

Gesture-speech temporal integration in language development

Alfonso Igualada Pérez

TESI DOCTORAL UPF / 2017

DIRECTOR DE LA TESI

Dra. Pilar Prieto i Dra. Laura Bosch (Departament Cognició,
Desenvolupament i Psicologia de l'Educació Facultat de Psicologia, Universitat
de Barcelona)

DEPARTAMENT DE TRADUCCIÓ I CIÈNCIES DEL
LLENGUATGE



A mi familia

ACKNOWLEDGMENTS

First of all, I would like to thank my supervisor Pilar Prieto from the Department of Translation and Language Sciences at the Universitat Pompeu Fabra. She has offered great guidance throughout these years. Thanks for always being available when needed, for being so encouraging and enthusiastic about work and research, for giving me the opportunity to participate in a wealth of scientific projects and conferences, for promoting teamwork and for giving support to my ideas. So many thanks. Second, I would like to thank my supervisor Laura Bosch from the Department of Cognition, Development and Educational Psychology from the Universitat de Barcelona. Thanks for opening the doors of your lab, for always providing wise insight, and for always participating with great precision when asked. I really feel very fortunate to have been able to count on the experience and knowledge of my two supervisors. Thanks a lot to both of you.

To my family, I would like to say thank you so much for always supporting me in whatever I have chosen to do. My heartfelt thanks to Carmen Pérez Herraiz, Pablo Gago López, Félix Fernández-Palacios Pérez, and Alfonso Igualada Pedraza. Thanks to all my friends. I love you. Thank you for looking out for me.

Special thanks go to Courtenay Norbury for hosting me at the Royal Holloway University of London, and to the Lilac team (Debbie Gooch, Charlotte Wray, Katie Whiteside, Rebecca Lucas, Harriet Maydew, Claire Sears and Clara Andrés Roqueta). Many thanks to all members of the Prosodic Studies Group, namely Joan Borràs Comes (special thanks for helping with stats), Maria del Mar Vanrell, Meghan Armstrong, Paolo Roseano, Rafèu Sichel Bazin, Núria Esteve-Gibert, Santiago González Fuente, Iris Hübscher, Olga Kushch, Evi Kiagia, Judith Llanes Coromina, Ingrid Vilà Giménez, Cristina Sanchez Conde and Florence Bails, and my other PhD colleagues at the Universitat Pompeu Fabra (Giorgia Zorzi, Toni Bassaganyas, Celia Alba, Aina Obis, Lieke van Maastricht, Kata Wohlmuth, Sara Cañas, Alexandra Spalek, Mihajlo Ignjatovic, Veronika Richtarcikova Alexandra Navarrete and Eugenio Vigo). Thanks to all of you for the great times.

Additional thanks are owed to the Department of Translation and Language Sciences at the Universitat Pompeu Fabra for their support during my pre-doctoral stage, as well as for funding my three-month research stay at the Royal Holloway University of London, and for the financial support that allowed me to go to many conferences. Special thanks go to Àlex Alsina, Cristina Gelpí, Toni Badia, Carmen Pérez, Aurora Bel, Louise McNally, Gemma Barberà Laia Mayol and Josep Maria Fontana. This Ph.D. was funded by a Recercaixa 2013-2015 project directed by Pilar Prieto and two projects from the Spanish MINECO (BFU2012-31995, under Pilar Prieto, and PSI-2011-25376, under Laura

Bosch). Thanks to the staff at ADAPEI-ASPRONA, the Escola Sant Martí, Escola La Farigola del Clot, and Escola Pública Dra. Estalella Graells for granting us access to and organizing the meetings with children. Also, many thanks to the children and their families for their participation in the experiments.

Thanks to all the members of the Laboratory in phonetics from CSIC, special thanks to Juana Gil Fernández, Victoria Marrero and Ana María Fernández Planas. I am indebted to APAL Lab members Ferran Pons, Jorgina Solé and Maria Teixido. I am grateful to my colleagues from the Faculty of Psychology and Education Sciences at the Universitat Oberta de Catalunya, especially to Llorenç Andreu, Teresa Guasch, Nati Cabrera, Teresa Romeu, Brigida Maestres and Gemma Abellán. Also many thanks to Mònica Sanz, M^a José Buj, Spyros Chistou, Laura Ferinu, Fernanda Pacheco and Nadia Ahufinger from the *Cognition and Language Research Group*.

Many thanks to anyone else with a connection to this thesis or who feel close to this work. Without all of you, it would have been much less fun.

ABSTRACT

In everyday interactions, speakers integrate gestures and speech sounds at a temporal level. One of the linguistic functions of temporally synchronous gesture-speech combinations is to provide prominence to specific parts of a discourse. While a bulk of evidence has explored the gesture-speech co-expressiveness at a semantic level, little is known about the children's ability to use synchronized gestural and prosodic prominences in the benefit of language. This PhD thesis investigates gesture-speech temporal integration abilities in development and its beneficial impact for children's language.

The dissertation includes three independent studies at different time points in development, each one described in one chapter. The first two studies aim at investigating the role of perceiving gesture-speech temporal synchronizations functioning as markers of prominence, and its linkage to language abilities. First, a study investigated whether three- to five- year- old children responded better to a word recall task when the word was presented with a contrast of prominence expressed with a synchronous beat gesture (i.e., a hand gesture synchronized with prominence in speech). The results indicated a beneficial local effect of the beat gesture on the recall of the temporally synchronous word. Second, a study examined whether six- to eight- year-old children processed

pragmatic inferences online more rapidly when the relevant information was presented together with a beat gesture. Additionally, this study investigated whether these potential benefits were due to the prominence expressed in the gesture or to its concomitant prosodic prominence. Results showed that children's processing of a pragmatic inference was positively improved by both prosodic and beat gesture prominence contributions to the discourse. The last study focused on the predictive role of the first infant's uses of temporally synchronous gesture-speech combinations on later language development. To do so, a longitudinal study correlated the infants' production of synchronous pointing gesture-speech combinations during controlled socio-communicative interactions at 12 months with linguistic measures at 18 months. Results demonstrated that synchronous productions positively correlated with lexical and grammatical development at 18 months of age.

Overall, the three studies show evidence that infant's synchronous gesture-speech abilities (a) function as multimodal markers of prominence; (b) when perceived in a discourse context synchronies have positive impact on children's word recall (Study 1) and pragmatic inference resolution (Study 2); and (c) infants' first productions of synchronous gesture-speech combinations serve a communicative strategy which is correlated to later language abilities (Study 3). The findings of the studies presented in this thesis point out the importance of synchronous gesture-speech

combinations in highlighting information, as well as their beneficial effects in language acquisition.

RESUM

En les interaccions quotidianes, els parlants integren temporalment els gestos i els sons de la parla. Una de les funcions lingüístiques de les sincronitzacions temporals de gest i parla és proporcionar prominència a parts concretes del discurs. Mentre que la major part d'estudis previs ha investigat la coexpressivitat entre gest i parla a nivell semàntic, se sap molt poc sobre la capacitat dels infants per utilitzar les prominències gestual i prosòdica sincronitzades en benefici del processament del llenguatge i la seva adquisició. Aquesta tesi doctoral investiga les habilitats de integració temporal entre gest i parla en el desenvolupament i el seu impacte beneficiós per al processament del llenguatge dels infants.

La tesi inclou tres estudis independents en moments diferents del desenvolupament, cadascun d'ells descrit en un capítol separat. Els dos primers estudis tenen com a objectiu investigar el paper de l'observació de les sincronitzacions de gest-parla com a marcadors de prominència, així com la seva relació en les habilitats del llenguatge. Al primer estudi s'investiga si els infants de tres a cinc anys recorden més les paraules en un discurs quan aquestes paraules es presenten amb un contrast de prominència mitjançant un gest rítmic sincrònic amb la parla (*beat gesture*). Els resultats indiquen un impacte local del gest, amb un increment del record només de la paraula, amb la qual està associada el gest. Al segon estudi s'avalua

si els nens de sis a vuit anys processen les inferències pragmàtiques en temps real més ràpidament quan la informació rellevant es presenta conjuntament amb el gest rítmic. A més, aquest estudi també investiga si aquests beneficis es deuen a la prominència expressada amb el gest o mitjançant la prosòdia. Els resultats mostren que el processament de les inferències pragmàtiques millora positivament amb les dues contribucions de prominència prosòdica i també gestual. L'últim estudi se centra en el paper predictiu dels primers usos lingüístics de les combinacions de gest i parla, sincronitzades temporalment en el desenvolupament posterior del llenguatge. Aquest estudi longitudinal demostra que hi ha una correlació entre l'ús, per part dels infants de 12 mesos, de la sincronització entre el gest d'assenyalar i la parla, i mesures lingüístiques als 18 mesos, específicament amb mesures de desenvolupament lèxic i gramatical.

En general, els resultats dels tres estudis mostren que la sincronització de gest i parla dels infants funcionen com a marcadors multimodals de prominència amb la funció de centrar l'atenció en posicions informatives importants. L'estudi de les associacions gest-parla dins d'un context discursiu ens ha permès observar que (a) es produeix un impacte positiu en el record de les paraules coordinades amb aquests marcadors multimodals (Estudi 1), (b) té un efecte beneficiós en processos de comprensió del llenguatge, com ara la resolució de inferències pragmàtiques (Estudi 2), i (c) les primeres combinacions de gestos sincronitzats amb parla funcionen com a una estratègia comunicativa que es correlaciona

amb habilitats posteriors del llenguatge (Estudi 3). En resum, els resultats dels estudis presentats en aquesta tesi posen en relleu la importància de les sincronitzacions de gest i parla com a marcadors de rellevància de la informació, així com els seus efectes beneficiosos en l'adquisició del llenguatge.

RESUMEN

En las interacciones cotidianas, los hablantes integran temporalmente los gestos y los sonidos del habla. Una de las funciones lingüísticas de las sincronizaciones temporales de gesto y habla es proporcionar prominencia en partes concretas del discurso. Mientras que la mayor parte de estudios previos ha investigado la coexpresividad entre gesto y habla a nivel semántico, se sabe muy poco sobre la capacidad de los niños para utilizar las prominencias gestual y prosódica sincronizadas en beneficio del procesamiento del lenguaje y su adquisición. Esta tesis doctoral investiga las habilidades de integración temporal entre gesto y habla en el desarrollo y su impacto beneficioso en el procesamiento del lenguaje de los niños.

La tesis incluye tres estudios independientes, cada uno de ellos descrito en un capítulo separado y centrado en un momento diferente del desarrollo. Los dos primeros estudios tienen como objetivo investigar el papel de la observación de sincronizaciones de gesto-habla como marcadores de prominencia, así como su relación con las habilidades del lenguaje. En el primer estudio se investiga si los niños de tres a cinco años recuerdan más las palabras en un discurso cuando estas se presentan mediante un contraste de prominencia producido a través de un gesto rítmico sincrónico con el habla (*beat gesture*). Los resultados indican un

impacto local del gesto, con un incremento del recuerdo solo de la palabra con la que está asociada el gesto. En el segundo estudio se evalúa si los niños de seis a ocho años procesan las inferencias pragmáticas en tiempo real más rápidamente, cuando la información relevante se presenta conjuntamente con el gesto rítmico. Además, este estudio también investiga si estos beneficios se deben a la prominencia expresada por el gesto o por la prosodia. Los resultados muestran que el procesamiento de las inferencias pragmáticas mejora positivamente con las contribuciones de prominencia prosódica y también gestual. El último estudio se centra en el papel predictivo de los primeros usos lingüísticos de las combinaciones de gesto y habla, sincronizadas temporalmente en el desarrollo posterior del lenguaje. Este estudio longitudinal demuestra que existe una correlación entre el uso que los niños de 12 meses hace de producciones sincronizadas gesto-habla y las medidas lingüísticas de desarrollo léxico y gramatical de esos mismos niños a los 18 meses.

En general, los resultados de los tres estudios muestran que la sincronización de gesto y habla funciona como un marcador multimodal de prominencia con la función de centrar la atención en posiciones informativas importantes. El estudio de las asociaciones gesto-habla en un contexto discursivo nos ha permitido observar que (a) que se produce un impacto positivo en el recuerdo de palabras coordinadas con estos marcadores multimodales (Estudio 1), (b) tiene un efecto beneficioso en procesos de comprensión del lenguaje, como por ejemplo en la resolución de inferencias

pragmáticas (Estudio 2), y (c) las primeras combinaciones de gestos sincronizados con el habla funcionan como una estrategia comunicativa que se correlaciona con habilidades posteriores del lenguaje (Estudio 3). En resumen, los resultados de los estudios presentados en esta tesis ponen de relieve la importancia de las combinaciones de gesto y habla como marcadores de relevancia de la información, así como sus efectos beneficiosos en la adquisición del lenguaje.

LIST OF ORIGINAL PUBLICATIONS

Chapter 2

Igualada, A., Esteve-Gibert, N. & Prieto, P. (2017). Beat gestures improve word recall in 3- to 5-year-old children. *Journal of Experimental Child Psychology*, 156, 99-112.

Chapter 4

Igualada, A., Bosch, L., Prieto, P. (2015). Language development at 18 months is related to multimodal communicative strategies at 12 months. *Infant Behavior and Development*, 39, 42-52.

TABLE OF CONTENTS

Acknowledgments.....	v
Abstract.....	ix
Resum.....	xiii
Resumen.....	xvii
List of original publications	xxi
Table of contents	xxii
Figures.....	xxv
Tables	xxvii
1. Introduction	1
1.1. Gesture and speech form an integrated system	1
1.2. Semantic co-expressiveness	5
1.3. Representational gestures are linked to language acquisition ..	11
1.4. Temporal synchronization between gestures and speech.....	18
1.5. Functional uses of multimodal markers of prominence in language	26
1.6. The development of prosodic and gestural prominence.....	30
1.7. Developmental benefits of temporally synchronized non- representational gestures	32
1.8. The benefits of gesture-speech temporal integration in language processing in development	38
a) Effects of beat gestures on word recall	38
b) Effects of beat gestures and its concomitant prosody on pragmatic inference abilities	39
c) Pointing gesture-speech temporal integration as a precursor or language	40
1.9. General objectives, research questions and hypotheses	41
2. Chapter 2. Beat gestures improve word recall in 3- to 5-year-old children.....	47

2.1.	Introduction	47
2.2.	Methods	56
	a) Participants	56
	b) Materials	58
	c) Procedure	63
	d) Coding	66
2.3.	Results	68
	a) Local word recall effects	68
	b) Global word recall effects	70
	c) Correlation analysis between word recall ability and age in months	71
2.4.	Conclusions	73
3.	Chapter 3. Benefits of beat gestures and prosody in children's online processing of pragmatic inferences	79
	3.1. Introduction	79
	3.2. Methods	85
	a) Participants	85
	b) Materials	86
	c) Procedure	96
	d) Analysis	97
	3.3. Results	99
	3.4. Conclusions	104
4.	Chapter 4. Language development at 18 months is related to multimodal communicative strategies at 12 months	109
	4.1. Introduction	109
	4.2. Methods	118
	a) Participants	118
	b) Experimental setting and materials	119
	c) Procedure	121
	d) Coding and reliability	125

4.3.	Results	130
a)	Effects of social condition on the use of pointing-speech combinations	131
b)	Predictive value of simultaneous pointing-speech combinations for expressive language outcomes at 18 months	135
4.4.	Conclusions	139
5.	General discussion and conclusions	147
5.1.	Summary of findings	147
5.2.	Temporal gesture-speech synchrony as a marker of prominence in language processing	150
5.3.	Effects of prosodic prominence and gestural prominence	155
5.4.	The temporal synchrony rule: A predictor of language acquisition	159
6.	References	163
	Appendix 1. Experiment 1 (Chapter 2)	183
	Appendix 2. Experiment 2 (Chapter 3)	189

FIGURES

Experiment 1 (Chapter 2)

Pg.

Figure 1. Example of a five-word list in both beat and no-beat conditions in which the central target word is underlined. The word in bold and capital letters was emphasized with a beat gesture in the accompanying video recording. 56

Figure 2. Still images from the stimulus video showing the actor producing the target word in the no-beat (left panel) and beat condition (right panel) while addressing Elmer the elephant. 60

Figure 3. Schematic representation of the procedure of the word recall task. 63

Figure 4. Mean number of items recalled in the target position as a function of condition and age. 66

Experiment 2 (Chapter 3)

Pg.

Figure 1. Still image from one of the stimulus videos which child participants were shown. In the center, the speaker is producing a beat gesture. Images in the four corners show a mouse (the target), duck (a competitor), door (item from the first sentence), and scissors (a distractor). 89

Figure 2. Examples of visual information and capture of the acoustic cues during the production of the specific clue *aire* ‘air’

in the control condition (left panel), the prosody-only condition (mid panel), and the beat+prosody condition (right panel). 90

Figure 3. Number of responses towards target and competitor areas of interest during TW2. 101

Experiment 3 (Chapter 4)

Pg.

Figure 1. Schematic representation of the central area within the testing room. The setting includes a curtain with six openings, three on each side, where six of the objects manipulated by an assistant hidden behind it were presented, two cameras (frontal and back position) and two additional objects placed in front of the curtain, on the floor, to the left and right of the experimenter. Locations of the experimenter, child, and caregiver are also indicated. 94

Figure 2. Mean number of occurrences of each type of communicative production=(speech-only, pointing-only, and pointing-speech combinations) per trial as a function of social condition (baseline, available, and unavailable conditions). Error bars: +/- 1 S.E. 107

TABLES

Experiment 1 (Chapter 2)

Pg.

Table 1. The sample population broken down into age groups, showing Mean (*M*) and Standard Deviation (*SD*) for age in months and memory span in number of words, as well as gender. 53

Table 2. The sample population separated into groups according to memory span in number of words and showing Mean (*M*), Standard Deviation (*SD*), median, minimum (*Min*) and maximum (*Max*) age in months, broken down by memory span. 54

Experiment 2 (Chapter 3)

Pg.

Table 1. The sample population broken down into age groups, showing Mean (*M*) and Standard Deviation (*SD*) for age in months, and gender..... 86

Table 2. P-values for the relevant main effects and interactions for the disambiguation of the AOI for all time windows (TW). 101

Experiment 3 (Chapter 4)

Pg.

Table 1. CDI scores of infants at 18 months as reported by parents.	112
Table 2. Multiple regression analyses of the most effective models predicting infants' vocabulary and morphosyntax measures at 18 months based on early communicative productions at 12 months during a specific social condition. R ² statistics and p-values are reported for each model.....	113

1. INTRODUCTION

1.1. Gesture and speech form an integrated system

In the last few decades, gesture studies have started to investigate the interplay between gestures and speech as being part of the same linguistic system. Within this field, gesture and speech have been argued to be closely linked, forming an integrated system (e.g., Kendon, 1980; McNeill, 1992; McNeill, 2005; Kelly, Ozyürek & Maris, 2010). According to McNeill (1992), gestures are defined as communicative acts which are closely related with the speech stream at different levels of linguistic analysis. The author suggested that gestures are co-expressive with speech at the phonological, semantic and pragmatic level and proposed the following three rules of synchronization between gesture and speech.

- The phonological synchrony rule predicts that prominences of gesture and speech are temporally aligned.
- The semantic synchrony rule predicts that gestures and speech are related to the same idea or concept.
- The pragmatic synchrony rule predicts that both modalities share the intentional function.

Gestures have traditionally been differentiated by their form and function. Following McNeill (1992) I will describe the four types (or dimensions) of gestures: representational, deictic, conventional, and last but not least, beat gestures.

Representational gestures (including iconic and metaphoric gestures) represent objects, actions, or relations by coding an aspect of their referent's shape or movement. Iconic gestures can express tangible information (e.g., representing the shape of an object with the hands, like producing a cup hand gesture to represent a glass). Metaphoric gestures are related to more abstract concepts (e.g., touching the head with two fingers to represent the verb "think", or moving the arm forward while producing circular movements with the hand to represent the concept "future").

Deictic gestures, such as pointing gestures, are hand and arm gestures which serve to direct attention toward a specific object or event of reference in the surrounding environment. These gestures include requesting (extending the arm toward an object, location or person, sometimes with a repeated opening and closing of the hand), showing (holding up an object in the adult's line of sight), giving (transferring an object to another person), and pointing (index finger or full hand extended towards an object, location, person or event).

Thirdly, conventional gestures are culturally shared symbols, with an arbitrary form and meaning within a given community. For example, a palm up hand gesture shaking with repetitive movements from right to left can serve to warn someone that he/she did something wrong according to the gesturer's opinion, with the potential consequence that there will be some kind of punishment. Another example would be a common “hi” hand gesture.

Finally, beat gestures have been defined as rhythmical movements of the hands which are timed together with speech, specifically with prosodic prominence in speech (e.g., Loehr, 2012; Shattuck-Hufnagel, Ren, Mathew, Yuen & Demuth, 2016). Beats have been typically described as not having a clear semantic meaning (i.e., non-representational gestures). In contrast with representational gestures, beat gestures do not add propositional content to a given utterance (McNeill, 1992, 2005; Kendon, 1995) but are rather used to mark “the word or phrase they accompany as being significant (...) for its discourse pragmatic content” (McNeill, 1992:15). Beat gestures have been associated with functions typically linked to prosody, such as, focus marking (i.e., prominence) and discourse structure marking (e.g., Loehr, 2012; Wagner, Malisz, & Kopp, 2014; Shattuck-Hufnagel et al., 2016). Research has shown that beat gestures are temporally synchronized with prosodic markers of prominence (i.e., pitch accents) (Yasinnik, Renwick & Shattuck-Hufnagel, 2004; Jannedy & Mendoza-Denton, 2005; Loehr, 2012; Shattuck-Hufnagel, et al., 2016). In addition, an increase in

prominence perception has been reported when words are produced together with hand gestures (Krahmer & Swerts, 2007) and head/facial beat gestures (Moubayed, Beskow & Granström, 2010). Typically, the movements of hand gestures occur together with head and eyebrow movements, which together signal the privileged status of a given piece of discourse in a multimodal fashion.

McNeill's classification has been extensively used in gesture research (for example, see Gullberg, DeBot & Volterra, 2008, or Cartmill, Demir & Goldin-Meadow, 2012, for a review). Even though gesture research has typically used this classification as a close categorization of gestures, examples are frequently found where more than one dimension can be observed at the same time. Regarding this conceptual issue on the use of this classification as dimensions that interplay rather than closed clusters, McNeill (1992) provided the following example: a pointing gesture may be temporally aligned with prosodic patterns and also act as a beat gesture. Thus, according to this view any type of gesture associated with a prosodic prominence might be functioning as a multimodal marker of prominence (i.e., a beat gesture).

1.2. Semantic co-expressiveness

A relevant question within the field of gesture research is to what extent are gestures and speech co-expressing semantic information (e.g., the semantic synchrony rule by McNeill, 2012). In this section, we summarize the formalizations proposed by theories and authors of the gesture-speech integration system at the semantic level. Various proposals postulate that the connection between speech and gesture occurs at different levels of depth and provide evidence that gestures are co-expressive with speech at the semantic level. For instance, the Growth Point Theory by McNeill (1992, 2005) and the Interface Hypothesis by Kita and Özyürek (2003) consider gesture and speech modalities to be an integrated unit functioning as a single system in communicative acts (see also Goldin-Meadow & Alibali, 2013; Gullberg, et al., 2008; and Wagner, et al., 2014 for a review).

According to the Growth Point Theory (McNeill, 1992, 2005; McNeill & Duncan, 2000; McNeill, Duncan, Cole, Gallagher & Bertenthal, 2008) gestures and speech function together in a single and integrated system. McNeill et al. (2008) pointed out that “gestures are integral components of language, not merely accompaniments. They are semantically and pragmatically co-expressive with speech, not redundant” (McNeill et al., 2008: 118). Gestures express global-synthetic meanings integrated in single units together with speech which articulate linear and segmented

representations of the information. Evidence for this view was provided from the analysis of gestures which are incomplete without speech accompaniment ('gesticulation', co-speech gestures) from the description of synchronous gesture-speech productions in situations in which the speaker cannot see its own hands, which results in full maintenance of synchronicity and co-expressiveness (see section 1.4 below).

