are equally affected by the alignment process. For instance, alignment in comprehension within the listener's linguistic representations leads to alignment in production,

Figure 2. *The Interactive Alignment Model (Pickering and Garrod, 2004)*



the reverse being also true (i.e. alignment occurring from production to comprehension). Consequently, the linguistic representations activated in the listener's semantic memory become more accessible for his production system and vice versa. This alignment between dialogue partners' representations increases the probability for each of them to reuse the linguistic inputs in subsequent productions and facilitate comprehension of these inputs in subsequent utterances interpretations. For example, when speaker A expresses his/her excitement about getting new socks, the bi-directionality of alignment means that as soon as speaker B understands the word, he/she will be aligned on his/her own production and use the same reference. The fact that both production and comprehension systems are supposed to be highly co-activated at the same time accounts for the speed at which interlocutors switch roles from listener to speaker and vice versa during a natural conversation. As noted by Pickering and Garrod (2004), listeners need to be constantly prepared to take their speech turn at the same time they are comprehending.

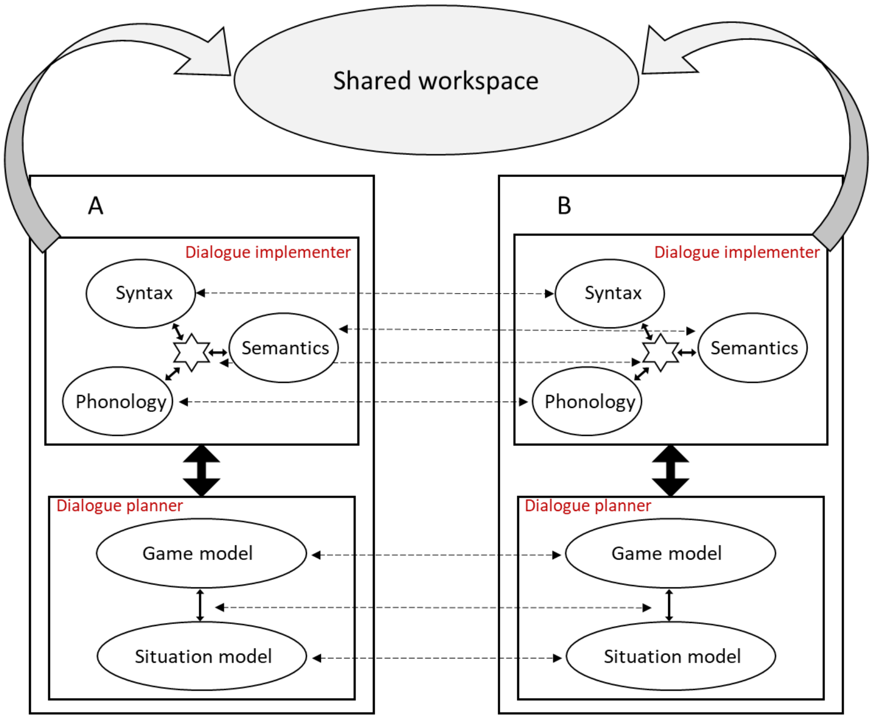
These mechanisms of interactive alignment provide an interesting explanation for some of the well-known phenomena described in this literature in dialogue. For instance, Pickering and Garrod (2004) suggested that what was previously considered as evidence of the lexical entrainment mechanism (Brennan & Clark, 1996; Garrod & Anderson, 1987) may be considered as evidence of alignment in lexical representations. Lexical alignment thus corresponds to the fact that a given lexical representation has been activated or primed to refer to a given object in the speaker's utterance, and is thus more likely to be reused in the following listener's answer to refer to that object again. For example, if A speaks about his new socks to his best friend B, the reason why B tend to reuse the term "new socks" instead of calling them the "beautiful socks" or "red socks" is because the reference "new socks" was activated by A's utterance and is now accessible in B's lexical representations. In this regard, the 'routines' proposed by Pickering and Garrod (2004) are also reminiscent of the notion of conceptual pact proposed by Brennan and Clark (1996), as they both constitute fixed references that are then reused by dialogue partners. Evidence of this alignment mechanism can also be found at other levels in the literature. For instance, syntactic priming was evidenced in studies where conversational partners tend to imitate each other's syntactic structures (Branigan et al., 2000; Hartsuiker et al., 2008). Pickering and Garrod (2004) also discussed alignment at the level of articulation, which had previously been characterised by Giles (1973; Giles, Taylor & Bourhis, 1973; Giles & Powesland, 1975) as part of the 'convergence' phenomenon (i.e. whereby individuals moves from their own socio-culturally marked accent or pronunciation to that of their partner). However, in the interactive alignment theory, convergence is seen as a consequence of the priming mechanism between dialogue partners' representations of articulation, according to which conversational partners prime each other's accent and speech rate until they align on the way words should be pronounced and at which pace.

To conclude on this first model, Pickering and Garrod (2004) based their assumptions on the behavioural findings of the collaborative approach to provide a mechanistic account of how the coordination between partner's linguistic behaviour could work. Crucially, this model suggests that during a dialogue, linguistic representations receive direct activations and that these activations could leave a trace in semantic memory such that they remain changed after dialogue. However, in the latter version of the model, the way linguistic representations are primed has been slightly changed. In the next section, we consequently review these changes that could impact the way linguistic representations could be affected by a dialogue setting.

2.2 Updated version of the interactive alignment model (2021)

Figure 3 illustrates the updated version of the interactive alignment model, as presented in the book by Pickering and Garrod (2021). As already mentioned, we will not review all the differences between the earlier and the later versions of the model (Pickering and Garrod, 2004;

Figure 3. *The Interactive Alignment Model from Pickering and Garrod (2021; Figure 6.5)*



2021), but instead highlight the key insights on the relation between the interactive alignment theory and memory representations in dialogue.

One of the major differences lies in the way each linguistic level interacts with the others. Indeed, the different linguistic levels are no longer believed to interact directly with each other but are instead supposed to interact via a 'binding node' (see the star in Figure 3). The general idea is that, as in the previous model, each linguistic representation corresponds to a stratum containing nodes that represent either concepts, words or phonemes depending on the stratum. When a speaker utters a word, the resulting activations of each stratum are thought to merge into the binding node to give rise to a 'bundle' of linguistic representations. This 'bundle' constitutes the schematic representations of the routines previously mentioned in the 2004 model and interacts with each individual stratum. However, the fact that the interactions between that node and the various levels are bidirectional means that the activation at one level may influence the activations in the binding node and vice versa. For example, the binding between a semantic representation and a lexical one may influence each of these two representations separately.

Another major difference is that the situation model is now included into a new level of representation, the 'dialogue planner', which combines the situation model with the 'game model'. The dialogue planner relies on these two components to conceptualise the situation in which dialogue takes place. On the one hand, the game model serves as a representation of the goal of the conversation and the details of the contributions. On the other hand, the situation model is used to keep track of the key components of the discussion. Its role is to search the concepts that are relevant to the current discussion in long-term representations. These representations are then stored in the long-term working memory, which corresponds to the memory that keeps a representation of the whole conversation. As such, the dialogue planner can be considered as the component that conceptualises the discussion and underpins dialogue planning, while the dialogue implementer which contains the different linguistic levels is used to make the conversion into linguistic representations and allow dialogue implementation. The distinction between these two components can be seen in the same way as the distinction between the commonality assessment and the message formation from the memory-based model (Horton & Gerrig, 2005a) in that the planner is meant to search for what can be said and the implementer is meant to search for how it can be said.

Despite these modifications, the mechanism which enables alignment remains the same, that is a mechanism of priming that may originate from the speaker to the listener and/or from one stratum of representation to the other. Interestingly, Pickering and Garrod (2021) posited that alignment between conversational partners may either be automatic, as when they share the same knowledge thanks to their common community membership, or result from a conscious process of alignment. In that case, partners may for example explicitly negotiate a reference, as it is often the case in most studies on which the collaborative approach relies (e.g.

the matching task; p.131). However, alignment is in both cases the consequence of the retrieval (either automatic or conscious) of the long-term memory representations that are relevant to the current discussion. In the next section, we examine the kind of representations that are thought to be retrieved from long-term memory.

2.3 The interactive alignment theory and memory processes

While the relation between dialogue coordination and memory representations was mainly subject to speculation in the collaborative approach, the interactive alignment model (Pickering and Garrod, 2021) provides more details as to the type of memory representations that may be involved in the process of alignment. First, the authors posited that the memory representations used at the various levels of the model are all stored in long-term memory. However, as it has been presented in the preface of this work, long-term memory contains several forms of memory representations. The authors thus focused on the declarative memory, meaning that they include both the representations stored in the episodic and semantic memory in their theory. The fact that the model includes both these registers is in line with the Multiple Perspective Theory (Heller and Brown-Schmidt, 2023) and with neuropsychological evidence (Yoon et al., 2017; Duff et al., 2006), which suggests that coordinating each other's utterances requires both the episodic representations and abstract representations. The model also provides more specific information about the use of each type of representation. Owing to its nature, the dialogue implementer is thought to rely only on linguistic representations. Conversely, the dialogue planner is thought to use both episodic and abstract representations (probably with the exception of the linguistic representations). In other words, the situation model of the planner contains all the 'background knowledge' that is relevant to the conversation.

With regard to this information, it is clear that the new version of the interactive alignment model (Pickering and Garrod, 2021) provides a more complete framework for the relation between dialogue and memory representations as it suggests the implication of both the episodic and abstract representations (including the linguistic representations) to sustain dialogue coordination. While this new version suggests that episodic representations interact with the linguistic representations (as represented by the bidirectional arrow between the dialogue planner and implementer on Figure 3), the idea by which alignment between dialogue partners can still occur directly at the level of the linguistic representations is in line with the previous version of the model (Pickering and Garrod, 2004). However, this proposal marks a break with the literature on semantic representations because it suggests that semantic representations could receive direct activations during a dialogue setting. While the model by Pickering and Garrod (2004; 2021) only suggests changes of activations in linguistic representations *during* dialogue, the literature on semantic representations informs us that such a change of activation could leave traces in the organisation of the semantic representations after dialogue. To better understand this idea, we detail the main lines of research in the literature on semantic representations in the next chapter.

Chapter 2 – Semantic representations

The fact that Pickering and Garrod (2021) mentioned the involvement of semantic representations in their model of dialogue is of particular interest, as the question of the relation between memory representations and dialogue is rather recent and has only focused on episodic representations rather than abstract representations of the discussion (Horton & Gerrig, 2005a; Nault et al., 2023). In the next part of this work, we review ways in which semantic representations can be modelled and evaluated in psycholinguistics. Since the aim of this thesis is not only to determine how to measure the access to semantic representations but also to better understand the way dialogue can impact their organisation, we present some of the most influential models of organisation of semantic representations. After this first section on the structure of semantic representations and how it can be evaluated, we present experimental evidence that shows how this structure can be adapted after exposure to linguistic inputs. This presentation sets the scene for examining how dialogue constitutes an ideal setting to change the organisation of semantic representations in adults.

---

1. The organisation of semantic representations

The first well-known attempt to model the organisation of semantic representations was Osgood's spatial model in 1952, in which semantic representations were computed as coordinates in a multidimensional space. This computation was created based on judgements of people evaluating word meanings according to various dimensions. Although this multidimensional space was not reused in following models, their method of computing data about judgement of participants on a series of words to represent the organisation of semantic representations was reused in subsequent models. The aim of these models was to imitate human data in tasks that explore the processing of semantic representations. However, a common limitation to these early models was their inability to account for the process of how these semantic representations and their structure are learnt. In this first sections, we review some of the most influential models on the conceptualisation of the organisation of semantic representations. We then describe how researchers have attempted to model the way these semantic representations can be learned and adapted, especially through linguistic exposure. The second half of this chapter is dedicated to the presentation of behavioural evidences of these changes in the organisation of semantic representations through linguistic experiences. Specifically, we provide some insights on the possibility to change the organisation of semantic representations by either adding new information or adapting already existing ones, in order to suggest, in light of what has been presented in the previous chapter, that dialogue may directly impact this organisation.

1.1 Semantic network models and semantic priming

Semantic network models are the first and probably the most influential models that were ever designed to explain the organisation of semantic representations. The proposition that is made on how semantic representations are organised was primarily thought to account for the so-called ‘semantic priming effect’. We consequently begin this section by reviewing how this effect allows researchers to study the relations between semantic representations, before going into the description of the semantic network models.

1.1.1 Evidence of the organisation of semantic representations from the semantic priming paradigm

As defined in McNarama’s book (2005, p.3), semantic priming refers to an improvement in either the speed or the accuracy of a response to a stimulus (word or picture), if the latter has been preceded by a semantically related stimulus (word, sentence, picture…). It was evidenced for the first time in 1971 (Meyer & Schvaneveldt) and has since been the object of much attention by researchers who used this effect as a relevant way to examine the access to semantic representation (Collins & Loftus, 1975; Ratcliff & McKoon, 1988). In a classical semantic priming paradigm, two words are successively presented and the first word, the prime, can be either related or not to the second word, the target. This paradigm is frequently coupled with a lexical decision task which is often considered as the best tool to examine word recognition. In this task, participants have to indicate whether the second stimulus is a real word or not (Neely, 1977; see McNamara, 2005 for a review). Importantly, the study of the semantic priming effect has raised the question of what can be considered as a semantic relation between words, which is a central question when investigating the organisation of the semantic representations and their interconnections.

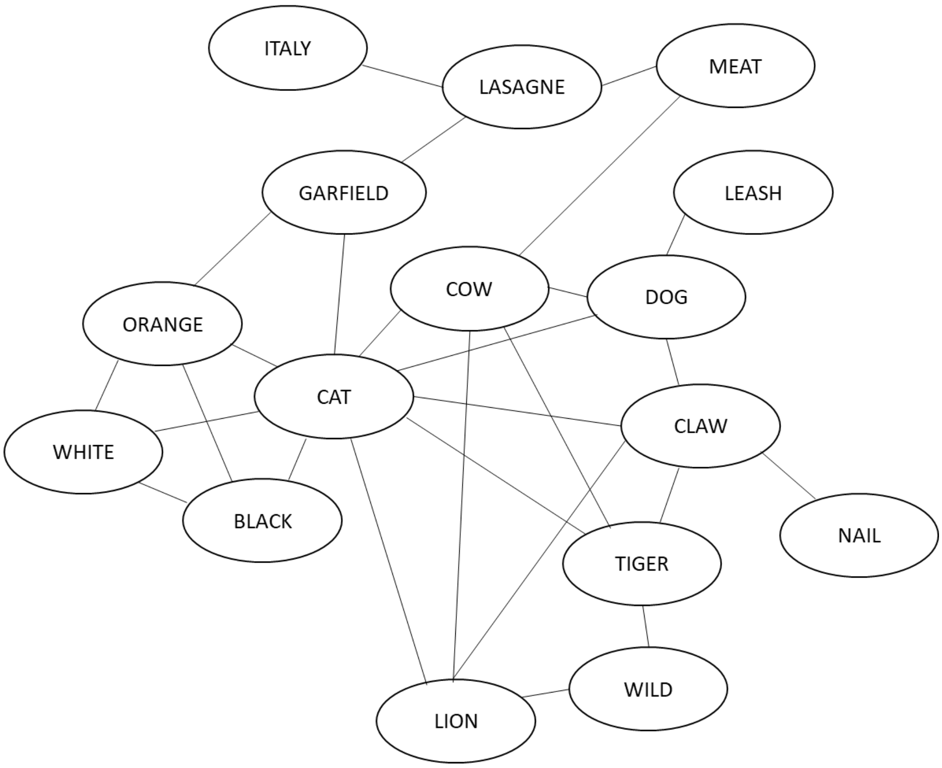
In particular, the distinction between semantic and associative relations between words has been the centre of much debate in the field. The term associative relation refers to the associative norms, which are obtained by asking participants to give the first word that comes to mind when they are provided a word (Nelson et al., 2004). For example, participants are shown the word 'cow' and are asked to provide the first word that comes to their mind (e.g. milk). Such norms are consequently informative of the way semantic representations can be related to one another in the semantic memory. However, the type of relation that is involved in these norms may be questioned. It is generally acknowledged that the nature of associative relations is twofold. It reflects at the same time 'pure' semantic relations as defined by feature similarities (i.e. perceptual, functional or categorical; McRae & Boisvert, 1998), and relations that can be inferred from statistical regularities in natural language (i.e. co-occurrences between words; Hutchison, 2003). In the previous example, while 'cow' and 'milk' do not share many semantic features, they still share a high co-occurrence frequency in English. The nature of the associative relations has made researchers wonder whether it was possible to distinguish the effects of statistical regularities from those of 'pure' semantic relations in terms of lexical access. This idea was tested by manipulating the stimulus onset asynchrony (SOA) between the prime and the target word, which represents the duration between the onset of the prime and the onset of the target word (de Groot, 1984; Lucas, 2000). Indeed, it has been observed that semantic relations seemed less impacted by the changes in SOA than associations based on co-occurrence frequencies (Roelke et al., 2018). While semantic priming and co-occurrence frequency effects seemed equally effective at short SOA (200ms), the effect of co-occurrence frequency was the only one to significantly increase with a longer SOA (1000ms). These results have been taken as evidence for distinct, yet complementary effects of the relation between pairs of words on the strength of the semantic priming effect. In addition, it has been shown that while purely semantic relations could be hard to evidence alone, the addition of an associative relation between words could 'boost' the effects of semantic priming and make it easier to evidence in a semantic priming paradigm (i.e. associative boost; Hutchison, 2003; Roelke et al., 2018). Interestingly, the same pattern was also observed in a study that used co-occurrence frequencies instead of associative relations to boost the purely semantic priming effect (Brunellière et al., 2017).

In sum, when one wants to study the relations between semantic representations by using lexical decision tasks and semantic priming paradigm, caution is advised in regards with the type of relations one wants to focus on, and with the SOA chosen to evidence these effects. In the next section, we present the first models of organisation of semantic representations that have been created to account for the early findings on semantic priming effects and which keeps influencing the way semantic priming effects are interpreted.

1.1.2 Spreading activation models

One of the earliest attempts to model semantic representations was the spreading activation theory by Quillian (1967). This model proposed a structure of semantic representations in terms of network of interconnected nodes. Each node was thought to be a semantic representation that was connected to the other nodes according to their semantic relations. According to this theory (and in line with its name), the activation of one node into the network was thought to be propagated along its connections, first towards the directly linked nodes, then to all the nodes linked to the direct neighbours of the originally activated node (see also Anderson, 1983 for similar proposals). The model also specifies other characteristics of the network, such as the various types of links that can characterize the relationship between two nodes. Although Quillian (1967) was the first to propose a spreading activation account of the organisation of semantic representations, the most influential model remains that of Collins and Loftus (1975). In this extended version of the theory, the basic principles of the original theory remained the same, with semantic representations being set in a network of highly interconnected nodes (see Figure 4). According to Collins and Loftus (1975), the strength of the spreading activation decreases along the connections, meaning that less related nodes (i.e. nodes related by the intermediate of a chain of one to multiple other concepts) receive less activation. When a related node receives such residual activation, it thus becomes more accessible for future retrieval. This assumption accounts for the semantic priming effects that can be found in lexical decision tasks for example. Indeed, in such task, when the prime word is presented, the activation of its corresponding node spreads along the semantic network and pre-activates the related node. Then, when a semantically related target is presented, its retrieval is facilitated. The response time needed to tell whether it is a real word or not is thus shorter than with a semantically unrelated target word. Importantly, the word associated with the semantic representation is thought to belong to a parallel yet related network, called the lexical

Figure 4. *Representation of the organisation of semantic representations according to the spreading activation theory (Collins & Loftus, 1975)*



network. Each node in the semantic network is believed to be linked with its corresponding word in the lexical network. When a word is encountered in a lexical decision task, its activation is propagated towards its related node in the semantic network. Conversely, when a node in the semantic network is pre-activated through the process of spreading activation, the pre-activation is sent to its corresponding node in the lexical network.

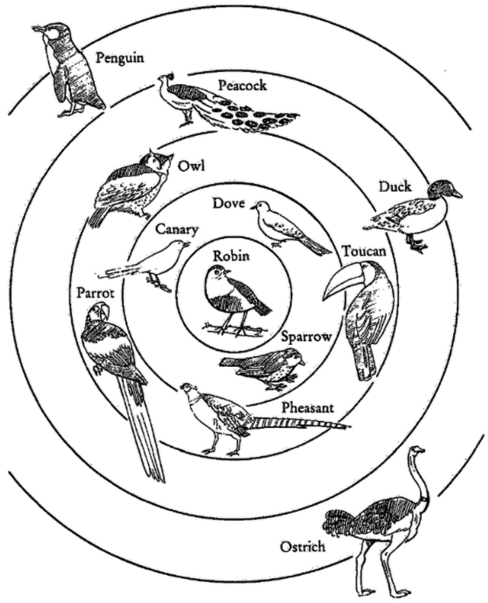
The interest of the network model by Collins and Loftus (1975) is twofold. First, the model is able to represent features for semantic representations (i.e. by connecting them into the network), and it also accounts for various other types of relations between concepts. Indeed, the connections between nodes drawn in the model represent the type of relation between the nodes. For instance, the connection existing between black and orange in Figure 4 is 'labelled' as a functional one. Second, it also provides an easy-to-understand account for the semantic priming effect and a way of quantifying the semantic relations between concepts, owing to their distance within the network. With such advantages, it is easy to understand why semantic network models continue to influence the field (De Deyne et al., 2016; Steyvers and Tenenbaum, 2005). However, this type of model is not the only attempt to explain the structure of semantic representations and other theories have emerged to account for data from other behavioural tasks such as naming or categorisation tasks. In the next section, we present one of these theories which proposes an organisation of the semantic representations in terms of semantic categories.

1.2 Organisation of the semantic representations into categories

1.2.1 Rosch's studies of semantic categories

From 1973, major papers were published to address the question of what constitutes the meaning of a word, leading to the creation of another very influential model of organisation of the semantic representations, i.e. the prototype theory (Rosch, 1973; 1975). This theory refers to the idea that semantic categories (e.g. trees, birds, cars) are created based on the extraction of clusters of features (Rosch, 1978). In other words, features that naturally co-occur in the environment form clusters that are then grouped under the same category label. For example, Figure 5 represents the category 'bird', which may have been created on the basis of natural co-occurrences between features such as 'have wings', 'fly' and 'have feathers'. The extraction of the average experience with each cluster of features (i.e. the mean of the possible variations for each feature) has been suggested to form a prototype of the category (Posner & Keele, 1968; Rosch, 1973). Prototypes are abstract representations, in the sense that they contain only the information that is relevant to the definition of the category (Hampton, 2003). They are thought to be the most central member of their category and possess the maximum number of features that define the category. As such, they serve as a point of comparison with the other members to determine their degree of typicality (Rosch and Mervis, 1975). Indeed, members of a semantic category can be rated on a continuum of typicality, with the most typical members

Figure 5. *Structure of the semantic category "bird" according to Rosch's theory (1973), as illustrated in Aitchison (2012; p.54, Figure 5.1)*



being the most similar to the prototype (i.e. the closest to the centre on Figure 5). Therefore, the more a member of a category is similar to the other members of its category, the more it is considered as typical. This notion has been called 'family resemblance' (Rosch & Mervis, 1975). Conversely, atypical members share fewer features with the other members of their category and more features with members of other categories (Rogers et al., 2015). They are consequently at the boundary of their semantic category (as illustrated on Figure 5). This graded structure of category membership can also be observed when asking participants to rate the goodness of each member of a category (Rosch, 1975; Rosch and Mervis, 1975). In fact, it has been shown that the judgements of the participants mapped closely the number of features that a member shares with the other members of its category. However, such tasks are not the only ones in which the graded structure of category membership can be found. Early studies evidenced that typical words are advantaged in terms of speed of processing. For example, typical words are categorized more quickly as belonging to their semantic category (Hampton, 1979), and they are also produced more frequently when people are asked to generate words for a given category from memory (Hampton & Gardiner, 1983; Mervis, Catlin, & Rosch, 1976). This variable access between typical and atypical words in memory has been evidenced in many cognitive and neuropsychological studies (Hampton, 1979; Kiran et al., 2007) and can easily be explained by the processes that rule the organisation of semantic representations, as detailed in the next section.

1.2.2 Typicality effects and semantic representations

Early studies on typicality demonstrated that healthy participants had easier access to typical than to atypical items of a semantic category, and this effect was demonstrated using several experimental paradigms (Casey, 1992; Hampton, 1979; Hampton & Gardiner, 1983). These results were replicated in later studies, confirming faster response times and lower error rates with typical items than with atypical items in naming tasks (i.e. participants are required to name an object from its picture; Holmes & Ellis, 2006) and category verification tasks (i.e. participants are required to check whether a target word belongs to the superordinate category presented as prime word; Kiran et al., 2007). Importantly, these findings are not incompatible with the spreading activation account presented previously. Indeed, since the earliest proposal of the prototype theory, it has been suggested that the structural evidence from Rosch's studies (1975) could easily be described in terms of spreading of activation. In a study that used the semantic priming paradigm, typical primes were shown to be more beneficial for the processing of a target from the same category than atypical ones (Brunellière & Bonnotte, 2018). More specifically, when participants were asked to perform a categorisation task or a semantic judgement on the target, only typical primes facilitated the processing of the target. This result is in line with a combination of the spreading activation theory (Collins & Loftus, 1975) and the prototype model (Rosch, 1975). Typical items are thought to be strongly and directly linked with their superordinate category within the network (e.g. 'sparrow' is strongly and directly related to 'bird'), but also with all the features that define that category (e.g. 'has wings', 'can fly', 'has feathers'…). When the activation spreads towards these features, all the items of the category receive several sources of activation that lead to their pre-activation. Conversely, atypical items are thought to have a weak connection with their superordinate category and to be linked to fewer core features of the category, thus producing less spreading of activation in the network. This explains why typical items are processed faster than atypical ones.

In the literature on the typicality effect, the difference between typical and atypical words has been shown not to originate only from a facilitation of typical words processing. In fact, it has also been evidenced that healthy participants proved to be better at remembering atypical words that have been presented in a previously performed task (i.e. category verification task), than typical ones (Alves & Raposo, 2015; Souza et al., 2022). More precisely, greater accuracy and faster reaction times were observed in the episodic recognition task where participants were asked to tell whether an item had been previously presented or not (i.e. recollection of a spatio-temporally marked information). These results were interpreted in terms of distinctiveness of the atypical items, which may have benefited the episodic memory encoding. Souza et al. (2022) suggested that since atypical items do not fit in their category, their processing relies on the recruitment of the same system that is in charge of the processing of new information, that is the episodic system. The graded category structure of semantic representations thus seems to influence the extent to which the episodic system is involved in word recognition tasks in healthy participants.

Taken together, these studies are in line with an organisation of semantic representations in terms of graded structure and semantic categories that are created on the basis of clusters of features. This organisation has a major impact on the way words are processed, as some of them may be easier to access, while others may require the recruitment of additional systems to be processed. It has to be noted however that the way features are seen in the literature on the typicality effect (i.e. descriptive properties), is not the only way features have been conceptualised in the literature. In the following section, we briefly review another class of models that have influenced the literature on semantic representations by organising them in terms of binary features.

1.3 Feature-based models

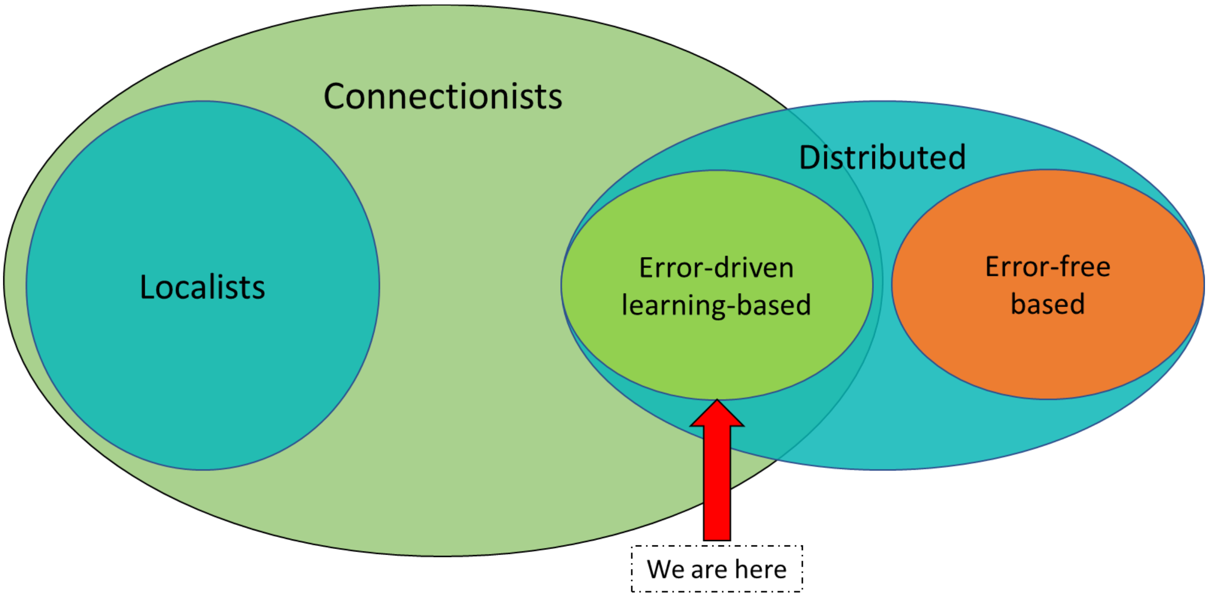
In the same decades that saw the creation of the two previously presented theories, Smith and collaborators (1974) conceptualised the first feature-based model (see also the contrast model from Tversky, 1977). Like all feature-based models, it proposed that each semantic representation can be represented as a distributed set of binary features that it may or may not possess. In other words, instead of describing a semantic representation only in terms of its possessed features as in the prototype theory, semantic representations are associated with the features they possess *and* do not possess (i.e. the semantic representation possesses/does not possess the feature). For instance, the representation of 'bird' is described as possessing the feature <has wings>, and not possessing the feature <has teeth>. In these models, the number of overlapping possessed features between two semantic representations thus determines the degree of relatedness between them. While these models have been criticised for their lack of representation of associative relations (Collins & Loftus, 1975; Kumar et al., 2022), the later versions have still been shown to successfully imitate human data in semantic priming paradigms, including for the typicality effect (McRae, 2004). However, one limitation of these models that is common with the spreading activation account, lies in their inability to explain the way humans acquire knowledge and extract features from their environment. Indeed, since the implementation of these models relies on norms obtained from human data, they can only represent the semantic representations at a given time and are not able to model how the relations between semantic representations evolve through learning. As the aim of this thesis is to determine the way semantic representations can change through exposure to dialogue settings, the following section is dedicated to the presentation of models that have been designed to understand the process of learning new semantic representations and their relations in the global structure of semantic memory. This approach is called the connectionist approach. While the network structure used in this approach originates from the spreading activation theory, its distributed structure resembles that proposed in the feature-based models.

2.2 Connectionist models of semantic memory

2.2.1 Terminology

The connectionist approach operationalises cognitive functions such as semantic memory in terms of parallel processing of networks of interconnected units (Rumelhart & McClelland, 1986). In this approach, models are able to represent learning processes by strengthening or weakening the interconnections within a network. While different types of connectionist models exist in the literature (see Figure 6), in the present thesis we focus only on the distributed 'error-driven learning-based' models, as those are the type of models often used to account for the adaptation in the organisation of semantic representations (e.g. Rodd et al., 2004; see Figure 6). The term 'distributed' refers to the fact that these models are based on the extraction of co-occurrence frequencies from large language corpora and operationalise semantic representations as distributed patterns of activation in the network, instead of representing them as nodes as in the spreading activation theory (Collins & Loftus, 1975). Interestingly, these distributed models are called 'error-driven learning-based' because they represent the learning process of semantic representations through the adjustment of the strength of connections between units (Rumelhard et al., 1986). Specifically, these models suggest that the update of the interconnections between units allows the system to 'learn' from its prediction errors. That is, the adjustments of interconnections are driven by the comparison between the output produced by the network and the predicted output. In the rest of this chapter, we will use the term 'connectionist models' to refer to error-driven learning-based distributional

Figure 6. *Schematic representation of the overlap between connectionist models and distributional models. The red arrow points toward the models we will be speaking about in the following section*



semantic models to simplify our explanations, although it has to be noted that these models can be referred to in other ways in the literature (e.g. 'error-driven learning models'; 'hybrid distributional models' as termed in Kumar, 2021).

2.2.2 Principles of connectionist models

Semantic networks are critical for describing the organisation of semantic representation, as they allow for quantitatively estimating the relation between semantic representations by calculating distances within the network. However, the fact that they are 'hand-built' makes them unable to represent how the organisation of semantic representations can evolve through time and exposure to the environment. On the other hand, feature-based models propose an interesting account for semantic representations based on a distributed representation of features. However, the same criticism could be made of them as they cannot 'learn'. To our knowledge, connectionist models are therefore the only class of models that attempt to bridge this gap.

Connectionist models characterise semantic representations as patterns of activation within a network of interconnected units (McNamara, 2005). In these models, the pattern of activation of a semantic representation accounts for the various aspects that characterise it, although each unit does not necessarily represent a nameable feature (Jones et al., 2006). As such, related semantic representations correspond to similar patterns of unit activity. Interestingly, all the units within a given pattern are not equally activated, but the amount of activation of each unit depends on the previous experience with a particular semantic representation. Connectionist models thus adapt the patterns of activation of a semantic representation by changing both the organisation of the units and the amount of activation given to each unit based on repeated exposure to the environment. This mechanism is one of the most important features of these models and is what makes them learnable. More precisely, in this approach past activations of patterns are not thought to be stored in semantic memory, but to leave a trace in the semantic network so that the connections between units adapt their weights after having been activated (McClelland & Cleeremans, 2009). The updated weights are then thought to influence the following activations of these units. The more a connection gains weight, the more activation will be propagated from one unit to the other within that connection. Therefore, connectionist models are continuously updating the organisation of interconnections between units through exposure to the environment, which has an impact on both the internal structure of semantic representations and the interactions between them.

The way in which connectionist models work suggests two important features of the semantic network: first, the access to a given semantic representation may depend on the weights of its interconnections within the semantic network; second, the organisation of the interconnections can be adapted by our exposure to linguistic inputs. However, while many studies on humans have explored the organisation of semantic representations, only a few of them have attempted to provide behavioural evidence that the organisation of the semantic network might indeed be shaped by everyday experiences, and even fewer have been carried out on adult participants. The aim of the following section is thus to provide a state-of-the-art on the adaptation of semantic representations.

2. Adaptation of semantic representations: state-of-the-art on behavioural data

The adaptation of semantic representations can be studied in two way: either by creating new representations, or by updating the organisation of pre-existing ones. As these two types of adaptations are examined in two separated sets of studies in the experimental part of this thesis, we present the two types of literatures in separated sections, even though they are highly related.

2.1 How are semantic representations adapted?

Just like children, adults are able to learn new meanings to adapt to changes in their environment. For instance, when using social networks became a daily activity, new meanings for known words such as ‘follow’ or ‘story’ were added to the lexico-semantic network of most social media users. These examples illustrate what can be taken as an adaptation of the lexico-semantic representations. One may argue that learning a new meaning provides a new pattern of activation into the lexico-semantic network without adapting the already existing representations. However, in light of the current view on how meanings are represented in the semantic network (i.e. distributed patterns of features that overlap depending on the semantic proximity in the network), this proposition seems unlikely. Instead, new lexico-semantic representations are more likely to interact with pre-existing ones in a way that changes the organisation of the lexico-semantic network. Evidences of such changes is presented in the following sections to show the adaptability of the lexico-semantic network.

2.1.1 Creation of new meanings

The study of ambiguous words is an interesting approach as it allows to manipulate different degrees of semantic relations without having to add new lexical entries. For a given lexical form, ambiguous words possess multiple semantic representations that vary in their degree of relatedness and their frequency of use. Specifically, ambiguous words may have unrelated meanings and be called homonyms (e.g. 'nail' as a tool or as a body part), or related meanings and be called polysemous words (e.g. 'fly' as an insect or as a way of travelling in a vehicle; Rodd et al., 2004). Manipulating these parameters can shed light on the processes that occur when the lexico-semantic network is subject to adaptation. The first study on that subject was the study by Rodd et al. (2012) which examined the impact of new meanings on already existing words (i.e. words that already possess a lexico-semantic mapping). In their study, they sought to determine whether a new recently learned meaning associated with a known word could be sufficiently integrated into the lexico-semantic network to compete with the original meaning of the word. To answer this question, they had participants learning new meanings for existing words by reading short paragraphs in which a new definition was given to an existing, non-ambiguous word. During the learning phase, they manipulated the relatedness of the new meaning in relation to the original meaning of the word. Their goal was indeed to determine whether new meanings that are related to the original meaning are easier to learn than new unrelated meanings (i.e. relatedness effect), and thus to conclude on the impact of new lexico-semantic representations on the adaptation of the whole network. This expected relatedness effect was thought to support the connectionist model proposed by Rodd et al. (2004).

Like any other connectionist model, the model by Rodd et al. (2004) postulated that the connections within a network of interconnected units were subject to an updating process of the weights of connections. It also described supplementary mechanisms to account for the difference between the time needed to process polysemous words and homonyms. According to this model, a word with multiple related meanings is represented with multiple overlapping patterns of semantic representations in the semantic network. These overlapping patterns implies that the mapping between the lexical form and the semantic pattern is reinforced by the proximity between meanings. In comparison, words with multiple non-related meanings are associated with multiple distinctive patterns of activation in the semantic network. The time needed to reach a stable state of activation for one of these patterns is therefore thought to be longer than with related meanings.

Coming back to the study by Rodd et al. (2012), their results showed a significant relatedness effect when participants had to explicitly recall the properties of the new meanings, with related meanings being significantly better recalled than unrelated ones. This result was found with both a short and a long training phase (i.e. one single training session vs. seven days of training) when assessing the new meanings with a cued recall test in which participants are explicitly asked to provide the semantic properties of the new meaning. However, when investigating the impact of learning with a lexical decision task, no significant relatedness effect was found except in the experiment in which the training was semantically engaging (e.g. writing sentences or stories using the new meaning) and sustained for four days. The difference between the results found in these two tasks nicely reflect the distinct processes occurring in the lexico-semantic network, depending on the task. Indeed, the results suggest that when participants are asked to strategically focus on the newly learnt meaning to produce an answer (i.e. cued-recall test), the new meaning is sufficiently integrated into the lexico-semantic network after only one learning phase to interact with pre-existing knowledge. However, when participants are only required to access the word (i.e. lexical decision task), which constitutes a faster and less strategic process, more semantic engagement is required during the learning phase to sufficiently anchor the new meaning into the lexico-semantic network. Overall, the results of this study are consistent with the connectionist model by Rodd et al. (2004) which suggested a disadvantage in the processing of words with multiple unrelated meanings in comparison with words with multiple related meanings.

Other studies that used similar procedures for learning new meanings for already existing words investigated how this integration into the network could impact subsequent processes of the already existing meaning (Fang & Perfetti, 2017; Fang & Perfetti, 2019; Maciejewski et al., 2020). Overall, the two studies by Fang and Perfetti (2017; 2019) showed that new semantic representations acquired through only one learning session interacted with the prior semantic representations of the words. Specifically, a momentary perturbation of the processing of the original meaning was observed as soon as the new meaning was added to the lexico-semantic network. This perturbation was observed at both the behavioural and neurophysiological levels, with longer decision times in a semantic judgement task and a larger negativity (i.e. N400) at the central midline cluster (Cz) for words with new meanings in comparison with control words (i.e. words without new meanings; Fang & Perfetti, 2017; 2019). In addition, in the 2019 study, they found slightly different patterns of results depending on word frequency, with the more frequent words being more negatively impacted by the new meaning immediately after the learning phase than the less frequent ones. The authors interpreted this effect as the consequence of the stronger connections between the word and its original meaning in the lexico-semantic network. Specifically, high frequency words are more accessible for reactivation and co-activation with new meanings, and this stronger co-activation may imply the need for stronger suppression of one of the two meanings in later access to the word.

In a follow-up study, Maciejewski et al. (2020) made their participants learn new meanings using the same procedure as the semantically engaging learning phase used by Rodd et al. (2012; e.g. sentence writing, story writing). In addition, they used a relatedness decision task in order to assess the impact of the new meaning on the processing of the original one, according to the degree of relatedness between both. In this task, participants had to decide whether a prime and a target word were related in meaning or not. Overall, they replicated both the perturbation effect and the relatedness advantage previously evidenced (Rodd et al., 2012; Fang & Perfetti, 2017; 2019), with new unrelated meanings producing stronger perturbation effect on the processing of the original meanings than new related meanings. Unlike the study by Rodd et al. (2012) however, even the related meanings produced a perturbation of the processing of the original meaning. Finally, Hulme (2018) questioned the difference of impact between incidental (i.e. unintended) and intentional (i.e. explicit) learning of a new meaning. Although they found substantial differences between participants, incidental learning led to a satisfactory level of learning of new meanings (38.5%), and this learning was still present after one week (Experiment 1). However, the intentional and more explicit learning proved to be a better learning medium for new meanings than incidental learning (experiment 2).

In the light of what has been previously presented, it seems that the organisation of lexico-semantic representations is able to change according to our recent exposure to linguistic inputs, at least in order to include new lexico-semantic representations and make them interact with the pre-existing ones. However, one may wonder whether the organisation of pre-existing lexico-semantic representations can change, without the need of additional new knowledge. For instance, we may wonder whether it is possible to change the weights of the connections between two pre-existing lexico-semantic representations. This question is addressed in the next section of this work.

2.1.2 Adaptation of the organisation of already existing meanings

In the previous section, we presented experiments in which the competition between two meanings –the original meaning of a word and a new one– was examined in order to assess the integration of new meanings into the lexico-semantic network (i.e. creation of new connections between the new meaning and the pre-existing network). Other studies however have been led to determine whether the competition between two *pre-existing* meanings of the same word could change with time and exposure to any of the meanings (i.e. adaptation of the weight of pre-existing connections). To do so, these studies rely on the examination of polysemous words. Most of the time, the meanings of polysemous words are not equivalent in terms of frequency of use. According to the model of Rodd et al. (2004), the more frequent the meaning, the more weight is attributed to its connection with its corresponding lexical entry. Conversely, a less frequent meaning is associated with a reduced connection weight. These are generally called the 'dominant meaning' and the 'subordinate meaning', respectively and impact the ease of access and retrieval to the meaning of a word.

In 2013, Rodd and collaborators questioned the lexico-semantic network's ability to adapt the organisation of pre-existing lexico-semantic representations of polysemous words depending on recent exposure, by changing the weights of connections between subordinate and dominant meanings. For instance, they wanted to test whether the subordinate meaning of the word 'fan' (i.e. supporter) could become less subordinate by strengthening its connection with its associated word form after a single exposure. They used a paradigm that they called the 'word-meaning priming paradigm', in which participants are first instructed to listen to sentences in which polysemous words are associated with their subordinate meanings (i.e.

priming phase; e.g. "the footballers were greeted by their adoring fans"). Then, the influence of this priming phase on the preferred meaning is tested using a word-association task, in which a target word is presented to participants who are then asked to type the first word that comes to their mind and is related to the word they have just heard. This paradigm allowed the authors to find a significant priming effect with words biased towards their subordinate meaning, eliciting more words associated with the subordinate meaning than the unprimed words. This effect was controlled for a possible episodic trace in a follow-up experiment (Experiment 2), in which the voice pronouncing the words changed between the priming phase and the test phase. This change did not produce any impact on the significant priming effect. The authors thus suggested that each encounter with an ambiguous word strengthens the connections between the word form and the meaning that is accessed during the priming phase, thus creating slight changes in the organisation of the pre-existing lexico-semantic representations.

This result was later replicated and extended by Rodd and collaborators (2016). The authors indeed demonstrated that the word-meaning priming effect could be found in both laboratory and ecological conditions, with participants being exposed to the subordinate meanings either by listening to a radio programme or during a lab-based experiment. In their third and fourth experiments, the authors also addressed the question of long-term exposure (i.e. repeated exposure over months or years of experience with the word-meaning mapping) to subordinate meanings on the organisation of lexico-semantic representations. To do so, Rodd et al. (2016) compared a group of rowers with a group of non-rowers on the meaning of words with subordinate meanings in the field of rowing. The impact of short- vs. long-term rowing experience and the role of recent exposure to the words (i.e. people who had practised on the day of the experiment or one day before) were also measured. As expected, long-term experience with the subordinate meanings produced more answers biased towards that meaning, in comparison with the non-rowing group (Experiment 3). More recent rowing practice (up to two days before the testing phase) also helped participants to produce significantly more answers related to the subordinate meaning than participants who had not rowed in the two days before the testing phase.

Altogether, these results suggest a cumulative effect of both immediate and long-term exposure to subordinate meanings on the organisation of lexico-semantic representations within the lexico-semantic network. More precisely, it is suggested that long-term lexico-semantic adaptations come from incremental changes in connection weights which occur after each new encounter with a word meaning mapping. However, this interpretation has to be qualified in light of the results obtained by Betts et al. (2018), who explored the impact of multiple recent encounters with word meanings. In three experiments, they compared the impact of one encounter with the subordinate meaning, with three encounters either massed or spaced out in time. While the one-encounter condition produced the effect already found in the literature (Rodd et al., 2013), the three encounters were found to enhance the meaning-priming effect, but only when these encounters were spaced out in time. This result confirms the cumulative effect of multiple encounters on the bias towards subordinate meanings, although this effect is subject to limitations such as the time needed between each presentation of the primed meaning to impact the connection weights within the network.

Overall, these findings are in line with the idea that pre-existing lexico-semantic representations can be shaped by the accumulation of encounters with linguistic input, even though these encounters have to be spaced over time to accumulate within the lexico-semantic network. In the following section, we examine how such adaptation works with the graded structure of semantic categories.

2.1.3 Adaptation of the organisation of semantic categories

In the previous section, we saw that long-term exposure to linguistic input pertaining to a particular domain of expertise can change the weights of connections between a word and its meanings. However, such effect of the expertise has also been evidenced in terms of how semantic categories are organised. Some studies investigated the way typicality can be affected by expertise or even by cultural differences. For instance, in the study by Lynch et al. (2000) people with different levels of expertise were asked to rate the "goodness of example" of trees (i.e. the extent to which a tree is representative of its category). While novices based their typicality judgement on familiarity, experts (i.e. taxonomists, landscapers and park maintenance personnel) based their judgements on ideals, which represent the features that a category member should possess to best serve its function (Barsalou, 1985). While the original proposal of category organization was based on perceptual features and seemed to be the preferred organisation for non experts, the organisation of semantic categories was improved by years of expertise and seems best accounted for by functional properties. These results were later replicated and extended to cultural effects (which may be considered as a particular kind of expertise; Bailenson et al., 2002), and were interpreted as a gradual alteration of long-term memory that occurs over long periods of time and affects the way semantic representations are retrieved according to the context (Dieciuc & Folstein, 2019).

While daily experiences can shape the semantic categories stored in long-term memory, it has also been shown that typicality has an impact on the creation of connections between lexico-semantic representations. In particular, studies that explored the ability of people with aphasia to improve their naming skills noted the importance of the graded structure of semantic categories. Aphasic patients often suffer from anomia, which is the difficulty to retrieve the word associated with a given semantic representation, despite the fact that this word-meaning mapping was accessible before the brain injury. One of the methods that is used to help these patients recover is to ask them to generate the semantic features of the target word (Gilmore et al., 2018; Kiran, 2008). By doing so, the connection between the word and its distributed semantic representation is strengthened. In line with the spreading activation theory (Collins & Loftus, 1975), the activation of the semantic representation results in the activation of other related representations, which in turn spread the activation towards their associated lexical representation. This is how patients generalise the learning to other members of the same semantic category, without the need to learn the word-meaning mapping of every member. Interestingly, several studies on these patients showed that training on atypical items was more beneficial for the generalisation process than training on typical items (Gilmore et al., 2018; Kiran, 2008; Kiran et al., 2011; Kiran & Johnson, 2008; Kiran & Thompson, 2003b). In particular, training on the features of atypical items led to greater generalisation to untrained within-category typical items. These results confirm the idea that atypical words provide more information about the structure of their semantic category, owing to their high degree of feature variation (both core features and distinctive ones), which consequently produces greater generalisation (Plaut, 1996). Crucially, it has been shown that among the various features possessed by atypical items, people seem to know more about their unique features than shared features with other category members (Malt and Smith 1982).

Altogether, these studies support the idea that atypical items are related with a more complex structure of features, both shared and unique, but that shared features are less salient than unique features within the semantic representation of an atypical item. In addition, studies on aphasic patients provide evidence that confirm the ability for the category structure of the semantic representations to be shaped by repeated encounters with linguistic input. However, unlike the adaptation of weights of meanings for polysemous words, these adaptations of the category structure have been demonstrated only in the very long term (i.e. several weeks of treatment), preventing us from drawing strong conclusions about the time course of these changes. Indeed, while Rodd and collaborators (2016) attempted to understand the time-course of the changes in the organisation of lexico-semantic representations, the underlying mechanisms of integration into the lexico-semantic network are still under debate. Although one of these mechanisms, called consolidation, has been known for a long time as a way of integrating new information into long-term memory, its role in the adaptation of already existing knowledge remains elusive. In the next section, we try to explain the current view on the involvement of consolidation in the adaptation of lexico-semantic representations.

2.2 The impact of consolidation on the adaptation of lexico-semantic representations

2.2.1 What is consolidation?

The concept of consolidation is not a recent idea, since it was first evidenced in healthy adults in 1900 by Mueller and Pilzecker (cited by Lechner et al., 1999) and has since been the object of many studies. In current research, consolidation is generally defined as the process of memory formation and strengthening that allows for newly encoded representations to become more stable long-term representations (Dumay & Gaskell, 2012; Klinzing et al., 2019). One of the key mechanisms that has been evidenced in the consolidation process is the importance of sleep episodes. Although this idea was first debated in the early 20th century, it is now generally acknowledged that sleep plays a crucial role in the learning of new linguistic content (Dumay & Gaskell, 2007; Tamminen, 2010; also see Fenn et al., 2003 for phonological aspects; Gaskell & Dumay, 2003 for lexical aspects; Gomez et al., 2006 for syntactical aspects; Mestres-Missé et al., 2007 for semantic aspects). There have been several hypotheses and models as to how memory consolidation benefits from sleep episodes. Interestingly, there has been evidence that the hippocampus plays a major role in the consolidation process (Eichenbaum, 2000; Squire & Alvarez, 1995), especially in linguistic learning (Davis & Gaskell, 2009). The hippocampus is traditionally believed to serve as a temporary mediator for new word-meaning mappings when learning new words. After a short period of time, the new linguistic information is consolidated into neocortical structures giving rise to a new lexico-semantic representation that is independent of the hippocampus (McClelland et al., 1995). However, this view has been challenged in recent proposals (Duff & Brown-Schmidt, 2017), to match the neuropsychological data obtained from amnesic patients. Specifically, it is now suggested that the hippocampus contributes not only to the learning of new lexico-semantic representations, but also to the updating of already existing information, especially in terms of strengthening and enrichment of pre-existing lexico-semantic representations (Klooster & Duff, 2015; Klooster et al., 2020). However, the idea that the hippocampus may be involved in the updating of already existing information contrasts with the theory according to which each encounter with the meaning of a word immediately creates a change in the lexico-semantic network and an immediate update of long-term connection weights (Gilbert et al., 2018; Rodd et al., 2016). The aim of the next section is therefore to determine which of these two hypotheses is the most likely to account for behavioural data on the mechanism that underpins both short-term and long-term adaptations of lexico-semantic representations.

3.2.2 Updating lexico-semantic representations: the impact of consolidation

In their original proposition, Rodd and collaborators (2016) suggested that the adaptation of the lexico-semantic network was due to direct modifications of the strength of connection within the semantic units of the semantic representations. However, this interpretation was in opposition with their earlier idea (Rodd et al., 2013) that the word-meaning priming effect emerged from the strengthening of the connection between the semantic representation and the word form (i.e. orthographic or phonologic). In a recent paper, Gilbert et al. (2018) thus attempted to resolve this debate by testing the impact of the priming modality (either oral or written) on the word-meaning priming effect. Indeed, whereas the proposition of Rodd et al. (2013) suggests that the word-meaning priming is modality-specific (i.e. only the modality of priming should be affected by the change in connection strength), the contrary is proposed by Rodd et al. (2016) (i.e. no modality-specificity of the word-meaning priming). The results by Gilbert et al. (2018) supported the latter hypothesis, suggesting that changes may directly occur within the semantic layer of the network. Interestingly, the authors also proposed two alternative interpretations to their results. The first one was the possibility that the activation of the mapping between the phonological form of the word and its semantic representation produces a simultaneous co-activation of the orthographic form of the word (and vice versa), thus preventing the word-meaning priming effect from being modality-specific. The second alternative is that instead of having only one system shaped by everyday experiences with words and their meanings, there could be two complementary systems. Specifically, it is posited that new leanings could be mediated by the hippocampal system instead of being directly integrated into the cortical (and more stable) system of representations, to protect the existing network from interference (Davis & Gaskell, 2009). This is in line with the proposition by Duff and Brown-Schmidt (2017) and may also account for the data reported by Rodd et al. (2016), which indicates a gradual decay in the strength of the word-meaning priming effect. Such an effect found after one minute was already reduced after a 20-minute delay, and even more so after 40 minutes. This progressive decay in the strength of the priming effect was supposed to be due to potential interference from intervening linguistic input between the prime and the target and could, as such, be the result of a change into a less stable structure than the originally proposed one.

To depart from these two views, Gaskell et al. (2019) designed two experiments that examined the consolidation of lexico-semantic adaptations in more detail. They used the word-meaning priming paradigm with homophones only (i.e. words with two semantically related meanings). In the first experiment, they manipulated the delay between the prime phase and the test (2h or 12h) and the state of the participants during this delay (awake or asleep). In the second experiment, the potential interference effect of subsequent linguistic input during the wake period was tested by priming participants either in the morning (with additional interference during the day before sleeping) or in the evening (a few moments before sleeping). Their aim was to depart from the debate between the immediate account, which states that new experience alters long-term cortical connections (Rodd et al., 2016), and the contextual binding account which posits that new learnings are mediated by the hippocampus and gradually integrated into long-term lexico-semantic representations during sleep (Duff & Brown-Schmidt, 2017). In the first experiment, the results showed that for both time intervals, participants who benefited from sleep elicited stronger priming effects than participants who stayed awake. Interestingly, in the second experiment the results demonstrated that the previously evidenced advantage of sleep periods was not due to the wake period being subject to interferences but to sleep periods actively promoting consolidation. This result was considered as more in line with the contextual binding account, according to which the priming phase does not alter the weights of long-term memory connections, but instead creates a hippocampal memory that is subject to interference or to offline consolidation through sleep. Duff & Brown-Schmidt (2017) also suggested that upon subsequent encounters with the primed words, people are able to make use of both long-term and hippocampal representations by mixing them in order to find the most appropriate interpretation. Altogether, it seems that the hippocampal memory can affect both short-term experiences with a linguistic input, and long-term experiences.

While this study suggests that new linguistic inputs are mediated by the hippocampus before impacting long-term memory connections, this has to date only been tested by means of a word-meaning priming paradigm. However, as suggested by Gaskell et al. (2019), the effect observed with the word-meaning priming paradigm could be attributable to more than one mechanism which have not all been investigated yet. In other words, the changes observed in the lexico-semantic network before and after consolidation could emerge from multiple mechanisms that take place during the processing of the word and the production of associated words. In order to determine whether the new linguistic inputs may have directly impacted long-term lexico-semantic representations, the assessment phase requires to prevent possible influences of the hippocampal representation on the processing of the word. Therefore, it may be useful to test the impact of the new linguistic input on long-term representations with a task which implicitly tests the access to the word.

In conclusion, we have seen that lexico-semantic representations are flexible, either in terms of connection strength or in terms of organisation in the network, and that they evolve with our linguistic experiences. This adaptation mechanism is observable both in the short term, to follow the immediate needs for specific lexico-semantic representations, and in the longer term through the consolidation process, so that our linguistic behaviours remain adapted to our environment. However, most results obtained in studies assessing the adaptability of lexico-semantic representations may have been affected by processes that do not directly reflect the long-term organisation of these representations. Indeed, while the majority of these studies used assessment tasks involving word production (i.e. word association task or cued recall task), the other tasks explicitly focus on the semantic properties of the words, thereby introducing attentional processes. The only known exception to this trend is the use of a lexical decision task in the experiments by Rodd et al. (2012). Crucially, these experiments suggested that without an extensive training on the new linguistic input, a task that implicitly assesses word access is not capable of showing changes in lexico-semantic network. However, these studies only attempted to change lexico-semantic representations through sentence comprehension tasks. Therefore, we can wonder to what extent dialogue, which has been suggested to produce activations in long-term linguistic representations due to the alignment mechanism (Pickering & Garrod, 2004), can impact the organisation of lexico-semantic representations such that their access remains subsequently changed. Specifically, the aim of this thesis is to assess whether the access to lexico-semantic representations may be changed after a dialogue in both the short and long term. In the next chapter, we specify the objectives of this thesis as well as our working hypothesis, before proceeding with the presentation of the experimental chapters that follow.

Chapter 3 – Aims and Hypotheses

1. Theoretical summary and aims of the thesis

During dialogue, people tend to adjust their linguistic behaviour to their partner to ensure mutual understanding. Such adjustments have been explored on a behavioural scale in several approaches including the collaborative approach, which interpreted them as a consequence of the need for coordination to make the dialogue successful (Clark, 1996; Clark & Brennan, 1991; Garrod & Anderson, 1987). Although some authors in the collaborative approach have highlighted the role of memory in dialogue coordination (Heller & Brown-Schmidt, 2023; Horton & Gerrig, 2005a), the only theory to date to suggest a relationship between dialogue and semantic representations is the interactive alignment theory (Pickering & Garrod, 2004, 2021). This theory hypothesises that the linguistic representations stored in semantic memory, including the semantic representations, become aligned between dialogue partners during the interaction. This alignment mechanism constitutes what distinguishes dialogue from other situations of isolated sentence comprehension. However, the theory does not explicitly question the extent to which the activation provoked by the alignment during dialogue can leave a trace in semantic memory so that semantic representations remain changed immediately after dialogue. We thus wanted to investigate whether the semantic representation of words, hereafter referred to as 'lexico-semantic' representations, could be changed immediately after a dialogue.

To determine if dialogue has an impact on those lexico-semantic representations, we looked at the literature on the organisation of lexico-semantic representations stored in long-term memory. We first looked at different models of organisation of lexico-semantic representations in long-term memory (Collins & Loftus, 1975; Jones et al., 2006; Rosch, 1973). Lexico-semantic representations have been suggested to be organised as a highly interconnected network (Collins & Loftus, 1975), itself organised in categories within which lexico-semantic representations follow a graded structure in terms of their typicality (Rosch, 1973; 1975). More recently, it has been hypothesised that both the relationships between lexico-semantic representations within the network and their category structure could change following exposure to language inputs (Lynch et al., 2000; Rodd et al., 2012; 2013). In the literature on word-meaning priming, we found evidence that linguistic information could affect lexico-semantic representations so that subsequent encounters with a word would be impacted, regardless of the context (Rodd et al., 2013). Finally, it has been proposed that changes in lexico-semantic representations could be, at least in part, first stored as hippocampal representations to be later consolidated during a sleep period (Gaskell et al., 2019; Rodd et al., 2016). Specifically, new encounters with a word-meaning mapping are thought to affect the weight of the already existing connections between lexico-semantic representations after a night of consolidation, thus changing the organisation of the lexico-semantic network.

Taken together, the literature on dialogue and lexico-semantic representations points toward the idea that activations in the long-term lexico-semantic network produced during a dialogue could impact its organisation such that the access to these representations remains changed after dialogue. However, this idea has never been tested. The objective of this thesis is thus to fill this gap. In particular, the literature on the organisation of lexico-semantic representations allows us to imagine two ways in which these representations could be impacted by dialogue:

(1) in line with the literature on word-meaning priming showing the flexibility of the connections between lexico-semantic representations (Rodd et al., 2012; 2013), we can imagine that new connections can be created between two pre-existing representations during a

dialogue. As a consequence, accessing one of the representations after dialogue should pre-activate the other and make it more accessible;

(2) in line with the literature showing the flexibility of the category structure of lexico-semantic representations (Lynch et al., 2000), we can imagine that the connections between an atypical item and its superordinate semantic category can be strengthened during a dialogue, such that the atypical item becomes less atypical after dialogue. As a consequence, before dialogue, accessing an atypical member of a semantic category should not pre-activate typical members of the same semantic category. However, after dialogue, accessing one of the discussed atypical members should pre-activate other typical members of the same semantic category and make them more accessible.

Finally, since the literature suggests that activations triggered by dialogue can impact access to the lexico-semantic representations both immediately afterward (Pickering & Garrod, 2021) and after a night of sleep promoting offline consolidation (Gaskell et al., 2019), each type of adaptation (i.e. objective 1 and 2) will be tested at these two time points (i.e. immediately after dialogue and one day later).

In order to test these hypotheses, we conducted two sets of experiments in which participants took part in two newly created dialogue tasks in which they were asked to speak about pictures (i.e. one dialogue task and one type of picture in each set of experiments). The organisation of the lexico-semantic representations of the participants was then assessed, either immediately after dialogue or one day later. In the following section, we present an overview of the experimental chapters in which these sets of studies are presented.

2. Overview of experiments and hypotheses

In our first set of experiments, we explored the possibility that dialogue might change the organisation of the lexico-semantic network by testing the possibility to create new lexico-semantic relationships between already existing lexico-semantic representations stored in the semantic memory. To do so, we created a new dialogue task in which abstract visual representations could be associated with two semantically unrelated words. We began by building a database of (abstract) tangram pictures and their associated names in order to find these visual representations that can be associated with two semantically unrelated labels. The database also provides information about the name agreement for each picture in order to ensure the plausibility of the chosen labels. In this thesis, we dedicate Chapter 4 to the presentation of the database. In Chapter 5, we consequently present the first set of experiments specifically designed to test the first hypothesis and assess the creation of *new* lexico-semantic relationships between pre-existing lexico-semantic representations after a dialogue. To do so, we examined whether or not a new relationship can be created between two words which are repeatedly associated with a common visual representation during a dialogue. Finally, Chapter 6 presents a second set of experiments designed to test the second hypothesis and determine whether dialogue can change the organisation of the lexico-semantic network, *without* creating new relationships. In particular, these experiments were designed to test whether it is possible or not to strengthen the already existing relationship between an atypical word and its superordinate semantic category. The detail of these chapters is provided in the following sections.

2.1 Overview of Chapter 4

Chapter 4 is dedicated to the presentation of our database of 332 tangram pictures (Experiment 1). This database provides most of the variables classically used in studies involving picture-naming tasks, in order to standardise the stimuli used in studies interested in using tangram pictures (i.e. image agreement, familiarity, visual complexity, image variability, and age of acquisition and concreteness; Alario, & Ferrand, 1999, Bonin et al., 2003; in Spanish, Dunabeitia et al., 2018, Manoiloff et al., 2010; in Greek, Dimitropoulou et al., 2009; in Russian, Tsaparina et al., 2011; in Persian, Ghasisin et al., 2014). The aim of creating this database was for us to be able to select the most appropriate stimuli for our first set of experiments. The key criterion for the picture selection was to find tangram pictures that could be labelled with two semantically unrelated yet plausible names. The question was therefore to determine whether tangram pictures could be associated with multiple consensual labels. To address this, analyses were performed on participants' labels for the pictures to obtain two different measures of the name agreement. Since the following set of experiments required to find pictures that can be easily seen as possessing two plausible labels, the measures of name agreement allowed us to select pictures with a satisfactory amount of name agreement. Additionally, the name agreement measures are discussed in relation to the literature on line-drawing pictures, to highlight the similarities and differences between the two types of stimuli and to demonstrate the relevance of tangram pictures in the experiments presented in Chapter 5.

It should be noted that due to the strict criterion of finding pictures with two plausible names, we were unable to control for all variables in the stimulus selection process for Chapter 5. As a consequence, Chapter 4 presents only the variables that were used to select the stimuli of our first set of experiments to focus on the purpose of the thesis (n.b., the article presenting the full database can be found in Appendix A).

2.2 Overview of Chapter 5

The obtention of standardised stimuli allowed us to run our first empirical set of experiments on the possibility to create new lexico-semantic relationships thanks to a dialogue setting. We created a new dialogue task called the 'Interactive Agreement Referential task' (IAR task). In this task, a naive participant interacts with a confederate, whose role is to propose two semantically unrelated names for a given tangram picture (e.g. 'tap' and 'dinosaur') and to repeat them together four times during the interaction. Crucially, the order in which the two words are given to name their pictures is alternated across the four repetitions in order to maximise the abstraction of the association and minimise the impact of an episodic learning (e.g. 'tap' is cited first in the first and third repetition while 'dinosaur' is comes first in the second and fourth repetitions). The goal is to create a new abstract lexico-semantic relationship between pairs of pre-existing semantically unrelated words by repeatedly associating them with a common visual representation. We hypothesised that pairs of words referring to a shared visual representation would become more semantically related than pairs of words referring to two distinct visual representations after dialogue. This hypothesis led to the creation of three different experiments.

In Experiments 2 and 3, we used an implicit task to measure the changes in lexico-semantic representations induced by the IAR task. We used a lexical decision task with a semantic priming paradigm to test whether the changes occurring during dialogue would create a new lexico-semantic relationship that is strong enough to propagate the activation from one lexico-semantic representation to the other. While the lexical decision task was performed immediately after dialogue in Experiment 2, it was performed after one night of sleep in Experiment 3. Therefore, the aim of Experiment 1 was to determine whether the impact of dialogue on the access to lexico-semantic representations would be observable immediately after the interaction, while Experiment 3 aimed at testing whether this impact was observable after one night of sleep promoting offline consolidation.

In Experiment 4 we explicitly led the participants to pay attention to the semantic relationship between pairs of words by using a semantic relatedness judgment task one day after dialogue. Thanks to this task, we investigated whether changes in pre-existing lexico-semantic representations through dialogue were observable when attention is focused on the semantic properties of the relationship. The task was only performed one day after dialogue in order to maximise our chances to determine whether lexico-semantic representations might be impacted by dialogue. Indeed, while the changes observed immediately after dialogue are a matter of debate (Duff & Brown-Schmidt, 2017; Gaskell et al., 2019; Pickering & Garrod, 2023), a consensus can be reached on the state of these changes observed after a period of offline consolidation. It has to be noted that in our three experiments, the modality of assessment of changes in lexico-semantic representations was different from the modality of creation of new lexico-semantic relationship (visual vs. oral modality). This was the case in order to ensure that our results could be interpreted in terms of changes in lexico-semantic representations stored in semantic memory, as we do not question episodic memory in the present manuscript. It can be noted that the article related to this first set of experiments (Fasquel et al., under review) can be found in Appendix B.

2.3 Overview of Chapter 6

While the previous chapter evaluated the possibility of creating new lexico-semantic relationships between already existing lexico-semantic representations, Chapter 6 assessed the possibility of adapting the category structure of lexico-semantic representations. In particular, the second set of experiments focused on the typicality effect and the possibility of changing the status of lexico-semantic representations so that atypical members of a semantic category (e.g. ‘penguin’ is an atypical member of the ‘bird’ category) become less atypical after a dialogue. To do so, we used a new dialogue task (different from the one used in Chapter 5) called the ‘Property-based matching task’, in which two naive participants discussed photo-realistic pictures presented on their respective screens to match the order in which they should be placed. Pictures referred either to an atypical or to a typical member of their belonging semantic category. Crucially, participants were asked not to use the names of the pictures to refer to them but sentences that represent the properties of their superordinate semantic category. The aim was to strengthen the lexico-semantic relationship between the atypical members and what constituted the core features of their superordinate semantic category. We hypothesised that atypical members would become less atypical after dialogue than before owing to new strengthened connections between them and their superordinate category. To test our hypothesis, a lexical decision task coupled with a semantic priming paradigm was performed before and after dialogue. The results of these two tasks were compared to assess the changes in the organisation of the already existing lexico-semantic representations. The reinforcement of the relationship between atypical members and the core features of their superordinate semantic category was expected to reduce the semantic distance between them. Although Experiment 5 was conducted to assess the changes in pre-existing lexico-semantic representations immediately after dialogue, Experiment 6 explored their persistence one day after dialogue. As in Chapter 5, the visual modality was chosen to assess the access to the lexico-semantic representations, to make it different from the oral modality of dialogue and ensure that our results could be interpreted in terms of changes in abstract representations rather than in terms of episodic traces.

2.4 Overview of Chapter 7

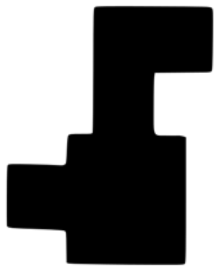
The last chapter of this manuscript is dedicated to the general discussion of the results obtained from experiments, in light of the literature on lexico-semantic representations and dialogue. In this chapter, we discuss the main results of Chapter 5 and 6 in light of the literature on the adaptation of lexico-semantic representations and review the methodological choices made in our experiments to propose adaptations for future studies on the subject. We conclude on the contribution of this work to the literature on dialogue and suggest directions to investigate the impact of dialogue on the adaptation of lexico-semantic representations.

Chapter 4 – Creation of a database for tangram pictures

1. Introduction and overview of Experiment 1

In experimental psychology, stimuli are often constituted of pictures. This is especially true for studies in the field of visual perception, memory and language (Jevtovic et al., 2019; Lopez-Madrona et al., 2022; Seeliger et al., 2018). While most of these fields use photo-realistic pictures or drawings of real objects, the field of dialogue psychology makes very often use of tangram pictures (e.g., Bangerter et al., 2020; Branigan et al., 2011; Brennan & Clark, 1996; Clark & Wilkes-Gibbs, 1986; Duff et al., 2006; Horton & Gerrig, 2002; Knutsen et al., 2018; Knutsen et al., 2019; Knutsen & Le Bigot, 2020; Yoon & Brown- Schmidt, 2014, 2019). Indeed, as tangram pictures are abstract visual representations (i.e. a set of geometrical figures assembled into one representation; see Figure 7), they may be perceived in various ways, leading people to discuss what the pictures should refer to. Thanks to these pictures, dialogue researchers can focus on the processes that allow dialogue partners to reach mutual understanding. The content of the conversations is then analysed to infer the nature of the processes involved in dialogue. However, the way people may discuss these pictures could be affected by their characteristics. For instance, picture's familiarity (i.e. the extent to which people are regularly in contact with the picture; Alario & Ferrand, 1999; Bonin et al., 2003; Ghasisin et al., 2014; Tsaparina et al., 2011) or discriminability (i.e. the extent to which a picture can easily be discriminated from others; Hupet et al., 1991) could play a role in the ease with which participants could find an agreement on how to name the pictures. One index that can reflect the ease with which interlocutors will be able to find an agreement on pictures' names during a dialogue is the name agreement. When building a database for pictures, name agreement reflects the extent to which the same name is used by all participants to refer to a

Figure 7. *Example of a tangram picture.*



given picture (Snodgrass & Vanderwart, 1980). As such, the higher is the name agreement in a database, the lower are the chances of disagreement during a dialogue focused on naming that picture. In sum, knowing the preferred name for tangram pictures and to what extent people agree on that name can greatly help researchers to select stimuli adapted to the dialogue setting they want to create to answer specific questions. However, there is to date no database specifically created to provide norms for tangram pictures. Thus, the aim of the present study was to fill this gap in the literature by providing a database for tangram pictures and their associated names. Specifically, the present chapter presents the first norms for a set of 332 tangram pictures and their possible names. Importantly, name agreement was calculated based on all the names provided in the database. The aim is twofold: first, to determine whether tangram pictures can be associated with two semantically unrelated labels that are sufficiently consensual to be used in our first set of experiments, and second, to use this measure as a tool to select the most appropriate pictures. At the end of this chapter, we also provide a discussion of the similarities and differences between name agreement for tangram pictures and for line drawing pictures (Alario & Ferrand, 1999; Bonin et al., 2003; Dimitropoulou et al., 2009; Ghasisin et al., 2014; Manoiloff et al., 2010; Snodgrass & Vanderwart,1980; Tsaparina et al., 2011). This comparison is performed to evidence that tangram pictures are the most suitable stimuli for our first set of experiments as we seek for pictures that can be seen in very different ways but are associated with a reasonable amount of agreement on their labels (see chapter 5).

The following pages present the methodology used to collect the data for the database as it is presented in Fasquel et al. (2023). It has to be noted first that all authors have given their consent for the methodology of the article to be presented in the present thesis. For the rest of this manuscript, all sections extracted from the article presented in Appendix A will be flanked by the symbols '╔' and '╗'. Second, as noted earlier in chapter 3, the whole database contains more variables than the one discussed in this chapter (i.e. image agreement, familiarity, visual complexity, image variability, concreteness and age of acquisition). However, these variables were not used in the selection of stimuli for the set of experiments presented in chapter 5 (see chapter 3 for justifications) and were thus not discussed in light of the research question of this thesis. This is why we choose to present and discuss only the results that concern the name agreement, which is the only variable extracted from the database that was used to select our stimuli. The entire article can however be found in Appendix A and the material can be found on the following OSF link: https://osf.io/bxkpa/?view_only=2fe0acd124e64df19a0195354ebe45b4.

╔ 2. Method

2.1 Participants

One hundred and ninety-three native French speakers took part in the study. They received course credit or monetary compensation (€20) for their participation. All participants had normal or corrected-to-normal vision and no history of language disorders. After their participation, 20 participants were removed from the data, owing to either an overly long or an overly short time of participation (longer than three hours or less than one hour). This was done to ensure that all participants performed the task correctly and dutifully without taking too much time or going too fast on the questionnaires. Following the same principle, two participants were removed from the analysis owing to a rate of similar responses greater than 10%. Two other participants were removed from the database to balance the number of participants across groups (see Materials for details). Therefore, data from 169 participants were examined (133 female and 36 male, 18–29 years old, M = 20.17; *SD* = 1.47). Before the beginning of the experiment, they were informed about the goal and duration of the study. They also validated an online written consent form which followed the Declaration of Helsinki.

2.2 Materials

2.2.1 Tangram pictures

We started by collecting a total of 375 monochrome (black) tangram pictures from booklets found in various tangram games. As specified below, not all 375 pictures were necessarily included in the final database. As in most studies involving tangram pictures (e.g., Knutsen et al., 2018), all the pictures were made of one square, two big triangles, two small triangles, one medium triangle, and one parallelogram. The pictures were then scanned and randomly divided into two sets, hereafter referred to as set A, which contained 187 pictures, and set B, which contained 188 pictures. All pictures were then uploaded to the online survey platform LimeSurvey (version 2.6; LimeSurvey GmbH, n.d.)). The largest side of the picture (length or width) was always 300 pixels long, and the size of the picture was automatically adjusted to maintain the original proportions of each picture.

2.2.2 Questionnaires

The variables examined in this study were split into three sets of questionnaires (hereafter questionnaires 1, 2, and 3). Two different versions of each questionnaire were then created, each corresponding to a different set of pictures (A or B). Six questionnaires (1A, 1B, 2A, 2B, 3A, and 3B) were thus created in total. We divided the pictures into two sets and the questions into three questionnaires to reduce the length of the experiment and to make sure that the collected data were reliable. In each questionnaire, the first question (which was always the same in all questionnaires) asked the participants to state the first word or expression which came to their mind when they saw the picture. That question was thus related to the name agreement variable. Due to the expected diversity in the labels provided by the participants to refer to a given tangram picture, the first question was always related to name agreement and the following questions were divided into different categories (questions related to how the picture may be referred to in dialogue in questionnaires 1A and 1B, the interface between the labels chosen and their visual representation in questionnaires 2A and 2B with the image agreement and image variability questions, and visual and conceptual properties of the pictures in questionnaires 3A and 3B).

Following the name agreement question, in questionnaires 1A and 1B, participants were asked to say whether they would use the label they had provided to describe the picture during a dialogue with another person, and if not, which label they would prefer to use. The purpose of this question was to determine how likely the labels provided by the participants were to be used in a dialogue setting. Participants were then asked to state whether any other label (i.e., word or expression) came to their mind when they looked at the picture. These two questions were respectively referred to as “use in dialogue” and “other label.” Participants had to answer “yes” or “no” to each of these two questions. They were required to provide an additional label if they had answered “no” to the “use in dialogue” question or if they had answered “yes” to the “other label” question.

In questionnaires 2A and 2B, the name agreement question was followed by questions on the conceptual characteristics of the pictures that assessed image agreement, image variability, and Age of Acquisition (AoA). The image agreement question asked participants to judge to what extent the picture's appearance was representative of the mental representation associated with the label they had provided, using a 5-point Likert scale from "very weakly representative" to "very highly representative." In the image variability question, the participants were instructed to rate whether the label they had provided was related to few or many different visual representations, using a 5-point Likert scale from "there are very few ways to visually represent this word or expression" to "there are many ways to visually represent this word or expression." Finally, in the AoA question, the participants were asked to estimate the age at which they thought they had learned the labels they had provided by selecting one age class among five: 0–3, 4–6, 7–9, 10–12, after 12.

In questionnaires 3A and 3B, the name agreement question was followed by questions about the visual properties of the pictures. The second question of these questionnaires sought to determine whether the picture had already been seen (referred to as the "already seen" question; e.g., the participants might have already seen the picture before taking part in the study). This question involved a "yes/no" answer. It was then followed by questions on familiarity, visual complexity, and concreteness, all rated on a 5-point Likert scale. The familiarity question consisted in indicating how familiar the participant was with the picture on a 5-point Likert scale from "unfamiliar" to "very familiar." Regarding the visual complexity question, the participants had to rate the picture on a 5-point scale from "very simple" to "very complex." Finally, the concreteness question required participants to rate the concept associated with the picture on a 5-point scale from "abstract" to "concrete." We asked participants to name the picture before rating the related concept on the concreteness scale. All questions used in each questionnaire are listed in Appendix A (p. 264; we provide the initial French wording as well as a translation in English; the questions are listed in the same order as in the initial questionnaires).

2.3 Procedure

To complete the online questionnaire, participants were asked to sit in a quiet room to avoid distractions such as music or noise and to answer the questions at their own pace. The first page of the questionnaire was the description of the experiment (goal and duration) and was followed by the consent form. Once participants had given their consent by answering "yes" to the question "Do you consent to take part in this study?" they were shown the instructions of the task. Pictures were then displayed one by one on the participants' screen. The pictures were alternately shown on a green or blue background, the alternation making it easier for the participants to understand that they had switched to a new picture. Each page of the questionnaire included one picture as well as all the questions the participant was required to answer. All questions were presented on the same page, below the picture. Each participant was shown only one of the six questionnaires (1A, 1B, 2A, 2B, 3A, or 3B). Therefore, they saw only the questions corresponding to the questionnaire they had been allocated. Each question included one sentence (the question itself) followed by a space to answer, or a Likert scale, depending on the type of question. When all the questions corresponding to a given picture had been answered, participants clicked on the "next" button to move on to the following picture. The order in which the pictures were presented was randomized across participants. At the end of the questionnaire, participants provided demographic information regarding their first language and other spoken languages, gender, age, and history of language disorders. The entire questionnaire took approximately 90 minutes to complete.

2.4 Data pre-processing on name agreement question

The data from the name agreement question were first examined by two native French speakers in order to correct spelling mistakes. Determiners of isolated words were removed, except for words with different meanings depending on their grammatical gender (e.g., “le vase”, which means “vase”, had to be distinguished from “la vase”, which means “mud”). In cases where the participant’s response consisted of a letter (“M”) or expressions such as “the letter M” (“la lettre M”), only the letter (capitalized) was kept (“M”). Regarding numbers, when they were used to count things, as in the expression “two mountains” (“deux montagnes” in French), they were written in words. In all other cases (e.g., when the picture was believed to represent a number), the number form was kept. Plural words were replaced by singular forms. Finally, when a participant used a quotation mark to express their answer (e.g., “?”, implying that the picture looked like a question mark), the quotation mark was written out in full (e.g., “question mark”). All other punctuation marks were removed. Words voluntarily written in English were not translated.

Secondly, labels indicating that the participants did not know the name of the label or did not want to respond were coded as “non-responses” (“abs” in the data files; e.g., “nothing”, “no idea”, “no opinion”). We also coded the following responses as NA: (a) when we did not understand the labels or when they presented lexical ambiguities, (b) when participants provided more than one label, (c) when labels referred to another picture which had been shown previously (e.g. “the same corridor as before but the door is closer”).

Sixteen pictures were removed from the analysis because they generated either more than 10% of NA responses, or more than 10% of non-responses. This represented 0.05% of the dataset. In the final dataset, 0.70% of responses were NA responses and 1.20% of responses were non-responses. In addition, 27 pictures were removed from the final dataset owing to an experimenter error. Therefore, the answers for the name agreement and all other variables were analyzed on the 332 remaining pictures (166 in set A and 166 in set B). Summary descriptive statistics for these 332 pictures are presented in ╗ Table 2. Appendix A (p. 265-274) ╔ provides the associated mean and standard deviation values for each given picture on the following ╗

Table 2. *Summary statistics for all variables*

	H1	H2	H3	Image Agreement
Mean	3.91	3.60	3.65	3.25
SD	0.82	0.93	0.96	0.55
Median	4.19	3.84	3.94	3.18
Min	0.64	0.22	0.00	2.10
Max	4.86	4.86	4.81	4.79
Range	4.21	4.64	4.81	2.69

Note. H1, H index or name agreement for questionnaire 1; H2, name agreement for questionnaire 2;H3, name agreement for questionnaire 3.

variables: percentage of name agreement, H index, image agreement, familiarity, visual complexity, image variability, AoA, and concreteness.

2.5 Data analyses on name agreement data: modal label, percentage of name agreement, and H index

In line with previous literature on name agreement data (e.g., Alario, & Ferrand, 1999; Bonin et al., 2003; Dimitropoulou et al., 2009; Dunabeitia et al., 2018; Ghasisin et al., 2014; Manoiloff et al., 2010; Tsaparina et al., 2011), three measures were calculated for each picture: the modal label, which was the label that most participants gave to refer to a given picture; the percentage of name agreement, which corresponded to the percentage of participants who gave the modal label as their answer; and the H index. The H index (Shannon & Weaver, 1949) reflects the diversity in the labels provided by participants to refer to a given picture. The H index was calculated for each picture using the following formula:

$$H = \sum_{i=1}^{k} p_i \, log_2(1/p_i)$$

where *k* refers to the number of different labels given to each picture and p represents the proportion of subjects who gave each label (Snodgrass & Vanderwart, 1980). More precisely, if all participants use the same label to refer to a given picture, the picture has an H index of 0 and its percentage of name agreement is 100. In contrast, when the variability in labels provided across participants increases, the value of the H index also increases, and the percentage of name agreement usually decreases as well. As already defined in the literature (Alario & Ferrand, 1999; Snodgrass & Vanderwart, 1980), the H index is calculated based on name agreement question because the H index captures the distribution of labels for each picture across participants better than the percentage of name agreement. The modal label, the percentage of name agreement, and the H index were computed on the two sets of pictures separately and on both sets combined (set A-B in Table 3). The results for the H index and the percentage of name agreement are presented in Table 3, while modal labels can be found online (see "ModalResponses-NamingAgreement" file from https://osf.io/bxkpa/?view_only=2fe0acd124e64df19a0195354ebe45b4).

Moreover, we explored whether the diversity in labels provided across participants for each given picture was concordant across the three sets of questionnaires. As pointed out by Snodgrass and Vanderwart (1980), the diversity in the participants' labels to refer to a given picture is best reflected by the H index. We therefore calculated Kendall's coefficient of concordance on the H indexes obtained in each questionnaire by comparing the different questionnaires of a same set of pictures (i.e., 1A, 2A, and 3A or 1B, 2B, and 3B). In this analysis, the pictures rather than the participants were the basic analysis unit. Given the high number of pictures, the tables of critical values for the Kendall's W statistic (Siegel & Castellan, 1988) were not appropriate to test W for statistical significance, and a chi-square test of significance was used instead.

Table 3. *Summary descriptive statistics for name agreement (percentage of name agreementand H index) in sets A and B taken separately and for both sets combined.*

	Set A		Set B		Set A-B	
	Name agreement (%)	H index	Name agreement (%)	H index	Name agreement (%)	H index
Mean	24.18	4.64	23.85	4.55	24.01	4.59
SD	18.18	1.15	17.64	1.14	17.89	1.14
Median	18.34	4.95	18.18	4.84	18.18	4.88
Range	25.88	1.71	20.24	1.55	23.40	1.62
Q1	9.41	3.82	10.71	3.87	9.64	3.86
Q3	35.29	5.53	30.95	5.42	33.04	5.48

Note. SD, standard deviation; *Q1*, 25th percentile; *Q3*, 75th percentile

We then examined the homogeneity of modal labels for each given picture across the three sets of questionnaires. In other words, for each given picture, we checked that the mode was the same regardless of the questionnaire in which this picture was presented. We calculated Krippendorff's alpha rather than Fleiss' kappa, as the comparison included three questionnaires for each set of pictures, A and B, and the dataset contained missing data (Zapf et al., 2016). Using the R-function kripp.alpha of the irr package (Gamer, Lemon, Fellows, & Singh, 2019), we compared the modal label obtained for each picture separately for both sets of pictures. This was performed in questionnaires 1A, 2A, and 3A on the one hand, and in questionnaires 1B, 2B, and 3B on the other, as sets A and B included different pictures.

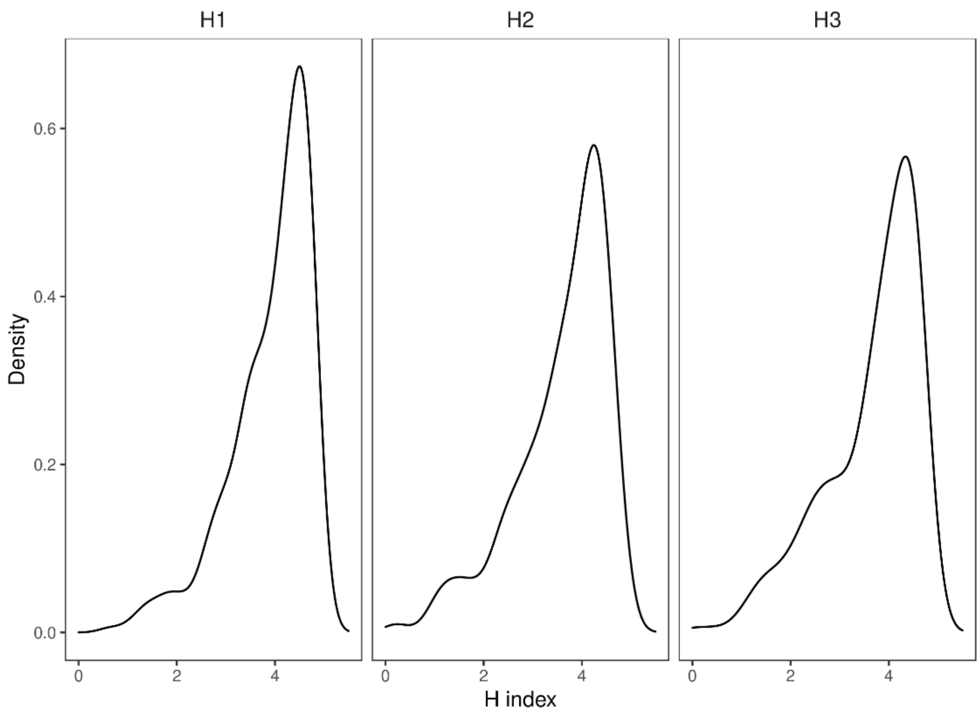
Importantly, 67 pictures from set A and 63 pictures from set B had more than one modal label (multiple-mode pictures, i.e., pictures for which two or more labels had been given the same number of times and were the most frequent labels). For example, picture A23 (set B) was named "bouteille" (i.e., bottle) by 12 participants and "maison" (i.e., house) by 12 other participants across all three questionnaires. In addition, 12 pictures from set A and 3 pictures from set B had no modal label. These were cases where each participant provided a different label to describe a picture. Each label was thus provided only once. Two different approaches were considered to solve these particularities in modal labels. The first approach consisted in including multiple-mode pictures in the analysis, but only taking one modal label per picture into account. We selected the mode included in the analysis using the following procedure: for each picture, if one of the multiple modes in one of the questionnaires was the same as the (unique) mode in another questionnaire, it was this mode which was included in the analysis, as it could be considered the most representative of the picture. All other possible modes were removed from the analysis. If none of the multiple modes matched the modes of the other questionnaires, then all multiple modes were replaced by an NA response and were not included in the analysis. The second approach consisted in removing, for each picture, the data from all three questionnaires if the picture was associated with multiple modes in at least one of the questionnaires. The modal responses for each approach are available online (see "ModalResponses-NamingAgreement" file from https://osf.io/bxkpa/viewonly=2fe0acd124e64df19a0195354ebe45b4). ╗

3. Results

The summary of descriptive statistics is presented in Table 3 and include the percentage of name agreement and the H index. As shown in Table 3, mean percentage of name agreement across participants from sets A and B combined was 24.01% (*SD* = 17.89), which reflect the high variability between participants in the labels used to refer to a given picture. In line with the percentage of name agreement, the average H index for both sets A and B combined was 4.59 (*SD* = 1.14), which again shows a large variability across participants in the labels given to each picture. Almost half of the pictures (151 tangram pictures) were associated with a modal response given by 10% to 30% of the participants, while only 11% of the pictures (37 pictures) had a modal response given by participants more than 50% of the time. One possible explanation for this lack of agreement lies in the opportunity for participants to use referential expressions instead of isolated words only to refer to the pictures. To address this possibility, an additional level of coding was used on the modal labels. When the modal label for a given picture was a word, we computed the total number of labels in which this word occurred alone or as part of an expression. However, when the modal label was a referential expression such as “a person who dances”, we calculated the total number of labels in which the content words included in the reference (in this case “person” and “to dance”) appeared separately or together. In cases when a picture was associated with at least two modal labels, we performed this additional level of coding for both labels and kept the most frequent one afterward. The aim of this additional coding was to examine whether this new counting method led to percentages of name agreement that were closer to previous studies on name agreement (e.g., Alario, & Ferrand, 1999; Bonin et al., 2003; Dimitropoulou et al., 2009; Dunabeitia et al., 2018; Ghasisin et al., 2014; Manoiloff et al., 2010; Tsaparina et al., 2011). Using this procedure, the maximal mean percentage of name agreement found in set A was 38.09% (*SD* = 24.28) and 31.95% (*SD* = 20.59) in set B, which was once again lower than the values reported in previous studies.

Kendall's coefficients of concordance were computed between the H indexes of each set of questionnaires within each set of pictures. The analysis revealed a coefficient of concordance between the three questionnaires of set A equal to 0.91 and the same coefficient (0.91) was obtained for the three questionnaires of set B. The chi-square test revealed a significant concordance for both sets A and B (respectively $\chi 2 = 451$, $p < 0.001$ and $\chi 2 = 448$, $p < 0.001$). This result implies that the H index of the 166 pictures of each set (A and B) could be ranked in approximately the same order in each of the three questionnaires (Siegel & Castellan, 1988). In other words, the agreement across participants regarding picture naming was concordant across the three questionnaires. This result is also well reflected in Figure 8 in which the distribution of the H index for each questionnaire has been plotted.

Figure 8. *Density plots of the H index for the three sets of questionnaires (1, 2, and 3). All three distributions show a maximum density of around 0.6 for a H index equal to four, indicating that the three sets of questionnaires elicit very similar patterns of density.*



The calculation of the Krippendorff's alpha which reflects the homogeneity of modal labels across the three sets of questionnaires revealed a reliability rate of 0.69 for set A, 0.66 for set B, and 0.68 for both sets combined. However, when multiple-mode pictures were removed from this analysis, we observed a higher reliability rate with 0.78 for set A, 0.71 for set B, and 0.74 for both sets combined. According to Krippendorff (2004), it is acceptable to draw tentative conclusions given that all the reliability coefficients in the present analyses are between 0.67 and 0.80 (Krippendorff, 2004). Thus, modal responses for a given picture were closely related across all three sets of questionnaires.

4. Discussion

In this chapter, we presented the first normative database for 332 tangram pictures and their associated French names. The aim of this database was for us to determine whether tangram pictures can be associated with several labels that are sufficiently consensual to be used in our initial experiments, and to obtain a measure of name agreement in order to select the most appropriate stimuli for our first set of experiments (see Chapter 5). By analysing the responses on the names provided by 169 participants, we were thus able to calculate the name agreement and the H index for each picture as indicators of the variability of names associated with each picture. The material of this study can be found in open-access and free from copyright restrictions for non-commercial purposes at the following URL: https://osf.io/bxkpa/view_only=2fe0acd124e64df19a0195354ebe45b4.

The main outcome of this study concerns the high variability in the labels given to most of the pictures in all three sets, as shown by the H index and the low percentage of name agreement. Crucially, this variability was shown to be highly reliable between our samples of participants, as shown by the Kendall's coefficient. Thus, this implies that the amount of variability in the labels given to each picture was consistent across the three groups of participants in both sets of pictures. It has also been shown that this high degree of variability could not be imputed to the format of the participants' responses. Indeed, it could be argued that allowing participants to use isolated words or expressions to name the pictures could increase the number of ways in which the same label may be expressed, thus accounting for the variability of our data. However, the analyses performed on our additional level of coding with the counting of word content into the expressions did not support this hypothesis. Specifically, even if the percentage of name agreement slightly increased in this analysis, it remained low in comparison with other studies on name agreement that used line-drawing pictures (e.g., Alario & Ferrand, 1999; Bonin et al., 2003; Dunabeitia et al., 2018).

Taken together, these results reinforce the idea that tangram pictures are suitable stimuli for our first set of experiments and for dialogue studies in general. As already mentioned, dialogue researchers often need to rely on stimuli that prompt natural negotiations between participants to study the structure and content of such exchanges (Bangerter et al., 2020; Branigan et al., 2011; Brennan & Clark, 1996; Clark & Wilkes-Gibbs, 1986; Duff et al., 2006; Horton & Gerrig, 2002; Knutsen et al., 2018; Knutsen et al., 2019; Yoon & Brown- Schmidt, 2014, 2019). As tangram pictures can be seen in various ways, they constitute an ideal setting for such experiments. Importantly, this database still provides tangram pictures with various degrees of name agreement, allowing researchers to purposefully choose the most suitable pictures as a function of the amount of negotiations they require. Moreover, researchers working with French speaking participants can also make use of the proposed names for each picture in order to control for the words that could be proposed during the interaction. We can imagine for instance a study in which one may want to prompt similar names for two pictures and see how participants deal with this situation (see Brennan & Clark, 1996 for similar studies). Ultimately, researchers working with non-French speaking participants may refer to the whole database provided in Fasquel et al. (2023; see Appendix A) to rely on other variables independent of the lexical label of the pictures in order to have a better control on their choices of tangram pictures. Indeed, this database also includes several parameters for each picture from visual properties to conceptual representations (name agreement, image agreement, familiarity, visual complexity, image variability, AoA, and concreteness).

In conclusion, the present study was designed to meet a need for standardised stimuli for dialogue studies that require pictures that can be seen in different ways and prompt natural negotiations. Although this need emerged from the idea of finding pictures that can be associated with two plausible but semantically unrelated labels, the resulting database should also be useful for the community of researchers working in the field of dialogue. In the next chapter, we present the set of three experiments for which this database has been primarily created.

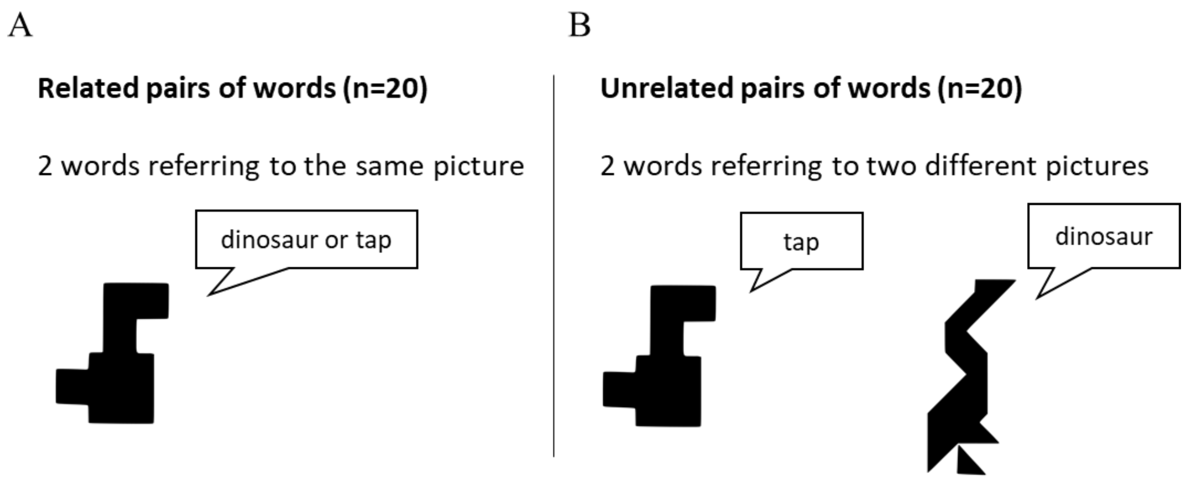
Chapter 5 – Creation of new lexico-semantic relationships in long-term memory through dialogue

1. Introduction and overview of the chapter

As we saw in the theoretical chapters, the model by Pickering and Garrod (2004; 2021) suggests that linguistic representations could change during a dialogue. Interestingly, we already know the organisation of representations stored in the lexico-semantic network can be adapted after exposure to isolated oral sentences and that these changes can be observed in the short and long term (Gaskell et al., 2019; Rodd et al., 2012; Rodd et al., 2016). However, the adaptation of lexico-semantic representations stored in long-term memory has never been tested after exposure to a dialogue situation. The aim of this set of experiments was thus to fill this gap by testing the possibility that dialogue can impact lexico-semantic representations such that they remain changed after dialogue. Specifically, we wanted to test whether new lexico-semantic relationships created during a dialogue could change already existing lexico-semantic representations immediately after dialogue but also after one night of sleep promoting offline consolidation. To do so, we associated two semantically unrelated words to a common visual representation and assessed the potential new relationship between those words immediately after dialogue or after a night of sleep to promote offline consolidation in long-term memory.

As semantically unrelated words are not meant to possess common visual properties, we use abstract pictures that could be named after various semantically unrelated words. Crucially, in the previous chapter, we have demonstrated that tangram pictures were ideal candidates in that they are generally associated with several plausible labels. We thus used the previously presented database for tangram pictures to create the present set of experiments. We ran three experiments based on these tangram pictures and created a new dialogue task. In this task called the Interactive Agreement Referential task (IAR task), a confederate and a participant interacted via an audio-conference device to find an agreement on how to name the tangram pictures. Thanks to the confederate, tangram pictures were named using either one word (e.g. 'tap') or two semantically unrelated words (e.g., 'dinosaur' and 'tap' referring to the same tangram picture, see Figure 9a). By doing so, we took advantage of the new word-picture mappings during dialogue to create new lexico-semantic relationships between pairs of semantically unrelated words. The other types of pairs of words were words associated to two distinct pictures during dialogue, thus creating the "unrelated condition". In line with the study by Betts et al., (2018) showing a boost of changes in lexico-semantic representations after three spaced repetitions of a linguistic input, the words referring to the tangram picture were repeated four times through different turn-takings. Moreover, to ensure that the number of words associated with each picture remained constant across all participants, the confederate followed a script while keeping the interaction as naturalistic as possible. In these scripts, the order of presentation of each word of a pair was alternated across the repetitions in order to maximise the abstraction of the new lexico-semantic relationship (as shown in Figure 9b). Crucially, the names proposed by the confederate to refer to the tangram pictures were chosen by selecting

Figure 9. *Schematic representation of a trial. The confederate and the participant saw the same three pictures and the confederate had to name them using either one (B) or two words (A).*



the more consensual names for each picture in the database presented in Chapter 4, to increase the probability that participants agree with them.

Following the IAR task, participants of Experiment 2 and 3 performed a semantic priming paradigm with a lexical decision task in order to assess the relationship between the words referring to a common tangram pictures with an implicit measure of the semantic relationship. The task was performed immediately after dialogue in Experiment 2 or one day later in Experiment 3. In this task, we used a very short SOA (i.e. 166ms) in order to favour interpretations in terms of spreading of activation rather than strategical decisions (Hutchison et al., 2001). In Experiment 4, the lexical decision task was replaced by a semantic relatedness judgement task performed one day after dialogue in order to assess the new lexico-semantic relationship in a more explicit way. The aim of this task was to determine whether changes in pre-existing lexico-semantic representations through dialogue were observable when attention is focused on the semantic properties of the relationship, by using a task which explicitly asks participants to focus on the semantic relationships between words. Finally, it is important to note that the presentation of the word in the lexical decision task as well as the semantic relatedness judgement task were performed in the visual modality in order to change the modality of presentation of the words between the dialogue and the assessment of the new relationship. This manipulation was meant to prevent our results to be interpreted as a new relation stored in the episodic memory, and favour an interpretation in terms of changes in the weights of lexico-semantic representations stored in semantic memory.

Overall, we expected that pairs of words referring to a shared visual representation become more semantically related than pairs of words referring to two distinct visual representations during dialogue (see Figure 9). More precisely, in line with the literature on word-meaning priming, we expected similar results in Experiment 2 when assessing the new relationship immediately after dialogue and in Experiment 3 in which the assessment took place after a night of sleep (Gaskell et al., 2019; Rodd et al., 2013; Rodd et al., 2016). In the semantic priming paradigm, we expected that targets belonging to pairs of words referring to the same tangram picture elicit faster reaction times following the presentation of the other word of the pair than targets belonging to pairs of words referring to two distinct tangram pictures. In Experiment 4 we expected greater semantic relatedness scores for pairs of words referring to the same tangram picture than for pairs of words referring to two distinct tangram pictures after dialogue.

In line with the literature on word-meaning priming showing a progressive decay in the strength of the priming until offline consolidation occurs (Gaskell et al., 2019; Rodd et al., 2016), we performed an additional control analysis on the time of the day at which participants from Experiment 3 and 4 took part in the dialogue task. Specifically, we explored whether participants performing the dialogue task earlier in the day could be susceptible to more decay in the strength of the changes in the pre-existing lexico-semantic representations in comparatively with participants performing the dialogue task later in the day. However, we did not specifically control for the delay between dialogue and sleep onset for all participants as the time course of the changes in the pre-existing lexico-semantic representations was not questioned in this thesis and we were only interested in determining whether or not these changes were observable after a night of sleep.

Finally, at the end of all three experiments, participants were asked to fill in two personality questionnaires: the Interpersonal Reactivity Index (IRI) questionnaire (Braun et al., 2015) and the Plymouth Sensory Imagery Questionnaire-French (PSIQ-F; Ceschi & Pictet, 2018). The two questionnaires were added as exploratory measures in order to check for potential inter-individual variability in the way dialogue impacts lexico-semantic representations. Specifically, they were added to evaluate a potential impact of empathy and mental imagery on the changes of lexico-semantic representations after dialogue. Regarding the

IRI questionnaire, we were particularly interested in the way the Perspective Taking dimension could play a role in the changes of lexico-semantic representations after dialogue. Indeed, by switching from their own perspective to the confederate's perspective, people with more Perspective Taking abilities could be subject to more changes of lexico-semantic representations. Concerning the PSIQ-F questionnaire, we wanted to check whether visual mental imagery might affect our results, given that participants had to name abstract visual representations (i.e., tangram pictures) with real words. The detailed methodology for this set of experiments is presented in the following section and extracted from the article presented in Appendix B (Fasquel et al., under review). As for the Chapter 4, the sections extracted from the article are flanked by the symbols '╔' and '╗'.

2. Experiment 2

╔ 2.1 Method

2.1.1 Participants

Forty-six native French speakers (35 female; 44 undergraduates and two community members; age, *M* = 21.23, Standard Deviation, *SD* = 2.53) were paid 15€ to take part in the experiment. All participants had normal or corrected-to-normal vision and no history of language disorder, auditory disorder or neurological disorder. They signed an informed consent form before the beginning of the experiment and were fully debriefed after the experiment. This work was approved by the Ethics Committee of the University of Lille (reference: 2021-512-S95). Six participants were excluded from the analysis of data (for more details, see the experimental design and data pre-processing section).

2.1.2 Confederates

During the dialogue, participants interacted with one of the two confederates of this experiment. Both were native French speaking women aged of 24 and 25 years old and had no history of language disorder, auditory disorder or neurological disorder. More precisely, twenty-two participants interacted with one of the confederates while eighteen participants interacted with the other.

2.1.3 Apparatus

The dialogues were recorded using a Zoom H4n with a dual-entry digital recorder (one per partner).

2.1.4 Stimuli

The interactive agreement referential task (IAR)

Two sets of 40 tangram pictures and their associated words were selected to be discussed by the confederate. A third set of 30 tangram pictures and their associated words was selected to be discussed by the participant. All the pictures came from the Fasquel et al. (2023) tangram picture database, previously described (see Chapter 4). In this database, any given picture could be associated with a single word or expression, or with several different words or expressions (i.e., in some cases, all participants provided the same expression to refer to the same picture; in other cases, different participants provided different expressions to refer to the same picture, implying that this picture was associated with different expressions in the database). This database enabled us to select pictures associated with at least two plausible words or expressions (e.g., a picture could be associated with the words "anchor" and "hat"). We then pre-selected 148 pictures which were those associated with at least two plausible words or expressions from this initial database and we asked 84 new native French speakers to name them by providing only isolated words (this new database is online available from the following url, https://doi.org/10.5281/zenodo.10419752).

The first set of 40 tangram pictures (set 1) and words was selected among these 148 pictures. In this set, each picture was associated with two semantically unrelated words. Thus, for a picture to be included in this set, it had to be associated with two plausible words which were not semantically related before the dialogue. The degree of relatedness between each pair of words was pre-tested by 29 other native French speakers on a 7-point Likert scale (1 = not related at all, 7 = strongly related). We ensured that each pair was rated below 4, meaning that they were all considered as semantically unrelated. Overall, the mean relatedness was 1.75 (*SD* = 0.71). Due to the written presentation of words during the lexical decision task, each word did not exceed a length of 10 letters (range: 3-10 letters). The psycholinguistic characteristics of each word within pairs are reported in Table 4.

In contrast with the first set in which two words were associated with the same picture, the second set of 40 tangram pictures associated with words was selected to create pairs of words associated with two different pictures. To do so, each pair of words from set 1 was divided into two so that one of the words remained associated with its corresponding picture from set 1

Table 4. *Psycholinguistic properties of primes and targets for experimental lists 1 and 2 separately.*

	Lexical frequency		Number of letters		Number of phonemes		Number of orthographic neighbours		Number of phonologic neighbours	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
List 1	Primes	45.40	88.84	5.90	1.59	4.40	1.43	3.10	3.18	6.60
	Targets	28.08	44.07	6.20	1.20	4.60	1.27	3.20	4.40	9.75
List 2	Primes	50.40	129.17	6.10	1.52	4.40	1.31	4.60	6.42	8.45
	Targets	30.78	45.65	6.20	1.77	4.60	1.53	3.60	5.43	9.15

Note. All properties were extracted from New et al. (2004); lexical frequency was based on movie subtitle corpora.

1, while the other word within each pair was associated with a new picture from the original database. Specifically, all these new pictures had been labelled at least once with their new associated word from set 1. More precisely, we selected these words from the Fasquel et al. (2023) database or the subsequent database of isolated words for 148 pictures. Each new picture from set 2 was therefore associated with only one word. In sets 1 and 2, all selected words were nouns. These two sets of pictures and words allowed to define our two experimental conditions with either related or unrelated pairs and to test our hypotheses.

For a picture to be selected in the third set of pictures and words, it had to be associated with words that were absent from sets 1 and 2. As explained below, these pictures were discussed by the participant rather than by the confederate. The third set allowed to make sure the participant remained engaged in the interactive agreement referential task. These pictures came from the Fasquel et al. (2023) database and their two most frequent words in the database were selected. Contrary to sets 1 and 2, the words could either be nouns or verbs and each pair of words associated with a picture could either be semantically related or not, as the data from the trials in which the pictures were described by the participant were not analysed.

The lexical decision task (LDT)

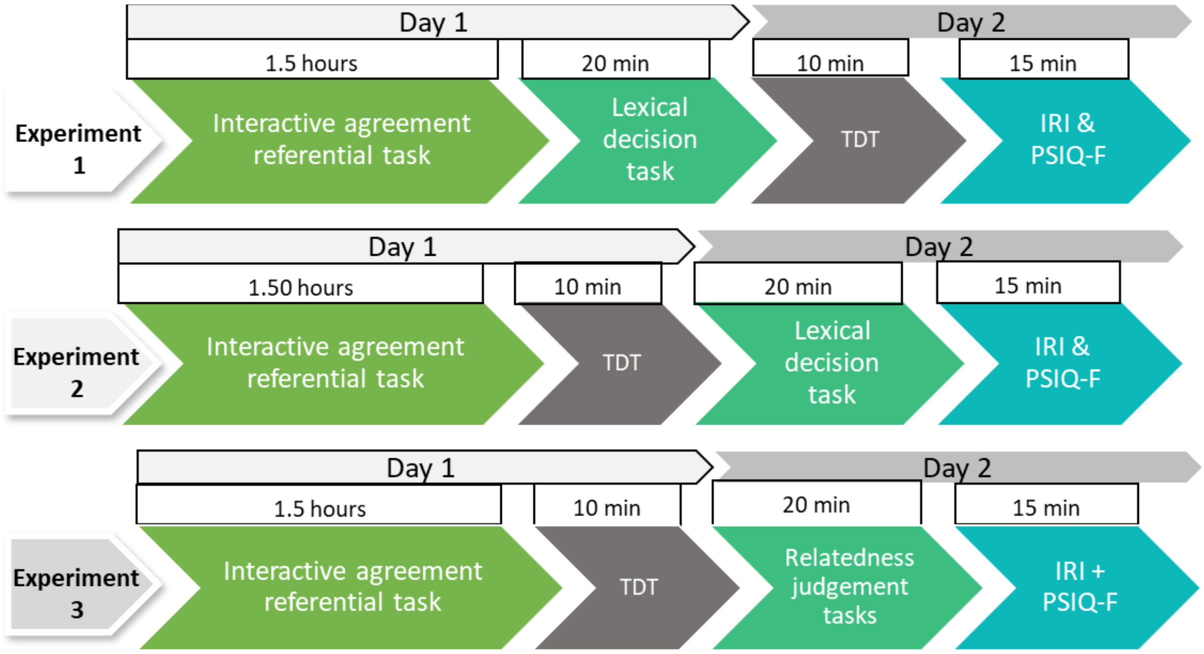
The 40 pairs of words associated with the first and second sets of tangram pictures were used as prime-target pairs in the lexical decision task. They were divided into two experimental lists so that each pair of words was presented only once per participant but was presented in the two experimental conditions across participants (20 pairs of words per experimental condition, related vs. unrelated). The properties of primes and targets were matched in terms of frequency, length and neighbourhood between the two experimental lists (see Table 4). Word pairs were also matched in terms of semantic relatedness and co-occurrence frequency between the two experimental lists (semantic relatedness: list 1, $M = 1.75$, $SD = 0.83$, list 2, $M = 1.76$, $SD = 0.58$;

co-occurrence frequency: list 1, $M = 0.35$, $SD = 0.93$, list 2, $M = 0.95$, $SD = 3.56$). The co-occurrence frequency was calculated with the same approach used in Brunellière et al. (2017) and in Brunellière and Bonnotte (2018). The same distribution in two lists was used for both the pictures of sets 1 and 2 in the IAR task and the words forming the experimental conditions in the lexical decision task. We added 40 filler pairs of words which were unrelated pairs of familiar nouns and 80 word-pseudoword pairs in each list. Importantly, forty words selected as primes of word-pseudoword pairs were produced in the IAR task, such that participants could not use the information related to repeated items during the dialogue to make their lexical decision on the targets. Pseudoword targets were orthographically legal and were constructed by replacing a letter in French words different from those used in the IAR task and from those used as fillers in the lexical decision task. Word and pseudoword targets were matched in length (number of letters, for words: $M = 6.76$, $SD = 1.68$; for pseudowords: $M = 7.18$, $SD = 1.57$).

2.1.5 Procedure

Experiment ╗ 2 ╔ took place on two consecutive days in order to keep the procedure constant across all three experiments (see Figure 10). In line with the main objective of the present experiment, the lexical decision task used in Experiment ╗ 3 ╔ and the semantic relatedness judgment task used in Experiment ╗ 4 ╔ were carried out one day after the IAR task (i.e., day 2 after the dialogue). While session 1 corresponded to the tasks performed during day 1 (i.e., interactive agreement referential task and lexical decision task, in Experiment ╗ 2 ╔*),* the session 2 consisted of tasks performed on the day 2 (i.e., filler task and questionnaires, in Experiment ╗ 2 ╔). Both sessions were performed in the same testing room and the overall experiment lasted around two and a half hours. ╗

Figure 10. *Recap of the procedure of the three experiments. TDT: Tone Deafness Test; IRI: Interpersonal Reactivity Index questionnaire; PSIQ-F: Plymouth Sensory Imagery Questionnaire in French.*



Session 1

Interactive agreement referential task (IAR)

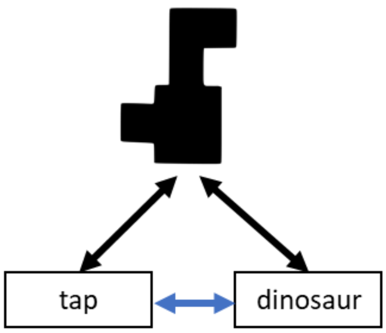
The participant and the confederate were placed in two separate rooms. The confederate read the scripts during the interaction, without being seen by the participant. The participant was informed that they would have to interact with another participant installed in another room using an audio-conference device and a headphone, and that the discussion would be recorded for the purpose of the experiment. They were also informed that, for the whole interaction they would be named "participant B" and that they will see the same pictures than the other person, named "participant A". The purpose was to reinforce the participant's impression that the confederate was also a naïve participant. The picture presentation was run on the computer screen using GNU Octave (version 6.4.0).

The IAR task was divided into 30 trials and a break was given to the participant after 15 trials. Each trial began with the instruction of which of the participant or the confederate had to propose words for the following pictures. The instruction was displayed on both the screen of the participant and the one of the confederates, and it was then followed by the presentation of three tangram pictures side by side. The instruction was divided into three steps. The first step (i.e., naming step) was to choose one or two words for each of the three pictures displayed on the screen. More precisely, the person proposing the words on that trial had to choose words among a list of six possible words and to write their choices on a response sheet before giving the choices of words. The six-word list contained the pairs of words associated with the filler tangram pictures (see the Stimuli section for more details). When the words chosen to refer to each of the three tangram pictures had been provided, the partners went to the second step (i.e. justification step) and the person who had chosen the words had to justify their choices (i.e. to what extent the picture reminds them of the proposed word) for the three pictures. The last step (i.e. agreement step) was to ask the partner if he or she agreed on the proposed words and write the partner's answer on the response sheet. Then, the trial ended and the confederate pressed the space bar to start the next trial and to control the transition between trials, while the participant was informed that the experimenter performed this task. Crucially, rather than following the same instructions as the participant, the confederate followed a script which enabled them to produce the critical words four times in trials in which it was their turn to provide the words. Thanks to this structure, the two partners were able to reach an agreement rapidly and explicitly.

For ten of the thirty trials, participants had to propose words to name tangram pictures. Half of these trials were presented in each block (five per block of 15 trials). The purpose of these trials was to encourage the participant to actively engage in the task (rather than simply listening to the confederate). The data from the trials in which the participant had to name the pictures were not analyzed further. The partners were given a training phase to familiarize themselves with the procedure. This training phase was composed of one trial with words given by the confederate, followed by one trial with words proposed by the participant.

For twenty of the thirty trials, the confederate had to propose the words to refer to the tangram pictures. As mentioned previously, in these trials, the confederate had to read the scripts as naturally as possible (see for an example of script at the following URL, https://doi.org/10.5281/zenodo.10419752). The scripts followed the same rules as those given to the participant on how to proceed when naming the pictures (see Figure 11). First, the confederate proposed the words associated with each picture (see the stimuli section for more details on the choices of words). When a picture was associated with a pair of words, these words were placed within a window of 10 words in the same sentence (see Brunellière et al., 2017, for justification). Words given to different pictures within the same trial were spaced by at least 10 words, such that the participant would have no doubts about the choice of words for a given picture. To do so, filler sentences such as "I think that's the best word" could be

Figure 11. *Adapted illustration of a trial. A: Example of two words proposed to name the picture in order to create a new lexico-semantic relationship between the two words (blue arrow). B: Example of a script with the alternance of the presentation of the words.*



Step 1. Naming: "I really think that this picture may be a **tap** and a **dinosaur**."

Participant's turn

Step 2. Justification: "For the **dinosaur** and the **tap**, the first one is obvious, because it has a little head at the top, with a long neck, and then the cube may represent its body, with its little tail on the left. Or we can see it reverse and on the left you have where the water fall, and on top the crank to open water".

Participant's turn

Step 3. Agreement: "Now, what do you think about the **tap** and **dinosaur?**

Participant's turn

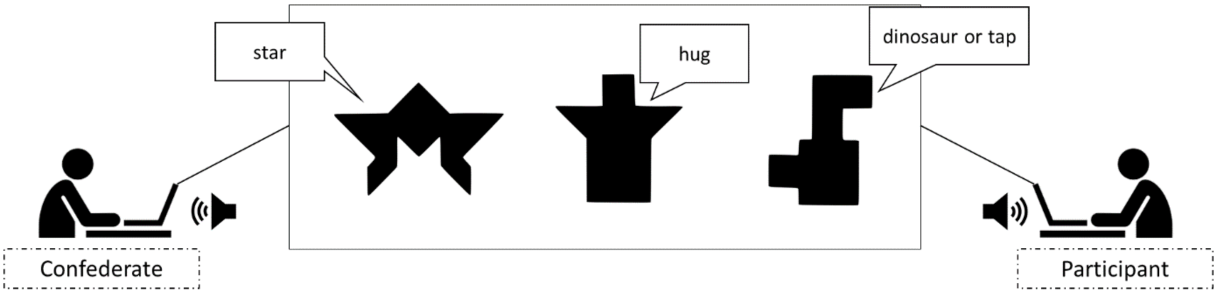
"[Check/cross] in the box, the **dinosaur** and the **tap** are [okay/not okay] ."

placed after each sentence containing a picture's word. After the picture-naming step, the confederate had to argue the choices of words. During this justification step, the words associated with each picture were repeated a second time and the picture was described using its geometrical shapes and global shape. The goal of the description was to justify the relationship between the picture and its given word. The descriptions included none of the words used to name the tangrams. After the justification step, the confederate had to ask the participants whether they agreed or not with the suggested words. This agreement step allowed the confederate to repeat a third and a fourth time the words associated to the pictures that is, for the agreement and then when the confederate confirmed that the words have been written on the response sheet. Therefore, the repetitions of words referring to a given picture were spaced through the different turn-takings during the three steps. Importantly, when pairs of words were proposed for a given picture, the order of presentation of the words of the same pair was alternated along the script.

Participants were randomly assigned to one of the four versions of the IAR task before their arrival. Two experimental lists made it possible to present the word pairs either in the related condition or in the unrelated condition and there were two different orders of presentation for the 20 trials for which the confederate had to ask to propose the words to refer to the tangram pictures. Two different versions were thus obtained for each experimental list. Although the order of these trials varied, the filler trials in which the participant described the pictures to the confederate were the same across all four versions of the task. Each version of the IAR task was associated with one script and the structure of the four scripts was always the same, and only the picture's word and its descriptions changed. The sequences of sentences were then adapted in order to make them as natural as possible.

In each trial in which the references were proposed by the confederate, the participants were exposed to three pictures: two pictures from set 1 (each picture was thus associated with

Figure 12. *Schematic representation of a trial. The confederate and the participant saw the same three pictures and the confederate had to name them using either one or two words*



two words) and one picture from set 2 (this picture was associated with just one word; see Figure 12). Importantly, 4 trials among the 20 did not follow these rules to prevent the participants from finding regularities in the composition of the trials. Two trials were thus composed of three pictures associated with only one word and two trials were composed of two pictures associated with two pairs of words. Additionally, each trial was composed of pictures of various shapes (rather square, long or height, straight or curved), complexity, and the category of objects represented by the corresponding words (either living or inanimate object). The composition of the filler trials also varied in terms of shape and complexity of the pictures.

The recordings of the dialogues were checked to make sure that words were correctly repeated four times by the confederate. The mean number of repetitions of the words referring to pictures by the confederate approached four (M = 3.94, SD = 0.34), confirming that the required number of repetitions were produced by the confederate. The number of disagreements given by the confederate on the words proposed by the participant was kept constant across participants. Specifically, the confederate had to disagree on one trial per block, but was able to choose the trial to disagree on to match with the relevance of the participant's justifications, in order to keep the interaction as naturalistic as possible. The mean percentage of disagreement given by the participants to the words proposed by the confederate was 8.63 % (SD = 7.04), showing that the participants easily reached an agreement with the confederate.

Lexical decision task (LDT)

After the IAR task, the participants performed an individual lexical decision task with a semantic priming paradigm. Stimuli were presented using the E-Prime software (E-Prime 2.0.10.356, Schneider et al., 2002a; 2002b). Participants first performed 16 practice trials before proceeding with the 160 experimental trials. They were divided into two blocks of 80 trials and a break between the blocks was proposed to the participant. Stimuli were displayed on the screen in lower case and in white font against a black background. Each trial consisted in the presentation of a fixation cross for 500 ms, followed by a prime word for 150 ms. A black screen was then displayed for 16 ms (leading to a SOA of 166 ms) before the presentation of the target stimulus, which remained on the screen until the participant's response. Participants had to indicate as quickly and accurately as possible whether or not the target stimulus was a real word. They had to respond by using a button box for which the assignment of the button responses depended on their handedness. At the end of each trial, a black screen was displayed for 1,500 ms before the next trial began. The task took approximately 15 minutes to be completed.

Session 2

Filler task and questionnaire

Participants came back the following day to perform a filler task and fill in personality questionnaires. The filler task was used to keep constant the timing of each session across Experiments ╗ 2 and 3 ╔ (2h on day 1 and 30 min on day 2). It consisted in an online citizen-science experiment designed by researchers at Yale University, called the Tone Deafness Test (see: https://www.themusiclab.org/quizzes/td). In this task, participants had to listen to sounds of different pitches and indicate whether the last sound went up or down. They were also asked questions regarding their experiences with music.

Regarding the questionnaires, participants were asked to fill in the IRI questionnaire (Braun et al., 2015) as a measure of cognitive and affective empathy, and the PSIQ-F questionnaire (Ceschi & Pictet, 2018) as a measure of mental imagery. Both questionnaires were implemented using an online survey administered through LimeSurvey 2.06 (LimeSurvey GmbH, n.d.). At the end of the questionnaires, participants answered two additional questions in order to determine if they were aware of the presence of a confederate. None of the participants spontaneously reported being aware of the presence of a confederate in the first question. In the second question, none of the participants reported being aware of the confederate due to the dialogue feeling unnatural.

2.1.6 Data analysis and pre-processing

Experimental design and data pre-processing

Reaction times from the lexical decision task were analyzed with linear mixed-effects models using lme4 package in R (Baayen et al., 2008). The dependent variable was the participants' reaction times on the lexical decision task. The independent variable, called the relatedness factor, was whether each pair of words shown in the lexical decision task was made of words which were both used to refer to the same picture (i.e., related condition) or which were used to refer to different pictures (i.e., unrelated condition). As in Barr et al. (2013), we started by implementing the maximal random effect structure justified by the design. This involved by-participants random intercepts, by-participants random slopes corresponding to relatedness, by-item random intercepts and by-item random slopes corresponding to the relatedness factor (note that in the present experiment, items corresponding to the target words of each pair on which participants made their lexical decision). Unfortunately, models including this maximal random effect structure systematically failed to converge. We thus attempted to simplify the structure of the models in order to improve model convergence, and we found

Table 5a. *Fixed effects, Experiment 2*

Fixed effect	χ^2	DF	*b*	*SE*	*p*
Relatedness_factor	0.18	1	2.67	6.37	0.68

Table 5b. *Random effects, Experiment 2*

Random effects	**Variance**	***SD***
By-participant		
Intercept	5246	72.43
By-item (targets)		
Intercept	1384	37.20
Residual	15996	126.47

that the only model to converge was a model including only by-participant and by-item random intercepts. The fixed effect remained as planned with the relatedness of the pair of words during the dialogue which was divided into two experimental conditions (related, or unrelated condition). The results from the main model are reported in Tables 5a and 5b.

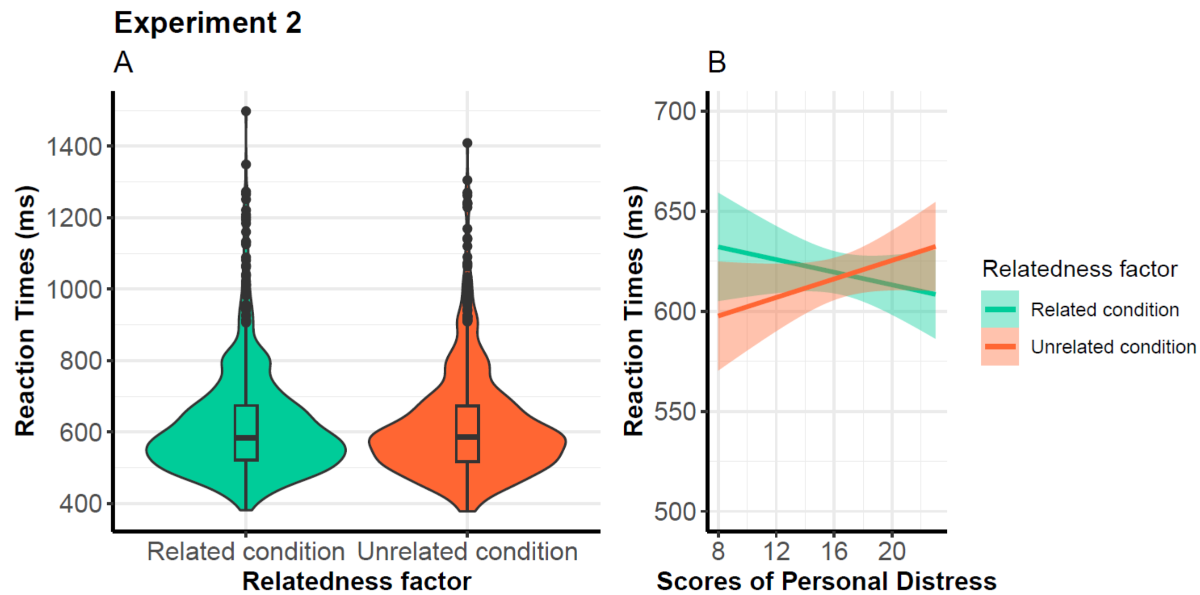
Two additional analyses were conducted to investigate the modulation effects of personality traits on the changes in lexico-semantic representations after the dialogue. These models both included by-participant and by-item random intercepts, following the same rationale as the main model. The first model included both the relatedness factor and the centred data for each dimension of the IRI questionnaire as fixed effects (Personal Distress, Perspective Taking, Empathic Concern and Fantasy). Centring the data for each dimension allowed to reduce the risk of multicollinearity. The second model also included the relatedness factor as fixed effect, as well as centred data for each dimension of the PSIQ-F as fixed effects (Vision, Sound, Smell, Taste, Touch, Body and Emotion). The results from these additional models are reported in Appendix B (p. 331).

╓ Finally, cut-offs were used to exclude outliers. These cut-offs were the same for the two experiments that relied on reaction times (i.e., Experiments ╖ 2 and 3 ╓). First, the data from 5 participants were excluded from the analyses due to a total response accuracy below 90% in order to keep participants with a number of measure points as close as possible to 20 trials per experimental condition for the analysis of reaction times (see also the conclusion from Diependaele et al., 2012). Data from another participant were also excluded due to a missing recording of the dialogue. Finally, the data from 40 participants were analysed. This sample size is in line with previous studies that used the lexical decision task as a measure of the relatedness between pairs of written words (Armstrong & Plaut, 2008; Brunellière et al., 2017; Shelton & Martin, 1992). Regarding the reaction times of the 40 remaining participants, incorrect responses (0.69%) and decision latencies longer than 1503 ms (5 standard deviation above the grand mean, for a similar approach, see Baayen et al., 2003) were excluded from data analysis (0.56%). The mean error rates for the related and unrelated conditions were 0.88% and 0.50% respectively. As our hypotheses did not concern error rates and as they were below the 5% threshold, no statistical analysis was conducted on them. ╖

2.2 Results

Figure 13a shows the results of Experiment 2. There were no significant differences in the lexical decision latencies between the pairs of words associated with the same picture during the dialogue (M = 619 ms, SD = 150 ms) and the pairs of words associated with two distinct pictures (M = 617 ms, SD = 150 ms), b = 2.67, SE = 6.37, χ^2 (1) = 0.18, p = .68. Concerning the analysis of the questionnaires, we found a significant interaction between the relatedness factor and the Personal Distress dimension of the IRI questionnaire (b = -4.85, SE = 1.99, χ^2 (1) = 5.96, p = .02). The b value suggests that this significant interaction come from the fact that an increase in the scores of personal distress is associated with a decrease in the reaction times to

Figure 13. *Reaction times as a function of the relatedness factor (A) and scores of Personal Distress (B) in Experiment 2.*



targets of pairs of words referring to the same picture (i.e. faster RT), but an increase in the reaction times to targets in pairs of words referring to different pictures (see Figure 13b). However, no significant interaction was found in the model including the data of the PSIQ-F and the relatedness factor.

2.3 Discussion

The aim of Experiment 2 was to investigate whether new lexico-semantic relationships created during a dialogue could change already existing lexico-semantic representations immediately after dialogue. To do so, a new dialogue task (IAR task) was created during which participants discussed tangram pictures with a confederate to reach an agreement on the best name to give to each picture. Crucially, the confederate had to refer to some pictures using two semantically unrelated words, thus creating a new lexico-semantic relationship between the words associated with the same visual representation. Immediately after dialogue, participants performed a lexical decision task coupled with a semantic priming paradigm. We expected to find a relatedness effect such that pairs referring to a common picture during dialogue elicit faster reaction times than pairs referring to two different pictures. However, the analysis of the data in the lexical decision task did not reveal any significant relatedness effect and our only significant result concerns the interaction between the relatedness of our pairs of words during dialogue and the personal distress dimension of the IRI questionnaire. According to Davis (1980) who first created this questionnaire, personal distress refers to a feeling of unease or distress in reaction to another's negative experience (e.g. physical or emotional pain). It has been shown that while people with higher scores of personal distress have better memories of socially encoded information (related to another person) than of non-socially encoded information (self-related), people with lower scores of personal distress exhibit the opposite pattern (Wagner et al., 2015). Our results seem to be in line with this idea by showing that when pairs of words referring to the same picture were provided by another person during dialogue, individuals with higher scores of personal distress were faster to access a word which was preceded by the other word in the pair. However, this result is difficult to interpret as evidence of changes in pre-existing lexico-semantic representations, since the results for pairs of words referring to different pictures (i.e. words for which no new relationship has been created) produced the opposite pattern to that of pairs of words referring to the same picture. Specifically, words preceded by another word used to refer to a different picture during dialogue elicited longer reaction times in individuals with higher scores of personal distress, despite the absence of any experimental manipulation on these pairs. Therefore, social abilities such as affective empathy seem to play a role in the way people consider lexico-semantic information provided during a dialogue when recognising words presented on a screen. Yet, the pattern of results does not bring straightforward evidence of changes in pre-existing lexico-semantic representations immediately after dialogue.

To sum up, the results of Experiment 2 do not bring clear support to our hypothesis. However, it is possible that the new lexico-semantic relationship requires to be consolidated by a night of sleep to be integrated in the lexico-semantic network and produce a propagation of activation from one word to the other (Dumay & Gaskell, 2012; Gaskell et al., 2019). To test this hypothesis, we conducted a third experiment in which we explored whether the changes in pre-existing lexico-semantic representations could be observed after a night of sleep.

3. Experiment 3

Experiment 3 was similar to Experiment 2 except that participants performed the lexical decision task one day after the IAR task. The aim of this experiment was to assess the changes in preexisting lexico-semantic representations stored in semantic memory one day after dialogue. (i.e., after one night of consolidation) due to the creation of a new semantic relationship during the dialogue. The hypotheses were identical to those of Experiment 2.

3.1 Method

╔ 3.1.1 Participants

Forty-two native French speakers (37 females; 40 undergraduates and two community members; age, *M* = 22.26, *SD* = 3.64) who did not participate in Experiment ╗ 2 ╔ were paid 15€ to take part in Experiment ╗ 3 ╔. Two participants were excluded from the analysis of data (for more details, see the experimental design and data pre-processing section). As in Experiment ╗ 2 ╔, participants interacted with one of the two same confederates. More precisely, 20 participants interacted with one of the confederates while 20 participants interacted with the other one.

3.1.2 Stimuli and procedure

The stimuli and tasks were the same than those used in Experiment ╖ 2 ╓. As in Experiment ╖ 2 ╓, the mean number of repetitions of the words produced by the confederate approached 4 (*M* = 3.98, *SD* = 0.24), confirming that the required number of repetitions were produced by the confederate. The mean percentage of disagreement given by the participants to the words proposed by the confederate was relatively low with 5.67 % of disagreement (*SD* = 4.60), showing the participants easily reached an agreement with the confederate. As in Experiment ╖ 2 ╓, the confederate had to disagree in one trial per block, but was able to choose on which trial to disagree with their partner keep the interaction as naturalistic as possible.

Contrary to Experiment ╖ 2 ╓, the IAR task was followed by the Tone Deafness Test to keep the timing of session 1 constant in Experiment ╖ 3 ╓ (see Figure 10). During the session 2, which took place the next day, participants performed the lexical decision task with the semantic priming paradigm one day after the IAR task. After this task, participants filled in the same questionnaires (IRI questionnaire and PSIQ-F questionnaire) as those used in Experiment 1. None of the participants spontaneously reported being aware of the presence of a confederate.

3.1.3 Data analysis and pre-processing

The same approach as in Experiment ╖ 2 ╓ was used to build the statistical models. Applying this rationale led to the same set of fixed and random effects as in Experiment ╖ 2 ╓. The results from the main model are reported in Tables 6a and 6b and the results of additional models can be found in ╖ Appendix B (p.333). ╓ In addition, the same cut-offs than those used in Experiment ╖ 2 ╓ were used to exclude data. Thus, the data from two participants were excluded from the analyses due to a total response accuracy below 90%. Regarding the reaction times of the 40 remaining participants, incorrect responses (0.83%) and decision latencies longer than 1513 ms (5 SD above the grand mean) were excluded (0.50%). ╖

Table 6a. *Fixed effects, Experiment 3*

Fixed effect	χ^2	**DF**	***b***	***SE***	***p***
Relatedness_factor	0.52	1	-4.72	6.55	0.47

Table 6b. *Random effects, Experiment 3*

Random effects	**Variance**	***SD***
By-participant		
Intercept	6563	81.01
By-item (targets)		
Intercept	1688	41.09
Residual	16914	130.05

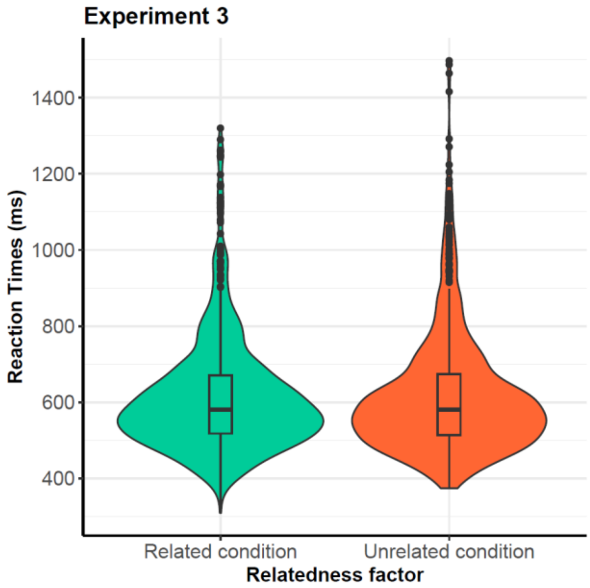
The mean error rates for the related and unrelated conditions were 1.19% and 0.48% respectively. As our hypotheses did not concern error rates and as they were below the 5% threshold, no statistical analysis was conducted on them.

3.2 Results

Results of Experiment 3 are presented in Figure 14. As in Experiment 2, we found no significant difference between pairs of words referring to the same picture during the dialogue (*M* = 615 ms, *SD* = 152 ms) and those referring to different pictures during the dialogue (*M* = 615 ms, *SD* = 164 ms), *b* = -4.72, *SE* = 6.55, $\chi^2(1) = 0.52$, *p* = .47. Regarding the models that included either the IRI or the PSIQ-F questionnaire, no interaction effect was found with the relatedness factor on the lexical decision latencies (see Appendix B, p. 333).

An additional analysis was conducted on the effect of a night of sleep on potential changes in participant’s lexico-semantic relationships. Specifically, we analysed the time of the day at which participants performed the dialogue task as an additional fixed effect in the model, together with the relatedness factor. Indeed, we were wondering whether participants performing the dialogue at different times of the day would benefit differentially from the night of sleep depending on the awakening time before sleep consolidation.

Figure 14. *Reaction times as a function of the relatedness factor in Experiment 3*



However, no significant effect of participation time was found, nor did this factor interact with relatedness (b = -1.03, SE = 2.50, $\chi^2(1)$ = 0.17, p = .68).

3.3 Discussion

The aim of Experiment 3 was to explore whether new lexico-semantic relationships created during dialogue between pre-existing lexico-semantic representations could persist after one night of sleep following dialogue thanks to consolidation processes. The dialogue task was the same as in Experiment 2, with an IAR task during which pairs of words were associated with either a common visual representation or two distinct visual representations (i.e. tangram pictures). One day later, participants performed a lexical decision task with a semantic priming paradigm and we compared the reaction times between the two types of pairs of words. However, the analysis did not reveal any significant relatedness effect. In addition, in comparison with Experiment 2 the analyses of the questionnaires did not reveal any significant interaction between the relatedness factor and the various dimensions of the IRI and PSIQ-F questionnaires. Finally, the time at which participants performed the dialogue task, which could reflect the delay between the dialogue phase and the night of sleep did not significantly interact with the relatedness effect.

From these results, we cannot conclude about a potential consolidation of the new lexico-semantic relationships created during dialogue. However, it is possible that the new relationships created during dialogue, even if integrated into the semantic network through offline consolidation processes, are too weak to produce a strong and consistent spreading of activation observable in our semantic priming paradigm with such a short SOA. Thus, in order to determine if the new lexico-semantic relationship could exist, even in a weaker state, we performed a fourth experiment. This experiment sought to assess if the degree of relatedness between our pairs of words could have changed after dialogue by the use of a more explicit task. Specifically, we changed our implicit assessment of the semantic relationships (i.e. lexical decision task) by a semantic relatedness judgement task in which participants were asked to focus on the semantic relatedness between our pairs of words. With this task, we aimed at amplifying the activation into the lexico-semantic network and highlighting weaker connections.

4. Experiment 4

The procedure of Experiment 4 was similar to the one used in Experiment 3, except that the visual lexical decision task was replaced by a visual semantic relatedness judgement task. In this task, participants were asked to explicitly focus on the degree of semantic relatedness between pairs of words. The goal was to determine if changes in pre-existing lexico-semantic representations could be observable when attentional processes could amplify the degree of activation into the lexico-semantic network (Dehaene et al., 2006). By assessing the new relationships one day after dialogue, we expected to find higher scores of semantic relatedness for the pairs of words that have been associated with the same tangram picture during dialogue than for pairs of words referring to two distinct tangram pictures. Crucially, a control group was also added in Experiment 4 to provide clear evidence that dialogue directly changes pre-existing lexico-semantic representations in semantic memory. More precisely, this control group performed the semantic relatedness judgement task without having been exposed to the dialogue task. We therefore expected to find a significant interaction between the relatedness factor (related vs. unrelated condition) and the group factor (i.e. exposed to dialogue vs. control group). In particular, the relatedness scores for the pairs of words referring to the same tangram picture should differ from those referring to different tangram pictures, but only for the group exposed to the dialogue task. Finally, a word-picture relatedness judgement task was performed by the participants in order to ensure that the group exposed to the dialogue task actually learned the associations between each word-picture pair during the IAR task.

4.1 Method

4.1.1 Participants

Forty French native speaking students (all undergraduates; 33 females; age, *M* = 20.50, *SD* = 2.43) were paid 15€ to be a part of the experimental group (i.e., group exposed to the dialogue). The overall recruitment procedure was the same as for Experiments 2 and 3. As in Experiments 2 and 3, participants interacted with one of the two same confederates. More precisely, 21 participants interacted with one of the confederates while 19 participants interacted with the other one. One participant was excluded from the analysis of data (for more details, see the experimental design and data pre-processing section). For the control group, forty native French speakers (33 females; age, *M* = 21.15, *SD* = 2.25) were recruited via prolific and social media. Participants in the control group, who performed only the semantic relatedness judgment and word-picture relatedness judgment tasks, were paid 5€ for their participation. The rest of the recruitment procedure was the same as for Experiments ╗2 and 3╔.

4.1.2 Stimuli

The stimuli in the IAR task were the same as those in Experiments ╗2 and 3╔. The 20 pairs of words referring to the same tangram picture and the 20 pairs of pairs of words referring to different tangram pictures during the dialogue were presented in the semantic relatedness judgment task. Twenty highly semantically related pairs of words were also selected from the stimuli of Brunellière and Bonnotte (2018) to make sure that the participants would use all the points on the scale of the semantic relatedness judgment task. Their mean semantic relatedness score was 6.47 on a 7-point scale (*SD* = 0.58) from the study of Brunellière and Bonnotte (2018). The mean lexical frequency and mean number of letters of the 20 highly semantically related pairs was matched with those of the two other experimental conditions across the experimental lists (see Table 7 for more details on the psycholinguistic properties). Thus, in total, 60 trials were performed in the semantic relatedness judgment task, with 20 trials in each critical condition (related vs. unrelated condition) and 20 highly semantically related pairs of words.

Two sets of 40 word-picture pairs were used in the word-picture relatedness judgment task. The aim of this task was to check that participants in the experimental group had learnt the associations between each of the two words used to name the same picture during the IAR task and their associated pictures, in comparison with participants in the control group who did not perform the IAR task. The first set of word-picture pairs consisted of the 20 pictures exposed during the dialogue as stimuli in the related condition of the IAR task. These 20 pictures were presented twice, each time using one of the two words with which the picture had been associated during the dialogue (i.e. in total, 40 word-pictures pairs). As in the first set, ╗

Table 7. *Psycholinguistic properties of pairs of words of the semantic relatedness judgement task for experimental lists 1 and 2 separately, Experiment 4*

		Lexical frequency		Number of letters		Number of phonemes		Number of orthographic neighbours		Number of phonological neighbours	
		Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
List 1	**Primes**	45.40	88.84	5.90	1.59	4.40	1.43	3.10	3.18	6.60	7.17
	Targets	28.08	44.07	6.20	1.20	4.60	1.27	3.20	4.40	9.75	10.40
List 2	**Primes**	50.40	129.17	6.10	1.52	4.40	1.31	4.60	6.42	8.45	9.13
	Targets	30.78	45.65	6.20	1.77	4.60	1.53	3.60	5.43	9.15	8.62
Highly semantically related	**Word 1**	44.01	76.37	6.10	1.52	4.25	1.37	3.05	3.75	10.45	8.79
	Word 2	23.53	34.94	6.30	2.66	4.85	2.23	2.95	4.24	8.75	9.95

Note. All properties were extracted from New et al. (2004); lexical frequency is based on movie subtitle corpora.

the pictures were presented twice in the second set and these word-picture pairs were created by combining 20 other pictures exposed during the dialogue with 40 new words. In other words, these pairs were never presented as words referring to pictures during the dialogue. Since half of the new words were words which had not been given by the participants enrolled in our database of pictures, these word-picture pairs were not related. The other half of new words were words which had been given by only one participant among the 84 respondents of the new database so that the word remained a plausible word to name the picture. The second group of word-picture pairs was created to make sure that participants could use the scale of the word-

Table 8. *Psycholinguistic properties of words presented in the word-picture relatedness judgement task corresponding to the primes and targets for experimental lists 1 and 2 separately, Experiment 4*

		Lexical frequency		Number of letters		Number of phonemes		Number of orthographic neighbors		Number of phonological neighbors	
		Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
List 1	Word 1. related	45.40	88.84	5.90	1.59	4.40	1.43	3.10	3.18	6.60	7.17
	Word 2. related	28.08	44.07	6.20	1.20	4.60	1.27	3.20	4.40	9.75	10.40
	Word 1. unrelated	44.72	53.00	5.90	2.07	4.45	1.76	2.95	3.68	9.55	8.67
	Word 2. unrelated	34.02	51.51	6.10	2.02	4.35	1.39	4.15	5.91	9.85	10.39
List 2	Word 1. related	50.40	129.17	6.10	1.52	4.40	1.31	4.60	6.42	8.45	9.13
	Word 2. related	30.78	45.65	6.20	1.77	4.60	1.53	3.60	5.43	9.15	8.62
	Word 1. unrelated	44.72	53.00	5.90	2.07	5.45	1.76	2.95	3.68	9.55	8.67
	Word 2. unrelated	32.66	44.48	6.40	1.73	4.70	1.78	2.85	3.87	6.80	6.45

Note. All properties were extracted from New et al. (2004); lexical frequency is based on movie subtitle corpora.

picture relatedness judgment task accurately. The 40 words of the first group were matched with those of the second group in terms of lexical frequency and number of letters (see Table 8 for more details on the psycholinguistic properties). ╗

4.1.3 Design

Experimental lists

The distribution of the 40 pairs of words associated with tangram pictures (i.e., 20 pairs of words referring to the same tangram picture and 20 pairs of words referring to two different tangram pictures) into two lists was kept constant across the IAR and the semantic relatedness judgment tasks. The 20 highly semantically related pairs of words were also added in both experimental lists of the semantic relatedness judgment task.

Regarding the word-picture relatedness judgment task, each list was composed of the 20 pairs of words referring to the same picture and these pictures were presented twice using one of the two words. In each list, the second set of the forty word-picture pairs was added.

4.1.4 Procedure

Experimental Group

Session 1

The procedure for day 1 was the same as in Experiment ╗ 3 ╔, with participants performing the IAR task first and then the Tone Deafness Test. Dialogues recordings were checked to make sure that words were repeated enough times by the confederate. The mean number of repetitions of the words produced by the confederate approached 4 (M = 3.99, SD = 0.17), confirming the required number of repetitions was produced by the confederate despite the naturalistic setting of the task. Regarding the level of disagreement provided by the participant on the words proposed by the confederate, the mean percentage was 5.92% (SD = 4.67). As in Experiments ╗ 2 and 3 ╔, this value was low, showing that an agreement was reached easily. Disagreement rules followed by the confederate were identical to Experiments ╗ 2 and 3 ╔.

Session 2

On day 2, participants came back to perform the semantic relatedness judgment task and the word-picture relatedness judgment task respectively. Both were presented using an online survey administered through LimeSurvey 2.06 (LimeSurvey GmbH, n.d.). Participants first received written instructions about the semantic relatedness judgment task and two examples describing the use of the Likert scale. Two training trials were then proposed before proceeding with the 60 experimental trials presented in random order. In each trial, pairs of words were presented one at a time. Words were displayed side by side on a white background and the 7-point Likert scale was positioned below them. Participants were asked to rate the degree of semantic relatedness for each pair of words before going on the next trial. At the end of the task, they were informed that they would have to do a second task, corresponding to the word-picture relatedness judgment task. This task followed the same structure as the semantic relatedness judgment task with the presentation of written instructions, two examples, and two training trials after which participants performed 80 experimental trials. In each trial, the picture and the word were presented one above the other on a white background. The 7-point Likert scale was placed above the word-picture pair and participants were asked to rate the degree of relatedness between the picture and the word. The order of completion of both tasks was not counterbalanced. The reason was because the main question concerned the new semantic relationship between the pairs of words by using the semantic relatedness judgment task, and the second task was only a control one. Moreover, we preferred to keep the semantic relatedness judgment task first to prevent any risk that the word-picture relatedness judgment task might influence the semantic relatedness judgment results between pairs of words. The completion of both tasks took approximately 15 minutes and was followed by the completion of the IRI and PSIQ-F questionnaires (Braun et al., 2015; Ceschi & Pictet, 2018). None of the participants spontaneously reported being aware of the presence of a confederate.

Control Group

Participants in the control group completed the semantic relatedness judgment task between the pairs of words, the word-picture relatedness judgement task and the two questionnaires (IRI and PSIQ-F) via an online survey. Both were presented using an online survey administered through LimeSurvey 2.06 (LimeSurvey GmbH, n.d.) and followed the same structure as that used for the experimental group.

4.1.5 Data analysis and pre-processing

Experimental design and data pre-processing

Data from the semantic relatedness judgment task were analysed in two steps. We first checked that participants from both the experimental and the control groups used the whole scale to rate the three types of word pairs. Both groups rated the highly semantically related pairs of words as accurately highly semantically related (experimental group, $M = 5.99$, $SD = 1.33$, control group, $M = 5.82$, $SD = 1.43$). After removing the data from the words that were not presented in the IAR task (as these were not of interest in the study), we ran a linear mixed-effects model using the lme4 package in R (Baayen et al., 2008). Random effects included by-participants random intercepts, by-participants random slopes corresponding to relatedness, by-item random intercepts and by-item random slopes corresponding to the relatedness factor and group. By-item intercepts were represented by the target words of each pair. The dependent variable was the centred participants' scores on the scale of the semantic relatedness judgment task between pairs of words. There were two independent variables with the relatedness factor (as in Experiments ╗ 2 and 3 ╔) and the group. Therefore, the fixed effects were the relatedness of the pairs of words during the dialogue, either related or unrelated, the group, experimental vs. control, and the interaction between these two factors. In line with our specific hypotheses on each group, Bonferroni-corrected pairwise comparisons were performed. The results from this main model are reported in Tables 9a and 9b. Two additional models were created to investigate the modulation effects of personality traits on the ratings from the semantic relatedness judgment task. These models both included by-participant and by-item random intercepts. The first model included both the relatedness of the pairs of words, the group and the centred data for each dimension of the IRI questionnaire as fixed effects (Personal Distress, Perspective Taking, Empathic Concern and Fantasy). The second model also included the relatedness of the pair of words as fixed effect, as well as centred data for each dimension of the PSIQ-F as fixed effects (Vision, Sound, Smell, Taste, Touch, Body and Emotion). ╗ The results of additional models can be found in Appendix B (p.335).

╔ Data from the word-picture relatedness judgment task were analysed the same way ╗

Table 9a. *Fixed effects of the semantic relatedness task, Experiment 4*

Fixed effect	χ^2	DF	*b*	*SE*	*p*
Relatedness_factor	1.20	1	-0.11	0.10	0.27
Group_participant	0.04	1	-0.04	0.19	0.84
Relatedness_factor * Group_participant	12.63	1	0.50	0.14	<0.01*

Table 9b. *Random effects of the semantic relatedness task, Experiment 4*

Random effects	Variance	*SD*
By-participant		
Intercept	0.62	0.79
Relatedness_factor	0.26	0.51
By-item		
Intercept	0.55	0.74
Relatedness_factor	0.02	0.16
Group_participant	0.06	0.24
Residual	1.38	1.18

╓ to ensure that words referring to the same picture had been learnt correctly with their associated pictures. Random effects included intercepts of participants and items, represented by the words of each word-picture pair (Barr et al., 2013), by-participants random slopes corresponding to word-picture relatedness and by-item random slopes corresponding to group. The dependent variable was the centred participants' scores on the scale of the word-picture relatedness judgment task. The fixed effects were the two factors, the word-picture relatedness, either word referring or not to a picture during the dialogue, and group, experimental vs. control. The results from this model are reported in Tables 10a and 10b. No further analyses were performed as the word-picture relatedness judgment task was only a control task.

Importantly, although the participants performed the word-picture relatedness judgment task after the semantic relatedness judgment task, we report the results from the word-picture relatedness judgment task first in the following section. Indeed, this task was simply a control. We then report the results from the semantic relatedness judgment task, as these enable us to test our hypothesis directly. ╖

Table 10a. *Fixed effects of the word-picture relatedness task, Experiment 4*

Fixed effect	χ^2	**DF**	***b***	***SE***	***p***
WP_relatedness	178.72	1	-2.68	0.20	<0.01*
Group_participant	7.33	1	0.36	0.13	0.01*
WP_relatedness * Group_participant	19.86	1	-0.76	0.17	<0.01*

Table 10b. *Random effects of the word-picture relatedness task, Experiment 4*

Random effects	**Variance**	***SD***
By-participant		
Intercept	0.25	0.50
WP_relatedness	0.39	0.63
By-item		
Intercept	1.11	1.05
Group_participant	0.17	0.41
Residual	1.98	1.41

4.2 Results

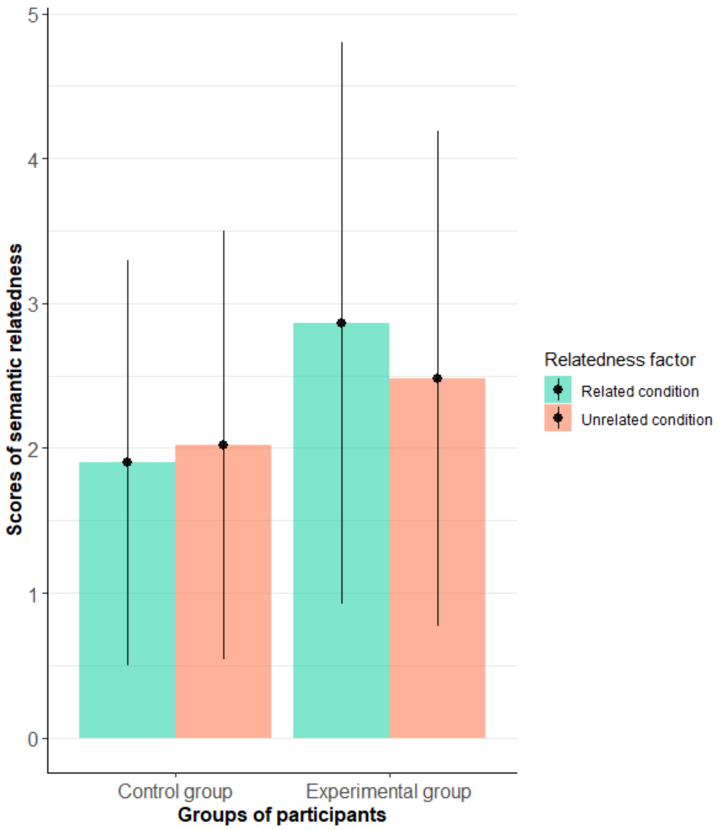
The analysis of word-picture relatedness scores revealed a significant interaction between the word-picture relatedness factor and the group factor. Specifically, there were higher word-picture relatedness scores for word-picture pairs associated during the dialogue than pairs not associated during the dialogue, and the *b* value suggests that this pattern is mainly observed in the experimental group (b = -0.76, SE = 0.17, χ^2 (1) = 19.86, $p < .001$; see Table 11).

As regards with our main analysis with the semantic relatedness judgement scores, we found a significant interaction between the relatedness factor (related vs. unrelated) and the group factor (experimental group vs. control group), $b = 0.50$, $SE = 0.14$, $\chi^2(1) = 12.63$, $p < 2.2e\text{-}16$; see Figure 15). Pairwise comparisons revealed that semantic relatedness scores for pairs of words associated with the same tangram picture during the dialogue were higher than those of words associated with distinct tangram pictures. This significant difference was however only present in the experimental group but not the control group (experimental: b = -0.38, $SE = 0.11$, $t(71.9) = -3.52$, $p = .0046$; control: $b = 0.10$, $SE = 0.11$, $t(71.9) = 0.97$, $p = 1.00$). As in Experiment 3, an additional analysis including participation time was conducted but no significant interaction with the relatedness factor was found ($b = 0.08$, SE = 0.05, $\chi^2(1) = 2.70$, $p = .10$).

Table 11. *Mean and standard deviation for the score of word-picture relatedness, Experiment 4*

	Experimental group		Control group	
	Mean	SD	Mean	SD
Related in the dialogue	6.13	1.23	5.28	1.93
Unrelated and not presented in the dialogue	2.70	2.02	2.60	2.04

Figure 15. *Semantic relatedness scores as a function of the groups of participants and the relatedness factor in Experiment 4. Note. Vertical bars correspond to standard deviation. Ns: non-significant. The asterisk denotes significance.*



4.3 Discussion

Experiment 4 aimed at investigating whether changes in pre-existing lexico-semantic representations occurring after a night of consolidation could be evidenced with a task in which attentional processes amplify the degree of activation into the lexico-semantic network. To do so, we used an explicit task in which participants had to focus on semantic relationships between pairs of words presented in the visual modality without time pressure. Specifically, after participants performed the same IAR task as in Experiment 2 and 3, they were asked to take part in a visual semantic relatedness judgement task. An additional word-picture relatedness task was also performed by the participants and allowed us to confirm that they had correctly learned the word-pictures associations created during the IAR task. This result implied that participants were sufficiently involved and paid sufficient attention during the IAR task to encode the new mappings between words and their new visual representation.

The results of this experiment revealed that pairs of words used to refer to the same tangram picture during dialogue were rated as significantly more semantically related than pairs of words used to refer to distinct tangram pictures. Yet, this result concerned only the participants who took part in the dialogue task and was not found with participants in the control group, who rated both types of pairs the same way. Altogether, these results are in line with our general hypothesis. Specifically, we have evidenced that dialogue could produce changes in the already existing lexico-semantic representations after one night of sleep promoting offline consolidation, although these changes are probably too weak to produce a strong spreading of activation into the lexico-semantic network and can consequently only be observed when attention is focused on the degree of semantic relatedness. In the next section, we will discuss these results together with the results from the previous experiments.

5. Discussion of Chapter 5

The aim of the present set of experiments was to examine to what extent new lexico-semantic relationships created during a dialogue changed already existing representations stored in long term memory after dialogue. We conducted three experiments in which participants were asked to take part in a new dialogue task, the Interactive Agreement Referential task (IAR), during which partners interacted via an audio-conference device to reach an agreement as to how tangram pictures presented on their screens had to be named. Crucially, one of the partners was a confederate who followed scripts in which each picture was named using either one word or two semantically unrelated words. Thanks to this manipulation, we created two types of pairs of words. The first half of them were related to the same visual representation during dialogue, thus creating the "related condition". By doing so, we took advantage of the new word-visual representation mappings during dialogue to create new lexico-semantic relationships between pairs of semantically unrelated words. The other types of pairs of words were words associated to two distinct pictures during dialogue, thus creating the "unrelated condition". We hypothesized that the creation of a new semantic relationship between pairs of known words used during dialogue could lead to the creation of a new lexico-semantic relationship between already existing representations. To test this hypothesis, participants of Experiment 2 and 3 performed a visual lexical decision task with a semantic priming paradigm thanks to which we tested whether changes in already existing lexico-semantic representations were associated with the establishment of strong connections into the lexico-semantic network. Participants in the Experiment 2 were tested immediately after dialogue in order to determine whether the new lexico-semantic relationship created during dialogue could impact the lexico-semantic representations stored in semantic memory immediately after dialogue. Participants in the Experiment 3 were tested one day after dialogue in order to determine whether the new lexico-semantic relationship has been integrated into long-term lexico-semantic representations after one night of sleep promoting offline consolidation. While these two experiments aimed at evidencing new strong connections by using an implicit task as assessment, Experiment 4 was created in an attempt to evidence the establishment of new weak yet reachable connections that could have been consolidated after one night of sleep. To do so, participants performed a semantic relatedness judgement task thanks to which attentional processes were meant to amplify the degree of activation into the network.

In light of the results from our three experiments, it seems that our hypothesis is only partially validated. We were indeed able to evidence a clear pattern of relatedness between the pairs of words associated with the same picture during dialogue for participants of Experiment 4. More precisely, participants who took part in the IAR task rated the pairs associated with the same tangram picture as more semantically related than pairs associated with two different tangram pictures. Crucially, this pattern was not found in participants who did not take part in the IAR task. These results support the idea that new lexico-semantic relationships created during dialogue have changed the already existing lexico-semantic representations stored in semantic memory, at least after one night of sleep. However, the findings from Experiment 2 and 3 contrasted with this idea, as they did not show a clear impact of dialogue on already existing lexico-semantic representations stored in semantic memory. Taken together, our findings support the idea that dialogue has an impact on lexico-semantic representations, but they also bring some interesting outcomes as to the way we can assess those changes. The implications of these results are discussed in the following sections in light of the literature on access to long-term lexico-semantic representations.

5.1 How to access the changes in pre-existing lexico-semantic representations?

The current study is the first to show that dialogue can impact the organization of the lexico-semantic network. Previous studies have demonstrated that multiple encounters with a word-meaning mapping can reinforce the relationship between those two in the lexico-semantic network (Betts et al., 2018; Rodd et al., 2012; 2016). Importantly, Betts and collaborators (2018) pointed out that for a repetition to be useful, the information had to be spaced over time instead of massed. In line with this study, we created scripts of dialogue in which the repetitions of the word-visual representation mappings were spaced by dialogue turns, leading to four repetitions. Specifically, participants were prompted by the confederate to speak after each repetition so that the following repetition would occur at an interval dictated by turn-taking. It has to be noted that such spacing did not provide long-term spacing, as task-oriented dialogue tends to prompt fast exchanges. Thus, the impact of such turn-taking spaces will have to be explored in further studies in order to confirm its role in changes in pre-existing lexico-semantic representations.

The present work also provides a new dialogic task which constitutes an ideal setting for the abstraction of the new semantic relationships between pairs of words. Indeed, by using pairs of words to refer to the same visual representation, and by varying the order of presentation of each word of the pair to refer to the picture, we aimed at creating not only an associative relationship between the pairs of words, but also a semantic one. Importantly also, the modality through which these words were conveyed during dialogue (i.e., the auditory modality) was different from the modality through which the relationship between the words was assessed (i.e., written modality). This manipulation allowed us to maximise our chances to observe changes in the semantic (but not episodic) memory. In line with the findings from the word-meaning priming paradigm, Experiment 4 highlighted a significant impact of our dialogue on the semantic relatedness judgement task, even in the cross-modal conditions. Our findings thus shed light on the importance of taking into account lexico-semantic representations in research domains at the interface of dialogue and memory. However, the exact nature of what has changed in the lexico-semantic network must be questioned.