An interesting question that has been addressed by many theories is the degree of interaction between gesture and speech. The Information Packaging Hypothesis (Kita, 2000) holds that gestures and speech interact at a level in which information is packaged. An extension from the two models cited above, the Interface Model Hypothesis (Kita & Özürek, 2003), argues that gestures and speech are two different subcomponents which interact in bidirectional ways at the time that information is being packaged. Bidirectionality is supported by evidence coming from cross-linguistic comparisons. Kita and Özürek (2003) assessed whether complex concept motion events which are underspecified in the linguistic code (verbal modality) of a language, and specified in others, might show differences on gesture production. They showed that gestures were more frequently used to express the motion events when the linguistic code was underspecified and when spatial information was not expressed in speech. They therefore claimed that gestures offered visuo-spatial information helping speakers to select specific items for verbalization. This is supported by empirical findings that show that in order to produce an

appropriate message, speakers will differently produce gestures depending on the message expressed on the verbal modality. For example, Hostetter, Alibali & Kita (2007) tested this hypothesis by asking participants to describe visuo-spatial information varying on easy (i.e., shapes were provided visually) or complex (i.e., visual information was not provided) contexts. Their results showed that participants gestured more frequently in complex visuo-spatial description tasks. Another example is provided by results in Hostetter and Alibali (2007) in which gesture productions increased for those participants who had strong visuo-spatial skills and weak verbal skills. Thus, the underlying rationale is that gesture production increases when the load of conceptual information is higher, and that the spatial and verbal abilities are tightly related to gesture production.

From the listener's point of view, the Integrated System Hypothesis explores gesture-speech integration in relation to language comprehension (Kelly, et al., 2010). Specifically, they claim that gestures and speech are integrated through mutual and obligatory interactions in language comprehension. This is supported by evidence from the two experiments included in Kelly et al. (2010). In Experiment 1, adults were exposed to video primes of an action (e.g., chopping vegetables), followed by targets which were representational gestures produced together with an isolated verb (i.e., a one word utterance). There were four targets varying in one modality, in which at least one of the modalities was congruent to

the action. There were two variations in the strength of incongruence (i.e., strong vs weak incongruities). There was a congruent condition (speech: “chop”; gesture: “chop”), a strongly incongruent speech condition (speech: “twist”; gesture: “chop”), a weakly incongruent speech condition (speech: “cut”; gesture “chop”), a strongly incongruent gesture condition (speech: “chop”; gesture “twist”), and a weakly incongruent gesture condition (speech: “chop”; gesture “cut”). The accuracy of responses showed significantly fewer errors in congruent responses irrespectively of the condition. These results drove the authors to conclude that both modalities are mutually influenced.

In their second experiment, Kelly, et al. (2010) replicated experiment 1 with different instructions. Participants this time were prompted to only fixate on speech information, by asking them to only say whether speech content in the targets was the same or different from the primes. The results in Experiment 2 showed that participants increased the error rates when the gestures were incongruent even when the task did not include instructions to attend to gestural information. The authors concluded that both modalities showed similar patterns of influence (i.e., mutual influence) (Experiment 1) and that there was an obligatory integration between modalities (Experiment 2).

All in all, the theories cited above support the hypothesis that gestures and speech form an integrated system by providing

evidence related to the integration of representational gestures and speech at a semantic level, both at the production and comprehension levels. These models have thus dealt with the processing of representational gestures, and the exploration of other synchrony rules (temporal and pragmatic) rather than the semantic synchrony rule have been outside the scope of these models.

Interestingly, there is one model, the Gesture-As-Simulated-Action Framework, in which motor actions are included as a relevant feature linked to language model (Hostetter & Alibali, 2008, 2010). These authors postulate that the production and perception of actions (i.e., movements done without a communicative purpose) are related to gestures, and gestures are activated by simulations of actions and mental imagery. Thus, the prediction in this theory is that embodied experience (actions) will affect gesture production. In Hostetter and Alibali (2010) participants described configurations of dot patterns to assess a potential change on gesture production when they previously recreated the patterns with physical shapes (action condition), or just observed them (observe condition). The results confirmed their hypothesis by showing a greater amount of representational gestures on their descriptions in the action condition than in the observe condition. However, the production of beat gestures did not change significantly between conditions. Thus, the physical experience to recreate actions only showed a significant impact on the production of representational gestures. Supporting evidence for the Gesture-As-Simulated-Action

framework is provided by Cook, Yip and Goldin-Meadow (2010). In their study, adults were asked to describe images that involved spatial movements and actions were assessed, with results showing that promoting the use of gestures during descriptions improves immediate recall and 3-weeks-delayed free recall. All in all, the embodied cognition paradigm (Barsalou, Simmons, Barbey & Wilson, 2003; Barsalou, 2008) applied to the field of gesture studies states that representational gestures are an expression of action simulations (Hostetter & Alibali, 2010).

In this thesis, we suggest that the impact of beat gestures, which can be conceived as gestures highlighting relevant parts of information but with a less strong semantic component than representational gestures, can have a strong impact in language production and comprehension processes. The role of synchronicity between beat gestures and speech on several language processes will be the focus of this PhD thesis.

1.3. Representational gestures are linked to language acquisition

This section will provide evidence on the linkage between gestures and language abilities from a semantic point of view. In development, gestures expressing a semantic or referential meaning, e.g. representational gestures, as well as pointing gestures, have been shown to have a pivotal role at several stages of the language acquisition process. The study of early word and gesture production in relation to the meanings conveyed in each modality was the goal in Butcher and Goldin-Meadow's (2000) study, which assessed longitudinal observations in spontaneous play situations in the interval from the one-word stage to the two-word stage. They found that at the beginning of the one-word period gestures were not combined with speech, but at the beginning of the two-word period children started producing gesture-speech combinations referring to only one meaning (i.e., complementary combinations). That is, children produced gestures that conveyed the same information as the accompanying word, e.g., *cookie + point at cookie*. Infants' complementary combinations of nouns and pointing gestures were analyzed in Cartmill, Hunsicker, and Goldin-Meadow (2014). Their hypothesis was that these combinations served a specifying function of the gesture that referred to that specific object "noun" and not to another. These complementary combinations, which could have been assumed to be redundant in their meaning, actually served a determiner function. That is, their results showed that early productions of complementary gesture-speech combinations were

correlated with earlier uses of determiners in nominal constituents (e.g., *the car*) later in development.

Research has also shown that supplementary semantic combinations of gesture and speech can also function as a predictor of syntactic development. For example, children might say “want” while pointing to a cookie. This type of gesture-speech combinations (i.e., supplementary combinations) convey two different meanings expressed in two different modalities. Interestingly, children’s supplementary gesture-speech combinations have been correlated with the later ability to produce two-word combinations in the oral modality (Goldin-Meadow & Butcher, 2003; Iverson & Goldin-Meadow, 2005; Iverson, Capirci, Volterra & Goldin-Meadow, 2008). Changes on the early ability to integrate supplementary combinations of semantic features in speech and gesture modality predict different construction types (Özçalışkan & Goldin-Meadow, 2005). In Özçalışkan and Goldin-Meadow (2005) children were observed longitudinally at 14, 18 and 22 months. First, the results showed that the number of supplementary combinations increased over time. Second, the early ability to combine certain supplementary constructions predicted the use of corresponding constructions in the verbal modality. For example, a two arguments production in supplementary combination (e.g., “mommy + *couch* referred with pointing gesture”) would predict the production of two arguments combinations (e.g., “mama chair”) in the verbal modality. Moreover, this prediction occurred for different argument complexities (e.g., three argument combinations, predicate +

argument combination, predicate + two arguments combinations, and predicate + predicate combinations with or without arguments). Thus, gestures allow children to produce increasingly complex meanings when they are not able to produce them in the verbal modality yet. Another example is provided by Murillo, Galera and Casla (2015), in which 24- to- 34- month- old children were involved in a game detecting the odd picture amongst a set of five pictures. Trials increased in complexity in terms of the property (i.e., color, size and spatial location) of the elements in the pictures. Similarly to previous studies, children functionally distributed the semantic information in the gesture modality depending on message complexity, with greater uses of gesture utterances in more complex situations.

Many studies have explored the positive impact of representational gestures in cognitive processing abilities in children. In Goldin-Meadow, Cook and Mitchell (2009), 8-9 year-old children were taught how to solve a mathematical problem in three different conditions depending on the gesture information, namely (a) the correct gesture condition, in which the use of a gesture explained how to group together two numbers, (b) the incorrect gesture condition, involving the same gesture but referring to non-relevant numbers, and (c) the speech-only non-gesture condition. The results showed that children made use of the gesture information in both gesture conditions to help them to solve the mathematical problems. Children performed better in the correct gesture condition, but the

gesture information was beneficial for learning even when used incorrectly. One of the main arguments to explain these results is that the information embedded in the manual modality may help cognitive processes, such as reducing working memory load.

Representational gestures also have been found to help memorization in second language learning by children (Tellier, 2008). In this study, twenty 4 to 5 year-old children were asked to remember words in two conditions, namely, a verbal utterance accompanied by a representational gesture of the word, or else accompanied by a picture image. The results showed that gestures and especially producing them increased children's memorization of words in a second language. The production of gestures which represent some kind of semantic information has been claimed to help reference learning, but less is known about whether gestures which do not express referential semantic information, such as beat gestures, have an effect on language processing.

In the domain of comprehension, the ability to integrate semantic features from gesture and speech modalities seem to follow specific developmental patterns. Sekine, Sowden and Kita (2015) assessed the children's ability to integrate co-expressive semantic features in the speech and gesture modalities. In this study, 3-to-5 year-old children's were assessed with recordings of an actor saying a short sentence with verb (speech-only condition), showing a representational gesture (gesture condition) or a combination of

both (multimodal condition). Crucially, information in the gesture modality expressed semantic features that constrained the meaning of the verb expressed in the verbal modality (e.g., saying the verb “throwing” while performing this gesture with a concrete type of object, such as a basketball). To assess children’s integration of semantic information from both modalities, first children saw the priming information (i.e., video recordings in one of the three conditions). Then they were asked to choose the best match among four pictures. The results showed that 5-year-old children and adults integrated better the information expressed with both modalities than 3-year-old children. Older children and adults were more capable of integrating semantic features from gesture and speech than younger children. In a second experiment, only 3-year-old children were assessed with a similar task but in live interaction with the experimenter. In this case, younger children benefited from the integration of both modalities when having to choose the correct answer during natural interactions with the adult. According to the authors, the results of this study support a developmental pattern occurring between age 3 and 5 in the ability to semantically integrate the two modalities (as previously proposed by Ramscar & Gitcho, 2007).

Regarding representational gesture’s integration to speech in older children, Sekine and Kita (2015) assessed 5-, 6-, 10-year-olds and adults in their ability to integrate multimodal information in narratives. In this case, participants watched videos of a speaker

narrating short stories (three sentences) while producing abstract representational gestures with a cohesive function. This gesture served to consistently refer to characters by locating the hand into one of two positions, one for each of the two characters in the story. In the first and second sentences of the story, the name of each character was produced while referring to its respective place in space, and in the third sentence the subject was omitted in the speech modality but referred to with the gesture. Thus, participants could only know who performed the target action mentioned in the third sentence by making an inference from the gesture. The results of the experiment showed that adults and 10-year-old children integrated the information significantly better than 5-year-old children. In a second experiment in Sekine and Kita (2015), the authors replicated Experiment 1 but by including pictures in the video which served to identify the characters. Only 5-year-old children were assessed in this experiment, with results showing better performances in this task than in the previous one. The results from these studies show that the ability to integrate gestures and speech at a semantic level changes across development, and depends on the complexity of the task.

Older children benefit from the use of representational gestures in more complex tasks too. Kirk, Pine and Ryder (2011) investigated whether information conveyed by representational gestures helped 7- to 8- year- old children with language impairment (and age-matched typically developing children) to understand a verbal message. Participants listened to short narrations in speech-only and

speech + gesture conditions in which they had to make an inference beyond what was explicitly stated verbally in the scenarios.¹ The results showed that children's comprehension, particularly for those with language impairment, was helped by representational gestures.

To conclude, research has extensively demonstrated that children across different age spans can successfully integrate information from representational gestures in different tasks. Moreover, children's ability to integrate gestures and speech at a semantic level is related to various linguistic abilities. However, scarce literature has investigated the potential beneficial effect of non-representational gestures (e.g., the 'beat gesture' dimension) on language abilities (see section 1.7 below for a summary). This thesis will test whether temporally synchronized gestures and speech functioning as multimodal markers of prominence have a potential effect on language processing (i.e., specifically on word recall, pragmatic inference resolution and early language acquisition).

¹ This is an example of a story with an inference from Kirk et al. (2011). "Freddie helped his dad **paint the bedroom**. Freddie had to put on his old clothes. Why did Freddie have to **put on his old clothes**?" *Gestures*: Right hand performing a painting action indicating paint the bedroom, and both hands come towards the body and down indicating outing clothes on; in this case an example of a correct answer would be: "Because his clothes that were very nice would get dirty and old clothes don't matter".

1.4. Temporal synchronization between gestures and speech

In everyday interactions, speakers integrate gestures and speech sounds at a temporal level. As mentioned before, the temporal synchrony between speech and gesture has been used as evidence of an integrated spoken language and gesture communication system (e.g., Wagner et al. 2014; Rusiewicz, Shaiman, Iverson & Szuminsky, 2013; Iverson & Thelen, 1999; McNeill, 1992; Rusiewicz & Esteve-Gibert, in press). This section overviews production and perception research reporting evidence on the natural mechanism to use gestures and speech in temporal synchrony in adult and infant populations.

McNeill (1992) suggested that the prominent parts of the gesture (such as the stroke phase of the gesture, which coincides with the interval of maximum effort in the gesture, as well as the apex of the gestures, which is the moment of greatest effort within the stroke phase) are relevant parts in the gestures anchoring to speech sounds. A variety of studies on the alignment of gesture and speech looking at precise measures of temporal coordination have shown that prominent parts of gestures and speech occur in tight synchrony. (e.g., De Ruiter, 2000; Esteve-Gibert & Prieto, 2013; Loehr, 2012; Rochet-Capellan, Laboissière, Galván, & Schwartz, 2008; Rusiewicz, 2010; Yasinnik, et al., 2004; see Wagner et al., 2014 for a review). For example, Yasinnik et al. (2004) showed that during conference talks between 60% and 90% of instances of the gesture

apexes occurred together with a pitch-accented syllable (see also Jannedy & Mendoza-Denton, 2005 for a review). Loehr's (2012) analysis of adult narrations showed that prominent accentuations at the intonation phrase level (i.e. pitch accents) were systematically coordinated at temporal level with the stroke of gestures. The temporal realization of intonation and pointing gestures was been investigated by Esteve-Gibert and Prieto (2013) with fifteen adults using a pointing-naming task.

They tested whether prominent parts of the pointing gesture (i.e., the stroke and the apex) and speech (different positions in the accented syllable, as well as the fundamental frequency peak of the rising pitch accent) were synchronized. Their results showed that the peak of the contrastive focus pitch accent was the most stable anchoring point for the onset of the stroke of the pointing gesture. All in all, even though there are some differences between studies on the exact sites of temporal alignment, there is a general consensus on the fact that the prominent part of the gesture, i.e., the stroke phase and the apex of the gesture, is temporally synchronized with the prominent parts in speech (i.e., pitch accents) (see Rusiewicz & Esteve-Gibert, in press for a review).

Furthermore, the synchronization between rhythmic hand movements and speech sounds have been shown to develop since early stages in development. At around 6 months of age, infants

tend to coordinate their hand/arm actions and vocalizations. Interestingly, this coordination seems to increase at the onset of the babbling period (Ejiri & Masataka, 2001; Iverson & Fagan, 2004). From a developmental perspective, the ability to temporally coordinate gesture and speech starts to be acquired with the emergence of referential communication, typically between 9 and 12 months of age. During this period, communicative gestures (e.g., pointing and reaching gestures) are more easily activated with a referential meaning than vocalizations. An important developmental period for gesture-speech interplay occurs when children begin to use synchronous gesture-speech combinations intentionally.

Dynamic System Theory provides a developmental framework in which gestures and speech develop together during the process of language acquisition (Parlade & Iverson, 2011; Iverson & Thelen, 1999). This model aims to explain how gestures and speech interrelate in development. According to the authors, there is a link in production between early hand-mouth coordination at a sensory-motor stage of an infant's development with more intentional and synchronized multimodal productions. This model focuses on understanding "how the dynamics of change in strength and stability of early vocal and motor skills can account for the emergence of the ability to link the two modalities in a single, coordinated behavior with common communicative intent" (Iverson, 2010:6). Thus, speech and gesture are aligned as a function of underlying motoric pulses. Findings such as those in Ejiri and Masataka (2001) and Iverson and Fagan (2004) show

evidence for the link between the early ability to coordinate rhythmic hand and arm movements and reduplicated babbling productions. They suggest that this early vocal-motor coordination seems to be a precursor of more mature pointing gesture-speech combinations. See Rusiewicz, Shaiman, Iverson and Szuminsky (2014) for evidence supporting this proposal with adult population regarding the pulse-based temporal entrainments of pointing gestures with regard to perturbation of speech, prosody and position of the target syllable.

Synchronous multimodal productions begin to occur near the end of the first year of life, a few months after the onset of canonical babbling (Butcher & Goldin-Meadow, 2000, Esteve-Gibert & Prieto, 2014; Murillo & Belinchón, 2013). Murillo and Belinchón (2013) analyzed productions of gestures, vocalizations and gesture-vocalization coordinations produced by eleven infants recorded longitudinally from 9- to 15- months of age. The results showed that the gesture-speech coordination (mostly reaching and pointing gestures) increased at 12 months of age, and continued with similar rates until 15-months of age, while vocalizations produced alone progressively decreased in their frequency rate between 9 and 15 months of age. The results pointed to a dynamic transition from the use of isolated modalities to the combination of both gesture and speech modalities. Esteve-Gibert and Prieto (2014) analyzed pointing gestures temporally aligned in their prominent parts (i.e., stroke phase) with speech prominences (i.e., accented syllable

boundaries). Data was extracted from a longitudinal sample of four infants recorded in naturalistic interactions with their caregivers. The results showed that infants already produced temporally synchronous gesture-speech combinations at the age of 11 months. Moreover, the sample of gesture-speech productions increased significantly at 15 months of age, and the time lag between gestures and prominent parts of speech was produced in similar time lags to adults. Murillo and Capilla, (2016) showed that acoustic features which have been extensively reported in literature as precursors of mature syllables (Oller, Niyogi, Gray, Richards, Gilkerson, Xu, Yapanel & Warren, 2010) are more frequently produced together with an deictic gesture. This was true for the gestures produced with an imperative intention but not for those gestures produced with a declarative function. Moreover, vocalizations synchronous to gestures with a declarative function had a significantly greater fundamental frequency than isolated vocalization with a declarative intention, or vocalizations with an imperative function with or without a gesture.

Unsurprisingly, both children and adults have been shown to be sensitive to temporal manipulations of gesture-speech combinations. Sensitiveness to gesture-speech temporal alignments have been assessed by creating asynchronies between prominent parts of both modalities. Leonard and Cummins (2011) found asymmetries in the sensitivity of listeners to the relative timing of beats gestures and speech. In their perception study, adult participants were exposed to temporally desynchronized

combinations of gestures and speech. Adults detected 200 ms desynchronized late gestures but had problems detecting early desynchronized productions. To our knowledge only one study has assessed the sensitivity to the temporal alignments between pointing gestures and speech in infant population. Esteve-Gibert, Prieto and Pons (2015) found that 9-month-old children were sensitive to misalignments of the apex of the pointing gestures and the stressed syllables of the word. Thus, infants are sensitive to gesture-speech alignment patterns well before they are able to produce them.

A few studies have investigated whether infants pointing gesture-speech synchronizations have a potential relation with later language abilities. In Murillo and Belinchón (2012), parent-infant dyads were recorded whilst interacting in a semi-structured play context at three longitudinal moments, namely at 9, 12, and 15 months. The results showed that the use of pointing gestures at 12 months, especially when accompanied by vocalizations and directed gaze on the part of the infant, correlated positively with vocabulary development at 15 months of age. In Wu and Gros-Louis (2014) the analysis of spontaneous interactions of 10- to 13-month-old infants with their mothers in fifty-one dyads showed that the infants' combinations of vocalization and pointing, and especially those produced when mothers were not attending to the target event, were related to the infants' subsequent comprehension skills at 15 months. All in all, evidence suggests that temporal alignment between pointing gestures and speech are produced (and perceived)

early in infancy, and this ability has been shown to be related to language abilities. However, to our knowledge no experimental work controlling the infant's pragmatic motivation to produce this temporal combination has been carried out.

On the other hand, later in development limited research has shown behavioral evidence on the effects of speech coordinated beat gestures on language abilities with opposite results. So, Chen-Hui and Wei-Shan (2012) found that while adults benefited from both beat and representational gestures in a word recall task, children benefited from representational gestures but not from beat gestures. The authors suggest that one possible explanation for the differences between adults and children would be a potential cognitive demand of the beat gestures. A second, and not excluding explanation of the differences might be methodologically motivated. First, in the experiment the words were presented in a list without context and thus without a pragmatically relevant discourse. Second, beat gestures were associated to every word in the list. This experimental manipulation might have reduced the potential effects of beat gestures, which provide prominence to a selected item and makes it more prominent in comparison with the surrounding elements. In another study, Austin and Sweller (2014) found that beat gestures did facilitate the recall of spatial directions in 3- to 4-year-old children. However in this experiment, the visual accessibility of the referents through the toy representation might have played a facilitating role in the recall of both location and action target words.

Overall, adults and children use and are aware of the temporal alignments of gestures and speech synchronicities. Although some studies have shown positive relations between language abilities and temporally synchronous gestures-speech combinations, further investigation is needed to know whether these temporal combinations function as multimodal highlighters of information. Following the evidence presented in this section, and in the following two sections, this thesis proposes that in the case of non-representational synchronous gesture-speech combinations, gestures are motor pulse-based entrainments of speech prominences (i.e., prosody). This thesis will test the proposal that these multimodal markers of prominence are developmentally rooted in early infants' ability to intentionally coordinate rhythmic limb motor impulses and oral sounds (see section 1.4 for a description of the Dynamic System Theory).

1.5. Functional uses of multimodal markers of prominence in language

In this section, we review the functional similarities between visual prominences expressed through beat gestures and prosodic prominences expressed through pitch accentuation. Recent research exploring the relationship between prosodic and gestural prominence has shown that prosodic prominence in speech (i.e., expressed through prosodic pitch accents) is typically produced simultaneously with more prominent gestural and articulatory features (such as head nods, eyebrow movements, beat gestures, exaggerated articulation, etc. (see, for example, Dohen, 2009; Ekman, 1979; Shattuck-Hufnagel et al., 2016; Swerts & Kraemer, 2008).

As mentioned before, prosody signals different functions that have been extensively studied across different language types. Among others, prosody functions to structure speech stream into different levels of organization (e.g., word segmentation, syntactic phrase boundaries, and discourse structure), as well as speaker-specific information (e.g., speakers' identity, speech style). Prosody can also signal pragmatic meanings related to the speakers' intentions, such as marking a phrase as a statement or a question or expressing degrees of commitment about the information expressed (e.g., Cole, 2015; Ladd, 2008; Pierrehumbert & Hirschberg, 1990). Prosody may also perform an important semantic function, such as signaling informational focus (or novelty marking) or contrastive focus.

Across languages, focus marking has been typically related to the use of pitch accentuation associated with the relevant words (Cole, 2015; Ladd, 2008; Pierrehumbert & Hirschberg, 1990).

Evidence increases on the need of studying beat gestures (or the beat gesture dimension) from a perspective that integrates those gestural markers with prominence functions related to prosody. Interestingly, beat gestures in natural speech have been reported to encode discourse structure functions, such as prosody. In an example, Guellaï, Langus and Nespors (2014) assessed whether gestures accompanying speech can also signal phrase boundaries, a typical prosodic function. The linguistic stimuli in this study were ambiguous sentences which could differ in their meaning depending on where the prosodic boundary was placed. Video recordings were manipulated to create mismatches between the alignment of prosody and gestures used in a discourse structuring function (chunking sentence). The results showed that adults tended to more often choose the meaning signaled by gestures during mismatches of the ambiguous sentences. The authors conclude that prosody seems not to be the exclusive marker of the prosodic phrasing function. Similarly, beat gestures have been associated with functions typically linked to prosody, such as focus marking (i.e., prominence) and discourse structure marking in both adult populations and children (e.g., Loehr, 2012; Wagner et al., 2014; Shattuck-Hufnagel et al., 2016) (See section 1.1 for the definition of beat gestures).