It has been shown that lexical decision task coupled with a semantic priming paradigm was more able to evidence strong connections into the lexico-semantic network than weak connections (Yap et al., 2016), thus excluding the possibility that the observed lexico-semantic relationships are of this nature. In addition, the findings of changes in the degree of lexico-semantic relationship when attention was focused on the semantic properties of the relationship are more in line with the formation of new weak connections into the lexico-semantic network. Top-down attention has been described as playing a role on the activation of networks of interconnected nodes by amplifying their level of activation in order to prioritize specific mental processes (Dehaene et al., 2006; Narhi-Martinez et al., 2023). In line with these views describing the impact of top-down attention, it has been shown that increasing the time of presentation of the words allowed for more strategical processes such as post-lexical decision processes (Hutchison, 2007; also called semantic matching by Neely et al., 1989). De Wit and Kinoshita (2014, 2015a, 2015b) also demonstrated that depending on the task used, either allowing the implication of attentional processes or not, the same prime-target pairs could lead to a significant priming effect or not. They explained their results in terms of the type of activation that is produced in the semantic network, which could vary depending on whether participants are explicitly asked to focus on the semantic properties of the words. Specifically, in a lexical decision task where the SOA is long enough to allow for strategical decisions (above 200ms; Hutchison, 2007), participants are expected to engage in more retrospective matching, that is to rely more on a reactivation of the prime after the appearance of the target (de Wit & Kinoshita, 2015b). A contrario, tasks focusing on the semantic status of an item are meant to promote more decisions based on evidence accumulation (Norris & Kinoshita, 2008) which consists in activating the semantic features of both the prime and the target and comparing them in order to accumulate evidences of overlap in the semantic features. This mechanism is supposed to take place only when semantic information is relevant for the task, as it is the case in the semantic relatedness judgement task. It is thus possible that the effect observed in Experiment 4 is due to a process of evidence accumulation between our pairs of words which have been associated to the same tangram picture, as suggested by de Wit and Kinoshita (2015b). In other words, when searching for evidence of a semantic relationship between the prime and the target one day after dialogue, under attentional control participants were able to find one in terms of visual semantic features and to conclude on a weak but existing connection into the lexico-semantic network.

As regards with Experiment 2 and 3, the lexical decision task does not require participants to search for evidence of semantic overlap. However, contrary to the comparison made by de Wit and Kinoshita (2015b), the results obtained in our lexical decision task can not be explained in terms of retrospective matching, as the SOA used in Experiment 2 and 3 has been chosen to prevent any strategic decision and to produce only a spreading of activation (i.e. SOA of 166ms). In addition, the mixed results found in Experiment 2 seem to be more in line with an impact of empathy on the encoding of the new lexico-semantic relationships (Wagner et al., 2015), but further studies should be led on this subject to better understand what happened during the interaction and why the effect disappeared after one night of sleep. As concerns with the absence of significant results in Experiment 3, we can only speculate on the possibility that the new lexico-semantic relationship added in semantic memory was too weak to propagate a strong activation from our prime to our target words.

Another possible interpretation can be found in the literature on semantic priming and lexical decision task. Stolz and collaborators (2005) suggested that in lexical decision tasks in which participants are not able to make strategical decisions on the target (i.e. short SOA and low proportion of related pairs), the magnitude of the semantic priming effect becomes highly variable across trials and participants. Specifically, results from their lexical decision task and semantic priming paradigm highlighted extremely low test-retest reliability at the level of item and participant in conditions promoting non strategical decisions. Such variability could explain why the relatedness effect was not significant despite the apparent impact of dialogue on the lexico-semantic representations stored in semantic memory. While we are not able to distinguish between these two interpretations in the current set of experiments, we think that the discussion about the variability in the magnitude of the semantic priming raises an important question as to the mechanisms at work when attempting to access changes in the lexico-semantic representations. This is all the more important given that the literature on lexico-semantic adaptations never succeeded in showing changes in already existing lexico-semantic representations by assessing word access (Rodd et al., 2012). As our study constitutes, to our knowledge, the first attempt to evidence such changes using a semantic priming paradigm with a lexical decision task, we will thus discuss in more detail the literature on the variability of the semantic priming in the general discussion of this manuscript. For the rest of this discussion however, we will focus on the significant results we have found in Experiment 4 in light of the literature on consolidation of lexico-semantic information and the model by Pickering and Garrod (2004; 2021)

5.2 Consolidation of the changes in pre-existing lexico-semantic representations

The fact that we were able to evidence changes in lexico-semantic representations one day after dialogue is in line with both the literature on word-meaning mapping (Gaskell et al., 2019; Maciejewski et al., 2020; Rodd et al., 2016) and with the idea that the activations in linguistic representations proposed by the model of Pickering and Garrod (2004; 2021). According to the latter, we could imagine that the alignment between interlocutors' linguistic representations during dialogue produces a direct activation within the linguistic representations stored in semantic memory that changes their organisation both immediately after dialogue and after a night of sleep. On the other hand, Gaskell et al. (2019) and Mak et al. (2023), proposed that encountering a word-meaning mapping creates first an episodic trace which, following a night of consolidation, changes the weight of the connections in the lexico-semantic network. By this account, the encounter of the new relationship between the pairs of words during dialogue may have produced a hippocampus-dependant episodic trace of the new lexico-semantic relationship. After a night of consolidation, the new connections between the pairs of words take on a certain weight. As this study did not aim to take a position in this debate, we focused solely on testing the adaptations in pre-existing lexico-semantic representations. We thus used an explicit task one day later to maximise our chances of observing an impact of dialogue on the relationship between pre-existing lexico-semantic representations. It could however be interesting in future studies to test for the immediate changes in order to provide further evidence in favour of one approach or the other.

5.3 Interim conclusion

In the present set of experiments, we showed that new content (i.e. a new lexico-semantic relationship) presented during dialogue impacts the organisation of pre-existing lexico-semantic representations stored in semantic memory. In this respect, our findings have important implications for dialogue research. In particular, while the collaborative approach suggests that dialogue partners build shared knowledge that allows us to adjust the content of their utterances to each other, our findings suggest that this shared knowledge may cause direct modifications in the pre-existing lexico-semantic representations of each interlocutor individually. The consequence of this finding is that the modifications occurring during a dialogue could not only impact subsequent interactions with the same interlocutor but also when changing of interlocutor.

While this research question emerged from word-meaning priming studies which demonstrated that adaptations of the lexico-semantic representations could be evidence after the creation of new lexico-semantic content (Fang & Perfetti, 2019; Maciejewski et al., 2020; Rodd et al., 2012), this literature also evidenced a form of adaptation without adding new content (Betts et al., 2018; Rodd et al., 2013; Rodd et al., 2016). Specifically, changes in the organisation of pre-existing meanings can occur by adapting the weights of connections between pre-existing words and meanings. In the following chapter, we consequently wanted to test whether such changes could occur as a consequence of the exposition to dialogue. It has to be noted however that in the continuation of this first set of experiments, the following chapter will test the access to lexico-semantic representations, and will consequently not make use of the word-meaning priming paradigm previously mentioned.

Chapter 6 – Changes in the organisation of semantic categories through dialogue: the typicality effect

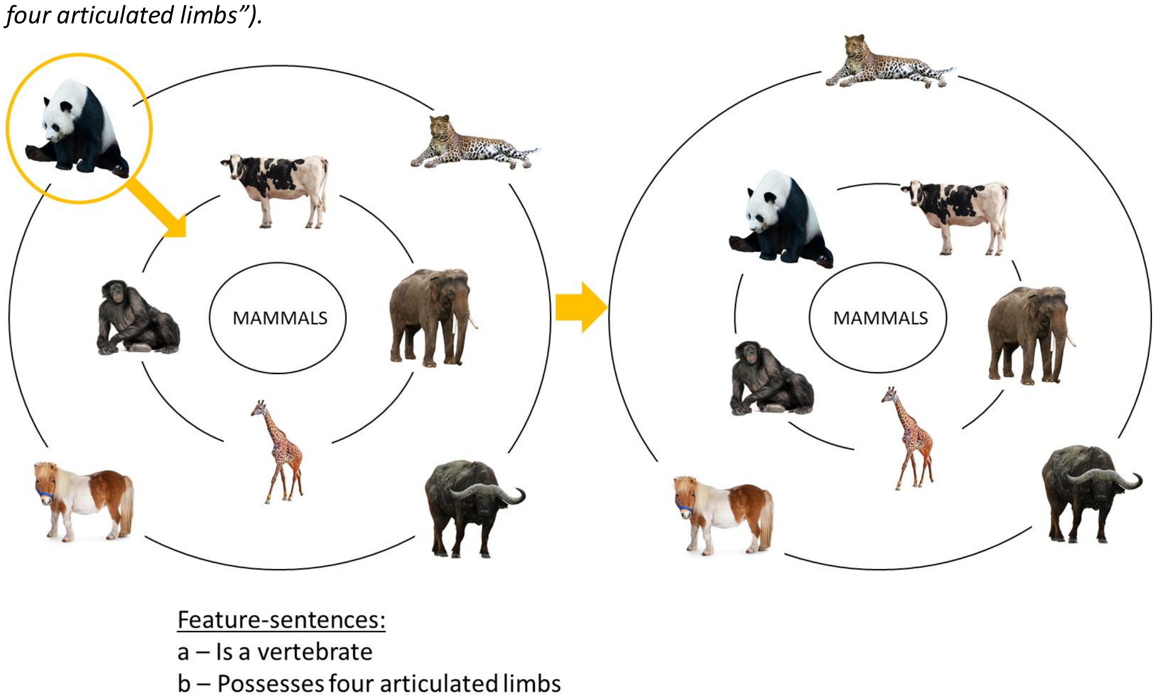
---

1. Introduction and overview of the chapter

The first set of experiments presented in Chapter 5 showed that dialogue can impact long-term lexico-semantic representations by creating new relationships between them. Although this finding required the use of an explicit measure as the new relationships were probably too weak to be evidenced with an implicit measure, we were still able to evidence the changes in long-term memory after dialogue. However, in line with our research question, it would be interesting to better identify the kinds of adaptation that may occur in the organisation of lexico-semantic representations after dialogue. While we demonstrated that it is possible to create new relationships between these representations, whether it is possible to change pre-existing ones without adding new content remained elusive. To answer this question, we focused on the findings that semantic representations are organised in categories that possess a graded structure (Rosch, 1975; Rosch and Mervis, 1975) and on the typicality effect. Specifically, we sought to determine whether lexico-semantic representations that are considered as atypical items of their belonging semantic category might become more typical after a dialogue.

To answer this question, we ran two experiments using on a new dialogue task based on the matching task already used in dialogue studies (Bangerter et al., 2020; Brennan & Clark, 1996; Clark & Wilkes-Gibbs, 1986; Duff et al., 2006; Horton & Gerrig, 2002) and called the Property-Based Matching task (PBM task). This new task was specifically designed to answer our research question, as it allowed us to reinforce the semantic relationship between atypical items and their superordinate semantic category, as well as other typical items of the category. Semantic categories have been suggested to be created on the basis of clusters of representative features (Rosch, 1973) that co-occur in the environment. Members of these semantic categories are thought to be organised according to the number and representativeness of their features. More precisely, the more representative features of its semantic category a member possesses, the more it is considered as typical, and conversely. By strengthening the association between the representative features of a category and its atypical items, it may thus be possible to reinforce the lexico-semantic relationship already existing between the atypical and the typical members of this category. For example, while ‘panda’ is considered as an atypical member of the ‘mammals’ category, its relationship with any other typical mammal (e.g. a cow) may be strengthened by reinforcing the fact that they share the same features representative of the ‘mammals’ category, such as ‘possessing four articulated limbs’ (see Figure 16). To evaluate this idea, we presented photo-realistic pictures of atypical and typical items from 10 semantic categories to dyads of naive participants (e.g. pictures of a panda and an elephant, respectively).

Figure 16. *Example of an atypical word (i.e. panda) becoming more typical after reinforcing its relationship with other typical members of the category thanks to their common properties (i.e. “is a vertebrate” and “possesses four articulated limbs”).*

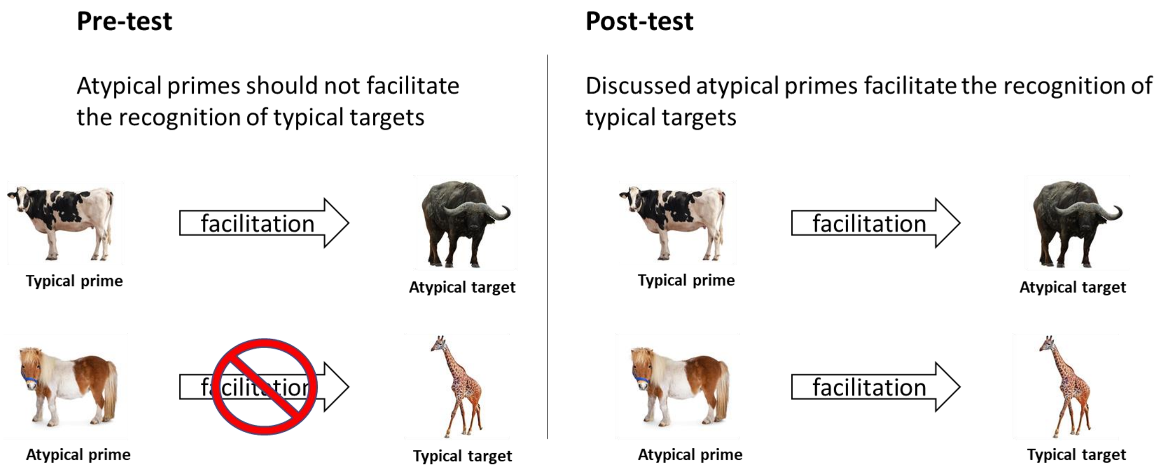


One of the participants was named ‘the director’ and had to make the other, called ‘the matcher’, guess the order in which the pictures should be placed. Crucially, the director was not allowed to name the pictures and had to refer to them by using the core features of their superordinate category (e.g. using ‘possesses four articulated limbs’ to designate the panda) that were proposed to him. Two properties for each semantic category were chosen and associated with all the subordinate items discussed during the dialogue task (i.e. typical and atypical). In line with the first set of experiments and with the study by Betts et al. (2018), each picture was associated two times with each of its two related properties during the 45-minute discussion (i.e. 4 presentations of each picture) to boost the changes in the lexico-semantic relationships.

To assess the potential changes in lexico-semantic representations, participants performed a visual semantic priming paradigm with a lexical decision task similar to the one used in the first set of experiments (see Chapter 5). The choice of keeping the lexical decision task and the semantic priming paradigm despite the non-significant results obtained in Experiments 2 and 3 comes from the idea that changes in the propagation of activation should be easier to observe when strengthening existing connections than when creating new weak connections. In addition, typicality effects have already been evidenced with a lexical decision task and a semantic priming paradigm (Brunellière & Bonnotte, 2018). To do so, the authors presented pairs of words belonging either to the same semantic category or to two distinct semantic categories. The pairs were composed of either typical or atypical primes and either typical or atypical targets. The results not only evidenced a semantic priming effect due to the category relationship, but also an interaction with the typicality status of the words. Crucially for the present set of experiments, when the target was a typical item, the semantic priming effect due to the category relationship was present only with typical primes but not atypical ones. To illustrate this result, let us take the example of a typical target word such as ‘cow’, which can be preceded either by an atypical prime of the same semantic category such as 'panda' or a typical prime such as 'elephant'. In this example, while the word 'cow' should successfully be primed by the word 'elephant' owing to their shared category membership, this effect should not be found when 'cow' is preceded by 'panda', despite their shared category membership. It is thus possible to evidence the status of a word in its semantic category (i.e. typical vs. atypical) by using such a paradigm.

Unlike the semantic priming paradigm used in the first set of experiments, participants had to perform the lexical decision task twice: first, one week before dialogue, and then immediately after dialogue (for participants taking part in Experiment 5) or one day after dialogue (for participants taking part in Experiment 6). This is due to the fact that we are this time assessing changes in pre-existing lexico-semantic relationships in comparison with the previous experiments which focused on the existence of new relationships. We therefore needed a measure of the access to pre-existing lexico-semantic relationships before the dialogue task in order to make a reliable post-test interpretation. In the pre-test of both experiments and in line with the results from Brunellière et and Bonnotte (2018), we expected pairs of words belonging to the same semantic category to elicit a significant semantic priming effect, unlike pairs of words belonging to two distinct semantic categories. We also expected to find a significant interaction between the typicality status of the prime and the semantic priming effect, with significant semantic priming effect for typical primes but not atypical primes. However, the general hypothesis for the post-test was that atypical primes should become more typical of their semantic category by being more semantically related to their superordinate category. As a consequence, we expected to find a significant semantic priming effect in the post-test, similar for both typical and atypical primes (see Figure 17). In line with the first set of experiments and with the literature on word-meaning priming (Gaskell et al., 2019; Rodd et al., 2013; Rodd et al., 2016), we expected similar results in Experiment 5 (post-test performed

Figure 17. *Expected results for the lexical decision task in the pre-test and the post-test*



immediately after dialogue) and in Experiment 6 (post-test performed after a night of sleep). While typical primes were still expected to elicit a significant facilitation of the access to the lexical status of the target when both words belonged to the same semantic category, atypical primes were expected to behave similarly and elicit a significant facilitation effect. No interaction between the category membership of the prime and its typicality was therefore expected to be found, and the word 'panda' was expected to produce a similar semantic priming effect as the word 'elephant' when they preceded a typical target such as 'cow'.

A final issue in this study concerns the change between the pre-test and the post-test and whether it might affect both partners equally. Indeed, interlocutors did not have the same instructions and, while the director had to pronounce the properties of each picture, the other had to listen to these properties. It may therefore be possible that dialogue did not have the same impact on the lexico-semantic representations of each interlocutor. This idea is investigated in the following set of experiments.

2. Experiment 5

2.1 Materials and methods

2.1.1 Participants

Forty-eight native French speaking students (38 females; age, *M* = 20.19, *SD* = 2.29) were paid 25€ to take part in the experiment. All of them had normal or corrected-to-normal vision and no history of language disorder, auditory disorder or neurological disorder. They signed an informed consent form before the beginning of the experiment and were fully debriefed after the experiment. This work was approved by the Ethics Committee of the University of Lille (reference: 2022-638-S110). Ten participants were excluded from the analysis of data (for more details, see the experimental design and data pre-processing section).

2.1.2 Apparatus

The dialogues were recorded using a Zoom H4n with a dual-entry digital recorder (one per partner).

2.1.3 Stimuli

The Property-Based Matching task (PBM)

Eighty French words were selected to be discussed by the participants. All words were selected from the 480 stimuli used in Brunellière and Bonnotte (2018). Among these words, half were considered as typical items of their respective category, and the other half as atypical items. The typicality of the words was established on the basis of the French database of Dubois and Poitou (2002). The latter provides judgements from 75 native French speakers who were asked to list, in the order in which they came to mind, the names of objects belonging to 22 semantic categories (i.e. nine categories of natural objects, 11 categories of artificial objects and two categories of activities). The words that were given more than 20% of the time were considered as typical items, while those given less than 7% of the time were considered as atypical items (see Brunellière & Bonnotte, 2018 for the criteria).

The 80 words selected belonged to 10 different semantic categories (three categories of natural objects and seven categories of artificial objects, see Appendix C). Each category was composed of four typical and four atypical words. These categories were selected from the 20 categories available in Brunellière and Bonnotte (2018) and on the basis of two additional criteria. The first one was the possibility to distinguish the various members of a category by means of a photo-realistic visual representation, as assessed by the experimenters. The words from categories such as trees, flowers, beverages and sports were therefore excluded due to the very specific knowledge required to visually distinguish each item of the category the others. The second criterion was the number of members available for each category. Four additional categories were consequently removed on the basis of their little number of items.

To ensure that participants were likely to know every discussed word, and in particular the atypical ones, the 40 selected atypical words were also selected on the basis of a pre-test performed by 12 French native speakers who were asked to rate 125 atypical words (extracted from Brunellière & Bonnotte, 2018) on two criteria. The first criterion was the familiarity of the word. Participants were asked to indicate whether or not they knew the word. If the answer was 'yes', they were invited to rate the word on the second criterion which was the knowledge of the visual representation of the word. For each of the 125 words, participants were thus asked to indicate whether they were able to visually represent the object designated by the word in their minds. If a participant did not know the word, its answer to this second question was automatically considered as a 'no' answer. A word was then selected if it obtained at least 79% of 'yes' answers in total, which represented at least 19 'yes' answers among the 24 obtained (i.e. two questions per 12 participants).

The 80 words were then associated with photo-realistic pictures found in pictures databases (i.e. 80 pictures in total). More precisely, 36 pictures came from the Bank Of Standardized Stimuli (BOSS; Brodeur et al., 2014), seven came from the POPORO database (Kovalenko et al., 2012), four from the database created by Moreno-Martinez and Montoro (2012), and the final set of 33 pictures was found on various websites providing royalty-free pictures (Adobe Stock, iStock, Freepik and others).

Forty additional filler pictures were selected from 10 semantic categories distinct from the previously mentioned ones, in order to be included in the PBM task (four filler pictures per category). Among the 10 categories, seven came from the database of Dubois and Poitou (2002) and three were created on the basis of the pictures available in the BOSS, Moreno-Martinez and Montoro, and POPORO databases. These pictures were not associated with a word, as they were not intended to be discussed nor to be presented in the Lexical Decision Task.

Twenty short sentences were created to be pronounced during the PBM task. These sentences corresponded to properties of the 10 semantic categories of interest (i.e. extracted from Brunellière and Bonnotte, 2018). Specifically, each semantic category was associated with two properties. For instance, the 'mammals' category was associated with 'possesses four articulated limbs' and 'is a vertebrate'. The properties were created on the basis of dictionary entries or features of the categories. As these properties were used during the PBM task to designate the pictures, they had to be representative of every picture of the corresponding category. The representativeness of each property was therefore pre-tested based on three criteria: (a) to which extent the property is representative of all the members of the category (from 1 to 7), (b) can the property be attributed to each of the eight pictures of the category (yes or no), (c) can the property be attributed to each of the eight words of the category (yes or no). These criteria were implemented into three distinct pre-tests, each taken by 16 French native speakers. The details of the materials (words per category, pictures and properties) and the results of all the pre-tests can be found in Appendix C. Overall, the representativeness of the properties in relation with (b) the pictures and (c) the words was high with a mean number of 'no' answers (i.e. the property is not representative of the picture) of 1.15 (SD = 2.10) and 0.93 (SD = 1.72) respectively. In addition, the mean score of representativeness of the properties in relation with (a) all members of a category is of 5.98 (SD = 0.63), which is also a high score.

Finally, four word-picture pairs were selected to serve as training pairs during dialogue. The words were picked from the semantic categories excluded during the selection process (i.e. buildings, insects, birds and fish) and followed the same pre-test procedure as the 80 other word-picture pairs. The only difference was that they were associated with only one property, as they were discussed only once during the training trial.

The lexical decision task (LDT)

The 80 words associated with a photo-realistic picture were used as primes in the lexical decision task (LDT). Although 80 additional words selected from the Brunellière and Bonnotte (2018) material were chosen to serve as targets in this task, they were not discussed during dialogue. These target words were typical words from the 10 semantic categories of interest, with eight typical words per semantic category. Each target was associated with two different semantic priming conditions, either related (i.e. belonging to the same semantic category as the prime) or unrelated. Half of the targets were systematically associated with a discussed typical prime and the other half with a discussed atypical prime, thus creating four possible types of pairs of words: typical prime with related typical target, typical prime with unrelated typical target, atypical prime with related typical target, and atypical prime with unrelated typical target. Finally, the 80 pairs of words were divided into two experimental lists so that each pair of words was presented only once per participant, but in both semantic priming conditions across participants (40 pairs of words per priming condition). The properties of primes and targets were matched in terms of frequency, length and neighbourhood between the two experimental lists (see Table 12a). Word pairs were also matched in terms of semantic relatedness and co-occurrence frequency between the two experimental lists (see Table 12b). The co-occurrence frequency was calculated with the same approach as in Brunellière et al. (2017) and in Brunellière and Bonnotte (2018). The semantic relatedness was evaluated by 20 native French speakers who did not took part in the other pre-tests.

Table 12a. *Psycholinguistic properties for typical and atypical primes and their paired typical targets in lists 1 and 2*

		Lexical frequency		Number of letters		Number of phonemes		Number of orthographic neighbours		Number of phonological neighbours		Typicality	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
List 1	Typical primes	30.04	39.13	5.95	1.70	4.30	1.03	3.00	4.29	8.20	8.11	59.07	23.67
	Typical targets paired with typical primes	18.70	26.35	6.55	2.28	5.30	2.18	2.30	4.14	5.75	8.75	41.13	17.33
	Atypical primes	3.46	2.96	6.85	1.42	5.2	1.06	1.65	2.25	2.75	3.97	2.20	1.25
	Typical targets paired with atypical primes	20.57	53.76	6.5	1.82	4.65	1.50	2.35	2.72	7.45	7.40	43.27	20.80
List 2	Typical primes	30.70	44.58	5.95	1.39	4.20	1.11	3.00	4.62	8.40	7.44	62.93	19.69
	Typical targets paired with typical primes	15.95	26.41	6.5	1.67	4.75	1.71	2.15	3.60	6.85	9.46	38.67	16.77
	Atypical primes	2.17	2.50	6.85	1.46	5.1	1.12	1.65	2.21	3.35	3.59	2.27	1.51
	Typical targets paired with atypical primes	18.50	35.36	6.55	1.93	4.80	1.77	2.30	2.94	7.35	8.94	46.40	20.90

Table 12b. *Co-occurrence frequencies and semantic relatedness for related pairs of typical and atypical words in lists 1 and 2*

		Co-occurrence frequency		Semantic relatedness	
		Mean	SD	Mean	SD
List 1	Typical prime + Typical target	0.15	0.49	5.31	1.00
	Atypical primes + Typical targets	0.20	0.41	4.20	0.99
List 2	Typical prime + Typical target	0.55	1.10	5.32	0.91
	Atypical primes + Typical targets	0.05	0.22	4.31	1.15

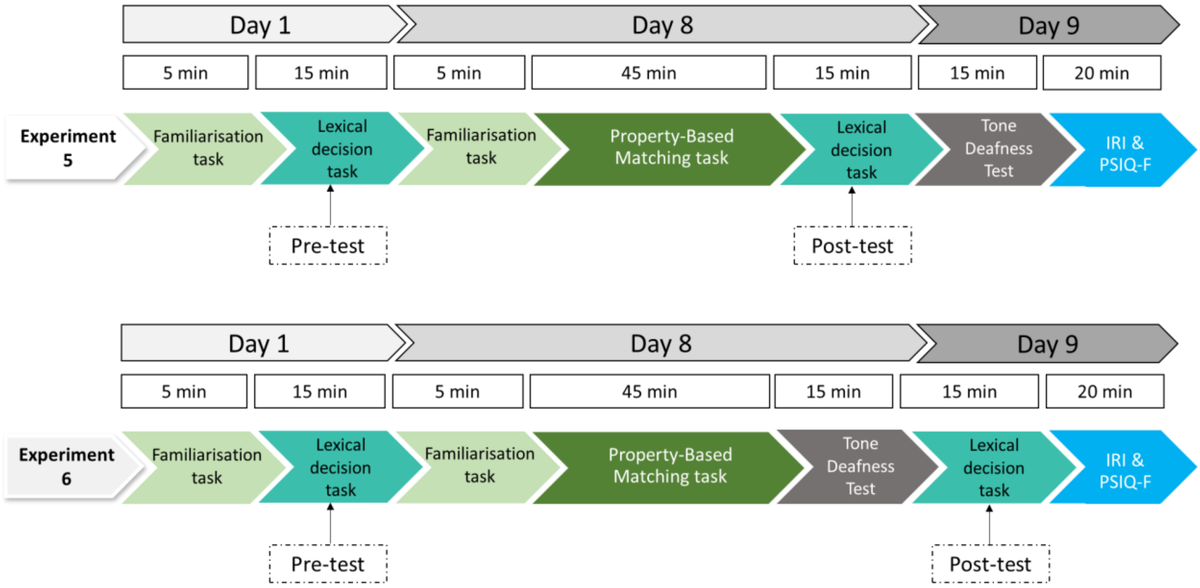
Another set of 80 unrelated pairs of words which were not discussed during the PBM task were added as fillers in each list, as well as 160 word-pseudoword pairs. The relatedness proportion was consequently equal to .25 (40 related pairs and 120 unrelated pairs) and the proportion of word-word pairs of .50. The unrelated filler pairs of words were added to prevent undiscussed prime words from always preceding pseudowords. Pseudoword targets were orthographically legal and were constructed by replacing a letter in French words different from those used in the PBM task and from those used as fillers in the LDT. Word and pseudoword targets were matched in length (number of letters, for words: M = 6.78, SD = 1.44; for pseudowords: M = 6.76, SD = 1.51).

Sixteen training trials were added following the same composition of related vs. unrelated pairs and word vs. pseudoword pairs. The related pairs of words were constituted of the four words used in the training session of the PBM task as prime words. Two of them were associated with typical words of their semantic category as targets and the two remaining words were associated with unrelated typical targets.

2.1.4 Procedure

Experiments 5 and 6 took place in three sessions over a nine-day period for methodological reasons and to keep the procedure constant across the two experiments (see Figure 18). For both experiments, session 1 took place on day 1 and served as a session of familiarization with the pictures to be shown in dialogue and for the pre-test of the LDT. However, as the familiarisation task consisted in the oral presentation of the words, it could have impacted the results of the pre-test. In order to make the comparison between the pre-test and the post-test possible, we consequently added the same familiarisation task before the post-

Figure 18. *Recap of the procedure of the two experiments. IRI: Interpersonal Reactivity Index questionnaire; PSIQ-F: Plymouth Sensory Imagery Questionnaire in French.*



test. Specifically, to keep the order of each task –except for the LDT– constant across our two experiments, the familiarisation task was always performed at the beginning of session 2.

In Experiment 5, session 2 took place on day 8 and consisted in the familiarisation task, the PBM task and the post-test. In Experiment 6, it consisted in the familiarisation task and the dialogue task as well, but the post-test was replaced by a filler task. In session 3 which took place on day 9, participants of Experiment 5 performed the filler task and answered some questionnaires while participants of Experiment 6 performed the post-test before filling the questionnaires one day after the PBM task. All three sessions were performed in the same testing room for a given participant, which was a different room for the matchers and the directors. Sessions 1 and 3 took place at different times for the two partners of a given dyad. Only session 2 was performed at the same time. The overall experiment lasted around two and a half hours for each participant.

Session 1 (day 1)

Familiarisation task

Participants were placed in a room according to the role they were randomly attributed prior to session 1. They performed the familiarisation task prior to the LDT. The stimuli for the familiarisation task were implemented on PychoPy2 (Peirce et al., 2019) and presented to the participants using the online Pavlovia platform (Bridges et al., 2020). Participants were presented 124 trials in random order (i.e. 84 discussed pictures and 40 filler pictures). A trial consisted in the presentation of a picture on the centre of the screen for 500ms. Then, the picture remained on the screen while a word was presented orally in the headset worn by the participants. This word corresponded to the picture that was presented on the screen. At the end of the oral stimulus, the picture remained on the screen for one more second, then disappeared and was replaced by a short text asking the participant to press the space bar to continue with the next trial. The instruction was for the participants to memorise the word associated with each picture. They were not given any practice trial nor any break as the experiment was self-paced and was not subject to any measure. The order of presentation of the pictures was randomised by the software.

Lexical decision task (LDT)

After the familiarisation task, participants performed a LDT with a semantic priming paradigm. Before their arrival, they were attributed one of the two experimental lists. The stimuli were presented using the E-Prime software (E-Prime 2.0.10.356, Schneider et al., 2002a; 2002b). Participants first performed 16 practice trials before proceeding with the 320 experimental trials (i.e. 160 word-pseudoword pairs, 80 unrelated filler pairs and 80 pairs of interest). They were divided into two blocks of 160 trials and a break was proposed between blocks. Stimuli were displayed on the screen in lower case and in white font on a black background. Each trial consisted in the presentation of a fixation cross for 500 ms, followed by a prime word for 150 ms. A black screen was then displayed for 16 ms (leading to a SOA of

166 ms) before the presentation of the target stimulus, which remained on the screen until the participant's response. Participants had to indicate as quickly and accurately as possible whether or not the target stimulus was a real word. They had to respond by using a button box for which the assignment of the button responses depended on their handedness. At the end of each trial, a black screen was displayed for 1,500 ms before the next trial began. The task took approximately 20 minutes to be completed. Importantly, the timings used for the trials of this task were the same as those used in the first set of experiments (see Chapter 5). However, more trials were presented in the present set of experiments, thus lengthening the duration of the task from 15 to 20 minutes.

Session 2 (day 8)

Familiarisation task

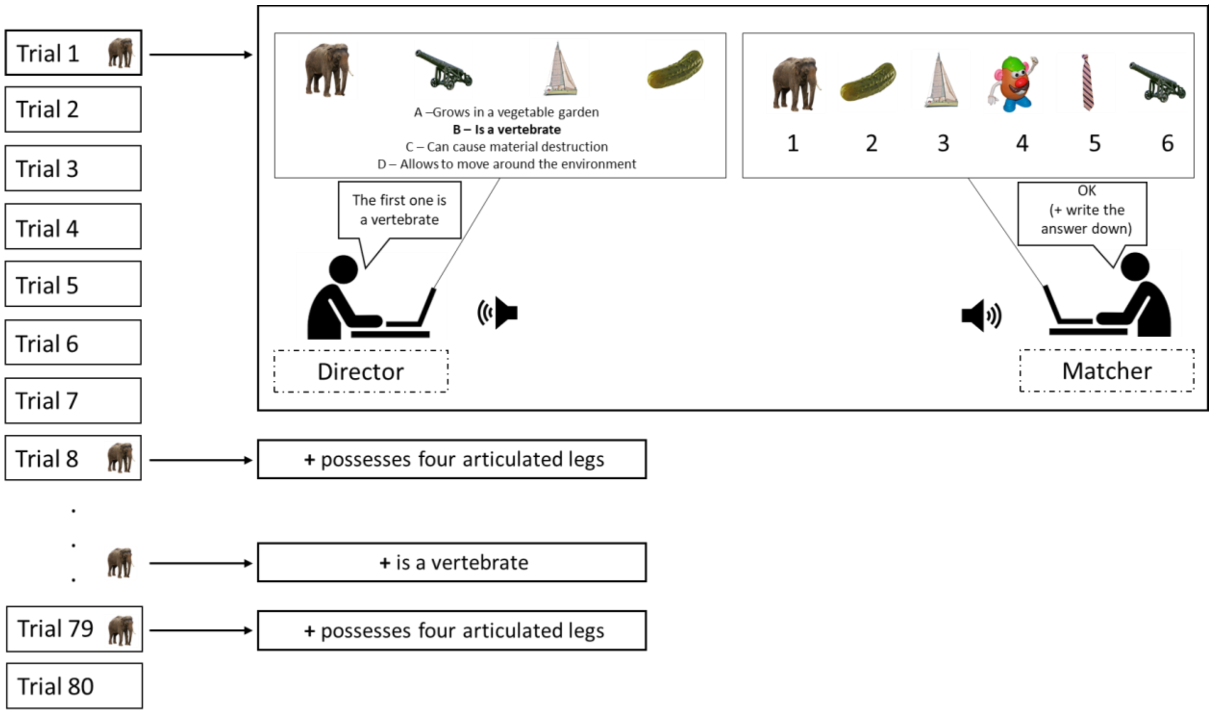
The same procedure as that used in session 1 was used in session 2.

Property-Based Matching task (PBM)

The director and the matcher were placed in their respective separate rooms. As in the first set of experiments (see Chapter 5), both participants were informed that they would have to interact with another participant located in another room using an audio-conference device and headphones, and that the discussion would be recorded for the purpose of the experiment. They were also told that they would have to discuss pictures displayed on their respective screens and that they would not see the same amount and order of pictures. They were informed of the role they would have to play during the conversation and that they would keep this role until the end of the task. The picture presentation was run on the computer screen using GNU Octave (version 6.4.0).

The PBM task was divided into 80 trials: in each trial, four of the 80 pictures associated with a prime word of the LDT were presented, with each picture being discussed four times during the PBM task (twice with each associated property). A break was provided after 40 trials. On the director's side, a trial consisted in the display of the four pictures of interest side by side in the upper half of the screen, and four sentences displayed one above the other on the lower half of the screen. On the matcher's side, the same trial consisted in the display of six pictures side by side on the middle of the screen and numbered 1 to 6 from left to right. Among these six pictures, four were the ones shown on the director's side and the two left were filler pictures. As the computers of both participants worked independently, the randomisation of the presentation order (from 1 to 4 or 1 to 6) was performed independently by the two versions of the software. For each trial, the director was instructed to give oral indications to the matcher to make him/her guess which of the six pictures was displayed on the director's screen and in which order they were displayed. To do so, the director was not allowed to name the pictures and had to use the sentences displayed on the screen. Specifically, the director had to mentally associate each sentence displayed on the screen with its corresponding picture (e.g. associating 'is a vertebrate' with the picture of the elephant) and then tell the sentences to the matcher in the order of their corresponding pictures (e.g. my first picture 'is a vertebrate'). The matcher was instructed to write the number of the picture (from 1 to 6) that corresponded to each place (from 1 to 4) on the response sheet. The matcher was allowed to produce feedback such as "ok" or "yes", but neither to ask for more details nor to say the name of the picture (see Figure 19). The addition of the two filler pictures on the matcher's side was meant to prevent him/her from guessing the fourth picture on the director's side without listening to its description. As soon as each of the four sentences had been given and each number written on the response sheet, both participants had to press the space bar to display the next trial. At the beginning of each trial, they were also highly encouraged to tell each other the number of the trial that was displayed

Figure 19. *Example of a trial and illustration of the alternance for the 'mammals' properties across the 80 trials in relation with the picture 'elephant'.*



in the bottom left corner of their screens, in order to make sure they were in sync. The partners were given a training phase to familiarise themselves with the procedure at the beginning of the task and the experimenter, who was always connected to the audio-conference device, could answer their questions at any time during the dialogue. During the dialogue, the experimenter listened to the interaction to correct the participants if they did not follow the instructions, and to write the order in which the director provided the sentences to the matcher.

The participants were randomly assigned to one of the two versions of the PBM task before the first session. Each version corresponded to one of the two possible orders of presentation of the 80 trials. The 80 trials designed for the first version of the task were thus kept constant. The order in which each trial was presented was the only factor that was changed to create the second version of the task. While these two versions were created to limit the impact of trial order on the incidental learning provoked by dialogue, we were not able to change the content of trials themselves due to the strict criteria used in terms of co-occurrence of the pictures and categories of interest within trials. First, at the category level, two pictures of the same semantic category could not occur in the same trial. Vegetables were not placed in the same trials as fruit to prevent any confusion in the properties associated with each category. Some tools were also excluded from trials where weapons were presented (e.g. the hammer could be confused with a weapon). Second, we also controlled for the number of times items of a given semantic category co-occurred in a trial with other items of other semantic categories, so that the number of co-occurrences between categories was kept between 9 and 13. For example, the number of times a mammal appeared in the same trial as a fruit was 12 across the 80 trials, and the number of times a tool appeared in the same trial as a vehicle was 10. The only exception to this rule was the co-occurrence between fruit and vegetables, which was equal to 0. At the word level, words that presented a strong associative relationship (such as 'ape' and 'banana') were not placed within the same trials. In addition, a trial was always composed of two typical and two atypical words of any of the 10 semantic categories. Finally, at the picture level, each of the 80 pictures was presented four times across the 80 trials but was never presented more than twice with another picture. This repetition made it possible to present each picture twice with each of the properties that defined their semantic category (see Figure 19). Each property was thus presented 16 times (twice with each of the eight members of their category). In addition, the presentation of the picture with its properties was alternated so that two consecutive presentations of a picture were not associated with the same property. The final step of the creation of the trials was to add two filler pictures in each trial. These filler pictures were added with the same criteria of not sharing a high degree of associative relationship with the four other pictures and not putting two filler pictures of the same semantic category in the same trial. Finally, we also avoided too much co-occurrences of the same semantic categories within trials.

Lexical decision task (LDT)

After the PBM task, the same LDT as in session 1 was performed. The experimental list and the procedure were kept constant in the pre- and post-tests.

Session 3

Filler task and questionnaires

Participants came back the following day to perform a filler task and fill in personality questionnaires. We used the same tasks as in the first set of experiments (see Chapter 5). The filler task was used once again to keep the timing of each session constant across Experiments 5 and 6 (1h15 on day 8 and 45min on day 9). However, it consisted this time in an *adapted* version of the online citizen-science experiment designed by researchers at Yale University, called the Tone Deafness Test (see: https://www.themusiclab.org/quizzes/td). This version was implemented on PychoPy2 (Peirce et al., 2019) and presented to the participants using the online Pavlovia platform (Bridges et al., 2020) with instructions translated in French. In this task, participants had to listen to sounds of various pitches and indicate whether the last sound went up or down. They were also asked questions (in French) regarding their experience with music, presented on a response sheet.

Regarding the questionnaires, participants were instructed to fill in the IRI questionnaire (Braun et al., 2015) as a measure of cognitive and affective empathy, and the PSIQ-F questionnaire (Ceschi & Pictet, 2018) as a measure of mental imagery. Both questionnaires were presented using an online survey administred through LimeSurvey 2.06 (LimeSurvey GmbH, n.d.). Participants then answered an additional question to determine whether they were aware of the goal of the study. None of them indicated being aware of the hypotheses.

2.1.5 Data analysis and pre-processing

Experimental design

The data from the pre- and post-tests were analysed separately. In both cases, reaction times from the LDT were analysed with linear mixed-effects models using lme4 package in R (Baayen et al., 2008). The dependent variable was the participants' reaction times on the LDT. The independent variables were the relatedness factor, corresponding to whether the pair of words belonged to the same semantic category, and the typicality factor, corresponding to whether the prime was a typical or an atypical word. As in Barr et al. (2013), we started by implementing the maximal random effects structure justified by the design. This involved by-participant random intercepts and by-participant random slopes corresponding to relatedness and typicality factors, as well a by-item random intercepts and by-item random slopes corresponding to the relatedness and typicality factors (in the present experiment, items corresponding to the target words of each pair on which participants made their lexical decision). Unfortunately, models including this maximal structure systematically failed to converge in both the pre- and post-tests. We thus attempted to simplify the structure of the models to improve model convergence, as detailed in the next section.

Regarding a potential difference of the impact of dialogue on the changes in lexico-semantic representations between the director and the matcher, we conducted a second analysis with data from the post-test. In order to determine whether producing information enhances the impact of dialogue, we built a linear mixed-effects model in which the role of the participant (either director or matcher) was added as an independent variable, in addition to the relatedness and the typicality factors. The dependent variable remained the same (i.e. reaction times on the LDT) and we began by implementing the maximal random effects structure justified by the design. This involved, by-participant random intercepts and by-participant random slopes corresponding to the role, relatedness and typicality factors, as well as by-item random intercepts and by-item random slopes corresponding to the role, relatedness and typicality factors. However, since the model also failed to converge when implementing this maximal structure, we attempted to simplify its structure as detailed in the next section.

Finally, two additional analyses were conducted with data from the post-test to investigate the modulation effects of personality traits on the changes in lexico-semantic representations after dialogue. These models both included by-participant and by-item random intercepts, following the same rationale as the main model. The first model included both the relatedness factor and the centred data for each dimension of the IRI questionnaire as fixed effects (Personal Distress, Perspective Taking, Empathic Concern and Fantasy). Centring the data for each dimension allowed to reduce the risk of multicolinearity. The second model also included the relatedness factor as fixed effect, as well as centred data for each dimension of the PSIQ-F as fixed effects (Vision, Sound, Smell, Taste, Touch, Body and Emotion). Since these two additional models also failed to converge when implementing this maximal structure, we attempted to simplify their structure.

Pre-test analysis

In the pre-test, the most complex structure of random effects that allowed the main model to converge included by-participant random intercepts, by-item random intercepts and by-item random slopes corresponding to the relatedness factor. Fixed effects for the model remained as planned, with the relatedness and the typicality factors.

Post-test analysis

In the post-test, the maximal structure of random effects that allowed the main model to converge included by-participant random intercepts and by-participant random slopes corresponding to relatedness, as well as by-item random intercepts and by-item random slopes corresponding to the relatedness factor. Fixed effects remained as planned, with the relatedness and the typicality factors.

Regarding the analysis of the effect of the role, the maximal structure of random effects that allowed the model to converge included by-participant random intercepts and by-participants random slopes corresponding to the relatedness factor, as well as by-item random intercepts and by-item random slopes corresponding to the relatedness factor. It also included the relatedness, the typicality and the role factors as fixed effects.

The first additional model including data from the personality questionnaires included by-participant random intercepts and by-participant random slopes corresponding to relatedness factor, as well as by-item random intercepts and by-item random slopes corresponding to the relatedness factors. It also included relatedness, typicality and centred data for each dimension of the IRI questionnaire as fixed effects (Personal Distress, Perspective Taking, Empathic Concern and Fantasy). The second model included by-participant random intercepts and by participant random slopes corresponding to the relatedness factor, as well as by-item random intercepts and two by-item random slopes corresponding to the relatedness and the typicality factors. Finally, it also included relatedness and typicality as fixed effects, as well as centred data for each dimension of the PSIQ-F as fixed effects (Vision, Sound, Smell, Taste, Touch, Body and Emotion).

Data pre-processing

The data from five dyads of participants were removed from the analyses before pre-processing, either because one of the participants did not follow the instructions in one of the tasks (PBM task or LDT) or owing to technical issues, leading to a total of 38 participants included in the analysis.

Regarding data pre-processing for the pre-test, only the reaction times and errors in this first LDT were taken into account. The data from two pairs of items giving rise to more than 30% of errors were removed from the analysis (see Brunellière and Bonnotte, 2018 for a similar approach). Regarding the reaction times for the remaining data, incorrect responses (3.34%) and decision latencies longer than 2120 ms (five standard deviations above the grand mean, for a similar approach, see Baayen et al., 2003) or under 200ms (Brunellière and Bonnotte, 2018) were excluded from data analysis (0.45%). The mean error rates for the related and unrelated conditions were 3.64% and 3.51% for the pairs with an atypical prime, and 1.89% and 3.92% for the pairs with a typical prime, respectively. As our hypotheses did not concern error rates and as they were below the 5% threshold, no statistical analysis was conducted on them.

Regarding data pre-processing for the post-test, errors from the dialogue task were taken into account in addition to the reaction times and errors from this second LDT. Specifically, we looked at two types of errors in the PBM task. The first type was an error in following the instructions of the task. The recordings of the dialogues were checked to control for possible errors committed by the director in the way they gave the properties (i.e. omissions or substitutions of words), or possible repetition of a property by either the director or the matcher within the same trial. The other type of error concerned the performance in the PBM task, calculated for directors and matchers separately. What counted as an error in terms of performance to the task on the director's side was the misattribution of a property to a given picture (e.g. the property for the spoon being attributed to the cradle or conversely). What counted as an error on the matcher's side was the misplacement of a picture, or when a picture was omitted or replaced by a filler picture more than once along its four repetitions. Finally, a whole category had to be removed from the analysis of both participants of a dyad if the director pronounced the name of the category instead of the property or discussed a property during the break. In the end, one dyad was excluded due to an error rate superior to 30% on the matcher's side. Incorrect responses from all other dyads (7.45%) were removed from the analysis on reaction times. These criteria were established to ensure that pictures were associated with their corresponding properties the right number of time and prevent uncontrolled associations between a picture and a property (i.e. too many associations with the same picture or wrong associations between a property and a picture).

Regarding the data from the LDT, the two pairs of words excluded in the pre-test for inducing more than 30% of errors were excluded again for the same reason (see Brunellière and Bonnotte, 2018 for a similar approach). Regarding the reaction times for the remaining data, incorrect responses (2.85%) were excluded from data analysis. The mean error rates for the related and unrelated conditions were 2.80% and 3.43% for the pairs with an atypical prime, and 1.83% and 2.79% for the pairs with a typical prime, respectively. As our hypotheses did not concern error rates and as they were below the 5% threshold, no statistical analysis was conducted on them. After the exclusion of data from the PBM task and the LDT, one dyad was also removed as one of the two participants scored fewer than 13 points of measure in at least one of the four conditions (i.e. unrelated pairs with atypical primes, unrelated pairs with typical primes, related pairs with atypical primes or related pairs with typical primes). Finally, decision latencies longer than 1818 ms (five standard deviations above the grand mean, for a similar approach, see Baayen et al., 2003) were excluded from data analysis (0.61%).

2.2 Results

The results of Experiment 5 are shown in Figure 20 and in Tables 13a and 13b. Unlike what was expected, the analysis of lexical decision latencies for the pre-test did not reveal any significant relatedness effect (related pairs, M = 661 ms, SD = 224 ms; unrelated pairs, M = 662 ms, SD = 212 ms), b = -1.33, SE = 10.58, χ^2 (1) = 0.02, p = .90. No significant typicality

Table 13a. *Fixed effects, Experiment 5, pre-test*

Fixed effect	χ^2	**DF**	***b***	***SE***	***p***
Relatedness factor	0.02	1	-1.33	10.58	0.90
Typicality factor	0.09	1	-5.28	17.70	0.77
Relatedness x Typicality	0.01	1	-1.30	14.93	0.93

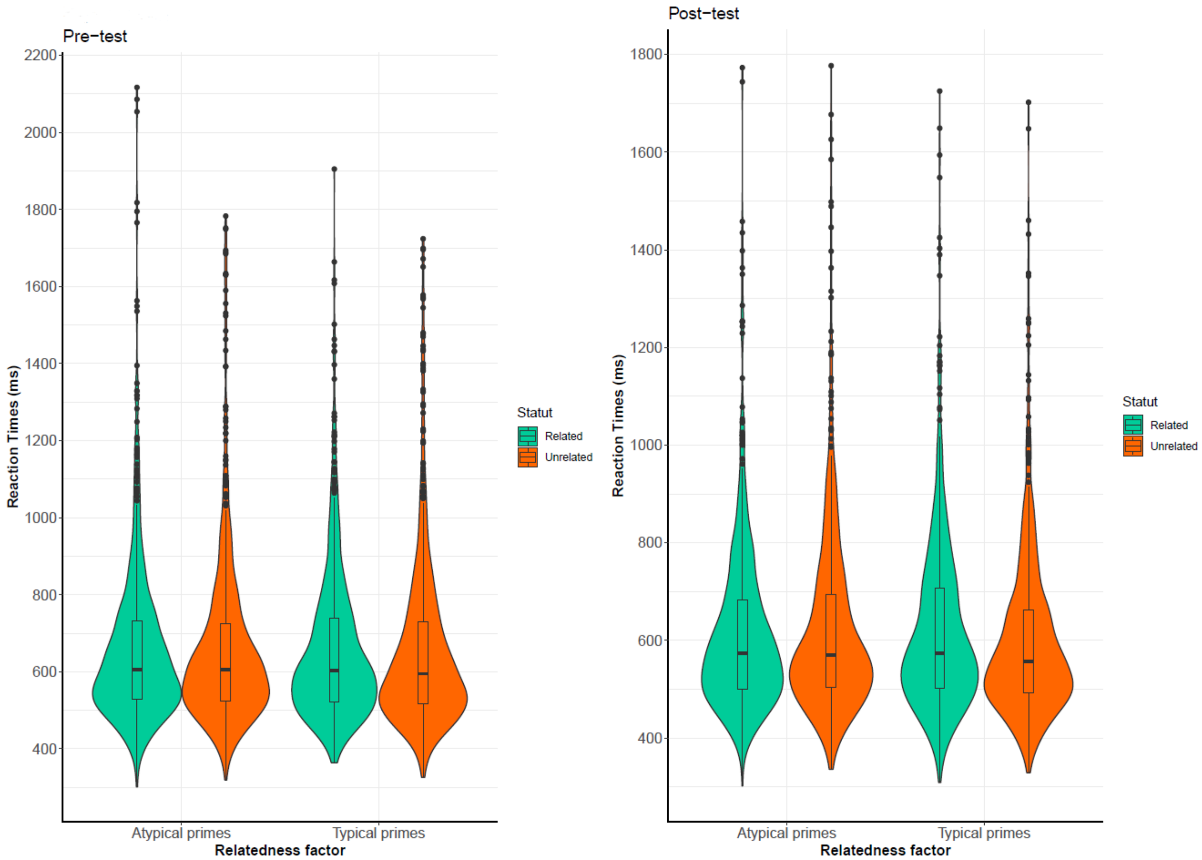
Table 13b. *Random effects, Experiment 5, pre-test*

Random effects	**Variance**	***SD***
By-participant		
Intercept	9330	96.59
By-item (target)		
Intercept	4177	64.63
Relatedness_factor	526	22.94
Residual	34779	186.49

effect was found either (typical prime, *M* = 659 ms, *SD* = 214 ms; atypical prime, *M* = 663 ms, *SD* = 223 ms), *b* = -5.28, *SE* = 17.70, χ^2 (1) = 0.09, *p* = .77. Finally, we found no significant interaction between the relatedness and the typicality factors (*b* = -1.30, *SE* = 14.93, χ^2 (1) = 0.01, *p* = .93).

Contrary to our hypotheses, the analysis of the lexical decision latencies for the post-test did not reveal any significant difference between the related pairs (*M* = 613 ms, *SD* = 187 ms) and the unrelated pairs (*M* = 621 ms, *SD* = 183 ms), *b* = 5.66, *SE* = 10.72, χ^2 (1) = 0.28, *p* = .60 (see Tables 14a and 14b). The difference between pairs containing a typical prime (*M* = 613 ms, *SD* = 182 ms) and pairs containing an atypical prime was not significant (*M* = 621 ms, *SD* = 189 ms), *b* = 4.03, *SE* = 15.70, χ^2 (1) = 0.07, *p* = .80. Finally, we found no significant interaction between the relatedness and the typicality factors (*b* = -26.07, *SE* = 14.64, χ^2 (1) = 3.17, *p* = .08).

Figure 20. *Reaction times as a function of the relatedness factor in Experiment 5 (pre- and post-test)*



The results of the analysis on the role of the participants did not reveal any significant main effect of role (b = -23.41, SE = 31.21, χ^2 (1) = 0.56, p = .45) or any significant interaction between the role and the relatedness (b = -16.62, SE = 14.20, χ^2 (1) = 1.37, p = .24) or the typicality factors (b = -14.60, SE = 12.52, χ^2 (1) = 1.36, p = .24). Regarding the two additional models, no interaction effect with the relatedness or the typicality factors was found. The results from these models are reported in Appendix D.

Table 14a. *Fixed effects, Experiment 5, post-test*

Fixed effect	χ^2	**DF**	***b***	***SE***	***p***
Relatedness factor	0.37	1	6.66	10.91	0.54
Typicality factor	0.06	1	3.93	15.71	0.80
Relatedness x Typicality	3.44	1	-27.07	14.60	-1.85

Table 14b. *Random effects, Experiment 5, post-test*

Random effects	Variance	*SD*
By-participant		
Intercept	7270	85.27
Relatedness_factor	390	19.75
By-item (target)		
Intercept	3239	56.91
Relatedness_factor	1005	31.69
Residual	23985	155

2.3 Discussion

In the present experiment, we investigated whether dialogue could impact the organisation of lexico-semantic representations into semantic categories by exploring the typicality effect. Dyads of participants took part in a dialogue task based on the matching task (Clark & Wilkes-Gibbs, 1986) during which one of them (i.e. the director) had to make his/her partner (i.e. the matcher) guess the order in which photo-realistic pictures had to be placed. These pictures could either represent typical or atypical items of a semantic category. Crucially, participants were not allowed to pronounce the pictures' names and had to refer to them by using sentences corresponding to the properties of their belonging semantic category instead. This allowed us to reinforce the relationship between atypical items and their semantic category, in order to make them less atypical after the dialogue task than before. Participants then performed a lexical decision task in a semantic priming paradigm both one week before the dialogue task and immediately after. We expected to find a significant relatedness effect in both measures, with faster reaction times for pairs of words belonging to the same semantic category than for those belonging to distinct semantic categories. In the pre-test one week before the dialogue task, we also expected to find a significant interaction between the relatedness and the typicality factors. More precisely, pairs of words primed with an atypical word were expected not to produce any difference in the reaction times between related and unrelated pairs of words, while pairs of words primed with a typical word were expected to elicit faster reaction times for related than for unrelated pairs. In the post-test however, atypical primes were expected to produce the same relatedness effect as typical primes, with faster reaction times for related pairs of words than for unrelated pairs of words. However, the analysis of the data in the lexical decision task did not reveal any significant relatedness effect or any significant interaction between the relatedness and the typicality factors, either in the pre- or in the post-test.

The absence of a relatedness effect in the pre-test and the absence of interaction between the relatedness and the typicality factors in the pre-test is not in line with the results by Brunellière and Bonnotte (2018). In their study, the lexical decision task demonstrated a significant relatedness effect with pairs of words belonging to the same semantic category, eliciting faster reaction times than pairs belonging to different semantic categories. The authors also evidenced a significant interaction between the two factors (i.e. relatedness and typicality), with typical primes producing a higher relatedness effect than atypical primes. This result was in line with the general account of spreading activation during the lexical decision task. Indeed, the prime is thought to pre-activate the target if they share a semantic relationship, which in this case was their shared category membership (Collins & Loftus, 1975; Rosch, 1975). However, it has been suggested that atypical items promote pre-activation less, as they are being less related to their superordinate category.

Interpreting these results in terms of the strength of the spreading of activation was assumed to be the most likely explanation for the semantic relatedness effect found in the lexical decision task. Indeed, the short stimulus onset asynchrony (SOA) (under 300 ms; Neely, 1977) and the low relatedness proportion (RP = .25) chosen by the authors in their lexical decision task are generally assumed to be related with relatively automatic processes. Conversely, a longer SOA and a higher RP are thought to be necessary for more strategic uses of the prime, such as expectation-based generation of target identity (i.e. using the identity of the prime to anticipate the lexical identity of the target; Hutchison et al., 2001; Hutchison, 2007; Neely, 1977). In the present experiment, the SOA and RP were the same as those used by Brunellière and Bonnotte (2018; SOA = 166 ms and RP = .25) and were in line with the results of Hutchison et al. (2001) that no strategy based on the relatedness proportion could emerge in this condition. The fact that we did not find any significant result is thus even more intriguing and is further discussed in light of the results of the sixth experiment.

In this sixth experiment, we explored the possibility that changes in the organisation of the lexico-semantic representations induced by dialogue could be consolidated by a night of sleep promoting consolidation and persist one day after dialogue (Dumay & Gaskell, 2012). As the data of Experiments 5 and 6 were collected simultaneously, the same lexical decision task was used to assess the changes in lexico-semantic representations (i.e. same paradigm, SOA and RP). This sixth experiment thus offers interesting insights as to the potential causes of the absence of significant effects in Experiment 5. Indeed, whether we replicate the results from Brunellière and Bonnotte (2018) in our pre-tests, or the results obtained in Experiment 5, the following experiment (Experiment 6) should help us draw interpretations regarding the impact of dialogue on the organisation of the semantic categories.

3. Experiment 6

Experiment 6 was similar to Experiment 5 except that participants performed the lexical decision task one day after the dialogue task. The aim of this experiment was to assess the changes in pre-existing lexico-semantic representations stored in semantic memory one day after dialogue. (i.e. after one night of consolidation). The hypotheses were identical to those of Experiment 5.

3.1 Materials and methods

3.1.1 Participants

Forty French native speakers (39 students and one community member, among which 29 females; age, $M = 20.35$, $SD = 2.41$) who did not take part in Experiment 5 were paid 25€ to take part in Experiment 6. As in Experiment 5, all participants had normal or corrected-to-normal vision and no history of language disorder, auditory disorder or neurological disorder. They signed an informed consent form before the beginning of the experiment and were fully debriefed after the experiment.

3.1.2 Stimuli and procedure

The stimuli and tasks were the same as in Experiment 5. Although session 1 remained unchanged compared to Experiment 5, in session 2 the PBM task was followed by the Tone Deafness Test instead of the post-test to keep the timing of this session constant between Experiments 5 and 6 (see Figure 18). During session 3, which took place on day 9, participants performed the LDT with the semantic priming paradigm one day after the PBM task. After this task, they filled in the same questionnaires (IRI and PSIQ-F questionnaires) as in Experiment 5. In line with the analyses performed in Experiments 3 and 4 from the previous set of experiments, an additional question was asked at the end of the questionnaires to explore the potential effects of the delay between the PBM task and the night of sleep on the consolidation of changes induced by dialogue. Finally, none of the participants indicated being aware of the hypotheses.

3.1.3 Data analysis and pre-processing

The same statistical models as those presented in Experiment 5 were planned. Unfortunately, models including the maximal structure failed to converge systematically in both the pre- and post-test. We thus attempted to simplify the structure of the models to improve model convergence.

Pre-test analysis

In the pre-test, the most complex structure of random effects that allowed the main model to converge included by-participant random intercepts, by-item random intercepts and by-item random slopes corresponding to the relatedness factor. Fixed effects for the model remained as planned, with the relatedness and the typicality factors.

Post-test analysis

In the post-test, the most complex structure of random effects that allowed the main model to converge included by-participant random intercepts and by-item random intercepts. Fixed effects for the model remained as planned, with the relatedness and typicality factors.

Regarding the analysis of the effect of role, the model included by-participant and by-item random intercepts, as well as by-item random slopes corresponding to the relatedness factor. It also included the relatedness, typicality and role factors as fixed effects.

The model with all dimensions of the IRI questionnaire included by-participant random intercepts, by-item random intercepts and by-item random slopes corresponding to the relatedness factors. It also included the relatedness factor, the typicality factor and the centred data for each dimension of the IRI questionnaire as fixed effects. Finally, the latter model included by-participant random slopes corresponding to the typicality factor and by-item random intercepts. It also included the relatedness and typicality factors, as well as centred data for each dimension of the PSIQ-F as fixed effects.

Data pre-processing

As in Experiment 5, the data from two pairs of items inducing to more than 30% of errors in the pre-test were removed from the analysis (see Brunellière and Bonnotte, 2018 for a similar approach). Regarding the reaction times for the remaining data, incorrect responses (3.39%) and decision latencies longer than 3040 ms (5 standard deviation above the grand mean, for a similar approach, see Baayen et al., 2003) were excluded from data analysis (0.47%). The mean error rates for the related and unrelated conditions were 2.95% and 3.59% for the pairs with an atypical prime, and 3.97% and 4.49% for the pairs with a typical prime respectively. As our hypotheses did not concern error rates and as they were below the 5% threshold, no statistical analysis was conducted on them.

Concerning the pre-processing of the post-test data, the criteria for data exclusion were the same as in Experiment 5, with the exclusion of the errors from the matching task and those from the LDT. Incorrect responses from the PBM task (4.36%) were removed from the analysis on reaction times.

Regarding the LDT, the data from two pairs of items inducing more than 30% of errors were removed from the analysis (see Brunellière and Bonnotte, 2018 for a similar approach). Concerning the reaction times for the remaining data, incorrect responses (3.17%) and decision latencies longer than 2376 ms (5 standard deviations above the grand mean, for a similar approach, see Baayen et al., 2003) were excluded from data analysis (0.54%). The mean error rates for the related and unrelated conditions were 3.63% and 1.64% for the pairs with an atypical prime, and 2.72% and 4.34% for the pairs with a typical prime, respectively. As our hypotheses did not concern error rates and as they were below the 5% threshold, no statistical analysis was conducted on them.

3.2 Results

The results of Experiment 6 are presented in Figure 21. The analysis of lexical decision latencies for the pre-test revealed a significant relatedness effect (related pairs, *M* = 672 ms, *SD* = 217 ms; unrelated pairs, *M* = 686 ms, *SD* = 230 ms; see Tables 15a & 15b), *b* = -23.38, *SE* = 10.37, χ^2 (1) = 5.08, *p* = .02. No significant typicality effect was found (typical primes, *M* = 677 ms, *SD* = 233 ms; atypical primes, *M* = 680 ms, *SD* = 215 ms), *b* = -11.43, *SE* = 23.16, χ^2 (1) = 0.24, *p* = .62. Finally, we found no significant interaction between the relatedness and the typicality factors (*b* = 14.75, *SE* = 14.70, χ^2 (1) = 1.01, *p* = .32).

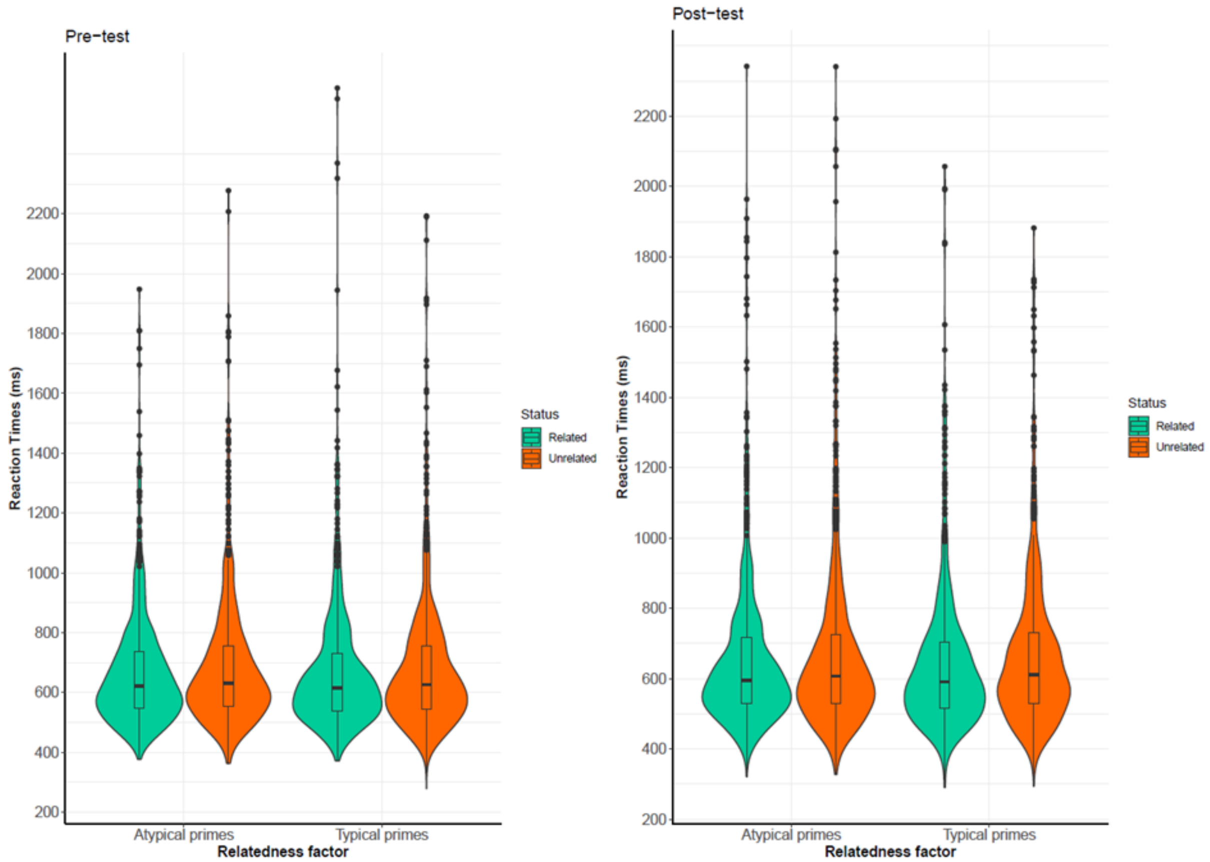
Table 15a. *Fixed effects, Experiment 6, pre-test*

Fixed effect	**χ^2**	**DF**	***b***	***SE***	***p***
Relatedness factor	5.08	1	-23.38	10.37	0.02*
Typicality factor	0.06	1	-11.43	23.16	0.62
Relatedness x Typicality	1.01	1	14.75	14.70	0.32

Table 15b. *Random effects, Experiment 6, pre-test*

Random effects	**Variance**	***SD***
By-participant		
Intercept	9645	98.21
By-item (target)		
Intercept	8636	92.93
Relatedness_factor	596	24.41
Residual	34545	185.86

Figure 21. *Reaction times as a function of the relatedness factor in Experiment 6 (pre- and post-test)*



The analysis of lexical decision latencies for the post-test revealed a significant relatedness effect (related pairs, *M* = 652 ms, *SD* = 223 ms; unrelated pairs, *M* = 673 ms, *SD* = 243 ms; see Tables 16a & 16b), *b* = -20.08, *SE* = 10.00, χ^2 (1) = 4.03, *p* = .04. The difference between pairs containing a typical prime (*M* = 657 ms, *SD* = 220 ms) and pairs containing an atypical prime was not significant (*M* = 667 ms, *SD* = 246 ms), *b* = -8.82, *SE* = 18.47, χ^2 (1) = 0.23, *p* = .63. Finally, we found no significant interaction between the relatedness and the typicality factors (*b* = -1.90, *SE* = 14.14, χ^2 (1) = 0.02, *p* = .89).

Table 16a. *Fixed effects, Experiment 6, post-test*

Fixed effect	χ^2	**DF**	***b***	***SE***	***P***
Relatedness factor	4.03	1	-20.08	10.00	0.04*
Typicality factor	0.23	1	-8.82	18.47	0.63
Relatedness x Typicality	0.02	1	-1.90	14.14	0.89

Table 17b. *Random effects, Experiment 6, post-test*

Random effects	**Variance**	***SD***
By-participant		
Intercept	16765	129.48
By-item (target)		
Intercept	4686	68.45
Residual	34246	185.06

Regarding the analysis of the role, no main effect (b = 53.87, SE = 43.61, χ^2 (1) = 1.53, p = .22) or interaction effect was found either with the relatedness (b = -12.73, SE = 14.11, χ^2 (1) = 0.81, p = .37) or with the typicality factor (b = -6.06, SE = 14.10, χ^2 (1) = 0.19, p = .67) on the lexical decision latencies (see Appendix D). In the analysis including the dimensions of the IRI questionnaire, no significant interaction was found (see Appendix D). However, in the analysis that included the dimensions of the PSIQ questionnaire, an interaction effect was found between the relatedness factor and the vision dimension of the questionnaire (b = -2.52, SE = 1.25, χ^2 (1) = 4.10, p = .04). This significant interaction comes from the fact that when the vision scores of the PSIQ-F increases, the reaction times to targets in unrelated pairs of words becomes significantly longer than the reaction times to targets in pairs of related words (see Figure 22).

Figure 22. *Reaction times as a function of the relatedness factor and scores of PSIQ_vision in Experiment 6.*

In line with the first set of experiments, an additional analysis was conducted in order to investigate further the effect of a night of sleep on potential changes in the participants' lexico-semantic representations. The detail of this analysis is provided in Appendix D. In this analysis, we included the time of day at which each participant performed the second session of the experiment (i.e. the PBM task) as an additional fixed effect in the model. The model also included relatedness as a fixed effect (as well as the interaction between relatedness and participation time); the outcome variable was the participant's reaction time on the lexical decision task. Indeed, dyads performing the first session at different times of the day meant that their awakening time before going to sleep on the night following the first session varied across dyads. However, no significant effect of participation time was found, nor did this factor interact with the relatedness or typicality factor; these results are therefore not discussed further in the remainder of this paper.

3.3 Discussion

In Experiment 6, we explored whether the organisation of lexico-semantic representations into semantic categories could change one day after a dialogue by exploring the phenomenon of consolidation generated during a night of sleep as well as the typicality effect. As in Experiment 5, dyads of participants performed the PBM task during which one of them (i.e. the director) had to make his/her partner (i.e. the matcher) guess the order in which photo-realistic pictures representing either typical or atypical items of their semantic category had to be placed. A lexical decision task with semantic priming paradigm was performed both one week before and one day after the PBM task in order to compare the state of the representations and assess the impact of dialogue. As expected from the literature but contrary to Experiment 5, the analysis of the data in the lexical decision task revealed a significant semantic priming effect with pairs belonging to the same semantic category eliciting faster reaction times than pairs belonging to distinct semantic categories. However, no interaction was found between the relatedness factor and the typicality factor in both the pre- and the post-tests. Additional analyses also revealed a significant interaction between the relatedness factor and the vision dimension of the PSIQ-F questionnaire. This dimension corresponds to the ability of someone to generate visual representations of a given concept in their mind (a friend, a cat climbing a tree or a campfire) and is evaluated on a scale from 1 to 10. The fact that this dimension interacted with the relatedness factor in the reaction times of the lexical decision task is both unpredicted and difficult to interpret considering the absence of literature on the subject. One possible interpretation of this effect lies in the use of visual properties to describe some of the pictures during the dialogue task (e.g. 'possesses four articulated limbs'). It is possible that the categorical relationships between items in a category for which visual properties were discussed became more strengthened in participants with higher abilities to generate visual representations. As a result, these strengthened relationships may have facilitated access to target words belonging to the same semantic category as their prime words during the lexical decision task.

The significant relatedness effect found in the pre-test of this sixth experiment is both in line with the literature on the typicality effect (Brunellière and Bonnotte, 2018; Rosch, 1973) and in contradiction with the results found in Experiment 5. A possible explanation to this discrepancy may be that the magnitude of the semantic priming in situations preventing strategical decisions is known to be subject to substantial variability. Indeed, it has already been suggested that the automatic activity occurring in semantic memory may be noisy and uncoordinated (Stolz et al., 2005). If this is the case, the results obtained with our SOA of 166ms may be subject to inconsistency, which would have prevented us from replicating the significant results of Brunellière and Bonnotte (2018). This idea is developed in more detail in the following section. Regarding the absence of interaction between the relatedness and the typicality factors in the pre-test, the result is not in line with our hypothesis or with the results of Brunellière and Bonnotte (2018), as we expected a significant semantic priming effect for typical primes but not atypical primes. Given the absence of an interaction between the relatedness and the typicality factors in the pre-test, the results of the post-test (i.e. a significant relatedness effect and an absence of interaction between with the two factors) are difficult to interpret. Although the relatedness effect observed in the pre-test was expected to remain in the post-test, the lack of interaction with the typicality factor was meant to be an evidence of changes in the organisation of lexico-semantic representations, compared to the pre-test. However, the absence of interaction between the relatedness and typicality factors in both the pre- and the post-test prevents us from drawing clear conclusions about the impact of dialogue on the typicality effect. In the following section, we discuss these results in the light of the literature on semantic priming.

4. Discussion of Chapter 6

The present set of experiments investigated to what extent dialogue may impact the organisation of lexico-semantic representations by exploring the typicality effect. In two experiments, a new dialogue task inspired by the matching task, the Property-Based Matching task (PBM), was created to allow two naive participants to discuss the order in which photo-realistic pictures had to be placed. Crucially, these pictures could be associated with either a typical word (i.e. picture of an elephant) or an atypical word (i.e. picture of a panda). During the task, one of the participants (i.e. the director) had to make the other one (i.e. matcher) guess the order in which pictures were placed on the screen using the properties of each picture's superordinate semantic category. By proposing the same properties for each picture of a given category, we aimed at strengthening the relationship between the atypical items and their semantic category. We hypothesised that the lexico-semantic representations associated with atypical items would become less atypical after dialogue. The changes in the organisation of these lexico-semantic representations were investigated by using a visual lexical decision task coupled with a semantic priming paradigm. While the changes in the pre-existing lexico-semantic representations were examined immediately after dialogue in Experiment 5, we tested the possibility that the changes might persist after one night of sleep thanks to a consolidation process in Experiment 6.

Although the results of these two experiments do not support our hypothesis that dialogue impacts pre-existing lexico-semantic representations, they do not point to an absence of impact either. Indeed, the inconsistency in the relatedness effect found in the pre-tests of both experiments in comparison with the literature suggests a high degree of inter-individual differences in the way the activation spreads into the lexico-semantic network during the lexical decision task. This inter-individual variability is well reflected by the comparison between the mean reaction times for related vs. unrelated pairs for each participant during the pre-tests. Numerically, among the 38 participants of Experiment 5, only 21 showed a facilitatory effect (55.26%), with mean reaction times for related pairs being shorter than for unrelated pairs. Even more striking is the range of the differences in mean reaction times for related and unrelated pairs, which spans from -147.18 ms to 107.71 ms (Mean = 0.26, SD = 46.82). In comparison, Experiment 6 showed a slightly higher proportion of participants exhibiting facilitatory effects, with 24 out of 40 participants (60%) showing faster mean reaction times for related pairs than for unrelated ones. In addition, the minimum difference between mean reaction times for related and unrelated pairs was closer to the mean of the distribution (Min = -54.24, Mean = 14.35, SD = 39.00), although the maximum difference was similar to that in Experiment 5 (Max = 107.43). These numbers align with the idea that our data reflects a high variability in the ease with which facilitatory effects are observed across participants. This idea is discussed in more depth in the next section.

4.1 Using a semantic priming paradigm to evidence the typicality effect

The hypotheses of the present set of experiments were based on the results of Brunellière and Bonnotte (2018), who used a lexical decision task coupled with a semantic priming paradigm to explore the typicality effect and the category structure of the semantic representations stored in long-term memory. While they found a main effect of relatedness when targets were typical items of their semantic category, they also evidenced a significant interaction between relatedness and typicality, with significant semantic priming for typical primes but not for atypical primes. This effect has been explained by a combination of the spreading activation theory (Collins & Loftus, 1975) and the prototype model (Rosch, 1975). Specifically, Brunellière and Bonnotte (2018) suggested that typical primes with stronger relations with their superordinate category and several shared features with the other members of the category should produce more spreading of activation towards the other members of the category than atypical primes. While this interpretation provides a coherent account of how the spreading activation and the typicality effect can be combined, we were not able to replicate such interaction in the pre-tests of our experiments. Yet, we do not think that our results invalidate those of Brunellière and Bonnotte (2018). The inconsistency of the relatedness effect found in the pre-tests of our two experiments prevents us from drawing clear conclusions based on the present dataset. Instead, we can only discuss this inconsistency in regard with the literature on semantic priming and lexical decision task.

As already mentioned, the use of a short SOA and a low relatedness proportion are thought to ensure the interpretation of the results of a lexical decision task in terms of spreading of activation, as in Brunellière and Bonnotte (2018). While this idea has been challenged in the literature by showing the impact of retrospective (strategic) decisions on the results of a lexical decision task in which a short SOA was used (de Wit & Kinoshita, 2014; 2015a; 2015b), these studies have always used an SOA longer than 200ms. Indeed, to our knowledge, there is no evidence of strategical decisions in studies using SOAs under 200ms. Consequently, the interpretations of Brunellière and Bonnotte (2018) in terms of spreading of activation remain unchallenged. The question thus arises concerning the reason why, when the spreading of activation into the lexico-semantic network is thought to be relatively automatic, this pattern is inconsistent across studies that use similar materials and tasks (i.e. Brunellière and Bonnotte, 2018 vs. the pre-tests from Experiments 5 and 6)? A first explanation may be that the familiarisation task that participants performed before doing the lexical decision task impacted the following encounter with the words presented in the latter. Indeed, this task constitutes one of the main differences between our set of experiments and that of Brunellière and Bonnotte (2018). However, the fact that we did not find the same pattern of results between the pre-tests of Experiments 5 and 6 challenges this hypothesis, as the familiarisation task was presented before them and should have produced the same bias on the results of both experiments. Another already mentioned explanation may lie in the substantial variability of the magnitude of the semantic priming observed in the literature within and between participants, in situations in which an interpretation in terms of spreading of activation is privileged (Stolz et al., 2005). Indeed, Stolz and collaborators (2005) showed significant test-retest or inter-item reliability in the lexical decision task but only with an RP equal to .50 and not with RPs equal to .25, independently of the SOA. The authors interpreted it as the consequence of noisy and uncoordinated activity in the semantic network. They suggested that although the spreading of activation is not a coherent pattern within each individual, the sum of these incoherent activations always results in a general priming effect. Moreover, strategic priming (i.e. higher RP and SOA) is thought to improve the test-retest reliability so that the priming observed at the group level emerges through the sum of more coherent (i.e. more strategic) activations. This lack of reliability in priming was also replicated in later studies that used a short SOA (i.e.

200ms) and a low RP (.25; Heyman et al., 2018), which may account for the variability in the results of our own pre-tests. If such variability exists in the way activation spreads in the semantic network, it might explain the fact that the sum of activations reaches the threshold of a significant priming effect in some groups but not all. However, Heyman and collaborators (2018) nuanced the interpretation of Stolz and collaborators (2005) by underlining that sample size may play a role in the lack of item-level reliability of the studies. By conducting an analysis on Tan and Yap's (2016) data, they demonstrated that when results are based on a large sample size (240 participants), the estimate is largely superior to those of random subsets of 40 participants extracted from the same sample, with a reliability ranging from .70 to only .28. Such a demonstration may account for the differences found in the results of our samples of 38 and 40 participants.

We therefore decided to run a supplementary analysis to determine whether a significant priming effect may be found with a larger sample of our data may bring out a significant priming effect. To do so, we combined the samples from the pre-tests of our Experiment 5 and 6 to obtain one large sample of 78 participants. The analysis was performed only on the pre-test, as all the participants from both experiments were treated equally at that stage. Using the same procedure than previously described to create the statistical models, we were not able to evidence a significant relatedness effect with this new sample of 78 participants (b = -7.12, SE = 9.36, χ^2 (1) = 0.58, p = .45). Although this sample size was not as large as that of Tan and Yap (2016), it was still larger than most of other samples (Brunellière et al., 2017; Brunellière & Bonnotte, 2018; De Wit & Kinoshita, 2015b; Heyman et al., 2018) without being able to show a significant priming effect. While this result suggests that the absence of a significant relatedness effect is not due to a lack of statistical power, further research is needed to clarify the question of the reliability of the results obtained with lexical decision tasks and semantic priming paradigms.