Regarding the prominence function, perception studies on this field have revealed that misalignments between visual beats and prosodic pitch accents have a detrimental effect on the perception of prominence (e.g., Leonard & Cummins, 2011; Treffner, Peter, & Kleidon, 2008). Krahmer and Swerts (2007) found that the presence of beat gestures has a significant effect on the perceived prominence independently from prosodic prominence. In their experiment, participants were exposed to the sentence “*Amanda goes to Malta*” in which the two target words were associated with combinations of the presence or absence of prosodic and gestural prominence. When participants saw a manual beat gesture on the focused word, and regardless of the presence or absence of prosodic prominence, this increased its perceived prominence and decreased the prominence of another target word. However, the strongest effect was produced by the multimodal combination of prosodic prominence and visual prominence. Thus, there is a strong perceptual connection between the presence of prosodic prominence (or pitch accentuation) and manual, facial and head beat gestures (Krahmer & Swerts, 2007; Moubayed, et al., 2010).

One of the linguistic functions of beat gestures is thus related to the signalling of prosodic focus, which has been compared to a yellow highlighter, that is, to emphasize information in the speech stream. McNeill (1992:15) stated that “the semiotic value of a beat lies in the fact that it indexes the word or phrase it accompanies as being

significant (. . .) for its discourse pragmatic content’’. As in the case of prosodic prominences, beat gestures typically align new information of the discourse context and highlight it. For example, when enumerating the features of a newly introduced character in the story: “his GIRLfriend, ALice, Alice WHIte” (example from McNeill, 2012).

In this thesis we are interested in exploring beat gestures as the gestural realization of prosodic accentuation as a marker of prominence. That is, to make particular items stand out from their surrounding non-prominent elements (e.g., Terken, 1991; Wagner, et al. 2015). We will test the effects of multimodal marking of prominence as a focus signaling strategy related to prosody. Based on previous evidence we expect that prominence provided by temporal gesture-speech synchronicities will positively impact on the development of language processing.

1.6. The development of prosodic and gestural prominence

In their review of the development of prosodic prominence, Speer and Ito (2009) proposed that the use of intonational prominence in childhood serves two main functions, novelty marking and contrastive focus. Intonation prominence can mark novel information with respect to previously known information (topic). The presence of prosodically exaggerated features in child-directed speech has been shown to facilitate reference resolution and word learning. Speer and Ito (2009) noted that children also make use of intonational prominence for contrastive information marking, and this may favor reference resolution. The use of prominent pitch accents casts a contrastive relationship between the entity with a prominence and its alternatives, making the information about alternatives more accessible. This contrastive relationship helps word resolution, as the scope of the contrast gains importance. Fernald and Mazzie, (1991) found that new information was marked with prosodic acoustic cues to a significantly greater extent in child-directed speech than in adult-directed speech. Grassmann and Tomasello (2007) used an eye-tracking methodology to compare how 2-year-old children's fixations to referents that were given relative to those that were new in the context varied depending on the placement of prosodic prominence. The results showed that children looked longer at new referents only when they were associated with prosodic highlighting (a LH* pitch accent).

On the other hand, research into the first uses of gestural prominence (e.g., beat gestures) during language development is fairly limited. Some production studies point out that children start producing beat gestures between the ages of 5 and 7, as their narrative skills develop (e.g., Colletta, Guidetti, Caprici, Cristilli, Demir, et al., 2015; Mathew, Yuen, Shattuck-Hufnagel, Ren & Demuth, 2014; Shattuck-Hufnagel et al., 2016). Mathew et al. (2014) showed that 6-year-old children produced beat gestures following similar patterns of temporal alignments than those produced by adults. Colleta et al., (2014) analyzed narrations of 5- and 10- year- old children. Their results showed that children increased the length and complexity of narratives with age. Regarding gesture production, while the general amount of gestures did not increase with age, the rate of beat gestures was significantly greater in older children. However, little is known about how children acquire the semantic-pragmatic functions associated with multimodal markers of prominence and whether they can use them in language memorization and comprehension processes.

1.7. Developmental benefits of temporally synchronized non-representational gestures

The previous sections have reported evidence on integration of the gesture by coupling its prominent parts (i.e, stroke phase and apex of the gesture) with prominent parts in speech (i.e., pitch accents and boundaries of the stressed syllables) (see section 1.4). This temporal integration has been related to perception of prominence when they are naturally aligned (see section 1.5). Studies on multimodal integration have shown how the coupling of visual input with oral speech sound provides redundancy of the event in a unitary instance, leading to an increase in intelligibility (Lewkowicz & Hansen-Tift, 2012; van Wassenhove, Grant & Poeppel, 2007) and gesture-speech sound integration (Esteve-Gibert, et al., 2015; Krahmer & Swerts, 2007). The ability to integrate unitary cues in a redundant way is proposed to be an adaptive strategy related to language comprehension. In fact, research with children with specific language impairments has suggested that children's difficulty in processing linguistic information might be related to difficulties in integrating multimodal cues of speech perception (Pons, Andreu, Sanz-Torren, Buil-Legaz & Lewkowicz, 2013). In line with these results, according to the Dual Coding Theory, memory traces are enhanced by the integration of multimodal channels, in this case redundancy between information conveyed by representational gestures and verbal information (Clark and Paivio, 1991). Moreover, we know from section 1.3 that semantic integration of representational gestures and speech follow specific

patterns in development related to different language production and comprehension processes. In this thesis, we would like to test whether the prominence function expressed through temporal gesture-speech alignments has an impact on language processing.

Evidence from studies assessing neurological activations during observations of beat gestures support the hypothesis that beat gestures might increase attention processes and activations of language related-brain areas. The following paragraphs review neurological evidence assessing the effects of beat gestures in a variety of linguistic tasks. Holle Obermeier, Schmidt-Kassow, Friederici, Ward & Gunter, (2012) assessed with event-related potential (ERP) technique the effect of beat gestures during a comprehension task of ambiguous sentences. The results showed that beat gestures co-occurring with the target word of the complex sentence facilitated syntactic processing. Wang and Chu, (2013) investigated the semantic processing of a critical target word produced in three conditions: produced together with a beat gesture, with a control hand movement or with speech alone. The results showed that only beat gestures activated N400 potentials suggesting a beat gesture facilitation of semantic processing of the target word co-occurring with the gesture. Moreover, beat gestures were interpreted as intentional communication as they elicited responses that were not activated with the control condition (i.e., hand movements). Hubbard, Wilson, Callan, and Dapretto (2009) investigated spontaneous uses of beat gestures using functional

magnetic resonance imaging (fMRI) technique in comparison to a still position and a nonsense movement condition. Their results found that beat gestures activated neural correlates related to multisensory and auditory processing of speech with respect to conditions with nonsense movements or still images. Regarding the study of beat gestures during naturalistic videos of political discourse, the functional neuroimaging study by Biau, Fernandez, Holle, Ávila, and Soto-Faraco (2016) showed that different brain areas were activated depending on whether speech was synchronized with beat gestures or with other non-gestural stimuli (i.e., discs/dots moving on a screen). While beat gestures activated language-related areas of the brain, non-gesture stimuli activated visual-perception areas. All these data support the idea that beat gestures can be distinguished from other potential visual highlighters because of their direct integration in the language system rather than a more general visual perceptual system.

Evidence suggests that, in order to have an impact on language processing, beat gestures (a) need to be synchronized with prosodic prominence in speech, and (b) need to be intentional co-speech gestures (as opposed to nonsense hand or visual movements). Biau et al. (2016) proposed that if beat gestures serve a common linguistic process with prosody to highlight pieces of information, video manipulations of asynchronies between beat gestures and prosody would affect processing in areas of multimodal integration. fMRI results showed greater activation when the beat gesture was synchronous to speech than in the asynchronous condition.

Moreover, the beat gestures' condition was compared to a visual control condition showing an unintentional movement (i.e., a dot that moved in the same direction patterns than the hands in the beat condition), which was also related to speech in synchronous and asynchronous conditions. The results showed an opposite effect of the non-intentional movement conditions in comparison to the beat gesture conditions. These results suggest that beat gestures are perceived as intentional actions closely related to prosodic features. The authors concluded that the emphasizing function of beat gestures in speech perception is instantiated through a specialized brain network sensitive to communicative intent conveyed by a speaker with his/her hands (Biau et al, 2016).

Recent findings by Dimitrova, Chu, Wang, Özyürek and Hagoort (2016) confirmed the above-mentioned findings. This study investigated whether there is a potential effect of beat gestures associated to prosody, and whether there is a differential effect of prosody or gestural markers of prominence on language processing. Prosodic prominence conditions varied on focus assigned to the target word with an accented focused target condition (in example, “did she receive a letter or an email from the teacher? She received an **EMAIL** from the teacher”) and an unaccented non-focused target condition (in example, “did she received an email from the teacher of from the rector? She received an **email** from the

TEACHER”)². Gesture conditions varied with three options, i.e., no gesture, beat gestures, and grooming hand movements. The results of this ERP study showed that attention processes increased in those words marked with focused target words synchronous with a beat gesture. That is there was a positive effect of beat gestures associated with prosody only when they accompanied the focus of the message (the target word). In fact, attention processes showed detrimental responses when an unfocused non-target word was presented with a beat gesture. Moreover, beat gestures, but not other type of non-intentional actions, function in temporal coordination to prosodic prominences, which facilitates speech processing.

Summing up, evidence on the temporal coordination of beat gestures’ prominences and prosodic prominence shows that beat gestures activate language processing. Importantly, beat gestures function as intentional communication rather than simple attention getters (e.g., in comparison to conditions with dots synchronized with speech sound, or unsynchronized non meaningful movements of the hands, or grooming gestures). Moreover, the last two studies provide clear evidence of the fact that beat gestures are intentional actions strongly linked to prosodic prominence function.

In our view, and based on previously presented evidence, studies on the integration of gesture and speech have mainly addressed the study of representational gestures. Following McNeill’s temporal

² Target words are in bold.

synchrony rule, the hypothesis underlying the studies presented in this thesis is that temporally synchronous gestures and speech (i.e., the beat dimension) are conceived as intentional motor coordinations reflecting prosodic prominences (see Gesture-As-Simulated-Action Framework by Hostetter & Alibali, 2008, 2010 in section 1.2). Thus, we hypothesize that the multimodal embodied experience of temporal synchronizations is related to important prosodic prominence functions in language, and should thus improve language processes.

Altogether, embodied experience of rhythmical arm and vocal activity might lead to a tight synchrony between gestures and speech with a unique communicative purpose. The developmental hypothesis defended in this thesis is that once infants have acquired this ability to synchronize intentional actions and speech sounds, they will be able to functionally make use of this ability. This embodied experience will dynamically develop and will be related to linguistic abilities. Thus the highlighting function of gestures which are synchronous to speech would develop depending on the complexity of the linguistic context and the children's age.

1.8. The benefits of gesture-speech temporal integration in language processing in development

This section includes the core relevant motivations to investigate the benefits of non-representational gestures in temporal synchrony with speech combinations on language recall, pragmatic inferential resolution and early language development (for an extended motivation of each study see sections 2.1, 3.1 and 4.1, respectively). In this thesis, specific abilities have been chosen in relation to children's specific points in development based on previous literature.

a) Effects of beat gestures on word recall

Research on gestures has extensively reported the beneficial results of representational gestures for various linguistic abilities (section 1.3 for references). While research has shown that adults can benefit from the presence of beat gestures in word recall tasks (see section 1.6), studies have failed to conclusively generalize these findings to preschool children. To our knowledge only two studies have investigated the effects of beat gestures synchronized with target word on language recall (So et al., 2012; Austin & Sweller, 2014). So et al. (2012) found positive effects of beat gestures on a word recall activity in adult participants but not in 4- to – 5- year- old children. In a spatial direction task, Austin and Sweller, (2014) provided an improvement on the ability to recall spatial directions when having a beat gesture associated to the target word. The study

which is presented in Chapter 2 aims at generalizing and extending previous findings to a task which assesses the role these multimodal markers of prominence.

b) Effects of beat gestures and its concomitant prosody on pragmatic inference abilities

Evidence from the previous studies has shown that beat gestures can improve children's language recall by reporting offline measures of behavioral responses (e.g., Austin & Sweller, 2014), and research has shown with adult population that observing beat gestures is related to language-related responses of the brain (see section 1.6). While previous studies have shown that pragmatic inference resolution can be improved by associating the target words to representational gestures (Kirk, et al., 2011) and prosodic markers of prominence (Ito, Bibyk, Wagner & Speer, 2014; Tomlinson, Gotzner, and Bott, in press), to our knowledge little is known about whether beat gestures can improve online resolution of pragmatic inferences. Moreover, no previous studies have used online eye responses (eye tracking technique) to disentangle whether the potential driving effect of the marker of prominence is due only to prosodic prominence, or to a multimodal effect of the beat gesture.

c) Pointing gesture-speech temporal integration as a precursor or language

Pointing gestures' temporal synchronization to speech has been shown to develop around the age of 12 months (see section 1.4). Moreover, children have been shown to modify their communicative responses depending on adult's joint attention to the children's object of interest during natural interaction (Miller & Lossia, 2013; Miller & Gros-Louis, 2013; Gros-Louis & Wu, 2012), and during controlled experimental settings (Liszkowski et al., 2008). However, less is known about whether children produced synchronized gesture-speech productions during experimentally controlled pragmatic context (i.e., varying on the adult's joint attention), and whether this ability is related to later language abilities (Murillo & Belinchón, 2012; Wu & Gros-Louis, 2014).

1.9. General objectives, research questions and hypotheses

This dissertation aims to investigate the role of synchronous gesture-speech integration on children's linguistic abilities. Specifically, we are interested in whether there is a potential beneficial effect of gesture-speech temporally synchronous combinations on a set of children's linguistic abilities at different points in their development.

The following three research questions will be addressed, each one in a separate chapter:

- 1) Do children respond better to a word recall task when the word is presented with a contrast of prominence expressed with a synchronous beat gesture?
- 2) Do children respond better to an online global coherence inference task when the relevant information is presented together with a synchronous beat gesture and its concomitant prosody?
- 3) Do infants' production of synchronous pointing gesture-speech combinations during controlled socio-communicative interactions relate to later language abilities?

Our general hypothesis is that synchronous gesture-speech abilities represent an effective pragmatic strategy which successfully

impacts on language abilities at different moments of children's development. Our specific hypotheses are the following. In the first study (Chapter 2), children will benefit from gestures synchronized with prominence in speech (i.e., beat gestures) in a word recall task. In the second study (Chapter 3), beat gestures and its concomitant prosody will have beneficial impact on the online processing of a pragmatic inference. In the third study (Chapter 4), infants will produce synchronous pointing gesture-speech combinations in the more demanding socio-communicative contexts, and importantly infants' ability to produce synchronous pointing gesture-speech combinations is related to better language abilities later in development.

The first study (Chapter 2) investigates whether the presence of beat gestures helps children to recall information when these gestures have the function of singling out a linguistic element in its discourse context. One hundred and six 3-to-5-year-old children were asked to recall a list of words within a pragmatically relevant context (i.e., a story-telling activity) in which the target word was accompanied or not by a beat gesture. Results showed that children recalled the target word significantly better when it was accompanied by a beat gesture than when it was not, indicating a local recall effect. Moreover, the recall of adjacent non-target words did not differ depending on the condition, revealing that beat gestures seem to have a strictly local highlighting function (i.e., no global recall effect). Interestingly, even though the effect was weak, the analyses show that the children's ability to recall the target word improved

with age only in the beat condition. These results demonstrate that preschoolers benefit from the pragmatic contribution offered by beat gestures when they function as multimodal markers of prominence.

The second study (Chapter 3) investigates whether beat gestures can show beneficial effects on the online processing of global coherence inferences expressed in discourse. In this study, seventy eight 6- to 8-year-old children participated in an eye-tracking experiment involving a pragmatic comprehension task designed to compare their sensitivity to multimodal markers of prominence (i.e., beat gestures and its concomitant prosody). Children's eye movements were recorded as they searched the correct image related to information extracted from a global coherence inference. The results indicated a stronger effect of markers of prominence on the online disambiguation of the target information. Thus, the processing of the inferences was more efficient in the prosody-only (i.e., L+H* pitch accent) and beat gesture (i.e., L+H* pitch accent associated to a beat gesture) conditions than in the control condition (i.e., L* pitch accent). Furthermore, the results showed near significance on the ability to disambiguate the target element depending on children's age. The evidence suggests a potential developmental shift towards a better sensitivity to prosodic cues of prominence by older children (7- and 8- year- old children). The results in this study show evidence on the role of beat gestures as

multimodal markers of prominence which help to improve children's online processing of language.

The third study (Chapter 4) investigates the degree to which an infants' use of synchronous gesture-speech combinations during controlled social interactions predicts later language development. Nineteen infants participated in a declarative pointing task involving three different social conditions: two experimental conditions (a) available, when the adult was visually attending to the infant but did not attend to the object of reference jointly with the child, and (b) unavailable, when the adult was not visually attending to either the infant or the object; and (c) a baseline condition, when the adult jointly engaged with the infant's object of reference. At 12 months of age measures related to infants' speech-only productions, pointing-only gestures, and synchronous pointing-speech combinations were obtained in each of the three social conditions. Each child's lexical and grammatical output was assessed at 18 months of age through parental report. Results revealed a significant interaction between social condition and type of communicative production. Specifically, only synchronous pointing-speech combinations increased in frequency during the available condition compared to baseline, while no differences were found for speech-only and pointing-only productions. Moreover, synchronous pointing-speech combinations in the available condition at 12 months positively correlated with lexical and grammatical development at 18 months of age. The ability to selectively use this multimodal communicative strategy to engage

the adult in joint attention by drawing his attention towards an unseen event or object reveals twelve-month-olds' clear understanding of referential cues that are relevant for language development. This strategy to successfully initiate and maintain joint attention is related to language development, as it increases learning opportunities from social interactions.

2. CHAPTER 2. BEAT GESTURES IMPROVE WORD RECALL IN 3- TO 5-YEAR-OLD CHILDREN

2.1. Introduction

In everyday communication, speakers use hand and body gestures to accompany speech. Beat gestures are a type of manual non-representational gesture which co-occurs with speech and which functions as a visual highlighter of information. In contrast with a representation gesture a beat gesture does not add propositional content to a given utterance (McNeill, 1992, 2005; Kendon, 1995) but rather is used to mark “the word or phrase it accompanies as being significant (...) for its discourse pragmatic content” (McNeill, 1992:15). Beat gestures have been defined as rhythmical movements of the hands which are timed together with prosodic prominence in speech (Loehr, 2012; Shattuck-Hufnagel, Ren, Mathew, Yuen & Demuth, 2016). Typically, the movements of hand gestures occur together with head and eyebrow movements, which together signal the privileged status of a given piece of discourse in a multimodal fashion (see, e.g., Cartmill, Demir & Goldin-Meadow, 2012; McNeill, 1992). In our study we investigate whether the use of beat gestures as a multimodal marker of prominence in a significant discourse context helps to improve language recall abilities in early childhood.

Research on gestures has extensively reported the beneficial results of representational gestures for various linguistic abilities, such as the improvement of narrative skills between the ages of 5 and 6 (Demir, Fisher, Goldin-Meadow & Levine, 2014) or the comprehension of complex syntactic abilities by 3- and 4-year-olds (Theakston, Coates & Holler, 2014). In parallel with this, the use of representational gestures has been shown to have cognitive benefits at different stages of children's cognitive development. For example, the benefits of representational gestures have been proven for 4-5-year-old children in the recall of words in the first language (Church, Kelly & Lynch, 2000; So, Chen-Hui & Wei-Shan, 2012; Thompson, Driscoll & Markson, 1998), for 5-year-old children learning of words in a second language (Tellier, 2008), and for 9-year-olds children solving arithmetic operations (Goldin-Meadow, Cook & Mitchell, 2009). By contrast, the potential beneficial effects of beat gestures, which by definition do not carry semantic meaning, has not been investigated in depth, particularly in development.

Adults seem to benefit from observing beat gestures when asked to recall lexical information in a first language (Kushch & Prieto, 2016; So, et al., 2012) or learn novel words in a second language (Kushch, Igualada & Prieto, under revision). This data suggests that beat gestures highlighting a specific target in the discourse benefit the recall and learning of that target, though less is known about the impact of this highlighting function of beat gestures on the processing of the co-occurring discourse information.

What underlies the possible cognitive advantage offered by beat gestures? Interestingly, several studies measuring event-related brain potentials (ERP) in adults have shown neural evidence of the activation of language-related areas when beat gestures are perceived, suggesting that they have an attentional effect (Biau & Soto-Faraco, 2013; Holle Obermeier, Schmidt-Kassow, Friederici, Ward & Gunter, 2012; Wang & Chu, 2013). The functional neuroimaging study by Biau, Moris Fernandez, Holle, Ávila, and Soto-Faraco (2015) showed that different brain areas were activated depending on whether speech was synchronized with beat gestures or with other non-gestural stimuli (i.e., discs/dots moving on a screen). While beat gestures activated language-related areas of the brain, non-gesture stimuli activated visual-perception areas. Hubbard, Wilson, Callan, and Dapretto (2009) found that beat gestures, and not nonsense movements or still images, enhanced auditory processing of speech. All these data support the idea that beat gestures can be distinguished from other potential visual highlighters because of their direct integration in the language system rather than a more general visual perceptual system.

Moreover, from a linguistic point of view, beat gestures have been shown to serve a focus-marking function (Yasinnik, Renwick & Shattuck-Hufnagel, 2004; Jannedy & Mendoza-Denton, 2005; Loehr, 2012; Shattuck-Hufnagel, et al., 2016). In addition, adult listeners have shown an increase in prominence perception when

words are produced together with hand gestures (Krahmer & Swerts, 2007) and head/facial beat gestures (Moubayed, Beskow & Granström, 2010). Apart from the abovementioned physiological and linguistic evidence, the positive cognitive effects of beat gestures are consistent with the embodied cognition paradigm, which underlines the relevance of the body movements and multimodal supporting channels in cognition and in favoring memory traces (see Paivio, 1990; and Barsalou et al., 2003, Bersalou, 2008).

Though there is strong evidence that beat gestures are related to language and cognitive abilities in adults, the benefits of beat gestures in childhood are less clear. First, research into the first uses of beat gestures during language development is fairly limited. Some production studies point out that children start producing beat gestures around their fifth year of life, as their narrative skills develop (e.g., McNeill, 1992; Stefanini, Bello, Caselli, Iverson & Volterra, 2009; Mathew, Yuen, Shattuck-Hufnagel, Ren & Demuth, 2014; Shattuck-Hufnagel et al., 2016). Regarding perception, to our knowledge only two studies—So et al. (2012) and Austin and Sweller (2014)—have investigated the effects of beat gestures on the recall of information in childhood, with opposite results.