Another line of research comes from the literature on associative relationships and highlights the question of the reliability of the semantic priming in the present set of experiments (Yap et al., 2016). As defined in the introduction of this manuscript, associative relationships are the relationships between words as they may be evidenced by associative norms, meaning that they reflect at the same time co-occurrences between words and 'pure' semantic relationships such as those manipulated in the present set of experiments (i.e. feature similarities). Yap et al. (2016) suggested that the strength of the associative relationship between a prime and a target in a semantic priming paradigm plays a role in the ease to evidence a semantic priming effect. Specifically, related pairs of words that are 'first associates' (i.e. the target is the most given response to the prime in a word association task) give rise to significantly more reliable semantic priming than 'other associates' (i.e. any word other than the most given one). Therefore, if the words are not closely related to one another in the network, the chances of observing a consistent spreading of activation within and across participants are low. Even if their RP was higher than the one used in the previously presented studies (.50), this effect has been observed with both a short SOA (i.e. 200ms) and a long SOA (i.e. 1200ms), thus limiting the impact of strategical responses. In the present set of experiments, we did not control for the associative strength of our prime-target pairs of words and chose to create related pairs of words based on their categorical (purely semantic) relationships. We also ensured that the pairs created for the lexical decision task did not co-occur highly (as presented in Table 12b). In such a context, we can hypothesise that the effects found by Yap and collaborators (2016) are mainly due to the strength of the co-occurrences between words rather than the strength of the purely semantic relationships, which may play only a minor role in their comparative study. The boost in the semantic priming effect may thus be due to higher statistical regularities between words rather than purely semantic relationships. This interpretation is in line with other studies on the effect of co-occurrence frequency on the strength of semantic priming (Brunellière et al., 2017). Coming back to our experiments, we can imagine that without an associative boost between pairs of words, the categorical relationship alone can hardly be evidenced in a semantic priming effect, owing to its low reliability. While the results by Brunellière and Bonnotte (2018) might challenge this idea, we believe that replication studies are needed in order to determine to which extent categorical relationships alone (i.e. without co-occurrences between words) can reliably be assessed with a semantic priming paradigm and a lexical decision task which prevent strategical decisions.

In conclusion, it seems that eliciting a constant semantic priming effect with a lexical decision task is difficult if one seeks to question the purely semantic relationships with spreading activation processes but without strategical processes. Consequently, while the semantic priming effect is arguably a meaningful tool to investigate the nature of the activations in the lexico-semantic network, it may not be the best one to evidence the impact of dialogue on lexico-semantic relationships. This conclusion is in line with that of the previous chapter and will be further discussed in the general discussion of this manuscript.

4.2 Interim conclusion

By studying the typicality effect, our primary aim was to investigate the impact of dialogue on the organisation of pre-existing lexico-semantic representations. Dialogue has already been suggested to interact with memory representations (Horton & Gerrig, 2005a; Pickering & Garrod, 2004; 2021). In particular, the interactive alignment model posits that alignment in the linguistic representations between dialogue partners can strengthen the co-activated representations. Creating a dialogue setting in which the relationship between an atypical item and its superordinate category is repeatedly mentioned may therefore strengthen the connections between their associated semantic representations. Although the present set of experiments was not able to provide data to support this hypothesis, it does not rule out the idea either. While we believe that the dialogue task was adapted to create changes in the organisation of lexico-semantic representations, these changes should be assessed with a more explicit task or by changing the parameters of the lexical decision task in order to allow more strategical processes to appear and provide more consistent data. Future studies should be performed in this respect to answer the question of the impact of dialogue on the organisation of the semantic representations. An additional line of research could be to try to replicate the findings of Brunellière and Bonnotte (2018) on the interaction between the typicality effect and the spreading activation in order to provide more consistent data on the subject.

Chapter 7 – General discussion

This last chapter is divided into five main sections, with the first section dedicated to a summary of the thesis and its main findings. From section two to section four, these results will be discussed in light of the literature presented in the first two chapters. Finally, the last section consists in the conclusion of the thesis.

---

1. Summary

1.1 Aims

The aim of this thesis was to investigate whether or not the organisation of pre-existing lexico-semantic representations stored in the semantic network could be changed after dialogue. Pickering and Garrod (2004; 2021) suggested that the linguistic representations of dialogue partners receive new activations during a conversation, owing to the mechanism of alignment (i.e. increase in the degree of similarity between mental representations). However, the question remained as to whether such activations could affect linguistic representations so that they remain changed after dialogue. To answer this question, we focused on lexico-semantic representations and their connections within the semantic network, as they are known to adapt after being exposed to specific linguistic inputs (Rodd et al., 2012; 2013; 2016). We examined whether changes in the lexico-semantic network caused by dialogue impact memory representations immediately after the interaction. As these changes are known to evolve with time such that they are susceptible to decay until they get consolidated by offline processes occurring during a night of sleep (Dumay & Gaskell, 2012; Gaskell et al., 2019; Mak et al., 2023), we also examined the changes in the lexico-semantic representations one day after dialogue. Before addressing this issue, Chapter 4 detailed the creation of a database of standardized tangram pictures and their associated names, so that the appropriate pictures could be selected for Chapter 5. Specifically, in the latter we wanted to create a new dialogue task called the Interactive Agreement Referential task to examine the possibility of creating new lexico-semantic relationships between already existing lexico-semantic representations. To do so, we selected tangram pictures that were plausible visual representations for two semantically unrelated words so that these words could be associated together during dialogue. Finally, in Chapter 6, we investigated the possibility of adapting the graded structure of pre-existing lexico-semantic representations within their semantic category. To this aim, another dialogue task was created and called the Property-Based Matching task. In the following section, we summarise our main results and discuss them in light of the literature already presented throughout this manuscript.

1.2 Summary of main findings

The main findings of this thesis were evidenced in the third experiment of Chapter 5. In this experiment, we showed that new relationships created during dialogue between pre-existing lexico-semantic representations could be evidenced one day later. This effect has been demonstrated using a semantic relatedness judgement task in which participants had to explicitly focus on the semantic similarities between pairs of words. However, these new relationships could not be evidenced when they were assessed with a lexical decision task and a semantic priming paradigm. We proposed that the changes occurring in the lexico-semantic network were too weak to be evidenced by a lexical decision task and that more explicit tasks are needed to explore the effects of dialogue. Similarly, no significant changes in the organisation of the semantic categories (Chapter 6) were observed, probably owing to the choice of tasks and parameters for the semantic priming paradigm (i.e. SOA and RP). Specifically, a short SOA and a low RP are thought to be optimal to prevent strategical decisions that are less representative of the organisation of the lexico-semantic network. However, these decisions may have affected the ease to observe subtle changes occurring in the network, as explained in the following section.

2. Choosing the appropriate measure to access lexico-semantic representations

2.1 Choosing the appropriate parameters for a semantic priming paradigm

As presented in the introduction of this manuscript, the notion of semantic priming effect is at the core of some of the most influential models of the organisation of semantic representations (Collins & Loftus, 1975; McNamara, 2005). Consequently, the semantic priming paradigm is often used as a tool to examine the access to lexico-semantic representations and their organisation (for reviews, see Hutchison, 2003; Lucas, 2000; McNamara, 2005; Neely, 1977). When using this paradigm, several choices have to be made to ensure that what is measured corresponds to what is being assessed. First, the semantic priming paradigm has to be associated with the appropriate task. In the present experiments, the semantic priming paradigm was coupled with a lexical decision task, as it is thought to be the best tool to explore word access, unlike semantic categorisation and pronunciation tasks which are used to explore word retrieval (Taikh and Lupker, 2020). Second, the semantic priming paradigm is associated with parameters which have major impacts on the way results can be interpreted. In the present thesis, three parameters were carefully calibrated: Stimulus Onset Asynchrony (SOA), the Relatedness Proportion (RP) and the presence of associative relationships between words, as defined by the co-occurrences between words in natural language. As we wanted to create strong relationships between lexico-semantic representations, we tried to evidence those relationships by examining the spreading of activation from one representation to another. We thus chose a very short SOA (166ms) and a low RP (.25), as these conditions are ideal to observe the spreading of activation without strategical activations into the lexico-semantic network (Hutchison et al., 2001; Hutchison, 2007; Neely, 1977).

Regarding the presence of co-occurrence frequencies, our two sets of experiments (Chapters 5 and 6) did not undergo the same treatment. In Chapter 5, we attempted to create a *new* lexico-semantic relationship that could compete with pre-existing ones in terms of the amount of activation it could receive from the prime word in the lexico-semantic network. Consequently, we chose to add co-occurrences between our newly semantically related words to boost the new relationship (Brunellière et al., 2017). More precisely, the new relationship was composed of a semantic dimension, owing to the new common visual representation between the words, and co-occurrence boost, owing to the words being repeated four times together during dialogue. We chose not to distinguish between semantic relationships and co-occurrence frequency in the interpretation of our results, in order to maximise our chances of finding a significant semantic priming effect. Conversely, we chose to focus on a specific type of semantic relationship in Chapter 6 (i.e. categorical relationships) to favour an interpretation of our results in terms of more pure semantic relationships.

While these choices were motivated by the need to answer specific research questions, they also produced limitations that may partly have led to the lack of significant results. Coming back to Chapter 6 in which we tried to assess changes in the pre-existing semantic relationships without manipulating co-occurrence frequency between pairs of words to boost the effect of the semantic priming. It may be argued that our chances of observing a significant semantic priming effect were not optimal. Indeed, purely semantic relationships have been shown to be harder to evidence without co-occurrence frequency, especially for categorical relationships as those assessed in the pre-tests of our experiments (Hutchison, 2003). However, as categorical relationships have already been evidenced with a semantic priming paradigm (Brunellière & Bonnotte, 2018), the lack of co-occurrence frequency between our pairs of words is likely not the main cause of the absence of significant results in Chapter 6. In line with this idea, when creating the semantic relationships in Chapter 5, co-occurrences between words were added. However, these new relationships did not produce a significant semantic priming effect and were considered as too weak to trigger spreading of activation. This interpretation is in line with the literature on associative norms, in which the strength of the association between words (which can be considered as a combination of semantic relationships and co-occurrences between words) is thought to play a role in the ease of evidencing reliable semantic priming effects (Yap et al., 2016). Indeed, 'first associates' are thought to produce reliable effects while 'other associates' are not. In the present experiments, new semantic relationships and co-occurrences between words were created that were more likely to produce variable semantic priming effects and consequently be harder to evidence.

In conclusion, the reason why we did not find significant semantic priming effects in Chapter 6 may come from the type of semantic relationships that have been tested here, although the literature on the subject remains divided on the possibility to elicit semantic priming effects with categorical relationships only. On the other hand, the absence of a significant semantic priming effect in Chapter 5 may be due to the compatibility of the task itself with the need to assess *new* lexico-semantic relationships rather than the presence of co-occurrences between words. In the following section, we develop this idea that new lexico-semantic relationships may require more explicit tasks to be evidenced.

2.2 Assessing new lexico-semantic relationships

2.2.1 Insights from the literature on semantic priming and word-meaning priming

While assessing semantic and associative changes in the lexico-semantic network by using a semantic priming paradigm and a lexical decision task may seem challenging owing to the parameters previously mentioned, it may be even more complex to use this tool to measure *new* changes. Using a lexical decision task to test new relationships between words has already been attempted in the field of episodic priming, in which the decision on the lexical status of the word is thought to rely on episodic traces of the new relationship rather than on abstract relationships. For instance, episodic priming can be assessed by keeping the same context (e.g. the same modality) between the learning and the assessment phases. However, no clear priming effect was ever found for the newly related pairs of words (Caroll & Kirsner, 1982; Dagenbach et al., 1990; Durgunoglu & Neely, 1987; see also Hutchison, 2003 for review). Interestingly, Dagenbach et al. (1990) observed a significant priming effect only when creating a new relationship with very rare words and after an extensive five-week training period. The implications of these results are twofold. First, according to the authors, the new relationship may have required becoming an abstract relationship and being integrated into the semantic memory before producing a significant priming effect. Episodic traces may thus not be able to produce a significant priming effect. Second, familiar words may already be too integrated within a rich semantic network composed of numerous connections with other lexico-semantic representations, and new relationships may have difficulty in competing with pre-existing ones. In the experiments presented in Chapter 5, we attempted to maximise the creation (i.e. learning) of new relationships that are abstract in nature (i.e. integrated into the semantic memory) by alternating the presentation of pairs of words during dialogue and associating it with visual features (the tangram pictures). In addition, by changing the modality of presentation between the dialogue task and the lexical decision task, we wanted to ensure that if a new lexico-semantic relationship was observed, it was stored in the semantic memory rather than in the episodic memory. In line with the hypothesis of Dagenbach and collaborators (1990), an extensive training phase with several dialogues about the same tangram pictures may be necessary to create a sufficiently strong connection between our newly related pairs of words to compete with pre-existing lexico-semantic network of each word independently.

Although Dagenbach et al. (1990) did not specifically assess the new relationships to ensure that they were abstract in nature, their study is to our knowledge the only one which attempted to measure new relationships within the semantic memory using a semantic priming paradigm and a lexical decision task. Outside the field of episodic priming, Rodd et al. (2012) are the only ones who attempted to use a lexical decision task to assess the integration of new meanings for known words. However, they did not use a semantic priming paradigm in their study. They asked participants to learn new meanings for already known words and manipulated the relatedness of the new meaning in relation to the original one (either related or not). After a five-day learning phase, participants were asked to perform a lexical decision task on words related to new meanings and their reaction times and error rates were measured. Results showed that new related meanings elicited faster reaction times than new unrelated meanings, but only when the learning phase was highly semantically engaging. This result was interpreted as evidence for the integration of the new meaning within long-term representations. Crucially, when the learning phase was not semantically engaging (i.e. not centred on the ability of the word to interact dynamically with other semantic representations), no significant effect was found with the lexical decision task but only in the results from a cued-recall test. This semantic engagement during the learning phase was operationalised by the fact that participants were asked to produce examples of sentences that were compatible with the definition of the new word, or to write a story that included the new word. In light of this study and that of Dagenbach

(1990), it therefore seems crucial for new lexico-semantic information to undergo an intensive training phase consisting of several days of exercises to be integrated in the semantic memory. Although this assumption matches well with the absence of significant semantic priming effect found in Chapter 5, the fact that we found an integration of the new lexico-semantic relationship when testing it with a semantic relatedness judgement task can be questioned.

In line with this latter result, evidence of adaptation without having trained participants on semantically engaging tasks has already been observed in the literature on word-meaning priming (Betts et al., 2018; Gaskell et al., 2019; Rodd et al., 2013; 2016). However, these studies used measures such as word-association tasks to observe the adaptations, in which participants were expected to implicitly retrieve the primed meaning of a word and provide an associated word. Their results suggest that a single exposure to new lexico-semantic information can induce changes in the way lexico-semantic representations are organised in the semantic memory. However, these changes were observed only when attention was focused on them. Therefore, although assessing changes in the lexico-semantic network may seem challenging, the literature offers us two different ways of doing it. First, participants may undergo an extensive learning phase focused on the semantic properties of the new lexico-semantic information to ensure that the latter has been sufficiently integrated in the lexico-semantic network and can be accessed with implicit tasks such as a lexical decision task. Second, if the new lexico-semantic information has not been extensively learnt, researchers may favour the use of a task which focus attentional processes of the participants on the semantic properties of the new lexico-semantic information.

This interpretation is also in line with studies which tested the impact of task demand on the ease to evidence a semantic priming effect, even in pre-existing lexico-semantic representations (de Wit and Kinoshita, 2014; 2015a; 2015b). In their 2015b study, the authors showed that while significant semantic priming was evidenced with unmasked priming in both a lexical decision task and a semantic categorisation task, the semantic priming effect was eliminated from the lexical decision task but not from the semantic categorisation task when masking the prime. According to the authors (de Wit and Kinoshita, 2014; 2015a; 2015b), the processes engaged when seeing the prime and target words are therefore not the same in the two tasks owing to the impact of attentional processes. Based on their analyses of RT distribution, they suggested that while an increased use of strategies in a lexical decision task could lead to more retrospective semantic matching (i.e. reactivation of the prime after the appearance of the target to determine whether they are related or not), tasks focusing on the semantic status of an item that allows the establishment of strategies rely more on evidence accumulation (i.e. comparison of the semantic features of the prime and the target; Norris & Kinoshita, 2008). The task itself is thus meant to modulate the way the lexico-semantic network is activated, thanks to distinct attentional focus. This interpretation is also in line with the results obtained in Experiments 3 and 4 from Chapter 5, although in Experiment 4 we not only changed the task but also removed the semantic priming paradigm so that the pairs of words were presented simultaneously and for as long as the participant needed. To our knowledge, the studies by de Wit and Kinoshita are however the only ones to have examined the impact of the task on the semantic relatedness between words. The attentional hypothesis is thus the only explanation from the literature that accounts for the differences between the results of Experiments 3 and 4. It may therefore be interesting to reproduce the results from Experiment 4 with a semantic categorisation task and a semantic priming paradigm, in order to ensure that our interpretations are in line with those of these authors.

2.2.2 Implications for Experiments 2 to 6

In light of the interpretations of de Wit and Kinoshita (2014; 2015a; 2015b) and the findings in the word-meaning priming literature (Betts et al., 2018; Gaskell et al., 2019; Rodd et al., 2013; 2016), we posit that the changes in lexico-semantic relationships having occurred during the dialogues of our five experiments were integrated into the lexico-semantic network. However, these changes — either the creation of a new relationship or the strengthening of a pre-existing one — were too weak to produce a spreading of activation from the prime to the target in a lexical decision task. This idea is in line with most spreading activation models (Anderson, 1983; Collins & Loftus, 1975), which suggest that the strength of the relationship is related to the level of activation of the target (Lorch, 1982). Unless the new lexico-semantic information has benefited from an extensive learning phase centred on the semantic properties (Rodd et al., 2012), the strength of the relationship is too weak to produce a sufficient level of activation from the prime to the target word. As a consequence, new lexico-semantic relationships cannot be evidenced using a lexical decision task with a semantic priming paradigm in which the parameters are set to observe the consequences of spreading of activation only (i.e. short SOA and low RP). Instead, using a task which focuses participant's attention on the newly acquired lexico-semantic relationship such as in the semantic relatedness judgement task used in Experiment 4, people are able to search for evidence of the new lexico-semantic relationship in the network. This is probably the mechanism involved in studies that use a word-meaning priming paradigm and a word association task (Betts et al., 2018; Maciejewski et al., 2020; Rodd et al., 2012; 2013; 2016). In the latter, participants explicitly focus on the meaning of the word and search for related meanings within the lexico-semantic network. Evidence is thus accumulated towards a meaning that is semantically related to the word presented on the screen. If the word has been primed towards its subordinate meaning, evidence towards that meaning may be strong enough to produce significantly more associates. Therefore, this task does not depend on spreading of activation but rather on more explicit processes during which attention is focused on the related meanings.

It would be interesting to examine the changes in the graded structure of semantic categories (Experiments 5 and 6) by using a task focusing on the shared properties of the prime and target words, such as the semantic categorisation task already used in the literature (de Wit & Kinoshita, 2014; 2015b). It would also be interesting to try to investigate the impact of several dialogues instead of one on the lexico-semantic network thanks to a lexical decision task. However, this would require more resources in an already demanding procedure in terms of recruitment of participants. Overall, the evidence presented here confirms the impact of dialogue on lexico-semantic representations, even though this impact is weaker than expected. In the next section, we discuss how these changes may occur within the lexico-semantic network in light of the literature on consolidation, word-meaning mapping and dialogue.

3 – How changes integrate the lexico-semantic network: insights from the contextual binding account

Although the present thesis hypothesised that dialogue could impact directly the lexico-semantic relationships stored in long-term semantic memory, our experiments were not designed to distinguish between this account (inspired by the model of Pickering and Garrod, 2021) and the contextual binding account proposed by Duff and Brown-Schmidt (2017) and extended by Gaskell et al. (2019). In the first part of this section, we consequently discuss our results in light of the contextual binding account as it provides an interesting approach to explain the results obtained in the present experiments, before examining them through the lens of the model by Pickering and Garrod (2021) in section 4. The argumentation follows the structure of the experimental section, with a first part centred on Experiment 2 (Chapter 5), then a discussion of the results of Experiments 3 and 4 (Chapter 5), and a final section centred on the experiments of Chapter 6 (i.e. the typicality effect).

3.1 Evidence from the present experiments

In 2019, Gaskell and collaborators based the interpretation of their results on the proposition by Duff and Brown-Schmidt (2017) that the hippocampus plays an important role in the integration of new semantic information, and on the finding that hippocampal memories can be consolidated after a sleep period (Bendor & Wilson, 2012; Remme et al., 2021). They suggested that newly acquired lexico-semantic information is temporarily stored in the hippocampal memory, where it is bound with long-term representations in a 'context-specific' memory. After a period of sleep, the latter is integrated into long-term representations thanks to offline consolidation, and this integration affects the connections between pre-existing representations by adjusting their weights. According to this view, the changes in lexico-semantic relationships created in our five experiments were first mapped into the hippocampal memory before being integrated into long-term lexico-semantic representation after a night of sleep to adjust the connections within the lexico-semantic network.

3.1.1 Immediate changes in lexico-semantic relationships

In the present experiments, the only measure performed immediately after dialogue was a lexical decision task, which is an implicit measure of the changes in lexico-semantic relationships. We chose not to perform an explicit task immediately after dialogue, as the aim of this thesis was to find changes in lexico-semantic representations, and we were more confident about finding these changes after one night of sleep. Although the possibility to find direct changes in lexico-semantic representations immediately after dialogue is predicted by the theory of Pickering and Garrod (2004; 2021), the same assumption cannot be made in light of the contextual binding account (Duff and Brown-Schmidt, 2017; Gaskell et al., 2019). The theory by Pickering and Garrod (2004; 2021) postulates direct modulations of the lexico-semantic representations stored in long-term semantic memory, which suggests a significant impact of dialogue on the recognition of words immediately after the interaction. A contrario, according to the contextual-binding account, newly encountered linguistic inputs are first stored in a context-specific format and are associated with hippocampal traces before being consolidated into long-term representations by offline processes (Duff & Brown-Schmidt, 2017; Gaskell et al., 2019). According to Mak et al. (2023), these context-specific information biases subsequent lexico-semantic processing so that lexico-semantic representations are retrieved and bound with the new context-specific information to produce the most relevant interpretation of the situation (Mak et al., 2023). Therefore, newly encountered linguistic information is probably not able to produce changes in the recognition of words immediately after dialogue, but only after being consolidated into long-term lexico-semantic representations. In line with this idea, the immediate changes in lexico-semantic representations observed in studies that used a word-meaning priming paradigm were mainly evidenced with tasks in which the binding between lexico-semantic representations and context-specific representations is possible (Betts et al., 2018; Maciejewski et al., 2020; Rodd et al., 2012; 2013; 2016).

Consequently, as the aim of this thesis was not to investigate the contextual binding account, further studies should be conducted on the topic to determine whether dialogue can immediately bias lexico-semantic representations in word recognition tasks. In particular, the experimental setting used in Chapter 5 could be reused and the semantic relatedness judgement task could be performed immediately after dialogue. In the next section, we continue to review what the contextual binding hypothesis suggests as to what may have occurred in our experiments in terms of integration of the new lexico-semantic information. This time, we focus on the results of Experiments 3 and 4 after one night of sleep consolidation.

3.1.2 Long-term changes in lexico-semantic relationships

Across the three experiments presented in Chapter 5, we found evidence of an impact of dialogue on lexico-semantic representations only when assessing new lexico-semantic relationships with a task which explicitly focuses on the relationships between lexico-semantic representations one day after the interaction. In a previous section of this discussion, we posited that this finding was probably due to the fact that the integration of new lexico-semantic relationships produced weak changes that cannot induce spreading of activation into the lexico-semantic network (see section 2.2). We suggested that using an explicit task is not the only way to evidence changes in lexico-semantic representations. Instead, we proposed that an extensive training focused on the semantic properties of words may strengthen the new information sufficiently, making it observable with more implicit tasks (Rodd et al., 2012).

The notion that an extended training phase strengthens lexico-semantic traces in memory aligns with several models of memory consolidation (Remme et al., 2021; Winocur et al., 2010). According to these models, each instance of a new information is initially encoded as a unique memory trace. The neocortex then uses this collection of memories to extract statistical regularities across occurrences and form a more abstract version of the information. While these models are able to account for the effect of the extensive training phase used by Rodd and colleagues (2012), they are also in line with the idea that the new context-dependant representations stored in the hippocampus are subject to interference and decay (Remme et al., 2021; Rodd et al., 2016). In Experiment 4, such decay (as assessed by the time at which dialogue was performed, either in the morning or in the afternoon), was however not evidenced. The absence of a significant impact of the delay between dialogue and the night of sleep can be explained by several factors. First, we did not control for the time at which participants went to sleep or for the duration of the sleep period of our participants. We referred only to the time at which the dialogue took place in the day to infer the delay between the interaction and the time of sleep. Therefore, we cannot be completely sure that all the participants having performed the dialogue task in the afternoon experienced less decay between the time of dialogue and the onset of the offline consolidation. This may have caused variability in the data and erased the potential effects of delay. However, all these assumptions are conditional on the fact that the new lexico-semantic information was first stored in a context-dependent format, which again was not tested here as it was not our primary objective.

To summarise these first two sections, we believe that the results from our first set of experiments can be explained mostly by the contextual-binding account, although the present thesis did not aim at validating it. In the next section, we focus on the results from the second set of experiments in light of the literature on the typicality effect.

3.1.3 Changes in the organisation of semantic categories

To our knowledge, very few studies have focused on immediate changes in the organisation of semantic representations in terms of category structure and these changes have never been evaluated in terms of direct modifications in abstract lexico-semantic representations. In 2019, Dieciuc and Folstein reviewed the literature on the distinction between the stable structure of these semantic categories, called ‘structural typicality’, and their more adaptative nature, named ‘functional typicality’. They suggested that semantic categories, as derived from decontextualized norms, are rather stable and resistant to change because they are built on the basis of correlations of features in the environment that do not change drastically on a daily basis. On the other hand, functional typicality allows for the representation of goal-derived categories, which are created on the fly but are not reflected in decontextualized norms of typicality (see also Barsalou, 1985). As an example, while the ‘bird’ category is a stable one because it is based on a cluster of features such as ‘have wings’ or ‘can fly’, goal driven categories such as ‘things to take camping’ are created on the fly because their members are all related to the same goal (e.g. ‘tent’ and ‘flash-light’ are both related to ‘camping’). These functional categories were suggested to be the consequence of the process of semantic representation in working memory (Folstein & Dieciuc, 2019). As the literature on working memory proposes the involvement of the hippocampus as part of the neural basis for this cognitive function (Li et al., 2023), we propose that the distinction between functional and structural typicality works in the same way as any other context-specific information that may bias the recognition of lexico-semantic representations. The context of an utterance may thus be stored in the working memory (represented at least in part by the hippocampus), in order to bias the process of lexico-semantic representations and create a functional semantic category. As such, the creation of a functional category and the way it is used to interpret subsequent lexico-semantic information could be, at least in part, consolidated into long-term lexico-semantic representations to apply slight changes to structural typicality. Such a hypothesis could account for the impact of expertise on the typicality effect already mentioned in the introduction of this manuscript (Lynch et al., 2000). In sum, rather than considering structural and functional typicality as two distinct mechanisms, we suggest that they are complementary stages of the integration of information.

Coming back to the two experiments presented in Chapter 6, although we did not find significant changes in the lexico-semantic representations following dialogue, we can still speculate on the changes that could have occurred during dialogue in light of the previously presented literature. In these experiments, we assessed whether dialogue could directly strengthen the connections between typical and atypical members of a semantic category by reinforcing the featural similarity between them. This idea is in line with the first set of experiments presented in Chapter 5 and with the general account of direct modulation of lexico-semantic representations stored in semantic memory. However, as already discussed, it is possible that the changes induced by dialogue (if we were able to evidence them) may first be stored in a less permanent format. If we transpose this idea to the literature on structural and functional typicality, it is possible that instead of directly modulating the structural typicality, dialogue may create a new functional category stored in a context-specific format before being integrated into long-term lexico-semantic representations after a night of consolidation. Crucially, as the information related to this new functional category is part of what has been discussed, then they are also part of the so-called ‘common ground’ where every discussed information is stored as a mutually shared information. As a consequence, common ground may play the same role as any other linguistic context in constraining the interpretation of lexico-semantic information and impacting long-term representations. This idea that common ground serves as the storage for context-specific information during dialogue is further discussed in the following section.

In these first three sections, we attempted to see what could have been the interpretations of our experiments in light of the literature on the contextual binding account if we had more significant results. Although we did not attempt to evaluate whether the changes in lexico-semantic representations were first stored in a temporary format before being consolidated, we found it interesting to explore this hypothesis as our experiments were not designed to deny this account and could even be compatible with it. The contextual binding account is even more relevant to explore when questioning the relationship between dialogue and memory representations, as it has common points with some models from the literature on dialogue. In the next section, we thus take a moment to go beyond the results of this thesis and discuss the mutual contribution of the contextual binding account and some models proposed in the literature on dialogue to the question of the impact of dialogue on lexico-semantic representations.

3.2 Changes in lexico-semantic representations due to dialogue

3.2.1 Insights from the memory-based and the constraint-based models

The idea according to which contextual information is stored temporarily in working memory to guide the process of long-term memory representations in sentence production and comprehension is well-supported in both dialogue studies (Brennan & Hanna, 2009; Hanna et al., 2003; Hanna & Tannenhaus, 2004; Horton & Gerrig, 2005a; 2005b; Yoon & Brown-Schmidt, 2019) and in studies of non-interactive sentence comprehension (Gaskell et al., 2019; Mak et al., 2023; Rodd et al., 2016; Van Berkum et al., 2005). In the field of dialogue, the memory-based model (Horton & Gerrig, 2005a) already attempted to bridge the gap between ordinary memory processes and context-specific information by proposing the episodic memory as the ordinary memory register that encodes these contextual experiences into the common ground. Crucially, Duff and Brown-Schmidt (2017) went a step further by suggesting the hippocampus as a candidate for supporting (at least in part) the creation of common ground in dialogue. Specifically, they proposed that context-specific memory representations stored in the hippocampus constitute the neural basis for common ground processes (Duff & Brown-Schmidt, 2017; Duff et al., 2011). In parallel, research in non-interactive sentence comprehension has shown that such hippocampal representations can be consolidated into long-term lexico-semantic representations (Gaskell et al., 2019). Taken together, evidence from these two fields suggests that context-specific lexico-semantic representations stored in the common ground during dialogue can be consolidated to affect long-term lexico-semantic representations.

Importantly, the content of the common ground was not directly tested in the present thesis. However, since every shared information during dialogue can be considered as part of the common ground, by examining the impact of dialogue on lexico-semantic representations we may have indirectly tested the impact of common ground on these memory representations. Given the significant changes observed in lexico-semantic representations one day after dialogue in the present thesis, we may have provided evidence that common ground, whether it requires consolidation or not, impacts lexico-semantic representations. This suggestion should however be approached with caution, as our dyads of participants interacted solely through an audio-conference device. Although limited in number, studies comparing face-to-face and phone-mediated conversations indeed highlighted differences in linguistic content, such as the way in which participants reuse references (Knutsen et al., 2016). It is therefore possible that the common ground created in our phone-mediated setting differs in nature from that created in face-to-face conversations. In the next section, we further discuss our results in relation with the notion of common ground and dialogue models of memory representations (Horton and Gerrig, 2005a; 2005b).

3.2.2 Rethinking the partner-specificity of the common ground

In the introduction of this manuscript, we reviewed the literature on dialogue and the fact that a speaker tends to adjust his/her utterances to his/her partner to ensure the understanding of the listener, who in turn interprets the linguistic content in light of the knowledge they are both aware of sharing (Brennan & Clark, 1996; Metzing & Brennan, 2003). From this proposition emerged the memory-based and the constraint-based models which state that this partner-specificity originates from the constraint exerted by the presence of the partner as a contextual cue for the retrieval of specific linguistic information (Hanna & Tanenhaus, 2004; Hanna et al., 2003; Horton & Brennan, 2016; Horton & Gerrig, 2005a; 2005b). According to these models, the partner-specificity of the conceptual pact arises from episodic traces that constrain the retrieval of long-term memory representations. However, in light of the results presented in this thesis, it seems reasonable to suggest that new context-specific representations may not only impact coordination immediately after their formation but may also lead to adjustments in subsequent dialogues with different partners. Indeed, we demonstrated that information shared during a dialogue between A and B impacts long-term lexico-semantic representations in such a way that they remain changed in A's semantic memory one day later. Consequently, it is possible that in subsequent dialogues, the way A interacts with C might differ after A's discussion with B, compared to before, due to the updates in the state of A's lexico-semantic representations. Although this proposition may be seen as being in contradiction with studies showing that the use of common ground is partner-specific, we do not believe that it is the case.

The partner-specificity of conceptual pacts has always been assessed either by using linguistic measures (i.e. the number of words, the type of content words or the use of definite references), or by exploring gaze behaviours in reaction to particular content words (Brennan & Clark, 1996; Metzing & Brennan, 2003; see also Bovet et al., 2023 for a review). For instance, it is known that speakers tend to use more definite references as soon as an object has already been discussed with a given partner, and to reuse always the same words to designate the object, in order to indicate that the reference belongs to their common ground (see also 1.1.3 from Chapter 1). As soon as speakers change partners, they tend to reuse indefinite references and to increase the number of words to refer to the same objects, until a new partner-specific conceptual pact is found. However, these measures of the conceptual pact reflect only the episodic traces mentioned by the memory-based model and not the use of more abstract representations. In the previous paragraph, we suggested that these episodic traces could change the organisation of lexico-semantic representations after a night of sleep, and this updated version of the lexico-semantic network in turn affects subsequent encounters with the same object, independently of the partner. However, this influence has never been assessed in the long term. It could thus be interesting to test whether a conceptual pact established on day one with a specific partner and potentially re-established across multiple conversations could impact the linguistic content of a conversation with a different partner one week later. The idea of re-establishing the conceptual pact across several conversations over consecutive days stems from previous evidence suggesting that an extensive training on the new lexico-semantic representation may be necessary to evidence changes in tasks that do not focus on the newly acquired lexico-semantic information, as it can be the case in most dialogue tasks (Rodd et al., 2012). However, the assumption that several dialogues may work the same way as several training sessions on the semantic properties should first be tested.

To conclude on section 3, we have seen that the context-specific hypothesis on lexico-semantic representations is compatible with the few significant results that we found in our experiments. It provides an interesting view on what may occur during a dialogue in terms of impact on long-term lexico-semantic representations. We also highlighted the similarities between this approach and other approaches of dialogue. However, these discussions are conditional on the idea that dialogue creates hippocampus-based context-specific representations that are only later consolidated into long-term representations through offline processes. It is therefore necessary to conduct further studies in order to test this idea. For instance, it could be interesting to replicate the experiments by Gaskell et al. (2019) while integrating the disambiguation sentences in a semi-scripted dialogue. In the following section, we review the implications of the previously discussed memory mechanisms in the literature on dialogue to suggest future lines of research in this domain.

4. Implications for the interactive alignment theory

In the interactive alignment model, dialogue partners reach mutual understanding by aligning their representations at various levels (Pickering & Garrod, 2004). Pickering and Garrod (2021) posited that while alignment at the level of the situation model is more akin to alignment in episodic memory, linguistic alignment is equivalent to alignment in semantic memory. Specifically, when dialogue partners interact, linguistic representations at the syntactic, phonological and semantic levels are primed either by the linguistic input from the speaker or by the activation at one of these three levels. If such priming produces a significant amount of activation, it triggers the activation of the representation. The consequence of the co-activation of representations is the strengthening of the relationship between the co-activated nodes. The idea by which dialogue leads to these strengthened relationships between co-activated nodes implies that dialogue may directly impact long-term representations.

In the present set of experiments, the results of Experiment 4 are in line with the idea that dialogue changed the strength of connections between lexico-semantic representations. They showed that by repeatedly associating two semantically unrelated words with a common visual representation, these words became more semantically related, as indicated by the semantic relatedness judgement task. In light of the model by Pickering and Garrod (2004; 2021), we can speculate that these results are a consequence of the alignment produced between the two partners' representations, although alignment was not directly assessed in this thesis. During dialogue, the enunciation of the pairs of unrelated words by the confederate in relation with their new common visual representation may have primed their co-activation in the linguistic representations of the naive participant. The consequence of this co-activation is the strengthening of the lexico-semantic relationship between these pairs of words, which may remain in the long-term or after one night of sleep at least, as it was the case in our experiments. It could be argued that the absence of a significant priming effect in the lexical decision task does not support this idea, as the strengthening of the new lexico-semantic relationship should have induced the spreading of activation from one word of the pair towards the other. However, as mentioned earlier in this discussion, the changes occurring in semantic memory may simply be too weak to be observable in a lexical decision task.

While modest, our results support the idea that dialogue impacts lexico-semantic representations stored in long-term memory. They are therefore in line with the view that linguistic representations stored in semantic memory may receive specific activations during dialogue, leading to changes in connection weights after dialogue. Although the present thesis did not aim at testing the nature of these activations during dialogue (i.e. whether or not they come from a priming mechanism) and the theory of Pickering and Garrod (2004; 2021), it still provides experimental evidence in line with their model.

5 – General conclusion

The aim of this thesis was to bridge the gap between the literature on the relationship between dialogue and memory, and the literature on the adaptation of lexico-semantic representations. Our goal was to determine whether lexico-semantic representations could be changed after a dialogue, either by including new lexico-semantic relationships or by changing the internal organisation of semantic categories. We also aimed at determining whether these potential changes occurring after only one dialogue could last over time. Finally, the novelty of our approach also lay in how the changes in lexico-semantic representations were measured, as we focused on assessing changes in *access* to these representations.

To answer our research question, we created several new methodological tools such as a database of tangram pictures and two new dialogue tasks (i.e. the IAR task and the PBM task), which will probably be also useful for future dialogue studies. Our results demonstrate only partially the impact of dialogue on lexico-semantic representations. Indeed, we were able to evidence the creation of new lexico-semantic relationships after dialogue, but only when measuring the changes with a task that explicitly focused the participants' attention on the semantic properties of words. Crucially, these changes were observed after one night of sleep, suggesting that they were consolidated through offline processes.

Our insights on these results are in line with the idea that dialogue can produce changes in the lexico-semantic network. However, these changes are too weak to be observed with an implicit task such as a lexical decision task. We therefore proposed several directions as to how the present methodology could be improved to provide more consistent evidence of the relationship between dialogue and lexico-semantic representations stored in long-term memory. Finally, we discussed these results in light of the models of dialogue that we considered the most relevant to explain the present data and more generally the findings of the literature on lexico-semantic adaptations.

In sum, although our results do not answer all the questions that arise concerning the impact of dialogue on lexico-semantic representations, this thesis provides interesting tools and methodologies for future studies on dialogue and access to lexico-semantic representations. As such, it constitutes a crucial step towards a better understanding of the memory representations that sustain one of the most common activities in our everyday life: dialogue.

References

Achim, A. M., Fossard, M., Couture, S., & Achim, A. (2015). Adjustment of speaker's referential expressions to an addressee's likely knowledge and link with theory of mind abilities. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00823

Alario, F.-X., & Ferrand, L. (1999). A set of 400 pictures standardized for French: Norms for name agreement, image agreement, familiarity, visual complexity, image variability, and age of acquisition. *Behavior Research Methods, Instruments, & Computers*, *31*(3), 531-552. https://doi.org/10.3758/BF03200732

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the Time Course of Spoken Word Recognition Using Eye Movements: Evidence for Continuous Mapping Models. *Journal of Memory and Language, 38*(4), 419-439. https://doi.org/10.1006/jmla.1997.2558

Alves, M., & Raposo, A. (2015). Is it a bird? Differential effects of concept typicality on semantic memory and episodic recollection. *Revista Portuguesa de Psicologia*, *44*, 65-79. https://doi.org/10.21631/rpp44_65

American Psychological Association. (2018). Language. In *APA dictionary of psychology*. https://dictionary.apa.org/language

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, *22*(3), 261-295. https://doi.org/10.1016/S0022-5371(83)90201-3

Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological Review*, *116*(3), 463-498. https://doi.org/10.1037/a0016261

Apperly, I. A. (2012). What is "theory of mind"? Concepts, cognitive processes and individual differences. *Quarterly Journal of Experimental Psychology*, *65*(5), 825-839. https://doi.org/10.1080/17470218.2012.676055

Armstrong, B. C., & Plaut, D. C. (2008). *Settling Dynamics in Distributed Networks Explain Task Differences in Semantic Ambiguity Effects: Computational and Behavioral Evidence*.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human Memory: A Proposed System and its Control Processes1. In K. W. Spence & J. T. Spence (Éds.), *Psychology of Learning and Motivation* (Vol. 2, p. 89-195). Academic Press. https://doi.org/10.1016/S0079-7421(08)60422-3

Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics using R*. Cambridge University Press.

Baayen, R. H., McQueen, J. M., Dijkstra, T., & Schreuder, R. (2003). Frequency effects in regular inflectional morphology: Revisiting Dutch plurals. In *Frequency effects in regular inflectional morphology: Revisiting Dutch plurals* (p. 355-390). De Gruyter Mouton. https://doi.org/10.1515/9783110910186.355

Bailenson, J. N., Shum, M. S., Atran, S., Medin, D. L., & Coley, J. D. (2002). A bird's eye view: Biological categorization and reasoning within and across cultures. *Cognition*, *84*(1), 1-53. https://doi.org/10.1016/S0010-0277(02)00011-2

Bangerter, A., Mayor, E., & Knutsen, D. (2020). Lexical entrainment without conceptual pacts? Revisiting the matching task. *Journal of Memory and Language*, *114*, 104129. https://doi.org/10.1016/j.jml.2020.104129

Barr, D. J., & Keysar, B. (2002). Anchoring Comprehension in Linguistic Precedents. *Journal of Memory and Language*, *46*(2), 391-418. https://doi.org/10.1006/jmla.2001.2815

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(4), 629-654. https://doi.org/10.1037/0278-7393.11.1-4.629

Bendor, D., & Wilson, M. A. (2012). Biasing the content of hippocampal replay during sleep. *Nature Neuroscience*, *15*(10), 1439-1444. https://doi.org/10.1038/nn.3203

Betts, H. N., Gilbert, R. A., Cai, Z. G., Okedara, Z. B., & Rodd, J. M. (2018). Retuning of Lexical-Semantic Representations: Repetition and Spacing Effects in Word-Meaning Priming. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *44*(7), 1130-1150. https://doi.org/10.1037/xlm0000507

Bonin, P., Peereman, R., Malardier, N., Méot, A., & Chalard, M. (2003). A new set of 299 pictures for psycholinguistic studies: French norms for name agreement, image agreement, conceptual familiarity, visual complexity, image variability, age of acquisition, and naming latencies. *Behavior Research Methods, Instruments, & Computers*, *35*(1), 158-167. https://doi.org/10.3758/BF03195507

Bovet, V., Knutsen, D., & Fossard, M. (2024). Direct and indirect linguistic measures of common ground in dialogue studies involving a matching task: A systematic review. *Psychonomic Bulletin & Review*, *31*(1), 122-136. https://doi.org/10.3758/s13423-023-02359-2

Branigan, H. P., Catchpole, C. M., & Pickering, M. J. (2011). What makes dialogues easy to understand? *Language and Cognitive Processes*, *26*(10), 1667-1686. https://doi.org/10.1080/01690965.2010.524765

Branigan, H. P., Pickering, M. J., Stewart, A. J., & Mclean, J. F. (2000). Syntactic priming in spoken production: Linguistic and temporal interference. *Memory & Cognition*, *28*(8), 1297-1302. https://doi.org/10.3758/BF03211830

Braun, S., Rosseel, Y., Kempenaers, C., Loas, G., & Linkowski, P. (2015). Self-Report of Empathy: A Shortened French Adaptation of the Interpersonal Reactivity Index (IRI) Using Two Large Belgian Samples. *Psychological Reports*, *117*(3), 735-753. https://doi.org/10.2466/08.02.PR0.117c23z6

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

Brennan, S. E., & Hanna, J. E. (2009). Partner-Specific Adaptation in Dialog. *Topics in Cognitive Science*, *1*(2), 274-291. https://doi.org/10.1111/j.1756-8765.2009.01019.x

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ*, *8*, e9414. https://doi.org/10.7717/peerj.9414

Brodeur, M. B., Guérard, K., & Bouras, M. (2014). Bank of Standardized Stimuli (BOSS) Phase II : 930 New Normative Photos. *PLOS ONE*, *9*(9), e106953. https://doi.org/10.1371/journal.pone.0106953

Brown, P. M., & Dell, G. S. (1987). Adapting production to comprehension: The explicit mention of instruments. *Cognitive Psychology*, *19*(4), 441-472. https://doi.org/10.1016/0010-0285(87)90015-6

Brown, R. (1958). How shall a thing be called? *Psychological Review*, *65*(1), 14-21. https://doi.org/10.1037/h0041727

Brown-Schmidt, S. (2012). Beyond common and privileged: Gradient representations of common ground in real-time language use. *Language and Cognitive Processes*, *27*(1), 62-89. https://doi.org/10.1080/01690965.2010.543363

Brunellière, A., & Bonnotte, I. (2018). To what extent does typicality boost semantic priming effects between members of their categories? *Journal of Cognitive Psychology*, *30*(7), 670-688. https://doi.org/10.1080/20445911.2018.1523174

Brunellière, A., Perre, L., Tran, T., & Bonnotte, I. (2017). Co-occurrence frequency evaluated with large language corpora boosts semantic priming effects. *Quarterly Journal of Experimental Psychology*, *70*(9), 1922-1934. https://doi.org/10.1080/17470218.2016.1215479

Carroll, M., & Kirsner, K. (1982). Context and repetition effects in lexical decision and recognition memory. *Journal of Verbal Learning and Verbal Behavior*, *21*(1), 55-69. https://doi.org/10.1016/S0022-5371(82)90445-5

Casey, P. J. (1992). A reexamination of the roles of typicality and category dominance in verifying category membership. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(4), 823-834. https://doi.org/10.1037/0278-7393.18.4.823

Ceschi, G., & Pictet, A. (2018). *Imagerie mentale et psychothérapie : Un ouvrage sur la psychopathologie cognitive*. Mardaga.

Clark, H. H. (1996). *Using Language*. Cambridge University Press.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In *Perspectives on socially shared cognition* (p. 127-149). American Psychological Association. https://doi.org/10.1037/10096-006

Clark, H. H., & Marshall, C. R. (1981). *Definite Reference and Mutual Knowledge*.

Clark, H. H., & Murphy, G. L. (1982). Audience Design in Meaning and Reference. In J.-F. Le Ny & W. Kintsch (Éds.), *Advances in Psychology* (Vol. 9, p. 287-299). North-Holland. https://doi.org/10.1016/S0166-4115(09)60059-5

Clark, H. H., & Schaefer, E. F. (1989). Collaborating on Contributions to Conversations. In R. Dietrich & C. F. Graumann (Éds.), *North-Holland Linguistic Series: Linguistic Variations* (Vol. 54, p. 123-152). Elsevier. https://doi.org/10.1016/B978-0-444-87144-2.50008-2

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

Cleland, A. A., & Pickering, M. J. (2003). The use of lexical and syntactic information in language production: Evidence from the priming of noun-phrase structure. *Journal of Memory and Language*, *49*(2), 214-230. https://doi.org/10.1016/S0749-596X(03)00060-3

Cohen, N. J., & Squire, L. R. (1980). Preserved learning and retention of pattern-analyzing skill in amnesia: Dissociation of knowing how and knowing that. *Science*, *210*(4466), 207-210. https://doi.org/10.1126/science.7414331

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407-428. https://doi.org/10.1037/0033-295X.82.6.407

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *8*(2), 240-247. https://doi.org/10.1016/S0022-5371(69)80069-1

Dagenbach, D., Horst, S., & Carr, T. H. (1990). Adding new information to semantic memory: How much learning is enough to produce automatic priming? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(4), 581-591. https://doi.org/10.1037/0278-7393.16.4.581

Davis, M. H. (1980). *Interpersonal Reactivity Index*. http://doi.apa.org/getdoi.cfm?doi=10.1037/t01093-000

Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning: Neural and behavioural evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1536), 3773-3800. https://doi.org/10.1098/rstb.2009.0111

De Deyne, S., Verheyen, S., & Storms, G. (2016). Structure and Organization of the Mental Lexicon: A Network Approach Derived from Syntactic Dependency Relations and Word

Associations. In A. Mehler, A. Lücking, S. Banisch, P. Blanchard, & B. Job (Éds.), *Towards a Theoretical Framework for Analyzing Complex Linguistic Networks* (p. 47-79). Springer. https://doi.org/10.1007/978-3-662-47238-5_3

de Groot, A. M. B. (1984). Primed lexical decision: Combined effects of the proportion of related prime-target pairs and the stimulus-onset asynchrony of prime and target. *The Quarterly Journal of Experimental Psychology Section A*, *36*(2), 253-280. https://doi.org/10.1080/14640748408402158

Dehaene, S., Changeux, J.-P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: A testable taxonomy. *Trends in Cognitive Sciences*, *10*(5), 204-211. https://doi.org/10.1016/j.tics.2006.03.007

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

de Wit, B., & Kinoshita, S. (2014). Relatedness proportion effects in semantic categorization: Reconsidering the automatic spreading activation process. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(6), 1733-1744. https://doi.org/10.1037/xlm0000004

de Wit, B., & Kinoshita, S. (2015a). An RT distribution analysis of relatedness proportion effects in lexical decision and semantic categorization reveals different mechanisms. *Memory & Cognition*, *43*(1), 99-110. https://doi.org/10.3758/s13421-014-0446-6

de Wit, B., & Kinoshita, S. (2015b). The masked semantic priming effect is task dependent: Reconsidering the automatic spreading activation process. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(4), 1062-1075. https://doi.org/10.1037/xlm0000074

Dieciuc, M. A., & Folstein, J. R. (2019). Typicality: Stable structures and flexible functions. *Psychonomic Bulletin & Review*, *26*(2), 491-505. https://doi.org/10.3758/s13423-018-1546-2

Diependaele, K., Brysbaert, M., & Neri, P. (2012). How Noisy is Lexical Decision? *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00348

Dijk, T. A. van, & Kintsch, W. (1983). *Strategies of Discourse Comprehension*. Academic Press.

Dimitropoulou, M., Duñabeitia, J. A., Blitsas, P., & Carreiras, M. (2009). A standardized set of 260 pictures for Modern Greek: Norms for name agreement, age of acquisition, and visual complexity. *Behavior Research Methods*, *41*(2), 584-589. https://doi.org/10.3758/BRM.41.2.584

Doherty-Sneddon, G., Anderson, A., O'Malley, C., Langton, S., Garrod, S., & Bruce, V. (1997). Face-to-face and video-mediated communication: A comparison of dialogue structure and task performance. *Journal of Experimental Psychology: Applied*, *3*(2), 105-125. https://doi.org/10.1037/1076-898X.3.2.105

Dubois, D., & Poitou, J. (2002). *Des « normes catégorielles » : Structuration cognitive et/ou linguistique des catégories sémantiques*. https://doi.org/10.3406/intel.2002.1668

Duff, M. C., & Brown-Schmidt, S. (2017). Hippocampal Contributions to Language Use and Processing. In D. E. Hannula & M. C. Duff (Éds.), *The Hippocampus from Cells to Systems : Structure, Connectivity, and Functional Contributions to Memory and Flexible Cognition* (p. 503-536). Springer International Publishing. https://doi.org/10.1007/978-3-319-50406-3_16

Duff, M. C., Gupta, R., Hengst, J. A., Tranel, D., & Cohen, N. J. (2011). The Use of Definite References Signals Declarative Memory: Evidence From Patients With Hippocampal Amnesia. *Psychological Science*, *22*(5), 666-673. https://doi.org/10.1177/0956797611404897

Duff, M. C., Hengst, J., Tranel, D., & Cohen, N. J. (2006). Development of shared information in communication despite hippocampal amnesia. *Nature Neuroscience*, *9*(1), 140-146. https://doi.org/10.1038/nn1601

Dumay, N., & Gareth Gaskell, M. (2012). Overnight lexical consolidation revealed by speech segmentation. *Cognition*, *123*(1), 119-132. https://doi.org/10.1016/j.cognition.2011.12.009

Dumay, N., & Gaskell, M. G. (2007). Sleep-Associated Changes in the Mental Representation of Spoken Words. *Psychological Science*, *18*(1), 35-39. https://doi.org/10.1111/j.1467-9280.2007.01845.x

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

Durgunoğlu, A. Y., & Neely, J. H. (1987). On obtaining episodic priming in a lexical decision task following paired-associate learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(2), 206-222. https://doi.org/10.1037/0278-7393.13.2.206

Eichenbaum, H. (2000). A cortical–hippocampal system for declarative memory. *Nature Reviews Neuroscience*, *1*(1), 41-50. https://doi.org/10.1038/35036213

Fang, X., & Perfetti, C. A. (2017). Perturbation of old knowledge precedes integration of new knowledge. *Neuropsychologia*, *99*, 270-278. https://doi.org/10.1016/j.neuropsychologia.2017.03.015

Fang, X., & Perfetti, C. A. (2019). Learning new meanings for known words: Perturbation of original meanings and retention of new meanings. *Memory & Cognition*, *47*(1), 130-144. https://doi.org/10.3758/s13421-018-0855-z

Farah, M. J., & McClelland, J. L. (1991). A computational model of semantic memory impairment: Modality specificity and emergent category specificity. *Journal of Experimental Psychology: General*, *120*(4), 339-357. https://doi.org/10.1037/0096-3445.120.4.339

Fasquel, A., Brunellière, A., & Knutsen, D. (2023). A modified procedure for naming 332 pictures and collecting norms: Using tangram pictures in psycholinguistic studies. *Behavior Research Methods*, *55*(5), 2297-2319. https://doi.org/10.3758/s13428-022-01871-y

Fenn, K. M., Nusbaum, H. C., & Margoliash, D. (2003). Consolidation during sleep of perceptual learning of spoken language. *Nature*, *425*(6958), 614-616. https://doi.org/10.1038/nature01951

Folstein, J. R., & Dieciuc, M. A. (2019). The Cognitive Neuroscience of Stable and Flexible Semantic Typicality. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.01265

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*(2), 178-210. https://doi.org/10.1016/0010-0285(82)90008-1

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

Gaskell, M. G., Cairney, S. A., & Rodd, J. M. (2019). Contextual priming of word meanings is stabilized over sleep. *Cognition*, *182*, 109-126. https://doi.org/10.1016/j.cognition.2018.09.007

Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition*, *89*(2), 105-132. https://doi.org/10.1016/S0010-0277(03)00070-2

Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating Form and Meaning: A Distributed Model of Speech Perception. *Language and Cognitive Processes*, *12*(5-6), 613-656. https://doi.org/10.1080/016909697386646

Ghasisin, L., Yadegari, F., Rahgozar, M., Nazari, A., & Rastegarianzade, N. (2015). A new set of 272 pictures for psycholinguistic studies: Persian norms for name agreement, image agreement, conceptual familiarity, visual complexity, and age of acquisition. *Behavior Research Methods*, *47*(4), 1148-1158. https://doi.org/10.3758/s13428-014-0537-0

Gilbert, R. A., Davis, M. H., Gaskell, M. G., & Rodd, J. M. (2018). Listeners and Readers Generalize Their Experience With Word Meanings Across Modalities. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *44*(10), 1533-1561. https://doi.org/10.1037/xlm0000532

Giles, H. (1973). Accent Mobility: A Model and Some Data. *Anthropological Linguistics*, *15*(2), 87-105.

Giles, H., & Powesland, P. F. (1975). *Speech style and social evaluation* (p. viii, 218). Academic Press.

Giles, H., Taylor, D. M., & Bourhis, R. (1973). Towards a theory of interpersonal accommodation through language: Some Canadian data. *Language in Society*, *2*(2), 177-192. https://doi.org/10.1017/S0047404500000701

Gilmore, N., Meier, E. L., Johnson, J. P., & Kiran, S. (2018). Typicality-based semantic treatment for anomia results in multiple levels of generalisation. *Neuropsychological Rehabilitation*, *30*(5), 802-828. https://doi.org/10.1080/09602011.2018.1499533

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*(2), 251-279. https://doi.org/10.1037/0033-295X.105.2.251

Gómez, R. L., Bootzin, R. R., & Nadel, L. (2006). Naps Promote Abstraction in Language-Learning Infants. *Psychological Science*, *17*(8), 670-674. https://doi.org/10.1111/j.1467-9280.2006.01764.x

Hampton, J. A. (1979). Polymorphous concepts in semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *18*(4), 441-461. https://doi.org/10.1016/S0022-5371(79)90246-9

Hampton, J. A., & Gardiner, M. M. (1983). Measures of internal category structure: A correlational analysis of normative data. *British Journal of Psychology*, *74*(4), 491-516. https://doi.org/10.1111/j.2044-8295.1983.tb01882.x

Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science*, *28*(1), 105-115. https://doi.org/10.1207/s15516709cog2801_5

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. https://doi.org/10.1016/S0749-596X(03)00022-6

Hartsuiker, R. J., Bernolet, S., Schoonbaert, S., Speybroeck, S., & Vanderelst, D. (2008). Syntactic priming persists while the lexical boost decays: Evidence from written and spoken dialogue. *Journal of Memory and Language*, *58*(2), 214-238. https://doi.org/10.1016/j.jml.2007.07.003

Heller, D., & Brown-Schmidt, S. (2023). The Multiple Perspectives Theory of Mental States in Communication. *Cognitive Science*, *47*(7), e13322. https://doi.org/10.1111/cogs.13322

Heyman, T., Bruninx, A., Hutchison, K. A., & Storms, G. (2018). The (un)reliability of item-level semantic priming effects. *Behavior Research Methods*, *50*(6), 2173-2183. https://doi.org/10.3758/s13428-018-1040-9

Hoffman, P., McClelland, J. L., & Lambon Ralph, M. A. (2018). Concepts, Control, and Context: A Connectionist Account of Normal and Disordered Semantic Cognition. *Psychological Review*, *125*(3), 293-328. https://doi.org/10.1037/rev0000094

Horton, W. S., & Gerrig, R. J. (2002). Speakers' experiences and audience design: Knowing *when* and knowing *how* to adjust utterances to addressees. *Journal of Memory and Language*, *47*(4), 589-606. https://doi.org/10.1016/S0749-596X(02)00019-0

Horton, W. S., & Gerrig, R. J. (2005a). Conversational Common Ground and Memory Processes in Language Production. *Discourse Processes*, *40*(1), 1-35. https://doi.org/10.1207/s15326950dp4001_1

Horton, W. S., & Gerrig, R. J. (2005b). The impact of memory demands on audience design during language production. *Cognition*, *96*(2), 127-142. https://doi.org/10.1016/j.cognition.2004.07.001

Horton, W. S., & Gerrig, R. J. (2016). Revisiting the Memory-Based Processing Approach to Common Ground. *Topics in Cognitive Science*, *8*(4), 780-795. https://doi.org/10.1111/tops.12216

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, *59*(1), 91-117. https://doi.org/10.1016/0010-0277(96)81418-1

Hulme, R. C. (2018). Incidental learning of new meanings for familiar words [Doctoral, UCL (University College London)]. In *Doctoral thesis, UCL (University College London).* UCL (University College London). https://discovery.ucl.ac.uk/id/eprint/10061270/

Hupet, M., Seron, X., & Chantraine, Y. (1991). The effects of the codability and discriminability of the referents on the collaborative referring procedure. *British Journal of Psychology*, *82*(4), 449-462. https://doi.org/10.1111/j.2044-8295.1991.tb02412.x

Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin & Review*, *10*(4), 785-813. https://doi.org/10.3758/BF03196544

Hutchison, K. A. (2007). Attentional control and the relatedness proportion effect in semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(4), 645-662. https://doi.org/10.1037/0278-7393.33.4.645

Hutchison, K. A., Neely, J. H., & Johnson, J. D. (2001). With great expectations, can two « wrongs » prime a « right »? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1451-1463. https://doi.org/10.1037/0278-7393.27.6.1451

Jevtović, M., Duñabeitia, J. A., & Bruin, A. de. (2020). How do bilinguals switch between languages in different interactional contexts? A comparison between voluntary and mandatory language switching. *Bilingualism: Language and Cognition*, *23*(2), 401-413. https://doi.org/10.1017/S1366728919000191

Jones, M. N., Kintsch, W., & Mewhort, D. J. K. (2006). High-dimensional semantic space accounts of priming. *Journal of Memory and Language*, *55*(4), 534-552. https://doi.org/10.1016/j.jml.2006.07.003

Karttunen, L., & Peters, S. (1979). *Conventional lmplicature*. Brill. https://doi.org/10.1163/9789004368880_002

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking Perspective in Conversation: The Role of Mutual Knowledge in Comprehension. *Psychological Science*, *11*(1), 32-38. https://doi.org/10.1111/1467-9280.00211

Keysar, B., Barr, D. J., Balin, J. A., & Paek, T. S. (1998). Definite Reference and Mutual Knowledge: Process Models of Common Ground in Comprehension. *Journal of Memory and Language*, *39*(1), 1-20. https://doi.org/10.1006/jmla.1998.2563

Kiran, S. (2008). Typicality of Inanimate Category Exemplars in Aphasia Treatment: Further Evidence for Semantic Complexity. *Journal of Speech, Language, and Hearing Research*, *51*(6), 1550-1568. https://doi.org/10.1044/1092-4388(2008/07-0038)

Kiran, S., & Johnson, L. (2008). Semantic Complexity in Treatment of Naming Deficits in Aphasia: Evidence From Well-Defined Categories. *American Journal of Speech-Language Pathology*, *17*(4), 389-400. https://doi.org/10.1044/1058-0360(2008/06-0085)

Kiran, S., Ntourou, K., & Eubank, M. (2007). The effect of typicality on online category verification of inanimate category exemplars in aphasia. *Aphasiology*, *21*(9), 844-866. https://doi.org/10.1080/02687030600743564

Kiran, S., Sandberg, C., & Sebastian, R. (2011). Treatment of Category Generation and Retrieval in Aphasia: Effect of Typicality of Category Items. *Journal of Speech, Language, and Hearing Research*, *54*(4), 1101-1117. https://doi.org/10.1044/1092-4388(2010/10-0117)

Kiran, S., & Thompson, C. K. (2003). The Role of Semantic Complexity in Treatment of Naming Deficits. *Journal of Speech, Language, and Hearing Research*, *46*(4), 773-787. https://doi.org/10.1044/1092-4388(2003/061)

Klinzing, J. G., Niethard, N., & Born, J. (2019). Mechanisms of systems memory consolidation during sleep. *Nature Neuroscience*, *22*(10), 1598-1610. https://doi.org/10.1038/s41593-019-0467-3

Klooster, N. B., & Duff, M. C. (2015). Remote semantic memory is impoverished in hippocampal amnesia. *Neuropsychologia*, *79*, 42-52. https://doi.org/10.1016/j.neuropsychologia.2015.10.017

Klooster, N. B., Tranel, D., & Duff, M. C. (2020). The hippocampus and semantic memory over time. *Brain and Language*, *201*, 104711. https://doi.org/10.1016/j.bandl.2019.104711

Knutsen, D., Bangerter, A., & Mayor, E. (2019). Procedural Coordination in the Matching Task. *Collabra: Psychology*, *5*(1), 3. https://doi.org/10.1525/collabra.188

Knutsen, D., & Le Bigot, L. (2014). Capturing egocentric biases in reference reuse during collaborative dialogue. *Psychonomic Bulletin & Review*, *21*(6), 1590-1599. https://doi.org/10.3758/s13423-014-0620-7

Knutsen, D., & Le Bigot, L. (2020). The influence of conceptual (mis)match on collaborative referring in dialogue. *Psychological Research*, *84*(2), 514-527. https://doi.org/10.1007/s00426-018-1060-1

Knutsen, D., Le Bigot, L., & Ros, C. (2017). Explicit feedback from users attenuates memory biases in human-system dialogue. *International Journal of Human-Computer Studies*, *97*, 77-87. https://doi.org/10.1016/j.ijhcs.2016.09.004

Knutsen, D., Ros, C., & Le Bigot, L. (2016). Generating References in Naturalistic Face-to-Face and Phone-Mediated Dialog Settings. *Topics in Cognitive Science*, *8*(4), 796-818. https://doi.org/10.1111/tops.12218

Knutsen, D., Ros, C., & Le Bigot, L. (2018). Spoilt for choice: Initially considering several referential expressions affects subsequent referential decisions. *Language, Cognition and Neuroscience*, *33*(5), 618-632. https://doi.org/10.1080/23273798.2017.1400080

Kovalenko, L. Y., Chaumon, M., & Busch, N. A. (2012). A Pool of Pairs of Related Objects (POPORO) for Investigating Visual Semantic Integration: Behavioral and Electrophysiological Validation. *Brain Topography*, *25*(3), 272-284. https://doi.org/10.1007/s10548-011-0216-8

Krauss, R. M., & Weinheimer, S. (1964). Changes in reference phrases as a function of frequency of usage in social interaction: A preliminary study. *Psychonomic Science*, *1*(1), 113-114. https://doi.org/10.3758/BF03342817

Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology*, *4*(3), 343-346. https://doi.org/10.1037/h0023705

Krcmar, M., & Haberkorn, K. (2020). Mental Representations. In *The International Encyclopedia of Media Psychology* (p. 1-17). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119011071.iemp0191

Krippendorff, K. (2004). *Content Analysis: An Introduction to Its Methodology*. SAGE.

Kumar, A. A. (2021). Semantic memory: A review of methods, models, and current challenges. *Psychonomic Bulletin & Review*, *28*(1), 40-80. https://doi.org/10.3758/s13423-020-01792-x

Kumar, A. A., Steyvers, M., & Balota, D. A. (2022). A Critical Review of Network-Based and Distributional Approaches to Semantic Memory Structure and Processes. *Topics in Cognitive Science*, *14*(1), 54-77. https://doi.org/10.1111/tops.12548

Lechner, H. A., Squire, L. R., & Byrne, J. H. (1999). 100 Years of Consolidation—Remembering Müller and Pilzecker. *Learning & Memory*, *6*(2), 77-87. https://doi.org/10.1101/lm.6.2.77

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1-38. https://doi.org/10.1017/S0140525X99001776

Li, J., Cao, D., Yu, S., Xiao, X., Imbach, L., Stieglitz, L., Sarnthein, J., & Jiang, T. (2023). Functional specialization and interaction in the amygdala-hippocampus circuit during working memory processing. *Nature Communications*, *14*(1), 2921. https://doi.org/10.1038/s41467-023-38571-w

LimeSurvey GmbH. (n.d.). (s. d.). *LimeSurvey : An open source survey tool*. LimeSurvey GmbH. (n.d.). https://www.limesurvey.org

López-Madrona, V. J., Medina Villalon, S., Badier, J.-M., Trébuchon, A., Jayabal, V., Bartolomei, F., Carron, R., Barborica, A., Vulliémoz, S., Alario, F.-X., & Bénar, C. G. (2022). Magnetoencephalography can reveal deep brain network activities linked to memory processes. *Human Brain Mapping*, *43*(15), 4733-4749. https://doi.org/10.1002/hbm.25987

Lorch, R. F. (1982). Priming and search processes in semantic memory: A test of three models of spreading activation. *Journal of Verbal Learning and Verbal Behavior*, *21*(4), 468-492. https://doi.org/10.1016/S0022-5371(82)90736-8

Lucas, M. (2000). Semantic priming without association: A meta-analytic review. *Psychonomic Bulletin & Review*, *7*(4), 618-630. https://doi.org/10.3758/BF03212999

Lynch, E. B., Coley, J. D., & Medin, D. L. (2000). Tall is typical: Central tendency, ideal dimensions, and graded category structure among tree experts and novices. *Memory & Cognition*, *28*(1), 41-50. https://doi.org/10.3758/BF03211575

Maciejewski, G., Rodd, J. M., Mon-Williams, M., & Klepousniotou, E. (2020). The cost of learning new meanings for familiar words. *Language, Cognition and Neuroscience*, *35*(2), 188-210. https://doi.org/10.1080/23273798.2019.1642500

Mak, M. H. C., Curtis, A. J., Rodd, J. M., & Gaskell, M. G. (2023). Episodic memory and sleep are involved in the maintenance of context-specific lexical information. *Journal of Experimental Psychology. General*, *152*(11), 3087-3115. https://doi.org/10.1037/xge0001435

Malt, B. C., & Smith, E. E. (1982). The role of familiarity in determining typicality. *Memory & Cognition*, *10*(1), 69-75. https://doi.org/10.3758/BF03197627

Manoiloff, L., Artstein, M., Canavoso, M. B., Fernández, L., & Segui, J. (2010). Expanded norms for 400 experimental pictures in an Argentinean Spanish-speaking population. *Behavior Research Methods*, *42*(2), 452-460. https://doi.org/10.3758/BRM.42.2.452

Martínez-Manrique, F. (2010). On the Distinction between Semantic and Conceptual Representation. *Dialectica*, *64*(1), 57-78. https://doi.org/10.1111/j.1746-8361.2010.01226.x

Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2012). *irr: Various Coefficients of Interrater Reliability and Agreement* (p. 0.84.1). https://doi.org/10.32614/CRAN.package.irr

Mcclelland, J., & Cleeremans, A. (2009). *Connectionist Models*.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419-457. https://doi.org/10.1037/0033-295X.102.3.419

McKinley, G. L., Brown-Schmidt, S., & Benjamin, A. S. (2017). Memory for conversation and the development of common ground. *Memory & Cognition*, *45*(8), 1281-1294. https://doi.org/10.3758/s13421-017-0730-3

McNamara, T. P. (2005). *Semantic Priming: Perspectives from Memory and Word Recognition*. Psychology Press. https://doi.org/10.4324/9780203338001

McRae, K. (2004). Semantic memory: Some insights from feature-based connectionist attractor networks. In *The psychology of learning and motivation: Advances in research and theory, Vol 45* (p. 41-86). Elsevier Academic Press.

McRae, K., & Boisvert, S. (1998). Automatic semantic similarity priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(3), 558-572. https://doi.org/10.1037/0278-7393.24.3.558

McRae, K., de Sa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General*, *126*(2), 99-130. https://doi.org/10.1037/0096-3445.126.2.99

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the Influence of Thematic Fit (and Other Constraints) in On-line Sentence Comprehension. *Journal of Memory and Language*, *38*(3), 283-312. https://doi.org/10.1006/jmla.1997.2543

Mervis, C. B., Catlin, J., & Rosch, E. (1976). Relationships among goodness-of-example, category norms, and word frequency. *Bulletin of the Psychonomic Society*, *7*(3), 283-284. https://doi.org/10.3758/BF03337190

Mestres-Missé, A., Rodriguez-Fornells, A., & Münte, T. F. (2007). Watching the Brain during Meaning Acquisition. *Cerebral Cortex*, *17*(8), 1858-1866. https://doi.org/10.1093/cercor/bhl094

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. https://doi.org/10.1016/S0749-596X(03)00028-7

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*(2), 227-234. https://doi.org/10.1037/h0031564

Moreno-Martínez, F. J., & Montoro, P. R. (2012). An Ecological Alternative to Snodgrass & Vanderwart : 360 High Quality Colour Images with Norms for Seven Psycholinguistic Variables. *PLOS ONE*, *7*(5), e37527. https://doi.org/10.1371/journal.pone.0037527

Narhi-Martinez, W., Dube, B., & Golomb, J. D. (2023). Attention as a multi-level system of weights and balances. *WIREs Cognitive Science*, *14*(1), e1633. https://doi.org/10.1002/wcs.1633

Nault, D. R., Voleti, R., Nicastro, M., & Munhall, K. G. (2023). Investigating the influence of local and personal common ground on memory for conversation using an online referential communication task. *Journal of Experimental Psychology: General*, *152*(6), 1598-1621. https://doi.org/10.1037/xge0001341

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, *106*(3), 226-254. https://doi.org/10.1037/0096-3445.106.3.226

Neely, J. H., Keefe, D. E., & Ross, K. L. (1989). Semantic priming in the lexical decision task: Roles of prospective prime-generated expectancies and retrospective semantic matching. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(6), 1003-1019. https://doi.org/10.1037/0278-7393.15.6.1003

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 402-407. https://doi.org/10.3758/BF03195588

Norris, D., & Kinoshita, S. (2008). Perception as evidence accumulation and Bayesian inference: Insights from masked priming. *Journal of Experimental Psychology: General*, *137*(3), 434-455. https://doi.org/10.1037/a0012799

Olson, D. R. (1970). Language and thought: Aspects of a cognitive theory of semantics. *Psychological Review*, *77*(4), 257-273. https://doi.org/10.1037/h0029436

Osgood, C. E. (1952). The nature and measurement of meaning. *Psychological Bulletin*, *49*(3), 197-237. https://doi.org/10.1037/h0055737

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195-203. https://doi.org/10.3758/s13428-018-01193-y

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2), 169-190. https://doi.org/10.1017/S0140525X04000056

Pickering, M. J., & Garrod, S. (2021). *Understanding Dialogue: Language Use and Social Interaction*. Cambridge University Press.

Plaut, D. C. (1996). Relearning after Damage in Connectionist Networks: Toward a Theory of Rehabilitation. *Brain and Language*, *52*(1), 25-82. https://doi.org/10.1006/brln.1996.0004

Plaut, D. C., & Booth, J. R. (2000). Individual and developmental differences in semantic priming : Empirical and computational support for a single-mechanism account of lexical processing. *Psychological Review, 107*(4), 786-823. https://doi.org/10.1037/0033-295X.107.4.786

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*(3, Pt.1), 353-363. https://doi.org/10.1037/h0025953

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(4), 515-526. https://doi.org/10.1017/S0140525X00076512

Quillian, M. R. (1967). Word concepts: A theory and simulation of some basic semantic capabilities. *Behavioral Science*, *12*(5), 410-430. https://doi.org/10.1002/bs.3830120511

Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological Review*, *95*(3), 385-408. https://doi.org/10.1037/0033-295X.95.3.385

Remme, M. W. H., Bergmann, U., Alevi, D., Schreiber, S., Sprekeler, H., & Kempter, R. (2021). Hebbian plasticity in parallel synaptic pathways: A circuit mechanism for systems memory consolidation. *PLOS Computational Biology*, *17*(12), e1009681. https://doi.org/10.1371/journal.pcbi.1009681

Rodd, J. M., Berriman, R., Landau, M., Lee, T., Ho, C., Gaskell, M. G., & Davis, M. H. (2012). Learning new meanings for old words: Effects of semantic relatedness. *Memory & Cognition*, *40*(7), 1095-1108. https://doi.org/10.3758/s13421-012-0209-1

Rodd, J. M., Cai, Z. G., Betts, H. N., Hanby, B., Hutchinson, C., & Adler, A. (2016). The impact of recent and long-term experience on access to word meanings: Evidence from large-scale internet-based experiments. *Journal of Memory and Language*, *87*, 16-37. https://doi.org/10.1016/j.jml.2015.10.006

Rodd, J. M., Gaskell, M. G., & Marslen-Wilson, W. D. (2004). Modelling the effects of semantic ambiguity in word recognition. *Cognitive Science*, *28*(1), 89-104. https://doi.org/10.1207/s15516709cog2801_4

Rodd, J. M., Lopez Cutrin, B., Kirsch, H., Millar, A., & Davis, M. H. (2013). Long-term priming of the meanings of ambiguous words. *Journal of Memory and Language*, *68*(2), 180-198. https://doi.org/10.1016/j.jml.2012.08.002

Roelke, A., Franke, N., Biemann, C., Radach, R., Jacobs, A. M., & Hofmann, M. J. (2018). A novel co-occurrence-based approach to predict pure associative and semantic priming.

Psychonomic Bulletin & Review, *25*(4), 1488-1493. https://doi.org/10.3758/s13423-018-1453-6

Rogers, T. T., Patterson, K., Jefferies, E., & Lambon Ralph, M. A. (2015). Disorders of representation and control in semantic cognition: Effects of familiarity, typicality, and specificity. *Neuropsychologia*, *76*, 220-239. https://doi.org/10.1016/j.neuropsychologia.2015.04.015

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, *104*(3), 192-233. https://doi.org/10.1037/0096-3445.104.3.192

Rosch, E. H. (1973). Natural categories. *Cognitive Psychology*, *4*(3), 328-350. https://doi.org/10.1016/0010-0285(73)90017-0

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*(4), 573-605. https://doi.org/10.1016/0010-0285(75)90024-9

Rossiter, C., & Best, W. (2013). "Penguins don't fly": An investigation into the effect of typicality on picture naming in people with aphasia. *Aphasiology*, *27*(7), 784-798. https://doi.org/10.1080/02687038.2012.751579

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel Distributed Processing: Explorations in the microstructure of cognition : Vol 1. Foundations*. MIT Press. https://mitpress.mit.edu/9780262680530/parallel-distributed-processing/

Rumelhart, D., Hinton, G., & Mcclelland, J. (1986). A General Framework for Parallel Distributed Processing. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, *1*(26), 45-76.

Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics*, *2*(9), 437-442. https://doi.org/10.3758/BF03208784

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language*, *50*(4), 696. https://doi.org/10.2307/412243

Saitta, L., & Hampton, J. A. (2003). Abstraction and context in concept representation. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*(1435), 1251-1259. https://doi.org/10.1098/rstb.2003.1314

Schneider, W., Eschman, A., & Zuccolotto, A. (2002a). *E-Prime Reference Guide*. Pittsburge, PA: Psychology Software Tools.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002b). *E-prime User's Guide*. Pittsburge, PA: Psychology Software Tools.

Seeliger, K., Fritsche, M., Güçlü, U., Schoenmakers, S., Schoffelen, J.-M., Bosch, S. E., & van Gerven, M. A. J. (2018). Convolutional neural network-based encoding and decoding of visual object recognition in space and time. *NeuroImage, 180*, 253-266. https://doi.org/10.1016/j.neuroimage.2017.07.018

Shannon, C. E., & Weaver, W. (1949). *The Mathematical Theory of Communication*. University of Illinois Press.

Shelton, J. R., & Martin, R. C. (1992). How semantic is automatic semantic priming? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(6), 1191-1210. https://doi.org/10.1037/0278-7393.18.6.1191

Siegel, S., & Castellan Jr., N. J. (1988). *Nonparametric statistics for the behavioral sciences, 2nd ed* (p. xxiii, 399). Mcgraw-Hill Book Company.

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, *81*(3), 214-241. https://doi.org/10.1037/h0036351

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. https://doi.org/10.1037/0278-7393.6.2.174

Souza, C., Garrido, M. V., Horchak, O. V., & Carmo, J. C. (2022). Conceptual knowledge modulates memory recognition of common items: The selective role of item-typicality. *Memory & Cognition*, *50*(1), 77-94. https://doi.org/10.3758/s13421-021-01213-x

Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(6), 1521-1543. https://doi.org/10.1037/0278-7393.24.6.1521

Squire, L. R., & Alvarez, P. (1995). Retrograde amnesia and memory consolidation: A neurobiological perspective. *Current Opinion in Neurobiology*, *5*(2), 169-177. https://doi.org/10.1016/0959-4388(95)80023-9

Stalnaker, R. C. (1978). *Assertion*. Brill. https://doi.org/10.1163/9789004368873_013

Steyvers, M., & Tenenbaum, J. B. (2005). The Large-Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth. *Cognitive Science*, *29*(1), 41-78. https://doi.org/10.1207/s15516709cog2901_3

Stolz, J. A., Besner, D., & Carr, T. H. (2005). Implications of measures of reliability for theories of priming: Activity in semantic memory is inherently noisy and uncoordinated. *Visual Cognition*, *12*(2), 284-336. https://doi.org/10.1080/13506280444000030A

Taikh, A., & Lupker, S. J. (2020). Do visible semantic primes preactivate lexical representations? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(8), 1533-1569. https://doi.org/10.1037/xlm0000825

Tamminen, J. (2010). *Learning new words: Effects of meaning, memory consolidation, and sleep* [Phd, University of York]. https://etheses.whiterose.ac.uk/866/

Tan, L. C., & Yap, M. J. (2016). Are individual differences in masked repetition and semantic priming reliable? *Visual Cognition*, *24*(2), 182-200. https://doi.org/10.1080/13506285.2016.1214201

Taylor, J. R. (2017). Lexical Semantics. In B. Dancygier (Éd.), *The Cambridge Handbook of Cognitive Linguistics* (p. 246-261). Cambridge University Press. https://doi.org/10.1017/9781316339732.017

Tsaparina, D., Bonin, P., & Méot, A. (2011). Russian norms for name agreement, image agreement for the colorized version of the Snodgrass and Vanderwart pictures and age of acquisition, conceptual familiarity, and imageability scores for modal object names. *Behavior Research Methods*, *43*(4), 1085-1099. https://doi.org/10.3758/s13428-011-0121-9

Tulving, E. (1972). Episodic and semantic memory. In *Organization of memory* (p. xiii, 423-xiii, 423). Academic Press.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*(4), 327-352. https://doi.org/10.1037/0033-295X.84.4.327

Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating Upcoming Words in Discourse: Evidence From ERPs and Reading Times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 443-467. https://doi.org/10.1037/0278-7393.31.3.443

Wagner, U., Handke, L., & Walter, H. (2015). The relationship between trait empathy and memory formation for social vs. Non-social information. *BMC Psychology*, *3*(1), 2. https://doi.org/10.1186/s40359-015-0058-3

Weber, A., & Scharenborg, O. (2012). Models of spoken-word recognition. *WIREs Cognitive Science*, *3*(3), 387-401. https://doi.org/10.1002/wcs.1178

Winocur, G., Moscovitch, M., & Bontempi, B. (2010). Memory formation and long-term retention in humans and animals: Convergence towards a transformation account of hippocampal–neocortical interactions. *Neuropsychologia*, *48*(8), 2339-2356. https://doi.org/10.1016/j.neuropsychologia.2010.04.016

Yap, M. J., Hutchison, K. A., & Tan, L. C. (2016). Individual Differences in Semantic Priming Performance : Insights from the Semantic Priming Project. In *Big Data in Cognitive Science*. Psychology Press.

Yoon, S. O., & Brown-Schmidt, S. (2014). Adjusting conceptual pacts in three-party conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(4), 919-937. https://doi.org/10.1037/a0036161

Yoon, S. O., & Brown-Schmidt, S. (2019). Contextual Integration in Multiparty Audience Design. *Cognitive Science*, *43*(12), e12807. https://doi.org/10.1111/cogs.12807

Yoon, S. O., Duff, M. C., & Brown-Schmidt, S. (2017). Learning and using knowledge about what other people do and don’t know despite amnesia. *Cortex*, *94*, 164-175. https://doi.org/10.1016/j.cortex.2017.06.020

Zapf, A., Castell, S., Morawietz, L., & Karch, A. (2016). Measuring inter-rater reliability for nominal data – which coefficients and confidence intervals are appropriate? *BMC Medical Research Methodology*, *16*(1), 93. https://doi.org/10.1186/s12874-016-0200-9

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*(2), 162-185. https://doi.org/10.1037/0033-2909.123.2.162

Résumé substantiel en français

Le dialogue se définit comme une activité conjointe entre au moins deux personnes qui cherchent à atteindre un but commun (e.g., décider d'une destination pour les prochaines vacances ; Clark, 1996). Contrairement à une situation de monologue, les partenaires de dialogue cherchent à coordonner leurs productions verbales et leurs connaissances pour assurer une compréhension mutuelle entre partenaires. Par exemple, une personne A peut être amenée à adapter son rythme de parole, son choix de mots ou les références qu'elle utilise au partenaire de dialogue à qui elle s'adresse (Clark, 1996 ; Clark & Brennan, 1991 ; Pickering & Garrod, 2004 ; 2021 ; Sacks et al., 1974).

L'une des conséquences observées de cette coordination est l'émergence de nouveaux usages pour des mots déjà connus, comme lorsque deux individus doivent trouver un consensus sur la meilleure manière de nommer un objet. Une fois le consensus trouvé sur la référence pour désigner l'objet, les partenaires tendent à réutiliser cette référence plutôt que d'en changer (Brennan & Clark, 1996). Ces deux phénomènes, l'attribution de nouvelles significations à des mots déjà connus (i.e. nommé pacte conceptuel ; Brennan & Clark, 1996) et l'utilisation répétée de mots pour référer aux objets au cours d'une conversation, font du dialogue une situation idéale pour l'intégration de nouvelles significations et de nouveaux usages en mémoire. Cependant, la possibilité que ces ajustements des comportements langagiers aient une influence sur les connaissances déjà présentes en mémoire concernant la signification des mots, n'a jamais été testée.

La question d'une possible adaptation des connaissances langagières préalables a déjà été abordée au cours d'études en compréhension isolée de phrases orales ou de textes écrits.

Ces études montrent que les représentations du sens des mots stockées en mémoire, appelées représentations sémantiques, peuvent être modifiées pour nous permettre de nous adapter aux contraintes de l'environnement et d'en maximiser la compréhension (Rodd et al., 2012 ; 2013 ; 2016). Cependant, ces adaptations n'ont encore jamais été démontrées suite à une situation de dialogue. L'objectif de cette thèse est donc de déterminer dans quelle mesure des adaptations peuvent s'opérer au niveau des représentations sémantiques en lien avec leur forme lexicale (ci-après nommées représentations lexico-sémantiques) suite à une situation de dialogue. Plus spécifiquement, nous nous intéressons à déterminer dans quelle mesure l'organisation de ces représentations lexico-sémantique, qui peut être pensée sous forme de réseau de représentations inter-connectées (Collins et Loftus, 1975), ou de catégories sémantiques (Rosch, 1975), peut être adaptée suite à un dialogue. De plus, étant donné que ce type d'adaptations peuvent bénéficier d'une nuit de sommeil pour être consolidées en mémoire et devenir plus résistantes à toute forme d'interférence (Gaskell et al., 2019), ces adaptations seront testées immédiatement après le dialogue mais aussi un jour plus tard.

Introduction générale

Le dialogue

Le dialogue, tel que défini dans le premier chapitre de cette thèse, nécessite d'ajuster ses usages linguistiques à son partenaire pour garantir une compréhension mutuelle. Les auteurs en psychologie du dialogue postulent que ces ajustements se fassent sur la base des connaissances partagées entre les interlocuteurs, incluant par exemple le contenu des informations précédemment fournies pendant le dialogue. Ces connaissances partagées sont alors considérées comme faisant partie du terrain commun, qui constitue l'ensemble des connaissances que les partenaires considèrent comme partagées (Clark & Marshall, 1981 ; Clark & Murphy, 1982). De par sa nature, le terrain commun est considéré comme étant spécifique au partenaire avec qui il est construit, c'est à dire que son contenu différera en fonction du partenaire avec qui l'interaction se fait. Cette notion de spécificité au partenaire est particulièrement importante dans la mesure où elle implique que les usages linguistiques pourront être relativement répétitifs tant que le partenaire avec qui le dialogue se fait reste inchangé, de par la contrainte exercée par leur terrain commun (Brennan and Clark, 1996 ; Clark & Wilkes-Gibbs, 1986). Par exemple, lorsque deux locuteurs établissent un pacte conceptuel, ils le font de manière spécifique de sorte que chaque fois qu'ils feront référence à l'objet du pacte conceptuel, le même mot ou la même expression seront toujours utilisés pour référer à cet objet.

Alors que le terrain commun a longtemps été considéré comme indispensable à la coordination des partenaires d'un dialogue, certains auteurs ont cherché à remettre en question cet usage systématique du terrain commun en pointant du doigt le coût cognitif que cela représenterait de mobiliser les connaissances partagées à chaque choix d'usage linguistique (Barr & Keysar, 2002 ; Keysar et al., 1998 ; Keysar et al., 2000). D'un point de vue cognitif, cette remise en question souligne la nécessité de comprendre la manière dont les informations linguistiques fournies au cours d'un dialogue seraient stockées en mémoire pour être ultérieurement récupérées. Partant de ce questionnement, des auteurs ont alors créé un modèle cognitif de la mémoire telle qu'elle serait utilisée au cours d'un dialogue (Horton & Gerrig, 2005a). Plus précisément, le modèle de Horton et Gerrig (2005a ; 2005b) postule que les représentations utilisées au cours d'un dialogue seraient associées en mémoire au contexte spatio-temporel dans lequel elles ont été encodées, ce qui inclus l'information associée au partenaire et au contexte dans lequel le dialogue s'est tenu. Ce modèle est le premier modèle qui s'appuie sur des fonctions cognitives communes à d'autres activités pour expliquer les comportements linguistiques observés en dialogue. Il fournit également un ensemble de mécanismes par lesquelles les représentations stockées en mémoire sont organisées et récupérées afin d'être réutilisées lors de la production d'un énoncé. Cependant, le modèle ne mentionne pas dans quelle mesure des représentations linguistiques déjà existantes en mémoire, telles que les représentations sémantiques (i.e. sens du mot), phonologiques (i.e. prononciation du mot) ou encore syntaxiques (i.e. organisation du mot au sein de la phrase) qui sont mobilisées pendant le dialogue, pourraient être impliqués dans la création de ces nouvelles traces en mémoire.

D'autres modèles ont depuis émergé pour expliquer les comportements linguistiques en dialogue par des mécanismes cognitifs généraux, mais cette fois en prenant en compte les représentations linguistiques déjà existantes. Parmi ces modèles, on retrouve par exemple celui de Heller et Brown-Schmidt (2023) qui apporte un regard nouveau sur le type de représentations qui peuvent être sollicitées en mémoire pour soutenir les usages linguistiques en dialogue. En effet, il est proposé que les représentations soutenant le dialogue ne soient pas uniquement des représentations marquées par le contexte spatio-temporel dans lequel elles ont été rencontrées, mais que des représentations plus abstraites (i.e. indépendantes du contexte spatio-temporel) pourraient également servir à garantir la coordination entre les interlocuteurs. Le modèle propose également l'existence d'une architecture cognitive s'appuyant sur trois composants (une représentation du soi, une représentation d'autrui et un processus de comparaison entre les deux), lesquels s'appuieraient sur les différents types de représentations précédemment évoqués afin de maximiser la coordination entre les interlocuteurs. Cependant, le modèle ne précise pas la manière dont chaque composant pourrait exploiter ces différents types de représentations.

En somme, bien que les modèles de Heller et Brown-Schmidt (2023) et Horton & Gerrig (2005a) suggèrent des mécanismes explicatifs de l'utilisation des fonctions cognitives dans la coordination en dialogue, la question de l'implication des représentations linguistiques préexistantes dans ces mécanismes demeure entière. C'est pourquoi, dans la seconde partie du Chapitre 1 de cette thèse, la proposition de Pickering et Garrod (2004 ; 2021) est présentée car elle offre une explication plus précise en termes de mécanismes utilisés et du rôle des représentations linguistiques pré-existantes dans la coordination. En effet, le modèle de Pickering et Garrod (2004, 2021) postule les partenaires d'un dialogue développent une compréhension mutuelle et se coordonnent en augmentant la similarité de leurs représentations mentales. Plus spécifiquement, il est proposé dans leur modèle que les représentations linguistiques pré-existantes, telles que les représentations sémantiques, partagent le même état d'activation entre les partenaires de dialogue pour s'aligner les unes aux autres et permettre ainsi la compréhension mutuelle. Ce mécanisme d'alignement fonctionnerait de manière relativement automatique de sorte que l'activation d'une représentation dans la mémoire de l'interlocuteur A pourrait amorcer et donc pré-activer cette même représentation dans la mémoire de son ou sa partenaire B, et la rendrait ainsi plus disponible pour être réutilisée dans la suite du dialogue entre A et B. Cependant, alors que le modèle de Horton et Gerrig (2005a) indiquait la manière dont les informations fournies pendant le dialogue perduraient après celui-ci en mémoire épisodique, le modèle de l'alignement interactif de Pickering et Garrod (2004 ; 2021) ne décrit que l'implication des représentations linguistiques pendant le dialogue et ne précise pas l'état dans lequel se trouvent ces mêmes représentations après le dialogue. Pourtant, alors que le modèle de Pickering et Garrod (2004 ; 2021) postule que des changements d'activations ont lieu dans les représentations linguistiques pré-existantes pendant l'interaction, la littérature sur les représentations sémantiques nous informe que des changements d'activations qui ont lieu au sein des représentations sémantiques pourraient laisser des traces au niveau de l'organisation de ces représentations. Pour mieux comprendre cette idée, le chapitre 2 de cette thèse décrivait comment les représentations sémantiques sont supposées être organisées puis dans un second temps comment cette organisation peut être adaptée par de nouveaux inputs linguistiques. La présentation de ce chapitre nous permettait alors d'introduire, dans le chapitre suivant, notre question de recherche qui consiste à nous demander comment un dialogue et les activations qu'il provoque au sein des représentations sémantiques pourrait amener à une adaptation de l'organisation de ces représentations.

Les représentations sémantiques

L'un des premiers modèles ayant fortement influencé la manière dont l'organisation des représentations sémantiques est pensée est celui de Collins et Quillian (1969), revu par Collins et Loftus (1975) postule une organisation en réseau de représentations hautement interconnectées. Ce modèle suggère que chaque représentation sémantique est modélisée sous la forme d'un nœud, et que chaque nœud est directement connecté à toutes les autres représentations sémantiques qui lui sont sémantiquement proches. La distance entre deux nœuds au sein du réseau (i.e. le nombre de nœuds qui les séparent) pourrait alors définir le degré de relation sémantique entre deux représentations. Par exemple, les représentations de 'chat' et de 'chien' étant pensées comme directement reliées au sein du réseau, leur distance est sensée être réduite. Le modèle de Collins et Loftus (1975) suggère également que l'activation d'une représentation au sein du réseau est susceptible de se propager le long de ses connections pour pré-activer les représentations à proximité. Pour reprendre l'exemple précédent, l'activation de la représentation de 'chien' pourrait se propager à la représentation de 'chat', puis de manière secondaire à celle de 'lion' qui serait lié au 'chat' car ils constituent tous deux des félins. Cependant, plus deux représentations sont éloignées dans le réseau, plus la quantité d'activation propagée est réduite, ce qui diminue les chances de provoquer la pré-activation des représentations les plus éloignées de la source d'activation. Ce modèle fournit une explication claire pour expliquer les effets d'amorçage sémantique qui consistent en un traitement plus rapide d'un mot cible lorsque celui-ci succède à un premier mot amorce qui lui est sémantiquement relié. Dans une tâche classique de décision lexicale qui consiste à indiquer si le stimulus cible est un mot ou non dans une langue donnée, cela résulte en des temps de réaction plus courts pour les cibles reliées sémantiquement à l'amorce que les cibles non-reliées sémantiquement à l'amorce.

Alors que ce modèle de propagation d'activation semble adapté pour décrire une partie des résultats de la littérature, une autre théorie compatible avec le modèle de Collins et Loftus (1975) a été créée qui permette d'expliquer notamment les résultats obtenus dans des tâches de catégorisation sémantique. En effet, d'après la théorie des prototypes de Rosch (1975), les représentations sémantiques pourraient suivre non seulement une organisation en réseau, mais être également organisées sous formes de catégories sémantiques au sein même de ce réseau. Plus spécifiquement, les catégories sémantiques seraient constituées de représentations sémantiques qui s'organisent de manière hiérarchique. Cette hiérarchie a pu être mise en évidence par des temps de réaction qui varient entre les items d'une même catégorie sémantique pendant des tâches de catégorisation sémantique, avec une plus grande accessibilité pour les items 'typiques' d'une catégorie sémantique que pour les items 'atypiques' de cette même catégorie (Casey, 1992 ; Hampton, 1979; Hampton & Gardiner, 1983). La notion de typicalité d'une représentation correspond au fait que certains items d'une catégorie sémantique sont jugés comme plus représentatifs de leur catégorie (i.e. plus typiques) que d'autres, parce qu'ils partagent plus de propriétés qui sont représentatives de leur catégorie sémantique d'appartenance. Ainsi, le fait de partager plus de propriétés représentatives de leur catégorie semble favoriser les items typiques d'une catégorie sémantique en termes d'accessibilité en mémoire. Par exemple, le 'chat' sera considéré comme un mammifère plus typique que le 'pangolin' car ces derniers ne possèdent pas de poils (i.e. caractéristique typique des mammifères), et sera donc plus facilement accessible en mémoire et plus rapidement catégorisé.

Ces résultats, qui décrivent donc une organisation hiérarchique des représentations sémantiques, peuvent s'expliquer en termes de distance dans un réseau sémantique, avec une plus grande proximité des items typiques vis-à-vis de leur catégorie d'appartenance que des items atypiques. Cette proximité des items typiques pourrait être dû à un plus grand partage de propriétés avec le reste de leur catégorie. Ainsi, lorsque l'activation se propage de la catégorie à ses items, on peut supposer que les items typiques reçoivent plus d'activation que les items atypiques. Enfin, l'étude de Brunellière et Bonnotte (2018) a également déjà mis en évidence que la propagation d'activation au sein du réseau de représentations sémantiques était sensible à l'organisation hiérarchique des catégories sémantiques. Par l'utilisation d'un paradigme d'amorçage sémantique et d'une tâche de décision lexicale, les auteures ont montré que lorsque l'amorce et la cible appartiennent à une même catégorie sémantique, une amorce atypique n'est pas en mesure de pré-activer une cible typique. En revanche, une amorce typique pourra pré-activer une cible typique (réduction du temps de réaction). Ces résultats vont donc dans le sens d'une relation plus forte entre les items typiques entre eux qu'entre un item typique et un item atypique, et donc d'une propagation d'activation plus importante dans le premier cas que le deuxième.

Les modèles de Collins et Loftus (1975) et de Rosch (1975) présentés en première partie du chapitre 2 semblent donc apporter chacun des indices sur la manière dont les représentations sémantiques peuvent être organisées en mémoire. Dans la seconde partie du Chapitre 2, nous détaillons les preuves expérimentales de la flexibilité de l'organisation des représentations provenant de la littérature en compréhension de phrases écrites ou orales. Dans un premier temps, nous avons présenté des preuves que l'apprentissage de *nouvelles* significations pouvait amener à des changements au sein du réseau de représentations lexico-sémantiques déjà existantes (Rodd et al., 2012 ; 2013 ; 2016). Plus spécifiquement, les auteurs ont montré que ces nouvelles significations étaient suffisamment intégrées au réseau sémantique pour interagir avec les autres représentations déjà existantes et donc recevoir de l'activation de la part des représentations auxquelles elles sont nouvellement reliées (Rodd et al., 2012 ; 2013 ; 2016).

Dans un second temps, nous avons également présenté les preuves expérimentales indiquant que l'ajout de nouvelles significations dans le réseau de représentations sémantiques n'est pas nécessaire pour adapter l'organisation du réseau, puisque des modifications dans l'organisation des représentations sémantiques pré-existantes peuvent également avoir lieu. Il a notamment déjà été montré que les effets d'expertise ou l'exposition à un milieu culturel donné pourraient entraîner une réorganisation des catégories sémantiques. Par exemple, alors que deux individus possèdent tous deux une représentation d'un moineau et d'une autruche, si l'un d'eux est ornithologue, il ne hiérarchisera pas ces deux oiseaux de la même manière au sein de leur catégorie qu'un individu qui n'est pas expert (Bailenson et al., 2002 ; Lynch et al., 2000).

Pris dans son ensemble, la littérature en dehors du champ du dialogue tend à montrer que les représentations lexico-semantiques peuvent être adaptées suite à une exposition plus ou moins prolongée à de nouvelles expériences langagières. Dans la dernière partie de ce Chapitre 2, il nous a paru important de présenter l'un des mécanismes cruciaux qui favoriserait les modifications des représentations lexico-sémantiques, à savoir le mécanisme de consolidation en mémoire par le biais du sommeil. En effet, il a déjà été montré que des périodes de sommeil étaient en mesure de promouvoir la consolidation des nouvelles informations apprises en mémoire, de sorte à en favoriser le maintien au cours du temps (Gaskel et al., 2019). Ainsi, il semble que des adaptations au sein des représentations lexico-sémantiques pourraient être observées non seulement immédiatement après le nouvel input linguistique (Rodd et al., 2016) pour correspondre à un besoin immédiat de son environnement, mais également sur le plus long terme après une nuit de sommeil de sorte à consolider un besoin d'adaptation sur le plus long terme. On peut alors se demander si ces adaptations, quelle que soit leur temporalité (à court ou long terme), pourraient être observées après un dialogue. En effet, comme le suggéraient Pickering et Garrod dans leurs modèles (2004 ; 2021), des changements d'activation pourraient avoir lieu au niveau des représentations linguistiques, suite à la mise en place d'un mécanisme d'alignement entre partenaires de dialogue. Ainsi, des changements d'activation au niveau des représentations lexico-sémantiques pourraient induire une adaptation de leur organisation au sein du réseau de représentation. A partir de cette question de recherche, le Chapitre 3 présente les objectifs et hypothèses de la thèse, ces dernières étant résumées dans la section suivante.

Hypothèses

A partir de la littérature sur le dialogue et de celle sur les représentations lexico-sémantiques, nous faisons l'hypothèse que des activations se produisant au sein du réseau de représentations lexico-sémantiques au cours d'un dialogue pourraient directement impacter leur organisation de sorte que ces représentations demeurent changées après le dialogue. Ainsi, comme suggéré précédemment, l'hypothèse selon laquelle l'accès aux représentations lexico-sémantiques pourrait être modifié suite à un dialogue sera examinée sous deux angles :

(1) En lien avec la littérature montrant la possibilité de créer un nouveau contenu au sein du réseau de représentations lexico-sémantiques (Rodd et al., 2012 ; 2013 ; 2016), nous faisons l'hypothèse que de nouvelles relations pourraient être créées entre deux représentations déjà existantes mais non préalablement liées avant le dialogue. Ainsi, le fait d'accéder à l'une de ces deux représentations après le dialogue devrait pré-activer la seconde et la rendre plus accessible en mémoire, en accord avec le modèle de Collins et Loftus (1975).

(2) En lien avec la littérature montrant la flexibilité de la structure catégorielle des représentations lexico-sémantiques (Lynch et al., 2000), nous faisons l'hypothèse que la

connexion entre les items atypiques et leur catégorie sémantique d'appartenance puisse être renforcé durant un dialogue de sorte que ses items deviennent plus typiques après le dialogue qu'avant. Le fait d'accéder à la représentation d'un item devenu plus typique après le dialogue devrait donc permettre de pré-activer un autre item typique de sa catégorie d'appartenance, de part leur nouvelle proximité sémantique au sein du réseau (i.e. renforcement des propriétés sémantiques partagées ; Brunellière et Bonnotte, 2018).

Enfin, étant donné les suggestions de la littérature concernant la possibilité d'observer des changements des représentations lexico-sémantique immédiatement après l'exposition à une nouvelle information (Rodd et al., 2016), et le bénéfice observé d'une nuit de sommeil sur les changements des représentations lexico-sémantiques (Gaskell et al., 2019), les hypothèses 1 et 2 seront testées à ces deux moments (i.e. immédiatement après le dialogue et un jour après).

Afin de tester ces hypothèses, nous avons conduit deux séries d'expériences durant lesquelles des participants ont pris part à une tâche de dialogue. Dans cette tâche, des dyades de participants devaient interagir à propos d'images dont le type (abstraites ou images photo-réalistes) variait en fonction de l'objectif des études. Suite à cette tâche de dialogue, l'organisation des représentations lexico-sémantique était évaluée, soit immédiatement après le dialogue, soit un jour après.

Concernant la première série d'expériences, la tâche de dialogue avait pour but de créer une nouvelle relation entre deux représentations lexico-sémantique pré-existantes. Pour cela, deux mots non sémantiquement reliés (e.g. dinosaure et robinet) étaient associés à une même représentation visuelle abstraite, en l'occurrence une image de tangram (i.e. représentations visuelles abstraites constituées de figures géométriques). Afin de trouver des images de tangram qui pourraient être associées à ce type de paires de mots, nous avons commencé par créer une nouvelle base de donnée pour des images de tangram et leurs dénominations associées. Cette base de données fournit également des informations concernant le taux d'accord sur les dénominations pour chaque image, afin de s'assurer que les noms choisis étaient suffisamment plausibles pour que les participants se les représentent facilement en voyant l'image. Grâce à cette base de données, la première série d'expériences a pu être créée afin de tester notre première hypothèse sur la possibilité de créer une *nouvelle* relation lexico-sémantiques entre des représentations déjà existantes. Enfin, la seconde série d'expériences nous a permis de tester notre seconde hypothèse, c'est-à-dire la possibilité de changer l'organisation du réseau lexico-sémantique *sans* créer de nouvelles relations. Plus spécifiquement, ces expériences avaient pour but de déterminer s'il est possible ou non de renforcer des relations déjà existantes entre un mot atypique et un mot typique de la même catégorie sémantique, et donc de conclure sur un éventuel renforcement de la relation entre le mot atypique et sa catégorie sémantique d'appartenance. Les trois axes principaux de cette thèse (i.e. création de la base de données, création d'un nouveau lien et réorganisation des relations au sein d'une catégorie sémantique) sont détaillés ci-après.

Expérience 1 : Création d'une base de données d'images de tangram

Le chapitre 4 détaille la création d'une base de données d'images de tangram. Les images de tangram sont fréquemment utilisées en psychologie du dialogue (e.g., Brennan & Clark, 1996 ; Clark & Wilkes-Gibbs, 1986 ; Horton & Gerrig, 2002 ; Knutsen et al., 2018 ; Yoon & Brown-Schmidt, 2014 ; 2019) car elles peuvent être associées à des dénominations variées, ce qui incite les interlocuteurs à discuter de ce qu'elles représentent. Ces images sont donc généralement utilisées pour étudier les processus de négociation et de compréhension mutuelle en dialogue. Cependant, à l'heure actuelle, aucune norme n'existe concernant les images de tangram, qui pourrait permettre aux chercheurs d'adapter leur sélection de stimuli au type de dialogue qu'ils souhaitent créer. En particulier, la possibilité de connaître à l'avance les dénominations les plus plausibles pour chaque image ou de connaître le taux d'accord sur les dénominations à une image était un enjeu crucial de la thèse puisqu'il nous fallait être en mesure de trouver des images pouvant être associées à deux dénominations sémantiquement non reliées mais suffisamment plausibles pour le premier ensemble d'expériences (Chapitre 5).

L'objectif de la création de cette base de données était donc de combler le manque de normes pour les images de tangram dans la littérature. Pour cela, nous avons créé les premières normes pour un ensemble de 332 images de tangram et leurs dénominations possibles en français, et effectué des calculs sur le taux d'accord de ces dénominations. Au total, 169 volontaires de langue maternelle française ont pris part à l'étude et rempli un questionnaire en ligne. La base de données étant constituée de 332 images de tangram, il a été fait le choix de ne présenter que la moitié des images à chaque participant pour réduire la durée de l'expérience. Les volontaires étaient donc aléatoirement attribués à l'une des 2 versions du questionnaire, c'est-à-dire que 85 participants ont été exposés au questionnaire A constitué de 187 images, et 84 participants ont été exposés au questionnaire B constitué de 188 images. Au total, 375 images de tangram ont donc été vues et dénommées par des participants, bien que 43 images aient été supprimées suite au pré-traitement des données, pour un total de 332 images dans la base de données finale. La consigne donnée aux participants était de donner le premier mot ou la première expression qui leur venait à l'esprit pour nommer l'image de tangram qui était présentée à l'écran. Les images étaient présentées une à une à l'écran et les participants devaient proposer un nom pour chacune des 187 ou 188 images présentées. L'ordre de présentation des images était rendu aléatoire entre les participants.

Les données concernant les dénominations ont d'abord été examinées par deux évaluateurs de langue maternelle française afin de corriger d'éventuelles fautes d'orthographe. La manière dont les réponses étaient données ont également été standardisées, de sorte que des réponses ayant le même sens prennent la même forme pour pouvoir être comptées comme identiques et ne pas fausser la mesure de l'accord sur la dénomination. Sur la base de la littérature existante (Alario & Ferrand, 1999 ; Bonin et al., 2003 ; Dimitropoulou et al., 2009 ; Dunabeitia et al., 2018), trois mesures ont été calculées pour chaque image :

1. **La dénomination modale**, qui correspond au nom le plus souvent donné par les participants pour désigner une image.
2. **Le pourcentage d'accord sur le nom**, qui représente le pourcentage de participants ayant utilisé la dénomination modale.
3. **L'indice H**, qui reflète la diversité des réponses (Shannon & Weaver, 1949). L'indice H augmente à mesure que la diversité des réponses augmente, tandis que le pourcentage d'accord sur le nom diminue. Un indice H de 0 signifie que tous les participants ont donné le même nom à l'image. L'indice H est donc plus précis que le pourcentage d'accord pour mesurer la distribution globale des noms donnés, car il ne reflète pas uniquement l'accord sur la dénomination modale.

Ces mesures ont été calculées séparément pour les deux séries d'images (set A et set B) mais aussi pour les deux séries combinées.

Les résultats de ces analyses indiquent que le pourcentage moyen d'accord sur le nom pour les séries A et B combinées était de 24,01 % (Écart-type = 17,89), ce qui reflète une grande variabilité entre les participants dans les noms attribués à une image donnée. De même, l'indice H moyen pour ces séries combinées était de 4,59 (Écart-type = 1,14), ce qui confirme encore cette variabilité. Près de la moitié des images (151 tangrams) étaient associées à une réponse modale donnée par 10 % à 30 % des participants, tandis que seulement 11 % des images (37 tangrams) avaient une réponse modale donnée par plus de 50 % des participants.

Le résultat principal de cette première expérience reflète la grande variabilité des dénominations données à la majorité des images, comme le montre l'indice H élevé. Cependant, cette variabilité des dénominations semble relativement stable à travers les images, puisque près de la moitié des images montrent un pourcentage de réponse modale compris entre 10 % et 30 %. Ces résultats renforcent l'idée que les images de tangram sont des stimuli adaptés pour notre première série d'expériences et pour les études en dialogue plus généralement, de part la possibilité de trouver facilement des images associées à plusieurs dénominations. Les chercheurs en dialogue ont en effet souvent besoin de stimuli qui suscitent des négociations naturelles entre les participants pour étudier la structure et le contenu de ces échanges. Malgré le fait que près de la moitié des images soient associées à une forte variabilité de dénomination, cette base de données fournit également un grand nombre d’images de tangram avec des niveaux d'accord variables, allant jusqu’à un fort accord de dénomination. Ce résultat est important car il permet aux chercheurs de choisir leurs stimuli en fonction de la quantité de négociations voulue dans la tâche. Les chercheurs travaillant avec des participants francophones peuvent également utiliser les dénominations proposées pour chaque image dans la base de données, afin de contrôler les mots qui pourraient être utilisés pendant l'interaction.

En conclusion, cette étude a été conçue pour répondre à un besoin de stimuli standardisés pour notre première série d’expériences, mais aussi plus généralement pour les études dans le champ du dialogue. La prochaine section est donc dédiée au résumé du Chapitre 5, lequel décrit la première série d’expérience susmentionnée.

Expériences 2 à 4 : Création de nouvelles relations lexico-sémantiques en mémoire à long-terme par le dialogue

Comme cela a été détaillé plus tôt, il a déjà été montré que l'organisation des représentations lexico-sémantique peut être modifiée après l'exposition à des phrases orales isolées ou à des textes, pour intégrer de nouvelles significations au sein de du réseau de représentations (Rodd et al., 2012 ; Rodd et al., 2016). De plus, on sait que ces changements peuvent être observés à court et à long terme (Gaskell et al., 2019). Cependant, l’adaptation de l’organisation des représentations lexico-sémantiques part l’ajout de nouveau contenu au sein du réseau n’a jamais été testée après une situation de dialogue. L'objectif de cette série d'expériences était donc de combler cette lacune dans la littérature en testant la possibilité que le dialogue puisse impacter l’organisation des représentations lexico-sémantiques par la création de nouvelles relations entre des représentations déjà existantes, qui perdurerait après le dialogue. Pour se faire, nous avons associé deux mots sémantiquement non-reliés à une représentation visuelle commune durant un dialogue et évalué la potentielle nouvelle relation entre ces mots soit immédiatement après le dialogue (Expérience 2) soit après une nuit de sommeil favorisant la consolidation (Expériences 3 et 4). Afin de trouver une représentation visuelle commune pour des mots sémantiquement non-reliés avant le dialogue, nous avons utilisé des images abstraites qui peuvent être dénommées de différentes manières. Nous avons donc sélectionné des images de tangram parmi celles présentes dans la base de données du chapitre 4, de sortes qu’elles soient associées à deux dénominations sémantiquement non-reliées mais néanmoins plausibles. Les images étaient ensuite utilisées au cours d’une tâche de dialogue nouvellement créée, pour fournir de support de discussion. Au cours de ce dialogue, un compère de l’expérimentateur et un participant naïf interagissaient pour trouver un accord sur la meilleure façon de nommer les images de tangram affichées sur leur écran. La présence du compère nous a permis de contrôler les noms attribués à chaque image de tangram, de sorte que certaines images soient associées à deux mots sémantiquement non-reliés (e.g. 'robinet' et 'dinosaure'), afin de créer une nouvelle relation lexico-sémantique entre ces paires de mots. Par ailleurs, certaines images de tangram étaient associées à un seul mot dit par le compère (e.g. 'robinet'), afin de créer des paires de mots « non-reliées », c'est-à-dire des mots associés à deux images de tangram distinctes durant le dialogue. Cette condition « non-reliée » nous servait alors de point de comparaison avec notre condition « reliée ».

La création de la nouvelle relation entre nos paires de mots reliés pendant le dialogue était testée soit immédiatement après la tâche de dialogue (Expérience 2), soit un jour après (Expériences 3 et 4). Dans les Expériences 2 et 3, nous avons utilisé une tâche de décision lexicale couplée à un paradigme d'amorçage sémantique afin de tester l'existence des nouvelles relations entre les paires de mots de manière implicite, c'est-à-dire sans que le participant ne prête explicitement attention aux relations de sens entre les mots. Dans l'Expérience 4, la mesure effectuée avait pour but de tester l'existence de cette nouvelle relation de manière explicite en focalisant l'attention des participants sur les relations de sens entre les paires de mots par l'usage d'une tâche de jugement de proximité sémantique. Cette manipulation nous permet de déterminer la force de la nouvelle relation créée au sein du réseau de représentations lexico-sémantique, puisqu'il est connu que l'attention portée à un endroit précis du réseau (ici au niveau des relations de sens entre des paires de mots) tend à amplifier le degré d'activation du réseau (Dehaene et al., 2006). Dans tous les cas, nous nous attendions au même pattern de résultats dans les trois expériences, c'est-à-dire que nous nous attendions à ce que des paires de mots associées à une même image de tangram deviennent plus sémantiquement reliées que des paires de mots associées à deux images de tangram distinctes. En effet, la littérature précédemment présentée nous permet de faire l'hypothèse que les résultats soient les mêmes immédiatement après le dialogue et un jour après (Gaskell et al., 2019 ; Rodd et al., 2013 ; Rodd et al., 2016).

Plus précisément, cela devrait se manifester, dans la tâche de décision lexicale, par des temps de réaction plus courts pour les cibles appartenant à une paire de mots de la condition « reliée » que pour une cible appartenant à une paire de mot de la condition « non-reliée ». Dans la tâche de jugement de proximité sémantique, nous nous attendions à des scores plus élevés pour les paires de mots de la condition « reliée » que pour celles de la condition « non-reliée ».

Ainsi, les différents participants des 3 expériences (40 participants pour chacune des Expériences 2, 3 et 4) ont pris part à la tâche de dialogue. Au cours de cette tâche, les participants étaient amenés à interagir avec un compère de l'expérimentateur pour négocier le nom à donner à chaque image de tangram présentée à l'écran. Plus spécifiquement, 90 images étaient discutées au cours de 30 essais. Au cours d'un essai, trois images étaient présentées côte à côte sur un écran d'ordinateur et il était indiqué, soit au compère, soit au participant naïf, de proposer un ou deux noms pour chacune des trois images et de justifier le choix de ces noms. Ensuite, la personne ayant proposé les noms devait demander à l'autre personne si elle était d'accord ou non avec les noms proposés. Au total, 60 images étaient discutées par le compère (20 essais composés de 3 images chacun), parmi lesquelles 20 images réparties au travers des essais étaient associées à deux noms sémantiquement non-reliés pour former les paires de mots de la condition « reliée ». Les 40 images restantes étaient donc associées à un seul nom, et permettaient de former les 20 paires de mots de la condition « non-reliée ». Le compère avait pour instruction de lire le plus naturellement possible des scripts afin de fournir toujours les mêmes noms et les mêmes justifications à tous les participants. A contrario, le participant naïf pouvait librement choisir les noms des images parmi une liste de 6 mots proposés pour chaque essai, et les justifiait à sa guise. Ce participant avait donc 30 images (10 essais) à dénommer et les données associées à ces essais n'étaient pas traitées puisqu'ils servaient uniquement à garder le participant impliqué dans la tâche.

À la fin du dialogue, les participants des Expériences 2 et 3 étaient donc amenés à effectuer la tâche de décision lexicale couplée au paradigme d'amorçage sémantique. Au cours de cette tâche, les 20 paires de mots de la condition « reliée » et les 20 paires de mots de la condition « non-reliée » étaient présentées à l'écrit parmi des paires de mot-pseudomot. Pour chaque paire, le mot amorce était présenté pendant 150ms, puis un écran noir apparaissait pendant 16ms et enfin la deuxième suite de lettre (mot ou pseudomot cible) apparaissait jusqu'à ce que les participants aient indiqué s'il s'agissait d'un mot ou non en français. Ce choix d'un très court intervalle de temps entre l'amorce et la cible (SOA pour *stimulus-onset asynchrony* en anglais), équivalent à 166ms, était fait dans le but de pouvoir favoriser une interprétation de nos résultats en termes de propagation d'activation plutôt que d'une décision stratégique, laquelle nécessiterait plus de temps pour être mise en place. Suite à la présentation de la deuxième suite de lettres, les temps de réaction des participants étaient mesurés. L'objectif était de déterminer si le deuxième mot d'une paire « reliée » était plus rapidement reconnu que le deuxième mot d'une paire « non-reliée ». En effet, si une nouvelle relation lexico-sémantique était créé entre les paires de mots associées à une même image pendant le dialogue, alors la présentation de l'un des mots de la paire en amorce devrait pré-activer la représentation du mot cible et en faciliter la reconnaissance, par rapport à une paire de mots pour laquelle aucune nouvelle relation n'a été créée et pour laquelle la présentation de l'amorce ne devrait donc pas pré-activer la représentation du mot cible.

Les participants de l'Expérience 4 étaient amenés à effectuer une tâche de jugement de proximité sémantique. Au cours de cette tâche, les deux mots de chaque paire étaient présentés à l'écran en même temps et il était demandé aux participants d'indiquer sur une échelle de 1 à 7 à quel point les deux mots étaient proches en termes de sens. Il est à noter que des paires de mots naturellement sémantiquement reliées (e.g. café et thé) étaient ajoutées aux paires de mots présentées dans le dialogue afin de s'assurer que les participants utilisent aussi toute l'échelle de Likert (i.e. très proches au niveau du sens). Enfin, les résultats des 40 participants de cette Expérience 4 étaient comparés à ceux de 40 participants faisant partie d'un groupe contrôle. Contrairement aux participants du groupe expérimental, ces participants du groupe contrôle n'ont pas effectué la tâche de dialogue avant de réaliser la tâche de jugement de proximité sémantique. Ce groupe a donc été ajouté afin de déterminer si la différence de score de jugement entre les paires de la condition « reliée » et celles de la condition « non-reliée » était significativement plus importante pour le groupe expérimental que le groupe contrôle, et ainsi de pouvoir conclure sur un impact du dialogue sur la relation entre les paires de mots. Ainsi, au lieu de s'attendre à un effet principal du type de paires de mots (relié vs. non-relié), nous nous attendions dans l'Expérience 4 à une interaction significative entre le type de paires de mots et le groupe d'appartenance des participants (expérimental vs. contrôle).

Les résultats des Expériences 2 et 3 n'ont montré aucune différence significative dans les temps de réaction entre les paires de mots associées à une même image durant le dialogue et les paires de mots associées à deux images distinctes durant le dialogue (b = 2.67, SE = 6.37, χ^2 (1) = 0.18, *p* = .68 et *b* = -4.72, *SE* = 6.55, $\chi^2(1)$ = 0.52, *p* = .47 respectivement). En revanche, les résultats de l'Expérience 4 ont permis de mettre en évidence un effet d'interaction significatif dans les scores de jugement de proximité sémantique entre le type de paires de mots et le groupe d'appartenance des participants (b = 0.50, SE = 0.14, $\chi^2(1)$ = 12.63, p < .01). Plus précisément, des comparaisons par paires (méthode de Tukey) ont révélé que les scores de proximité sémantique pour les paires de mots associées à une même image étaient significativement plus élevés que pour les paires de mots associées à deux images distinctes pendant le dialogue. Cependant, cette différence n'est significative que pour le groupe de participant ayant pris part au dialogue (expérimental : *b* = -0.38, SE = 0.11, *t*(71.9) = -3.52, p = .0046; contrôle : b = 0.10, SE = 0.11, *t*(71.9) = 0.97, *p* = 1.00).

À la lumière des résultats de nos trois expériences, il semble que notre hypothèse soit partiellement validée. Nous avons effectivement pu mettre en évidence la création de nouvelles relations entre les paires de mots associées à la même image lors du dialogue pour les participants de l'expérience 4. Plus précisément, les participants ayant participé à la tâche de dialogue ont évalué les paires de mots associées à une même image pendant le dialogue comme étant plus sémantiquement liées que celles associées à deux images différentes. Fait crucial, cette nouvelle relation n'a pas été observée chez les participants qui n'ont pas pris part à la tâche de dialogue. Ces résultats soutiennent donc l'idée que de nouvelles relations lexico-sémantiques peuvent être créées entre des représentations lexico-sémantiques déjà existantes pendant le dialogue et être intégrées au réseau de représentations de sorte à être observées un jour après le dialogue. Cependant, les résultats des expériences 2 et 3 nous amènent à nuancer cette interprétation. En effet, lorsque l’existence de nouvelles relations était testée avec une tâche implicite, aucun impact du dialogue n’a été mis en évidence sur l’organisation des représentations lexico-sémantiques. Pris ensemble, nos résultats soutiennent l'idée que le dialogue a un impact sur l’organisation des représentations lexico-sémantiques, bien que les adaptations observées semblent trop faibles pour pouvoir être observées sans l’influence de processus attentionnels, lesquels sont connus pour augmenter le degré d’activation des représentations dans le réseau et donc les rendre plus accessibles (Dehaene et al., 2006). A contrario, lorsque l’attention du participant est focalisée sur ces nouvelles relations lexico-sémantique, comme dans l’Expérience 4, les participants parviennent à accéder à la partie dédiée du réseau au moment de répondre à la tâche.

En conclusion, cette série d’expériences nous aura non seulement permis de concevoir une nouvelle tâche de dialogue permettant de créer de nouvelles relations entre des représentations lexico-sémantiques déjà existantes, mais également de mettre en évidence que ce type d’adaptations au sein de l’organisation des représentations peut être observé un jour après le dialogue. Ces résultats ont donc des implications importantes pour la recherche sur le dialogue. En effet, ils impliquent que lors d'un dialogue, les informations partagées entre deux interlocuteurs ne sont pas uniquement stockées en mémoire pour être réutilisées ultérieurement dans un dialogue avec le même interlocuteur (Horton et Gerrig, 2005a). En effet, nos résultats tendent à montrer que ces informations partagées pourraient impacter directement les représentations lexico-sémantiques préexistantes des interlocuteurs. Or, ces représentations lexico-sémantiques étant des représentations abstraites, c'est-à-dire stockées en mémoire de manière indépendante à leur contexte d'exposition, elles pourraient être récupérées en mémoire en l'absence du contexte dans lequel elles ont été rencontrées. Cela signifie que l'impact des adaptations qui ont eu lieu pendant un dialogue ne se limite pas aux interactions avec le même interlocuteur, mais pourraient également affecter les dialogues futurs avec d'autres interlocuteurs. Cependant, au cours de ces expériences, nous avons uniquement testé l'hypothèse qu'un dialogue puisse engendrer l'ajout de nouvelles relations entre des représentations lexico-sémantiques déjà existantes. Dans la prochaine section, notre objectif sera de voir si un dialogue pourrait également amener des modifications de l'organisation de ces représentations sans ajout de nouveau contenu tel qu'une nouvelle relation entre des représentations lexico-sémantiques, notamment en adaptant la structure hiérarchique des catégories sémantiques en termes de typicalité.

Expériences 5 et 6 : Adaptation de l'organisation des catégories sémantiques par le dialogue – l'effet de typicalité

La première série d'expériences a montré que le dialogue pouvait modifier les représentations lexico-sémantiques en créant de nouvelles relations entre elles. Cependant, la mise en évidence des adaptations au sein du réseau de représentations lexico-sémantique n'a pu se faire que par le biais d'une mesure explicite. En effet, il est probable que les nouvelles relations créées dans le réseau aient été trop faibles pour être détectées avec une mesure implicite. Cependant, on peut se demander si des adaptations au niveau de relations déjà existantes entre des représentations lexico-sémantiques, par exemple en renforçant une relation déjà ancrée au sein du réseau de représentations, pourraient être mises en évidence avec une mesure implicite. Pour répondre à cette question, nous nous sommes concentrés sur l'idée que les représentations sémantiques sont organisées en catégories dotées d'une structure hiérarchique (Rosch, 1975 ; Rosch et Mervis, 1975), c'est-à-dire que certains items de leur catégorie sont considérés comme plus typiques que d'autres. Par exemple, la « poussette » est considérée comme un item atypique de la catégorie des véhicules, tandis que la « voiture » en sera un item typique, et ce statut dépend de la quantité de propriétés représentatives de leur catégorie que les items possèdent. Nous avons alors cherché à déterminer si des représentations lexico-sémantiques considérées comme des items atypiques de leur catégorie sémantique pouvaient devenir plus typiques de leur catégorie immédiatement après un dialogue ou un jour après celui-ci, en renforçant leur relation avec leur catégorie d'appartenance.

Pour répondre à cette question, deux expériences ont été menées au cours desquelles une nouvelle tâche de dialogue a été utilisée pour répondre à notre question de recherche. En effet, cette tâche permet de renforcer la relation sémantique entre des items atypiques et leur catégorie sémantique, ainsi qu'avec d'autres items typiques de cette catégorie. Pour cela, nous avons créé une tâche de dialogue au cours de laquelle les propriétés représentatives de leur catégorie sémantique sont répétitivement associées par les participants aux items atypiques de la catégorie. En renforçant la relation entre un item atypique et les propriétés qui définissent sa catégorie d'appartenance, nous faisons l'hypothèse que cela renforcera la relation déjà existante entre les items atypiques et leur catégorie d'appartenance, et par extension les items typiques de cette catégorie.

Pour évaluer l'existence d'une adaptation de l'organisation des représentations lexico-sémantiques à l'issue du dialogue, les participants étaient amenés à effectuer une tâche de décision lexicale couplée avec un paradigme d'amorçage sémantique, similaire à celle utilisée dans la première série d'expériences. Le choix de conserver cette tâche et ce paradigme, malgré les résultats non significatifs des expériences 2 et 3, repose sur les résultats de la littérature montrant que le statut d'un mot, typique ou atypique, peut être déterminé grâce à ce type de paradigme. Comme précédemment présenté, Brunellière et Bonnotte (2018) ont mis en évidence que lorsque la cible est un item typique, l'effet d'amorçage sémantique ne se manifeste que si l'amorce est typique mais pas si elle est atypique. Les participants effectuaient donc d'abord un pré-test individuellement, avant de revenir une semaine plus tard pour effectuer le dialogue ensemble, puis étaient évalués en post-test soit immédiatement après le dialogue (Expérience 5), soit un jour après (Expérience 6). L'existence d'un pré-test et d'un post-test est due au fait que ces expériences visent à évaluer des changements dans les relations lexico-sémantiques préexistantes. Ainsi, il est nécessaire d'obtenir une première mesure de l'accès aux représentations déjà existantes avant le dialogue pour interpréter correctement les résultats obtenus après le dialogue. Dans le pré-test, on s'attendait donc à ce que les paires de mots appartenant à une même catégorie sémantique produisent un effet d'amorçage sémantique uniquement lorsque l'amorce et la cible étaient typiques, mais pas lorsque l'amorce était atypique et la cible typique (Brunellière et Bonnotte, 2018). Après le dialogue, on s'attendait à ce que les amorces atypiques discutées deviennent plus typiques, produisant un effet de facilitation comparable à celui des amorces typiques. Des résultats similaires étaient attendus immédiatement après le dialogue (Expérience 5) et après une nuit de sommeil (Expérience 6).

Au total, les données de 78 participants ont été récoltées et analysées dans cette seconde série d'expériences (38 pour l'Expérience 5 et 40 pour l'Expérience 6). Concernant la tâche de dialogue, des dyades de participants naïfs étaient amenés à discuter d'images photo-réalistes représentant des items typiques et atypiques de 10 catégories sémantiques différentes. L'un des participants voyait une suite de quatre images référant à des items de quatre catégories sémantiques différentes sur son écran et devait faire deviner à l'autre participant l'ordre dans lequel ces images étaient placées. Pour cela, il ne pouvait pas nommer les images directement. A la place, une liste de quatre propriétés étaient affichées à l'écran, une pour chaque image et donc une par catégorie sémantique, et il ou elle devait trouver la propriété associée à chaque image, puis la donner à l'oral à son ou sa partenaire en indiquant à quelle position l'image correspondant à cette propriété se trouvait (première, deuxième, troisième ou quatrième). De son côté, l'autre personne devait choisir, parmi les images affichées sur son écran, laquelle correspondait à chacune des propriétés énoncées et indiquer sur une feuille de réponse l'ordre dans lequel son ou sa partenaire voyait ces images. Au total, il existait deux propriétés différentes par catégorie sémantique et chaque image était présentée deux fois avec chaque propriété de sa catégorie d'appartenance, donnant un total de quatre présentations de chaque image pour une discussion de 45 minutes. Ainsi, toutes les images d'une même catégorie sémantique (typique comme atypique) étaient associées aux deux mêmes propriétés. Cette répétition de la présentation des images avec leurs propriétés avait pour but de renforcer les changements dans les relations lexico-sémantiques.

Concernant les pré- et post-tests, ils consistaient donc en une tâche de décision lexicale couplée au paradigme d'amorçage sémantique. Dans cette tâche, les paires de mots utilisées étaient constituées des mots associés à une image pendant la tâche de dialogue en tant qu'amorce, et de cibles typiques soit de la même catégorie sémantique que l'amorce, soit d'une catégorie sémantique différente. Les cibles n'étaient pas discutées pendant le dialogue. De plus, des paires de mot-pseudomot étaient présentées au cours de certains essais pour les besoins de la tâche. De la même manière que pour les d'expériences 2, 3 et 4, les participants devaient indiquer, pour chaque paire de stimuli, si la deuxième suite de lettre de la paire présentée constituait un mot ou non en français. Les temps de réaction des participants étaient alors mesurés. L'objectif du pré-test était de confirmer, en lien avec les résultats de la littérature (Brunellière et Bonnotte, 2018), que l'effet de facilitation de reconnaissance de la cible (i.e. temps de réactions plus courts) retrouvé pour les paires de mots appartenant à une même catégorie sémantique n'était visible que lorsque l'amorce était un mot typique, et pas lorsque celle-ci était un mot atypique. Nous nous attentions donc à un effet principal de la relation sémantique, avec les paires de mots appartenant à une même catégorie sémantique produisant des temps de réaction plus courts que les paires appartenant à deux catégories différentes. En lien avec nos hypothèses, nous nous attendions aussi une interaction significative entre la relation sémantique et le statut de l'amorce (typique ou atypique), avec des temps de réaction plus courts pour les paires appartenant à la même catégorie sémantique mais uniquement lorsque l'amorce est typique et pas lorsqu'elle est atypique. A contrario, l'objectif du post-test était de déterminer si les amorces atypiques associées à des propriétés représentatives de leur catégorie sémantique d'appartenance au cours du dialogue étaient devenues plus typiques de leur catégorie, et étaient donc en mesure de produire un effet de facilitation similaire aux amorces typiques. Nous nous attendions donc toujours à un effet principal de la relation sémantique, mais cette fois à une absence d'interaction entre la relation sémantique et le statut de l'amorce, c'est-à-dire que les temps de réaction devraient être plus courts pour les paires de mots appartenant à une même catégorie sémantique, et ceci indépendamment du statut de l'amorce, que pour les paires appartenant à des catégories différentes.

Contrairement à ce qui était attendu, les résultats de l'Expérience 5 n'ont montré aucun effet de la relation sémantique, ni pour les amorces typiques, ni pour les amorces atypiques et aucune interaction significative n'a été trouvée avec le statut de l'amorce (typique ou atypique). Cela est d'autant plus marquant que cette absence d'interaction significative se retrouvait à la fois dans le pré-test (b = -1.30, SE = 14.93, χ^2 (1) = 0.01, *p* = .93) et le post-test ((b = -26.07, SE = 14.64, χ^2 (1) = 3.17, *p* = .08). En revanche, les résultats de l'Expérience 6 nous ont montré un pattern de résultats différent puisqu'un effet principal de l'amorçage sémantique significatif a été trouvé à la fois dans le pré-test (b = -23.38, SE = 10.37, χ^2 (1) = 5.08, *p* = .02) et le post-test (b = -20.08, SE = 10.00, χ^2 (1) = 4.03, *p* = .04). Cependant, aucune interaction significative n'a été mise en évidence ni en pré-test (b = 14.75, SE = 14.70, χ^2 (1) = 1.01, *p* = .32), ni en post-test (b = -1.90, SE = 14.14, χ^2 (1) = 0.02, *p* = .89), ce qui signifie que l'effet d'amorçage sémantique était le même pour les amorces typiques et atypiques, avant et après le dialogue.

A la lumière des résultats de nos deux expériences, il semble que nos hypothèses ne soient pas validées. Cependant, l'inconsistance des résultats de nos pré-tests entre l'Expérience 5 et la 6 ne nous permet pas de tirer de conclusion. En effet, les résultats des pré-tests suggèrent plutôt l'existence d'une grande variabilité individuelle dans la façon dont l'activation se propage dans le réseau lexico-sémantique lors de la tâche de décision lexicale, contrairement à ce qui est attendu dans la littérature. Cette variabilité est bien illustrée par la variabilité inter-participant en termes de différence de temps de réaction entre les paires de mots appartenant à une même catégorie sémantique et celles appartenant à différentes catégories sémantiques. Dans l'Expérience 5 par exemple, seulement 21 des 38 participants (55,26%) ont montré un effet de facilitation, c'est-à-dire des temps de réaction plus courts pour les paires partageant une relation catégorielle que pour les paires ne partageant pas de relation catégorielle. La différence de temps de réaction entre les paires reliées et non reliées variait de -147,18 ms à 107,71 ms (Moyenne = 0,26, SD = 46,82). Pour l'Expérience 6, une proportion légèrement plus élevée de participants (24 sur 40, soit 60%) a montré des effets de facilitation, avec une plage de différences de temps de réaction allant de -54,24 ms à 107,43 ms (Moyenne = 14,35, SD =

39,00). Cette variabilité a d’ailleurs déjà été discutée dans la littérature, et notamment dans l’article de Stolz et collaborateurs (2005). Ces auteurs suggèrent en effet que les activations ayant cours au sein du réseau de représentations sémantiques ne sont pas uniformes et que la présentation d’un même stimulus à deux reprises pourrait ne pas donner lieu à la même activité dans le réseau. Si une telle variabilité existe dans la manière dont l'activation se propage dans le réseau sémantique, cela pourrait expliquer pourquoi un effet de facilitation significatif est retrouvé dans certains échantillons de participants et pas d’autres. En effet, la moyenne des activations de l’échantillon dépendra forcément de la variabilité interindividuelle, en particulier dans des petits échantillons comme le nôtre (n = 38 et n = 40), comme cela a déjà pu être proposé par Heyman et al. (2018).

En conclusion, il semble que l’effet de facilitation sémantique souvent observé dans la littérature à l’aide d’une tâche de décision lexicale ne soit difficile à répliquer de manière systématique, surtout en situation d’échantillons de taille inférieur à 200 participants. Par conséquent, bien que l'effet de facilitation sémantique soit un outil pertinent pour étudier l’organisation de représentations lexico-sémantique bien établies dans le réseau sémantique, il pourrait ne pas être le plus adapté pour étudier les adaptations qui peuvent avoir lieu au sein de ces représentations et qui pourraient être moins bien ancrées aux représentations déjà existantes. Cette conclusion est en accord avec celle du chapitre précédent et est approfondie dans la discussion générale qui suit.

Discussion générale

L'objectif de cette thèse était d'examiner si l'organisation des représentations lexico-sémantiques préexistantes, pouvait être modifiée après un dialogue. Dans leur modèle d’alignement interactif, Pickering et Garrod (2004 ; 2021) suggéraient la possibilité que des changements d’activation pouvaient avoir lieu au niveau des représentations linguistiques des partenaires d’un dialogue de sorte que ces représentations s’alignent (c'est-à-dire que leurs niveaux d’activation deviennent plus similaires). Cependant, la question restait de savoir si ces activations pouvaient impacter les représentations linguistiques de sorte que leur organisation demeure changée après le dialogue. Pour y répondre, nous nous sommes concentrés sur les représentations lexico-sémantiques et leurs relations en terme de réseau de connections, car ces représentations sont connues pour s'adapter suite à des expositions linguistiques récentes (Rodd et al., 2012; 2013; 2016) et que de telles adaptations peuvent être consolidées en mémoire suite à une nuit de sommeil (Gaskell et al., 2019). Nous avons donc cherché à déterminer dans quelle mesure le dialogue pouvait impacter les représentations lexico-sémantiques de sorte que celles-ci demeurent changées, mais également dans quelle mesure ces changements pouvaient être observés un jour après le dialogue.

Les principaux résultats de cette thèse proviennent de l’Expérience 4 (Chapitre 5), et semblent indiquer que de nouvelles relations créées pendant un dialogue entre des représentations lexico-sémantiques préexistantes pouvaient être observées un jour après le dialogue. Cet effet a été trouvé à l'aide d'une tâche de jugement de proximité sémantique dans laquelle les participants étaient explicitement amenés à se concentrer sur les similitudes de sens entre des paires de mots précédemment associées à une même image au cours du dialogue. Cependant, aucun effet significatif n'a été mis en évidence lorsque l’existence de ces nouvelles relations était testée avec une tâche de décision lexicale et un paradigme d’amorçage sémantique (Chapitre 5). De la même manière, aucun effet significatif n’a été trouvé dans le Chapitre 6, dans lequel nous avons tenté de mettre en évidence des changements dans l’organisation des catégories sémantiques avec cette même tâche et ce même paradigme. Ces résultats suggèrent donc que le dialogue a en effet un impact sur les représentations lexico-sémantiques de sorte que leur organisation soit changée un jour après le dialogue, mais que ces changements (ici les nouvelles relations entre des représentations pré-existantes) étaient possiblement trop faibles pour être détectés par une tâche de décision lexicale. En revanche, en utilisant une tâche favorisant un focus attentionnel sur les propriétés sémantiques communes des représentations, il est possible d'amplifier l'activation des nouvelles relations au sein du réseau et donc de les rendre visibles.

Après ce résumé du manuscrit et de ses résultats, la discussion s'oriente vers les raisons pour lesquelles, malgré les résultats observés dans la littérature utilisant une tâche de décision lexicale et un paradigme d'amorçage sémantique, nos études et plus spécifiquement celles du chapitre 6, ont échoué à mettre en évidence les effets de facilitation attendus avec ce paradigme. Plus spécifiquement, nous y discutons les trois paramètres principaux qui ont été choisis avec ce paradigme et qui pourraient avoir eu des conséquences sur l'absence de résultat significatifs dans nos études : l'intervalle de temps entre l'amorce et la cible (SOA), la proportion de paires reliées (RP pour *relatedness proportion*) et la présence de relations associatives, c'est-à-dire les relations entre des paires de mots décrites dans les normes associatives et qui ne représentent donc pas 'purement' les relations sémantiques entre les mots. Cette section vise donc à expliciter les choix qui ont été faits concernant ces trois paramètres au sein de nos cinq expériences, et notamment en quoi le fait d'avoir testé des relations purement sémantiques de type catégorielles dans les Expériences 5 et 6 a pu être délétère dans la mise en évidence des adaptations. En effet, la mise en évidence de ce type de relation par l'utilisation d'un paradigme d'amorçage sémantique n'a pas toujours été un succès dans la littérature, comme l'a soulevé Hutchison (2003) dans sa revue de littérature.

Nous mentionnons dans la suite de cette discussion le fait que tester de nouvelles relations entre des représentations déjà existantes à l'aide d'un paradigme d'amorçage sémantique, comme cela a été fait dans le Chapitre 5, pouvait également représenter un challenge. En effet, en croisant des éléments de la littérature en amorçage épisodique et dans le champ des adaptations des représentations lexico-sémantiques, il semble que seules des relations déjà bien établies et suffisamment fortes entre les représentations lexico-sémantiques peuvent être observées par un paradigme d'amorçage sémantique (Dagenbach et al., 1990). En d'autres termes, il aurait été nécessaire d'exposer les participants à un apprentissage intensif basé sur plusieurs dialogues durant lesquels les mêmes images de tangram sont répétitivement présentées avec leurs noms associés afin de pouvoir observer un effet significatif de l'amorçage sémantique au niveau des nouvelles relations créées dans le réseau.

D'un autre côté, des preuves d'adaptation dans l'organisation des représentations lexico-sémantique ont déjà été trouvées en l'absence d'un apprentissage intensif de la nouvelle relation (Betts et al., 2018 ; Gaskell et al., 2019 ; Rodd et al., 2013 ; 2016). Cependant, ces études utilisaient des mesures comme les tâches d'association de mots pour observer les adaptations, où les participants devaient récupérer le sens amorcé d'un mot et fournir un mot associé. Ce type de tâche implique donc des processus très différents de ceux employés lors d'une tâche de décision lexicale couplée à un paradigme d'amorçage sémantique, notamment par le fait que l'attention des participants soit explicitement portée sur les relations entre les mots. Ainsi, à la lumière de la littérature, nous pouvons trouver au moins deux moyens de tester l'existence d'une adaptation dans l'organisation des représentations lexico-sémantiques. Premièrement, les participants peuvent passer par une phase d'apprentissage intensif axée sur les propriétés sémantiques des nouvelles informations pour s'assurer qu'elles sont suffisamment intégrées et accessibles via des tâches implicites comme la décision lexicale. Deuxièmement, si les nouvelles informations ne sont pas apprises de manière intensive, il est préférable d'utiliser une tâche qui concentre l'attention des participants sur les propriétés sémantiques des nouvelles informations lexico-sémantiques.

En ce qui concerne notre propre paradigme expérimental, nous suggérons donc que des changements ont bien eu lieu au sein du réseau de représentations lexico-sémantiques mais que ces changements – qu'ils soient de la création de nouvelles relations ou du renforcement de relations déjà existantes – étaient trop faibles pour produire une propagation d'activation de l'amorce à la cible et ne pouvaient donc être observés que par une tâche ancrant l'attention des participants sur les propriétés sémantiques des mots. Dans la suite de notre discussion, nous essayons de déterminer dans quelle mesure nos résultats peuvent être expliqués par des théories récentes concernant la manière dont les représentations lexico-sémantiques s'adaptent après chaque nouvelle expérience linguistique. La discussion commence par étudier les propositions du '*contextual-binding account*' (Mak et al., 2023). D'après cette théorie, les nouvelles expériences linguistiques sont d'abord stockées sous un format temporaire, mêlant représentations linguistiques et informations contextuelles, avant d'impacter les représentations linguistiques suite au processus de consolidation en mémoire. Le manuscrit ne présentant pas de résultat significatif de l'adaptation des représentations lexico-sémantiques immédiatement après le dialogue, la discussion ne tend pas à démontrer ou réfuter cette théorie mais plutôt de proposer des pistes sur des études qui pourraient être menées, en lien avec les méthodologies développées dans la thèse, afin de déterminer dans quelle mesure cette théorie pourrait expliquer les adaptations observées un jour après le dialogue. Par la suite, nous discutons également nos résultats en lien avec la littérature sur la structure catégorielle des représentations lexico-sémantiques. En particulier, l'article de Folstein et Dieciuc (2019) est particulièrement mis en avant car il reprend cette idée que les catégories sémantiques pourraient exister sous deux états. Le premier serait un état flexible qui peut faire penser au stockage contextuel temporaire de l'information mentionné par le *contextual-binding account*. Le second est un état stable qui, de notre point de vue, fait plutôt écho à la structure en réseau des représentations linguistiques telles qu’elles se présentent en mémoire à long-terme et qui ne serait altérée par nos expériences linguistiques qu’après consolidation.

Finalement, la discussion s’ouvre sur les implications de nos résultats pour la littérature en dialogue. En effet, le modèle de l’alignement interactif (Pickering et Garrod, 2004) suggérait que des changements dans le degré d’activation des représentations sémantique déjà existantes pouvait avoir lieu. Alors que nos résultats sont en accord avec cette idée, ils nous permettent d’aller encore plus loin dans la démonstration d’un impact du dialogue sur les représentations sémantiques, puisqu’ils suggèrent que le réseau de représentations soit changé un jour après le dialogue. Par ailleurs, le fait que des adaptations des représentations lexico-sémantiques soient trouvées un jour après le dialogue en l’absence du partenaire avec qui les informations ont été partagées remet en question l’effet de spécificité au partenaire et ses modèles associés présentés dans l’introduction du manuscrit. Pour rappel, la littérature en dialogue qui se base sur la notion de terrain commun proposait que les informations partagées en dialogue étaient spécifiques au partenaire (Brennan et Clark, 1996). Suite à cette proposition le modèle basé sur la mémoire a suggéré l’idée que les informations partagées en dialogue étaient récupérées en mémoire grâce à la présence du partenaire, qui sert d’indice de récupération des informations, afin de faciliter la coordination avec ce même partenaire pendant le dialogue. Cependant, à la lumière de nos résultats, il semble raisonnable de penser que les informations partagées au cours d’un dialogue pourraient également impacter les interactions avec des partenaires de dialogue différents, dans la mesure où les représentations linguistiques sollicitées ont été changées.

En résumé, cette thèse est le premier travail de recherche faisant le lien entre les littératures en dialogue et sur les représentations sémantiques. Elle présente également la création de méthodologies pertinentes pour adresser la question du lien entre le dialogue et l'accès aux représentations lexico-sémantiques, lesquelles pourront être réutilisées pour de futures études.

En ce sens, elle constitue une étape cruciale vers une meilleure compréhension des représentations en mémoire qui soutiennent le dialogue.

Appendices

Title: A modified procedure for naming 332 pictures and collecting norms: using tangram pictures in psycholinguistic studies.

Authors: Alicia Fasquel, Angèle Brunellière and Dominique Knutsen

Affiliation: Univ. Lille, CNRS, UMR 9193 - SCALab - Sciences Cognitives et Sciences Affectives, F-59000 Lille, France

Corresponding author:

Alicia Fasquel

SCALab, CNRS UMR 9193, Université de Lille, Domaine Universitaire du Pont de Bois, BP 60149, 59653 Villeneuve d'Ascq, France

fasquel.a.97@gmail.com

Abstract:

Tangram pictures are abstract pictures which may be used as stimuli in various fields of experimental psychology and are often used in the field of dialogue psychology. The present study provides the first norms for a set of 332 tangram pictures. These pictures were standardized on a set of variables classically used in the literature on cognitive processes, such as visual perception, language and memory: name agreement, image agreement, familiarity, visual complexity, image variability and age of acquisition. Furthermore, norms for concreteness were also provided owing to the influence of this variable on the processes involved in lexical production. Correlational analyses on all variables were performed on the data collected from French native speakers. This new set of standardized pictures constitutes a reliable database for researchers when they select tangram pictures. Given the abstract nature of tangram pictures, this paper also discusses the similarities and differences with the literature on line drawings, and highlights their value for dialogue psychology studies, for psycholinguistics studies and for cognitive psychology in general.

Keywords: Picture database, Tangrams, Name agreement, Image agreement, Familiarity, Visual complexity, Image variability, Age of acquisition, Concreteness

Introduction

Pictures are often used as stimuli in studies on visual perception, memory and language. In the field of dialogue psychology, which focuses on the mental representations and the psychological processes which enable (at least) two people to reach mutual comprehension as they interact, various methodologies have been developed. They range from the experimental study of genuine dialogues (e.g., Clark & Krych, 2004; Clark & Wilkes-Gibbs, 1986; Isaacs & Clark, 1987; Knutsen & Le Bigot, 2012; Kraut et al., 2003; Rossnagel, 2000) to the use of the visual world paradigm in interactive settings (e.g., Barr & Keysar, 2002; Brown-Schmidt, 2009a, 2009b; Metzing & Brennan, 2003). Interestingly, a number of these studies use similar pictures, that is, tangram pictures (e.g., Bangerter et al., 2020; Bard et al., 2014; Branigan et al., 2011; Brennan & Clark, 1996; Clark & Wilkes-Gibbs, 1986; Fox Tree & Clark, 2013; Horton & Gerrig, 2002; Hupet et al., 1991; Hupet & Chantraine, 1992; Knutsen, Col, et al., 2018; Knutsen et al., 2019; Knutsen, Ros, et al., 2018; Knutsen & Le Bigot, 2018; Lysander & Horton, 2012; Murfitt & McAllister, 2001; Ntsame-Mba & Caron, 1999; Rogers et al., 2013; Rogers & Fay, 2016; Russell & Schober, 1999; Schober & Clark, 1989; Swets et al., 2013; Wilkes-Gibbs & Clark, 1992; Yoon & Brown-Schmidt, 2014, 2019) (see Figure 1). Tangram pictures are abstract pictures which are usually made of seven smaller geometrical black and white figures (one square, two big triangles, two small triangles, one medium triangle and one parallelogram), although some studies have used more figures to create bigger tangram pictures (e.g., Bard et al., 2014).

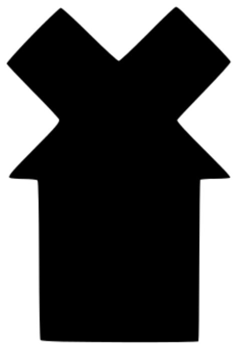


Figure 1. Example of a tangram picture made of seven figures

These pictures are used as stimuli in dialogue studies as they may be perceived in different ways, leading dialogue partners to engage in discussion and negotiation to reach an agreement as to how each picture should be referred to. The content of the conversations is then analyzed to infer the nature of the processes involved in dialogue. For instance, a decrease in the number of words and speech turns when a tangram picture is referred to repeatedly may reflect the emergence of "common ground" (i.e., knowledge that two people share and are aware of sharing) between both dialogue partners (Clark, 1996; Clark & Wilkes-Gibbs, 1986).

A potential issue with the use of tangram pictures as stimuli is that their characteristics are likely to affect the way in which people talk about them, and therefore the results of the studies they are used in. For instance, Hupet et al. (1991) found that picture codability (i.e. the ease with which each picture can be interpreted and verbally expressed) and discriminability (i.e. the ease with which a picture can be discriminated from others) made it more or less difficult for participants to reach an agreement as to how to name the pictures. Other characteristics may also play a role. For instance, some tangram pictures may be visually simpler than others. Likewise, some tangram pictures are perceived as more "consensual" than others, that is, many people may come up with the same label to refer to them, whereas other pictures may be associated with several different labels. However, the characteristics of tangram pictures are seldom controlled for in dialogue studies. Murfitt and McAllister (2001) controlled their pictures for codability and discriminability. Swets et al. (2013) selected their tangram pictures based on a norming pre-study in which they assessed the mean naming time for each picture. Knutsen, Ros, et al. (2018) attempted to control for label consensus by considering the frequency with which each label was used in their study in a post-hoc control. Nonetheless, it seems that in most studies, tangram pictures are chosen randomly, based only on the category they belong to (e.g., human-like pictures, animals or objects). This is problematic because, as mentioned above, most of the dialogue studies which involve tangram pictures base their conclusions on the analysis of the chacacteristics of the participants' speech, but these characteristics may also depend on the features of the tangram pictures under discussion. For instance, the decrease over time in the number of words needed to repeatedly describe a referent may reflect not only common ground construction, but also the ease with which the picture may be described, due for instance to its familiarity or its concreteness. In other words, without a strict control of the stimuli used, it is difficult to determine which conclusions can be drawn from the study of dialogues about tangram pictures.

In sum, although tangram pictures are often used in dialogue research, little is known regarding the characteristics of these pictures or the way in which these characteristics may affect the interaction between participants. The current study aimed to examine several variables associated with tangram pictures in order to enable researchers to control for them in their experiments. We specifically focused on variables related to the visual properties of the pictures, to the way they are represented conceptually, and to the ease with which the label used to refer to each picture may be retrieved. Indeed, picture-naming is known to involve various cognitive processes, such as converting the visual stimulus into a conceptual representation, label/lexical retrieval and lexical production (Dell et al., 1997). Variables that may affect each of these processing steps must thus be identified and controlled for.

Snodgrass and Vanderwart (1980) were the first to collect data on four distinct characteristics of a set of pictures of common objects (black-and-white line drawings): name agreement, image agreement, familiarity, and visual complexity. Name agreement reflects the extent to which participants use the same label to refer to a given picture. Image agreement is defined as the extent to which the picture's appearance is representative of the mental representation that participants associate with the picture's label. Familiarity refers to the extent to which participants are in contact with the picture in everyday life. Visual complexity is defined as the amount of visual detail in the picture (e.g., the number of lines in the drawing). Following Snodgrass and Vanderwart's (1980) study, many studies involving picture-naming tasks were conducted in a wide variety of languages (e.g., in French, Alario, & Ferrand, 1999, Bonin et al., 2003; in Spanish, Duñabeitia et al., 2018, Manoiloff et al., 2010; in Greek, Dimitropoulou et al., 2009; in Russian, Tsaparina et al., 2011; in Persian, Ghasisin et al., 2015) and in various populations including children, younger and older adults (e.g., Cannard et al., 2005, Yoon et al., 2004). Normative data on picture characteristics usually include additional variables, such as image variability and age of acquisition. Image variability reflects whether the label used to refer to a picture is related to a small or a large number of different images. Age of acquisition is the age at which participants believe that they learnt the label and is one of the main determinants of the speed of lexical retrieval and production (e.g., Cortese & Khanna, 2007).

Among the variables usually explored in normative studies, some are significantly correlated with each other. For instance, name agreement and image agreement are correlated negatively, suggesting that a stronger agreement between the picture's label and its appearance is related to a smaller number of labels provided, owing to the selection of similar labels across participants (Alario, & Ferrand, 1999; Bonin et al., 2003; Dimitropoulou et al., 2009; Ghasisin et al., 2015; Manoiloff et al., 2010; Snodgrass & Vanderwart,1980; Tsaparina et al., 2011). Moreover, age of acquisition is correlated positively with name agreement (Alario, & Ferrand, 1999; Bonin et al., 2003; Dimitropoulou et al., 2009; Duñabeitia et al., 2018; Ghasisin et al., 2015; Manoiloff et al., 2010; Tsaparina et al., 2011) and negatively with familiarity and image variability (Alario, & Ferrand, 1999; Bonin et al., 2003; Ghasisin et al., 2015; Manoiloff et al., 2010; Tsaparina et al., 2011). In other words, when labels are acquired early on, fewer labels are provided and the level of agreement across participants is high. In addition, when labels are acquired early on, more pictures are judged as familiar, and the labels provided are associated with more different images.

To constitute a normative database for tangram pictures, we focused on the same variables as in other studies involving picture-naming tasks: naming agreement, image agreement, familiarity, visual complexity, image variability and age of acquisition (see Table 1). We adapted the instructions used in name agreement tasks to obtain labels which were not necessarily isolated words, as participants in dialogue studies may use simple (e.g., "the guy") or more complex referential expressions to refer to tangram pictures (e.g., "the guy walking and wearing a hat"). Concreteness, which is known to influence the processes involved in lexical production (e.g., Hanley et al., 2013), was also measured in this study. Most studies on concreteness focus on word concreteness (e.g., a participant is shown a word and is asked to say how concrete the word is). In the current study, it reflected the extent to which a picture was judged as representing a concrete concept as opposed to a more abstract entity. The remainder of the collected data sought to examine the use of the labels associated with the pictures in interactive dialogue settings, the use of alternative labels to name the pictures, and the prior knowledge of pictures outside the study. Importantly, due to the nature of the stimuli used, we also adapted the instructions usually used in studies focusing on picture-naming tasks (Alario, & Ferrand, 1999; Bonin et al., 2003; Manoiloff et al., 2010; Snodgrass & Vanderwart,1980; Tsaparina et al., 2011) by asking questions on the visual and conceptual properties of the pictures *immediately* after the participant provided a label. Each participant thus focused on their choice of a label before answering the questions on the visual and conceptual properties of the picture. In picture-naming studies involving line drawings (Alario, & Ferrand, 1999; Bonin et al., 2003; Manoiloff et al., 2010; Snodgrass & Vanderwart,1980; Tsaparina et al., 2011), participants are usually asked to perform judgements about the visual and conceptual properties of the labels provided most frequently by another set of participants. Our approach made it possible to collect data from the same participants including the labels they had chosen as well as their judgments on the visual and conceptual properties of each picture.

Despite these methodological differences, similar cognitive processes are likely to be involved in picture-naming tasks using both line drawings and tangram pictures (e.g., converting the visual stimulus into a conceptual representation, label/lexical retrieval and lexical production). When using tangram pictures, we thus expect to find the same correlations between the variables usually taken into account in picture-naming studies involving line drawings. However, it is noteworthy that tangram pictures may be labeled in several different ways; thus, we expect a stronger diversity in the labels provided by participants to refer to a given tangram picture than those provided to refer to line drawings. Another difference with line drawings could be related to the amount of visual detail in tangram pictures being smaller than in line drawings.

Methods

Participants

One hundred and ninety-three native French speakers took part in the study. They received course credits or monetary compensation (20€) for their participation. All participants had normal or corrected-to-normal vision and no history of language disorders. After their participation, 20 participants were removed from the data, owing to either an overly long or an overly short time of participation (longer than three hours or less than one hour). This was done to ensure that all participants performed the task correctly and dutifully without taking too much time or going too fast on the questionnaires. Following the same principle, two participants were removed from the analysis owing to a rate of similar responses greater than 10%. Two other participants were removed from the database to balance the number of participants across groups (see Materials for details). Therefore, data from 169 participants were examined (133 females and 36 males, 18-29 years old, M = 20.17; *SD* = 1.47). Before the beginning of the experiment, they were informed about the goal and duration of the study. They also validated an online written consent form which followed the Declaration of Helsinki.

Materials

Tangram pictures

We started by collecting a total of 375 monochrome (black) tangram pictures from booklets found in various tangram games. As specified below, all 375 pictures were not necessarily included in the final database. As in most studies involving tangram pictures (e.g., Knutsen, Ros, et al., 2018), all the pictures were made of one square, two big triangles, two small triangles, one medium triangle and one parallelogram. The pictures were then scanned and randomly divided into two sets, hereafter named set A which contained 187 pictures, and set B which contained 188 pictures. All pictures were then uploaded to the online survey platform LimeSurvey (version 2.6). The largest side of the picture (length or width) was always 300 pixels long and the size of the picture was automatically adjusted to maintain the original proportions of each picture.

Questionnaires

The variables examined in this study were split into three sets of questionnaires (hereafter questionnaires 1, 2 and 3). Two different versions of each questionnaire were then created, each corresponding to a different set of pictures (A or B). Six questionnaires (1A, 1B, 2A, 2B, 3A and 3B) were thus created in total. We divided the pictures into two sets and the questions into three questionnaires to reduce the length of the experiment and to make sure that the collected data were reliable. In each questionnaire, the first question (which was always the same in all questionnaires) asked the participants to state the first word or expression which came to their mind when they saw the picture. That question was thus related to the name agreement variable[1]. Due to the expected diversity in the labels provided by the participants to refer to a given tangram picture, the first question was always related to name agreement and the following questions were divided into different categories (questions related to how the picture may be referred to in dialogue in questionnaires 1A and 1B, the interface between the labels chosen and their visual representation in questionnaires 2A and 2B with the image agreement and image variability questions, and visual and conceptual properties of the pictures in questionnaires 3A and 3B).

Following the name agreement question, in questionnaires 1A and 1B, participants were asked to say whether they would use the label they had provided to describe the picture during a dialogue with another person and, if not, which label they would prefer to use. The purpose of this question was to determine how likely the labels provided by the participants were to be used in a dialogue setting. Participants were then asked to state whether any other label (i.e., word or expression) came to their mind when they looked at the picture. These two questions were respectively referred to as “use in dialogue” and “other label”. Participants had to answer “yes” or “no” to each of these two questions. They were required to provide an additional label if they had

1 Participants were not asked to indicate their agreement with a label proposed by the experimenter, as the term "agreement" may suggest. As mentioned above, they were asked to come up with a label when they were shown the picture: the agreement between participants was then measured by calculating the number of different names given to a particular picture across participants.

answered “no” to the “use in dialogue” question or if they had answered “yes” to the “other label” question.

In questionnaires 2A and 2B, the name agreement question was followed by questions on the conceptual characteristics of the pictures that assessed image agreement, image variability and age of acquisition. The image agreement question asked participants to judge to what extent the picture’s appearance was representative of the mental representation associated with the label they had provided, using a 5-point Likert scale from “very weakly representative” to “very highly representative”. In the image variability question, the participants were instructed to rate whether the label they had provided was related to few or many different visual representations, using a 5-point Likert scale from “there are very few ways to visually represent this word or expression” to “there are many ways to visually represent this word or expression”. Finally, in the age of acquisition (AoA) question, the participants were asked to estimate the age at which they thought they had learnt the labels they had provided by selecting one age class among five: 0-3, 4-6, 7-9, 10-12, after 12.

In questionnaires 3A and 3B, the name agreement question was followed by questions about the visual properties of the pictures. The second question of these questionnaires sought to determine whether the picture had already been seen (referred to as the “already seen” question; e.g., the participants might have already seen the picture before taking part in the study). This question involved a “yes/no” answer. It was then followed by questions on familiarity, visual complexity, and concreteness, all rated on a 5-point Likert scale. The familiarity question consisted in indicating how familiar the participant was with the picture on a 5-point Likert scale from “unfamiliar” to “very familiar”. Regarding the visual complexity question, the participants had to rate the picture on a 5-point scale from “very simple” to “very complex”. Finally, the concreteness question required participants to rate the concept associated with the picture on a 5-point scale from “abstract” to “concrete”. We asked participants to name the picture before rating the related concept on the concreteness scale. All questions used in each questionnaire are listed in Appendix A (we provide the initial French wording as well as a translation in English; the questions are listed in the same order as in the initial questionnaires).

Procedure

To complete the online questionnaire, participants were asked to sit in a quiet room to avoid distractions such as music or noise and to answer the questions at their own pace. The first page of the questionnaire was the description of the experiment (goal and duration) and was followed by the consent form. Once participants had given their consent by answering “yes” to the question “Do you consent to take part in this study?”, they were shown the instructions of the task. Pictures were then displayed one by one on the participants’ screen. The pictures were alternately shown on a green or blue background, the alternation making it easier for the participants to understand that they had switched to a new picture. Each page of the questionnaire included one picture as well as all the questions the participant was required to answer. All questions were presented on the same page, below the picture. Each participant was shown only one of the six questionnaires (1A, 1B, 2A, 2B, 3A or 3B). Therefore, they saw only the questions corresponding to the questionnaire they had been allocated. Each question included one sentence (the question itself) followed by a space to answer, or a Likert scale, depending on the type of question. When all the questions corresponding to a given picture had been answered, participants clicked on the “next” button to move on to the following picture. The order in which the pictures were presented was randomized across participants. At the end of the questionnaire, participants provided demographic information regarding their first language and other spoken languages, gender, age, and history of language disorders. The entire questionnaire took approximately 90 minutes to complete.