With regard to the first of these studies, So et al. (2012) found that adults benefited from both beat and representational gestures in a word recall task, while children benefited from co-speech

representational gestures but not from beat gestures. In a first experiment, adults were shown a video presentation of a list of 10 verbs in three conditions (accompanied by a representational gesture, a beat gesture, or no gesture). In the two conditions containing gestures, each verb co-occurred either with a gesture representing some semantic feature of the action described (the representational condition) or with a simple open-palm downward beat gesture (the beat condition). The results showed that adults recalled a greater number of verbs in either the representational or the beat gesture conditions than in the no gesture condition. The second experiment replicated the first study but with 4- to 5-year-old children as subjects. The procedure was similar except for the number of words presented in the list, which was reduced from 10 to 5 in order to accommodate the shorter mnemonic span of the children. The results revealed that the children's ability to recall words benefited from the presence of representational gestures but not from the presence of beat gestures. There are two possible reasons for these negative results. First, in the experiment every word in the list was accompanied by a beat gesture, which may have reduced the highlighting function of the beat gestures. Typically, beat gestures in natural speech do not appear in with every word but rather are used to make particular items stand out from surrounding non-prominent elements (Terken, 1991; Wagner, Origlia, Avesani, Christodoulides, Cutugno, D'Imperio, Escudero Mancebo, Gili Fivela, Lacheret, Ludusan, Moniz, Ní Chasaide, Niebuhr, Rousier-Vercruyssen, Simon, Šimko, Tesser & Vainio,

2015). Moreover, the list of target words was presented in isolation and without a child-relevant discourse context in which the presence of beat gestures might have been pragmatically motivated. From a perspective of discourse assessment, a task in which linguistic units are presented without contextual support may lack pragmatic motivation (e.g., Ito, Jincho, Minai, Yamane & Mazuka, 2012).³

With regard to the second study, Austin and Sweller (2014) found that beat gestures did facilitate the recall of spatial directions in 3- to 4-year-old children. Participants first visually examined a toy representation of a landscape. Then, the experimenter provided different directions that the children had to recall afterwards. These directions consisted of locations and actions accompanied by either a co-speech gesture, a beat gesture, or no gesture. In the no gesture condition, spatial directions were described verbally without gestures; in the co-speech gesture condition, different types of gestures (i.e., beat, deictic, metaphoric, and iconic gestures) were produced with affiliated target words; and in the beat condition, target words co-occurred with a beat gesture. Results showed that, on average, children recalled the information about the spatial directions better in either the co-speech gesture condition or the beat condition than in the no-gesture condition. In this experiment, the visual accessibility of the referents through the toy representation

³ Moreover, in her chapter about methodologies to explore the processing of prosodic focus by children, Ito (2014) recommends the use of a “conversational task that gives salience to contrastable referential candidates” (p. 206).

might have played a facilitating role in the recall of both location and action target words. Indeed, other studies assessing children's comprehension of prosodic prominence have demonstrated that the presence of visual stimuli facilitates the recognition of target items by children (Ito, et al., 2012; Ito, 2014). However, given that in many natural contexts the visual referents associated with beat gestures are typically not present in the immediate context, it would be of interest to generalize and extend these findings to a task which does not involve a concomitant visual presence of the referents. This is what we will endeavor to do in the present study.

The first goal of our study is to investigate preschoolers' general ability to use beat gestures in a word recall task. In order to address the methodological issues that we feel were raised by the two earlier studies, the experimental task was designed with the following features: (a) beat gestures will function as local highlighters of target words which will contrast with adjacent words produced without beat gestures; (b) the task will be embedded in a discourse context that is pragmatically relevant for 3- to 5-year-old children; and (c) the target words will not refer to objects that are visually present in the experimental setting. We expect that children will interpret words associated with salient beat gestures as pragmatically more relevant than the others, which will enhance their recall of them. Thus, we hypothesize that children aged from 3 to 5 will make use of the highlighting function of a beat gesture when having to recall the word affiliated with that gesture.

A second aim of our study is to disentangle whether this potentially enhanced recall effect impacts not only on the target word associated with the beat gesture (i.e., a local recall effect) but also on the recall of adjacent words in the list which are not associated with beat gestures (i.e., a global recall effect). If beat gestures work exclusively as local highlighters of information conveyed through speech, adjacent words in the list should not be affected by the presence of a beat gesture highlighting a target word. This would be consistent with the results of previous studies showing language-related brain responses temporally associated with the words produced with a beat gesture (Hubbard et al. 2009; Biau & Soto-Faraco, 2013; Holle et al., 2012; Wang & Chu, 2013; Biau et al. 2015). If, by contrast, beat gestures induce a global recall effect, a beat gesture co-occurring with a target word will enhance the recall of adjacent words in the list.

Third, while previous studies have reported that children start producing beat gestures at around 5 years of age (e.g., McNeill, 1992; Stefanini et al. 2009; Mathew et al. 2014; Shattuck-Hufnagel, et al. 2016), and that they process other markers of prominence (like pitch accentuation) in an adult-like way around 6 years of age (Ito et al. 2014), we would like to assess the developmental pattern of these effects in younger children aged 3–5. If the effect of beat gestures on memory recall abilities is sensitive to age differences,

this study will provide important data on how multimodal markers of prominence are processed during development.

2.2. Methods

a) Participants

One hundred and thirteen 3-to 5-year-old children were initially recruited for the study. However, seven of these children had to be excluded from the study: one child was diagnosed with language pathology by the school services, four children did not want to collaborate during the experiment, and two children were not tested because their memory span was greater than the length of the lists prepared for the experiment (see section 2.3.1 below). The total sample of the study thus comprised 106 children aged 3, 4, and 5. Table 1 offers details of the subject sample broken into groups according to age, gender, and memory span. All the participants were preschoolers enrolled at three public nursery schools located in the province of Barcelona, Spain.⁴ In these nursery schools, the main language of instruction is Catalan. Parents were informed about the experiment's goal and signed a participation consent form. Furthermore, language exposure questionnaires (based on Bosch and Sebastián-Gallés, 2001) were administered to the caregivers in order to ensure that children were predominantly exposed to Catalan at home on a daily basis (mean exposure to Catalan as a percentage of all daily language exposure: $M = 87.30$, $SD = 13.09$). Parental questionnaires reported that children were healthy and had normal hearing.

⁴ Escola Sant Martí in Arenys de Munt, Escola La Farigola del Clot in Barcelona, and Escola Pública Dr. Estalella Graells in Vilafranca del Penedès.

Children	<i>N</i>	Age in months	Girls	Memory span in number of words
3-year-olds	34	$M = 41.74$ ($SD = 3.58$)	12	$M = 2.68$ ($SD = .64$). 2 words ($N = 14$); 3 words ($N = 17$); 4 words ($N = 3$)
4-year-olds	40	$M = 53.93$ ($SD = 3.79$)	18	$M = 3.20$ ($SD = .56$). 2 words ($N = 3$); 3 words ($N = 26$); 4 words ($N = 11$)
5-year-olds	32	$M = 64.91$ ($SD = 3.16$)	17	$M = 3.84$ ($SD = .45$). 2 words ($N = 1$); 3 words ($N = 3$); 4 words ($N = 28$)

Table 1. The sample population broken down into age groups, showing Mean (M) and Standard Deviation (SD) for age in months and memory span in number of words, as well as gender.

Memory span	<i>N</i>	Age in months
2 words	18	$M = 42.78$, $SD = 7.08$, Median = 41.50, Min = 35, Max = 64
3 words	46	$M = 50.15$, $SD = 7.03$, Median = 50.00, Min = 36, Max = 67

4 words	42	$M = 61.33, SD = 6.86,$ Median = 63.00, Min = 42, Max = 71
---------	----	---

Table 2. The sample population separated into groups according to memory span in number of words and showing Mean (M), Standard Deviation (SD), median, minimum (Min) and maximum (Max) age in months, broken down by memory span.

b) Materials

The materials consisted of the audio-visual recordings of sentences produced by a Catalan female actor. Each sentence consisted of an introductory phrase followed by a list of nouns that an elephant named ‘Elmer’ (taken from the children’s book series about *Elmer the Patchwork Elephant* by David McKee) was supposed to remember (see Figure 1 for an example). All nouns in each list belonged to the same semantic context (things to buy at the market, animals to visit at a zoo, objects to tidy up in a room, or pictures to be drawn) and were controlled for frequency of usage. All the nouns included in the lists appeared in the Catalan version of the MacArthur-Bates Communicative Development Inventory (CDI) for children aged 16-30 months (Serrat, Sanz-Torrent, Badia, Aguilar, Olmo, Lara, Andreu & Serra, 2010) (see Appendix 1 for a complete list of sentences). They were all disyllabic with word stress either word-initial or word-final. The audiovisual recording

was performed using a PMD660 Marantz professional portable digital video recorder and a Rode NTG2 condenser microphone.

Adobe Premiere Pro CS6 editing software was used to splice together the various video sequences (introductory sentence followed by list) and also to embed a drawing of an elephant into the video image. Finally, each video was embedded in a PowerPoint presentation.

Example: “Elmer! Before leaving for your vacation, go to the zoo and say goodbye to the...

BEAT CONDITION: ducks, birds, **PARROTS**, horses, bears

NO-BEAT CONDITION: ducks, birds, parrots, horses, bears

Figure 1. Example of a five-word list in both beat and no-beat conditions in which the central target word is underlined. The word in bold and capital letters was emphasized with a beat gesture in the accompanying video recording.

In a within-subject experimental design, the test materials were presented to the children in two different audiovisual conditions, a beat condition and a no-beat condition. In the beat condition, in order to create a contrast of prominence among the nouns in the list only one word (the target word) was accompanied by a beat gesture.

In the no-beat condition, the target noun was not accompanied by a beat gesture. In order to have a target list that was appropriate for the memory span of each particular child, three types of test lists were created, namely three-noun lists, four-noun lists, and five-noun lists, (see section 2.3.1). Table 1 above shows memory span expressed as the number of words a child could recall, broken down by age in years, and table 2 shows the age in months, broken down by memory span ability. For both conditions we controlled for serial sequential effects (i.e., the tendency to remember more easily the first or last items in a list) by placing the target word in the central position in the list (as seen in Figure 1). The critical items were placed in the second position in a three-word length list, in the second position in a four-word length list, and in the third position in a five-word length list.

The beat gesture consisted of a downward hand movement associated with a head nod, opened eyes, and raised eyebrows. More specifically, the hands were raised from a low position near the waist to chin level, with an open-palm vertically oriented hand shape. They were then lowered to chest level as the actor nodded, opened her eyes widely, and raised her eyebrows. Finally, the actor returned her hands to the initial rest position (Figure 2). The specific type of beat gesture used in the materials was defined after conducting a Discourse Completion Task⁵ (Billmyer & Varghese, 2000) with 10 adult Catalan-speakers who were asked to imagine they had to prompt a 6-year-old child to remember a list of three

⁵ The Discourse Completion Task is a data elicitation method based on discourse contexts which has been applied for many years in research on pragmatics and sociolinguistics (see, e.g., Prieto and Roseano, 2010).

words by saying something like *Joan! Hem de comprar pomes, iogurt, cebes* ‘John! We need to buy apples, yogurt, onions’ This was done in two ways. First, the adults were given no explicit instructions about how to form the sentence they would tell the child, and second, they were told to emphasize the target words in the list. Results of the Task showed that the adults produced significantly more beat gestures when asked to highlight the target words ($t(238) = -2.696, p < .01$), and that the most commonly occurring beat gesture was an up-down arm movement (100% of instances) with palms open and fingers spread (45% of instances; palm open with fingers curled in 27.8%, okay gesture shape 10%, palm open with fingers touching 10%, and pointing with index finger 2.5%). Typically, this manual gesture was accompanied by a head nod (75.9%) (see Figure 2, right panel). This combination of up-down arm movement, open palm with fingers spread, and head nod was therefore the beat gesture used in all instances by the actor in the video recording.

The recording of each target word was first produced in the beat condition followed by the recording of the no-beat condition in order to avoid variance across conditions. All words in all conditions were produced with the same flat and high pitch contour (H*). Measurements of the recorded signal showed that the pitch range (measured in hertz) of the words did not change across conditions: ($t(38) = .824, p = .415$; beat condition: $M = 53.55, SD = 21.21$; no-beat condition: $M = 48.97, SD = 12.91$). By contrast, the

duration of the stressed vowel ($t(38) = 2.092, p < .05$; beat condition: $M = 301.29, SD = 69.54$; no-beat condition: $M = 256.85, SD = 64.75$) and duration of the word ($t(38) = 3.643, p = .01$; beat condition: $M = 1050.85, SD = 237.82$; no-beat condition: $M = 843.75, SD = 89.84$) changed across conditions. Throughout each recording, the gaze of the actor producing the sentence was directed at the bottom left corner of the screen to make it appear as though she was addressing Elmer the elephant, a colorful drawing of which was embedded into this area of the screen.

In the recall experiment, the children were exposed to a total of six trials: two practice trials aimed at familiarizing the child with the task (i.e., one trial per condition) and four experimental trials (i.e., two trials per condition). A total of 36 PowerPoint presentations were created by varying the following factors: (a) the number of words per list (three, four, or five words); (b) three random changes of the word order position in the lists to prevent words from always occurring in the same position; (c) two counterbalanced orders of condition presentation (i.e., one group of children first received an experimental trial in the beat condition followed by an experimental trial in a no-beat condition, this order being repeated across all trials, and the other group of children received all trials in the opposite order of presentation); and (d) a counterbalanced order of presentation of the semantic contexts to prevent a potential semantic priming effect (i.e., some children were exposed first to market and zoo contexts, others first to room and picture drawing contexts). In sum, a total of 36 PowerPoint presentations ($3 \text{ list lengths} \times 3 \text{ word}$

orders in the list \times 2 orders of conditions \times 2 orders of semantic contexts = 36 PowerPoint presentations) were produced.

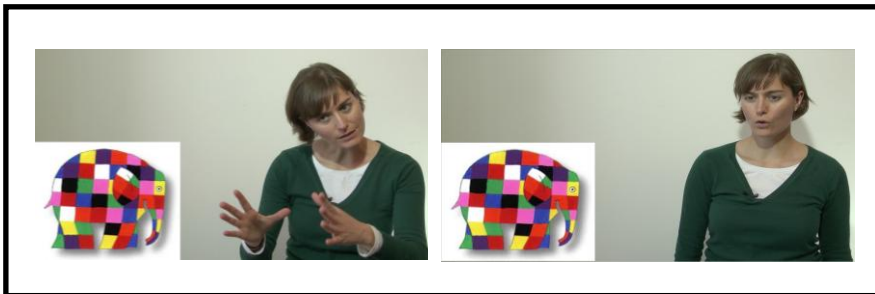


Figure 2. Still images from the stimulus video showing the actor producing the target word in the no-beat (left panel) and beat condition (right panel) while addressing Elmer the elephant.

c) Procedure

The children were tested individually in a quiet classroom at their nursery school. The full session lasted approximately 15-20 minutes and consisted of two tasks, a memory span task and a word recall task as detailed below. The memory span task allows us to assign each child to the list type appropriate to their memory span, namely three, four or five word list lengths. Then they were tested with the word recall task.

Memory span task

First, each child was asked to play a game intended to measure their memory span in which they were supposed to repeat a list of words spoken by the experimenter. Following Henry, Messer, Luger-Klein and Crane's (2012) procedure, memory span was measured in terms of the maximum number of words from the list that the child could recall. For all participants, the memory span test started with a list of one item, which was followed by a list of two, a list of three, and so on. This procedure continued until the child could no longer succeed in recalling all the words in the list. Once the maximum list length seemed to have been reached, the child was told four lists of this length but consisting of different words. This number of words was regarded as the child's memory span if all the items were recalled in at least three out of the four lists. The memory span threshold thus measured was what determined the length of the list in the subsequent word recall task, such that if the child's memory span was equal to two words, children were presented with three-word lists in the word recall task. Thus, word span ability was used as a control measure to adjust the demands of the word recall task to participants' memory abilities. Within those parameters, children were randomly assigned to one of the PowerPoint presentations previously prepared (see section 2.2).

Word recall task

The word recall task was performed during the narration of a story about an elephant that enjoys travelling, in which children were told that they would have to recall a list of things that the elephant has to do before travelling. The word recall task consisted of two phases: a presentation phase, where the characters and plot of the story were introduced, and a test phase, which involved a repetitive sequence of trials of three sub-phases: a word list exposure phase, a word list recall phase, and a story-resumption or concluding scene (see Figure 3 for a schematic diagram and Appendix 1 for a detailed script of the word recall task).

In the presentation phase, the experimenter used the initial slide of a PowerPoint presentation to introduce Elmer the elephant and his friends, a group of elephants and a female human (here a photo of the actor amid drawings of elephants appeared). The experimenter then went on to explain that Elmer always forgot things but was very lucky because his good human friend always helped him to remember things. The child was encouraged to help Elmer too by repeating the list of things that the friend would remind him to do.

After this background to the story was presented to the child, the test phase began. In each trial the plot of the story continued with a set of sequences that all followed the same pattern: (a) a word list exposure phase, in which a video embedded within the PowerPoint

showed the actor described a context and said a list of words; (b) a word list recall phase, in which the experimenter asked the child a prompt question and the child attempted to repeat the word list; and (c) a story-resumption or concluding scene with a distractor scene, which was intended to refresh the child's memory load and motivate the resumption of the narrative with a drawing of the character in a new scene (e.g., the beach, mountains, or desert), or the scene that brought the story to a close, with Elmer playing with his elephant friends because he had managed to finish everything he needed to do before leaving on his trip.

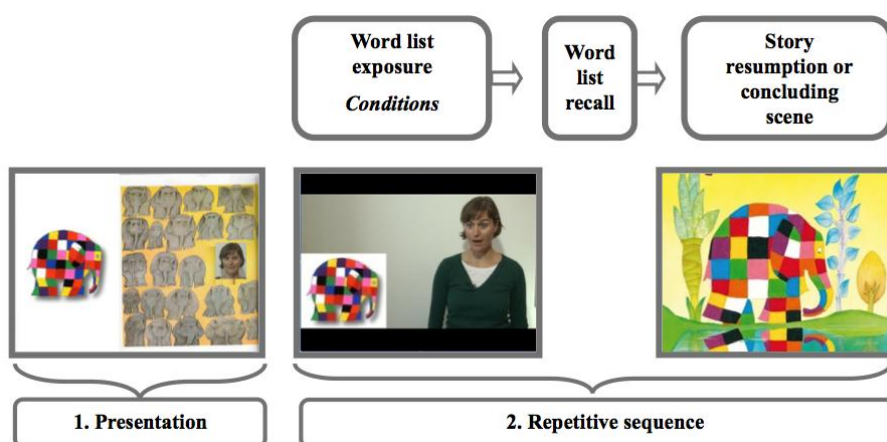


Figure 3. Schematic representation of the procedure of the word recall task.

d) Coding

An assistant was seated behind the child to code the responses. Children's responses were systematically coded in an answer sheet

which included the lists of words in the same order of appearance than in the presentation. Before each target word there was a small check box in which the assistant indicated with a check mark a word that was mentioned by the child. The experimenter double-checked the words with the assistant after each word list recall phase by repeating what the child had said in the same order.

2.3. Results

a) Local word recall effects

In order to assess the effect of beat gestures on the children's ability to recall the target word from among the items in each word list, a Generalized Linear Mixed Model (GLMM) was applied to the data. The dependent variable was the number of target words recalled in each trial (1–recalled, 0–not recalled) by children during the test phase. Gesture condition (two levels: no-beat gesture and beat gesture), age (three levels: 3-, 4- and 5- year-olds), and all their possible interactions were set as fixed factors; trial, condition order, and participant were set as random factors. Bonferroni pairwise comparisons were carried out for the significant main effects and interactions.

The results of the GLMM analysis showed a main effect of condition ($F(1, 418) = 4.009, p < .05$), with a better recall of the target item in the beat condition than in the no-beat condition. The mean (M) and standard deviation (SD) values of the recall of the target item were as follows: no-beat condition: $M = .38, SD = .48$; beat condition: $M = .49, SD = .5$. The results did not show a main effect of age ($F(2, 418) = 2.804, p = .062$). No significant interaction between gesture condition and age was found ($F(2, 418) = .112, p = .894$). Figure 4 shows that at all ages, target items

accompanied by a beat gesture were recalled better than target items not accompanied by a beat gesture.⁶

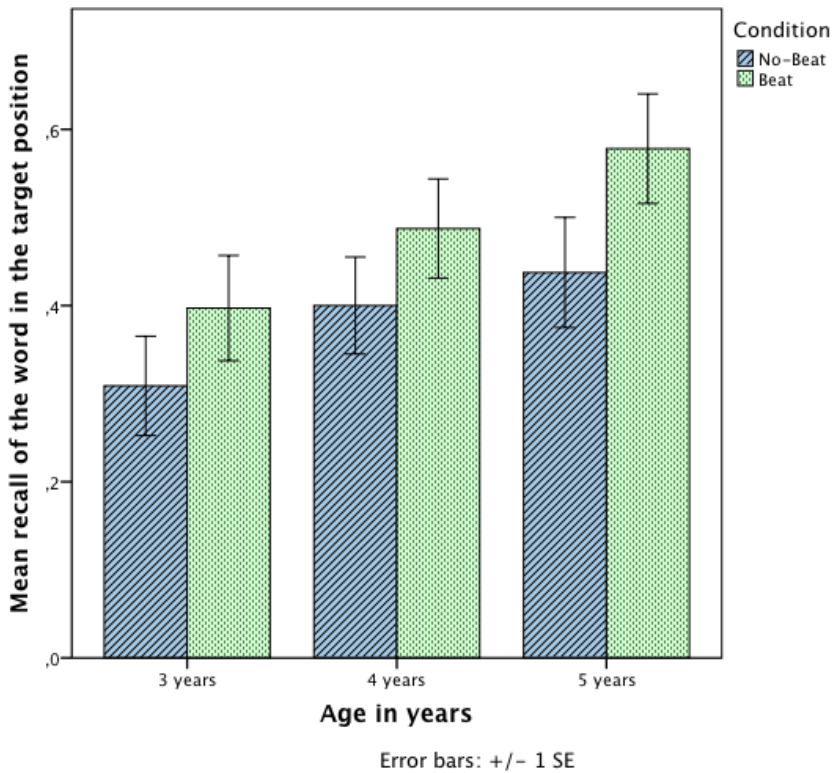


Figure 4. Mean number of items recalled in the target position as a function of condition and age.

⁶ Following a reviewer's suggestion to test for a potential gender difference in the ability to recall the target word, a GLMM analysis was applied to the dependent measure of the ability to recall the target word. Gender (two levels: male and female), condition (two levels: no-beat and beat gesture), age (three levels: 3-, 4- and 5- year-olds), and all their possible interactions were set as fixed factors; trial, condition order and participant were set as random factors. Bonferroni pairwise comparisons were carried out for the significant main effects and interactions. The results of the GLMM did not show a main significant effect of gender ($F(1, 412) = 0.273, p = .602$), nor an interaction effect between gender and condition ($F(1, 412) = 1.587, p = .208$), nor a triple interaction effect between gender, condition and age ($F(2, 412) = 0.424, p = .654$). These results back up our decision not to include gender as a fixed factor in our main GLMM model.

b) Global word recall effects

In order to assess what we call the global effects of beat gestures, that is, the potential effect of the presence of beat gestures on the recall of the non-target items in the list, another GLMM analysis was run. The dependent variable was calculated as a proportion of the number of non-target words recalled in each trial divided by the maximum number of items in the list. Gesture condition (two levels: no-beat gesture and beat gesture), age (three levels: 3-, 4- and 5-year-olds), and all their possible interactions were set as fixed factors; trial, condition order and participant were set as random factors. Bonferroni pairwise comparisons were carried out for the significant main effects and interactions.

The results of the analysis showed that recall of non-target items was not significantly affected by the gesture condition ($F(1, 418) = .165, p = .695$). A main effect of age was found ($F(2, 418) = 21.697, p < .001$), and Bonferroni pairwise comparisons revealed that 5-year-old children recalled the items in non-target positions better than 3-year-olds ($p < .001$) and 4-year-olds ($p < .01$), and 4-year-old children recalled non-target items better than 3-year-old children ($p < .01$). No significant interaction was found between condition and age ($F(2, 418) = 1.505, p = .223$). The mean (M) and standard deviation (SD) values for the recall of non-target items were as follows: for 3-year-olds, no-beat condition: $M = .35, SD = .23$; beat condition: $M = .30, SD = .21$; for 4-year-olds, no-beat condition: $M = .42, SD = .19$; beat condition: $M = .40, SD = .18$;

and for 5-year-olds, no-beat condition: $M = .48$, $SD = .16$; beat condition: $M = .52$, $SD = .16$.

c) Correlation analysis between word recall ability and age in months

In the previous set of results in section 3.1, no main effect of age nor a significant interaction between age and condition was found on the ability to recall the target word. Nevertheless, visual inspection of Figure 4 and the fact that the effect of age was near-significant ($F(2, 418) = 2.804$, $p = .062$) signals the importance of age in the ability to recall words and suggests that this developmental pattern is stronger in the beat condition than in the no-beat condition.⁷

In order to explore these effects, two separate pairwise correlation analyses were run with children's age in months in relation to the ability to recall the target item in first the beat condition and then in the no-beat condition (that is, a separate correlation analysis for each gesture condition). The dependent variable was the average recall score for each child per condition. This measure was calculated by adding the total number of words that a child recalled

⁷ The mean and standard deviation values for recall of the target item were as follows: for 3-year-olds, no-beat condition: $M = .31$, $SD = .46$; beat condition: $M = .40$, $SD = .49$; for 4-year-olds, no-beat condition: $M = .4$, $SD = .49$; beat condition: $M = .49$, $SD = .5$; and for 5-year-olds, no-beat condition: $M = .44$, $SD = .5$; beat condition: $M = .58$, $SD = .49$.

in the two trials associated with one condition and then dividing that figure by two (the number of trials), resulting in possible averages of 0, .5, or 1.