Data preprocessing on name agreement question

The data from the name agreement question were first examined by two native French speakers in order to correct spelling mistakes. Determiners of isolated words were removed, except for words with different meanings depending on their grammatical gender (e.g., "le vase", which means "vase", had to be distinguished from "la vase", which means "mud"). In cases where the participant's response consisted of a letter ("M") or expressions such as "the letter M" ("la lettre M"), only the letter (in capital) was kept ("M"). Regarding numbers, when they were used to count things, as in the expression "two mountains" ("deux montagnes" in French), they were written in words. In all other cases (e.g., when the picture was believed to represent a number), the number form was kept. Plural words were replaced by singular forms. Finally, when a participant used a quotation mark to express their answer (e.g., "?", implying that the picture looked like a question mark), the quotation mark was written out in full (e.g., "question mark"). All other punctuation marks were removed. Words voluntarily written in English were not translated.

Secondly, labels indicating that the participants did not know the name of the label or did not want to respond were coded as "non-responses" ("abs" in the data files; e.g., "nothing", "no idea", "no opinion"). We also coded the following responses as NA: (a) when we did not understand the labels or when they presented lexical ambiguities, (b) when participants provided more than one label, (c) when labels referred to another picture which had been shown previously (e.g. "the same corridor as before but the door is closer")[2].

Sixteen pictures were removed from the analysis because they generated either more than 10% of NA responses, or more than 10% of non-responses. This represented 0.05% of the dataset. In the final dataset, 0.70% of responses were NA responses and 1.20% of responses were non-responses. In addition, 27 pictures were removed from the final dataset owing to an experimenter error. Therefore, the answers for the name agreement and all other variables were analyzed on the 332 remaining pictures (166 in set A and 166 in set B). Summary descriptive statistics for these 332

2 When participants provided more than one description, it was not possible to determine which of the two labels was the first to come to mind. Thus, we could not arbitrarily decide which one would count as the answer to the name agreement question and which one would count as the answer to the other label question. We thus decided to be conservative and to exclude these data from the analysis. Regarding cases where the label referred to a picture shown previously, our goal was to assess the first word or expression which came to the participants' mind when they saw the picture itself and not the comparison of labels between different pictures.

pictures are presented in Table 1. Appendix B provides the associated mean and standard deviation values for each given picture on the following variables: percentage of name agreement, H index, image agreement, familiarity, visual complexity, image variability, AoA and concreteness.

Table 1. Summary statistics for all variables

<Insert Table 1 here>

Data analyses

Analyses on name agreement data: modal label, percentage of name agreement and H index

In line with previous literature on name agreement data (e.g., Alario, & Ferrand, 1999; Bonin et al., 2003; Dimitropoulou et al., 2009; Duñabeitia et al., 2018; Ghasisin et al., 2015; Manoiloff et al., 2010; Tsaparina et al., 2011), three measures were calculated for each picture: the modal label, which was the label that most participants gave to refer to a given picture; the percentage of name agreement, which corresponded to the percentage of participants who gave the modal label as their answer, and the H index. The H index (Shannon & Weaver, 1949) reflects the diversity in the labels provided by participants to refer to a given picture. The H index was calculated for each picture using the following formula:

$$H = \sum_{i=1}^{k} p_i \log_2(1/p_i)$$

where *k* refers to the number of different labels given to each picture and *p* represents the proportion of subjects who gave each label (Snodgrass & Vanderwart, 1980). More precisely, if all participants use the same label to refer to a given picture, the picture has an H index of 0 and its percentage of name agreement is 100. In contrast, when the variability in labels provided across participants increases, the value of the H index also, increases, and the percentage of name agreement usually decreases as well. As already defined in the literature (Alario & Ferrand, 1999; Snodgrass & Vanderwart, 1980), the H index is calculated based on name agreement because the H index captures the distribution of labels for each picture across participants better than the name agreement measure. The modal label, the percentage of name agreement and the H index were computed on the two sets of pictures separately and on both sets combined (set A-B in Table 2). The results for the H index and the percentage of name agreement are presented in Table 2, while modal labels can be found online (see “ModalResponses-NamingAgreement” file).

Table 2

Summary descriptive statistics for name agreement (percentage of name agreement and H index) in sets A and B taken separately and for both sets combined.

<Insert Table 2 here>

Moreover, we explored whether the diversity in labels provided across participants for each given picture was concordant across the three sets of questionnaires. As pointed out by Snodgrass and Vanderwart (1980), the diversity in the participants’ labels to refer to a given picture is best reflected by the H index. We therefore calculated Kendall's coefficient of concordance on the H indexes obtained in each questionnaire by comparing the different questionnaires of a same set of pictures (i.e., 1A, 2A and 3A or 1B, 2B and 3B). In this analysis, the pictures rather than the participants were the basic analysis unit. Given the high number of pictures, the tables of critical values for the Kendall’s W statistic (Siegel & Castellan, 1988) were not appropriate to test W for statistical significance, and a chi-square test of significance was used instead.

We then examined the homogeneity of modal labels for each given picture across the three sets of questionnaires. In other words, for each given picture, we checked that the mode was the same regardless of the questionnaire in which this picture was presented. We calculated Krippendorff’s alpha rather than Fleiss’ kappa, as the comparison included three questionnaires for each set of pictures, A and B, and the dataset contained missing data (Zapf, Castell, Morawietz & Karch, 2016). Using the R-function kripp.alpha of the irr package (Gamer, Lemon, Fellows, & Singh, 2019), we compared the modal label obtained for each picture separately for both sets of pictures. This was performed in questionnaires 1A, 2A and 3A on the one hand, and in questionnaires 1B, 2B and 3B on the other, as sets A and B included different pictures.

Importantly, 67 pictures from set A and 63 pictures from set B had more than one modal label (multiple mode pictures, i.e. pictures for which two or more labels had been given the same number of times and were the most frequent labels). For example, picture A23 (set B) was named “bouteille” (i.e. bottle) by 12 participants or “maison” (i.e. house) by 12 other participants across all three questionnaires. In addition, 12 pictures from set A and 3 pictures from set B had no modal label. These were cases where each participant provided a different label to describe a picture. Each label was thus provided only once. Two different approaches were considered to solve these particularities in modal labels. The first approach consisted in including multiple mode pictures in the analysis, but only taking one modal label per picture into account. We selected the mode included in the analysis using the following procedure: for each picture, if one of the multiple modes in one of the questionnaires was the same as the (unique) mode in another questionnaire, it was this mode which was included in the analysis, as it could be considered as the most representative of the picture. All other possible modes were removed from the analysis. If none of the multiple modes matched the modes of the other questionnaires, then all multiple modes were replaced by an NA response and were not included in the analysis. The second approach consisted in removing, for each picture, the data from all three questionnaires if the picture was associated with multiple modes in at least one of the questionnaires. The modal responses for each approach are available online (see “ModalResponses-NamingAgreement” file).

Reliability for image agreement, AoA, image variability, familiarity, visual complexity and concreteness

The reliability of the ratings was assessed by calculating intra-class correlation coefficients (ICCs) on image agreement, AoA, image variability, familiarity, visual complexity and concreteness. For each variable, the ICC was obtained by using two-way random effects as the model, consistency as the definition and multiple raters/measurements as the type (see McGraw and Wong, 1996). The analysis sought to assess inter-participant reliability for each variable within each set of pictures.

Correlational analyses between H index, image agreement, AoA, image variability, familiarity, visual complexity and concreteness

Correlations between the H index, image agreement, AoA, image variability, familiarity, visual complexity, and concreteness were performed for the two sets of pictures separately and for both sets combined. The purpose of these correlations was to determine whether the correlational results reported in the literature on black-and-white drawings (e.g., Alario & Ferrand, 1999; Bonin et al., 2003) were also found with the tangram pictures used in the current study. The percentage of name agreement was not included in these analyses as the diversity in the participants' labels to refer to a given picture is best reflected by the H index. Image agreement, AoA, image variability, familiarity, visual complexity and concreteness scores were obtained by averaging the numerical responses to each picture across participants. The average scores for each picture are included in appendix B, and the corresponding raw data are available online (see "OtherVariables" file).

Analyses on three additional exploratory variables: "use in Dialogue", "other Label", and "already Seen" variables

Descriptive statistics were calculated for the "use in dialogue" and the "other label" variables in order to provide information on how the participants would have named the pictures in dialogue settings and on whether the participants would have provided any other label (i.e., word or expression) as an alternative to their first label. Analyses on these variables are presented as exploratory since this is the first norming paper to examine such variables, to our knowledge. Indeed, the pictures examined in other norming studies are not systematically used in dialogue research, thus making questions about how they would be referred to in dialogue less relevant. We therefore computed the percentage of "no" answers, that is when an additional label was provided as being more likely to be used in a dialogue setting than the first label (hereafter called the percentage of "use in dialogue") across all pictures. Regarding the "other label" question, we calculated the percentage of "yes" answers for which an alternative label was provided (hereafter called the percentage of "other label") across all pictures. This enabled us to determine how often the first label provided in response to the naming question (a) was not the label participants would have used in a dialogue setting or (b) was not the only label to come to the participants' mind to describe the picture.

Finally, the proportion of "yes" responses to the "already seen" question was also calculated in order to perform correlational analyses between these data and the other numerical variables measured in the questionnaires (i.e. questionnaires 3A and 3B, in which H index, familiarity, visual complexity and concreteness were examined). This analysis was performed to quantify the effect of prior knowledge on the other variables explored in this study.

Results

Analyses on name agreement data: modal label, percentage of name agreement and H index

Table 2 presents a summary of descriptive statistics related to the percentage of name agreement and the H index, including the 25th (*Q1*) percentile and the 75th (*Q3*) percentile. As shown in Table 2, the mean percentage of name agreement across participants from sets A and B combined was 24.01% (*SD* = 17.89), reflecting the fact that situations in which most participants allproduced the same label to refer to a given picture were quite rare. In line with the percentage of name agreement, the average H index for both sets A and B combined was 4.59 (*SD* = 1.14), which implies that the variability in naming was large across participants. For almost half of the pictures (151 tangram pictures), the modal response to each picture was given by 10% to 30% of the participants, and only 11% of the pictures (37 pictures) had a modal response given by participants more than 50% of the time. One possible explanation for this lack of agreement lies in theopportunity for participants to use full referential expressions (instead of isolated words only to refer to the pictures). To address this possibility, we used an additional level of coding on the modallabels associated with each picture. If a modal label was a word, we counted the total number of labels in which this word occurred alone or as part of an expression. However, if the modal label was an expression such as “a person who dances”, we counted the total number of labels in which the corresponding content words (“person” and “to dance”) appeared separately or together. If a picture was associated with more than one modal label, we used the most frequent label after having counted the number of times each modal label occurred as an isolated word or as a content word within a full referential expression. The goal was to determine whether a new coding grouping labels as isolated entities led to percentages of name agreement that were closer to previous studies on name agreement (e.g., Alario, & Ferrand, 1999; Bonin et al., 2003; Dimitropoulou et al., 2009; Duñabeitia et al., 2018; Ghasisin et al., 2015; Manoiloff et al., 2010; Tsaparina et al., 2011). Using this procedure, the maximal mean percentage of name agreement found in set A was 38.09% (*SD* = 24.28) and 31.95% (*SD* = 20.59) in set B. This was once again lower than the values reported in previous studies.

Kendall's coefficients of concordance were calculated between the H indexes of each set of questionnaires within each set of pictures. The analysis revealed a coefficient of concordance of 0.91 for the three questionnaires of the set A and the same coefficient of concordance (0.91) for the three questionnaires of the set B. The chi-square test revealed a significant concordance for both sets A and B (respectively $\chi^2 = 451$, $p < 0.001$ and $\chi^2 = 448$, $p < 0.001$). This result implies that for both sets A and B, the H index of each of the 166 pictures could be ranked in approximately the same order for each of the three questionnaires (Siegel & Castellan, 1988). In other words, the agreement across participants regarding picture naming was concordant across the three questionnaires. In line with this result, the distribution of the H index for each questionnaire appeared as quite similar across questionnaires (see Figure 2).

Figure 2. Density plots of the H index for the three questionnaires (1, 2 and 3).

< Insert Figure 2 here >

The analysis conducted using Krippendorff’s alpha revealed a reliability rate of 0.69 for set A, 0.66 for set B and 0.68 for both sets combined, when multiple mode pictures were included in the analysis of homogeneity of modal labels across the three sets of questionnaires. When multiple mode pictures were removed from the analysis, we observed a higher reliability rate of 0.78 for set A, 0.71 for set B and 0.74 for both sets combined. It is acceptable to draw tentative conclusions given that these reliability coefficients are between 0.67 and 0.80 (Krippendorff, 2004). Modal responses for a given picture were thus closely related across all three sets of questionnaires.

Reliability for image agreement, AoA, image variability, familiarity, visual complexity and concreteness

Intra-class correlational analyses (ICC) were performed on image agreement, AoA, image variability, familiarity, visual complexity and concreteness, in order to assess the inter-participant reliability for each of these variables. Table 3 reports the results for each set of pictures (A and B). All ICC values reflect moderate to excellent reliability.

Table 3. ICC index for image agreement, AoA, image variability, familiarity, visual complexity and concreteness.

< Insert Table 3 here >

Correlational analyses between H index, image agreement, AoA, image variability, familiarity, visual complexity and concreteness

Correlations between the variables examined in questionnaires 2 (2A and 2B) and 3 (3A and 3B) were calculated. These correlations were calculated using Kendall's tau, as all variables were not distributed normally, except for the AoA. Table 4 shows the correlation matrix for all variables. Since the data associated with each variable were not collected using the same sample of participants, the correlation matrix was divided into three parts. Matrix A shows correlations between the variables and the H index obtained using questionnaire 2. Matrix B shows correlations between the variables and the H index obtained using questionnaire 3. Finally, Matrix C shows correlations between the variables obtained using both questionnaires.

Table 4
Summary of all correlations performed within and across questionnaires 2 and 3

< Insert Table 4>

As shown in Table 4, almost all variables were significantly correlated with each other. High correlations were found between concreteness and familiarity (.73) and between concreteness and image agreement (.67). This suggests that pictures rated as representing a concrete concept had an appearance judged as highly representative of the participants' associated mental representations of the labels. Pictures rated as representing a concrete concept were also judged as more familiar.

As expected, name agreement, as measured by the H index, and image agreement were correlated negatively (-.43). Name agreement was also correlated negatively with familiarity (-.41) and concreteness (-.42), suggesting that pictures with a high level of agreement were judged as familiar and representing a concrete concept. Age of acquisition was also correlated positively with name agreement (.25) and negatively with familiarity (-.14), image variability (-.30), image agreement (-.18) and concreteness (-.15). This suggests that when labels are acquired early on, fewer labels are provided and the level of agreement is high across participants. In addition, when labels are acquired early on, more pictures are judged as familiar and representing a concrete concept and the labels produced evoked a larger number of different images. Labels acquired early on were also associated with pictures whose appearance was judged as highly representative of the participants' associated mental representations of the labels. Image agreement was correlated positively with familiarity (.62), suggesting that pictures whose appearance was judged as highly representative of the participants' associated mental representations of the labels were also rated as familiar.

There were also significant (but small or moderate) correlations between visual complexity and several other variables. Visual complexity was correlated positively with image variability (.21) and name agreement (.30), but negatively with image agreement (-.30), familiarity (-.48) and concreteness (-.37). This means that complex pictures were associated with labels which evoked a larger number of different images and led participants to provide more labels. However, complex pictures were also judged as less representative of the participants' associated mental representations of the labels and were rated as unfamiliar and more abstract.

Analyses on three additional exploratory variables: "use in dialogue", "other label" and "already seen"

Finally, we explored the data from the "use in dialogue", "other label", and "already seen" variables. The "use in dialogue variable" reflects whether a participant would use another word than the label provided to describe the picture to another person in an interactive dialogue setting. For example, for picture A23, a participant used the word "tour" (i.e. tower) to name the picture but indicated that they would use the word "robe" (i.e. dress) to describe this picture in a dialogue setting. As for the "other label" variable, participants had the opportunity to suggest another label from the one they had initially provided (i.e. their answer to the name agreement question). For example, for the A23 picture, one participant proposed the label "maison avec une cheminée" (i.e. house with a chimney) for the picture, but added another label "bouteille" (i.e. bottle) as their response to the "other label" question.

The percentage of "no" answers for the "use in dialogue" variable (i.e., when an additional label was provided as being more likely to be used in a dialogue setting than the first label) and "yes" answers "other label" (i.e., "yes" answers, when an alternative label was provided) variable was relatively low. The mean percentage of "no" answers for the "use in dialogue" variable was 17.58% (*SD* = 7.53) and 27.59 % for the "yes" answers of the "other label" variable (*SD* = 10.48). This means that most participants considered that their response to the naming question could be used in a dialogue setting. In addition, it was rare that participants gave another label after providing the first label that came to mind. Regarding the "already seen" variable, the mean percentage of "yes" responses was 12.04% (*SD*=12.36). This could be due to some of the participants having already played tangram games before. To develop the analysis of the data from questionnaire 3, we also computed correlations between the percentage of "yes" answers to the "already seen" variable and the other variables from this set. All these correlations were significant (Table 5). There was a negative correlation between the "already seen" variable and the H index (-.30), suggesting that pictures which had already been seen led to less diversity in the labels provided. As revealed by a negative correlation between the "already seen" variable and visual complexity (-.46) and a positive correlation between the "already seen" variable and familiarity (.42) and concreteness variables (.28), pictures which had already been seen were also judged as less complex, more familiar and as representing a more concrete concept.

Table 5
Correlations between the percentage of « yes » answers to the "already seen" question and the other variables measured in questionnaire 3.

<Insert Table 5 here>

Discussion

We present the first French normative database for 332 tangram pictures that can be used in dialogue research and other research fields. This database includes norming data on several characteristics, from visual properties to conceptual representations (name agreement, image agreement, familiarity, visual complexity, image variability, age of acquisition and concreteness).

The first important result of this study concerns the high variability in the labels given to each picture in all three sets, as shown by the high H index and the low percentage of name agreement. The reliability analysis revealed a sufficient level of agreement between the three sets of questions regarding how the pictures were named in all groups of participants. This variability cannot be attributed to the format of the participants' responses. Indeed, one may argue that while most norming studies (e.g., Alario, & Ferrand, 1999, Bonin et al., 2003; Dimitropoulou et al., 2009 ; Duñabeitia et al., 2018 ; Ghasisin et al., 2015 ; Manoiloff et al., 2010; Tsaparina et al., 2011) usually offer only isolated words as labels to refer to pictures, allowing participants to use either isolated words or expressions to respond increases the number of ways in which the same label may be expressed, thus explaining the variability in data. However, the additional analysis performed by grouping participants' responses by word content yielded results which do not support this hypothesis. Although the percentage of name agreement increased in this analysis, it remained low in comparison with the results on name agreement reported in the literature (e.g., Alario & Ferrand, 1999; Bonin et al., 2003; Duñabeitia et al., 2018). This reinforces and helps quantify the idea that tangram pictures are suitable for dialogue studies. Indeed, as mentioned in the introduction, dialogue research needs stimuli which can be perceived in different ways to encourage participants to negotiate how to refer to them. This enables researchers to study collaboration and the emergence of mutual knowledge and common ground (Clark, 1996; Clark & Wilkes-Gibbs, 1986). In this context, our database may be used to select stimuli based on name agreement – e.g., a researcher may be interested in selecting tangram pictures with a high level of name agreement (which could lead participants to reach an agreement promptly) vs. tangram pictures with a low level of name agreement (which could lead to more negotiation among participants, thus potentially causing the dialogue to last longer). The high level of variability in the labels provided by the participants also highlights that dividing tangram pictures into categories (e.g., people, cats, boats, etc. as it is often done in the dialogue literature) is not as straightforward as it may seem.

The second important result concerns the correlations obtained between the name agreement, calculated using the H index, and the other variables examined in the study. We found that name agreement was correlated with all other variables except for image variability, as expected from prior literature (Alario, & Ferrand, 1999; Bonin et al., 2003; Dimitropoulou et al., 2009; Ghasisin et al., 2015; Manoiloff et al., 2010; Snodgrass & Vanderwart,1980; Tsaparina et al., 2011). Name agreement and image agreement were negatively correlated, suggesting that a stronger agreement between a picture's label and its appearance is related to a lower number of labels provided. As in previous studies (Alario, & Ferrand, 1999; Bonin et al., 2003; Dimitropoulou et al., 2009; Ghasisin et al., 2015; Manoiloff et al., 2010; Tsaparina et al., 2011), we found that pictures associated with a smaller agreement rate were also those for which the label was acquired later.

These pictures were also judged as less familiar and more complex. These results also match the findings of Bonin et al. (2003). Therefore, while the level of name agreement may be lower for tangram pictures than for line drawing pictures, correlations between name agreement and other variables pertaining to the characteristics of the pictures seem to remain relatively stable. Correlations between the other variables than name agreement are discussed below.

Negative correlations were also found between AoA and familiarity on the one hand, and between AoA and image variability on the other, as in previous studies (Alario, & Ferrand, 1999; Bonin et al., 2003; Ghasisin et al., 2015; Manoiloff et al., 2010; Tsaparina et al., 2011). Regarding the correlation between AoA and familiarity, it suggests that later in life a word is learned, the less familiar people are with the shape of its visual representation, probably because they have been exposed to the word less. AoA was also negatively correlated with image agreement, as in the work by Alario, & Ferrand (1999), Ghasisin et al. (2015) and Tsaparina et al. (2011). Words acquired early thus tend to evoke more visual representations and are perceived as more familiar in our study, they are also associated with more representative pictures than words acquired late. The difference with previous results (Alario and Ferrand, 1999; Ghasisin et al., 2015; Manoiloff et al., 2010; Tsaparina et al., 2011) is the absence of correlation between AoA and visual complexity in our data, although it was not found either in the study by Bonin et al. (2003).

Image variability was positively correlated with visual complexity in our study, which means that labels evoking a large number of different images were associated with more complex pictures. This correlation was not significant in the work by Bonin et al. (2003) and it was negative in that of Alario and Ferrand (1999) and Manoiloff et al. (2010), since more complex pictures were associated with labels evoking a small number of images. This difference may lie in the nature of the pictures used in both studies, and in the subsequent interpretation of visual complexity. In line with drawing pictures, visual complexity may arise from a greater number of lines and thus be associated with a more detailed picture, leading to fewer possible representations. However, in tangram figures, increased visual complexity can be associated with a loss of clarity of the visual forms and a greater number of possible representations for the same picture. In accordance with this interpretation, it is interesting to note that the mean value of visual complexity in our study is similar to that found in previous studies (Alario and Ferrand, 1999; Bonin et al., 2003; Manoiloff et al., 2010) although line drawings usually include more visual details than tangram pictures.

Image agreement was positively correlated with familiarity as in research by Ghasisin et al. (2015), unlike in other studies (Alario and Ferrand, 1999; Bonin et al.,2003; Manoiloff et al. 2010; Tsaparina et al., 2011). This correlation means that the more a picture is considered as familiar, the more it is perceived as representative of its label. Image agreement was also negatively correlated in our study with visual complexity, and visual complexity was negatively correlated with familiarity. In other words, the more a tangram picture is perceived as complex, the less the picture is considered as representative of its label (Ghasisin et al., 2015; Tsaparina et al., 2011). Similar to previous studies using line drawing pictures (Alario, & Ferrand, 1999; Bonin et al., 2003; Ghasisin et al., 2015; Manoiloff et al., 2010; Tsaparina et al., 2011), pictures rated as more complex were judged as less familiar. As for name agreement, correlations between visual complexity and other variables pertaining to the characteristics of the pictures seem to remain relatively stable with respect to line drawing pictures, except for the correlation between visual complexity and image variability.

This is also the first study to explore concreteness for picture norms. This variable, which reflects the extent to which a picture is judged as representing a concrete concept, as opposed to a more abstract entity, is known to influence the processes involved in lexical production (e.g., Hanley et al., 2013). In this study on tangram pictures, concreteness was negatively correlated with name agreement and visual complexity, and was positively correlated with familiarity. These correlations mean that the more a concept was perceived as concrete, the more familiar and less complex the picture was perceived to be. Pictures which were judged as representing a more concrete concept were also associated with a lower name agreement rate. Our findings about concreteness are not surprising considering that (a) concreteness and imageability (i.e. the ease with which a mental image can be generated in response to the presentation of a written word) are known to be highly correlated (Paivio et al., 1968) and (b) the same correlations between the variables previously cited and imageability have already been found in the literature (Tsaparina et al., 2011). Therefore, if the perception of the concreteness of a concept is partly linked with the ease to generate a mental image of a given word, it seems reasonable to assume that these two variables will correlate in the same way with name agreement, visual complexity and familiarity. Moreover, it is interesting to note that tangram pictures were rated as representing abstract concepts (with a median value of 2.83). Hence, this study shows that tangram pictures are an interesting tool for presenting abstract concepts, even though presenting abstract concepts is not usually considered as feasible (Hanley et al., 2013).

Regarding the “already seen” variable, there were two negative correlations (with the name agreement and the visual complexity variables) and two positive correlations (with familiarity and concreteness variables). This means that the way participants perceived pictures was influenced by their previous experiences with these pictures. Consequently, the more a picture had already been seen by participants, the more it was perceived as familiar and representing a concrete concept, and the less it was perceived as complex. The already seen pictures also led to more consensus in the participants’ answers, as shown by the correlation with the name agreement.

In conclusion, the present database provides the first French norms for a new set of 332 tangram pictures. The analyses of this database show that the characteristics of tangram pictures are very similar to those of line drawings regarding many variables. Nevertheless, tangram pictures also have specific characteristics. Overall, the results confirm that tangram pictures are particularly well adapted for dialogue studies. As they can be perceived in several different ways, they lead participants to use a wide range of labels to refer to them, implying that they must discuss and negotiate in order to reach an agreement regarding how to refer to them. In addition, the characteristics assessed in this study may be particularly useful for dialogue researchers. Indeed, the data collected through the naming question may be used to anticipate how a sample of participants may refer to a given picture, and/or how likely it is for someone to use a given label to refer to a specific tangram picture. This could be particularly useful in studies in which the consensualness of tangram pictures must be taken into account (e.g., Knutsen, Ros, et al., 2018), or when attempting to write plausible scripts for studies involving confederates (see Kuhlen & Brennan, 2013). Moreover, the other variables examined in this study may help to control the materials used in dialogue studies better, as the visual and conceptual characteristics of tangram pictures are known to affect the way in which participants talk about them (e.g., Hupet et al., 1991). Specifically, the variables taken into account in the current study were not the same as those examined by Hupet et al. (who focused on codability or discriminability). Discriminability could not be assessed in the current study, as this variable involves comparing a tangram picture with other tangram pictures. However, codability, which is defined as the ease with which a tangram picture can be verbally expressed, may be related to a number of our variables (e.g., name agreement, as a picture whose name agreement is high may be perceived as easy to express verbally; pictures with a high level of image agreement may also be perceived as easy to express verbally). Thus, it would be interesting to examine whether these two variables affect dialogue in the same way as codability. More generally speaking, we hope that the variables measured in this study will be used by dialogue researchers to anticipate how picture characteristics might affect dialogue characteristics. This should enable researchers to control for a number of item-related phenomena (e.g., the fact that labels associated with more complex or less familiar pictures may take longer to negotiate) in order to obtain a better understanding of the processes which underlie dialogue.

Unlike the work by Duñabeitia et al. (2018) providing norms for 750 drawings in six European languages, our database concerns French labels, thus providing information on a less studied language than English but the part about the visual properties of pictures can still be used in any other languages and the concreteness ratings can also be generalized to other languages which have the same classification of concepts. Above all, the methodological approach adopted in this study offers interesting ways to develop new databases of abstract pictures in other languages. For example, the procedure we used to assess image agreement, image variability and age of acquisition (these were assessed following the choice of a label to refer to the picture) strongly differs from the procedure usually used with line drawing pictures (in which modal labels of pictures are judged in terms of image agreement, image variability and age of acquisition; Alario and Ferrand, 1999; Bonin et al., 2003; Manoiloff et al., 2010; Tsaparina et al., 2011). We chose to adapt the procedure due to the way in which tangram pictures are used in the dialogue literature, in which participants are usually required to come up with a word or expression to refer to the tangram pictures they are shown. In this kind of context, both the features of the tangram picture per se as well as the features of the label/concept chosen are likely to affect the way in which the tangram picture is referred to in dialogue. This is why we decided to ask the participant to focus on the concept they had chosen (even if this did not finally correspond to the modal label for the picture) rather than on modal labels. A limitation of this study is the fact that data were collected in a situation other than a dialogue setting. This led us to create the "use in dialogue" and "other label" variables. In future studies, it would thus be interesting to verify that the modal responses found in our name agreement question are used in genuine dialogue situations, or whether participants favor the use of the "other labels" listed in this study (in the latter case, it would also be interesting to determine which factors lead participants to favor the use of "other labels" over modal labels).

Outside the field of dialogue research, researchers focusing on the cognitive processes involved in the processing of pictures and concepts, on the interface of picture recognition and language processing, or on language production in general may want to use our database to control in their studies the variables we have measured in our study. Moreover, the fact that tangram pictures may be associated with several different labels may enable researchers to answer questions such as how lexical production is affected by the competition between multiple concepts and labels evoked by pictures, as well as the nature of the factors that influence the selection of concepts and labels, or even the cognitive mechanisms associated with the activation of abstract concepts after the presentation of pictures.

Regarding the materials used in this study, they are open-access and free from copyright restrictions for non-commercial purposes, to facilitate their use for further research and the exploration of other aspects of the pictures. Researchers can find the materials online at the following url https://osf.io/bxkpa/?view_only=2fe0acd124e64df19a0195354ebe45b4, with all the featured pictures and corresponding norms. The modal responses using the various methods presented in this article are also available.

Acknowledgements

This research was supported by the *Maison Européenne des Sciences de l'Homme et de la Société (*Emergents project-PréDi) and the French National Research Agency (ANR-19-CE28-0006). The authors are grateful to Laurent Ott, Julie Lecerf, Loïc Noël and Charlotte Stinkeste for their help during data collection and preprocessing. They are also grateful to Melvin Yap and two anonymous reviewers for their help during the peer review process. The manuscript was proofread by a native English-speaking copyeditor.

References

Alario, FX., Ferrand, L (1999). A set of 400 pictures standardized for French: Norms for name agreement, image agreement, familiarity, visual complexity, image variability, and age of acquisition. *Behavior Research Methods, Instruments, & Computers* 31, 531–552. https://doi.org/10.3758/BF03200732

Bangerter, A., Mayor, E., & Knutsen, D. (2020). Lexical entrainment without conceptual pacts? Revisiting the matching task. *Journal of Memory and Language*, *114*, 104129. https://doi.org/10.1016/j.jml.2020.104129

Bard, E. G., Hill, R. L., Foster, M. E., & Arai, M. (2014). Tuning accessibility of referring expressions in situated dialogue. *Language, Cognition and Neuroscience*, *29*(8), 928–949. https://doi.org/10.1080/23273798.2014.895845

Barr, D. J., & Keysar, B. (2002). Anchoring comprehension in linguistic precedents. *Journal of Memory and Language*, *46*(2), 391–418. https://doi.org/10.1006/jmla.2001.2815

Bonin, P., Peereman, R., Malardier, N., Méot, A., & Chalard, M. (2003). A new set of 299 pictures for psycholinguistic studies: French norms for name agreement, image agreement, conceptual familiarity, visual complexity, image variability, age of acquisition, and naming latencies. *Behavior Research Methods, Instruments, & Computers* 35, 158–167. https://doi.org/10.3758/BF03195507

Branigan, H. P., Catchpole, C. M., & Pickering, M. J. (2011). What makes dialogues easy to understand? *Language and Cognitive Processes*, *26*(10), 1667–1686. https://doi.org/10.1080/01690965.

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

Brown-Schmidt, S. (2009). Partner-specific interpretation of maintained referential precedents during interactive dialog. *Journal of Memory and Language*, *61*(2), 171–190. https://doi.org/10.1016/j.jml.2009.04.003

Brown-Schmidt, S. (2009). The role of executive function in perspective taking during online language comprehension. *Psychonomic Bulletin and Review*, *16*(5), 893–900. https://doi.org/10.3758/PBR.16.5.893

Cannard, C., Blaye, A., Scheuner, N., & Bonthoux, F. (2005). Picture naming in 3- To 8-year-old French children: Methodological considerations for name agreement. *Behavior Research Methods*, *37*(3), 417–425. https://doi.org/10.3758/BF03192710

Clark, H. H. (1996). Using language. *New York: Cambridge University Press*.

Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, *50*(1), 62–81. https://doi.org/10.1016/j.jml.2003.08.004

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

Cortese, M. J., & Khanna, M. M. (2007). Age of acquisition predicts naming and lexical-decision performance above and beyond 22 other predictor variables: An analysis of 2,342 words. *Quarterly Journal of Experimental Psychology*, *60*(8), 1072–1082.

https://doi.org/10.1080/17470210701315467

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical Access in Aphasic and Nonaphasic Speakers. *Psychological Review*, *104*(4), 801–838. https://doi.org/10.1037/0033-295X.104.4.801

Dimitropoulou, M., Duñabeitia, J. A., Blitsas, P., & Carreiras, M. (2009). A standardized set of 260 pictures for Modern Greek: Norms for name agreement, age of acquisition, and visual complexity. *Behavior Research Methods*, *41*(2), 584–589. https://doi.org/10.3758/BRM.41.2.584

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 (2006)*, *71*(4), 808–816. https://doi.org/10.1080/17470218.2017.1310261

Fox Tree, J. E., & Clark, N. B. (2013). Communicative Effectiveness of Written Versus Spoken Feedback. *Discourse Processes*, *50*(5), 339–359. https://doi.org/10.1080/0163853X.2013.797241

Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2019). irr: Various Coefficients of Interrater Reliability and Agreement. *R Package Version 0.84*.

Ghasisin, L., Yadegari, F., Rahgozar, M., Nazari, A., & Rastegarianzade, N. (2014). A new set of 272 pictures for psycholinguistic studies: Persian norms for name agreement, image agreement, conceptual familiarity, visual complexity, and age of acquisition. *Behavior Research Methods*, *47*(4), 1148–1158. https://doi.org/10.3758/s13428-014-0537-0

Hanley, J. R., Hunt, R. P., Steed, D. A., & Jackman, S. (2013). Concreteness and word production. *Memory and Cognition*, *41*(3), 365–377. https://doi.org/10.3758/s13421-012-0266-5

Horton, W. S., & Gerrig, R. J. (2002). Speakers' experiences and audience design: Knowing when and knowing how to adjust utterances to addressees. *Journal of Memory and Language*, *47*(4), 589–606. https://doi.org/10.1016/S0749-596X(02)00019-0

Hupet, M., Seron, X., & Chantraine, Y. (1991). The effects of the codability and discriminability of the referents on the collaborative referring procedure. *British Journal of Psychology*, *82*(4), 449–462. https://doi.org/10.1111/j.2044-8295.1991.tb02412.x

Hupet, M., Chantraine, Y. (1992). Changes in repeated references: Collaboration or repetition effects?. *Journal of Psycholinguistic Research* 21, 485–496. https://doi.org/10.1007/BF01067526

Isaacs, E. A., & Clark, H. H. (1987). References in Conversation Between Experts and Novices. *Journal of Experimental Psychology: General*, *116*(1), 26–37. https://doi.org/10.1037/0096-3445.116.1.26

Knutsen, D., Bangerter, A., & Mayor, E. (2019). Procedural coordination in the matching task. *Collabra: Psychology*, *5*(1). https://doi.org/10.1525/collabra.188

Knutsen, D., Col, G., & Le Bigot, L. (2018). An investigation of the determinants of dialogue navigation in joint activities. *Applied Psycholinguistics*, *39*(6), 1345–1371. https://doi.org/10.1017/S0142716418000358

Knutsen, D., & Le Bigot, L. (2018). The influence of conceptual (mis)match on collaborative referring in dialogue. *Psychological Research*, *84*, 514–527. https://doi.org/https://doi.org/10.1007/s00426-018-1060-1

Knutsen, D., & Le Bigot, L. (2012). Managing dialogue: How information availability affects collaborative reference production. *Journal of Memory and Language*, *67*(3), 326–341. https://doi.org/10.1016/j.jml.2012.06.001

Knutsen, D., Ros, C., & Le Bigot, L. (2018). Spoilt for choice: initially considering several referential expressions affects subsequent referential decisions. *Language, Cognition and Neuroscience*, *33*(5), 618–632. https://doi.org/10.1080/23273798.2017.1400080

Kraut, R. E., Fussell, S. R., & Siegel, J. (2003). Visual information as a conversational resource in collaborative physical tasks. *Human-Computer Interaction*, *18*(1–2), 13–49. https://doi.org/10.1207/S15327051HCI1812_2

Krippendorff, K. (2004). Content analysis: An introduction to its methodology (2nd ed.). *Thousand Oaks, California: Sage.*

Kuhlen, A. K., & Brennan, S. E. (2013). Language in dialogue: When confederates might be hazardous to your data. *Psychonomic Bulletin and Review*, *20*(1), 54–72. https://doi.org/10.3758/s13423-012-0341-8

Lysander, K., & Horton, W. S. (2012). Conversational Grounding in Younger and Older Adults: The Effect of Partner Visibility and Referent Abstractness in Task-Oriented Dialogue. *Discourse Processes*, *49*(1), 29–60. https://doi.org/10.1080/0163853X.2011.625547

Manoiloff, L., Artstein, M., Canavoso, M. B., Fernández, L., & Segui, J. (2010). Expanded norms for 400 experimental pictures in an Argentinean Spanish-speaking population. *Behavior Research Methods*, *42*(2), 452–460. https://doi.org/10.3758/BRM.42.2.452

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(1), 30–46. https://doi.org/10.1037/1082-989X.1.1.30

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. https://doi.org/10.1016/S0749-596X(03)00028-7

Murfitt, T., & McAllister, J. (2001). The effect of production variables in monolog and dialog on comprehension by novel listeners. *Language and Speech*, *44*(3), 325–350. https://doi.org/10.1177/00238309010440030201

Ntsame-Mba, F., & Caron, J. (1999). Construction et gestion du terrain commun chez des enfants de sept à quatorze ans et des adultes. *Archives de Psychologie (Genève)*, *67*(260).

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, Imagery, and Meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, *76*(1 PART 2), 1–25. https://doi.org/10.1037/h0025327

Rogers, S. L., & Fay, N. (2016). Stick or switch: A selection heuristic predicts when people take the perspective of others or communicate egocentrically. *PLoS ONE*, *11*(7), e0159570. https://doi.org/10.1371/journal.pone.0159570

Rogers, S. L., Fay, N., & Maybery, M. (2013). Audience Design through Social Interaction during

Group Discussion. *PLoS ONE*, *8*(2), 57211. https://doi.org/10.1371/journal.pone.0057211

Roxβnagel, C. (2000). Cognitive load and perspective-taking: Applying the automatic- controlled distinction to verbal communication. *European Journal of Social Psychology,30*, 429–445. https://doi.org/10.1002/(SICI)1099-0992(200005/06)30:3

Russell, A. W., & Schober, M. F. (1999). How beliefs about a partner's goals affect referring in goal-discrepant conversations. *Discourse Processes*, *27*(1), 1–33. https://doi.org/10.1080/01638539909545048

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, *21*(2), 211–232. https://doi.org/10.1016/0010-0285(89)90008-X

Shannon, C. E., & Weaver, W. (1949). The Mathematical Theory of Communication. *Urbana, IL: University of Illinois Press*.

Siegel, S., & Castellan, N. J. J. (1988). Nonparametric statistics for the behavioral sciences (2nd ed.). *Mcgraw-Hill Book Company*.

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. https://doi.org/10.1037/0278-7393.6.2.174

Swets, B., Jacovina, M. E., & Gerrig, R. J. (2013). Effects of Conversational Pressures on Speech Planning. *Discourse Processes*, *50*(1), 23–51. https://doi.org/10.1080/0163853X.2012.727719

Tsaparina, D., Bonin, P., & Méot, A. (2011). Russian norms for name agreement, image agreement for the colorized version of the Snodgrass and Vanderwart pictures and age of acquisition, conceptual familiarity, and imageability scores for modal object names. *Behavior Research Methods*, *43*(4), 1085–1099. https://doi.org/10.3758/s13428-011-0121-9

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

Yoon, C., Feinberg, F., Luo, T., Hedden, T., Gutchess, A. H., Chen, H. Y. M., Mikels, J. A., Jiao, S., & Park, D. C. (2004). A cross-culturally standardized set of pictures for younger and older adults: American and Chinese norms for name agreement, concept agreement, and familiarity. *Behavior Research Methods, Instruments, and Computers*, *36*(4), 639–649. https://doi.org/10.3758/BF03206545

Yoon, S. O., & Brown-Schmidt, S. (2019). Contextual Integration in Multiparty Audience Design. *Cognitive Science*, *43*(12), 12807. https://doi.org/10.1111/cogs.12807

Yoon, S. O., & Brown-Schmidt, S. (2014). Adjusting conceptual pacts in three-party conversation. *Journal of Experimental Psychology: Learning Memory and Cognition*, *40*(4), 919–937. https://doi.org/10.1037/a0036161

Zapf, A., Castell, S., Morawietz, L., & Karch, A. (2016). Measuring inter-rater reliability for nominal data - Which coefficients and confidence intervals are appropriate? *BMC Medical Research Methodology*, *16*(1), 1–10. https://doi.org/10.1186/s12874-016-0200-9

Table 1 Summary statistics for all variables

	H1	H2	H3	Image agreement	Familiarity	Visual complexity	Image variability	AoA	Concreteness
Mean	3.91	3.60	3.65	3.25	2.94	2.74	3.14	2.38	2.90
SD	0.82	0.93	0.96	0.55	0.56	0.59	0.35	0.36	0.61
Median	4.19	3.84	3.94	3.18	2.89	2.73	3.18	2.39	2.83
Min	0.64	0.22	0.00	2.10	1.71	1.11	1.96	1.39	1.48
Max	4.86	4.86	4.81	4.79	4.68	4.14	3.97	3.71	4.46
Range	4.21	4.64	4.81	2.69	2.96	3.04	2.00	2.32	2.98

H1, name agreement for questionnaire 1 with the H index; H2, name agreement for questionnaire 2 with the H index; H3, name agreement for questionnaire 3 with the H index; AoA, age of acquisition

Table 2 Summary descriptive statistics for name agreement (percentage of name agreement and H index) in sets A and B taken separately and for both sets combined.

	Set A		Set B		Set A-B	
	% Name agreement	H index	% Name agreement	H index	% Name agreement	H index
Mean	24.18	4.64	23.85	4.55	24.01	4.59
SD	18.18	1.15	17.64	1.14	17.89	1.14
Median	18.34	4.95	18.18	4.84	18.18	4.88
Range	25.88	1.71	20.24	1.55	23.40	1.62
Q1	9.41	3.82	10.71	3.87	9.64	3.86
Q3	35.29	5.53	30.95	5.42	33.04	5.48

SD, standard deviation; *Q1*, 25th percentile; *Q3*, 75th percentile

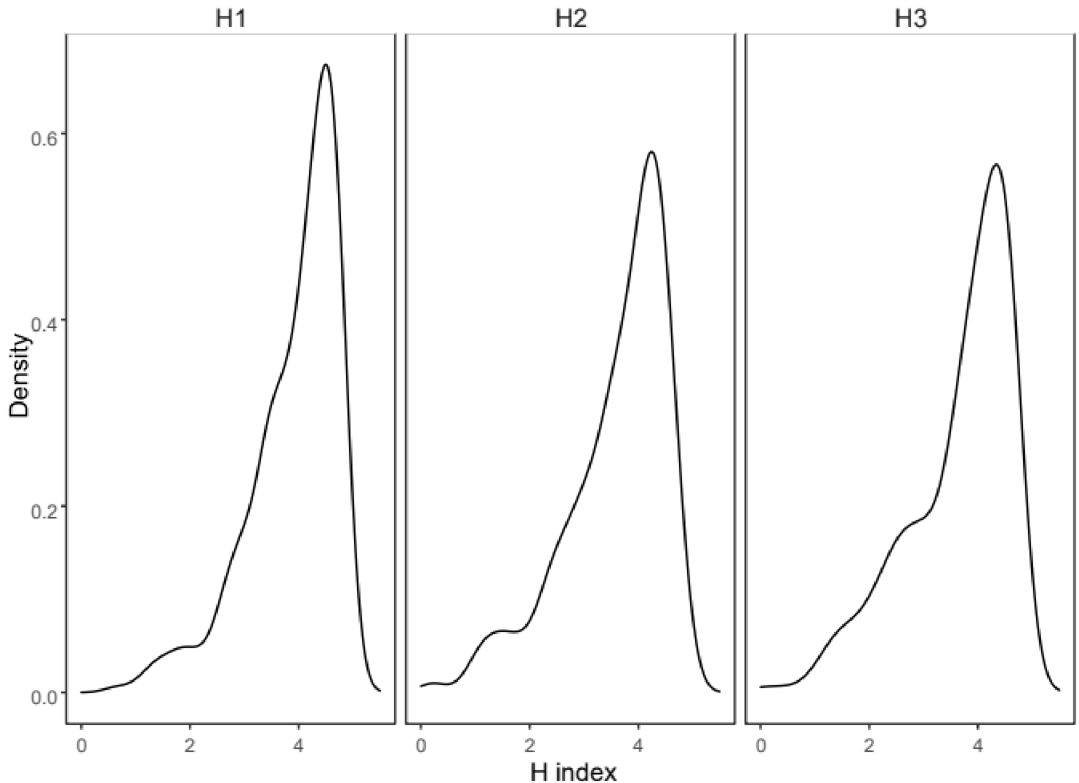


Figure 2.

Table 3
ICC index for image agreement, AoA, image variability, familiarity, visual complexity and concreteness.

Variables	Set A	Set B
Image Agreement	0.91	0.89
AoA	0.74	0.78
Image Variability	0.72	0.65
Familiarity	0.87	0.88
Visual Complexity	0.90	0.92
Concreteness	0.87	0.91

Note. AoA, Age of Acquisition.

Table 4

Summary of all correlations performed within and across questionnaires 2 and 3

Matrix A. Correlations between the variables and the H index obtained using questionnaire 2

Variables	H2	AoA	Image Variability	Image Agreement
H2				
AoA	0.25*			
Image Variability	-0.01	-0.30*		
Image Agreement	-0.43*	-0.18*	0.04	

Matrix B. Correlations between the variables and the H index obtained using questionnaire 3

Variables	H3	Visual Complexity	Familiarity	Concreteness
H3				
Visual Complexity	0.30*			
Familiarity	-0.41*	-0.48*		
Concreteness	-0.42*	-0.37*	0.73*	

Matrix C. Correlations between the variables obtained using both questionnaires 2 and 3

Variables	AoA	Image Variability	Image Agreement
Visual Complexity	0.02	0.21*	-0.30*
Familiarity	-0.14*	0.05	0.62*
Concreteness	-0.15*	-0.003	0.67*

Note. H2, name agreement for questionnaire 2; AoA, Age of Acquisition; H3, name agreement for questionnaire 3.

Table 5

Correlations between the percentage of "yes" answers to the "already seen" question and the other variables measured in questionnaire 3.

Variables	H3	Visual Complexity	Familiarity	Concreteness
"Already Seen"	-0.30*	-0.46*	0.42*	0.28*

Note. H3, name agreement for questionnaire 3 with the H index.

Appendix A.a. Detailed information about the collected variables

The table below provides the following information: the concerned variable in the ”Variables” column; the questionnaires in which the variable occurs in the ”Questionnaire” column; the French (original) version of the question in the ”French Version” column; the English version of the question in the ”English Version” column.

Variables	Questionnaires	French Version	English Version
Name Agreement	Q1, Q2, Q3	Quel est le premier mot ou la première expression qui vous vient à l'esprit pour décrire cette image ?	What is the first word or expression which comes to your mind to describe this picture?
Use in Dialogue	Q1	Est-ce que vous utiliseriez ce mot ou cette expression pour décrire cette image à quelqu'un d'autre en situation de dialogue ? Si non, quel mot ou expression utiliseriez-vous ?	Would you use this word or expression to describe this picture to someone else in a dialogue situation? If not, what word or expression would you use?
Other Label	Q1	Est-ce qu'un autre mot ou une autre expression vous vient à l'esprit pour décrire cette image ? Si oui, lequel ou laquelle ?	Is there another word or expression that comes to your mind to describe this picture? If so, which one?
Image Agreement	Q2	Est-ce que l'image vous semble être représentative de ce mot ou de cette expression ?	Do you think the picture is representative of this word or expression?
AoA	Q2	A quel â ge pensez-vous avoir appris ce mot ou cette expression ?	At what age do you think you learned this word or expression?
Image Variability	Q2	Dans la vie de tous les jours, ce mot ou cette expression peut-il ou elle être représenté(e) visuellement de plusieurs façons différentes ?	In everyday life, can this word or expression be represented visually in several different ways?
Already Seen	Q3	Aviez-vous déjà vu cette image auparavant ?	Have you ever seen this picture before?
Familiarity	Q3	Est-ce que la forme de l'image vous semble familière ?	Does the shape of the picture look familiar to you?
Visual Complexity	Q3	Est-ce que l'image vous semble simple ou complexe ?	Does the picture seem simple or complex to you?
Concreteness	Q3	Est-ce que l'image vous semble représenter un concept abstrait ou concret ?	Does the picture seem to represent an abstract or concret concept to you?

Appendix A.b. Values obtained on the various variables for each picture

The table below provides the following information: the set and identification code of each picture are provided in the Set ImgCode column; the overall H index (obtained by taking into account the data from all three questionnaires) is provided in the ”H” column; the H index for questionnaires 1, 2 and 3 separately is provided in columns ”H1”, ”H2” and ”H3”; the age of acquistion is provided in the ”AoA” column. SD: Standard deviation.

Set_imgCode	% of agreement	H	H1	H2	H3	Image agreement		Familiarity		Visual Complexity		Image Variability		AoA		Concreteness	
						Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
A_A02	56.47	2.78	3.57	2.16	1.27	3.66	1.14	3.57	1.20	2.04	0.74	3.28	1.19	2.17	0.76	3.68	1.19
A_A04	20.73	4.80	3.86	3.81	4.03	2.90	0.77	2.68	1.12	2.57	0.88	3.24	0.95	2.34	1.08	2.56	1.09
A_A10	15.29	5.04	4.56	3.78	4.18	3.59	0.91	3.04	1.26	2.43	1.07	3.10	1.29	1.97	1.02	3.36	1.16
A_A13	6.33	5.41	4.42	4.39	4.39	2.76	1.09	2.57	1.14	2.86	0.97	3.28	1.10	2.55	0.95	2.43	1.20
A_A14	15.85	4.94	4.12	4.02	4.04	3.24	1.02	3.21	1.07	2.46	0.84	3.45	1.06	1.97	0.98	2.71	1.05
A_A16	6.49	5.59	4.65	4.36	4.37	2.86	0.99	3.00	1.22	2.68	1.12	3.10	1.14	2.66	0.97	2.56	1.12
A_A30	14.12	5.04	4.25	4.19	3.98	2.62	1.01	2.57	1.10	2.71	0.98	2.97	1.15	2.45	1.02	2.30	0.99
A_A31	27.06	4.28	3.62	3.62	3.75	2.79	1.08	2.71	1.38	2.82	1.06	2.76	1.18	2.59	0.95	2.39	1.10
A_A32	7.14	5.60	4.72	4.30	4.50	3.07	1.07	2.86	1.21	2.86	1.04	3.14	1.16	2.34	1.20	2.63	1.11
A_A34	3.66	6.20	4.79	4.74	4.62	2.39	1.03	2.18	1.16	3.46	1.04	3.46	1.14	2.71	1.12	2.28	1.37
A_A35	8.33	5.50	4.32	4.30	4.50	2.83	1.00	2.75	1.29	2.93	0.90	3.10	1.11	2.48	1.30	2.63	1.15
A_A38	7.23	5.78	4.56	4.49	4.49	2.97	1.24	3.11	1.34	2.39	1.13	2.76	1.21	2.55	1.15	2.79	1.26
A_A52	4.71	5.95	4.72	4.58	4.81	2.55	0.99	2.21	1.17	3.04	1.20	2.28	1.07	2.41	0.95	1.89	0.80
A_A60	4.82	5.94	4.72	4.57	4.43	2.28	1.19	2.07	1.05	3.64	1.03	2.45	1.21	2.24	1.21	2.04	0.98
A_A66	14.12	5.19	4.22	4.02	4.38	3.21	1.01	1.93	0.98	3.79	1.03	3.07	1.39	3.10	1.08	1.48	0.75
A_A67	10.59	5.54	4.86	3.85	4.38	2.79	0.90	2.46	1.04	3.46	0.96	3.45	1.15	2.14	0.79	2.25	1.14
A_A69	5.88	6.01	4.63	4.86	4.74	2.79	1.26	2.57	1.32	3.29	1.15	2.41	0.95	2.72	1.28	2.41	1.08
A_A73	7.23	5.75	4.72	4.36	4.49	2.83	1.10	3.25	1.14	1.96	0.84	2.90	1.18	2.41	1.05	2.82	1.28
A_A74	9.41	5.13	4.42	4.18	4.18	3.14	1.09	3.07	1.33	1.93	0.72	2.93	1.28	2.86	0.99	3.11	1.31
A_A76	9.41	5.60	4.63	4.51	4.32	2.72	1.19	2.61	1.40	2.36	1.13	2.69	1.23	2.59	1.24	2.44	1.34
A_A78	10.71	5.34	4.51	4.30	4.38	3.48	1.24	2.86	1.35	2.29	1.24	2.72	1.46	2.62	1.15	2.70	1.51
A_A84	6.10	5.88	4.72	4.49	4.68	3.24	0.83	2.89	1.17	2.43	0.92	3.31	1.11	2.31	1.04	2.96	1.20
A_A90	49.41	3.35	3.48	2.79	2.08	3.62	0.98	3.54	1.14	2.14	0.65	3.24	1.02	2.45	1.06	3.50	1.20
A_A91	38.82	3.64	3.64	2.95	2.68	3.62	0.68	3.54	1.10	2.71	0.94	3.45	0.91	2.10	0.86	3.50	1.14
A_A92	20.00	5.10	4.22	3.89	4.31	2.82	1.22	3.07	1.44	2.14	1.11	3.57	1.45	2.07	1.21	2.78	1.31
A_A93	51.76	2.61	2.67	2.38	1.58	4.28	1.07	4.50	0.69	1.89	0.83	3.90	1.11	1.72	1.07	4.29	0.98
A_A94	16.47	3.91	3.27	3.44	3.62	3.66	1.01	3.54	1.10	1.68	0.72	2.76	1.21	2.24	0.99	3.32	1.33
A_A95	28.24	4.55	3.95	3.69	3.48	3.83	0.60	3.00	1.19	2.57	1.10	3.69	0.97	1.90	0.98	2.89	0.92
A_A96	50.59	3.52	3.37	2.92	2.13	2.82	0.98	2.71	1.27	3.29	1.12	3.71	1.12	1.71	0.81	2.67	1.30
A_A98	7.06	5.71	4.38	4.28	4.59	3.28	1.00	2.68	1.06	3.25	0.97	3.45	1.12	2.41	0.98	2.79	1.17
A_A100	8.33	5.68	4.79	4.42	4.38	3.72	1.13	3.18	1.09	2.93	0.94	3.14	0.99	2.52	0.87	3.22	1.19
A_A106	7.14	5.67	4.72	4.11	4.81	3.64	0.83	3.36	1.22	3.25	1.08	3.54	1.10	1.75	0.70	3.43	1.23
A_A109	8.24	5.68	4.79	4.28	4.35	3.45	1.12	2.96	1.23	3.36	1.19	3.14	1.13	2.10	0.90	3.00	1.28
A_A110	7.06	5.72	4.72	4.39	4.57	3.55	0.69	3.21	1.10	3.11	1.10	3.72	0.88	1.93	0.80	3.18	1.19
A_A111	3.53	6.31	4.86	4.79	4.81	3.07	1.07	2.57	1.35	3.64	1.10	3.52	1.12	2.45	1.15	2.50	1.17

Appendix A.b. Values obtained on the various variables for each picture (continued)

Set_imgCode	% of agreement	H	H1	H2	H3	Image agreement		Familiarity		Visual Complexity		Image Variability		AoA		Concreteness	
						Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
A_A113	10.59	5.70	4.56	4.38	4.59	2.76	1.02	2.89	0.96	3.00	0.98	3.55	1.09	2.34	1.04	2.61	1.13
A_A114	14.12	5.74	4.58	4.39	4.64	3.93	0.75	3.50	0.96	3.04	1.07	3.45	1.09	1.93	0.92	3.25	1.24
A_A117	10.59	5.40	4.32	4.49	4.49	3.86	0.69	3.18	0.98	2.86	0.93	3.55	1.12	2.10	0.90	3.50	1.00
A_A118	4.71	5.96	4.65	4.79	4.66	3.93	0.59	3.43	1.07	3.11	1.17	3.28	1.03	2.34	0.94	3.57	1.14
A_A120	8.24	5.81	4.58	4.39	4.66	4.28	0.75	3.68	1.06	2.50	1.04	3.59	1.12	2.52	1.18	3.36	1.06
A_A121	9.41	5.70	4.86	4.25	4.52	3.83	0.80	3.07	1.18	3.39	1.17	3.38	0.98	2.21	1.01	3.36	1.16
A_A122	3.53	6.19	4.86	4.65	4.81	2.48	1.06	2.25	1.11	3.89	0.99	3.21	1.15	2.62	1.01	1.79	0.92
A_A124	20.00	4.89	4.02	3.95	4.05	3.83	0.54	2.89	1.17	2.96	1.00	3.97	0.98	2.07	1.10	3.00	1.22
A_A125	71.76	1.89	1.71	1.11	1.98	4.38	0.49	3.89	0.83	2.57	1.26	3.55	0.95	1.52	0.83	4.07	0.86
A_A126	67.06	1.83	2.40	0.99	1.23	4.21	0.62	3.75	1.11	2.39	1.03	3.03	1.09	1.55	0.78	3.93	1.05
A_A127	74.12	1.69	1.80	1.13	1.34	4.28	0.59	3.71	1.15	2.71	1.24	3.38	0.94	1.59	0.78	3.96	1.17
A_A128	75.29	1.58	1.34	1.40	1.30	4.62	0.68	4.04	1.04	2.57	1.37	2.69	0.93	2.72	0.80	4.46	0.84
A_A130	11.76	5.19	4.49	4.12	3.87	2.79	0.82	2.39	1.26	3.29	1.05	3.24	1.33	2.17	1.07	2.73	1.19
A_A133	29.41	3.30	3.43	2.49	3.05	4.21	0.73	3.64	0.99	2.79	0.99	3.14	1.03	2.00	1.07	3.54	1.26
A_A134	54.12	2.67	2.91	1.54	2.28	4.00	0.46	3.14	1.04	2.93	1.05	3.97	1.02	1.59	0.82	3.46	1.00
A_A137	8.24	5.86	4.69	4.56	4.66	3.07	1.13	2.64	1.16	3.57	0.96	3.41	1.15	2.34	1.23	2.63	1.15
A_A138	25.88	4.48	3.69	3.56	3.79	2.97	1.27	2.29	1.12	3.32	1.09	2.97	1.05	2.10	0.90	2.32	1.09
A_A139	38.82	2.96	2.99	2.73	2.29	3.93	0.46	3.46	0.96	2.68	0.86	3.28	1.36	2.10	0.82	3.50	1.04
A_A141	42.35	3.19	3.18	2.35	2.62	3.24	0.95	2.89	1.17	3.32	0.90	3.41	1.12	1.83	0.97	2.89	1.23
A_A146	9.52	4.91	4.28	3.84	4.28	3.31	1.14	2.32	1.02	3.64	1.13	3.38	1.29	2.21	1.05	2.57	1.29
A_A148	80.00	1.49	1.06	1.27	1.30	3.90	0.94	2.82	1.06	2.61	1.03	2.83	1.07	1.86	0.83	3.29	1.08
A_A151	41.18	2.86	3.04	2.38	1.90	4.14	0.44	3.71	1.15	2.14	0.93	3.41	0.95	2.10	0.86	3.93	1.18
A_A157	20.00	4.88	4.11	4.16	3.56	2.69	1.07	2.57	1.23	3.64	0.91	3.31	1.17	2.45	0.95	2.44	1.25
A_A161	34.12	2.90	3.03	2.29	2.01	4.21	0.82	3.46	1.04	2.61	0.99	3.45	1.09	2.17	1.04	3.50	1.11
A_A162	31.76	3.66	3.27	3.26	3.02	3.41	1.05	3.29	1.01	2.43	0.84	3.21	1.24	2.14	1.03	3.32	1.12
A_A164	35.29	3.23	2.94	2.99	2.43	3.83	0.54	3.79	0.83	1.79	0.50	3.34	0.90	2.31	0.85	4.14	0.85
A_A166	21.43	4.96	4.25	4.13	3.88	3.17	1.04	2.36	1.19	2.89	1.10	2.72	1.10	2.14	0.79	2.25	1.08
A_A179	6.02	5.91	4.51	4.53	4.74	2.93	1.10	2.68	1.36	3.61	0.99	3.17	1.31	2.76	1.06	2.52	1.25
A_A180	7.06	6.09	4.72	4.72	4.52	3.03	1.18	2.61	1.20	3.43	1.03	3.03	1.21	2.69	1.14	2.57	1.26
A_A183	9.41	5.39	4.46	4.31	4.23	3.00	1.16	2.46	1.26	3.86	1.01	2.79	1.32	2.79	1.24	2.29	1.08
A_A184	4.71	5.98	4.86	4.65	4.42	3.45	1.02	2.43	1.07	3.32	1.28	3.48	1.09	2.52	1.27	2.61	1.10
A_A192	16.67	4.60	4.00	3.51	3.82	3.00	1.04	2.93	1.15	3.04	0.96	3.24	0.87	2.14	1.06	2.82	0.94
A_A193	14.12	5.13	4.04	4.06	4.57	2.69	1.07	2.68	1.25	2.89	1.17	2.83	1.17	2.31	1.00	2.22	1.22
A_A198	71.76	1.89	1.41	1.94	1.65	3.90	0.82	3.86	0.80	2.36	1.06	2.62	1.01	2.45	0.87	3.61	1.03
A_A201	10.59	5.49	4.63	4.35	4.42	2.52	1.24	1.71	0.98	3.21	1.32	3.00	1.22	2.38	0.94	1.81	1.11

Appendix A.b. Values obtained on the various variables for each picture (continued)

Set imgCode	% of agreement	H	H1	H2	H3	Image agreement		Familiarity		Visual Complexity		Image Variability		AoA		Concreteness	
						Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
A_B01	16.47	5.26	4.19	4.22	4.40	3.79	0.73	3.04	1.20	3.25	1.04	3.62	1.21	2.10	0.82	2.93	1.15
A_B02	9.41	5.58	4.63	4.22	4.31	2.93	1.03	2.71	1.15	3.79	0.88	3.66	1.11	2.21	1.15	2.48	1.25
A_B04	16.47	5.01	4.22	3.53	4.08	3.79	1.05	2.96	1.20	3.25	1.11	3.69	0.93	1.83	0.97	2.71	1.38
A_B05	9.41	5.47	4.58	4.03	4.52	3.10	0.98	2.68	1.31	3.07	1.02	3.41	1.21	2.17	1.10	2.64	1.06
A_B07	11.76	5.53	4.72	4.25	4.08	3.45	0.91	3.00	1.25	3.36	1.10	3.79	1.15	2.41	1.12	2.86	1.15
A_B08	5.95	6.08	4.86	4.65	4.61	2.76	1.21	2.36	1.03	3.43	0.96	3.59	1.09	2.07	1.00	2.44	1.15
A_B10	4.71	5.96	4.79	4.72	4.49	3.66	0.97	3.07	1.09	3.18	1.19	3.62	1.15	2.31	1.07	3.04	1.20
A_B12	12.94	5.81	4.63	4.39	4.64	3.38	1.15	2.79	1.29	3.54	1.07	3.10	1.35	2.07	1.03	2.82	1.28
A_B13	11.76	5.16	4.65	3.85	4.40	3.59	0.91	3.00	1.22	3.32	1.12	3.31	1.34	2.03	1.05	2.93	1.21
A_B14	18.82	4.93	4.38	3.62	4.11	3.90	0.77	3.57	1.10	3.07	1.21	3.79	0.86	1.93	1.07	3.04	1.26
A_B15	16.47	5.46	4.39	3.95	4.49	3.72	0.88	2.86	1.18	3.07	1.15	3.24	1.09	2.55	1.06	3.11	1.17
A_B18	7.06	5.61	4.44	4.65	4.57	3.45	0.74	3.07	1.05	2.75	1.04	3.31	1.04	2.24	1.12	2.75	1.08
A_B20	20.00	5.37	4.63	3.81	4.25	3.24	1.18	2.64	1.06	2.61	1.13	3.28	1.31	2.38	1.15	2.41	0.97
A_B22	17.65	5.31	4.04	4.09	4.52	2.62	0.90	2.54	1.29	2.82	1.02	3.00	1.16	2.17	1.10	2.59	1.12
A_B32	8.33	5.95	4.86	4.39	4.74	2.41	1.12	2.46	0.96	3.11	0.92	3.10	1.08	2.31	1.31	2.11	0.80
A_B33	10.59	5.31	4.39	4.00	4.32	3.07	0.96	2.71	1.21	2.32	0.77	2.93	1.19	3.00	1.34	2.54	1.23
A_B37	9.41	5.51	4.69	4.30	4.28	2.79	0.94	2.64	1.03	3.18	0.90	3.17	1.17	2.34	1.11	2.22	0.85
A_B38	12.94	5.45	4.79	4.22	4.11	3.14	0.88	2.79	1.17	2.75	1.04	3.24	1.24	2.59	1.05	2.57	1.00
A_B39	11.90	5.47	4.63	4.32	4.32	3.10	0.90	2.79	1.29	2.89	1.20	3.17	0.85	2.17	0.97	2.75	1.27
A_B40	25.88	4.79	4.03	3.10	4.16	2.97	0.98	2.32	1.12	3.61	0.96	2.97	0.98	2.52	1.02	2.56	1.23
A_B43	14.29	5.36	4.32	3.78	4.33	2.62	1.12	2.54	1.04	2.96	1.04	3.38	1.29	2.59	1.02	2.71	1.18
A_B46	10.59	5.14	4.42	4.09	4.14	2.79	1.01	2.21	1.07	2.68	1.02	2.93	1.03	2.72	1.16	2.39	1.03
A_B48	27.06	3.79	3.34	3.20	3.02	3.07	1.28	2.82	1.16	3.07	1.02	2.97	1.02	2.48	1.06	2.88	1.48
A_B51	21.18	4.68	3.46	4.24	3.62	2.55	0.99	2.82	1.47	2.50	1.17	2.72	1.22	2.38	0.78	2.43	1.20
A_B53	7.14	5.45	4.39	4.28	4.38	2.72	1.13	2.68	1.19	2.68	1.09	2.93	1.16	2.31	1.04	2.48	0.98
A_B54	36.47	4.25	3.74	3.64	2.99	3.34	0.90	3.50	1.29	1.93	0.72	2.83	1.26	2.21	1.01	3.32	1.33
A_B55	15.29	4.53	4.16	3.66	3.77	3.17	1.04	2.96	1.32	1.93	1.09	2.62	1.05	1.83	0.76	2.46	1.26
A_B56	35.29	4.11	3.60	3.41	3.48	3.41	1.18	3.43	1.17	1.75	0.65	3.31	1.26	1.79	0.86	2.86	1.21
A_B57	61.18	2.76	2.70	2.01	1.85	3.59	0.98	2.93	1.30	2.21	0.99	2.52	1.06	3.03	0.82	2.71	1.15
A_B59	15.29	5.61	4.53	4.03	4.57	2.69	1.31	2.14	1.01	2.64	1.19	2.66	1.37	2.45	1.02	2.00	0.96
A_B60	17.86	5.03	4.32	4.16	3.94	2.86	1.27	2.43	1.07	2.43	1.03	2.76	0.99	2.93	1.13	2.26	1.35
A_B61	11.90	5.52	4.53	4.38	4.49	3.14	1.09	2.25	1.11	2.54	1.23	2.76	1.30	2.07	0.88	2.14	1.11
A_B63	12.94	5.63	4.53	4.72	4.11	2.86	0.99	2.64	1.22	2.32	1.16	3.45	1.27	2.45	1.06	2.19	1.08
A_B64	41.18	4.00	3.96	2.88	3.02	4.41	0.68	3.14	1.21	2.11	0.79	3.21	1.32	2.21	0.94	2.93	1.30
A_B67	9.64	5.21	4.25	4.14	4.49	2.93	1.10	2.86	1.11	2.18	0.98	3.03	1.12	2.41	1.24	2.39	1.13

Appendix A.b. Values obtained on the various variables for each picture (continued)

Set imgCode	% of agreement	H	H1	H2	H3	Image agreement		Familiarity		Visual Complexity		Image Variability		AoA		Concreteness	
						Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
A_B74	39.29	4.18	3.89	2.52	3.38	4.07	0.88	4.29	0.71	1.50	0.64	2.72	1.31	2.59	1.15	4.00	1.22
A_B77	30.59	3.77	2.80	3.23	3.41	4.38	0.86	4.14	0.93	1.93	1.09	2.72	1.19	2.52	1.06	3.54	1.40
A_B78	54.12	2.96	3.20	1.55	2.64	4.62	0.73	4.36	0.87	1.43	0.84	2.79	1.45	1.90	0.72	4.11	1.31
A_B79	27.38	4.43	4.06	3.48	3.38	3.71	0.81	3.64	1.19	1.68	0.72	2.64	0.99	2.64	0.95	3.39	1.23
A_B81	50.59	3.41	3.71	1.81	2.90	4.79	0.62	4.39	0.99	1.43	0.88	2.93	1.46	2.34	1.01	4.00	1.36
A_B82	56.47	3.18	3.81	0.85	2.83	4.55	0.99	4.43	0.96	1.43	0.88	2.90	1.26	2.03	0.78	4.29	1.30
A_B90	5.88	5.84	4.72	4.63	4.59	2.79	1.01	2.57	1.14	3.43	0.88	3.41	1.09	2.48	1.09	2.56	1.01
A_B94	40.00	4.22	4.32	2.39	3.38	3.03	1.05	2.82	1.22	3.32	0.86	3.41	1.12	2.07	1.00	2.39	1.07
A_B95	41.18	4.09	3.81	2.88	3.02	3.21	0.94	2.71	1.05	2.71	1.05	3.66	1.11	1.69	0.93	2.68	1.06
A_B96	40.96	3.75	3.57	2.81	2.98	2.97	0.94	2.39	1.10	3.54	1.04	3.41	1.02	2.14	1.03	2.20	0.91
A_B97	49.38	3.54	3.20	3.16	2.29	3.00	0.98	2.46	1.26	3.82	0.86	3.36	1.37	2.07	1.15	2.29	1.12
A_B98	42.35	3.93	3.41	3.24	2.55	3.07	1.13	2.71	1.05	3.25	1.04	3.34	1.29	1.93	1.03	2.69	1.01
A_B99	32.94	4.51	4.16	3.64	3.38	2.43	1.17	2.32	1.22	3.61	0.99	3.54	1.04	1.79	0.79	2.23	1.11
A_B100	30.95	4.02	3.35	3.64	3.09	3.21	1.01	3.11	1.07	2.29	0.90	3.21	1.15	2.34	1.01	2.89	1.13
A_B101	25.88	3.25	3.05	2.87	2.31	4.17	0.80	3.46	1.10	2.14	0.59	3.24	1.09	2.38	1.08	3.50	1.17
A_B102	37.65	3.10	3.07	2.48	2.19	3.72	0.92	3.43	1.10	2.14	0.65	3.55	1.35	1.69	0.97	3.11	0.96
A_B103	46.99	3.63	3.51	3.06	2.24	3.24	0.83	2.79	0.96	2.39	0.96	3.48	1.15	1.90	1.01	2.88	1.17
A_B105	35.29	3.68	4.25	2.91	2.25	3.38	0.90	2.57	1.10	2.71	0.85	3.83	1.04	1.76	0.95	2.57	1.03
A_B109	43.53	2.78	1.82	2.49	2.61	3.90	0.98	3.86	1.11	1.96	0.92	3.52	1.21	1.90	1.21	3.63	1.11
A_B111	48.24	3.52	2.95	2.88	2.64	3.07	1.07	3.07	1.21	2.21	0.88	3.45	1.30	2.07	1.19	3.18	0.98
A_B112	44.71	3.28	2.97	2.51	2.67	4.10	0.62	3.68	0.98	1.79	0.57	3.59	1.05	2.10	1.18	3.64	1.10
A_B114	52.94	3.40	3.37	3.02	2.25	3.41	0.98	2.79	1.17	3.11	0.96	3.28	1.13	2.14	1.27	2.96	1.23
A_B116	40.00	4.27	4.18	3.02	2.90	3.24	1.09	2.82	1.19	3.14	0.89	3.76	1.15	1.97	1.15	2.89	1.20
A_B121	22.35	5.10	4.49	3.87	3.94	2.45	1.06	2.21	0.99	3.71	0.85	3.69	1.17	2.17	1.17	1.93	0.90
A_B122	22.62	4.81	4.28	3.97	3.80	2.52	0.91	2.32	1.09	3.61	0.92	3.03	1.12	2.17	1.20	2.38	0.85
A_B123	12.94	5.53	4.46	4.12	4.28	2.34	0.90	1.86	1.04	4.07	0.90	3.07	1.13	2.31	1.14	2.13	1.15
A_B124	40.74	4.17	3.71	3.35	2.90	2.32	1.22	2.21	1.03	3.57	0.84	3.39	1.37	2.14	1.18	2.42	1.14
A_B133	8.24	5.64	4.72	4.32	4.18	3.00	1.13	2.89	1.07	2.39	0.83	3.03	1.15	2.69	1.04	2.43	1.03
A_B134	37.65	3.52	3.54	2.67	2.76	3.69	0.81	3.82	1.06	1.64	0.73	3.21	1.26	2.00	0.80	3.39	1.20
A_B139	22.08	4.95	4.16	3.82	3.90	2.97	1.05	2.50	1.20	2.89	0.99	3.10	1.08	2.93	1.03	2.58	1.17
A_B141	10.59	5.24	4.58	4.24	3.87	2.93	1.21	2.57	1.32	2.86	0.97	3.28	1.21	2.14	1.21	2.32	1.02
A_B142	19.75	4.93	4.22	3.73	4.00	3.21	1.15	2.79	1.07	2.89	1.10	3.17	1.07	2.34	0.97	2.56	1.15
A_B146	18.82	4.45	4.28	3.63	3.54	3.38	0.82	3.18	1.16	2.54	0.88	3.34	1.17	2.17	0.85	2.63	1.21
A_B147	22.89	4.84	4.25	3.56	3.78	3.14	1.09	3.29	1.15	2.18	0.86	2.55	1.06	2.72	1.31	2.86	1.27
A_B149	10.71	5.36	4.44	4.22	4.50	3.21	0.94	2.71	1.05	2.82	0.98	2.90	1.21	2.31	0.89	2.43	1.00

Appendix A.b. Values obtained on the various variables for each picture (continued)

Set imgCode	% of agreement	H	H1	H2	H3	Image agreement		Familiarity		Visual Complexity		Image Variability		AoA		Concreteness	
						Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
A_B150	28.92	4.48	3.90	3.11	3.79	2.93	1.07	2.61	1.26	3.14	1.04	3.14	1.19	2.41	0.82	2.59	1.22
A_B151	31.76	4.40	3.73	3.48	3.57	3.28	0.80	3.36	1.03	2.36	0.95	2.97	0.94	2.17	0.76	3.00	1.22
A_B153	57.14	2.53	1.86	1.81	2.57	3.72	1.07	3.54	1.14	2.07	0.81	2.69	1.14	2.79	1.01	3.46	1.14
A_B154	33.33	2.79	2.67	2.19	2.55	3.38	0.94	3.61	0.96	2.43	0.92	2.62	1.05	2.93	1.19	3.32	0.94
A_B156	22.35	4.45	3.65	3.57	3.62	3.07	0.96	3.36	1.03	2.04	0.64	3.00	1.34	2.34	1.17	3.21	1.10
A_B157	62.35	2.42	1.60	2.80	1.09	4.00	0.96	4.04	1.00	1.82	0.82	2.55	1.06	2.31	1.00	4.00	1.09
A_B158	22.35	4.50	3.47	3.57	4.22	3.41	0.98	3.25	1.04	2.18	0.86	2.45	1.09	2.69	1.20	3.25	1.04
A_B159	25.88	4.79	4.19	3.77	4.03	2.86	0.88	3.21	1.29	2.54	0.92	2.59	1.09	2.34	0.94	3.11	1.09
A_B160	7.14	5.46	4.63	4.35	4.33	2.59	1.09	2.54	1.17	2.25	0.89	2.69	1.11	2.48	1.09	2.44	1.09
A_B161	13.58	5.03	3.98	4.06	4.00	3.03	0.98	2.43	1.14	2.79	0.88	2.90	1.21	2.34	1.04	2.54	1.14
A_B168	9.41	5.66	4.65	4.16	4.59	3.45	0.99	2.64	1.22	3.07	1.02	3.69	1.14	1.93	1.07	2.71	1.08
A_B169	16.47	5.30	4.32	4.19	4.23	2.34	1.04	2.04	1.10	3.50	1.00	2.86	0.95	2.72	1.10	1.77	0.76
A_B170	23.53	4.10	3.69	2.90	3.33	3.03	1.02	2.21	0.96	2.79	0.96	3.41	1.09	2.28	1.13	2.39	0.88
A_B178	32.14	4.40	4.42	3.38	2.91	2.75	1.04	2.18	1.19	2.86	1.01	3.57	1.03	1.93	1.02	2.52	1.12
A_B185	3.53	6.00	4.72	4.65	4.74	2.34	1.04	2.54	1.20	3.25	0.93	3.34	1.47	2.14	1.03	2.61	1.07
A_B192	43.53	3.31	3.50	2.31	2.45	3.48	0.83	3.50	1.00	2.32	0.72	3.38	1.15	1.69	0.76	3.36	1.03
A_B194	4.76	5.91	4.86	4.58	4.66	2.72	1.10	2.11	1.07	3.11	1.29	2.69	0.97	2.76	1.12	1.93	1.04
A_B195	30.59	4.42	3.91	3.00	3.79	3.57	0.69	2.93	1.12	2.57	1.00	3.61	1.20	1.61	0.79	2.64	1.16
A_B196	11.39	5.49	4.72	4.28	3.91	2.62	0.94	2.21	1.03	3.04	1.10	3.03	1.02	2.66	1.08	2.44	1.12
A_B199	20.48	5.03	4.32	3.86	4.01	2.48	1.12	2.46	1.23	2.71	1.21	3.03	1.18	2.38	1.01	2.68	1.35
A_B200	32.14	4.55	3.81	3.81	3.79	2.72	1.31	2.61	1.23	2.93	1.12	2.38	1.24	2.59	1.15	2.19	1.08
A_B201	6.02	5.95	4.72	4.42	4.57	2.10	1.14	1.93	0.98	2.89	1.37	2.79	1.35	2.38	1.15	1.76	0.88
A_B205	23.81	4.92	4.19	4.04	3.81	2.93	0.72	2.64	0.91	2.39	0.99	3.21	0.99	2.25	1.17	2.54	1.07
A_C03	36.47	3.78	3.75	2.98	2.38	3.52	0.63	3.32	1.02	2.29	0.90	3.41	0.82	2.31	1.17	3.14	1.01
A_C07	74.12	1.72	1.81	0.22	1.71	4.38	0.56	3.79	1.17	2.14	0.80	3.10	0.86	1.83	0.71	3.82	0.90
A_C08	16.47	4.76	4.05	4.25	3.65	2.90	1.14	2.96	1.26	2.79	0.99	2.97	1.30	2.45	1.24	2.46	1.10
B_A01	22.62	4.03	3.39	3.11	3.74	3.29	0.85	3.36	1.19	2.21	0.74	2.93	0.98	2.82	1.16	3.04	0.96
B_A03	48.81	3.41	3.03	3.46	1.51	3.18	1.06	3.46	1.14	2.39	0.99	2.89	1.20	2.39	0.88	3.50	1.20
B_A05	8.33	5.29	4.25	4.32	4.32	2.36	0.95	2.39	1.20	2.61	0.96	3.29	1.21	2.54	1.07	2.25	1.04
B_A06	25.00	4.37	3.74	3.41	3.55	3.43	1.29	3.07	1.25	2.11	0.74	3.25	1.17	2.14	0.93	3.15	1.20
B_A07	15.48	4.78	3.97	4.04	3.87	3.25	0.84	3.18	1.09	2.29	0.98	2.79	1.03	2.18	0.90	3.25	1.00
B_A08	22.89	4.60	4.19	3.51	3.96	2.70	0.95	2.79	1.29	2.50	0.88	3.22	0.97	2.56	1.12	2.85	1.05
B_A09	18.07	4.34	3.44	3.56	3.96	3.54	1.04	3.32	1.19	2.21	1.03	2.61	1.13	2.04	0.74	3.14	1.41
B_A11	15.48	4.86	3.95	4.28	3.80	2.82	1.28	3.07	1.12	2.36	0.78	2.79	1.34	2.54	0.96	2.68	1.16
B_A12	13.10	4.92	4.31	4.01	4.18	3.32	0.94	3.32	1.22	2.14	0.76	2.75	1.00	2.07	1.02	3.50	0.92

Appendix A.b. Values obtained on the various variables for each picture (continued)

Set imgCode	% of agreement	H	H1	H2	H3	Image agreement		Familiarity		Visual Complexity		Image Variability		AoA		Concreteness	
						Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
B_A15	12.05	5.21	4.56	4.25	4.14	3.14	0.89	3.04	1.26	2.07	0.66	3.11	0.96	2.75	1.00	2.86	1.11
B_A19	22.08	4.46	3.95	3.38	3.91	3.00	1.09	3.07	1.15	2.64	1.16	2.75	0.97	3.71	1.21	3.00	1.05
B_A20	16.05	5.17	4.39	4.18	3.95	2.46	1.14	2.43	1.26	3.18	0.98	3.18	0.82	2.64	0.95	2.69	1.23
B_A22	4.88	6.01	4.56	4.59	4.75	3.00	1.02	2.93	1.12	2.29	0.85	2.57	1.00	3.07	1.09	2.68	1.09
B_A23	13.41	4.89	4.39	3.77	4.13	2.71	0.90	2.79	1.17	2.39	0.79	3.36	1.28	2.21	0.99	2.68	1.16
B_A25	4.82	5.94	4.63	4.52	4.58	3.00	1.31	2.46	1.07	3.00	0.86	3.00	1.25	2.71	1.15	2.26	1.02
B_A26	28.57	4.81	3.38	4.03	4.25	3.29	1.27	3.00	1.36	2.57	1.00	3.25	1.32	2.46	1.07	3.04	1.40
B_A27	10.71	5.42	4.49	4.38	4.45	2.79	1.17	2.46	1.29	3.00	1.02	3.32	1.19	2.50	0.88	2.07	1.05
B_A28	14.29	5.27	4.46	4.35	4.11	3.14	0.85	2.79	1.20	2.43	0.79	3.14	1.30	2.68	1.19	2.57	1.00
B_A29	14.29	5.09	3.95	3.79	4.74	2.93	1.21	2.93	1.05	3.11	0.88	2.50	1.00	2.61	1.03	2.32	1.02
B_A33	8.33	5.81	4.72	4.35	4.64	2.86	1.08	2.21	1.07	3.54	1.04	3.25	1.27	2.68	0.86	2.37	1.18
B_A36	19.05	4.94	4.22	3.69	4.11	2.64	0.91	2.61	1.31	3.43	0.92	3.46	1.07	2.36	1.10	2.18	0.90
B_A37	10.84	5.55	4.65	4.31	4.25	2.61	1.17	2.36	1.06	2.32	0.94	2.93	1.25	2.68	1.06	2.29	1.15
B_A39	7.14	5.98	4.72	4.57	4.66	2.50	0.96	2.46	1.29	2.64	1.10	3.07	1.21	2.79	1.13	2.21	1.26
B_A43	6.10	5.82	4.79	4.07	4.55	2.81	1.02	2.25	1.32	2.46	1.04	2.58	1.17	2.96	1.22	2.21	1.29
B_A46	5.95	6.05	4.79	4.45	4.66	2.68	1.16	2.00	1.19	2.61	1.10	3.18	1.22	2.25	0.93	2.07	1.17
B_A53	9.52	5.34	4.22	4.08	4.57	2.33	1.04	2.18	1.33	3.07	0.90	3.26	1.20	2.22	0.80	2.07	1.15
B_A54	4.76	5.96	4.79	4.66	4.42	2.61	1.26	2.39	1.17	3.46	1.04	2.79	1.23	2.64	0.95	2.15	1.03
B_A57	12.20	5.36	4.58	4.28	4.15	2.61	0.96	2.54	1.14	3.11	0.96	2.61	0.88	2.82	1.02	2.29	1.18
B_A64	15.58	5.44	4.35	4.25	4.3	2.68	1.06	2.64	1.22	3.25	0.93	3.29	1.24	2.82	1.12	2.46	1.04
B_A65	11.90	5.29	4.49	3.98	4.35	3.18	1.09	2.46	1.23	3.07	1.15	3.57	1.29	2.36	0.99	2.39	1.07
B_A68	63.10	2.24	2.49	1.23	2.04	3.89	1.07	3.64	0.95	2.50	1.00	3.00	1.09	2.29	0.76	3.57	1.00
B_A70	22.22	4.55	4.09	3.48	3.54	3.14	1.18	2.96	1.14	2.96	0.96	3.75	1.04	2.39	1.07	2.74	1.20
B_A71	8.33	5.59	4.58	4.57	4.31	2.82	1.12	2.61	1.23	2.75	0.97	3.46	1.00	2.57	1.10	2.29	1.12
B_A72	7.23	5.60	4.72	4.53	4.35	2.85	1.06	2.43	1.20	2.39	0.96	2.96	1.43	2.85	1.06	2.07	1.02
B_A77	19.28	5.19	4.56	3.79	4.15	3.26	1.10	3.00	1.22	2.14	1.21	2.93	1.30	2.89	0.97	2.61	1.23
B_A87	17.50	4.57	3.91	3.79	3.59	3.36	1.03	3.14	1.18	1.93	0.81	2.71	1.24	2.64	1.13	3.18	1.16
B_A88	39.29	3.38	3.04	2.97	2.98	3.50	0.79	3.50	0.92	2.57	1.03	3.43	1.07	2.54	0.96	3.79	1.03
B_A89	51.19	2.56	2.66	2.24	1.70	4.11	0.79	4.32	0.72	1.79	0.63	3.71	1.12	2.00	1.02	4.21	0.79
B_A101	13.10	5.00	4.51	4.07	3.82	3.57	0.96	2.96	1.07	2.86	0.97	3.32	1.06	2.68	1.02	3.39	0.99
B_A105	29.76	4.32	3.53	3.34	3.75	3.93	0.90	3.61	0.83	2.32	0.72	3.32	1.16	2.36	1.16	3.71	0.81
B_A107	10.71	5.33	4.63	4.04	4.11	4.32	0.72	3.71	1.08	2.82	1.31	3.25	1.17	2.54	1.00	4.18	0.98
B_A112	9.52	5.65	4.46	4.38	4.49	3.36	1.10	2.61	1.31	3.46	1.07	3.39	1.03	2.39	1.13	2.81	1.39
B_A115	11.90	5.69	4.51	4.57	4.49	4.00	0.77	2.96	0.96	3.29	1.18	3.32	1.33	2.61	1.29	3.43	1.03
B_A119	4.76	6.16	4.69	4.81	4.74	3.11	1.03	2.93	1.18	3.11	0.92	3.18	1.12	2.64	0.91	3.11	1.17

Appendix A.b. Values obtained on the various variables for each picture (continued)

Set imgCode	% of agreement	H	H1	H2	H3	Image agreement		Familiarity		Visual Complexity		Image Variability		AoA		Concreteness	
						Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
B_A123	8.33	5.50	4.58	4.16	4.66	3.68	0.86	2.71	1.05	3.36	1.06	3.86	1.08	2.00	0.82	3.29	1.27
B_A129	34.52	2.99	2.80	2.65	2.34	3.89	0.88	3.54	0.92	2.64	0.91	3.18	1.25	2.14	0.97	3.86	0.80
B_A131	36.90	3.57	3.71	2.68	2.59	3.82	0.72	3.68	0.98	2.89	0.96	2.86	0.97	2.04	0.79	3.93	0.90
B_A132	51.19	2.91	2.88	2.31	2.49	3.75	0.80	3.75	0.93	2.61	1.17	3.18	0.9	1.93	0.77	4.14	0.85
B_A135	80.95	1.18	1.20	1.36	0.44	4.18	0.94	3.36	1.22	2.71	1.21	2.39	0.92	2.61	0.79	4.07	0.81
B_A136	19.05	4.70	4.49	3.51	3.72	4.04	0.74	3.46	0.96	3.21	0.99	3.11	1.03	2.64	0.91	3.82	0.90
B_A140	58.33	2.37	2.53	1.84	1.68	4.00	0.94	3.50	1.29	2.93	1.30	2.36	0.99	2.82	0.98	3.89	0.96
B_A142	27.38	4.03	4.10	3.26	2.92	4.50	0.58	3.64	1.10	3.00	1.41	3.00	0.94	2.71	0.98	4.25	0.84
B_A143	29.76	3.90	3.61	3.15	2.96	3.48	1.09	3.36	1.10	3.25	1.08	3.19	1.21	1.89	0.75	3.14	1.01
B_A145	26.19	4.66	4.28	3.71	3.45	3.32	1.09	3.04	1.14	3.46	0.96	3.21	1.32	2.57	1.10	3.39	0.63
B_A147	64.29	2.14	2.19	1.44	1.89	4.18	0.90	3.96	0.96	2.61	1.17	3.57	1.20	1.50	0.58	4.18	0.86
B_A149	20.24	3.86	3.48	3.16	3.48	3.79	0.83	3.54	1.07	2.29	0.90	3.21	1.26	1.79	0.79	3.96	0.84
B_A150	63.10	1.85	1.44	1.66	1.60	4.32	0.61	3.46	1.14	3.21	1.13	2.57	0.92	2.43	1.00	3.93	0.90
B_A152	21.43	4.14	3.50	3.84	3.37	3.46	1.00	3.43	1.00	2.96	1.26	3.11	1.10	2.29	1.18	3.46	1.07
B_A153	28.57	3.80	3.45	2.79	3.44	3.18	0.86	2.86	1.15	2.54	0.88	2.75	1.00	1.89	0.92	3.29	1.01
B_A154	26.19	4.06	3.39	3.77	3.48	3.32	0.94	3.04	1.10	2.29	0.85	3.00	1.12	2.43	1.07	3.29	1.12
B_A155	24.10	3.72	3.52	3.12	3.28	4.04	0.88	3.07	1.15	2.93	0.94	3.00	0.98	2.11	1.07	3.50	0.88
B_A156	70.24	2.14	2.65	1.09	1.44	3.89	0.99	3.18	1.06	2.79	0.96	2.79	1.13	2.11	0.83	3.64	0.99
B_A158	39.02	2.09	2.73	1.48	1.59	3.96	0.96	3.89	0.96	2.43	1.07	3.46	1.29	2.50	1.26	3.64	1.06
B_A159	32.14	4.20	3.63	3.43	3.34	2.96	1.04	2.11	1.07	3.64	0.83	3.54	1.00	2.64	1.34	2.39	1.13
B_A160	37.35	2.34	2.97	1.67	2.00	4.64	0.56	4.25	0.75	2.18	1.25	3.75	1.32	2.43	1.07	4.39	0.63
B_A165	36.90	4.28	3.64	3.94	2.83	2.85	0.77	2.39	1.07	2.68	0.98	3.30	1.07	2.44	1.05	2.68	1.16
B_A167	32.14	3.30	2.73	2.46	3.37	3.18	0.86	2.96	1.23	2.11	0.63	3.18	1.06	1.86	0.97	2.96	1.10
B_A172	4.88	5.89	4.56	4.59	4.75	2.62	1.17	2.64	1.22	2.93	0.98	3.08	1.09	2.69	0.97	2.50	1.04
B_A174	6.41	5.64	4.56	4.21	4.36	2.59	1.15	2.21	1.10	3.29	1.01	2.96	1.19	2.44	0.97	2.15	1.01
B_A175	8.33	5.62	4.56	4.52	4.45	2.61	1.10	2.21	1.03	3.57	0.88	3.14	1.18	2.50	0.84	2.18	0.82
B_A176	13.10	5.40	4.24	4.08	4.52	2.82	1.09	2.18	1.12	3.54	0.84	3.04	1.14	2.82	0.98	1.89	1.01
B_A177	10.71	5.34	4.25	4.28	4.32	2.82	1.06	1.75	0.80	3.50	1.17	3.18	1.31	2.36	0.87	2.19	1.08
B_A178	4.76	6.09	4.79	4.64	4.66	2.61	1.13	2.46	1.26	3.96	0.96	3.29	1.15	2.54	0.84	2.68	1.33
B_A181	9.52	5.65	4.72	4.04	4.59	3.71	1.08	3.25	1.21	3.21	1.29	3.14	1.11	2.46	1.04	3.64	1.19
B_A182	4.76	6.09	4.79	4.81	4.66	2.39	1.13	1.96	1.04	3.96	1.00	3.21	1.32	2.46	1.00	2.07	1.25
B_A185	14.46	5.44	4.56	4.21	4.11	3.50	0.96	2.39	1.23	3.57	0.96	3.32	1.25	2.39	1.17	2.89	1.31
B_A187	54.76	3.18	2.39	2.64	2.54	3.50	0.92	3.14	1.08	2.21	0.88	3.36	0.99	2.75	1.17	3.04	1.14
B_A188	9.52	5.23	4.51	4.04	4.45	3.14	0.89	3.71	0.90	1.64	0.68	2.86	1.15	2.25	0.65	3.36	0.95
B_A189	15.66	4.85	3.88	3.96	3.80	3.54	0.88	3.46	1.20	2.00	0.82	3.00	1.19	2.96	1.26	3.75	0.93

Appendix A.b. Values obtained on the various variables for each picture (continued)

Set imgCode	% of agreement	H	H1	H2	H3	Image agreement		Familiarity		Visual Complexity		Image Variability		AoA		Concreteness	
						Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
B A191	15.48	4.50	4.09	3.84	3.72	3.32	1.09	3.18	1.33	2.11	0.79	3.21	0.99	2.18	1.06	3.18	0.94
B A194	30.95	3.72	2.71	2.71	3.77	3.21	1.03	3.11	1.10	2.96	0.92	3.18	1.19	2.39	1.10	3.18	1.12
B A195	13.75	4.57	3.91	3.90	3.81	3.00	0.77	3.00	1.05	2.39	0.88	3.04	1.07	2.21	1.03	2.57	1.03
B A196	14.29	4.94	4.39	4.14	4.01	2.89	1.20	2.79	1.13	3.04	1.04	2.89	0.99	2.93	1.15	2.71	1.21
B A197	27.71	3.74	3.41	3.22	3.28	3.39	0.88	2.96	1.14	3.04	1.07	3.07	1.15	2.75	1.00	2.93	1.18
B A199	3.57	5.83	4.72	4.31	4.66	2.68	0.98	2.21	1.17	3.39	0.88	3.18	1.25	2.39	0.83	2.11	0.92
B A202	15.38	5.09	4.23	4.21	3.86	2.86	0.93	3.21	1.26	2.21	0.99	3.00	1.05	2.79	0.96	3.25	1.11
B B03	7.14	5.48	4.56	4.42	4.31	2.93	1.21	2.39	0.96	3.71	1.12	3.14	1.18	2.21	0.96	2.54	1.29
B B06	9.52	5.56	4.86	4.38	4.15	3.50	0.96	3.25	1.08	3.07	1.05	3.25	1.35	2.64	1.13	3.50	1.04
B B09	4.76	6.12	4.79	4.66	4.64	2.79	1.10	2.04	1.29	4.14	0.93	3.14	1.21	2.68	1.02	2.14	1.11
B B11	14.29	5.46	4.56	4.28	4.08	2.93	0.98	2.39	1.23	3.79	1.23	3.04	1.23	2.21	0.74	2.54	1.23
B B16	15.48	5.16	4.24	4.38	4.08	3.43	1.00	2.86	1.35	3.25	1.08	3.04	1.10	2.79	1.17	3.39	1.37
B B17	21.43	4.43	4.00	3.73	3.24	3.29	1.05	3.86	0.85	1.82	0.72	2.89	1.20	2.64	0.83	3.43	0.96
B B19	76.19	1.56	1.48	1.09	1.02	4.29	0.85	3.93	1.02	1.79	0.88	2.68	1.12	2.93	0.90	4.07	0.90
B B21	17.07	5.23	4.56	4.24	3.79	2.82	1.12	3.50	1.07	2.43	0.88	2.89	1.23	2.39	1.31	3.18	1.09
B B23	9.52	5.50	4.65	4.42	4.14	3.00	0.98	2.96	1.10	2.11	0.88	3.21	1.07	2.39	0.96	2.96	1.00
B B24	8.33	5.33	4.21	4.64	4.11	3.11	1.10	3.39	1.29	2.57	0.88	3.25	1.17	2.68	0.90	3.07	1.18
B B28	16.87	4.83	4.31	3.70	4.01	3.43	0.88	2.96	1.20	2.14	0.71	3.32	0.98	2.57	1.03	3.21	1.17
B B29	25.00	4.58	3.44	4.15	3.64	2.93	0.90	2.71	1.15	2.43	0.92	3.14	1.18	3.00	1.33	2.64	1.03
B B30	50.00	3.18	3.14	2.18	2.25	3.54	1.04	3.71	1.08	2.07	0.94	2.89	1.23	2.79	0.92	3.44	1.12
B B31	7.14	5.65	4.58	4.32	4.42	2.64	1.10	2.68	1.02	2.71	0.98	3.11	1.13	2.79	1.13	2.50	1.11
B B35	7.14	5.46	4.63	4.52	4.52	3.14	1.08	2.68	1.16	2.71	1.01	3.18	1.06	2.68	1.12	3.11	1.07
B B36	34.18	4.00	3.70	2.73	3.41	3.21	1.03	3.14	1.30	2.18	0.94	2.54	1.00	2.96	1.26	2.96	1.20
B B41	54.22	2.98	2.60	1.77	2.49	3.25	1.08	2.79	1.29	3.64	1.06	2.89	0.96	3.61	0.96	3.07	1.21
B B42	26.19	4.26	3.98	3.05	3.68	3.21	0.92	2.71	1.21	2.75	0.80	3.46	1.07	2.71	1.05	2.61	1.07
B B44	26.51	4.35	3.41	3.51	3.78	3.18	1.15	2.89	1.17	2.21	0.69	2.79	1.15	2.79	1.15	3.26	1.02
B B47	13.10	4.93	4.00	4.09	4.04	3.11	1.10	2.57	1.14	2.86	1.04	3.04	0.92	3.18	1.19	2.86	1.18
B B49	15.48	4.88	4.19	3.46	4.21	2.93	1.09	3.11	1.23	1.89	0.69	2.93	1.12	2.79	0.99	3.11	1.15
B B50	30.95	4.62	3.13	3.41	4.32	3.11	1.26	2.64	1.22	1.86	0.85	2.54	1.32	2.32	1.12	2.75	1.11
B B52	38.10	3.40	2.60	2.82	2.87	4.29	1.08	4.14	1.18	1.50	0.64	1.96	1.26	2.07	0.81	4.04	1.04
B B58	26.19	4.95	4.46	3.55	4.11	3.07	1.05	2.54	1.23	2.21	0.99	2.82	1.28	3.18	1.02	2.57	1.17
B B62	7.14	5.88	4.56	4.66	4.64	2.54	1.17	2.43	1.32	2.71	0.98	2.93	1.25	2.89	1.29	2.04	1.07
B B65	29.76	3.80	3.22	3.27	3.02	4.07	1.02	3.32	1.28	1.36	0.68	2.46	1.20	2.25	0.70	3.32	1.33
B B66	47.62	3.08	2.88	2.49	2.57	4.07	0.98	3.39	1.29	1.50	0.58	2.86	1.48	1.86	0.65	3.61	1.10
B B68	18.29	5.15	4.35	3.95	4.11	3.07	1.12	2.86	1.15	1.79	0.92	3.11	1.20	2.43	0.88	2.96	1.29

Appendix A.b. Values obtained on the various variables for each picture (continued)

Set imgCode	% of agreement	H	H1	H2	H3	Image agreement		Familiarity		Visual Complexity		Image Variability		AoA		Concreteness	
						Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
B B69	31.33	3.53	2.90	3.07	2.78	3.54	1.10	3.61	1.07	1.57	0.74	3.21	1.13	2.36	1.10	3.29	1.12
B B70	65.48	1.87	2.04	1.17	0.92	4.71	0.85	4.68	0.61	1.11	0.31	2.11	1.50	1.89	0.83	3.86	1.41
B B71	15.48	4.88	4.05	4.09	3.77	3.04	1.32	2.96	1.17	2.00	0.98	2.75	1.21	2.39	1.03	2.57	1.20
B B73	21.43	4.80	4.30	3.47	3.94	3.21	1.07	3.36	1.10	1.79	0.96	2.68	1.22	2.29	0.81	2.75	1.14
B B76	58.33	2.96	2.12	2.13	2.64	4.11	0.99	3.82	1.22	1.64	0.73	3.11	1.45	2.64	1.03	3.54	1.23
B B80	69.05	2.14	2.11	1.55	1.44	4.14	0.59	3.82	0.98	2.00	0.86	3.25	1.24	2.39	0.79	3.68	1.09
B B87	25.00	4.56	3.67	3.72	3.90	3.64	0.85	3.89	0.99	2.18	1.03	3.46	0.95	2.46	0.96	4.07	0.82
B B92	16.87	5.46	4.69	4.45	3.85	2.54	1.03	2.57	1.30	3.79	0.86	3.00	1.20	2.43	1.03	2.43	1.10
B B93	45.24	3.85	3.31	2.64	2.99	2.71	1.09	2.68	1.50	3.46	0.98	3.75	0.96	2.14	0.93	2.36	1.22
B B106	50.00	2.90	2.97	2.15	2.03	3.50	1.04	2.96	1.10	2.14	0.59	3.64	1.13	2.04	0.88	3.39	1.07
B B108	35.71	3.95	3.84	2.73	3.26	3.32	0.94	3.00	1.05	2.36	0.73	3.57	1.00	2.36	1.28	3.07	1.09
B B110	20.24	5.30	4.19	4.03	4.45	2.43	1.07	2.18	1.06	3.79	0.99	3.21	1.13	2.21	0.88	2.33	1.14
B B117	19.51	5.24	4.04	4.31	4.38	2.74	1.13	2.04	1.10	3.89	0.83	3.37	1.28	2.48	0.98	2.11	1.09
B B118	38.55	4.18	3.55	3.38	3.28	2.75	1.04	2.32	1.25	3.54	0.96	3.79	1.26	2.07	0.90	2.21	1.03
B B119	30.95	4.68	3.81	3.71	3.81	3.29	1.27	2.54	1.04	3.50	0.88	3.00	1.12	2.82	1.22	2.86	1.01
B B120	50.60	3.36	2.70	3.31	2.52	3.36	1.03	2.79	1.17	2.96	0.96	3.07	1.33	2.29	1.08	3.00	0.98
B B125	48.81	3.63	3.31	2.38	2.95	2.71	0.94	3.04	1.14	3.32	0.98	3.68	1.02	2.32	0.94	2.75	1.17
B B126	15.48	5.24	4.42	4.04	4.16	2.71	0.98	2.57	1.26	3.32	1.09	3.43	1.17	2.68	1.09	2.43	1.07
B B128	16.67	4.94	4.14	3.66	4.28	3.25	0.84	2.68	1.19	2.57	0.79	3.00	1.15	2.00	0.86	2.86	1.11
B B130	23.81	4.38	3.50	3.40	3.67	3.18	1.09	3.54	1.17	2.07	0.66	3.54	1.04	2.18	0.90	3.43	0.92
B B131	10.71	4.85	3.89	4.07	4.08	3.11	0.88	3.21	0.92	2.14	0.71	3.18	1.09	2.61	0.88	3.14	0.89
B B135	24.10	4.74	4.25	3.64	3.73	3.25	0.80	3.11	1.20	2.25	0.75	3.14	1.27	2.43	0.88	3.00	1.11
B B143	22.62	4.20	3.37	3.18	3.62	3.75	1.00	3.07	1.05	2.54	0.84	3.21	1.07	2.25	0.84	2.96	1.14
B B144	8.33	5.42	4.28	4.42	4.42	3.89	0.79	3.64	1.03	2.54	0.88	2.93	0.90	3.00	1.15	3.57	1.00
B B145	28.57	4.16	3.88	3.31	3.22	3.21	0.99	3.39	1.29	2.36	0.87	3.14	1.04	2.54	0.84	3.54	0.96
B B148	37.35	3.82	2.99	3.65	2.90	3.59	0.93	2.96	1.00	2.68	0.82	3.26	1.06	2.48	1.19	2.93	1.04
B B152	15.48	4.54	3.76	3.82	3.45	3.57	1.10	3.21	1.13	2.32	0.82	2.46	1.14	2.68	1.36	3.00	0.98
B B155	59.52	2.86	2.01	2.52	1.98	3.71	1.08	3.57	1.26	1.93	0.90	2.43	1.00	2.36	1.10	3.82	0.94
B B163	14.29	4.85	3.91	4.23	3.66	2.93	1.25	2.89	1.20	2.25	0.93	2.39	1.13	3.25	1.14	2.39	1.10
B B164	9.52	5.98	4.79	4.52	4.64	2.86	1.18	2.89	1.17	2.89	0.96	2.61	1.10	3.43	1.32	2.86	1.15
B B165	20.24	4.87	4.16	4.08	3.87	3.32	0.86	2.61	1.34	2.93	0.98	3.54	1.14	2.29	1.08	2.89	0.99
B B166	6.02	5.55	4.51	4.49	4.50	2.86	1.01	2.75	1.11	3.07	0.94	3.32	1.09	2.14	1.01	2.59	1.08
B B167	11.90	5.70	4.58	4.42	4.47	2.61	1.07	2.39	1.31	3.25	1.21	3.61	1.10	2.57	1.14	2.29	1.30
B B171	15.48	5.31	4.63	3.82	4.15	2.75	1.11	2.64	1.45	2.93	1.02	2.96	1.23	2.43	1.14	2.63	1.28
B B172	13.25	5.49	4.49	4.15	4.49	2.39	1.07	2.11	1.26	3.71	0.90	3.32	1.25	2.50	1.26	1.82	1.06

Appendix A.b. Values obtained on the various variables for each picture (continued)

Set imgCode	% of agreement	H	H1	H2	H3	Image agreement		Familiarity		Visual Complexity		Image Variability		AoA		Concreteness	
						Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
B_B173	12.05	5.22	4.44	3.86	4.21	2.68	1.16	2.32	1.28	3.39	1.03	3.32	1.16	2.46	1.17	2.11	1.13
B_B174	9.52	5.48	4.56	4.45	4.38	2.93	1.05	2.71	1.24	3.21	1.13	3.00	1.25	2.82	1.28	2.14	1.21
B_B175	35.71	3.39	3.48	2.53	2.46	4.25	0.80	4.07	0.90	2.07	1.09	2.54	0.96	3.21	0.96	4.07	0.94
B_B176	45.24	3.04	2.40	2.57	2.64	3.68	0.82	3.36	1.03	2.82	0.94	3.32	1.33	2.18	1.12	3.04	1.10
B_B177	25.00	4.02	3.55	3.34	3.08	3.43	0.96	2.93	1.27	3.07	0.98	3.36	1.19	2.43	1.20	3.00	1.19
B_B179	26.19	4.16	3.69	3.48	3.13	3.18	1.09	2.79	1.10	3.04	0.92	3.18	0.90	2.71	1.18	3.21	1.17
B_B180	26.19	4.73	3.85	3.65	3.86	3.43	1.07	2.89	1.31	2.68	1.09	3.00	0.94	2.57	1.03	2.59	0.89
B_B187	22.62	3.86	3.31	3.12	3.53	4.14	0.93	3.54	0.96	2.54	0.88	3.21	1.07	2.18	0.90	3.71	0.85
B_B193	3.61	5.94	4.56	4.68	4.42	2.68	1.02	2.29	0.98	2.68	0.94	3.39	1.23	2.89	1.20	2.07	1.02
B_B197	13.25	5.11	3.95	4.50	4.35	2.75	0.97	2.64	1.16	3.00	0.94	3.25	1.17	2.50	0.88	2.39	1.07
B_B198	35.71	3.57	3.24	3.01	2.57	3.46	0.74	2.64	1.13	3.36	0.99	3.50	1.23	2.36	0.78	2.93	1.12
B_B202	9.64	5.67	4.56	4.49	4.53	2.46	1.07	1.82	0.82	3.00	1.15	2.96	1.07	2.61	0.96	1.85	1.03
B_B203	28.57	4.43	3.74	3.39	3.59	2.56	0.93	2.39	1.10	2.93	0.86	3.52	1.19	2.04	1.06	2.26	0.94
B_B204	9.52	5.29	4.16	4.32	4.53	3.22	1.05	2.07	1.09	3.04	1.00	3.30	1.27	2.70	0.95	2.22	1.25
B_B206	8.43	5.89	4.25	4.49	4.74	2.77	1.14	1.86	1.21	3.46	1.04	2.81	1.33	2.50	1.03	2.07	1.17
B_C01	91.67	0.65	0.64	0.22	0.00	4.39	1.00	4.07	1.06	2.32	1.25	3.11	1.15	1.39	0.57	4.32	0.91
B_C02	6.02	5.94	4.79	4.57	4.68	2.46	1.10	2.29	1.09	3.18	1.21	3.54	1.10	2.39	0.96	1.92	1.02

Appendix B. Author version of the Fasquel et al. submitted paper

Title: How does the creation of new semantic relationships during dialogue impact long-term semantic representations after the dialogue?

Authors: Alicia Fasquel[1], Wassila El Mardi[2], Isabelle Bonnotte[1], Dominique Knutsen[1], & Angèle Brunellière[1]

Affiliation:

[1] Univ. Lille, CNRS, UMR 9193 - SCALab - Sciences Cognitives et Sciences Affectives, F-59000 Lille, France

[2] Laboratoire DysCo, Université Paris 8, Saint-Denis, France

Corresponding author:

Alicia Fasquel,

SCALab, CNRS UMR 9193, Université de Lille, Domaine Universitaire du Pont de Bois, BP 60149, 59653 Villeneuve d'Ascq, France

fasquel.a.97@gmail.com

Angèle Brunellière,

SCALab, CNRS UMR 9193, Université de Lille, Domaine Universitaire du Pont de Bois, BP 60149, 59653 Villeneuve d'Ascq, France

Tel: (+33)3 20 41 72 04

angele.brunelliere@univ-lille.fr

Abstract

To bridge the gap between dialogue studies and long-term semantic representations, three experiments were conducted to understand the extent to which new semantic relationships created during dialogue may change preexisting representations in long-term semantic memory after the dialogue. For this purpose, we developed an interactive agreement referential task to create new semantic relationships in dialogue. Then, participants performed either a semantic priming paradigm with a lexical decision task immediately or one day after a dialogue during which pairs of words referred or not to a single picture, or a semantic relatedness judgment task one day later. No significant effect of relatedness was found in the lexical decision task. However, when the attention was focused on semantic relationships, participants in the semantic relatedness judgment task rated the newly related words as more related semantically. The three experiments bear major implications for understanding on the links between dialogue and memory.

Max: 150 words

Keywords: dialogue, semantic relationships, long-term semantic representations, semantic memory

Introduction

Dialogue is a joint activity involving at least two partners who interact using language to reach a common goal (e.g., choosing the hour of a meeting or the destination of a trip). This daily activity requires the partners to coordinate the timing and content of their conversation to ensure easy and efficient communication and understand each other (Clark, 1996; Clark & Brennan, 1991; Pickering & Garrod, 2004, 2021; Sacks, Schegloff, & Jefferson, 1974). According to the collaborative approach, partners attempt to minimize the effort put into dialogue by producing utterances designed to their interlocutor to reach a mutual understanding (Clark, 1996; Clark & Brennan, 1991; Clark & Schaefer, 1989; Clark & Wilkes-Gibbs, 1986; Nault et al., 2023). For instance, they attempt to favor the production of words and utterances which are easily understandable for their current partner (Clark & Murphy, 1982). The implications of this process, referred to as audience design, are twofold. First, the production of such words may result in the emergence of new uses for known words and new meanings during dialogue. Second, partners tend to reuse the same words or expressions to refer repeatedly to the same objects throughout the conversation in order to improve mutual understanding (Brennan & Clark, 1996; Clark, 1996). Dialogue thus constitutes an ideal setting to integrate new uses for known words and new meanings into long-term memory. A key question concerning the interface between memory and language is understanding to what extent new semantic relationships created in dialogue change preexisting lexico-semantic representations in long-term memory after the dialogue. We address this question in the present study.

Dialogue and memory

The links between dialogue and memory have particularly been explored with the concept of common ground and the study of the use of referential labels in dialogue (e.g., Knutsen & Le Bigot, 2014; McKinley et al., 2017; Nault et al., 2023). Common ground includes knowledge and beliefs that conversational partners share and are aware of sharing (Brennan & Clark, 1996; Clark, 1996; Clark & Marshall, 1981). One of the main sources used to determine whether an information belongs to the common ground or not (along with other sources such as visual copresence or community memberships; Clark & Marshall, 1981) is the partners' memory of the ongoing interaction or past interactions. Common ground is used by dialogue participants to tailor utterances to their audience, as it enables them to determine whether their current partner is capable of understanding a given word or expression.

Over the past decades, common ground has been studied by examining the choice of words used to refer to objects and how this choice impacts referential decisions made later during the interaction, or during subsequent interactions (Brennan & Clark, 1996; Clark & Brennan, 1991; Clark & Wilkes-Gibbs, 1986; Metzing & Brennan, 2003). If speakers A and B talk about A's new pair of socks, they may do so by using a variety of referential labels such as "A's new socks", "the socks with a dog on it" or "A's red socks". However, once speaker's A and B found an agreement on the best reference to use, they tend to reuse the same label to refer to these socks throughout the conversation and during subsequent interactions (Brennan & Clark, 1996; Clark & Wilkes-Gibbs, 1986; Metzing & Brennan, 2003). This tendency to coordinate perspectives on an object by using a given reference is called, 'lexical entrainment', and originates from the creation of conceptual pacts between partners (Brennan & Clark, 1996; Brennan & Hanna, 2009). Conceptual pacts are defined as temporary agreements on how to refer to an object, a person or an entity (Brennan & Clark, 1996). They can be studied in experiments involving a referential communication task (initially developed by Krauss and Weinheimer, 1964; see also Clark & Wilkes-Gibbs, 1986) in which participants must discuss the location of pictures (often tangram pictures) in a grid. The result is that they quickly reach an agreement on how to name each picture. Once the conceptual pact is indeed created, it tends to be reused to refer to the same picture again during the following trials, as it henceforth belongs to the partners' common ground.

Studies involving a referential communication task thus suggest that dialogue partners rely on their memory to determine the references to use to refer to things as they interact. However, a number of questions regarding the interplay between memory and dialogue remain to be answered. Particularly, although research has examined participants' ability to store references in memory, the question of how conceptual pacts may affect one's preexisting representations in long-term memory after the dialogue has received very little attention. For instance, if speakers A and B agree on the use of the word "anchor" to refer to the picture shown in Figure 1, this new conceptual pact may create a new mapping in A and B's long-term memory between the picture and the word at the representational level, both during and after the dialogue. However, the possibility that preexisting representations in long-term memory change after the dialogue has never been addressed in dialogue research.

< Insert Figure 1 here >

The creation of a new mapping between a word and an object mentioned in a dialogue raises the theoretical question of how this mapping is stored in long-term memory. According to the classic view of declarative memory (Tulving, 1972), there is a clear distinction between semantic memory and episodic memory in long-term representations. Although episodic memory stores detailed events including the context in which the event took place in space and time, and the sensory modalities through which the information was conveyed, semantic memory is usually characterized by the storage of contextually-independent information. This distinction must be considered with caution because grounding models of semantic memory posit that sensorimotor modalities and context can play a crucial role in the construction of meaning (Barsalou, 2008; Matheson & Barsalou, 2018). Importantly, most of the current research on the link between dialogue and storage in long-term memory focuses on episodic memory. Some authors suggest that conceptual pacts are partner-specific (Brennan & Clark, 1996; Brennan & Hanna, 2009; Clark, 1996; Metzing & Brennan, 2003), or that specific dialogue partners act as retrieval cues for information shared during the dialogue through a process called resonance (Horton & Gerrig, 2002, 2005a, 2005b, 2016). More precisely, these studies explored what people remember from past interactions, how cues present at the time of the initial interaction may impact subsequent information retrieval from the common ground, and the nature of the cognitive processes that support information encoding and retrieval during dialogue. A more recent study by Nault et al. (2023) examined what people remember with verbatim and semantic recall memory performance after a referential communication task and its link with the strength of the common ground. Therefore, it remains to be understood how dialogue may affect people's preexisting representations stored in semantic memory, such that the representations changed after the dialogue would be activated independently of the presence of the partner or the recall of the dialogue situation.

The interactive alignment model of Pickering and Garrod (2004, 2021) provides interesting proposals on the way in which dialogue affects preexisting mental representations. This model proposed an alignment between linguistic representations in long-term memory during the dialogue such that similar representations between dialogue partners at different levels (semantic, syntactic and phonological) become activated as the interaction unfolds. It also suggests that dialogue partners become aligned on dialogue models, i.e., on their representations of the discussion at hand (Pickering & Garrod, 2021). According to the alignment model, information mentioned during the dialogue can be stored in both episodic and semantic memory. However, this model mainly explains the influence of dialogue on preexisting linguistic representations stored in long-term semantic memory by the change in their level of activation with increased activation of representations on which partners interact. Importantly, the model makes no assumptions as to whether these representations are modified after the dialogue.

In the present study, we want to address the question of changes in preexisting representations stored in long-term semantic memory following dialogue. More precisely, we examined whether and how the creation of a new semantic relationship between pairs of words referring to the same object in dialogue can impact long-term semantic representations after the dialogue (i.e., preexisting representations stored in long-term semantic memory). Because we want to explore the activation level of long-term semantic representations, we investigated the access to long-term semantic representations with classical methods by which we can describe the strength of activation into a large lexico-semantic network owing to a semantic relationship between pairs of words. In the following section, we describe methods to study the access to long-term semantic representations that we hereafter refer to lexico-semantic representations, and present some empirical evidence of their flexibility in memory.

Long-term semantic representations and their flexibility

Lexico-semantic representations are often conceptualized as a network of highly interconnected nodes that can be activated in order to retrieve the meanings of words (Collins and Quillian, 1969; Collins & Loftus, 1975). The structure of the lexico-semantic network has mainly been proposed to account for the well-known semantic priming effect, evidenced for the first time by Meyer & Schvaneveldt in 1971. The semantic priming effect corresponds to the fact that a word preceded by a semantically related word is recognized more quickly than a word preceded by a semantically unrelated word. This effect can be evidenced by the use of a lexical decision task, in which a prime is presented before a semantically related target about which participants have to decide whether or not if the target is a word (e.g., Brunellière et al., 2017; Brunellière & Bonnotte, 2018). Reaction times are then compared with trials in which the prime and the target are semantically unrelated. The main advantage of the lexical decision task is that it explores the access to long-term semantic representations without an explicit judgment on the degree of semantic relatedness between words. The semantic priming effect has thus been described as the result of semantic spreading activation through a lexico-semantic network (Hutchison, 2003; Lucas, 2000;

Neely, 1991; de Wit & Kinoshita, 2015), although additional interpretations (*retrospective semantic matching*, *post-access coherence checking*, *evidence accumulation*) have also been suggested as contributing to a semantic priming effect. Other methods have been developed to examine the access to long-term semantic representations, such as, the categorization task or the semantic relatedness judgment task (e.g., Brunellière et al., 2017; Brunellière & Bonnotte, 2018; de Wit & Kinoshita, 2015). The latter consists in the presentation of two words on which participants are explicitly asked to rate their degree of semantic relatedness with a Likert scale. In this task, there is an activation into the lexico-semantic network through which the top-down attention is focused on the degree of lexico-semantic relationships between words. Attention has been described as a multi-level system of weights and balances to prioritize specific mental processes (Narhi-Martinez, Dube, & Golomb, 2023). As proposed by Dehaene et al. (2006), top-down attention acts by amplifying the degree of activation into networks of interconnected nodes. Therefore, the semantic relatedness judgment task is thought to amplify the activation of the lexico-semantic network and help to track weaker connections into the lexico-semantic network, whereas an implicit task such as the lexical decision task would ensure access to stronger connections.

Although multiple tasks have questioned different aspects of the organization of semantic representations, the exact structure of the lexico-semantic network is still under debate (see Hutchison, 2003; Lucas, 2000; Kumar, 2021; McNamara, 2005; Neely, 1991). The usual consensus in the literature on the organization of semantic representations is that the degree of relatedness between words depends on the number of interconnections between nodes (Anderson, 1983; Collins and Loftus, 1975; Farah and McClelland, 1991; McRae, de Sa & Seidenberg, 1997; Plaut and Booth, 2000). Interestingly, these interconnections have been shown to be flexible, so that new meanings in language comprehension can change preexisting lexico-semantic representations and can be integrated into the lexico-semantic network (e.g., Rodd et al., 2012, 2013, 2016). In particular, Rodd and collaborators (2012) worked on the integration of new word-meanings mapping into the lexico-semantic network by reading paragraphs that described these meanings. Participants were informed that they would be exposed to new meanings for existing words and they were able to recall the new properties of these words, showing that new word-meaning mappings were learnt. For instance, participants had to learn that a “widow” (an already existing word) was “an animal that is forced out of their group”, which constitutes one among five properties of the new meaning attributed to the word in short paragraphs. Interestingly, novel meanings that were related to the original ones were easier to recall than unrelated novel meanings. These results highlight the integration of new word-meaning mappings into the preexisting lexico-semantic network. A similar pattern was found using incidental learning, in which participants were not informed that they would be exposed to new meanings and were not asked to learn them (Hulme et al., 2019). They read short paragraphs that biased the meaning of existing words towards new meanings and the more participants encountered the new meanings, the better their memory of them was in a cued recall task (see also in spoken language exposure with a word association task, Rodd et al., 2013; 2016). Sleep intervals have already been shown to improve the learning of new linguistic knowledge, such as phonology (Dumay & Gaskell, 2007), semantics (Tham et al., 2015) and grammar (Nieuwenhuis et al., 2013). Using the word-meaning priming paradigm in which participants were asked to generate an associated word either after ambiguous words primed by sentence exposure or in an unprimed condition, Gaskell et al. (2019) showed that the word-meaning priming effect was significantly stronger in participants who performed the task after a period of sleep, than in those who performed the task after a period of wake.

However, all the studies on the flexibility of lexico-semantic representations have been conducted in language comprehension (i.e., in non-interactive settings). Interestingly, people have the opportunity to create new mappings between words and meanings as they interact during dialogue. Further research is therefore necessary to better understand how dialogue can change preexisting lexico-semantic representations stored in long-term semantic memory. In the present study, we want to know whether dialogue may affect people's preexisting representations stored in semantic memory, such that the representations changed after the dialogue would be activated independently of the presence of the partner or the recall of the dialogue situation. We also question the persistence of such changes at the representation level one day after the dialogue and examine the nature of changes into the lexico-semantic network by using implicit or explicit tasks on the degree of relatedness.

The present study

The aim of the present set of experiments was to investigate how new semantic relationships created in dialogue change preexisting lexico-semantic representations in semantic memory after the dialogue. In order to do this, we developed a new methodological approach combining an interactive task and various tasks accessing the activation of lexico-semantic representations in long term memory after a dialogue. Importantly, we created a new experimental paradigm in dialogue that we call the interactive agreement referential task (IAR) during which partners interact via an audio-conference device to reach an agreement on how to name tangram pictures presented on their screens. This task allows participants to create new semantic relationships in dialogue by taking advantage of new word-visual representation mappings. Indeed, as illustrated in Figure 2, the purpose of this task is to create a new semantic relationship between two words by representing them with the same abstract visual representation (e.g., 'anchor' and 'hat' referring to the same tangram picture). In everyday life, a word can be detected as having a similar meaning as another word if they both refer to the same object. In this task, partners had to suggest words to name abstract pictures and justify their choices to their dialogue partner. They then must reach an agreement on their choices of words (see Figure 2). This task was thus organized so that the partners reach an agreement rapidly and explicitly. In line with the study by Betts et al., (2018) showing a boost of changes in lexico-semantic representations after three repetitions when their presentations were spaced out, pairs of words referring to the same abstract visual representation were repeated four times through different turn-takings (see below for a detailed description and Figure 2). Moreover, to ensure that the number of words associated with each picture remained constant across all participants, one of the partners was a confederate who followed a script while keeping the interaction as naturalistic as possible.

< Insert Figure 2 here >

As our research question concerns the possibility that new semantic relationships created in dialogue change preexisting lexico-semantic representations in semantic memory after the dialogue, our experimental paradigm in dialogue helped the abstraction of the new semantic relationship between pairs of words referring to the same abstract visual representation by changing the order of presentation of each word of the pair referring to the same visual representation. In addition, the modality through which these words were conveyed during the dialogue (i.e., the auditory modality) was different from the one in which the access to the lexico-semantic representations was assessed after the dialogue (i.e., written modality). This change in the modality of presentation between the dialogue phase and the assessment of lexico-semantic representations allowed to study the impact of dialogue on abstract representations at a lexico-semantic level stored in semantic memory.

Our working hypothesis was that pairs of words referring to a shared visual representation become more semantically related than pairs of words referring to two distinct visual representations after the dialogue. We manipulated the dialogues so that half of the pictures referred to one word only, while the other half referred to two words which were not semantically related before the dialogue (see Figure 3). To show evidence of changes in preexisting lexico-semantic representations after the dialogue, we compared the processing of pairs of words referring to a shared visual representation to that of pairs of words referring to two distinct visual representations after the dialogue. An effect of relatedness corresponded to the fact that pairs of words referring to a shared visual representation (i.e., pairs related during the dialogue that we called the ‘related condition’) were processed or judged as being more semantically related than pairs of words referring to two distinct visual representations (i.e., pairs unrelated during the dialogue that we called the ‘unrelated condition’).

< Insert Figure 3 here >

To address the question of the persistence of such changes at representation level one day after the dialogue and the nature of changes into the lexico-semantic network, we conducted three experiments involving the IAR task but varying according to the task assessing the access to preexisting lexico-semantic representations in semantic memory. In Experiments 1 and 2, we used an implicit task on the degree of relatedness with a visual lexical decision task during a semantic priming paradigm. We thus investigated whether changes in preexisting lexico-semantic representations in semantic memory were associated with the establishment of strong connections into the lexico-semantic network. This aspect was also ensured by the use of short presentation interval time (usually called stimulus onset asynchrony between pairs of stimuli, SOA), thus reducing the development of attentional processes (Neely, 1977, 1991). Although Experiment 1 was conducted to examine changes in preexisting lexico-semantic representations in semantic memory immediately after the dialogue, Experiment 2 explored the persistence in the changes in preexisting lexico-semantic representations stored in semantic memory one day after the dialogue. In the semantic priming paradigm with a lexical decision task, one of the words of the pair served as prime, while the other served as target. We expected targets belonging to pairs of words referring to a shared visual representation to be more quickly recognized following the presentation of the other word of the pair than targets belonging to pairs of words referring to two distinct visual representations after the dialogue. In line with the main objective of the present study, we explicitly led the participants in Experiment 3 to pay attention to the semantic relationship between the pairs of words by using a semantic relatedness judgment task one day after the dialogue. Thanks to this task, we explored whether changes in preexisting lexico-semantic representations in semantic memory were related to new weak yet reachable connections into the lexico-semantic network under the action of attentional processes, by amplifying the degree of activation into the network. It also assessed whether these changes persisted for one day after the end of the dialogue. In Experiment 3, we expected greater semantic relatedness scores for pairs of words referring to a shared visual representation than for pairs of words referring to two distinct visual representations after the dialogue.

At the end of each experiment, participants had to complete the Interpersonal Reactivity Index (IRI) questionnaire (Braun et al., 2015) and the Plymouth Sensory Imagery Questionnaire-French (PSIQ-F; Ceschi & Pictet, 2018). The latter were added as exploratory measures to evaluate a potential impact of empathy and mental imagery on the changes of lexico-semantic representations after the dialogue. Regarding the IRI questionnaire, one or several sub-dimensions of empathy may play a role in the changes of lexico-semantic representations after the dialogue, especially the Perspective Taking dimension. In fact, by switching from their own perspective to the confederate's perspective, people with more Perspective Taking abilities could be subject to more changes of lexico-semantic representations. With regards to the PSIQ-F questionnaire, we wanted to check whether visual mental imagery might affect our results, given that participants had to name abstract visual representations (i.e., tangram pictures) with real words.

EXPERIMENT 1

In Experiment 1, we assessed changes in preexisting lexico-semantic representations in semantic memory immediately after a dialogue during which pairs of words were associated with the same visual representation (i.e., the same tangram picture) to produce the creation of a new semantic relationship. Pairs of written words were presented in a lexical decision task using a semantic priming paradigm. We expected faster reaction times for pairs of words referring to same visual representation than for pairs of words referring to distinct visual representations, corresponding to an effect of relatedness.

Materials and methods

Participants

Forty-six native French speakers (35 female; 44 undergraduates and two community members; age, $M = 21.23$, Standard Deviation, $SD = 2.53$) were paid 15€ to take part in the experiment. All participants had normal or corrected-to-normal vision and no history of language disorder, auditory disorder or neurological disorder. They signed an informed consent form before the beginning of the experiment and were fully debriefed after the experiment. This work was approved by the Ethics Committee of the University of Lille (reference: 2021-512-S95). Six participants were excluded from the analysis of data (for more details, see the experimental design and data pre-processing section).

Confederates

During the dialogue, participants interacted with one of the two confederates of this experiment. Both were native French speaking women aged of 24 and 25 years old and had no history of language disorder, auditory disorder or neurological disorder. More precisely, twenty-two participants interacted with one of the confederates while eighteen participants interacted with the other.

Apparatus

The dialogues were recorded using a Zoom H4n with a dual-entry digital recorder (one per partner).

Stimuli

The interactive agreement referential task (IAR)

Two sets of 40 tangram pictures and their associated words were selected to be discussed by the confederate. A third set of 30 tangram pictures and their associated words was selected to be discussed by the participant. All the pictures came from the Fasquel et al. (2023) tangram picture database, which contains 332 tangram pictures labelled with either words or expressions, collected from 84 native French speakers. In this database, any given picture could be associated with a single word or expression, or with several different words or expressions (i.e., in some cases, all participants provided the same expression to refer to the same picture; in other cases, different participants provided different expressions to refer to the same picture, implying that this picture was associated with different expressions in the database). This database enabled us to select pictures associated with at least two plausible words or expressions (e.g., a picture could be associated with the words "anchor" and "hat"). We then pre-selected 148 pictures which were those associated with at least two plausible words or expressions from this initial database and we asked 84 new native French speakers to name them by providing only isolated words (this new database is online available from the following url, https://doi.org/10.5281/zenodo.10419752).

The first set of 40 tangram pictures (set 1) and words was selected among these 148 pictures. In this set, each picture was associated with two semantically unrelated words. Thus, for a picture to be included in this set, it had to be associated with two plausible words which were not semantically related before the dialogue. The degree of relatedness between each pair of words was pre-tested by 29 other native French speakers on a 7-point Likert scale (1 = not related at all, 7 = strongly related). We ensured that each pair was rated below 4, meaning that they were all considered as semantically unrelated. Overall, the mean relatedness was 1.75 (*SD* = 0.71). Due to the written presentation of words during the lexical decision task, each word did not exceed a length of 10 letters (range: 3-10 letters). The psycholinguistic characteristics of each word within pairs are reported in Table 1.

< Insert Table 1 here >

In contrast with the first set in which two words were associated with the same picture, the second set of 40 tangram pictures associated with words was selected to create pairs of words associated with two different pictures. To do so, each pair of words from the set 1 was divided into two so that one of the words remained associated with its corresponding picture from set 1, while the other word within each pair was associated with a new picture from the original database. Specifically, all these new pictures had been labelled at least once with their new associated word from set 1. More precisely, we selected these words from the Fasquel et al. (2023) database or the subsequent database of isolated words for 148 pictures. Each new picture from set 2 was therefore associated with only one word. In sets 1 and 2, all selected words were nouns. These two sets of pictures and words allowed to define our two experimental conditions with either related or unrelated pairs and to test our hypotheses.

For a picture to be selected in the third set of pictures and words, it had to be associated with words that were absent from sets 1 and 2. As explained below, these pictures were discussed by the participant rather than by the confederate. The third set allowed to make sure the participant remained engaged in the interactive agreement referential task. These pictures came from the Fasquel et al. (2023) database and their two most frequent words in the database were selected. Contrary to sets 1 and 2, the words could either be nouns or verbs and each pair of words associated with a picture could either be semantically related or not, as the data from the trials in which the pictures were described by the participant were not analyzed.

<u>The lexical decision task (LDT)</u>

The 40 pairs of words associated with the first and second sets of tangram pictures were used as prime-target pairs in the lexical decision task. They were divided into two experimental lists so that each pair of words was presented only once per participant but was presented in the two experimental conditions across participants (20 pairs of words per experimental condition, related vs. unrelated). The properties of primes and targets were matched in terms of frequency, length and neighbourhood between the two experimental lists (see Table 1). Word pairs were also matched in terms of semantic relatedness and co-occurrence frequency between the two experimental lists (semantic relatedness: list 1, $M = 1.75$, $SD = 0.83$, list 2, $M = 1.76$, $SD = 0.58$; co-occurrence frequency: list 1, $M = 0.35$, $SD = 0.93$, list 2, $M = 0.95$, $SD = 3.56$). The co-occurrence frequency was calculated with the same approach used in Brunellière et al. (2017) and in Brunellière and Bonnotte (2018). The same distribution in two lists was used for both the pictures of sets 1 and 2 in the IAR task and the words forming the experimental conditions in the lexical decision task. We added 40 filler pairs of words which were unrelated pairs of familiar nouns and 80 word-pseudoword pairs in each list. Importantly, forty words selected as primes of word-pseudoword pairs were produced in the IAR task, such that participants could not use the information related to repeated items during the dialogue to make their lexical decision on the targets. Pseudoword targets were orthographically legal and were constructed by replacing a letter in French words different from those used in the IAR task and from those used as fillers in the lexical decision task. Word and pseudoword targets were matched in length (number of letters, for words: $M = 6.76$, $SD = 1.68$; for pseudowords: $M = 7.18$, $SD = 1.57$).

Procedure

Experiment 1 took place on two consecutive days in order to keep the procedure constant across all three experiments (see Figure 4). In line with the main objective of the present study, the lexical decision task used in Experiment 2 and the semantic relatedness judgment task used in

Experiment 3 were carried out one day after the IAR task (i.e., day 2 after the dialogue). While session 1 corresponded to the tasks performed during day 1 (i.e., interactive agreement referential task and lexical decision task, in Experiment 1*)*, the session 2 consisted of tasks performed on the day 2 (i.e., filler task and questionnaires, in Experiment 1). Both sessions were performed in the same testing room and the overall experiment lasted around two and a half hours.

< Insert Figure 4 here >

Session 1

Interactive agreement referential task (IAR)

The participant and the confederate were placed in two separate rooms. The confederate read the scripts during the interaction, without being seen by the participant. The participant was informed that they would have to interact with another participant installed in another room using an audio-conference device and a headphone, and that the discussion would be recorded for the purpose of the experiment. They were also informed that, for the whole interaction they would be named “participant B” and that they will see the same pictures than the other person, named “participant A”. The purpose was to reinforce the participant’s impression that the confederate was also a naïve participant. The picture presentation was run on the computer screen using GNU Octave (version 6.4.0).

The IAR task was divided into 30 trials and a break was given to the participant after 15 trials. Each trial began with the instruction of which of the participant or the confederate had to propose words for the following pictures. The instruction was displayed on both the screen of the participant and the one of the confederates, and it was then followed by the presentation of three tangram pictures side by side. The instruction was divided into three steps (see Figure 2). The first step (i.e., naming step) was to choose one or two words for each of the three pictures displayed on the screen. More precisely, the person proposing the words on that trial had to choose words among a list of six possible words and to write their choices on a response sheet before giving the choices of words. The six-word list contained the pairs of words associated with the filler tangram pictures (see the Stimuli section for more details). When the words chosen to refer to each of the three tangram pictures had been provided, the partners went to the second step (i.e. justification step) and the person who had chosen the words had to justify their choices (i.e. to what extent the picture reminds them of the proposed word) for the three pictures. The last step (i.e. agreement step) was to ask the partner if he or she agreed on the proposed words and write the partner's answer on the response sheet. Then, the trial ended and the confederate pressed the space bar to start the next trial and to control the transition between trials, while the participant was informed that the experimenter performed this task. Crucially, rather than following the same instructions as the participant, the confederate followed a script which enabled them to produce the critical words four times in trials in which it was their turn to provide the words. Thanks to this structure, the two partners were able to reach an agreement rapidly and explicitly.

For ten of the thirty trials, participants had to propose words to name tangram pictures. Half of these trials were presented in each block (five per block of 15 trials). The purpose of these trials was to encourage the participant to actively engage in the task (rather than simply listening to the confederate). The data from the trials in which the participant had to name the pictures were not analyzed further. The partners were given a training phase to familiarize themselves with the procedure. This training phase was composed of one trial with words given by the confederate, followed by one trial with words proposed by the participant.

For twenty of the thirty trials, the confederate had to propose the words to refer to the tangram pictures. As mentioned previously, in these trials, the confederate had to read the scripts as naturally as possible (see for an example of script at the following URL, https://doi.org/10.5281/zenodo.10419752). The scripts followed the same rules as those given to the participant on how to proceed when naming the pictures (see Figure 2). First, the confederate proposed the words associated with each picture (see the stimuli section for more details on the choices of words). When a picture was associated with a pair of words, these words were placed within a window of 10 words in the same sentence (see Brunellière et al., 2017, for justification). Words given to different pictures within the same trial were spaced by at least 10 words, such that the participant would have no doubts about the choice of words for a given picture. To do so, filler sentences such as “I think that’s the best word” could be placed after each sentence containing a picture’s word. After the picture-naming step, the confederate had to argue the choices of words. During this justification step, the words associated with each picture were repeated a second time and the picture was described using its geometrical shapes and global shape. The goal of the description was to justify the relationship between the picture and its given word. The descriptions included none of the words used to name the tangrams. After the justification step, the confederate had to ask the participants whether they agreed or not with the suggested words. This agreement step allowed the confederate to repeat a third and a fourth time the words associated to the pictures that is, for the agreement and then when the confederate confirmed that the words have been written on the response sheet. Therefore, the repetitions of words referring to a given picture were spaced through the different turn-takings during the three steps. Importantly, when pairs of words were proposed for a given picture, the order of presentation of the words of the same pair was alternated along the script.

Participants were randomly assigned to one of the four versions of the IAR task before their arrival. Two experimental lists made it possible to present the word pairs either in the related condition or in the unrelated condition and there were two different orders of presentation for the 20 trials for which the confederate had to ask to propose the words to refer to the tangram pictures. Two different versions were thus obtained for each experimental list. Although the order of these trials varied, the filler trials in which the participant described the pictures to the confederate were the same across all four versions of the task. Each version of the IAR task was associated with one script and the structure of the four scripts was always the same, and only the picture's word and its descriptions changed. The sequences of sentences were then adapted in order to make them as natural as possible.

In each trial in which the referential choices were made by the confederate, the participants were exposed to three pictures: two pictures from set 1 (each picture was thus associated with two words) and one picture from set 2 (this picture was associated with just one word; see Figure 5). Importantly, 4 trials among the 20 did not follow these rules to prevent the participants from finding regularities in the composition of the trials. Two trials were thus composed of three pictures associated with only one word and two trials were composed of two pictures associated with two pairs of words. Additionally, each trial was composed of pictures of various shapes (rather square, long or height, straight or curved), complexity, and the category of objects represented by the corresponding words (either living or inanimate object). The composition of the filler trials also varied in terms of shape and complexity of the pictures.

< Insert Figure 5 here >

The recordings of the dialogues were checked to make sure that words were correctly repeated four times by the confederate. The mean number of repetitions of the words referring to pictures by the confederate approached four (M = 3.94, SD = 0.34), confirming that the required number of repetitions were produced by the confederate. The number of disagreements given by the confederate on the words proposed by the participant was kept constant across participants. Specifically, the confederate had to disagree on one trial per block, but was able to choose the trial to disagree on to match with the relevance of the participant's justifications, in order to keep the interaction as naturalistic as possible. The mean percentage of disagreement given by the participants to the words proposed by the confederate was 8.63 % (*SD* = 7.04), showing that the participants easily reached an agreement with the confederate.

Lexical decision task (LDT)

After the IAR task, the participants performed an individual lexical decision task with a semantic priming paradigm. Stimuli were presented using the E-Prime software (E-Prime 2.0.10.356, Schneider et al., 2002a; 2002b). Participants first performed 16 practice trials before proceeding with the 160 experimental trials. They were divided into two blocks of 80 trials and a break between the blocks was proposed to the participant. Stimuli were displayed on the screen in lower case and in white font against a black background. Each trial consisted in the presentation of a fixation cross for 500 ms, followed by a prime word for 150 ms. A black screen was then displayed for 16 ms (leading to a SOA of 166 ms) before the presentation of the target stimulus, which remained on the screen until the participant's response. Participants had to indicate as quickly and accurately as possible whether or not the target stimulus was a real word. They had to respond by using a button box for which the assignment of the button responses depended on their handedness. At the end of each trial, a black screen was displayed for 1,500 ms before the next trial began. The task took approximately 15 minutes to be completed.

Session 2

Filler task and questionnaires

Participants came back the following day to perform a filler task and fill in personality questionnaires. The filler task was used to keep constant the timing of each session across Experiments 1 and 2 (2h on day 1 and 30 min on day 2). It consisted in an online citizen-science experiment designed by researchers at Yale University, called the Tone Deafness Test (see: https://www.themusiclab.org/quizzes/td). In this task, participants had to listen to sounds of different pitches and indicate whether the last sound went up or down. They were also asked questions regarding their experiences with music.

Regarding the questionnaires, participants were asked to fill in the IRI questionnaire (Braun et al., 2015) as a measure of cognitive and affective empathy, and the PSIQ-F questionnaire (Ceschi & Pictet, 2018) as a measure of mental imagery. Both questionnaires were implemented on LimeSurvey 2.06. At the end of the questionnaires, participants answered two additional questions in order to determine if they were aware of the presence of a confederate. None of the participants spontaneously reported being aware of the presence of a confederate in the first question. In the second question, none of the participants reported being aware of the confederate due to the dialogue feeling unnatural.

Data analysis

Experimental design and data pre-processing

Reaction times from the lexical decision task were analyzed with linear mixed-effects models using lme4 package in R (Baayen et al., 2008). The dependent variable was the participants' reaction times on the lexical decision task. The independent variable, called the relatedness factor, was whether each pair of words shown in the lexical decision task was made of words which were both used to refer to the same picture (i.e., related condition) or which were used to refer to different pictures (i.e., unrelated condition). As in Barr et al. (2013), we started by implementing the maximal random effect structure justified by the design. This involved by-participants random intercepts, by-participants random slopes corresponding to relatedness, by-item random intercepts and by-item random slopes corresponding to the relatedness factor (note that in the present study, items corresponding to the target words of each pair on which participants made their lexical decision). Unfortunately, models including this maximal random effect structure systematically failed to converge. We thus attempted to simplify the structure of the models in order to improve model convergence, and we found that the only model to converge was a model including only by-participant and by-item random intercepts. The fixed effect remained as planned with the relatedness of the pair of words during the dialogue which was divided into two experimental conditions (related, or unrelated condition). The results from the main model are reported in Tables 2 and 3.

< Insert Tables 2 and 3 here >

Two additional analyses were conducted to investigate the modulation effects of personality traits on the changes in lexico-semantic representations after the dialogue. These models both included by-participant and by-item random intercepts, following the same rationale as the main model. The first model included both the relatedness factor and the centered data for each dimension of the IRI questionnaire as fixed effects (Personal Distress, Perspective Taking, Empathic Concern and Fantasy). Centering the data for each dimension allowed to reduce the risk of multicolinearity. The second model also included the relatedness factor as fixed effect, as well as centered data for each dimension of the PSIQ-F as fixed effects (Vision, Sound, Smell, Taste, Touch, Body and Emotion). The results from these additional models are reported in Appendix A.

Finally, cut-offs were used to exclude outliers. These cut-offs were the same for the two experiments that relied on reaction times (i.e., Experiments 1 and 2). First, the data from 5 participants were excluded from the analyses due to a total response accuracy below 90% in order to keep participants with a number of measure points as close as possible to 20 trials per experimental condition for the analysis of reaction times (see also the conclusion from Diependaele et al., 2012). Data from another participant were also excluded due to a missing recording of the dialogue. Finally, the data from 40 participants were analyzed. This sample size is in line with previous studies that used the lexical decision task as a measure of the relatedness between pairs of written words (Armstrong & Plaut, 2008; Brunellière et al., 2017; Shelton & Martin, 1992).

Regarding the reaction times of the 40 remaining participants, incorrect responses (0.69%) and decision latencies longer than 1503 ms (5 standard deviation above the grand mean, for a similar approach, see Baayen et al., 2003) were excluded from data analysis (0.56%). The mean error rates for the related and unrelated conditions were 0.88% and 0.50% respectively. As our hypotheses did not concern error rates and as they were below the 5% threshold, no statistical analysis was conducted on them.

Results

The results of Experiment 1 are shown in Figure 6. The analysis of lexical decision latencies did not reveal any significant difference between the pairs of words referring to the same picture during the dialogue (*M* = 619 ms, *SD* = 150 ms) and those referring to different pictures during the dialogue (*M* = 617 ms, *SD* = 150 ms), *b* = 2.67, *SE* = 6.37, χ^2 (1) = 0.18, *p* = .68. However, a significant interaction between the relatedness factor and one dimension of the IRI questionnaire was found in the additional model that included the four dimensions of the IRI questionnaire. The personal distress subscale of the IRI questionnaire interacted significantly with the relatedness factor, *b* = -4.85, *SE* = 1.99, χ^2 (1) = 5.96, *p* = .015. The *b* value suggests that the significant interaction between the relatedness and the personal distress factors was because when the personal distress scores increased, the reaction times to targets in pairs of words referring to the same picture were faster, while the reaction times to targets in pairs of words referring to different pictures during the dialogue were longer (see, Figure 6). Finally, the last model that included the data of the PSIQ-F questionnaire did not show any significant interaction between the relatedness factor and the seven dimensions of the PSIQ-F questionnaire.

< Insert Figure 6 here >

Discussion

In the present experiment, we investigated whether new semantic relationships created in dialogue changed the preexisting lexico-semantic representations in semantic memory immediately after the dialogue. Participants took part in a novel task (IAR task) during which they discussed tangram pictures with a confederate to reach an agreement on the best word to give to each picture. Crucially, the confederate had to refer to each picture using two semantically unrelated words, thus creating a new semantic relationship between the words mapped with the same visual representation. Immediately after the dialogue, participants performed a lexical decision task in a semantic priming paradigm, in which we expected to find a relatedness effect with faster reaction times with pairs related referring to the same picture during the dialogue than with the pairs referring to the two different pictures. However, the analysis of the data in the lexical decision task did not reveal any significant relatedness effect. The only significant result reported in this study was an interaction between the related and unrelated word pairs during the dialogue and the personal distress subscale from the IRI questionnaire. Personal distress corresponds to a feeling of unease or distress when seeing someone else's physical or emotional pain and this is defined as playing a role of affective empathy. Wagner et al. (2015) demonstrated that people with higher scores of personal distress have better memories of socially encoded information (related to another person) than of non-socially encoded information (self-related), while people with lower scores of personal distress show the opposite pattern. In line with this study, our results demonstrate that people with a higher level of personal distress exhibited faster word recognition when the word was preceded by another word referring to the same picture during the dialogue. This was particularly interesting to observe this pattern of results after the dialogue, since pairs of words referring to the same picture was provided by another person during the dialogue. However, it is difficult to interpret this result in terms of the changes in preexisting lexico-semantic representations, since the processing of pairs of words referring to different pictures produced the reverse pattern to that of pairs of words referring to the same picture. More precisely, when a word was preceded by another word referring to different pictures during the dialogue, its recognition was longer in people with a higher personal distress. Therefore, people can take into account the new semantic relationship and the words referring to tangram pictures according to their social abilities to recognize words presented on a screen.

However, the results of Experiment 1 do not bring a clear support to our working hypothesis and there was no experimental evidence of the changes in preexisting lexico-semantic representations stored in semantic memory immediately after the dialogue. The lack of a significant relatedness effect may be explained by various factors. First, it is possible that the new semantic relationship may need to be consolidated by a night of sleep before it can be integrated into the lexico-semantic network and producing any propagation of activation from one word to the other (Dumay & Gaskell, 2012). Second, as mentioned in the introduction, the strength of new connections may not be integrated strongly enough into the lexico-semantic network to be accessed via an implicit task such as the lexical decision task. These two possible explanations were tested in the following experiments. Specifically, in the following experiment 2, we explored whether the changes in preexisting lexico-semantic representations were observed one day after the dialogue, following a night of consolidation.

EXPERIMENT 2

Experiment 2 was similar to Experiment 1 except that participants performed the lexical decision task one day after the IAR task. The aim of this experiment was to assess the changes in preexisting lexico-semantic representations stored in semantic memory one day after the dialogue. (i.e., after one night of consolidation) due to the creation of a new semantic relationship during the dialogue. The hypotheses were identical to those of Experiment 1.

Materials and methods

Participants

Forty-two native French speakers (37 females; 40 undergraduates and two community members; age, *M* = 22.26, *SD* = 3.64) who did not participate in Experiment 1 were paid 15€ to take part in Experiment 2. Two participants were excluded from the analysis of data (for more details, see the experimental design and data pre-processing section). As in Experiment 1, participants interacted with one of the two same confederates. More precisely, 20 participants interacted with one of the confederates while 20 participants interacted with the other one.

Stimuli and procedure

The stimuli and tasks were the same than those used in Experiment 1. As in Experiment 1, the mean number of repetitions of the words produced by the confederate approached 4 (*M* = 3.98, *SD* = 0.24), confirming that the required number of repetitions were produced by the confederate. The mean percentage of disagreement given by the participants to the words proposed by the confederate was relatively low with 5.67 % of disagreement (*SD* = 4.60), showing the participants easily reached an agreement with the confederate. As in experiment 1, the confederate had to disagree in one trial per block, but was able to choose on which trial to disagree with their partner keep the interaction as naturalistic as possible.

Contrary to Experiment 1, the IAR task was followed by the Tone Deafness Test to keep the timing of session 1 constant in Experiment 2 (see Figure 4). During the session 2, which took place the next day, participants performed the lexical decision task with the semantic priming paradigm one day after the IAR task. After this task, participants filled in the same questionnaires (IRI questionnaire and PSIQ-F questionnaire) as those used in Experiment 1. None of the participants spontaneously reported being aware of the presence of a confederate.

Data analysis and preprocessing

The same approach as in Experiment 1 was used to build the statistical models. Applying this rationale led to the same set of fixed and random effects as in Experiment 1. The results from the main model are reported in Tables 4 and 5 and the results of additional models can be found in Appendix B. In addition, the same cut-offs than those used in experiment 1 were used to exclude data. Thus, the data from two participants were excluded from the analyses due to a total response accuracy below 90%. Regarding the reaction times of the 40 remaining participants, incorrect responses (0.83%) and decision latencies longer than 1513 ms (5 SD above the grand mean) were excluded (0.50%). The mean error rates for the related and unrelated conditions were 1.19% and 0.48% respectively. As our hypotheses did not concern error rates and as they were below the 5% threshold, no statistical analysis was conducted on them.

< Insert Tables 4 and 5 here >

Results

The results of Experiment 2 are shown in Figure 6. As in Experiment 1, there was no significant difference between pairs of words referring to the same picture during the dialogue (M = 615 ms, SD = 152 ms) and those referring to different pictures during the dialogue (M = 615 ms, SD = 164 ms), b = -4.72, SE = 6.55, $\chi^2(1) = 0.52$, p = .47. Regarding the models that included either the IRI or the PSIQ-F questionnaire, no interaction effect was found with the relatedness factor on the lexical decision latencies.

An additional analysis was conducted in order to investigate further the effect of a night of sleep on potential changes in the participants' lexico-semantic network. The detail of this analysis is provided in Appendix B. In this analysis, we included the time of day at which each participant performed the first session of the experiment as an additional fixed effect in the model. Just like in the main analysis, the model also included relatedness as a fixed effect (as well as the interaction between relatedness and participation time); the outcome variable was the participant's reaction time on the lexical decision task. Indeed, dyads performing the first session at different times of the day meant that their awakening time before going to sleep on the night following the first session varied across dyads. However, no significant effect of participation time was found, nor did this factor interact with relatedness; these results are therefore not discussed further in the remainder of this paper.

Discussion

In Experiment 2, we explored whether preexisting lexico-semantic representations stored in semantic memory changed one day after the dialogue thanks to the phenomenon of consolidation generated during a night of sleep. As in Experiment 1, participants performed an IAR task during which pairs of words were either referring to the same tangram picture or to two different tangram pictures. One day later, participants performed the same lexical decision task with semantic priming paradigm. The analysis of the data in the lexical decision task did not reveal any significant relatedness effect nor any interaction between relatedness factor and the various subscales of the IRI and PSIQ-F questionnaires. From these results, we cannot conclude about any consolidation of changes in preexisting lexico-semantic representations after one night of sleep.

To draw clear conclusions regarding what happens in the lexico-semantic network after the dialogue, we conducted a third experiment in which the lexical decision task was replaced by a semantic relatedness judgment task. We sought to address the nature of the changes into the lexico-semantic network by using an explicit task on the degree of relatedness. Contrary to an implicit task such as a lexical decision task which would ensure access to stronger connections into the lexico-semantic network, the use of the semantic relatedness judgment task would amplify the activation of the lexico-semantic network and help to track weaker connections into the lexico-semantic network.

EXPERIMENT 3

Experiment 3 followed a similar procedure to that of Experiment 2, except that the visual lexical decision task was replaced by a visual semantic relatedness judgment task, during which participants must explicitly pay attention to the degree of relatedness between pairs of written words. We explored whether changes in preexisting lexico-semantic representations in semantic memory were present when attentional processes could amplify the degree of activation into the lexico-semantic network. One day after the dialogue, we expected higher scores of semantic relatedness for pairs of words referring to the same tangram picture during the dialogue than for pairs of words referring to two different tangram pictures during the dialogue. Crucially, a control group was also added in Experiment 3 to provide clear evidence that dialogue directly changes preexisting lexico-semantic representations in semantic memory. This group was not exposed to the dialogue. A significant interaction between relatedness (related vs. unrelated condition) and group (group exposed to the dialogue vs. control group) was therefore expected to be found. Contrary to the group exposed to the dialogue (i.e., experimental group), we expected that the scores of semantic relatedness for pairs of words referring to the same tangram picture did not differ with those referring to two different tangram pictures in the control group. In addition, both groups were asked to perform a word-picture relatedness judgment task as a control task to ensure that the participants had learnt the associations between each of the two words used to name the same picture during the dialogue (i.e., the IAR task) and their associated picture.

Materials and methods

Participants

Forty French native speaking students (all undergraduates; 33 female; age, $M = 20.50$, $SD = 2.43$) were paid 15€ to be a part of the experimental group (i.e., group exposed to the dialogue). The overall recruitment procedure was the same as for Experiments 1 and 2. As in Experiments 1 and 2, participants interacted with one of the two same confederates. More precisely, 21 participants interacted with one of the confederates while 19 participants interacted with the other one. One participant was excluded from the analysis of data (for more details, see the experimental design and data pre-processing section). For the control group, forty native French speakers (33 female; age, *M* = 21.15, *SD* = 2.25) were recruited via prolific and social media. Participants in the control group, who performed only the semantic relatedness judgment and word-picture relatedness judgment tasks, were paid 5€ for their participation. The rest of the recruitment procedure was the same as for Experiments 1 and 2.

Stimuli

The stimuli in the IAR task were the same as those in Experiments 1 and 2. The 20 pairs of words referring to the same tangram picture and the 20 pairs of pairs of words referring to different tangram pictures during the dialogue were presented in the semantic relatedness judgment task. Twenty highly semantically related pairs of words were also selected from the stimuli of Brunellière and Bonnotte (2018) to make sure that the participants would use all the points on the scale of the semantic relatedness judgment task. Their mean semantic relatedness score was 6.47 on a 7-point scale (*SD* = 0.58) from the study of Brunellière and Bonnotte (2018). The mean lexical frequency and mean number of letters of the 20 highly semantically related pairs was matched with those of the two other experimental conditions across the experimental lists (see Table 6 for more details on the psycholinguistic properties). Thus, in total, 60 trials were performed in the semantic relatedness judgment task, with 20 trials in each critical condition (related vs. unrelated condition) and 20 highly semantically related pairs of words.

< Insert Table 6 here >

Two sets of 40 word-picture pairs were used in the word-picture relatedness judgment task. The aim of this task was to check that participants in the experimental group had learnt the associations between each of the two words used to name the same picture during the IAR task and their associated pictures, in comparison with participants in the control group who did not perform the IAR task. The first set of word-picture pairs consisted of the 20 picture exposed during the dialogue as stimuli in the related condition of the IAR task. These 20 pictures were presented twice, each time using one of the two words with which the picture had been associated during the dialogue (i.e. in total, 40 word-pictures pairs). As in the first set, the pictures were presented twice in the second set and these word-picture pairs were created by combining 20 other pictures exposed during the dialogue with 40 new words. In other words, these pairs were never presented as words referring to pictures during the dialogue. Since half of the new words were words which had not been given by the participants enrolled in our database of pictures, these word-picture pairs were not related. The other half of new words were words which had been given by only one participant among the 84 respondents of the new database so that the word remained a plausible word to name the picture. The second group of word-picture pairs was created to make sure that participants could use the scale of the word-picture relatedness judgment task accurately. The 40 words of the first group were matched with those of the second group in terms of lexical frequency and mean number of letters (see Table 7 for more details on the psycholinguistic properties).

< Insert Table 7 here >

Design

<u>Experimental lists</u>

The distribution of the 40 pairs of words associated with tangram pictures (i.e., 20 pairs of words referring to the same tangram picture and 20 pairs of words referring to two different tangram pictures) into two lists was kept constant across the IAR and the semantic relatedness judgment tasks. The 20 highly semantically related pairs of words were also added in both experimental lists of the semantic relatedness judgment task.

Regarding the word-picture relatedness judgment task, each list was composed of the 20 pairs of words referring to the same picture and these pictures were presented twice using one of the two words. In each list, the second set of the forty word-picture pairs was added.

Procedure

Experimental Group

Session 1

The procedure for day 1 was the same as in Experiment 2, with participants performing the IAR task first and then the Tone Deafness Test. Dialogues recordings were checked to make sure that words were repeated enough times by the confederate. The mean number of repetitions of the words produced by the confederate approached 4 (*M* = 3.99, *SD* = 0.17), confirming the required number of repetitions was produced by the confederate despite the naturalistic setting of the task. Regarding the level of disagreement provided by the participant on the words proposed by the confederate, the mean percentage was 5.92% (*SD* = 4.67). As in Experiments 1 and 2, this value was low, showing that an agreement was reached easily. Disagreement rules followed by the confederate were identical to Experiments 1 and 2.

Session 2

On day 2, participants came back to perform the semantic relatedness judgment task and the word-picture relatedness judgment task respectively. Both were presented on LimeSurvey 2.06. Participants first received written instructions about the semantic relatedness judgment task and two examples describing the use of the Likert scale. Two training trials were then proposed before proceeding with the 60 experimental trials presented in random order. In each trial, pairs of words were presented one at a time. Words were displayed side by side on a white background and the 7-point Likert scale was positioned below them. Participants were asked to rate the degree of semantic relatedness for each pair of words before going on the next trial. At the end of the task, they were informed that they would have to do a second task, corresponding to the word-picture relatedness judgment task. This task followed the same structure as the semantic relatedness judgment task with the presentation of written instructions, two examples, and two training trials after which participants performed 80 experimental trials. In each trial, the picture and the word were presented one above the other on a white background. The 7-point Likert scale was placed above the word-picture pair and participants were asked to rate the degree of relatedness between the picture and the word. The order of completion of both tasks was not counterbalanced. The reason was because the main question concerned the new semantic relationship between the pairs of words by using the semantic relatedness judgment task, and the second task was only a control one. Moreover, we preferred to keep the semantic relatedness judgment task first to prevent any risk that the word-picture relatedness judgment task might influence the semantic relatedness judgment results between pairs of words. The completion of both tasks took approximately 15 minutes and was followed by the completion of the IRI and PSIQ-F questionnaires (Braun et al., 2015; Ceschi & Pictet, 2018). None of the participants spontaneously reported being aware of the presence of a confederate.

Control Group

Participants in the control group completed the semantic relatedness judgment task between the pairs of words, the word-picture relatedness judgement task and the two questionnaires (IRI and PSIQ-F) via an online survey. The survey was supported by LimeSurvey 2.06 and followed the same structure as that used for the experimental group.

Data analysis

Experimental design and data pre-processing

Data from the semantic relatedness judgment task were analyzed in two steps. We first checked that participants from both the experimental and the control groups used the whole scale to rate the three types of word pairs. Both groups rated the highly semantically related pairs of words as accurately highly semantically related (experimental group, $M = 5.99$, $SD = 1.33$, control group, $M = 5.82$, $SD = 1.43$). After removing the data from the words that were not presented in the IAR task (as these were not of interest in the study), we ran a linear mixed-effects model using the lme4 package in R (Baayen et al., 2008). Random effects included by-participants random intercepts, by-participants random slopes corresponding to relatedness, by-item random intercepts and by-item random slopes corresponding to the relatedness factor and group. By-item intercepts were represented by the target words of each pair. The dependent variable was the centered participants' scores on the scale of the semantic relatedness judgment task between pairs of words. There were two independent variables with the relatedness factor (as in Experiments and 2) and the group. Therefore, the fixed effects were the relatedness of the pairs of words during the dialogue, either related or unrelated, the group, experimental vs. control, and the interaction between these two factors. In line with our specific hypotheses on each group, Bonferroni-corrected pairwise comparisons were performed. The results from this main model are reported in Tables 8 and 9. Two additional models were created to investigate the modulation effects of personality traits on the ratings from the semantic relatedness judgment task. These models both included by-participant and by-item random intercepts. The first model included both the relatedness of the pairs of words, the group and the centered data for each dimension of the IRI questionnaire as fixed effects (Personal Distress, Perspective Taking, Empathic Concern and Fantasy). The second model also included the relatedness of the pair of words as fixed effect, as well as centered data for each dimension of the

PSIQ-F as fixed effects (Vision, Sound, Smell, Taste, Touch, Body and Emotion). The results of additional models can be found in Appendix C.