The results of the first correlation analysis showed that the ability to recall the target word in the beat condition was positively and significantly correlated with age in months $r(104) = .234$ ($p < .05$). Conversely, children's age in months was not significantly correlated with recall of the target word in the no-beat condition $r(104) = .155$ ($p = .112$). These results confirm that the ability to recall words in the beat condition develops more strongly with age than the ability to recall words in the no-beat condition.

2.4. Conclusions

In this study we set out to investigate whether the presence of beat gestures increases children's word recall in a list of words when beat gestures function as multimodal highlighters in a child-relevant discourse context. Second, we investigated whether the impact of beat gestures on recall is limited to words that co-occur with a beat gesture (i.e., it has a merely local effect), or whether this effect extends also to adjacent words in the discourse (a global effect). Moreover, we were interested in investigating whether children's age influences their ability to benefit from perceiving beat.

The results of a word recall task performed by one hundred and six 3-to-5-year-old children showed that the children recalled the target words significantly better when they were accompanied by a beat gesture than when they were not. This demonstrates that preschool children benefit from the presence of beat gestures when the gesture marks an item as being more salient than others. This evidence represents a valuable addition to the hitherto contradictory results reported in the literature on the benefits of beat gestures in children. On the one hand, So et al. (2012) reported results showing that preschoolers recalled more words when they were associated with iconic gestures than when they were not, but this facilitating effect was not found with beat gestures. The clear effects of the beat condition in our study, in comparison with the negative effects found in So et al. (2012), might have been influenced by a set of

factors: (a) in our materials, beat gestures functioned as local highlighters of a target word which contrasted with adjacent words produced without beat gestures; (b) the task was embedded in a discourse context that was pragmatically relevant for preschoolers; and (c) we followed a naturalistic approach in the design of the materials and the beat gestures included hand movements together with other gestural markers of prominence, such as head and eyebrow movements. It is well known that in natural communication, beat gestures typically co-occur with other prosodic and gestural markers of prominence such as hand and head movements, and specific facial cues. In So et al.'s experiment, beat gestures were limited to hand movements only. In our study, on the other hand, beat gestures were multimodal. This multimodality may have boosted the children's perception of prominent elements within the discourse, thus explaining the difference in our results.

With respect to this multimodal encoding of beat gestures, future research will be needed to address the question of whether certain visual markers of prominence (i.e., particular head, eyebrow, or hand movements) are more powerful than others, or act like prosodic markers of prominence. As is well known, there is a tight temporal synchrony between beat gestures and prosodic prominence (Leonard & Cummins, 2010; Loehr, 2012; Shattuck-Hufnagel, et al., 2016). However, it is also possible that other non-manual multimodal markers of prominence like head movements worked together with the hand beat gestures to trigger the word recall effect. Recent results seem to suggest that adults learn novel words better

when they are synchronized with a beat gesture, but only if the words are accompanied by prosodic prominence (Kushch, Igalada & Prieto, 2015). Moreover, while prosodic prominence triggers attentional processes, only visual prominence (e.g., beat gestures) facilitates semantic processing of the co-occurring words during the perception (Wang & Chu, 2013).

In contrast with So et al.'s results, and in line with Austin and Sweller (2014), we found that preschoolers recalled significantly more target words when they were accompanied by a beat gesture. As noted above, in Austin and Sweller (2014) the visual presence of referents might have played a facilitating role in the recall task. In this regard, our study expands on their results by showing that beat gestures strongly favor word recall even when the referent is not visually present in the conversational context. Presumably the ability to access the highlighting function of beat gestures is developing in children of this age range, and therefore accompanying every word with a beat gesture (as in So et al.'s experiment) makes it more difficult for children to access the highlighting and contrasting function of such gestures. Importantly, in spontaneous discourse some words are accompanied by beats and others are not (Terken, 1991; Austin & Sweller, 2014; Wagner, et al., 2015), and this saliency feature of beat gestures might be important for the child to identify what is the important information in a discourse.

Our results also point to a local effect of beat gestures, since we found no overall increase in the children's recall of adjacent words in the beat condition. These results are in line with recent neurophysiological data showing a clear local time alignment between brain responses and target speech associated with beat gestures (Hubbard et al., 2009; Wang and Chu, 2013; Biau, et al., 2015). Beat gestures seem to act as attentional local highlighters that help children focus their attention on a particular piece of information, which in turn helps them improve their performance in a recall task.

With respect to the developmental factor, our results indicated that 3- to 5-year-old children performed similarly in the recall task regardless of age when age was expressed in years. A more fine-grained analysis showed a weak but significant positive correlation between age expressed in months and the ability to recall a target word in the beat condition. The correlation between age in months and the ability to recall target words in the no-beat condition, however, was not significant. Interestingly, the ability to recall words associated with beat gestures seems to appear earlier than children's own first productions of beat gestures, which takes place around the age of 5 (McNeill, 1992; Stefanini et al., 2009; Mathew et al., 2014). Future research could further explore how beat gestures develop in parallel with other cognitive and linguistic abilities, and whether the sensitivity to beat gestures at age 3 can predict a greater use of these gestures (and other grammatical markers of saliency) at later ages. Oral strategies to express saliency

in discourse (by means of, for example, degree modifiers, syntactic word order, or contrastive pitch accent) seem to develop at around 4-5 years of age (Chen, 2011; Ito et al., 2012; Järvikivi, Pyykkönen-Klauck, Schimke, Colonna & Hemforth, 2015; Tribushinina, 2004), and our results show a successful interpretation of beat gestures as early linguistic highlighting devices at even younger ages, pointing to a potential scaffolding role for beat gestures as multimodal markers of prominence in a discourse.

In sum, the main novelty of our study is the evidence it provides that children can benefit from the presence of beat gestures functioning as multimodal markers of prominence in a pragmatically appropriate context. Why is it that beat gestures help listeners recall the target words that they accompany? We think the answer lies in the fact that gestures mark referents as being prominent in a multimodal way, and this saliency increases the attention that a listener pays to a particular piece of information. In fact, the ability to selectively attend to specific elements of speech while disregarding others (i.e., temporal attention) has recently been argued to facilitate language development in its early stages (de Diego-Balaguer, Martinez-Alvarez & Pons, 2016). In the realm of applied linguistics, beat gestures could be used as a teaching strategy to cue relevant information in a discourse context in both educational and therapeutic settings.

3. CHAPTER 3. BENEFITS OF BEAT GESTURES AND PROSODY IN CHILDREN'S ONLINE PROCESSING OF PRAGMATIC INFERENCES

3.1. Introduction

Research on gestures has extensively reported that deictic and representational gestures⁸ are beneficial in promoting language comprehension across different ages and for various linguistic abilities. For example, 12-month-old infants are able to comprehend an adult's indications about the location of a hidden object when they are accompanied by a pointing gesture (Behne, Liszkowski, Carpenter & Tomasello, 2012). Research has also assessed the impact of children's ability to integrate semantic features expressed in speech and gesture. For example, Sekine, Sowden and Kita (2015) explored 3- to 5-year-old children and adults' ability to integrate iconic gestures with information expressed in speech (e.g., a speaker makes a gesture describing a particular type of object while saying the verb "throwing"). While children aged five and adults were successful with the comprehension task when it involved video-recorded input, younger children were not.

⁸ Representational gestures express semantic features of entities and events. Deictic gestures are (e.g., pointing gestures) intended to direct a listener's attention to a referent of interest.

However, in a second experiment 3-year-old children succeeded in integrating gesture-speech semantic information during live interactions. Similarly, other studies have reported evidence on the impact of the use of abstract gestures on language comprehension. Theakston, Coates, and Holler (2014) used place holder gestures (i.e., gestures which refer to two concepts by associating them with two different positions in the gesturing area) to improve 3- and 4-year-old children's ability to comprehend complex syntactic structures. This was done by consistently relating the agent and patient of the sentence to two different locations in the gesturing area.

Other research has focused on beat gestures, which do not express specific semantic meanings. Beat gestures have been defined as rhythmical movements of the hands which are timed together with prosodic prominence in speech. In one study analyzing speaker behaviors, for example, Yasinnik, Renwick and Shattuck-Hufnagel (2004) showed that between 65-90% of beat gesture apices occurred together with a pitch-accented syllable. Typically, the movements of hand gestures occur together with head and eyebrow movements (see, e.g., Cartmill, Demir & Goldin-Meadow, 2012; McNeill, 1992). Beat gestures have been typically associated with rhythmic marking, focus marking, and discourse structure marking functions (e.g., Yasinnik, et al., 2004; Jannedy & Mendoza-Denton, 2005; Loehr, 2012; Shattuck-Hufnagel, Ren, Mathew, Yuen & Demuth, 2016). Beat gestures have also been shown to positively influence the recall of information by adults in a first language (So,

Chen-Hui & Wei-Shan, 2012; Kushch & Prieto, 2016) and in a second language (Kushch, Igalada & Prieto, under revision), and also in child populations (Austin & Sweller 2014; Igalada, Esteve-Gibert & Prieto, 2017). However, less is known about the potential beneficial effects of beat gestures accompanying prosody for language (and pragmatic) comprehension.

Regarding the role of prosodic prominence, research has reported that pitch accentuation has a positive effect on the memorization of information (e.g., Fraundorf, Watson & Benjamin, 2010; Kushch & Prieto, 2016) and also L2 word learning (Kushch et al. under revision) by adult populations. Various authors agree that intonation constrains the online resolution of inferences in association with different pragmatic meanings. For example, research has shown that pitch accentuation in English (H*) can rapidly disambiguate utterances by integrating information from referents that have previously been mentioned in the discourse in both adults (Dahan, Chambers & Tannenhaus, 2002) and children (Ito, Bibyk, Wagner & Speer, 2014). Tomlinson, Gotzner, and Bott (in press) showed that adults associate pitch accents with informative (H* pitch accent) and uninformative (L* pitch accent) pragmatic meanings. Grassmann and Tomasello (2007) used an eye-tracking methodology to compare how 2-year-old children's fixations to referents that were given relative to those that were new in the context varied depending on the placement of prosodic prominence. The results showed that children looked longer at new referents only when they were associated with prosodic highlighting (a LH*

pitch accent). Ito, Bibyk, Wagner, and Speer (2014) also used eye-tracking to assess the effect of a contrastive pitch accent (L+H*) and non-emphatic accent (L*) on the resolution of a referent. Their results showed that 6- to 11-year-old children's detection of the target element was facilitated by the more prominent L+H* accent.

An important milestone in the development of comprehension abilities in children is that of inference resolution. Comprehension abilities have been shown to be related to inference resolution abilities in written (Cain, Oakhill & Bryant, 2004; Oakhill & Cain, 2012) as well as oral language (Currie & Cain, 2015). Currie and Cain (2015) assessed whether 6- to 10-year-old children's vocabulary comprehension and working memory abilities predicted their ability to infer information from elements that were not explicitly mentioned in the oral discourse (see (1) in methods section below for an example similar to the one used in Currie & Cain, 2015). The results of their study showed that this inference resolution ability develops between ages 6 and 8. Moreover, children's increasing comprehension of vocabulary as they got older mediated their ability to infer information. Regarding the impact of co-speech gestures on pragmatic inferential abilities, Kelly (2001) found that 3- to 5-year-old children could make use of deictic gesture information in contexts of pragmatic inference, such as indirect requests (e.g., saying "Don't forget, it's raining" while pointing to a raincoat). Kirk, Pine, and Ryder (2011) investigated whether information conveyed by representational gestures helped children with language impairment (and aged-matched typically developing children) to understand a verbal message. Participants

listened to short narrations in speech-only and speech+gesture conditions. The results showed that children's comprehension, particularly for those with language impairment, was helped by representational gestures. Thus, observing gestures which are semantically co-expressive with speech improve pragmatic inference resolution seems to boost children's language comprehension as well as their pragmatic ability to generate inferences. However, to our knowledge, no study has assessed the potential beneficial effect of beat gestures, which are devoid of semantic content, on the ability to make a pragmatic inference.

The first goal of our study is therefore to investigate whether children's ability to process information that is not explicitly stated verbally (i.e., a global coherence inference; Schmidt & Paris, 1983; Currie, 2014) can be improved by online exposure to multimodal markers of prominence (i.e., beat gestures combined with prosody). Given the evidence noted above regarding the beneficial effect of markers of prominence on language abilities in general, we expect that such markers will also enhance the ability to make inferences. In other words, we expect an improvement in the processing of a global coherence inference when the relevant clues in the discourse are marked with beat gestures and prosodic markers of prominence.

A second aim of our study is to disentangle the effect of prosodic marking alone from the effect of prosodic marking accompanied by

beat gestures. If we assume that gesture and speech belong to the same system (McNeill, 1992) in accordance with embodied cognition theory, which underlines the importance of body movements and multimodal supporting channels in cognition (see Barsalou et al., 2003, Barsalou, 2008; Hostetter & Alibali, 2010), we expect stronger effects when prominence is expressed multimodally rather than unimodally. Otherwise, if both prosodic cues alone trigger a similar effect than visual beat cues, this can be interpreted as a strong indicator that prosodic features drive a beneficial effect on inferential resolution, as has been shown by previous studies. Moreover, the study will also explore on the interaction between a child's age and the effect of multimodal markers of prominence on their ability to make an inference.

3.2. Methods

a) Participants

Seventy-eight 6- to 8-year-old children participated in the study (see Table 1). This particular age range was chosen because it is during this period that not only inference-making abilities (Currie & Cain, 2015) but also the ability to use beat gestures (Mc Neill, 1992; Shattuck-Hufnagel, Ren, Mathew, Yuen & Demuth, 2016; Igualada, Esteve-Gibert & Prieto, 2017) is undergoing development. All participants were recruited from two Catalan public schools located in the province of Barcelona, Spain.⁹ Parents were informed about the experiment's goal and signed a participation consent form. Furthermore, language exposure questionnaires (based on Bosch & Sebastián-Gallés, 2001) were administered to the caregivers in order to ensure that the children were predominantly exposed to Catalan on a daily basis (mean exposure to Catalan as a percentage of all daily language exposure: $M = 85.65$, $SD = 17.51$). Parental questionnaires reported that children were healthy and had normal hearing and vision.

⁹ Escola Sant Martí in Arenys de Munt and Escola Pública Dr. Estalella Graells in Vilafranca del Penedès.

Children	N	Age in months	Boys	Girls
6-year-olds	26	$M = 75.73$ ($SD = 6.39$)	17	9
7-year-olds	26	$M = 87.46$ ($SD = 3.99$)	12	14
8-year-olds	26	$M = 105.12$ ($SD = 5.74$)	13	13
Total sample	78	$M = 89.44$ ($SD = 13.3$)	42	36

Table 1. The sample population broken down into age groups, showing Mean (M) and Standard Deviation (SD) for age in months, and gender.

b) Materials

Thirteen stories (i.e., one familiarization trial and twelve experimental trials) were created to assess whether the children would be able to infer a specific concept that was not explicitly stated verbally (i.e., to generate a global coherence inference; Schmidt & Paris, 1983; Currie, 2014) on the basis of two lexical clues. Each story consisted of six short sentences that contained the two clues (an opening clue in the fourth sentence and a specific clue in the fifth sentence) that allowed the children to make the global coherence inference (see sample story (1)). The stories'

macrostructure and the position of all the lexical clues within the sentences and discourse were controlled for, with similar structures across stimuli (all the stories used in the experiment are reproduced in full in the Appendix).

Let us examine the translation of a sample story in (1). In order to make the appropriate inference leading to the target concept (in this case “mouse”), the listener would need to activate the semantic concepts related to the two lexical clues, “animal” and “cheese”, which were both marked for prominence by the speaker. The opening clue “animal” provides general information about the concept in the fourth sentence, and the specific clue “cheese” serves to disambiguate the concept in the fifth sentence. No other words in the story (e.g., verbs related to the concept and pronouns) provide any clues about the target. A third word, in the first sentence there is one word that is verbally expressed in the story but completely unrelated to the target concept (i.e., the literal concept). In this case the word “door” in the first sentence (i.e., the information expressed literally in the story), was also marked for prominence in order to control for all the items in the visual world paradigm. The grammatical category of the three words receiving prominence was controlled for (they were all nouns), as was their syntactic relation to the verb (all complements), and their position in the sentence (final position). The rhythmic pattern of the specific clues (nine monosyllabic and four disyllabic words) was also controlled for by having the stressed syllable in the first syllable. At the end of each

story there was an inference-tapping question intended to prompt children to make the target inference.

(1) English translation of a sample story. Words assigned prosodic prominence appear in capital letters.

Pau was waiting near the DOOR. (Literal word)

(He) saw his father running,

and (he) thought about what he could do.

His father was trying to catch an ANIMAL. (general clue)

(He) wanted to catch it with some CHEESE. (specific clue)

He always has great ideas!

Inference-tapping question: Pau's father, what did he want to catch?

Target concept: mouse

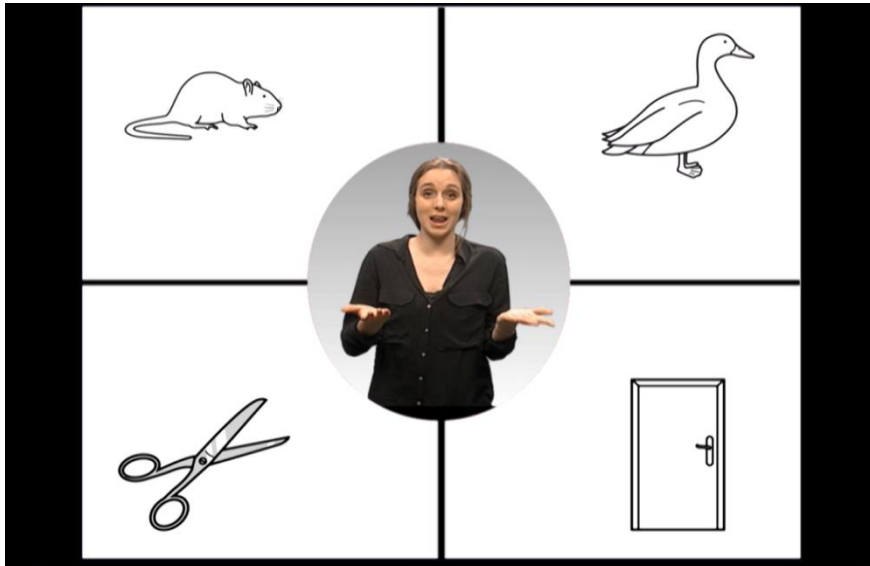


Figure 1. Still image from one of the stimulus videos which child participants were shown. In the center, the speaker is producing a beat gesture. Images in the four corners show a mouse (the target), duck (a competitor), door (item from the first sentence), and scissors (a distractor).

The stories were presented to the children in a within-subjects experimental design under three audiovisual conditions. The variable distinguishing the three conditions was the type of prosodic and gestural marking applied to the three target words in each story. In the control condition, the target words were produced without gestural information and with a flat low pitch accent (L^*) prosodic realization (see Figure 2, left panel). In the prosody-only condition, the target words were produced without gestural information but with a high (H^*) pitch accent over each of the target words (mid

panel). In the beat+prosody condition, the target words were produced with both the high (H*) prosodic realization and a beat hand gesture (right panel).

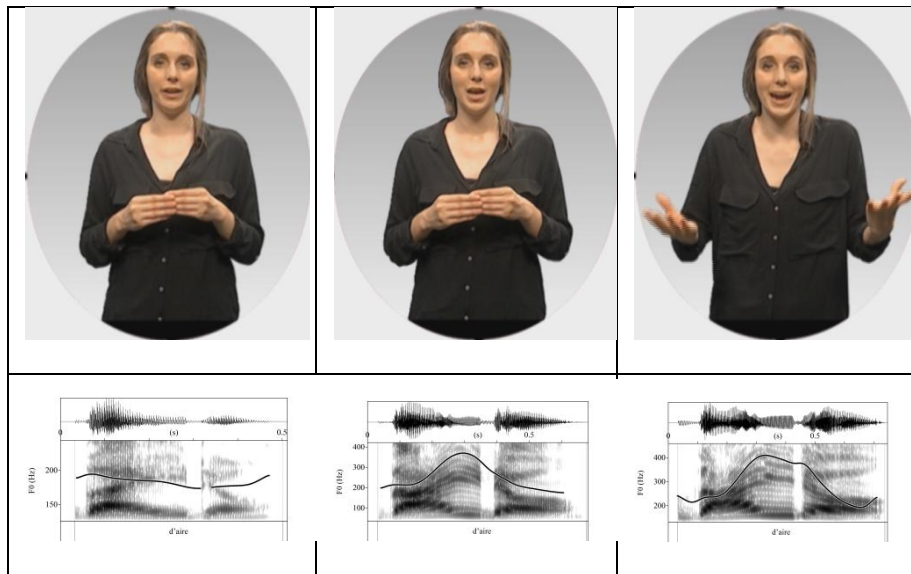


Figure 2. Examples of visual information and capture of the acoustic cues during the production of the specific clue *aire* ‘air’ in the control condition (left panel), the prosody-only condition (mid panel), and the beat+prosody condition (right panel).

The specific type of beat gesture to be used in the recordings of the materials was decided upon after conducting a Discourse Completion Task¹⁰ (Billmyer & Varghese, 2000) with 10 adult

¹⁰The Discourse Completion Task is a data elicitation method based on discourse contexts which has been applied for many years in research on pragmatics and sociolinguistics (see, e.g., Prieto and Roseano, 2010).

Catalan speakers, who were asked to memorize the two sentences from each story which contained the inference-tapping clues (e.g., sentences four and five in (2)) and then repeat them while looking at the video camera that was recording them as if speaking to a child. The adults did this twice for the full set of sentences. The first time, they were given no explicit instructions regarding their performance. The resulting recordings constituted the control condition for the Task. Then they were asked to repeat their performance but to stress the two target words. The recording that resulted in the case constituted the emphatic condition for the Task. Adults were not given explicit instructions about how to perform. Each story contained two key words, yielding a total of 24 multimodal productions per participant. These recordings were then analyzed (6 stories \times 2 words in each story \times 2 conditions) using the MIT Gesture Studies Coding Manual.¹¹

The results of this analysis showed that the adults produced significantly more beat gestures when asked to highlight the target words than in the control condition ($t(238) = -3.284, p < .01$), but they did not produce different amounts of representational gestures across conditions ($t(238) = -.413, p = .680$). The most commonly occurring beat gesture in the emphasis condition was the open palm gesture (60.5% of instances), with a forward palm orientation (52.2% of instances) and with an outwards trajectory movement (56.5% of the instances). Typically, this manual gesture was

¹¹ <http://web.mit.edu/pelire/www/manual/>

significantly most often accompanied by a widening of the eyes along with a raising of the eyebrows ($t(54) = -2.535, p < .05$; 12.5% of eye opening in control condition, and 65.2% in emphasis condition). However, this difference was not significantly different for the production of a head nod in the emphasis condition ($t(54) = -2.535, p < .05$; 25% of head nod productions in control condition, and 43.5% in emphasis condition). In accordance with these results, the beat gesture selected for the study consisted of an outward hand movement, with palm open, widened eyes, and raised eyebrows. To produce the full gesture, the hands were initially held close to the waist, palms inward and fingertips touching. Then forearms were swung outwards with elbows held close to the body until the palms were exposed, facing upwards at the apex of the gesture. The hands and arms then returned to their initial position (see Figure 2, right panel).