< Insert Tables 8 and 9 here >

Data from the word-picture relatedness judgment task were analyzed to ensure that words referring to the same picture had been learnt correctly with their associated pictures. We used a linear mixed-effects model, performed with the lme4 package in R (Baayen et al., 2008). Random effects included intercepts of participants and items, represented by the words of each word-picture pair (Barr et al., 2013), by-participants random slopes corresponding to word-picture relatedness and by-item random slopes corresponding to group. The dependent variable was the centered participants' scores on the scale of the word-picture relatedness judgment task. The fixed effects were the two factors, the word-picture relatedness, either word referring or not to a picture during the dialogue, and group, experimental vs. control. The results from this model are reported in Tables 10 and 11. No further analyses were performed as the word-picture relatedness judgment task was only a control task.

Importantly, although the participants performed the word-picture relatedness judgment task after the semantic relatedness judgment task, we report the results from the word-picture relatedness judgment task first in the following section. Indeed, this task was simply a control. We then report the results from the semantic relatedness judgment task, as these enable us to test our hypothesis directly.

< Insert Tables 10 and 11 here >

Results

The mixed model used to analyze the word-picture relatedness scores revealed a significant interaction between the word-picture relatedness and the group of participants, b = -0.76, SE = 0.17, χ^2 (1) = 19.86, p < .001). The b value suggests that this significant interaction was because the word-picture relatedness scores (Table 12) were higher for word-picture pairs were presented during the dialogue (i.e., stimuli presented as word pairs referring to the same picture) in the experimental group than in the control group. This difference between the two groups of participants was stronger for word-picture pairs presented during the dialogue (i.e., stimuli presented as word pairs referring to the same picture) in comparison with that for word-picture pairs not presented during the dialogue.

< Insert Table 12 here >

Crucially to the objective of the present study, the analysis of the semantic relatedness judgment scores between word of pairs showed a significant interaction between the relatedness factor (related vs. unrelated condition) and the group of participants (either included in the experimental group or in the control group), b = 0.50, SE = 0.14, χ^2(1) = 12.63, p < 2.2e-16). This interaction was illustrated in Figure 7. The pairwise comparisons revealed that as expected, the scores of semantic relatedness for pairs of words referring to the same tangram picture during the dialogue were higher than for those referring to two different tangram pictures during the dialogue in the experimental group (b = -0.38, SE = 0.11, t(71.9) = -3.52, p = .0046). In contrast, there was no significant effect of the relatedness factor in the control group (b = 0.10, SE = 0.11, t(71.9) = 0.97, p = 1.00). Finally, the two additional models created to investigate the modulation effects of personality traits showed no significant interaction between relatedness, groups of participants and the one of the two questionnaires (IRI or PSIQ-F).

As in Experiment 2, an additional analysis including participation time was conducted. This analysis of the semantic relatedness judgment scores is reported in Appendix B. As in Experiment 2, no significant effect of participation time was found; these results are therefore not discussed further in the remainder of this paper.

< Insert Figure 7 here >

Discussion

The aim of Experiment 3 was to explore whether changes in preexisting lexico-semantic representations in semantic memory occurred after one night of consolidation when attentional processes could amplify the degree of activation into the lexico-semantic network by using an explicit task focusing attention on semantic relationships between pairs of words. To do so, participants performed the same IAR task as in Experiments 1 and 2, and returned one day later to perform a visual semantic relatedness judgment task between pairs of words. Interestingly, the additional word-picture relatedness judgment task revealed that participants had correctly learned the word-picture associations between each of the two words used to name the same picture during the IAR task and their associated pictures, confirming that they payed attention to the two pairs of words referring the same picture during the dialogue and they encoded the new mapping between a word and an object used in a dialogue. Crucially to the objective of the present study, participants who took part in the dialogue rated the pairs of words associated with the same picture during the dialogue as more semantically related than the pairs of words associated with two different pictures. In contrast, participants who did not take part in the dialogue rated both types of pairs the same way. Altogether, these results suggest that preexisting lexico-semantic representations were changed after one night of consolidation and the persistence of such changes in semantic memory due to the dialogue was found when attention was focused on the degree of semantic relatedness between pairs of words.

General discussion

The aim of the present study was to examine to what extent new semantic relationships created in dialogue changed preexisting representations stored in semantic memory after the dialogue. In three experiments, we asked participants to take part in a new experimental paradigm in dialogue, the interactive agreement referential task (IAR), during which partners interacting via an audio-conference device came to an agreement on how to name tangram pictures presented on their screens. We took advantage of new word-visual representation mappings during the dialogue to create new semantic relationships between pairs of words. The role of the confederate was to name each picture using either one word or two words, and to justify the relationship between each word and its corresponding picture. By doing so, two types of pairs of words were created: half of them were related to the same visual representation during the dialogue, and the other half were used to describe different pictures during the dialogue. We hypothesized that the creation of a new semantic relationship between pairs of words used as words during the dialogue could change in preexisting lexico-semantic representations associated with these words after the dialogue. In Experiments 1 and 2, we investigated whether changes in preexisting lexico-semantic representations stored in semantic memory were associated with the establishment of strong connections into the lexico-semantic network by using a visual lexical decision task coupled with a semantic priming paradigm. Although Experiment 1 was conducted to examine changes in preexisting lexico-semantic representations in semantic memory immediately after the dialogue, Experiment 2 was more in line with the main objective of the study by exploring the persistence in the changes of preexisting lexico-semantic representations stored in semantic memory could persist one day after the dialogue. More particularly, we investigated the persistence of such changes at the representation level one day after the dialogue and the nature of these changes into the lexico-semantic network by using implicit or explicit tasks on the degree of semantic relatedness. Experiment 3 was created to assess whether changes in preexisting lexico-semantic representations in semantic memory one day after the dialogue were related to new weak yet reachable connections under the action of attentional processes by amplifying the degree of activation into the network.

The results from the three experiments partially support our hypothesis that dialogue can affect people's preexisting representations stored in semantic memory. Indeed, the results of Experiment 3 showed a clear pattern of relatedness between the pairs of words that were associated with the same picture during the dialogue. Crucially, the relatedness effect evidenced in Experiment 3 was present only in the experimental group who did took part in the dialogue. These participants in the semantic relatedness judgment task rated the newly related words as more semantically related. Such findings support the idea that new semantic relationships created during the dialogue changed preexisting lexico-semantic representations stored in semantic memory after a night of sleep. The findings of Experiments 1 and 2 with a semantic priming paradigm contrasted with this view, as they showed no clear impact of the dialogue on preexisting lexico-semantic representations stored in semantic memory. Taken together, our findings shed light on the importance of the attention focused on the degree of semantic relatedness and the nature of changes into the lexico-semantic network. The implications of these results are discussed in the following section, first in relation to the literature on the access to long-term semantic representations and then in relation to the field of dialogue research.

Long-term semantic representations

The current study is the first to show that dialogue can impact the organization of the lexico-semantic network one day after the interaction. Various methodologies have demonstrated that the more word-meanings are encountered, the more they are reinforced into the lexico-semantic network (Betts et al., 2018; Rodd et al., 2012; 2016). Importantly, Betts and collaborators pointed out that for a repetition to be useful, the information had to be spaced over time instead of massed. For instance, they showed a boost of changes in lexico-semantic representations after 3 repetitions when their presentation of sentences was spaced. In line with this study, in our work, we characterized the spacing in terms of dialogue turns between each repetition, leading to four repetitions. Specifically, participants were prompted to speak after each repetition so that the following repetition would occur at an interval dictated by turn-taking. The length of turn-taking between partners could however not be controlled in our work. The impact of such spacing will have to be explored in further studies in order to confirm its role in changes in preexisting lexico-semantic representations.

More generally, most of the studies regarding the flexibility of lexico-semantic representations have been conducted in language comprehension (Betts et al., 2018; Eligio & Kaschak, 2020; Hulme et al., 2019; Fang & Perfetti, 2019; Rodd et al., 2012; 2013; 2016). Central to our work is the idea that people have the opportunity to create new mappings between words and meanings as they interact during dialogue. Therefore, dialogue constitutes an ideal setting for the integration of new uses for words and new meanings into semantic memory. Our dialogic experimental paradigm allowed for the abstraction of the new semantic relationship between pairs of words referring to the same visual representation by varying the order of presentation of each word of the pair referring to the same visual representation. Importantly, the modality through which these words were conveyed during the dialogue (i.e., the auditory modality) was different from the one in which the access to the lexico-semantic representations was assessed after the dialogue (i.e., written modality). Our findings thus shed light on the importance of taking into account semantic memory in the links between dialogue and memory. However, this finding raises the question of whether such changes occur in the same way as word-meaning mapping learning in language comprehension. Further studies using dialogue and comprehension settings are thus needed to explore their respective impact on changes in lexico-semantic representations stored in semantic memory.

Even if our findings demonstrated that changes in preexisting lexico-semantic representations persist one day after the dialogue (as shown in Experiment 3), the exact nature of what has changed in semantic memory must be questioned. While evidence of changes in preexisting lexico-semantic representations in semantic memory with an implicit task such a lexical decision task would ensure the access to stronger connections into the lexico-semantic network, evidence of changes in preexisting lexico-semantic representations under the action of attentional processes suggests the formation of new weak connections into the lexico-semantic network. Top-down attention has been described as playing a role on the activation of networks of interconnected nodes by amplifying their level of activation in order to prioritize specific mental processes (Dehaene et al., 2006; Narhi-Martinez et al., 2023). In line with these views describing the impact of top-down attention, de Wit and Kinoshita (2014, 2015) demonstrated that depending on the task used, the same prime-target pairs could lead to significant priming effect or not when the primes were masked or unmasked. They explained their results in terms of type of activation that is produced in the semantic network, depending on whether participants have to explicitly check for semantic properties of the prime. While the information extracted from the prime in a lexical decision task is of little relevance, the checking on the overlap of semantic features between the prime and the target is very crucial in a semantic categorization task. In the context of our own study, the lexico-semantic relationship created during the IAR task can be characterized as a new overlap between the visual semantic features between the prime and the target. The observation of changes in preexisting lexico-semantic representations one day after the dialogue when the attention was focused on the degree of semantic relatedness suggest that there were weak connections were built into the lexico-semantic network. Similar to previous studies exploring the learning of new meanings on preexisting lexico-semantic representations (Hulme et al., 2019; Rodd et al., 2012), the learning effect is most often evidenced with the use of explicit recall of meaning such as cued recall tests.

The persistence of changes in lexico-semantic representations one day after the dialogue is in line with the results of Rodd et al. (2016) and those of Gaskell et al. (2019) who explained their results by referring to a connectionist network model of ambiguous word representation and processing. They proposed an interpretation in terms of weight of the connections between words and meanings. According to this approach, each encounter of a word-meaning mapping creates an activation in the lexico-semantic network that changes the weight of the connections after a night of consolidation, particularly that of the recently activated connections. By this account, the encounter of the new relationship between the pairs of words during the dialogue is thought to produce new connections between both words into the lexico-semantic network. After a night of consolidation, the new connections between the pairs of words take on a certain weight. In the current study, the fact that participants encountered this new relationship few times probably could explain why this weight was too weak to produce sufficient activation to be detectable with a semantic priming paradigm. However, in a real life setting, words and meanings are encountered multiple times, and they may be encountered in different conversations taking place on different days, giving raise to changes in the weights between connections into the lexico-semantic network.

Dialogue and memory

Studying the interface between dialogue and memory is not new (Brennan & Hanna, 2009; Horton & Gerrig, 2005; 2016; Knutsen & Le Bigot, 2014; McKinley et al., 2017; Nault et al., 2023). However, the way in which the two processes interact remains unclear in the literature. Although we know that dialogue partners tend to change their language habits to establish conceptual pacts (Brennan & Clark, 1996), adjust utterances to specific dialogue partners and maximize mutual understanding (Clark, 1996; Clark & Schaefer, 1989), little is known about potential changes at representational level after the dialogue. The only attempt to conceptualize the changes in preexisting linguistic representations due to the dialogue is the alignment model by Pickering & Garrod (2004, 2021). In this model, the same linguistic representations become activated in both partners during the interaction, as they become increasingly aligned. In contrast, the current study did not seek to examine changes in the level of activation of people's representations during the dialogue, but rather to address how new semantic relationship created during a dialogue may interact with preexisting knowledge, causing changes in lexico-semantic representations stored in semantic memory.

Our findings have important implications for dialogue research. In particular, they help shed light on what people actually do when they build shared knowledge as they interact. According to the collaborative approach to dialogue (Brennan & Clark, 1991; Clark, 1996; Clark & Schaefer, 1987; Clark & Wilkes-Gibbs, 1986), when two people interact, the information they mention is added to their common ground, which includes the knowledge they share and are aware of sharing (Clark & Marshall, 1981). Dialogue participants may then resort to their common ground to adjust the content of their utterances to each other's dialogic needs, thus contributing directly to mutual understanding. Our findings suggest that the impact of dialogue on interlocutors is not limited to the establishment of shared knowledge and that dialogue also causes direct modifications in people's preexisting representations in semantic memory. In the current study, such changes were only examined in one dialogue partner (that is, the participant, as the other dialogue partner was a confederate). However, in experimental dialogue settings involving two naïve participants, or in everyday life conversation, such changes are likely to affect both partners, which could imply that dialogue causes both partners' representations to become increasingly similar over the time course of the interaction. One intriguing question is whether such changes in partners' representations depend on their role (speaker vs. listener) during the dialogue. A growing body of research suggests that dialogue partners are subject to a self-production effect, whereby self-produced information is remembered better than partner-produced information during and after the end of a dialogue (e.g., Knutsen et al., 2016, 2017; Knutsen & Le Bigot, 2014; 2020; McKinley et al., 2017; Zormpa et al., 2023). In a similar way, changes in lexico-semantic representations could be stronger in the partner who created new meaning than when the partner only received this information from the other person. This suggestion could not be tested in the current study, and further studies are needed to explore this question. Thus, even if the presence of a confederate aware of the working hypothesis was necessary to ensure enough experimental control in the current study, future studies should focus on more ecological settings and explore the case of dialogues between naïve participants. In addition, such ecological settings may involve goal-oriented dialogue (as in the current study) or not, allowing us to investigate the generalisability of our findings to various kinds of communication settings.

The current study provides new perspectives on the links of dialogue and semantic memory and open up new questions about episodic memory and semantic memory in contact with dialogue. Interestingly, a new mapping between a word and an object used in a dialogue is deemed part of the partners' common ground. For instance, information mentioned during a dialogue between partners A and B is deemed part of A and B's common ground; both A and B can subsequently resort to this information to produce partner-adjusted utterances. This common ground is partner-specific. The framework that we propose is that placing partner-specific information in the common ground during the interaction gives each participant the opportunity to change preexisting representations stored in semantic memory when these representations were used with different usages during the interaction. Because these changes concern representations in semantic memory, participants may then resort to these representations even in the absence of their initial dialogue partner. In other words, although common ground is partner-specific, the changes incurred in each participant's semantic memory are not necessarily partner-specific, and they might affect subsequent interactions regardless of who these take place with. Further studies should be led to explore in which interactive situations linguistic representations can change specifically or not to the partner and at what extent such changes occur during and after the dialogue.

Conclusion

The three experiments reported here demonstrate that new knowledge about how dialogue can affect pre-existing representations by changing the organization of the lexico-semantic network. While such changes are not visible with a lexical decision task combining with a semantic paradigm while the use of semantic relatedness judgment task evidenced changes of lexico-semantic representations one day after the dialogue. This finding suggests that new knowledge added to the common ground of conversation partners changes lexico-semantic representations with new connections into the lexico-semantic network. Nonetheless, attention is required to capture the activation of new connections into semantic memory. This result implies that in real life conversations, when conceptual pacts are reused through multiple interactions with the same partner, the word-meaning mapping could be created and then reinforced, resulting in persistent changes in the lexico-semantic network.

Supplementary materials

The data and R scripts of the three experiments can be found at the following url https://doi.org/10.5281/zenodo.10419752, which includes an example of confederate's script, and the pictures used in three experiments. The new picture database described in this paper is also online available at the same url.

Acknowledgments

This research was supported by a grant awarded by the French National Research Agency (ANR-19-CE28–0006) to Angèle Brunellière. We are very grateful to Laura Boulongne, Emilie Clad, Rémi Lefèvre, Laurent Ott and Carla Ridel for their help in analyzing the productions of confederates in the IAR task and setting up the three experiments. The manuscript was proofread by a native-speaking English copyeditor.

Declaration of interest statement

All authors declare that they have no conflicts of interest.

References

Anderson, J. R. (2013). *The Architecture of Cognition*. Psychology Press. https://doi.org/10.4324/9781315799438

Armstrong, B. C, & Plaut, D. C. (2008). Settling Dynamics in Distributed Networks Explain Task Differences in Semantic Ambiguity Effects: Computational and Behavioral Evidence. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 30. Retrieved from https://escholarship.org/uc/item/6rv2c5hh

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

Baayen, R. H., McQueen, J. M., Dijkstra, T., & Schreuder, R. (2003). Frequency effects in regular inflectional morphology: Revisiting Dutch plurals. *Trends in Linguistics Studies and Monographs*, *151*, 355-390.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Barsalou, L. W. (2008). Grounded Cognition. *Annual Review of Psychology*, *59*(1), 617-645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Betts, H. N., Gilbert, R. A., Cai, Z. G., Okedara, Z. B., & Rodd, J. M. (2018). Retuning of lexical-semantic representations: Repetition and spacing effects in word-meaning priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(7), 1130-1150. https://doi.org/10.1037/xlm0000507

Braun, S., Rosseel, Y., Kempenaers, C., Loas, G., & Linkowski, P. (2015). Self-Report of Empathy : A Shortened French Adaptation of the Interpersonal Reactivity Index (IRI) Using Two Large

Belgian Samples. *Psychological Reports*, *117*(3), 735-753. https://doi.org/10.2466/08.02.PR0.117c23z6

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

Brennan, S. E., & Hanna, J. E. (2009). Partner-Specific Adaptation in Dialog. *Topics in Cognitive Science*, *1*(2), 274-291. https://doi.org/10.1111/j.1756-8765.2009.01019.x

Brunellière, A., & Bonnotte, I. (2018). To what extent does typicality boost semantic priming effects between members of their categories? *Journal of Cognitive Psychology*, *30*(7), 670-688. https://doi.org/10.1080/20445911.2018.1523174

Brunellière, A., Perre, L., Tran, T., & Bonnotte, I. (2017). Co-occurrence frequency evaluated with large language corpora boosts semantic priming effects. *Quarterly Journal of Experimental Psychology*, *70*(9), 1922-1934. https://doi.org/10.1080/17470218.2016.1215479

Ceschi, G., & Pictet, A. (2018). *Imagerie mentale et psychothérapie : Un ouvrage sur la psychopathologie cognitive*. Mardaga.

Clark, H. H. (1996). *Using Language*. Cambridge University Press.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In *Perspectives on socially shared cognition* (p. 127-149). American Psychological Association. https://doi.org/10.1037/10096-006

Clark, H. H., & Marshall, C. R. (1981). *Definite Reference and Mutual Knowledge*.

Clark, H. H., & Murphy, G. L. (1982). Audience Design in Meaning and Reference. In J.-F. Le Ny & W. Kintsch (Eds.), *Advances in Psychology* (Vol. 9, p. 287-299). North-Holland. https://doi.org/10.1016/S0166-4115(09)60059-5

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science*, *13*(2), 259-294. https://doi.org/10.1016/0364-0213(89)90008-6

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

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407-428. https://doi.org/10.1037/0033-295X.82.6.407

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *8*(2), 240-247. https://doi.org/10.1016/S0022-5371(69)80069-1

Dehaene, S., Changeux, J. P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: A testable taxonomy. *Trends in Cognitive Sciences, 10*, 204-211. https://doi.org/10.1016/j.tics.2006.03.007

de Wit, B., & Kinoshita, S. (2014). Relatedness proportion effects in semantic categorization: Reconsidering the automatic spreading activation process. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(6), 1733-1744. https://doi.org/10.1037/xlm0000004

de Wit, B., & Kinoshita, S. (2015). The masked semantic priming effect is task dependent: Reconsidering the automatic spreading activation process. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(4), 1062-1075. https://doi.org/10.1037/xlm0000074

Diependaele K, Brysbaert M and Neri P (2012) How noisy is lexical decision? *Front. Psychology* **3**:348. doi: 10.3389/fpsyg.2012.00348

Dumay, N., & Gareth Gaskell, M. (2012). Overnight lexical consolidation revealed by speech segmentation. *Cognition*, *123*(1), 119-132. https://doi.org/10.1016/j.cognition.2011.12.009

Dumay, N., & Gaskell, M. G. (2007). Sleep-Associated Changes in the Mental Representation of Spoken Words. *Psychological Science*, *18*(1), 35-39. https://doi.org/10.1111/j.1467-9280.2007.01845.x

Eligio, R. B., & Kaschak, M. P. (2020). Gaming experience affects the interpretation of ambiguous words. *PLOS ONE*, *15*(12), e0243512. https://doi.org/10.1371/journal.pone.0243512

Fang, X., & Perfetti, C. A. (2019). Learning new meanings for known words: Perturbation of original meanings and retention of new meanings. *Memory & Cognition*, *47*(1), 130-144. https://doi.org/10.3758/s13421-018-0855-z

Farah, M. J., & McClelland, J. L. (1991). A computational model of semantic memory impairment: Modality specificity and emergent category specificity. *Journal of Experimental Psychology: General*, *120*(4), 339-357. https://doi.org/10.1037/0096-3445.120.4.339

Fasquel, A., Brunellière, A., & Knutsen, D. (2023). A modified procedure for naming 332 pictures and collecting norms: Using tangram pictures in psycholinguistic studies. *Behavior Research Methods*, *55*(5), 2297-2319. https://doi.org/10.3758/s13428-022-01871-y

Gaskell, M. G., Cairney, S. A., & Rodd, J. M. (2019). Contextual priming of word meanings is stabilized over sleep. *Cognition*, *182*, 109-126. https://doi.org/10.1016/j.cognition.2018.09.007

Horton, W. S., & Gerrig, R. J. (2002). Speakers' experiences and audience design: Knowing when and knowing how to adjust utterances to addressees. *Journal of Memory and Language*, *47*(4), 589-606. https://doi.org/10.1016/S0749-596X(02)00019-0

Horton, W. S., & Gerrig, R. J. (2005a). Conversational common ground and memory processes in language production. *Discourse Processes*, *40*, 1-35. https://doi.org/10.1111/tops.12216

Horton, W. S., & Gerrig, R. J. (2005b). The impact of memory demands on audience design during language production. *Cognition*, *96*(2), 127-142. https://doi.org/10.1016/j.cognition.2004.07.001

Horton, W. S., & Gerrig, R. J. (2016). Revisiting the memory-based processing approach to common ground. *Topics in Cognitive Science*, *8*(4), 780-795. https://doi.org/10.1111/tops.12216

Hulme, R. C., Barsky, D., & Rodd, J. M. (2019). Incidental Learning and Long-Term Retention of New Word Meanings From Stories: The Effect of Number of Exposures. *Language Learning*, *69*(1), 18-43. https://doi.org/10.1111/lang.12313

Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin & Review*, *10*(4), 785-813. https://doi.org/10.3758/BF03196544

Knutsen, D., & Le Bigot, L. (2014). Capturing egocentric biases in reference reuse during collaborative dialogue. *Psychonomic Bulletin & Review*, *21*(6), 1590-1599. https://doi.org/10.3758/s13423-014-0620-7

Knutsen, D., & Le Bigot, L. (2020). The influence of conceptual (mis)match on collaborative referring in dialogue. *Psychological Research*, *84*(2), 514-527. https://doi.org/10.1007/s00426-018-1060-1

Knutsen, D., Ros, C., & Le Bigot, L. (2016). Generating References in Naturalistic Face-to-Face and Phone-Mediated Dialog Settings. *Topics in Cognitive Science*, *8*(4), 796-818. https://doi.org/10.1111/tops.12218

Knutsen, D., Le Bigot, L., & Ros, C. (2017). Eliciting explicit feedback from users to attenuate memory biases in human-system dialogue. *International Journal of Human-Computer Studies, 97*, 77-87. https://doi.org/10.1016/j.ijhcs.2016.09.004

Krauss, R. M., & Weinheimer, S. (1964). Changes in reference phrases as a function of frequency of usage in social interaction: A preliminary study. *Psychonomic Science*, *1*(1-12), 113-114. https://doi.org/10.3758/BF03342817

Kumar, A. A. (2021). Semantic memory: A review of methods, models, and current challenges. *Psychonomic Bulletin & Review*, *28*(1), 40-80. https://doi.org/10.3758/s13423-020-01792-x

Lucas, M. (2000). Semantic priming without association: A meta-analytic review. *Psychonomic Bulletin & Review*, *7*(4), 618-630. https://doi.org/10.3758/BF03212999

Mak, M. H. C., Curtis, A. J., Rodd, J. M., & Gaskell, M. G. (2023). Episodic memory and sleep are involved in the maintenance of context-specific lexical information. *Journal of Experimental Psychology: General*, *152*(11), 3087-3115. https://doi.org/10.1037/xge0001435

Matheson, H. E., & Barsalou, L. W. (2018). Embodiment and Grounding in Cognitive Neuroscience. In J. T. Wixted (Ed.), *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience* (1st ed., p. 1-27). Wiley. https://doi.org/10.1002/9781119170174.epcn310

McKinley, G. L., Brown-Schmidt, S., & Benjamin, A. S. (2017). Memory for conversation and the development of common ground. *Memory & Cognition*, *45*(8), 1281-1294. https://doi.org/10.3758/s13421-017-0730-3

McNamara, T. P. (2005). *Semantic Priming: Perspectives from Memory and Word Recognition*. Psychology Press.

McRae, K., de Sa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General*, *126*(2), 99-130. https://doi.org/10.1037/0096-3445.126.2.99

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. https://doi.org/10.1016/S0749-596X(03)00028-7

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*(2), 227-234. https://doi.org/10.1037/h0031564

Narhi-Martinez, W., Dube, B., & Golomb, J. D. (2023). Attention as a multi-level system of weights and balances. *WIREs Cognitive Science*, *14*(1), e1633. https://doi.org/10.1002/wcs.1633

Nault, D. R., Voleti, R., Nicastro, M., & Munhall, K. G. (2023). Investigating the influence of local and personal common ground on memory for conversation using an online referential communication task. *Journal of Experimental Psychology: General*, *152*(6), 1598-1621. https://doi.org/10.1037/xge0001341

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, *106*(3), 226-254. https://doi.org/10.1037/0096-3445.106.3.226

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading* (pp. 264-336). Hillsdale, NJ: Erlbaum.

New, B., Pallier, C., Brysbaert, M. *et al. Lexique 2*: A new French lexical database. *Behavior Research Methods, Instruments, & Computers* 36, 516–524 (2004). https://doi.org/10.3758/BF03195598

Nieuwenhuis, I. L. C., Folia, V., Forkstam, C., Jensen, O., & Petersson, K. M. (2013). Sleep Promotes the Extraction of Grammatical Rules. *PLOS ONE*, *8*(6), e65046. https://doi.org/10.1371/journal.pone.0065046

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2), 169-190. https://doi.org/10.1017/S0140525X04000056

Pickering, M. J., & Garrod, S. (2021). *Understanding Dialogue: Language Use and Social Interaction*. Cambridge University Press.

Rodd, J. M., Berriman, R., Landau, M., Lee, T., Ho, C., Gaskell, M. G., & Davis, M. H. (2012). Learning new meanings for old words: Effects of semantic relatedness. *Memory & Cognition*, *40*(7), 1095-1108. https://doi.org/10.3758/s13421-012-0209-1

Rodd, J. M., Cai, Z. G., Betts, H. N., Hanby, B., Hutchinson, C., & Adler, A. (2016). The impact of recent and long-term experience on access to word meanings: Evidence from large-scale internet-based experiments. *Journal of Memory and Language*, *87*, 16-37. https://doi.org/10.1016/j.jml.2015.10.006

Rodd, J. M., Lopez Cutrin, B., Kirsch, H., Millar, A., & Davis, M. H. (2013). Long-term priming of the meanings of ambiguous words. *Journal of Memory and Language*, *68*(2), 180-198. https://doi.org/10.1016/j.jml.2012.08.002

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language*, *50*(4), 696-735. https://doi.org/10.2307/412243

Schneider, W., Eschman, A., & Zuccolotto, A. (2002) E-Prime User's Guide. Pittsburgh: Psychology Software Tools Inc.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002) E-Prime Reference Guide. Pittsburgh: Psychology Software Tools Inc.

Shelton, J. R., & Martin, R. C. (1992). How semantic is automatic semantic priming? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(6), 1191–1210. https://doi.org/10.1037/0278-7393.18.6.1191

Tham, E. K. H., Lindsay, S., & Gaskell, M. G. (2015). Markers of automaticity in sleep-associated consolidation of novel words. *Neuropsychologia*, *71*, 146-157. https://doi.org/10.1016/j.neuropsychologia.2015.03.025

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson, *Organization of memory.* Academic Press.

Wagner, U., Handke, L., & Walter, H. (2015). The relationship between trait empathy and memory formation for social vs. Non-social information. *BMC Psychology*, *3*(1), 2. https://doi.org/10.1186/s40359-015-0058-3

Appendix

Appendix B.a. Tables of fixed and random effects for analyses in Experiment 2.

Table A1
Fixed effects of the analysis including the IRI questionnaire, Experiment 1

Fixed effect	χ^2	DF	*b*	*SE*	*p*
Relatedness_factor	0.18	1	2.69	6.36	0.67
Personal_Distress	0.62	1	3.08	3.91	0.43
Perspective_Taking	0.06	1	-1.05	4.14	0.80
Fantasy	0.39	1	-2.73	4.37	0.53
Empathic_Concern	1.25	1	-4.93	4.40	0.26
Personal_Distress * Relatedness_factor	5.96	1	-4.85	1.99	0.01*
Perspective_Taking * Relatedness_factor	0.17	1	-0.86	2.10	0.68
Fantasy * Relatedness_factor	0.77	1	1.96	2.24	0.38
Empathic_Concern * Relatedness_factor	0.01	1	0.18	2.24	0.94

Table A2
Random effects of the analysis including the IRI questionnaire, Experiment 1

Random effects	Variance	*SD*
By-participant		
Intercept	5536	74.41
By-item (targets)		
Intercept	1379	37.14
Residual	15976	126.39

Table A3
Fixed effects of the analysis including the PSIQ-F questionnaire, Experiment 1

Fixed effect	χ^2	DF	*b*	*SE*	*p*
Relatedness_factor	0.18	1	2.67	6.37	0.67
PSIQ_vision	0.78	1	-2.54	2.87	0.38
PSIQ_sound	0.60	1	2.09	2.68	0.44
PSIQ_smell	0.99	1	-1.47	1.48	0.32
PSIQ_taste	0.65	1	1.61	1.99	0.42
PSIQ_touch	0.06	1	0.62	2.54	0.81
PSIQ_body	1.84	1	2.93	2.16	0.17
PSIQ_emo	0.88	1	-1.70	1.81	0.35
PSIQ_vision * Relatedness_factor	1.46	1	1.77	1.46	0.23
PSIQ_sound * Relatedness_factor	1.45	1	-1.64	1.36	0.23
PSIQ_smell * Relatedness_factor	0.23	1	0.35	0.75	0.63

PSIQ_taste * Relatedness_factor	2.37	1	-1.55	1.01	0.12
PSIQ_touch * Relatedness_factor	1.00	1	1.31	1.31	0.32
PSIQ_body * Relatedness_factor	<0.01	1	-0.06	1.08	0.95
PSIQ_emo * Relatedness_factor	0.27	1	0.49	0.95	0.60

Table A4
Random effects of the analysis including the PSIQ-F questionnaire, Experiment 1

Random effects	**Variance**	***SD***
By-participant		
Intercept	5610	74.90
By-item (target)		
Intercept	1412	37.58
Residual	16006	126.52

Appendix B.b. Tables of fixed and random effects for analyses in Experiment 3.

Table B1

Fixed effects of the analysis including the IRI questionnaire, Experiment 2

Fixed effect	χ^2	**DF**	***b***	***SE***	***p***
Relatedness_factor	0.52	1	-4.70	6.54	0.47
Personal_Distress	0.21	1	-1.67	3.61	0.64
Perspective_Taking	1.37	1	-4.69	4.01	0.24
Fantasy	1.26	1	4.05	3.61	0.26
Empathic_Concern	0.29	1	-1.93	3.56	0.59
Personal_Distress * Relatedness_factor	2.07	1	-2.53	1.76	0.15
Perspective_Taking * Relatedness_factor	1.33	1	-2.26	1.96	0.25
Fantasy * Relatedness_factor	1.12	1	1.92	1.82	0.29
Empathic_Concern * Relatedness_factor	2.36	1	2.67	1.74	0.12

Table B2

Random effects of the analysis including the IRI questionnaire, Experiment 2

Random effects	**Variance**	***SD***
By-participant		
Intercept	6388	79.92
By-item (target)		
Intercept	1693	41.15
Residual	16897	129.99

Table B3

Fixed effects of the analysis including the PSIQ-F questionnaire, Experiment 2

Fixed effect	χ^2	**DF**	***b***	***SE***	***p***
Relatedness_factor	0.52	1	-4.74	6.56	0.47
PSIQ_vision	11.43	1	-9.72	2.88	<0.01*
PSIQ_sound	0.08	1	-0.53	1.86	0.78
PSIQ_smell	4.56	1	3.48	1.63	0.03*
PSIQ_taste	0.36	1	-1.21	2.03	0.55
PSIQ_touch	2.80	1	3.65	2.18	0.09
PSIQ_body	0.39	1	-2.35	3.75	0.53
PSIQ_emo	1.21	1	-2.77	2.52	0.27
PSIQ_vision * Relatedness_factor	0.18	1	0.68	1.58	0.67
PSIQ_sound * Relatedness_factor	0.07	1	0.26	1.00	0.79
PSIQ_smell * Relatedness_factor	0.01	1	0.09	0.87	0.92
PSIQ_taste * Relatedness_factor	0.18	1	0.46	1.09	0.67
PSIQ_touch * Relatedness_factor	0.36	1	0.73	1.22	0.55
PSIQ_body * Relatedness_factor	0.62	1	-1.57	2.00	0.43
PSIQ_emo * Relatedness_factor	0.01	1	0.16	1.35	0.91

Table B4
Random effects of the analysis including the PSIQ-F questionnaire, Experiment 2

Random effects	**Variance**	***SD***
By-participant		
Intercept	5247	72.44
By-item (target)		
Intercept	1692	130.26
Residual	16968	130.26

Table B5
Fixed effects of the supplementary analysis, Experiment 2

Fixed effect	χ^2	**DF**	***b***	***SE***	***p***
Relatedness_factor	0.07	1	9.15	34.32	0.79
session1_time	0.04	1	-1.07	5.20	0.84
Relatedness_factor * session1_time	0.17	1	-1.03	2.50	0.68

Table B6
Random effects of the supplementary analysis, Experiment 2

Random effects	**Variance**	***SD***
By-participant		
Intercept	6728	82.02
By-item (target)		
Intercept	1688	41.08
Residual	16924	130.09

Appendix B.c. Tables of fixed and random effects for analyses in Experiment 4.

Table C1
Statistical results for the fixed effects of the analysis including the IRI questionnaire, Experiment 3

Fixed effect	χ^2	DF	*b*	*SE*	*p*
Relatedness_factor	1.16	1	-0.11	0.10	0.28
Group_participant	0.04	1	-0.04	0.19	0.84
Perspective_Taking	0.01	1	<0.01	0.03	0.93
Fantasy	0.18	1	0.01	0.03	0.67
Empathic_Concern	0.54	1	-0.03	0.03	0.46
Personal_Distress	1.19	1	-0.03	0.03	0.27
Group_participant * Relatedness_factor	12.25	1	0.50	0.14	<0.01*
Perspective_Taking * Relatedness_factor	0.09	1	-0.01	0.02	0.76
Fantasy * Relatedness_factor	<0.01	1	<0.01	0.02	0.95
Empathic_Concern * Relatedness_factor	0.19	1	-0.01	0.02	0.66
Personal_Distress * Relatedness_factor	0.04	1	<0.01	0.02	0.83
Perspective_Taking * Group_participant	<0.01	1	<0.01	0.05	0.96
Fantasy * Group_participant	1.05	1	-0.05	0.05	0.31
Empathic_Concern * Group_participant	0.43	1	0.04	0.06	0.51
Personal_Distress * Group_participant	0.13	1	0.02	0.04	0.72
Personal_Distress * Relatedness_factor*Group_participant	0.71	1	-0.02	0.03	0.40

Table C2
Random effects of the analysis including the IRI questionnaire, Experiment 3

Random effects	Variance	*SD*
By-participant		
Intercept	0.66	0.81
Relatedness_factor	0.27	0.52
By-item		
Intercept	0.55	0.74
Relatedness_factor	0.02	0.16
Group_participant	0.06	0.24
Residual	1.38	1.18

Table C3
Fixed effects of the analysis including the PSIQ-F questionnaire, Experiment 3

Fixed effect	***χ²***	**DF**	***b***	***SE***	***p***
Relatedness_factor	0.94	1	-0.10	0.11	0.33
Group_participant	0.30	1	-0.11	0.20	0.59
PSIQ_vision	0.58	1	-0.03	0.04	0.45
PSIQ_sound	0.81	1	-0.02	0.02	0.37
PSIQ_smell	0.01	1	<0.01	0.02	0.91
PSIQ_taste	0.80	1	0.02	0.02	0.37
PSIQ_touch	1.85	1	0.04	0.03	0.17
PSIQ_body	0.23	1	-0.01	0.03	0.63
PSIQ_emo	<0.01	1	<0.01	0.02	0.97
Relatedness_factor:Group_participant	10.21	1	0.49	0.15	<0.01*
Relatedness_factor:PSIQ_vision	0.06	1	<0.01	0.02	0.81
Relatedness_factor:PSIQ_sound	0.04	1	<0.01	0.01	0.85
Relatedness_factor:PSIQ_smell	0.46	1	-0.01	0.01	0.50
Relatedness_factor:PSIQ_taste	0.01	1	<0.01	0.01	0.94
Relatedness_factor:PSIQ_touch	0.01	1	<0.01	0.01	0.92
Relatedness_factor:PSIQ_body	0.27	1	0.01	0.02	0.60
Relatedness_factor:PSIQ_emo	0.78	1	-0.01	0.01	0.38
Group_participant:PSIQ_vision	0.36	1	-0.03	0.05	0.55
Group_participant:PSIQ_sound	0.88	1	0.03	0.04	0.35
Group_participant:PSIQ_smell	0.15	1	-0.01	0.03	0.70
Group_participant:PSIQ_taste	0.41	1	-0.02	0.03	0.52
Group_participant:PSIQ_touch	1.45	1	-0.05	0.04	0.23
Group_participant:PSIQ_body	1.91	1	0.06	0.04	0.17
Group_participant:PSIQ_emo	0.02	1	<0.01	0.03	0.88

Table C4
Random effects of the analysis including the PSIQ-F questionnaire, Experiment 3

Random effects	**Variance**	***SD***
By-participant		
Intercept	0.66	0.82
Relatedness_factor	0.29	0.54
By-item		
Intercept	0.55	0.74
Relatedness_factor	0.02	0.16
Group_participant	0.06	0.24
Residual	1.38	1.18

Table C5
Fixed effects of the supplementary analysis, Experiment 3

Fixed effect	χ^2	DF	*b*	*SE*	*p*
Relatedness_factor	1.14	1	-0.76	0.71	0.29
session1_time	4.14	1	0.12	0.06	0.04*
Relatedness_factor * session1_time	2.70	1	0.08	0.05	0.10

Table C6
Random effects of the supplementary analysis, Experiment 3

Random effects	Variance	*SD*
By-participant		
Intercept	0.71	0.85
Relatedness_factor	0.49	0.70
By-item		
Intercept	0.64	0.80
Relatedness_factor	0.03	0.18
Residual	1.56	1.25

Tables

Table 1. Psycholinguistic properties of primes and targets for experimental lists 1 and 2 separately.

		Lexical frequency		Number of letters		Number of phonemes		Number of orthographic neighbors		Number of phonological neighbors	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
List 1	Primes	45.40	88.84	5.90	1.59	4.40	1.43	3.10	3.18	6.60	7.17
	Targets	28.08	44.07	6.20	1.20	4.60	1.27	3.20	4.40	9.75	10.40
List 2	Primes	50.40	129.17	6.10	1.52	4.40	1.31	4.60	6.42	8.45	9.13
	Targets	30.78	45.65	6.20	1.77	4.60	1.53	3.60	5.43	9.15	8.62

Note. All properties were extracted from New et al. (2004); lexical frequency was based on movie subtitle corpora.

Table 2. Fixed effects, Experiment 1

Fixed effect	χ^2	DF	*b*	*SE*	*p*
Relatedness_factor	0.18	1	2.67	6.37	0.68

Table 3. Random effects, Experiment 1

Random effects	**Variance**	***SD***
By-participant		
Intercept	5246	72.43
By-item (targets)		
Intercept	1384	37.20
Residual	15996	126.47

Table 4. Fixed effects, Experiment 2

Fixed effect	χ^2	**DF**	***b***	***SE***	***p***
Relatedness_factor	0.52	1	-4.72	6.55	0.47

Table 5. Random effects, Experiment 2

Random effects	**Variance**	***SD***
By-participant		
Intercept	6563	81.01
By-item (targets)		
Intercept	1688	41.09
Residual	16914	130.05

Table 6. Psycholinguistic properties of pairs of words of the semantic relatedness judgement task for experimental lists 1 and 2 separately, Experiment 3

		Lexical frequency		Number of letters		Number of phonemes		Number of orthographic neighbors		Number of phonological neighbors	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
List 1	Primes	45.40	88.84	5.90	1.59	4.40	1.43	3.10	3.18	6.60	7.17
	Targets	28.08	44.07	6.20	1.20	4.60	1.27	3.20	4.40	9.75	10.40
List 2	Primes	50.40	129.17	6.10	1.52	4.40	1.31	4.60	6.42	8.45	9.13
	Targets	30.78	45.65	6.20	1.77	4.60	1.53	3.60	5.43	9.15	8.62
Highly semantically	Word 1	44.01	76.37	6.10	1.52	4.25	1.37	3.05	3.75	10.45	8.79
related	Word 2	23.53	34.94	6.30	2.66	4.85	2.23	2.95	4.24	8.75	9.95

Note. All properties were extracted from New et al. (2004); lexical frequency was based on movie subtitle corpora.

Table 7. Psycholinguistic properties of words presented in the word-picture relatedness judgement task corresponding to the primes and targets for experimental lists 1 and 2 separately, Experiment 3

		Lexical frequency		Number of letters		Number of phonemes		Number of orthographic neighbors		Number of phonological neighbors	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
List 1	Word 1. related	45.40	88.84	5.90	1.59	4.40	1.43	3.10	3.18	6.60	7.17
	Word 2. related	28.08	44.07	6.20	1.20	4.60	1.27	3.20	4.40	9.75	10.40
	Word 1. unrelated	44.72	53.00	5.90	2.07	4.45	1.76	2.95	3.68	9.55	8.67
	Word 2. unrelated	34.02	51.51	6.10	2.02	4.35	1.39	4.15	5.91	9.85	10.39
List 2	Word 1. related	50.40	129.17	6.10	1.52	4.40	1.31	4.60	6.42	8.45	9.13
	Word 2. related	30.78	45.65	6.20	1.77	4.60	1.53	3.60	5.43	9.15	8.62
	Word 1. unrelated	44.72	53.00	5.90	2.07	5.45	1.76	2.95	3.68	9.55	8.67
	Word 2. unrelated	32.66	44.48	6.40	1.73	4.70	1.78	2.85	3.87	6.80	6.45

Note. All properties were extracted from New et al. (2004); lexical frequency was based on movie subtitle corpora.

Table 8. Fixed effects of the semantic relatedness task, Experiment 3

Fixed effect	χ^2	**DF**	***b***	***SE***	***p***
Relatedness_factor	1.20	1	-0.11	0.10	0.27
Group_participant	0.04	1	-0.04	0.19	0.84
Relatedness_factor * Group_participant	12.63	1	0.50	0.14	<0.01*

Table 9. Random effects of the semantic relatedness task, Experiment 3

Random effects	**Variance**	***SD***
By-participant		
Intercept	0.62	0.79
Relatedness_factor	0.26	0.51
By-item		
Intercept	0.55	0.74
Relatedness_factor	0.02	0.16
Group_participant	0.06	0.24

Table 10. Fixed effects of the word-picture relatedness task, Experiment 3

Fixed effect	χ^2	**DF**	***b***	***SE***	***p***
WP_relatedness	178.72	1	-2.68	0.20	<0.01*
Group_participant	7.33	1	0.36	0.13	0.01*
WP_relatedness * Group_participant	19.86	1	-0.76	0.17	<0.01*

Table 11. Random effects of the word-picture relatedness task, Experiment 3

Random effects	**Variance**	***SD***
By-participant		
Intercept	0.25	0.50
WP_relatedness	0.39	0.63
By-item		
Intercept	1.11	1.05
Group_participant	0.17	0.41
Residual	1.98	1.41

Table 12. Mean and standard deviation for the score of word-picture relatedness, Experiment 3

	Experimental group		**Control group**	
	Mean	**SD**	**Mean**	**SD**
Related in the dialogue	6.13	1.23	5.28	1.93
Unrelated and not presented in the dialogue	2.70	2.02	2.60	2.04

Note. Scores of word-picture relatedness for each group of participants (Experimental or Control) and each type of word-picture pair (“Related in the dialogue” or “Unrelated and not presented in the dialogue).

Figure 1. Example of a tangram picture used in this study. This example could be referred to as an "anchor".

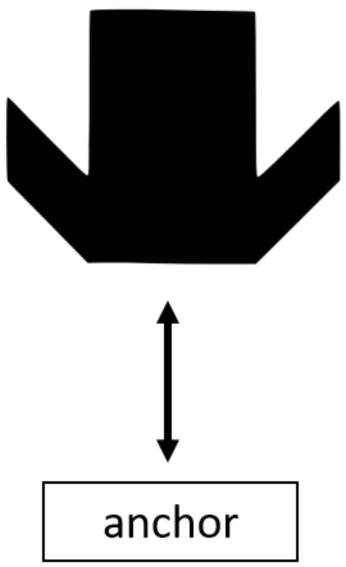


Figure 2. Adapted illustration of a trial. A: example of two words proposed to name the picture in order to create a new lexico-semantic relationship between the two words (blue arrow). B: illustration of the three steps of the IAR task.

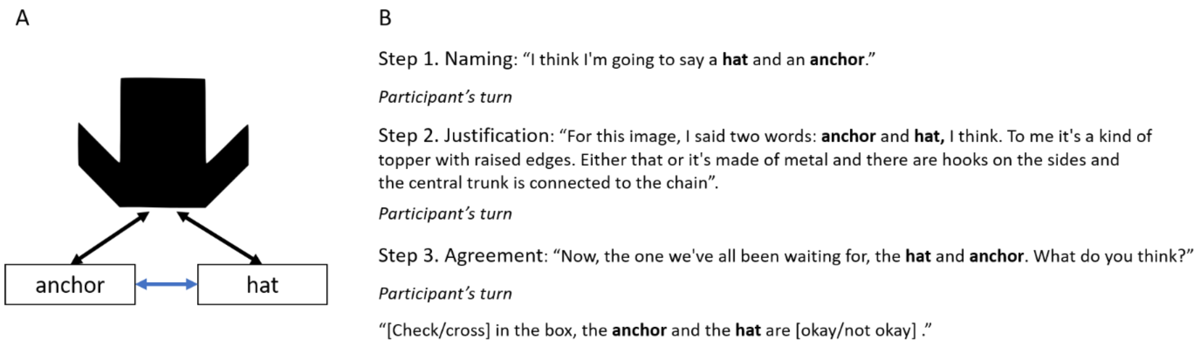


Figure 3. Illustration of the two experimental conditions of the relatedness factor.

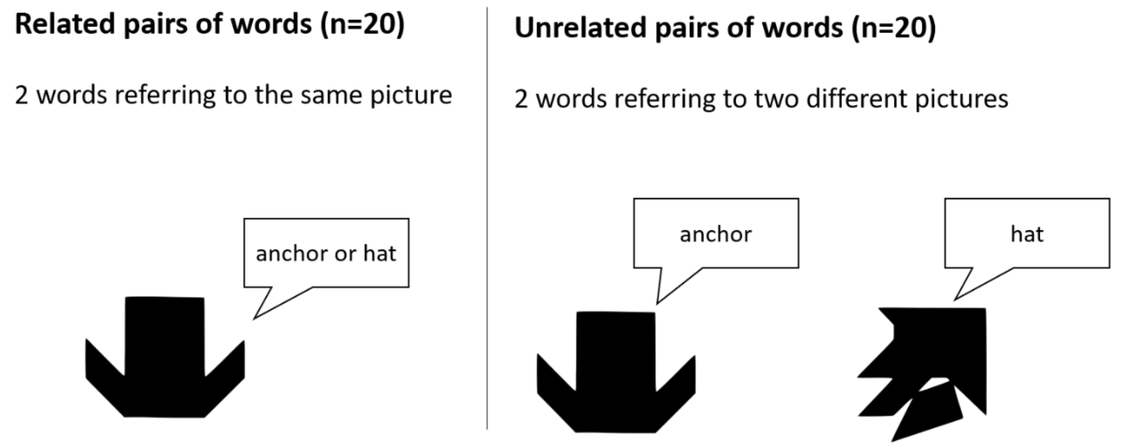


Figure 4. Recap of the procedure of the three experiments. IRI: Interpersonal Reactivity Index questionnaire; PSIQ-F: Plymouth Sensory Imagery Questionnaire in French.

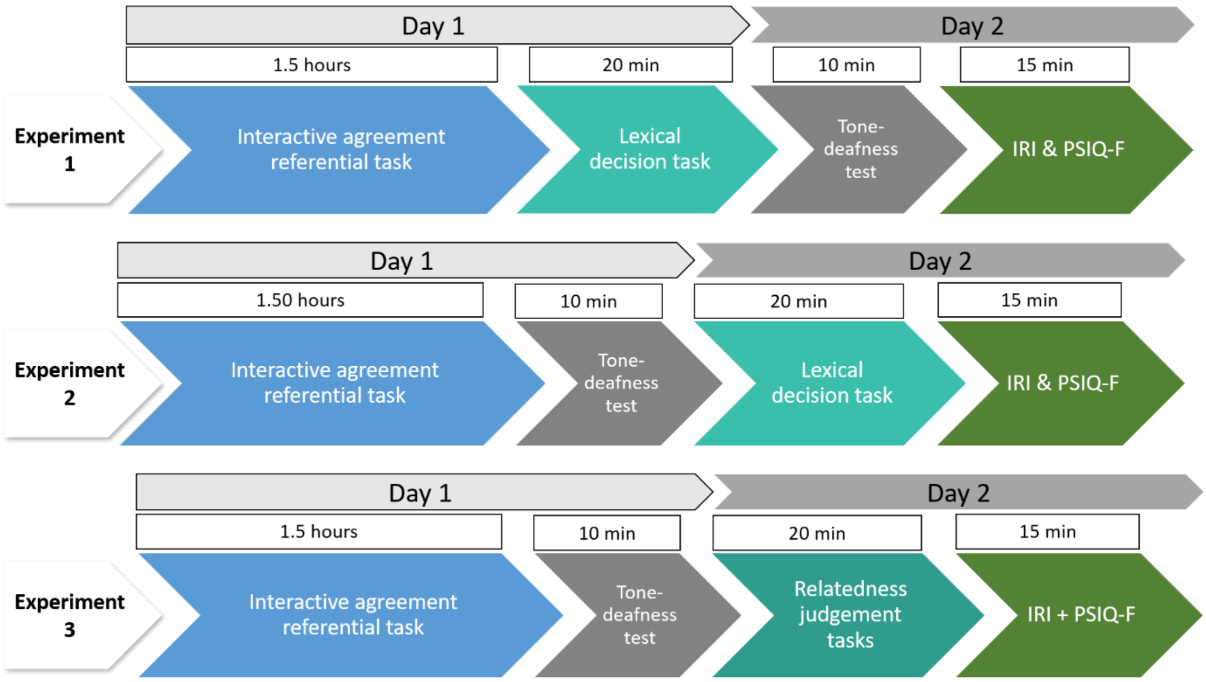


Figure 5. Schematic representation of a trial. The confederate and the participant saw the same three pictures and the confederate had to name them using either one or two words.

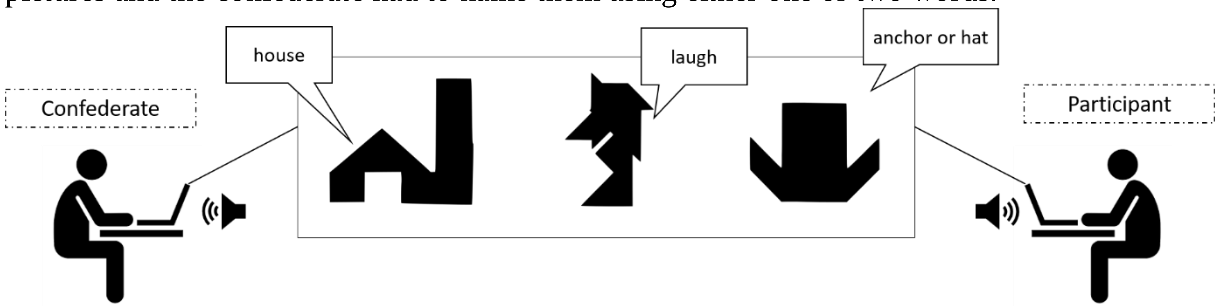


Figure 6. Reaction times as a function of the relatedness factor in Experiment 1 and 2 (A and C) and scores of Personal Distress (B) in Experiment 1.

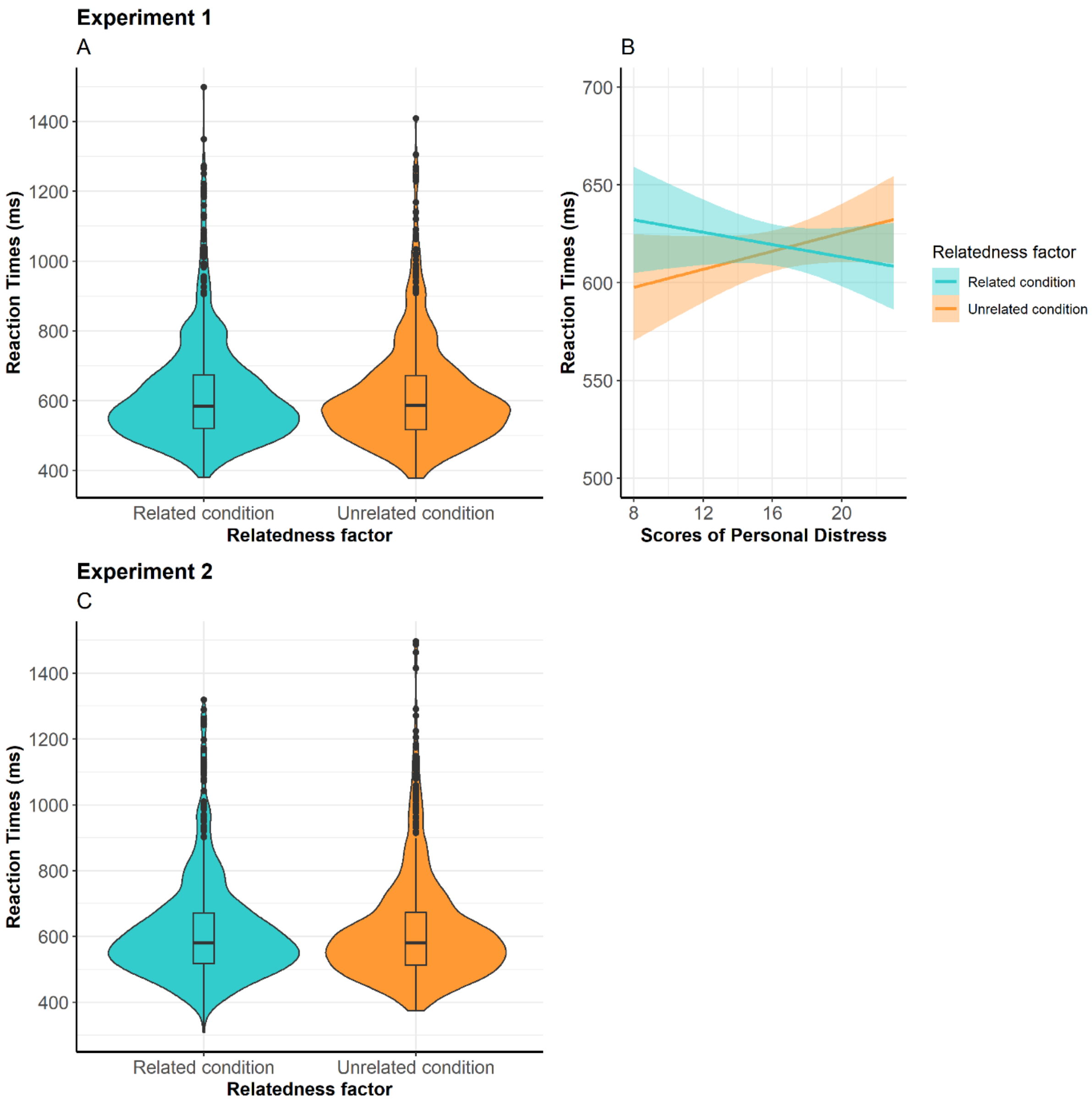
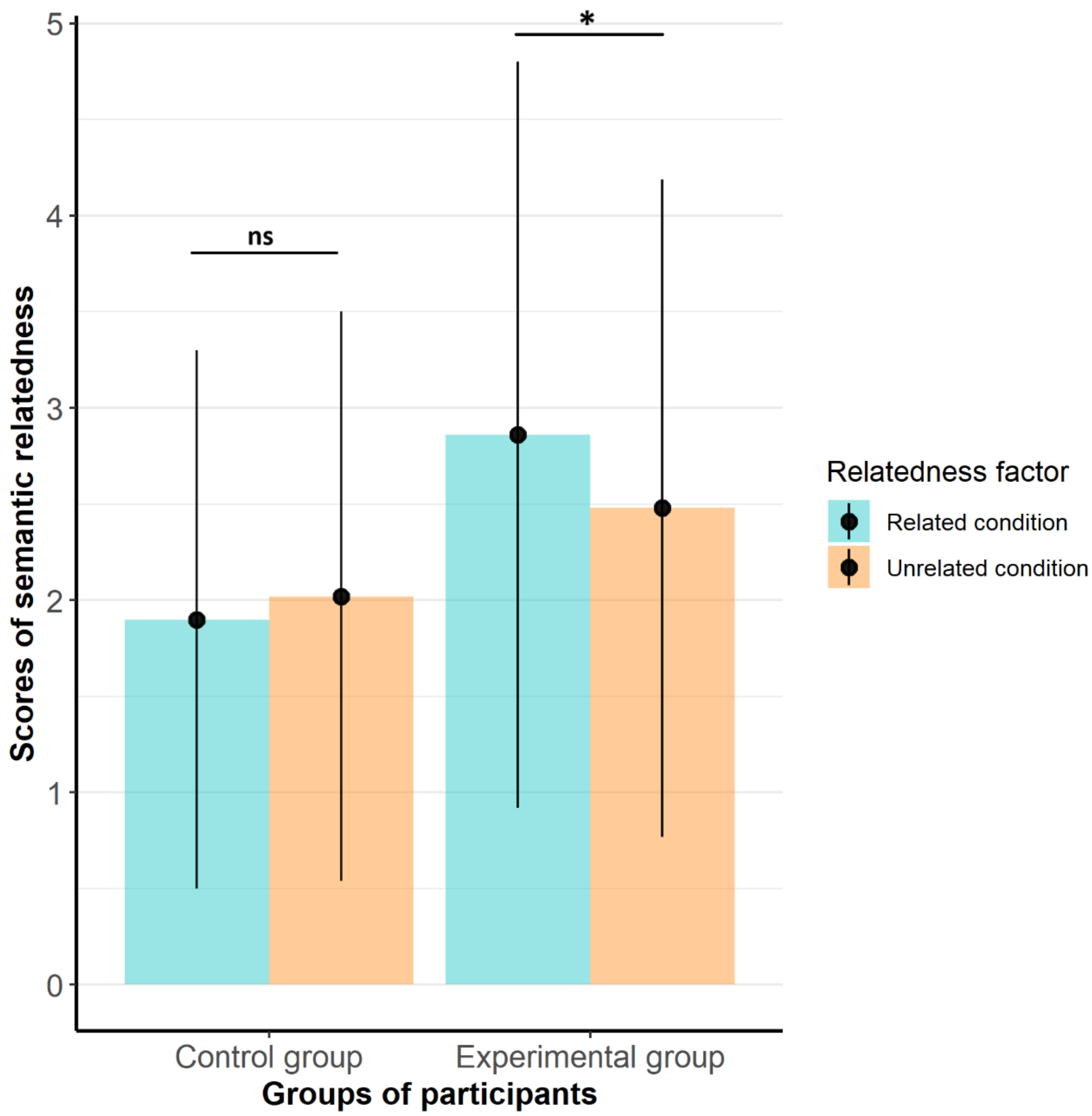
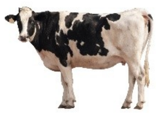


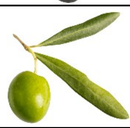
Figure 7. Semantic relatedness scores as a function of the groups of participants and the relatedness factor in Experiment 3. Note. Vertical bars correspond to standard deviation. Ns: non-significant. The asterisk denotes significance.



Note. Vertical bars correspond to standard deviation. Ns: non-significant. The asterisk denotes significance.

Appendix C. Detail of the words (in French and English), pictures and properties for each of the 10 semantic categories, and the results to the pre-tests. Pre-test 'a': representativeness of the properties in relation to all members of the category ; pre-test 'b': representativeness of the properties in relation to the pictures ; pre-test 'c': representativeness of the propertiess in relation to the words

Mammals				Pre-test a		Pre-test b		Pre-test c	
	Words in French	**Words in English**	**Pictures**	Mean property 1 (SD)	Mean property 2 (SD)	Property 1	Property 2	Property 1	Property 2
Typical words	girafe	giraffe				100.00%	100.00%	100.00%	100.00%
	singe	ape				100.00%	100.00%	100.00%	100.00%
	éléphant	elephant				100.00%	100.00%	100.00%	100.00%
	vache	cow				100.00%	100.00%	100.00%	100.00%
Atypical words	poney	pony				100.00%	100.00%	100.00%	100.00%
	panda	panda				100.00%	100.00%	100.00%	100.00%
	jaguar	jaguar				100.00%	100.00%	100.00%	100.00%
	buffle	buffalo				93.75%	93.75%	100.00%	100.00%
Properties in French: 'Est un vertébré' (1) / 'Possède quatre membres articulés' (2)				5.73 (1.10)	5.33 (1.40)				
Properties in English: 'Is a vertebrate' (1) / 'Possesses four articulated limbs' (2)									
Weapons				Pre-test a		Pre-test b		Pre-test c	
	Words in French	**Words in English**	**Pictures**						
Typical words	fusil	rifle				100.00%	100.00%	100.00%	100.00%
	grenade	grenade				100.00%	100.00%	100.00%	100.00%
	canon	cannon				100.00%	100.00%	93.75%	100.00%

	couteau	knife				93.75%	93.75%	100.00%	87.5%
Atypical words	javelot	javelin				93.75%	100.00%	81.25%	68.75%
	boulet	cannonball				93.75%	87.5%	100.00%	100.00%
	glaive	glaive				93.75%	100.00%	100.00%	100.00%
	roquette	rocket				87.5%	75,00 %	93.75%	93.75%
Properties in French: 'Peut causer une destruction matérielle' (1)/ 'Peut être utilisé lors de combats et de guerres' (2)				6.07 (0.88)	6.60 (0.63)				
Properties in English: 'Can cause material destruction' (1) / 'Can be used in combat and war' (2)									
Fruits				Pre-test a		Pre-test b		Pre-test c	
	Words in French	**Words in English**	**Pictures**						
Typical words	orange	orange				62.5%	100.00%	100.00%	100.00%
	banane	banana				50,00 %	62.5%	87.5%	75,00 %
	fraise	strawberry				81.25%	93.75	100.00%	93.75%
	cerise	cherry				68.75%	100.00%	93.75%	100.00%
Atypical words	olive	olive				56.25%	100.00%	100.00%	100.00%
	citrouille	pumpkin				50,00 %	100.00%	81.25%	100.00%
	gland	acorn				62.5%	75,00 %	81.25%	68.75%
	pistache	pistachio				50,00 %	68.75%	62.5%	62.5%

Properties in French: 'Se forme à partir de fleurs pollinisées' (1) / 'Possède un pépin, une graine, un noyau ou de la pulpe' (2)				5.40 (1.59)	6.27 (0.88)				
Properties in English: 'Forms from pollinated flowers' (1) / 'Has a pip, a seed, a kernel or pulp' (2)									
Musical instruments				Pre-test a		Pre-test b		Pre-test c	
	Words in French	**Words in English**	**Pictures**						
Typical words	trompette	trumpet				100.00%	75,00 %	100.00%	87.5%
	piano	piano				100.00%	75,00 %	100.00%	87.5%
	violon	violin				100.00%	87.5%	100.00%	87.5%
	guitare	guitar				100.00%	81.25%	100.00%	87.5%
Atypical words	maracas	maracas				100.00%	62.5%	100.00%	87.5%
	sifflet	whistle				100.00%	68.75%	93.75%	81.25%
	grelots	sleight bells				100.00%	62.5%	100.00%	68.75%
	timbale	kettledrum				100.00%	68.75%	100.00%	87.5%
Properties in French: 'Peut produire un son contrôlé par un musicien' (1) / 'Appartient à une famille (cordes, vents ou percussions)' (2)				6.60 (0.74)	6.67 (0.72)				
Properties in English: 'Can produce a sound controlled by a musician' (1) / 'Belongs to a family (strings, winds or percussion)' (2)									
Vegetables				Pre-test a		Pre-test b		Pre-test c	
	Words in French	**Words in English**	**Pictures**						
Typical words	chou	cabbage				100.00%	100.00%	100.00%	100.00%
	navet	turnip				100.00%	93.75%	100.00%	100.00%
	salade	lettuce				100.00%	100.00%	100.00%	100.00%

	haricot	bean				100.00%	100.00%	100.00%	100.00%
Atypical words	asperge	asparagus				100.00%	100.00%	100.00%	100.00%
	piment	chilli				93.75%	81.25%	100.00%	87.5%
	cornichon	pickle				87.5%	87.5%	87.5%	93.75%
	maïs	corn				75,00 %	100.00%	93.75%	100.00%
Properties in French: 'Pousse dans un potager' / 'Est plutôt consommé en entrée ou en plat'				6.00 (1.13)	5.60 (1.30)				
Properties in English: 'Grows in a vegetable garden' / 'Is best eaten as a starter or main course'									
Furniture				Pre-test a		Pre-test b		Pre-test c	
	Words in French	**Words in English**	**Pictures**						
Typical words	table	table				93.75%	100.00%	93.75%	100.00%
	bureau	desk				100.00%	100.00%	100.00%	100.00%
	chaise	chair				93.75%	100.00%	100.00%	100.00%
	lit	bed				100.00%	100.00%	100.00%	100.00%
Atypical words	berceau	cradle				87.5%	50,00 %	100.00%	81.25%
	comptoir	counter				87.5%	93.75%	81.25%	75,00 %
	vitrine	display cabinet				100.00%	100.00%	75,00 %	68.75%
	établi	workbench				81.25%	50,00 %	93.75%	50,00 %
Properties in French: 'Sert au confort ou au rangement de l'utilisateur' / 'Peut contribuer à la décoration du logement'				5.73 (1.44)	5.60 (1.24)				

Properties in English: 'Provides comfort or storage for the user' / 'Can contribute to the decoration of the home'									
Tools				Pre-test a		Pre-test b		Pre-test c	
	Words in French	**Words in English**	**Pictures**						
Typical words	marteau	hammer				100.00%	100.00%	100.00%	100.00%
	pelle	shovel				100.00%	100.00%	100.00%	100.00%
	pioche	pickaxe				100.00%	100.00%	100.00%	100.00%
	pince	pliers				100.00%	100.00%	100.00%	100.00%
Atypical words	épingle	pin				68.75%	93.75%	81.25%	68.75%
	fourche	pitchfork				100.00%	100.00%	100.00%	87.5%
	scalpel	scalpel				100.00%	100.00%	100.00%	100.00%
	maillet	mallet				100.00%	100.00%	100.00%	100.00%
Properties in French: 'Permet de fabriquer des objets ou de réaliser une activité humaine' / 'Améliore l'efficacité d'une action ou la rend possible'				6.27 (0.80)	5.27 (1.16)				
Properties in English: 'Enables to make objects or to carry out a human activity' / 'Improves the efficiency of an action or makes it possible'									
Containers				Pre-test a		Pre-test b		Pre-test c	
	Words in French	**Words in English**	**Pictures**						
Typical words	bol	bowl				100.00%	93.75%	100.00%	100.00%
	assiette	plate				56.25%	100.00%	87.5%	93.75%
	vase	vase				100.00%	100.00%	100.00%	100.00%
	cuillère	spoon				75,00 %	93.75%	100.00%	87.5%

Atypical words	gobelet	cup				100.00%	100.00%	100.00%	93.75%
	biberon	baby bottle				93.75%	93.75%	93.75%	87.5%
	chaudron	cauldron				100.00%	100.00%	100.00%	100.00%
	saucière	gravy boat				100.00%	100.00%	100.00%	93.75%
Properties in French: 'Est un objet creux, incurvé' / 'Est destiné à recevoir des produits solides, liquides ou gazeux'				4.33 (1.72)	4.93 (1.71)				
Properties in English: 'Is a hollow, curved object' / 'Is designed to hold solid, liquid or gaseous products'									
Vehicle				Pre-test a		Pre-test b		Pre-test c	
	Words in French	**Words in English**	**Pictures**						
Typical words	vélo	bike				100.00%	87.5%	100.00%	81.25%
	camion	truck				100.00%	100.00%	100.00%	100.00%
	auto	car				100.00%	100.00%	100.00%	100.00%
	avion	plane				100.00%	100.00%	100.00%	100.00%
Atypical words	traîneau	sleigh				100.00%	75,00 %	100.00%	68.75%
	caravane	caravan				93.75%	100.00%	81.25%	87.5%
	wagon	coach				87.5%	100.00%	100.00%	93.75%
	poussette	stroller				100.00%	87.5%	100.00%	50,00 %
Properties in French: 'Permet le déplacement dans l'environnement' / 'Nécessite une source d'énergie (motrice, thermique ou électrique)'				6.47 (0.92)	5.73 (1.33)				

Properties in English: 'Enables to move in the environment' / 'Requires a source of energy (motor, thermal or electrical)'									
Clothing				Pre-test a		Pre-test b		Pre-test c	
	Words in French	**Words in English**	**Pictures**						
Typical words	chaussette	sock				100.00%	100.00%	100.00%	100.00%
	chapeau	hat				100.00%	100.00%	100.00%	100.00%
	pantalon	pants				100.00%	100.00%	100.00%	100.00%
	chemise	shirt				100.00%	100.00%	100.00%	100.00%
Atypical words	casquette	cap				100.00%	100.00%	100.00%	100.00%
	kilt	kilt				100.00%	100.00%	100.00%	100.00%
	pantoufle	slipper				100.00%	100.00%	100.00%	100.00%
	béret	beret				100.00%	100.00%	100.00%	100.00%
Properties in French: 'Sert à couvrir une partie du corps humain' / 'Est constitué de matériaux (le coton, le polyester, la laine, etc...)'				6.80 (0.41)	6.00 (0.85)				
Properties in English: 'Is used to cover parts of the human body' / 'Is made of materials (cotton, polyester, wool, etc.)'									

Appendix D. Tables of fixed and random effects for analyses of Experiments 5 and 6.

Table D1.

Fixed effects of the analysis including the role factor, Experiment 1, post-test

Fixed effect	χ^2	DF	*b*	*SE*	*p*
Relatedness_factor	1.31	1	14.86	12.98	0.25
Typicality_factor	0.43	1	11.10	16.88	0.51
Role_factor	0.56	1	-23.41	31.21	0.45
Relatedness_factor*Typicality_factor	3.42	1	-26.99	14.60	0.06
Relatedness_factor*Role_factor	1.37	1	-16.62	14.20	0.24
Typicality_factor*Role_factor	1.36	1	-14.60	12.52	0.24

Table D2.

Random effects of the analysis including the role factor, Experiment 1, post-test

Random effects	Variance	*SD*
By-participant		
Intercept	7271.60	85.27
Relatedness_factor	378.60	19.46
By-item (target)		
Intercept	3236.80	56.89
Relatedness_factor	1003.30	31.67
Residual	23980.40	154.86

Table D3

Statistical results for the fixed effects of the analysis including the IRI questionnaire, Experiment 1, post-test

Fixed effect	χ^2	**DF**	***b***	***SE***	***p***
Relatedness_factor	1.16	1	7.08	10.86	0.51
Typicality_factor	0.04	1	4.01	15.72	0.80
Perspective_Taking	0.01	1	-12.37	4.79	<0.01*
Fantasy	0.18	1	0.55	3.80	0.89
Empathic_Concern	0.54	1	2.55	4.09	0.52
Personal_Distress	1.19	1	1.92	3.36	0.45
Typicality_factor * Relatedness_factor	3.44	1	-27.15	14.63	0.06
Perspective_Taking * Relatedness_factor	0.70	1	1.92	2.29	0.40
Fantasy * Relatedness_factor	0.08	1	-0.50	1.78	0.77
Empathic_Concern * Relatedness_factor	0.15	1	0.77	1.95	0.69
Personal_Distress * Relatedness_factor	3.27	1	-2.89	1.60	0.07
Perspective_Taking * Typicality_factor	0.84	1	1.85	2.02	0.36
Fantasy * Typicality_factor	0.01	1	0.18	1.58	0.91
Empathic_Concern * Typicality_factor	<0.01	1	0.06	1.72	0.97
Personal_Distress * Typicality_factor	0.04	1	-0.27	1.43	0.84

Table D4

Random effects of the analysis including the IRI questionnaire, Experiment 1, post-test

Random effects	**Variance**	***SD***
By-participant		
Intercept	6530.80	80.81
Relatedness_factor	341.10	18.47
By-item (target)		
Intercept	3241.90	65.94
Relatedness_factor	1016.20	31.88

Residual	24001.2	154.92

Table D5

Fixed effects of the analysis including the PSIQ-F questionnaire, Experiment 1, post-test

Fixed effect	χ^2	DF	*b*	*SE*	*p*
Relatedness_factor	0.51	1	7.75	10.82	0.47
Typicality_factor	0.06	1	3.87	15.68	0.81
PSIQ_vision	0.14	1	-1.31	3.53	0.71
PSIQ_sound	1.84	1	-3.09	2.27	0.17
PSIQ_smell	1.70	1	2.80	2.15	0.19
PSIQ_taste	1.62	1	-4.45	3.49	0.20
PSIQ_touch	1.75	1	5.86	4.42	0.18
PSIQ_body	<0.01	1	-0.13	4.56	0.98
PSIQ_emo	0.92	1	-2.44	2.55	0.34
Relatedness_factor*Typicality_factor	3.41	1	-27.08	14.65	0.06
Relatedness_factor*PSIQ_vision	2.61	1	-2.64	1.63	0.11
Relatedness_factor*PSIQ_sound	0.36	1	0.62	1.04	0.55
Relatedness_factor*PSIQ_smell	0.53	1	-0.70	0.96	0.46
Relatedness_factor*PSIQ_taste	0.60	1	1.22	1.57	0.44
Relatedness_factor*PSIQ_touch	0.28	1	-1.12	2.13	0.60
Relatedness_factor*PSIQ_body	0.93	1	2.03	2.11	0.33
Relatedness_factor*PSIQ_emo	4.09	1	-2.36	1.17	0.04*
Typicality_factor*PSIQ_vision	2.59	1	-2.39	1.49	0.11
Typicality_factor*PSIQ_sound	0.03	1	-0.16	0.94	0.87
Typicality_factor*PSIQ_smell	0.48	1	-0.60	0.86	0.49
Typicality_factor*PSIQ_taste	0.03	1	-0.25	1.41	0.86
Typicality_factor*PSIQ_touch	0.01	1	0.22	1.84	0.91
Typicality_factor*PSIQ_body	1.01	1	1.87	1.86	0.31
Typicality_factor*PSIQ_emo	<0.01	1	-0.06	1.05	0.95

Table D6.

Random effects of the analysis including the PSIQ-F questionnaire, Experiment 1, post-test

Random effects	**Variance**	***SD***
By-participant		
Intercept	7116.80	84.36
Relatedness_factor	289.70	17.02
By-item (target)		
Intercept	3220.40	56.75
Relatedness_factor	1031.10	32.11
Residual	23996.10	154.91

Table D7.

Fixed effects of the analysis including the role factor, Experiment 2, post-test

Fixed effect	χ^2	**DF**	***b***	***SE***	***p***
Relatedness_factor	1.35	1	-14.62	12.56	0.24
Typicality_factor	0.08	1	-6.02	21.50	0.78
Role_factor	1.53	1	53.87	43.61	0.22
Relatedness_factor*Typicality_factor	0.02	1	-1.91	14.76	0.90
Relatedness_factor*Role_factor	0.81	1	-12.73	14.11	0.37
Typicality_factor*Role_factor	0.19	1	-6.06	14.10	0.67

Table D8.

Random effects of the analysis including the role factor, Experiment 2, post-test

Random effects	**Variance**	***SD***
By-participant		
Intercept	16657.40	129.06
By-item (target)		
Intercept	6103.80	78.13
Relatedness_factor	356.20	18.87
Residual	34175.00	184.86

Table D9.

Statistical results for the fixed effects of the analysis including the IRI questionnaire, Experiment 2, post-test

Fixed effect	χ^2	**DF**	*b*	*SE*	*p*
Relatedness_factor	4.01	1	-20.94	10.46	0.04*
Typicality_factor	0.20	1	-9.06	20.33	0.66
Perspective_Taking	0.84	1	4.29	4.67	0.36
Fantasy	0.67	1	-5.13	6.28	0.41
Empathic_Concern	0.39	1	4.43	7.12	0.53
Personal_Distress	0.01	1	0.60	4.95	0.90
Typicality_factor * Relatedness_factor	0.01	1	-1.69	14.79	0.91
Perspective_Taking * Relatedness_factor	0.21	1	-0.69	1.48	0.64
Fantasy * Relatedness_factor	<0.01	1	0.01	1.97	0.99
Empathic_Concern * Relatedness_factor	0.13	1	-0.83	2.34	0.72
Personal_Distress * Relatedness_factor	0.08	1	0.44	1.59	0.78
Perspective_Taking * Typicality_factor	0.07	1	0.38	1.48	0.80
Fantasy * Typicality_factor	0.88	1	1.84	1.96	0.35
Empathic_Concern * Typicality_factor	0.17	1	0.92	2.23	0.68
Personal_Distress * Typicality_factor	0.74	1	-1.33	1.55	0.39

Table D10.

Random effects of the analysis including the IRI questionnaire, Experiment 2, post-test

Random effects	**Variance**	*SD*
By-participant		
Intercept	17748.60	133.22
By-item (target)		
Intercept	6096.60	78.08
Relatedness_factor	365.70	19.12
Residual	34242.3	185.05

Table D11.

Fixed effects of the analysis including the PSIQ-F questionnaire, Experiment 2, post-test

Fixed effect	χ^2	**DF**	***b***	***SE***	***p***
Relatedness_factor	4.13	1	-20.32	9.99	0.04*
Typicality_factor	0.25	1	-9.34	18.72	0.62
PSIQ_vision	0.05	1	0.88	3.98	0.82
PSIQ_sound	0.44	1	2.53	3.83	0.51
PSIQ_smell	0.30	1	-1.56	2.83	0.58
PSIQ_taste	0.24	1	1.65	3.34	0.62
PSIQ_touch	1.67	1	5.84	4.53	0.20
PSIQ_body	1.28	1	-4.77	4.22	0.26
PSIQ_emo	0.25	1	-1.20	2.40	0.62
Relatedness_factor*Typicality_factor	0.01	1	-1.51	14.15	0.91
Relatedness_factor*PSIQ_vision	4.10	1	-2.52	1.25	0.04*
Relatedness_factor*PSIQ_sound	0.36	1	0.72	1.20	0.55
Relatedness_factor*PSIQ_smell	0.19	1	-0.39	0.90	0.67
Relatedness_factor*PSIQ_taste	0.06	1	-0.26	1.06	0.81
Relatedness_factor*PSIQ_touch	0.21	1	-0.67	1.46	0.65
Relatedness_factor*PSIQ_body	1.94	1	1.89	1.36	0.16
Relatedness_factor*PSIQ_emo	0.67	1	0.61	0.75	0.41
Typicality_factor*PSIQ_vision	1.10	1	1.42	1.35	0.29
Typicality_factor*PSIQ_sound	0.15	1	0.50	1.30	0.70
Typicality_factor*PSIQ_smell	0.16	1	-0.38	0.96	0.69
Typicality_factor*PSIQ_taste	0.54	1	-0.84	1.14	0.46
Typicality_factor*PSIQ_touch	0.01	1	0.18	1.54	0.90
Typicality_factor*PSIQ_body	<0.01	1	0.04	1.47	0.98
Typicality_factor*PSIQ_emo	0.16	1	0.33	0.81	0.69

Table D12.

Random effects of the analysis including the PSIQ-F questionnaire, Experiment 2, post-test

Random effects	**Variance**	***SD***
By-participant		
Intercept	18140.50	134.69
Typicality_factor	349.20	18.69
By-item (target)		
Intercept	4693.40	68.51
Residual	34233.60	185.02