On the basis of this preliminary study, stimulus materials were prepared which consisted of video-recordings that included both an embedded video showing a speaker in the center of the frame and still images in the four corners. The video recordings of the speaker were made at a TV recording studio at the Universitat Pompeu Fabra and edited with the AVID video-editing programme. A female speaker was trained to produce the target sentences, reading them off a teleprompter. First, she recorded the frame story, and then the first, fourth and fifth sentences (which contained the target words) were recorded separately. She was trained to carefully produce the target words with identical prosodically prominent H*

realizations for the experimental conditions (i.e., prosody-only and beat+prosody conditions) and with a flat L* realization for the control condition. Thus a total of three recordings of the each target sentence were made, one with prominence marked by prosody only, another with prominence marked by prosody accompanied by the beat gesture described above, and the third without prominence marking of any sort. To control for the consistency of prosodic and gestural productions across conditions and stories, the first and last authors were present during the recording sessions. Subsequently, the first author compared the recordings of the experimental conditions from an auditory point of view and identified the best match for prosody between the two experimental conditions (i.e., prosody-only and beat+prosody). Acoustic analyses were performed by the first author using Praat (Boersma & Weenink, 2012) to analyse the pitch range, average intensity, and duration of the stressed syllable of the three target words in each of the selected sentences. The results of an ANOVA analysis showed the experimental/control conditions differed across the three acoustic measures, namely pitch range ($F(1, 35) = 36.201, p < .05$), intensity ($F(1, 35) = 21.259, p < .001$), and duration ($F(1, 35) = 9.372, p < .01$). As expected, Tukey post-hoc analysis revealed that the three acoustic parameters were similar between the prosody-only and beat+prosody conditions, but significantly different between the control condition and the two experimental conditions. The mean (M) and standard deviation (SD) values for the analysis of the acoustic features of the stress syllable in the target words across

conditions were as follows: for pitch range, control condition: $M = 41.23$, $SD = 24.85$; prosody-only condition: $M = 173.49$, $SD = 49.87$; beat+prosody condition $M = 171.44$, $SD = 51.04$; for intensity, control condition: $M = 72.24$, $SD = 4.25$; prosody-only condition: $M = 78.58$, $SD = 1.31$; beat+prosody condition $M = 79.35$, $SD = 2.45$; and for duration analysis, control condition: $M = 269.82$, $SD = 60.3$; prosody-only condition: $M = 386.97$, $SD = 110.52$; beat+prosody condition $M = 421.5$, $SD = 91.79$.

To prepare the visual world experiment, for each story, the target concept was paired with a competitor concept (i.e., a concept sharing the same semantic relation to the opening clue but very tenuously related to the specific clue; for example, in (1) *mouse* and *duck* are both animals), and a distractor (i.e., an concept unrelated to any of the inference-tapping clues; in (1), a pair of scissors has no connection with a mouse). (All these items are listed in the Appendix.). Items were matched according to frequency of occurrence in a corpus of written Catalan using the NIM¹² search engine (Guasch, Boada, Ferré & Sánchez-Casas, 2013). The relative frequency of the word (i.e., frequency of occurrence of the word in parts per million) was significantly similar between the opening and specific clues ($t(24) = .386$, $p = .703$; opening clue: $M = 98.17$, $SD = 126.27$; specific clue: $M = 78.72$, $SD = 130.55$). The relative frequency of the word was also significantly similar between the interactions of the target ($p = 1.00$), competitor ($p = 1.00$) and

¹² NIM <http://psico.fcep.urv.es/utilitats/nim/eng/valorescat.php>

distractor ($p = 1.00$) lexical items ($F(1, 2) = .147, p = .86$; target $M = 30.77, SD = 35.59$; competitor $M = 24.02, SD = 29.14$; distractor $M = 28.16, SD = 31.03$).

Finally, Adobe Premiere and Photoshop Pro CS6 software were used to edit the audiovisual materials in a visual world paradigm similar to that used in Silverman, Bennetto, Campana, and Tanenhaus (2010) (see Figure 1). Black and white images illustrating the target, competitor, literal, and distractor concepts were retrieved from the Aragonese Portal of Augmentative and Alternative Communication¹³. The pixel weight load of images was also controlled for to ensure that one image was not more likely to attract a viewer's attention than another between target images, the analysis showed that their weight in pixels was statistically similar ($F(1, 3) = 2.088, p = .115$; target $M = 251.71, SD = 57.64$; competitor $M = 287.50, SD = 43.06$; literal $M = 228.18, SD = 84.28$; distractor $M = 249.04, SD = 42.25$). The audiovisual recordings of the speaker were manipulated in each of the 13 stories to make sure that there was a pause of 1000 milliseconds (i.e., 25 frames) following each of the three target words, which were always in sentence-final position, and a pause of 3000 milliseconds (i.e., 75 frames) following each inference-tapping question. Copies of frames with images in rest position and muted sound were duplicated when the pause produced naturally by the speaker was not sufficient to complete the 25 or 75 frames. Finally, sentences

¹³ ARASAAC <http://arasaac.org/index.php>

were concatenated to build the story. The second, third, and sixth sentences also the inference-tapping question were copies of the same file across conditions.

Three presentations were created to counterbalance a potential effect of story item and condition. To do this, the twelve experimental stories were grouped into three blocks of four stories (i.e., block 1 included items 1, 2, 3, and 4; block 2 included items 5, 6, 7, and 8; and block 3 included items 9, 10, 11, and 12). In presentation 1, the control condition was assigned to block 1, the prosody-only condition to block 2, and beat+prosody condition to block 3; in presentation 2, the beat+prosody condition was assigned to block 1, the control condition to block 2, and the prosody-only condition to block 3; and in presentation 3, the prosody-only condition was assigned to block 1, the beat+prosody condition to block 2, and the control condition to block 3.

c) Procedure

At the beginning of the session each child was randomly assigned to one of the three presentations. Participants were tested individually in a quiet school classroom. They were asked to wear headphones and to sit approximately 50 cm. in front of a Tobii X2-60 Eye Tracker, which was attached to a laptop computer. Tobii Studio 3.2.2. was used to present the video stimuli. Stimuli videos were centered and adjusted to a screen size of 1920×1080 pixels. The

visual angle of each object subtended approximately 15°, well above the 0.5° accuracy of the eye tracker.

Once each participant was seated, a five-point calibration procedure lasting approximately 20 s was automatically carried out by Tobii Studio. Participants then started with the global coherence inference task. They were told by the experimenter that they would hear a person telling stories with a riddle, and that they should listen carefully because after each story they would be asked to guess the riddle. After asking the inference-tapping question, the experimenter took note of the offline response on an answer sheet. A first trial served to get the child familiarized with the task before they were presented with a total of 12 experimental trials. Tobii Studio was set to present the videos in a random order within three blocks of videos. At the end of each block, a short presentation of a colored drawing served to encourage the child to continue with the task. The position of the four images was randomized across trials to avoid children's anticipation of the position of the target item. The full session lasted approximately 12-15 minutes.

d) Analysis

Gaze data was extracted using Tobii Studio 3.2.2. The outputs included information about the name of the recording (Recording Name), the timestamp counted from the start to the end of each video recording (Recording Timestamp), the type of eye movement

(i.e., fixation, saccade, and unclassified events) classified by the filter settings applied during the gaze data export (Gaze Event Type), and the horizontal and vertical coordinates of the averaged left and right eye gaze point on the screen (Gaze Point X and Gaze point Y, respectively). The beginning of a pause was set as the target moment of disambiguation (i.e., silence after the specific clue; see example in (2)) in which audiovisual information was similar for all three conditions. The areas of interest were defined for each target picture as the set of all screen coordinates that fell within the squared shape surrounding the picture (see Fig. XXX). The eye tracking device is able to record 60 instances per second, with each gaze instance equal to 16.66 milliseconds. In this study, gaze data which the Tobbi software qualified as saccade or "unclassified event" was excluded, and only gaze instances classified as "fixation" were analyzed. Then gaze instances were grouped into 200 millisecond bins (see Arnold, Eisendband, Brown-Schmidt & Trueswell, 2000) and time intervals prior to and during pauses was broken into time windows (TW), with one prior to the pause (TW -1) and five TWs constituting the pause itself (TW 1, 2, 3, 4, and 5). Thus, a maximum of twelve gaze instances could fall within a 200 milliseconds bin. Finally, trials with wrong responses in the offline answer to the inference-tapping question were excluded from the analysis as they were assumed to possibly lead to different patterns of gaze response.

3.3. Results

In order to assess the effect of condition (i.e., control, prosody-only and beat+prosody conditions) and age on the children's ability to disambiguate target images from competitor images, a series of Generalized Linear Mixed Model (GLMM) were applied to the data for all TW (i.e., TW -1, 1, 2, 3, 4, and 5). In each case, the dependent variable was set as the number of gaze instances, in all stories (Poisson distribution, log link), for each participant and condition. Areas of interest (henceforth AOI, two levels: target and competitor), Condition (three levels: control, prosody-only and beat+prosody gesture), Age (three levels: 6, 7, and 8 years of age), and their interactions were set as fixed factors. Participant was set as a random factor. Pairwise comparisons were carried out for significant main effects and interactions.

The results of the GLMM analysis run for TW -1 (i.e., from -200 ms to the beginning of the pause) did not reach significance levels for any of its main effects (Condition $F(2, 274) = .455, p = .635$; AOI $F(1, 274) = .401, p = .527$) or interactions (Condition \times AOI $F(2, 274) = 2.060, p = .129$, Condition \times AOI \times Age $F(4, 274) = 1.584, p = .179$) (See table 2). Regarding the main effect of AOI, the GLMM analysis run for TW1 to TW5 showed a significant effect of AOI in the temporal windows 1 to 5 (i.e., beginning to end of the pause after disambiguation of the specific clue). There were a greater number of instances of gaze directed towards the target

image than to the competitor image at TW 1 ($F(1, 320) = 5.436, p < .05$), TW2 ($F(1, 344) = 16.102, p < .001$), TW3 ($F(1, 358) = 75.561, p < .001$), TW4 ($F(1, 338) = 84.907, p < .001$), and TW5 ($F(1, 304) = 74.097, p < .001$). These results showed an online preference of gaze directed towards the target image after hearing the specific clue, which helped to resolve the inference. Apart from the significant effect of AOI, the results of the GLMM analysis related to TW1 did not show a significant effect of Condition ($F(1, 320) = .835, p = .435$), interaction of Condition x AOI ($F(2, 320) = .676, p = .509$), or interaction of Condition \times AOI \times Age ($F(4, 320) = 1.789, p = .131$).

Main effects	Previous	Disambiguation during the pause				
	TW -1	TW 1	TW 2	TW 3	TW 4	TW 5
Condition	$p = .635$	$p = .435$	$p = .258$	$p = .535$	$p = .911$	$p = .911$
AOI	$p = .527$	$p < .05$	$p < .001$	$p < .001$	$p < .001$	$p < .001$
Condition \times AOI	$p = .129$	$p = .509$	$p = .055$	$p = .420$	$p = .838$	$p = .335$
Triple interaction	$p = .179$	$p = .131$	$p < .05$	$p = .082$	$p = .140$	$p = .140$

Table 2. P-values for the relevant main effects and interactions for the disambiguation of the AOI for all time windows (TW).

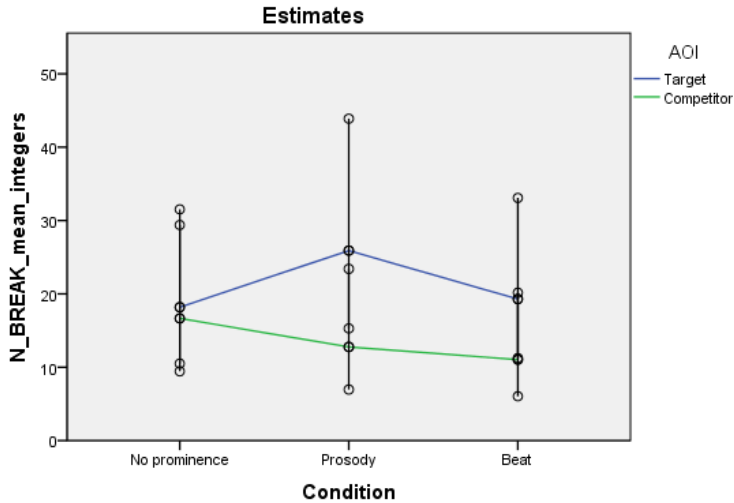


Figure 3. Number of responses towards target and competitor areas of interest during TW2.

See Figure 3 for a graph depicting the number of gazes directed at target and competitor areas of interest during TW2. The GLMM analysis with the number of instances during TW 2 (i.e., from 200 to 400 ms after disambiguation) did not show a significant effect of Condition $F(2, 344) = 1.358, p = .258$). However, this was the only analysis showing a near significant value in the interaction between Condition and AOI $F(2, 344) = 2.933, p = .055$). Post-hoc analysis of this interaction showed a tendency for children’s gazes to be

directed toward the target image in greater proportion than to the competitor image for both the prosody-only ($p < .05$) and the beat+prosody ($p < .05$) conditions, but the preference for the target image in TW2 was not significant for the control condition ($p = .634$). Moreover, the interaction between Condition, AOI, and Age reached significant levels $F(2, 344) = 3.098, p < .05$. Post-hoc analyses showed that 7-year-old children's gaze instances were significantly more often directed towards the target than to the competitor image only in the prosody Condition ($p < .05$). No other post-hoc analysis of the interaction between Condition, AOI, and Age reached significance levels.

The GLMM analysis run with the number of gaze instances during TW3 (i.e., 400 to 600 ms after disambiguation) did not show a significant effect of Condition $F(2, 358) = .626, p = .535$ or interaction between Condition and AOI $F(2, 358) = .870, p = .420$. However, the triple interaction between Condition, AOI, and Age reached a close to significant level $F(2, 358) = 2.085, p = .082$, which suggests that the disambiguation of the inference may depend on Age and Condition. Post-hoc analyses showed that 6-year-olds looked significantly more at the target than at the competitor image in the beat+prosody condition ($p < .05$), but not in the control ($p = .063$) and prosody-only ($p = .055$) conditions. Seven-year-old children looked more at the target image in the prosody-only condition ($p < .05$), but not in the control ($p = .056$) and beat+prosody ($p = .110$) conditions. And 8-year-old children looked more at the target than at the competitor image in both prosody-only

($p < .05$) and beat+prosody ($p < .05$) conditions, but not in the control condition ($p = .060$).

The GLMM analysis run with number of gaze instances during TW4 and TW5 did not show significant effects other than AOI as specified before. For TW4, there was not a significant effect of Condition ($F(2, 338) = .093$ $p = .911$), Condition \times AOI ($F(2, 338) = .177$ $p = .838$) or Condition \times AOI \times age ($F(4, 338) = 1.745$ $p = .140$). For TW5, there was not a significant effect of Condition ($F(2, 304) = .093$ $p = .911$), Condition \times AOI ($F(2, 304) = 1.098$ $p = .335$) or Condition \times AOI \times Age ($F(2, 304) = 1.746$ $p = .140$).

3.4. Conclusions

The aim of this study was to investigate the role of multimodal markers of prominence (i.e., prosody accompanied by beat gestures) in children's ability to draw an inference on the basis of information not explicitly mentioned in an oral discourse (i.e., their ability to make a global coherence inference). For example, if an adult hears that someone wants to catch an animal (general clue) by using some cheese (specific clue), they will infer that the target animal is a mouse (target concept). To detect the accuracy of their inferences online, we showed 78 6- to 8-year-old children video recordings of a speaker telling stories in the three different conditions with four images only one of which depicted the correct inference to be drawn and then measured the length of time they spent looking at these images.

Overall, the results of the inference resolution task showed that the children directed their gazes more often towards the target image in comparison to the competitor image. This occurred in the pause after they had heard the specific clue. As this was the first time in the story that children could possibly infer the target concept, we interpreted this result as evidence of online processing of the inference. To our knowledge, no other studies have previously used this method to assess children's online processing of an inference in a discourse and therefore complements previous studies (Currie, 2014) in which the online processing of inference was measured with reaction time rather than gaze time measures.

An important question addressed by this study was the relative benefit of gestural markers of prominence (i.e., beat gestures) in association with prosody versus prosodic prominence alone, and also in comparison with no prominence marking at all, in the online processing of pragmatic inferences. Previous studies assessing the effect of beat gestures by means of behavioral tasks (Austin & Sweller, 2004; Igualada et al., 2017; Llanes, et al., under revision) did not disentangle whether the beneficial effects of beat gestures on language processing are due to a multimodal effect on perception, or to a potential understanding of prosodic features which serve to provide prominence.

Regarding the potential effect of prominence markers (i.e., beat gestures and prosody) in the online resolution of the inference, the results showed near-significant and significant effects of condition in interaction with the target and competitor areas of interest during the second time window (i.e., TW 2, 200-400 ms) in the pause after participants heard the specific clue. Although values did not achieve significance ($p = .055$), there was a clear tendency to increase the number of gazes towards the target image in the beat-prosody and prosody-only conditions but not in the control condition. We interpret this tendency as showing that prosody and beat prominence markers have a facilitating effect on the online resolution of inference and therefore more generally facilitate language comprehension. Thus, our first hypothesis that the

multimodal condition (i.e., beat gestures plus prosody) would improve recall was not confirmed, and children were not more efficient in their responses by having this information in both speech and gestural modalities (see Barsalou et al., 2003, Barsalou, 2008; Hostetter & Alibali, 2010). However, prosody-only and beat+prosody conditions had similar positive tendencies on the online resolution of the inference, which the control condition did not.

With respect to of the relation between children's age on their online processing of pragmatic inference, 7- year-old children showed a significant effect (during TW 2) of prosodic prominence on their ability to disambiguate the target referent online. This can be interpreted as further evidence that prosodic features are a driving force in inferential resolution, as has been shown by previous studies (Fraundorf, et al. 2010; Ito, et al. 2014; Tomlinson, et al. in press). The results here are in fact similar to recent findings reported in Kushch, et al. (under revision), whereby second language memorization seems to be enhanced by beat gestures only when they are associated with patterns of prosodic prominence. While prosody and beat gestures have been shown to create different neurological activations related to language development (e.g., Holle Obermeier, Schmidt-Kassow, Friederici, Ward & Gunter, 2012; Hubbard, Wilson, Callan, & Dapretto, 2009; Wang & Chu, 2013), future studies still need to replicate this finding with behavioral evidence.

In this study, we used a replication of Silverman et al.'s (2010) visual world paradigm to assess online resolution of the inference at the precise moment of disambiguation. Future studies might attempt to measure a potentially stronger effect by controlling for the fact that viewers were faced with competing targets (i.e., the video of the speaker and the images) for their visual attention during the study, namely the speaker and the still images depicting objects (Gullberg & Holmqvist, 2006). Nevertheless, all in all the evidence provided in this study supports the notion that beat gestures and prosody operating together as a multimodal marker of prominence tend to enhance children's online pragmatic processing of language.

4. CHAPTER 4. LANGUAGE DEVELOPMENT AT 18 MONTHS IS RELATED TO MULTIMODAL COMMUNICATIVE STRATEGIES AT 12 MONTHS

4.1. Introduction

Gesture-speech integration is an important feature of human communication. As McNeill (1992) noted, in human languages gesture and speech modalities are coordinated not only at the temporal and phonological levels (i.e., the most prominent part of the gesture is typically aligned with the most prominent part of speech), but also at the semantic and pragmatic levels (i.e., the two components can share similar semantic functions). Infants begin to use simultaneous gesture-speech combinations intentionally near the end of the first year of life, a few months after the onset of canonical babbling and typically preceding the beginning of the one-word production stage (Carpenter, Mastergeorge & Coggins, 1983; Butcher & Goldin-Meadow, 2000; Esteve-Gibert & Prieto, 2014). The presence of these combined, multimodal communicative behaviors have been taken as an indicator of intentional communication, representing a step further on the way towards linguistic communication (Bates, Camaioni, & Volterra, 1975; Bates, Benigni, Bretherton, Camaioni, & Volterra, 1979; Wetherby & Prizant, 1989). But research is still needed concerning the prevalence and pragmatic function of simultaneous gesture-speech

combinations in specific socio-communicative contexts and their potential predictive value for later language development.

The developmental pathway of simultaneous gesture-speech combinations was studied in Esteve-Gibert and Prieto (2014). The study showed that at 11 months infants already produced simultaneous gesture-speech combinations, but pointing without speech still occurred more frequently. In their longitudinal sample they also found a significant increase in gesture-speech productions by 15 months of age. These multimodal productions mostly involved pointing and reaching gestures with a declarative communicative purpose, and when combined with speech, the two modalities were temporally coordinated in an adult-like way. The use of simultaneous gesture-speech combinations may serve to provide redundant information about the same referent through multimodal means, thereby highlighting a particular piece of information and minimizing joint effort in a communicative context (see Wagner, Malisz, & Kopp, 2014, for a review). In other words, infants may intentionally use multimodal strategies to mark a prominence in their communicative productions, a behavior that favors joint attention processes.

There is a considerable body of evidence that infants' joint attention abilities are linked to later language development (Tomasello & Farrar, 1986; Mundy & Gomes, 1998; Tomasello, 1988; Laakso, Poikkeus, Katajamäki, & Lyytinen, 1999; Kristen, Sodian,

Thoermer, & Perst, 2011). Studies have provided evidence that caregivers' contingent interactions (e.g., those that follow on the infant's focus of attention) tend to elicit more pointing and speech combinations by infants (e.g., Miller & Lossia, 2013; Miller & Gros-Louis, 2013) and also lead to better language abilities later in development (Tomasello & Farrar, 1986; Rollins, 2003; Tamis-LeMonda, Bornstein, & Baumwell, 2001; McGillion, Herbert, Pine, Keren-Portnoy, Vihman, & Matthews, 2013). These results provide indirect evidence about the potential relationship between an infant's multimodal communicative ability to initiate joint attention (i.e., to communicate and influence an adult's attention regarding an intended referent) on the one hand and the infant's later language abilities on the other.

Literature addressing early infants' communication abilities has typically focused on separate analyses of either gestures (and mainly pointing gestures) or speech modality but not both. For example, the ability to use pointing gestures has been regarded as a clear and powerful non-verbal strategy to also initiate joint attention between the infant and the adult with regard to an object or event (Tomasello, Carpenter, & Liszkowski, 2007). Likewise, research on infants' gesture production has shown that communicative gestures (e.g., iconic and pointing gestures) signal intentional communication (Bates et al., 1979; Bavin, Prior, Reilly, Bretherton, Williams, Eadie, Barret, & Ukoumunne, 2008; Caselli, Rinaldi, Stefanini, & Volterra, 2012) and that pointing gestures with a

declarative intention are a good predictor the emergence of verbal language (Colonnesi, Stams, Koster, & Noom, 2010;). On the other hand, literature on speech development has also documented that acoustic measures of early infants' vocalizations vary depending according to their communicative intentionality (Esteve-Gibert & Prieto, 2013) and that vocalizations coordinated with gaze directed at the referent affect adult-infant social interactions and support language learning (Goldstein, Schwade, Briesch, & Syal, 2010; Gros-Louis, West, & King, 2014). While some studies with slightly older infants have shown that one particular use of supplementary gesture-speech combinations--that in which the gesture modality conveys a different meaning than the one conveyed by speech--predicts the onset of grammatical development (Capirci, Iverson, Pizzuto, & Volterra, 1996; Iverson & Goldin-Meadow, 2005; Özçaliskan & Goldin-Meadow, 2005; Pizzuto, Capobianco, & Devescovi, 2005; Rowe & Goldin-Meadow, 2009), the emergence of simultaneous gesture-speech combinations (i.e., gesture co-occurring with speech to express the same meaning) and their relation to later language development has not been extensively analyzed.

In this study we are interested in exploring the link between the early ability to intentionally produce simultaneous pointing-speech combinations in specific communicative contexts and later language development. To our knowledge only two studies have explored the predictive role of early simultaneous gesture-speech combinations on later language development. In Murillo and Belinchón (2012), a

sample of eleven parent-infant dyads were recorded interacting in a semi-structured play context at three longitudinal moments, namely at 9, 12, and 15 months. The results showed that the use of pointing gestures at 12 months, especially when accompanied by vocalizations and directed gaze on the part of the infant, correlated positively with vocabulary development at 15 months of age. In a recent study, Wu and Gros-Louis (2014) analyzed the spontaneous interactions of 10- to 13-month-old infants with their mothers in fifty-one dyads and showed that the infants' combinations of vocalization and pointing, and especially those produced when mothers were not attending to the target event, were related to the infants' subsequent comprehension skills at 15 months. It should be noted that both of the studies mentioned above are based on the analysis of spontaneous interactions, where it is difficult to behaviorally control for two important aspects of early communicative patterns, namely, (a) the pragmatic intention or motive behind children's use of pointing gestures to comment on an event or object; and (b) the social interaction gaze patterns used by the adult during the communication. In this study we will attempt to address this issue by controlling for these two factors. To do so, we will examine pointing gestures that express a declarative intention (i.e., the communicator engages with the recipient to share information with him/her about something) by using a task that was specifically designed to elicit this behavior in infants, namely the declarative pointing task (Carpenter, Nagell, & Tomasello, 1998).

Liszkowski et al., (2008) showed that the communicative behaviors of 12- and 18-month-olds were affected by the patterns of adult attention to both the child and the event of reference. The authors measured children's communicative responses in two experimental social interaction conditions involving differences in an adult's availability in relation to the infant. The behavioral procedure used in the study consisted of a declarative pointing task which took into account different social conditions in order to control for the adult's joint attention patterns (Carpenter et al., 1998; Matthews, Behne, Lieven, & Tomasello, 2012). In the baseline condition the adult jointly engaged with the infant's event of reference, while in the critical conditions, the adult either looked at the infant but not at the object of reference (available condition), or was not visually attending to either of them (unavailable condition). The results of the study revealed that infants pointed significantly more, and produced more vocalizations during and after the infant's first point in the available and unavailable conditions than in the baseline condition. Moreover, the adult's social interaction patterns during the unavailable condition triggered less pointing behavior than during the available condition. Therefore, infant communicative responses changed depending on adult attention behaviors. Similarly, Gros-Louis and Wu's (2012) analysis of 12-month-old infants' interactions with their mothers showed that the children were more likely to combine vocalizations with pointing when mothers were not looking at the target event. These studies suggest that infants use pointing gestures and speech intentionally in specific communicative situations, and that they seem to efficiently

adapt their communicative behavior to the adults' availability for joint attention.

The current study investigates the predictive value of simultaneous pointing-speech combinations in different social interaction contexts. We will test whether the use of multimodal cues to attract an adults' attention in a communicative context is an important ability related to later language development. Following Liszkowski et al. (2008) and using a similar procedure (i.e., a declarative pointing task carried out under social conditions that differ according to adult joint attention patterns), we aim more specifically to investigate the role of adult interaction patterns in the integrated use of pointing gesture and speech by 12-month-old infants. As already mentioned, the declarative pointing task is especially suited to our purposes for two main reasons: (a) it elicits the production of declarative pointing in a situation in which infants are likely to initiate joint attention with the adult about an event of reference (Carpenter et al., 1998 Matthews et al., 2012); and (b) it controls for the adult's joint attention patterns, thus increasing the child's opportunity to produce simultaneous gesture-speech combinations in more demanding social conditions where the adult is either available (but not jointly attending to the child's object of reference) or unavailable. According to previous research, we expect a greater use of early multimodal productions during the available condition compared to baseline, but also differences in the use of gesture and speech combinations between the available and unavailable

conditions. Liszkowski et al. (2008) reported measures of the vocalizations produced during and after the first pointing behavior. Since in their study the vocalizations produced during the first point (multimodal productions) and after the first point (unimodal productions) were grouped together, their results do not reflect a clear measure of the use of pointing-speech combinations which are the focus of the present research. The first goal of our study is thus to replicate and extend prior findings on the role of different social conditions in triggering multimodal communicative strategies, such as simultaneous pointing-speech combinations.

The second goal of the study is to explore the degree to which an infant's early ability to use multimodal, simultaneous pointing-speech combinations at 12 months of age predicts subsequent vocabulary acquisition, with measures at 18 months of age. By using an experimental task which favors a specific declarative intention from the child and which controls for the adult's patterns of responses, we aim at extending results elicited in research by Murillo and Belinchón (2012) and Wu and Gros-Louis (2014) through an experimental task where infants are not interacting with their mothers. This controlled scenario will allow us to more thoroughly analyze the connection between infants' communicative strategies and their language outcomes six months later (measures in those studies were obtained just three months later, at 15 months of age). It is important to point out that analyzing a child's interaction with an unfamiliar adult (as opposed to mothers or other habitual caregivers) can provide a stronger assessment of infants'

communicative abilities because their behavior in the task will not be influenced by prior experience or shared routines in infant-caregiver interactions. In line with previous research showing that the use of combinations of gestures and words are good predictors of both lexical and grammatical development (Iverson & Goldin-Meadow, 2005; Özçaliskan & Goldin-Meadow, 2005; Rowe & Goldin-Meadow, 2009), we expect that the early ability to use simultaneous multimodal combinations at 12 months of age in communicatively demanding social interaction situations will be positively correlated with measures of language growth at 18 months of age. It is suggested that infants' ability to successfully engage the adult in joint attention using a combined, multimodal strategy can increase language-learning opportunities from social interaction contexts.

4.2. Methods

a) Participants

A final sample of nineteen infants (N = 19; 12 boys and 7 girls) participated in this study. They had all been born at term, were healthy, and had normal hearing. They were followed longitudinally, first being tested at 12 months of age (mean age: 12;12; range: 11;23-12;27) on a task involving different interaction conditions and then contacted again at 18 months of age to obtain language outcome measures from parental reports. An additional seven infants were initially recruited and tested but had to be excluded from the final sample because of oral habits which interfered with the pointing activity (2) or crying (1), or because follow-up data on their language outcomes at 18 months was unavailable (4). All participants were raised in monolingual Spanish-speaking homes: six of them were recruited in a monolingual Spanish-speaking area (Albacete) from public nurseries and thirteen participants were recruited from the APAL Infant Lab database in Barcelona. Results from the language exposure questionnaire (Bosch & Sebastián-Gallés, 2001) administered to the caregivers ensured that even the infants recruited in Barcelona, a linguistically mixed city, were predominately spoken to by their parents in Spanish on a daily basis. Exposure to a different language, if present, was sporadic and restricted to encounters with people outside the home environment (percentage of overall exposure to Spanish: median: 100 %; mean

90.4%; SD: 12.2). Since the minimum exposure to a second language at home required to qualify infants as bilingual is 25% (Bosch & Sebastián-Gallés, 2001), participants in the present study could be safely qualified as monolinguals. When initially contacted, all families reported that their infant had already begun to point at objects, an essential eligibility criterion for participation in this experiment.

b) Experimental setting and materials

The experimental setting was based on Liszkowski et al. (2008). Experimental sessions took place in a 2 x 3.5 m distracter-free testing room where the portable set-up could be easily placed. A large opaque curtain divided the room into two unequal areas, a central one where the caregiver and experimenter were seated and a small one behind the curtain, where an assistant, hidden from the infants' view, manipulated the objects to be presented during the task. The infant sat on his or her caregiver's lap in the central area at a distance of 2 m from the curtain, facing the experimenter, who had the curtain behind him. A small table was placed between the experimenter and the caregiver (see Figure 1). The caregiver wore a pair of earphones which continuously played music to distract him or her from the activity and avoid interference with the child's spontaneous behavior.

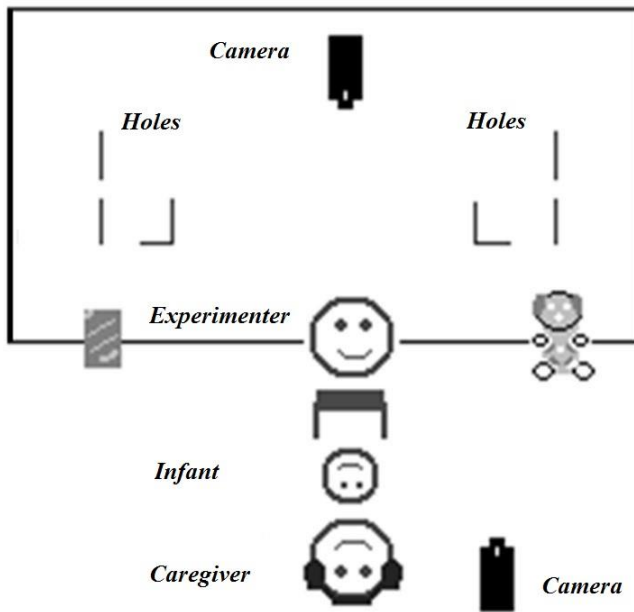


Figure 1. Schematic representation of the central area within the testing room. The setting includes a curtain with six openings, three on each side, where six of the objects manipulated by an assistant hidden behind it were presented, two cameras (frontal and back position) and two additional objects placed in front of the curtain, on the floor, to the left and right of the experimenter. Locations of the experimenter, child, and caregiver are also indicated.

A video camera was positioned so that it could record the child's reactions through a large opening in the upper center of the curtain while a second camera was positioned at the back of the room to record the sequence of events as seen from the child's perspective. The curtain had three lateral openings on each side through which

the puppets were made visible to the child, one at a time (see Figure 1). These openings (four of them at a distance of 60 cm from the floor and two of them at 100 cm) were symmetrically positioned at about 30° and 25° respectively to the left and right of the infant's direct frontal view. A total of ten different stimuli were manipulated by an assistant hidden behind the screen. These stimuli followed the same characteristics as in Liszkowski et al. (2008), namely six similar hand puppets (a cat, a frog, a cow, a rooster, a sun, and a snail), two different hand puppets (an articulated mouth and a grandmother), and two electronic stimuli (a dancing pig and a light). The latter two were located on the floor approximately 30° to the infant's left and right. Both electronic devices had switches which allowed the assistant to activate and deactivate them from behind the curtain. These electronic stimuli remained inactive except for the trials in which they were used. The labels of all these objects are included in the Spanish version of MacArthur's Communicative Development Inventory for children aged 8 to 15 months (López-Ornat, Gallego, Gallo, Karousou, Mariscal, & Martínez, 2005). A moveable bead toy and a pair of books were used between conditions to return the infant's attention to the experimenter.

c) Procedure

Liszkowski et al.'s (2008) procedure was adapted to elicit a range of infant communicative behaviors (i.e., pointing-speech combinations, pointing-only, and speech-only productions) through

an enjoyable event, by presenting puppets or toys from behind the experimenter. The procedure involved social interaction in two experimental conditions (i.e., available and unavailable) and a baseline condition, these three conditions differing in terms of the adult's joint attention patterns as in Liszkowski et al.'s study:

- In the *baseline condition*, the experimenter jointly engaged with the infant and the stimulus. First, the stimulus was activated and the experimenter looked at the infant, ignoring the stimulus until the infant had pointed to it. After this first pointing the experimenter reacted with joint attention (i.e., by looking back and forth between the stimulus and the infant's face), pointed to the object, and said things like, "Oh..., look, it's a cat!" "Look! It's saying hi to you!" "Oh, a cat!"
- In the *available condition*, the experimenter looked at the infant but did not look at the stimulus. First, the stimulus was activated and the experimenter looked at the infant, ignoring the stimulus until the infant had pointed to it. After this first pointing the experimenter maintained eye contact with the infant and did not look at the stimulus, while saying "Hmm? What? What's there? Hmm?"
- In the *unavailable condition*, the experimenter attended to neither the infant nor the object. When the stimulus was activated the experimenter's attention was directed at neither the infant nor the stimulus, but rather at the book. Even when the infant pointed to the stimuli, the experimenter

continued looking at the book while saying “Hmm? What? What’s there? Hmm?”

The testing session was organized in the following way. First, caregivers were informed about the experimental procedure and permission to record was obtained. Then they received general instructions on how to behave during the task: they were instructed to hold the child gently but firmly on their lap in order to maintain the infant’s position constant but to avoid all interaction with the child during testing and to avoid looking at the curtain and the objects that would be appearing there. They were encouraged to sit calmly while listening to music through the headphones.

After instructions were given, the warm-up period began. This took place in a separate room and consisted of an enjoyable play activity (lasting from 5 to 20 minutes) between the experimenter and the infant. Then, accompanied by the caregiver, they moved to the testing room, where the pointing task was carried out. Before baseline trials, there was a short play period with the bead toy to keep the infant interested in the experimenter as a social partner. When the experimenter judged that the infant was relaxed and attentive, he gradually withdrew from the interaction and signaled to the assistant by means of snapping his fingers out of the infant’s sight that the first stimulus could be activated. The infant had 20 seconds within which to initiate a pointing gesture. When the infant

pointed for the first time, the stimulus continued to be activated for another 20 seconds or until he/she showed that he/she was no longer interested in the object by ceasing to look at it by more than 10 seconds. If no communicative behavior was produced in reaction to the stimuli (i.e., no gestures, vocalizations, or any combination thereof), the stimulus was withdrawn after the first 20 seconds had elapsed. In all cases, the assistant, who could see the behavior of the infant from one of the holes in the curtain, monitored the duration of the trial and signaled to the experimenter when the trial was finished by clucking her tongue. Before experimental trials, the experimenter and the infant shared a book activity until the infant was relaxed and attentive; then the experimenter gradually withdrew the activity, signaled to the assistant by means of a finger snap to activate the next stimulus while continuing to look at the book, which he held at the opposite side of the infant's field of view relative to where the stimulus was going to appear in front of the screen.

The within-subjects experimental design was organized as sequences of three different types of trials, always starting with a baseline condition, followed by the available and unavailable conditions in a counterbalanced order across participants (i.e., ten participants were tested in the Baseline-Available-Unavailable order and nine participants were tested in the Baseline-Unavailable-Available order). Each sequence was repeated 5 times, so that each child completed a total of 15 trials. The right or left side where the first stimulus appeared was also presented in a counterbalanced

order across participants. Side presentation was alternated from right to left in successive trials until 15 trials were completed. The electronic stimuli were placed at infant's left side (light stimuli) and the right side (dancing pig stimuli), they were only used in those trials when the assistant alternated the position to the left or right side, respectively. The other experimental stimuli (handheld puppets) were randomly protruded by the assistant through one of the three holes of each side. Five of the 10 stimuli were used twice in order to complete 15 trials. The order of presentation of the stimuli and the stimuli to be used for a second time were also randomly chosen by the assistant. The full experimental session lasted approximately 18 minutes. Following the session, parents were given instructions on how to complete the Spanish version of the MacArthur-Bates Communicative Development Inventories, Words, and Sentences section of the 16-30 months CDI (López-Ornat et al., 2005). They were contacted again and asked to fill out and return the form six months later, when their child was 18 months of age.

d) Coding and reliability

Coding was performed using ELAN software (Lausberg & Sloetjes, 2009), which is especially well suited for video annotations. Measures of communicative modality were separately obtained for baseline, available, and unavailable conditions. Behaviors corresponding to three different modalities were registered, namely,

pointing-only, speech-only, and pointing-speech combinations. In what follows, the specific criteria used for coding the infants' behavior are described.

Pointing-only. Only instances of pointing directed at the target stimulus of the trial were coded, while other communicative gestures (e.g., waving the hand to say “hello” or clapping hands) were not taken into account.

We followed Liszkowski et al.'s (2008) coding of pointing gestures (isolated or in combination with speech), in which pointing gestures were coded when the infant extended the arm (either fully or slightly bent) and index finger or open hand downwards (similarly to Brooks & Meltzoff, 2008, and Cartmill, Demir, & Goldin-Meadow, 2012).

Speech-only. This category included any vocalization produced by the infant except infants' fixed signals (e.g., cries, shouts, laughs or groans) or vegetative sounds (e.g., sneezes or burps) (Oller, Niyogi, Gray, Richards, Gilkerson, Xu, Yapanel, & Warren, 2010; Nathani & Oller, 2001). Following these authors, we coded vocalizations as independent utterances when they were separated by a silence longer than 300 ms. Then, following Goldstein, Schwade, Briesch, & Syal (2010) and Gros-Louis, West, & King (2014), we coded infants' gaze direction as pertaining to one of the following

categories: stimulus-directed (looking at the target stimulus presented in the trial), experimenter-directed (looking at the experimenter after seeing the target stimulus), looking-caregiver (looking at the caregiver in the moment after seeing the target stimulus), looking-other (looking at other objects in of the room like the books or the bead toy, to a stimulus visually present in the room like the pig or light, or to the caregiver/experimenter when they were not previously looking at the target stimulus). We only included as vocalizations those that clearly referred to the target stimulus, that is, were stimulus-directed, and also those that were experimenter-directed after the child had seen the stimulus.

Pointing-speech combinations. Simultaneous pointing-speech combinations are defined as sharing all pragmatic function, semantic content, and phonological temporal cues (McNeill, 1992; Butcher & Goldin-Meadow, 2000). In the latter regard, the stroke phase of the gesture must coincide with the interval of maximum effort in the gesture. We therefore classified communicative productions as simultaneous pointing-speech combinations by looking at their temporal alignment, so that vocalizations which overlapped with at least some portion of the stroke of the pointing gesture were coded as simultaneous (following McNeill, 1992; Gros-Louis & Wu, 2012; Esteve-Gibert, & Prieto, 2014). Such combinations were counted as such only when they were clearly directed at the target stimulus.

After coding, the number of occurrences of each communicative modality (speech-only, pointing-only, or pointing-speech combinations) per trial was obtained and their frequency was computed. Inter-rater reliability was assessed by two observers who had been trained for 2 hours in the coding procedure. Observers assessed a total number of 61 trials, which corresponds to 21.4% of the trials across conditions. Agreement for presence/absence of communicative productions in each trial was very high: overall agreement was 96% and the fixed-marginal kappa statistic was 0.90. Observers assessed a total number of 141 infant productions, which corresponded to 43% of the data. The overall agreement for the classification of communicative acts (141 items) into one of the three categories (namely, *speech-only*, *gesture-only* and *pointing-speech combinations*) was 95% and the fixed-marginal kappa was 0.94, indicating that there was substantial agreement among independent coders. Overall agreement for the classification by coders of infant gaze behaviors into one of the 4 categories (namely, *stimulus-directed*, *experimenter-directed*, *caregiver-directed*, and *other-directed*) was 88% and the fixed-marginal kappa statistic was 0.78, indicating that there was substantial agreement among independent coders.

Finally, to control for consistency of the experimenter's behavior within a given condition, we assessed whether the experimenter's expected speech and gesture performance in the three conditions was as defined in the procedure section. We did this by monitoring the experimenter's gaze, gestures, and speech across conditions. The

results showed that the experimenter used speech and gesture behaviors as defined in the procedure in 100% and 93.8% of cases, respectively. The following descriptive information shows that in the baseline condition the number of gazes directed at the stimuli after the infant's first pointing (Median = 4; mode = 4; Mean = 3.78; SD = 1.25) and the number of pointing gestures directed at the stimuli (Median = 2; Mode = 2; Mean = 2.14; SD = .787) was consistent across participants. In the experimental trials, the gaze behavior before and after infants' first point was correctly performed by the experimenter in 97.9% and 100%, respectively. In trials in which there was no pointing gesture, the experimenter directed his gaze correctly in 98.9% of the trials (that is, at the infant in the baseline and available conditions and at the book in the unavailable condition). Thus the experimenter consistently performed according to the prescribed behaviors in each condition.

4.3. Results

The results section is divided in two subsections, which correspond to the two main goals of this research: (1) the effects of social conditions on the production of simultaneous pointing-speech combinations at 12 months of age and (2) the predictive value of early pointing-speech combinations with regard to language development measures at 18 months.¹⁴

First, to check for the potential effects of the stimuli used on the infants' communicative productions, three GLMMs were run with number of communicative productions as a dependent variable. The results of the first analysis show that the number and modality of infants' communicative productions were not affected by the frequency of appearance of each stimulus type ($F(18, 807) = 1.243$, $p = .22$). A second analysis revealed that infants' communicative productions were not influenced by stimuli which were presented once (new) or twice (given) ($F(1, 835) = .287$, $p = .593$). The third analysis revealed that infants' communicative productions were not affected by the side of the screen in which the stimuli were activated ($F(1, 817) = .001$, $p = .973$). This shows that the variation in the number and location of the stimuli used in the procedure did not influence the number of infants' communicative productions.

¹⁴ The application of Generalized Linear Mixed Models (GLMM) is especially suitable for our data because this technique extends the linear model so that the target is linearly related to the factors and covariates via a specified link function, thus allowing the target to have a non-normal distribution and the observations to be correlated (e.g., West, Welch, & Galecki, 2007; Nouri, 2010). All statistical analyses were performed using SPSS Statistics 19.0 (SPSS Inc., Chicago IL).

a) Effects of social condition on the use of pointing-speech combinations

In order to assess the effects of the different social conditions on the number of occurrences per trial of speech-only, pointing-only, and pointing-speech combinations that were elicited in the baseline and the two experimental conditions (i.e., available and unavailable), we conducted a Generalized Linear Mixed Model (GLMM) with number of communicative productions (three levels: speech-only, pointing-only, and pointing-speech combinations) as the dependent variable (Poisson distribution, log link); social conditions (three levels: baseline, available, and unavailable conditions), communicative modality (three levels: speech-only, pointing-only, and pointing-speech combinations), and all their possible interactions as fixed factors; and subject, trial, and social conditions as random factors. Bonferroni paired post-hoc tests were carried out for the significant main effects and interactions.

A total of 322 communicative behaviors were coded, including speech-only productions ($N = 174$), pointing-only productions ($N = 76$), and pointing-speech combinations ($N = 72$). The following infants' behaviors were excluded from analysis: communicative behaviors not directed at the stimuli activated/protruded during the trial (speech-only $N = 90$; pointing-only $N = 23$; and pointing-speech combinations $N = 24$) (e.g., books, bead toy, pig, light, holes at the opposite side of the screen, or other points in the room); vegetative

and fixed vocal signals like cries or laughter produced during the trial ($N = 49$); and other communicative but non-stimulus-oriented gestures produced during the trial ($N = 62$) (e.g., greeting gestures, clapping gestures, or beat gestures accompanying speech).

The results of the GLMM analysis showed a main effect of communicative modalities ($F(2, 846) = 16.772, p < .001$), with a greater production of speech-only productions than the other two communicative modalities, that is, pointing-speech combinations and pointing-only productions ($F(2, 846) = 14.833, p < .001$). The mean number of communicative productions were the following. Baseline trials (speech-only: .46 [SD = .85]; pointing-only: .39 [SD = .69]; pointing-speech combination: .21 [SD = .52]); available trials (speech-only: .71 [SD = 1.21]; pointing-only: .26 [SD = .53]; pointing-speech combination: .41 [SD = .87]); unavailable trials (speech-only means = .61 [SD = 1.09]; pointing-only means = .17 [SD = .54]; pointing-speech combination means = .13 [SD = .36]).(see Figure 2).

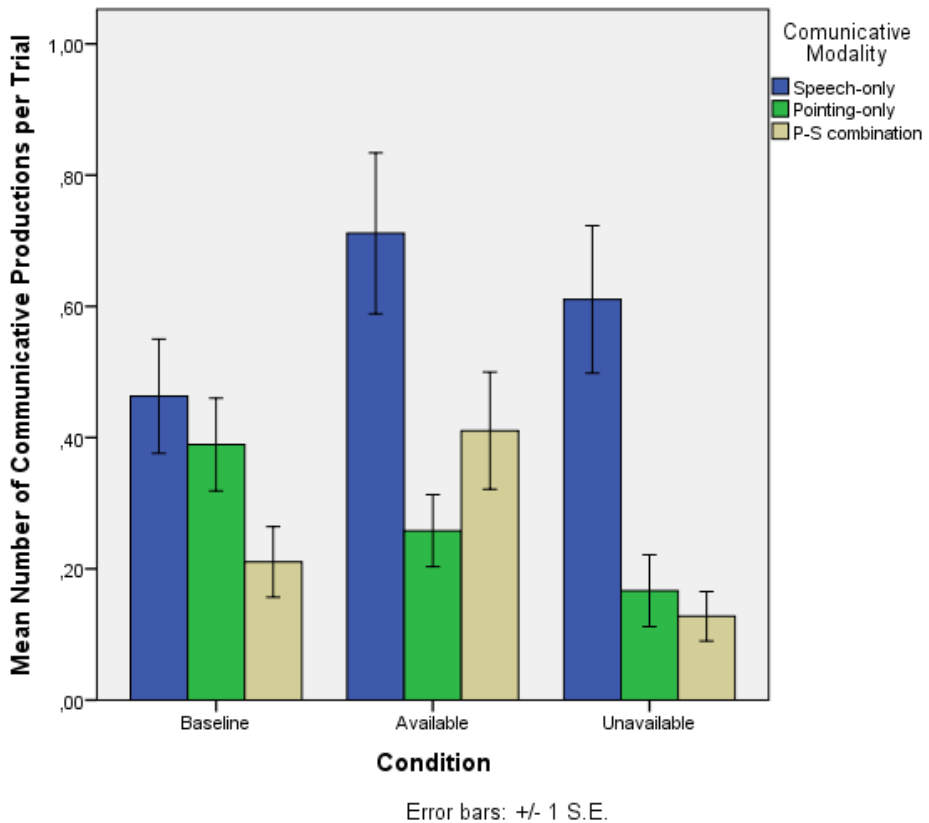


Figure 2. Mean number of occurrences of each type of communicative production (speech-only, pointing-only, and pointing-speech combinations) per trial as a function of social condition (baseline, available, and unavailable conditions). Error bars: +/- 1 S.E.

With respect to the distribution of communicative modalities in relation to the social conditions presented in the task, there was a

significant main interaction between social condition and communicative modality, ($F(4, 846) = 3.754, p < .01$), indicating that communicative productions behave differently depending on the social condition. More interestingly for our purposes, post-hoc analyses of the effect of social condition on the distribution of communicative modalities showed significant effects of simultaneous pointing-speech combinations, with a greater number of productions in the available condition than in the baseline or unavailable conditions, ($F(2, 846) = 6.292, p < .01$). By contrast, speech-only and pointing-only productions did not significantly differ between social conditions, respectively ($F(2, 846) = 2.816, p = .06$ and $F(2, 846) = .947, p = .388$). Thus, although speech-only productions noticeably increased from available to baseline condition, only simultaneous pointing-speech combinations increased in number in the available condition with respect to the baseline condition.

Post-hoc analyses revealed that communicative modalities did not show significant differences among themselves in frequency of production in the baseline condition (baseline $F(2, 846) = 2.111, p < .01$). Pointing-speech combinations were more frequently produced than pointing-only in available condition, and pointing-only was more frequently produced than pointing-speech combinations in the unavailable condition (available $F(2, 846) = 9.853, p < .001$; unavailable $F(2, 846) = 11.800, p < .001$). Again, speech-only productions occurred more frequently than the other

communicative productions in available and unavailable conditions, but they did not differ significantly between conditions.

Additionally, taking into account that there were two subgroups exposed to different orders of social condition, i.e., ten infants in order 1 (Baseline-Available-Unavailable trial order), and nine infants in order 2 (Baseline-Unavailable-Available trial order), the possibility of order effects has been analyzed. Infants' communicative productions were significantly different depending on social condition's order of presentation ($F(2, 837) = 5.342, p = .005$). Post-hoc analyses revealed that (a) children tested with order 2 produced a greater amount of speech-only production than those tested with order 1 ($F(1, 837) = 11.594, p = .001$), (b) infants tested with order 1 produced a greater amount of pointing-speech combinations than infants tested with order 2 ($F(1, 837) = 6.515, p = .010$), and (b) infants tested with both orders produced similar amounts of pointing-only productions ($F(1, 837) = .338, p = .561$).

b) Predictive value of simultaneous pointing-speech combinations for expressive language outcomes at 18 months

The results of the Spanish version of the MacArthur-Bates CDI that were obtained at 18 months of age (as reported by parents) are provided in Table 1. The expressive vocabulary section measures the total number of words that each child produced at that age,

while the other section measures the child’s ability to produce morphosyntactic features in their utterances. Table 1 shows the median, average, standard deviation (SD), minimum, and maximum CDI scores on vocabulary and grammar section for our sample of nineteen infants at 18 months. By that time, all participants had already begun to produce words, but morphosyntactic markers were still absent in some of them as reported by their parents (note the ranges in the last two right-hand columns). We did not include word-ending measures in the analysis because such responses were very infrequent (median = 0; mean = 1.63; SD = 2.4; min = 0; max = 8).

	Median	Mean	SD	Min	Max
<i>Expressive vocabulary section</i>	44	65.42	82.82	3	367
<i>Morphosyntax section</i>	6	11.21	15.07	0	67

Table 1. CDI scores of infants at 18 months as reported by parents.

For each participant the dependent variables, defined as the total number of communicative productions (separated into speech-only, pointing-only, and pointing-speech combination productions) in each of the three social conditions (baseline, available, and unavailable conditions) were obtained.

In order to analyze the predictive value of earlier pointing-speech combinations in different social conditions, one multiple regression analysis was run with each language measure (expressive vocabulary and morphosyntax) as dependent variables. Thus total of two multiple regression analyses were run with communicative productions (i.e., speech-only, pointing-only, and pointing-speech combination) uttered during different social conditions (i.e., baseline, available, and unavailable conditions) as independent variables, using a step-wise method. Table 2 shows the results of the two most effective models of communicative production uttered in trials with different social conditions (independent variables) for the prediction of different language measures at 18 months (dependent variables).

<i>Social condition</i>	<i>Communicative production included in the model</i>	<i>R² statistic (%)</i>	<i>β-typified</i>
<i>Model with expressive vocabulary measure at 18 months as dependent variable</i>			
Available	P-S combinations	30.2 *	.550 *
<i>Model with morphosyntactic measure at 18 months as dependent variable</i>			
Available	P-S combinations	29.5 *	.543 *

*p-value: p < .05 **

Table 2. Multiple regression analyses of the most effective models predicting infants' vocabulary and morphosyntax measures at 18 months based on early communicative productions at 12 months during a specific social condition. R² statistics and p-values are reported for each model.

As can be seen in the table, the results of the first regression model, which included communicative productions elicited in the baseline, available, and available conditions, showed significant effects of pointing-speech combinations during the available condition; specially, the model indicated that the number of pointing-speech combinations in the available condition at 12 months explained 30.2% of the expressive vocabulary variance ($R^2 = .302$, $F(1, 18) = 7.365$, $p < .05$). It was found that simultaneous pointing-speech combinations produced during the available condition were the best predictor of vocabulary measures at 18 months ($\beta = .55$, $p < .05$). The best model relating communicative productions to morphosyntactic measures reported significant differences in the measures of pointing-speech combinations in the available condition. The results of the regression model indicated that the number of pointing-speech combinations found in the available condition at 12 months explained 29.5% of the morphosyntax section variance at 18 months ($R^2 = .295$, $F(1, 18) = 7.365$, $p < .05$). In sum, it was found that the number of simultaneous pointing-speech combinations produced during the available condition at 12 months was the best predictor of both vocabulary measures ($\beta = .55$, $p < .05$) and morphosyntax measures at 18 months ($\beta = .543$, $p < .05$).

4.4. Conclusions

This study set out to investigate the effects of specific contexts of social interaction on the elicitation of multimodal communicative abilities by 12-month-olds, and more importantly, to investigate the predictive value of the integrated use of speech and gesture during these social interaction contexts for later language abilities. We used a declarative pointing task (Carpenter et al., 1998) with controlled social interactions (as used in Liszkowski et al. 2008) to measure infants' communicative productions (i.e., speech-only, pointing-only, and pointing-speech combinations). The first set of results showed significant effects of social condition (baseline, available, and unavailable conditions) on the type of communicative productions produced by children. Infants used pointing-speech combinations in the available condition, i.e., the condition in which the adult looked at the infant and did not look at the event of reference, more frequently than in the baseline condition, i.e., the condition in which the adult showed gaze engagement between the infant and the event of reference. By contrast, the frequency of use of pointing-only productions did not change significantly between baseline condition and either available or unavailable condition. And though speech-only productions occurred more frequently than pointing-only and pointing-speech combinations, the frequency of oral productions did not differ between baseline condition and either experimental condition. Thus, simultaneous pointing-speech combinations were the only type of communicative production that

was significantly more frequent in the available condition than in the baseline condition.

Despite the observed interaction between order of presentation and communicative productions (see section 3.1), related to infants ability to adapt to the adult responses, the overall predominance of pointing-speech combinations in the available condition remains unaffected.

Interestingly, the ability to fully integrate gesture and speech did not increase when the adult failed to show joint attention in response to a communicative intent of the infant (i.e., in the unavailable condition). There was no increase in any communicative modality between the baseline and unavailable conditions. This result differs from Liszkowski et al.'s (2004, 2008) result, which showed an increase in the number of pointing productions when the adult was not available to communicate about a referent (unavailable condition) in comparison to trials where the adult shared joint attention with the infant about a referent (baseline condition). In their studies, the experimental conditions were tested between participants, so that trials were presented in the baseline and available conditions for one group of participants and in the baseline and unavailable conditions for a different group. In our study we tested trials belonging to three social conditions within participants (baseline and both experimental conditions). Therefore, the infants' social interaction opportunities to communicate with

adults in our study differed from those in previous studies. It is possible that testing the three conditions within participants (e.g., two trials in which the adult was available and one more trial in which the adult was not available to communicate) might have changed infants' communicative strategies such that they would employ predominately pointing behavior when the adult was ready to communicate. This could explain why the frequency of pointing by infants in this study was lower when the adult was not available than in previous studies. Nevertheless, our main results clearly show evidence that infants' use of pointing-speech combinations is dependent on adult attention patterns, with an increase in the number of multimodal productions when the adult is available to communicate but does not look at the referent of the infants' interest.

Our results extend those of Liszkowski et al. (2008) in confirming that the ability to use simultaneous multimodal combinations is employed by children as a repairing strategy to reinforce information related to their communicative goal when the adult does not share attention to the referent but is available to communicate with the infant. On the other hand, the difference between the amount of speech-only productions in the available and baseline conditions only approached but did not reach significance. This result differs from the one related to pointing behavior, whereby the difference in the number of pointing-speech combinations across the two conditions was significant. This

difference may be explained by the fact that the experimenter responded to pointing gestures in the baseline and experimental conditions but ignored vocalizations (e.g., the experimenter started to communicate only after the infant's first point). This might have reinforced and promoted the use of pointing gestures. Nevertheless, these results go in the same direction as those yielded by Gros-Louis and Wu (2012) and Wu and Gros-Louis (2014) through naturalistic observations, since they also noted an increase in simultaneous gesture-speech combinations when the adult was available to the infant but not attending to the event or object of interest. Therefore, the boosting effect of multimodal gesture-speech combinations in the available condition may be interpreted as a signal of the intentional ability of the child in a situation in which the adult has yet not seen the object but is crucially expressing an interest by looking at the child. This ability to deploy multimodal means might thus be a good reflection of a better ability to intentionally pursue a communicative goal (Bates et al., 1975; Bates et al., 1979; Wetherby & Prizant, 1989; Liszkowski et al., 2008).

Importantly, the second set of results revealed a predictive relationship between the capacity to produce early multimodal communicative productions at 12 months and language measures at 18 months. Two multiple regression analyses were run to test whether the type and frequency of multimodal communicative productions expressed during a specific social condition significantly predicted later expressive vocabulary and sentence

measures at 18 months. The use of simultaneous pointing-speech combinations elicited during the available condition at 12 months was the variable most predictive of expressive vocabulary and morphosyntactic measures at 18 months. Overall, results showed that these two measures of language development were significantly related to early multimodal productions: a total of 30.2% and 29.5% of the variance of vocabulary and morphosyntactic development at 18 months, respectively, could be explained by the frequency of use of simultaneous pointing-speech combinations obtained during the available condition at 12 months. Therefore, the ability to produce pointing-speech combinations was positively correlated to later language measures extracted from parental reports.

The results of this study support the hypothesis that pointing gestures synchronized with speech constitute evidence of a powerful joint engagement ability for infants which is related to later language development. In line with our results, the observation of naturalistic interactions in Wu and Gros-Louis (2014) also revealed that an infant's ability to produce multimodal utterances when an adult looked at the infant but not at an object of interest predicted later language abilities. We have thus extended previous results with the use of an experimental task which controls for social interactions with an unfamiliar adult and favors a given pragmatic intention from the child during the task (e.g., a declarative intention). One of the positive outcomes of our experiment is to show that a set of infants with different family

communicative backgrounds react in similar ways to specific social conditions. The fact that infants were interacting with an unfamiliar adult neutralized the possible influence of prior caregiver-child routines and shaped interaction. Moreover, with regard to the potential effects of the use of multimodal productions on later language development, our results extend the results yielded by Murillo and Belinchón (2012) and Wu and Gros-Louis (2014) to the age of 18 months, when early production of gesture-speech combinations are found to correlate with lexical and grammatical output at that stage of development.

Our first finding on the positive effects of the available condition for the use of simultaneous pointing-speech combinations backs up the hypothesis that infants' sensitiveness to the common conceptual ground of the interlocutor is expressed through multimodal cues (de Ruiter, 2000; Holler & Stevens, 2009; Tomasello, 2008). That is, infants at an early stage of pointing-speech multimodal development are able to adjust their response to their interlocutor's knowledge of their shared space. Moreover, this finding also relates to a substantial body of literature on how adult-infant joint attention processes affect infants' communication and language abilities (Tomasello & Farrar, 1986; Carpenter et al., 1998; Hoff, 2006). Yu and Smith (2012) noticed that cooperation between adults and infants favors the creation of optimal visual moments of language learning which reduce distracters from the scene. Likewise, recent studies have found that adult contingent responses (i.e., adult communication following the infant's focus of attention) are linked

to the ability to produce simultaneous gesture-speech combinations (Miller & Lossia, 2013; Miller & Gros-Louis, 2013). We still do not know how the adults' contingent interactions might affect and the infant's ability to persist with their communicative goal, as well as his or her multimodal abilities. Future studies could test whether or not caregiver-infant contingent interactions are related to the infants' abilities to achieve their communicative goals and their successful use of simultaneous gesture-speech combinations. The fact that the recipient's auditory and visual sensory channels are activated to share attention with the adult may serve as a strategy to reduce the number of distracters from the context. Our interpretation is that simultaneous gesture-speech combinations may work as an effective communicative strategy to highlight a piece of information and reduce ambiguity (see Wagner et al., 2014, for a review). As Goldin-Meadow, Goodrich, and Iverson (2007) suggested, pointing gestures reinforce speech by singling out the referent indicated by the accompanying speech.

Importantly, the results of the present study have shown that the ability to use gesture-speech multimodal integration as a communicative strategy at 12 months is related to later language development. Though firmer conclusions could be drawn on the basis of a greater sample, the results of this study provide important information about early language precursors. Gesture-speech integration may be an early indicator of intentional communicative efficiency in those situations where an infant intends to highlight a

piece of information and draw an adult's attention towards an object. This research has shown the importance of an infant's capacity to convey meaning simultaneously in two distinct modalities, gesture and speech, as a precursor of language development. That is, pointing in combination with early speech may be an important signal of intentional communication, in which semantic, pragmatic, and phonological information is integrated for the first time.

5. GENERAL DISCUSSION AND CONCLUSIONS

5.1. Summary of findings

The general aim of this thesis was to investigate whether temporal gesture-speech integration has a beneficial effect on a set of different linguistic abilities (i.e., word recall, pragmatic inference resolution) and whether the early production of gesture-speech integration can be related to later language outcomes assessed at 18 months of age. Three independent studies were carried out, each one presented in a different chapter. We adopted different methodological approaches to assess offline and online behavioral responses from children at three different points in development. Adopting these approaches allowed us to investigate children's responses to naturalistic tasks containing experimentally controlled features which are relevant to the object of analysis in this thesis: temporal gesture-speech integration.

The first two studies focused on whether perceiving temporal gesture-speech synchronizations functioning as markers of prominence had a beneficial effect on children's word recall and pragmatic inference resolution abilities (see Chapters 2 and 3). The third study investigated whether infants' first uses of temporally

synchronous gesture-speech instances (e.g., pointing-speech combinations) predicted later language development (see Chapter 4).

Regarding improvement in language abilities when perceiving a target word synchronized with a beat gesture and its concomitant prosody, two main results were obtained. First, in Chapter 2 we found that 3- to 5-year-old children improved recall of a target word synchronized with a beat gesture in a child-relevant discourse context. In this study, we also found that the beneficial effect in the recall of words was bound to the recall of the synchronized target word (i.e., local effect), and not to the adjacent words. Second, in Chapter 3 we found that 6- to 8-year-old children's processing times of a pragmatic inference were reduced when the target word (i.e., a specific clue which solved the inference) was temporally synchronized with a beat gesture and its concomitant prosody. We also found that the beneficial effect on the processing of the inference was similarly impacted by prominences expressed both by prosody and beat gestures.

Finally, with respect to the first uses of intentional synchronous gesture-speech combinations, two main results are reported in Chapter 4. First, we found that young infants (i.e., 12 months of age) produce pointing gestures synchronized to speech sounds more frequently depending on social interaction, concretely when the adult interlocutor visually attended to him/her but was not aware of

the child's object of interest. Second, we showed that this early communicative strategy to use synchronous gesture-speech productions with the purpose of sharing joint attention with the adult was predictive of later language abilities at 18 months of age. As a whole, this chapter provides evidence on the beneficial impact of this early multimodal prominence strategy on later language development.

In the next sections I will discuss the findings of this thesis in relation to the previous literature and show how they contribute to the existing body of research specifically with regard to (a) the impact of temporally synchronous prominences on language processing, (b) the interaction between gesture and prosody, and finally (c) its linkage to language development.

5.2. Temporal gesture-speech synchrony as a marker of prominence in language processing

Research on the integration of gesture and speech has typically provided evidence supporting their interaction at a semantic level. That is, representational gestures and speech modalities are semantically co-expressive, have been shown to be related and thus benefit different language abilities across the lifespan. In this thesis, we hypothesize that non-representational gestures also play a relevant role in language processing and language development.

Two main conclusions can be drawn from the results of the first two studies included in this doctoral dissertation. First, that gesture-speech prominence markers temporally aligned with a target word improve, at the very least, the two linguistic processes assessed in this thesis, namely word recall and global coherence inference resolution. While the previous behavioral evidence showed contradictory results with respect to a potential impact of beat gestures on word recall (So et al., 2012; Austin & Sweller, 2014), our findings showed that prominent gesture-speech alignments embedded in a discourse context enhance not only word recall but also pragmatic comprehension processes. In line with this, recent evidence from research carried out in our team supports these results with positive effects of beat gestures on language comprehension (Llanes-Coromina et al., under revision) and in narrative abilities in preschool children (Vilà-Giménez, Igualada & Prieto, under revision).

In Llanes-Coromina et al.'s (under revision) first experiment, 4-year-old children observed stories with contrastive words (e.g., "...in the lake there were fishes and ducks, and you picked up the fishes... What did he pick up?") in three conditions, namely prominence in speech and gesture, prominence only in speech, non-prominence in speech and gesture, prominence only in speech, non-prominent speech. The results of a recall task showed a positive impact of prominence in speech and gesture on the recall of the target words with respect to the other two conditions. However, no significant differences were found between the recall rates for the non-prominent prosody and prominent prosody conditions. In the second experiment, 5-year-old children were exposed to recordings of child-directed narratives in two conditions, i.e., with target words presented with prominence in speech together with beat gestures (the beat condition) or without gestures (the no-beat condition). Children's responses showed results of better comprehension of the story when the narratives were presented in the beat gesture condition. In Vilà-Giménez et al. (under revision), 5- to 6-year-old children were exposed to a short training session in which they were exposed to a set of narratives in two between-subject conditions, namely the beat condition (i.e., in which the focal elements were highlighted with a beat gesture) and the no-beat condition (i.e., in which the focal elements were not highlighted with beat gestures). The results showed that children in the beat condition produced narratives with better narrative structure and better fluency scores than children in the no-beat condition. All in all, children's

behavioral responses showed a beneficial effect of observing gesture-speech temporal prominences on language narrative production.

The second conclusion is related to one of the findings in the first study, which shows that the multimodal prominence effect exclusively improves word recall of the co-occurring element (e.g., a local effect), not recall of the adjacent information. These results are in line with recent neurophysiological data showing a clear local time alignment between brain responses and target speech associated with beat gestures (Biau et al., 2016; Dimitrova, et al., 2016). Current neurophysiological evidence has related this beneficial effect on language processing to an enhancement of attention towards the target information co-occurring with the gesture (Biau et al., 2016; Holle et al., 2012; Hubbard et al., 2009; Wang & Chu, 2013). Moreover, Dimitrova et al. (2016) showed that temporal gesture-speech prominences function for specific words related to the context of the sentence and not others. In fact, this study showed that attentional responses decreased when non-target words were presented with a beat gesture. Thus, the scope of the beat gesture is temporally aligned with its co-occurring information, but crucially beat gestures are functionally dependent on pragmatic context.

Why is it that beat gestures facilitate language processing abilities? One potential explanation is that redundant multimodal integration

cues facilitate speech perception, thus facilitating language processing (Lewkowicz & Hansen-Tift, 2012; van Wassenhove, et al., 2007). Thus, the effect of temporally synchronous gestures and speech on language processing could be attributed to the integration of cross-modal perception processes (Biau, et al., 2016; Hubbard, et al., 2009). In fact, the ability to selectively attend to specific elements of speech while disregarding others (i.e., temporal attention) has recently been argued to facilitate language development in its early stages (de Diego-Balaguer, Martinez-Alvarez & Pons, 2016). A second explanation, complementary to the previous one, is that the visual and speech prominence encoded by beat gestures marking linguistically relevant functions (e.g., focus marking) have a potential effect on language processing. This means that, first, in a situation of prominence, a particular element needs to stand out from the surrounding elements (Terken, 1991), and second, that the focused element reflects certain properties of the discourse context (Büring, 2007). The proposal in this thesis is that prominence expressed with beat gestures is heavily dependent on surrounding speech elements, as well as discourse context, to fully express the semiotic value of the beat gesture (McNeill, 1992).

To conclude this section, although some research has proposed that beat gestures' semantic meanings express concepts, such as thoughts or ideas (i.e., positive and negative points about ideas in political discourse) (Casasanto & Jasmin, 2009), or are related to spatial information expressed in the verbal content (Yap & Casasanto,

2017), in my view the evidence suggests that gestures can be multidimensional (i.e., serve different functions, such as deictic, representational, conventional or beat functions). However, when beat gestures are isolated from semantic information, they serve to add pragmatic information regarding the relevance of a particular element (McNeill, 1992).

5.3. Effects of prosodic prominence and gestural prominence

As mentioned in the previous section, there is a tight temporal synchrony between beat gestures and prosodic prominence (i.e., pitch accents) (Krahmer & Swerts, 2007; Loehr, 2012; Shattuck-Hufnagel, et al., 2016; Yasinnik et al., 2004). Temporal gesture-speech synchronizations have been shown to naturally occur in association with prosodic markers of prominence in production (e.g., Loehr, 2012; Yasinnik et al., 2004). Moreover, both adults and infants are sensitive to asynchronies between gestures and speech, which is usually perceived as less prominent (Leonard & Cummins, 2011). These temporal synchronizations have been shown to serve a prominence function (e.g., Krahmer and Swerts, 2007; Rochet-Capellan, et al., 2008; Rusiewicz, 2010; Yasinnik, et al. 2004).

Despite this interdependence, previous studies assessing the effects of beat gestures on a variety of behavioral tasks (Austin & Sweller, 2014; Igualada et al. 2017; So et al., 2012; Vilà-Giménez, under revision) have not disentangled whether the beneficial effect of beat gestures on language processing abilities was due to either a multimodal effect on perception, or to a potential effect of prosodic prominence.

The question of the potential difference in weight between gestural and prosodic cues of prominence was addressed by the second study presented in this thesis. This study compared the potential effects of three conditions in the online processing of pragmatic inferences by children, namely, the effects of gestural markers of prominence (i.e., beat gestures) associated with their natural prominent realization, as compared to the effects of prosodic-only markers of prominence and the control non-prominent patterns. To our knowledge, this is the first study to explore this research question with a child population. In this study, the results showed that children performed better (at near significant levels) in both prominence conditions (i.e., the beat gesture condition and the prosody-only condition) than in the control condition. Moreover, children were not more efficient in their responses when they had this information in both modalities (e.g., speech and gesture), as would have been expected in this thesis. Since the prosody-only and beat conditions had similar facilitating effects on the online resolution of pragmatic inferences, this can be interpreted as a strong indicator that prosodic features drive a clear beneficial effect on inferential resolution, as has been shown by previous studies with adult populations (Fraundorf, et al. 2010; Ito, et al. 2014; Tomlinson, et al. in press).

Research has still not agreed on the potential differences of the weight of prosodic and gestural prominence markers. To our knowledge, only three behavioral studies have controlled for the effects of prosodic prominence and gestural prominence on

linguistic processing abilities. First, in Kushch et al. (under revision), adults participated in a second-language word learning task to assess whether the potential effect of the beat gesture was related to gestural information or to prosodic prominence (i.e., focal pitch accent). Second, in Kushch & Prieto (2016) adults were asked to recall specific words from a contrastive discourse. The results in both studies showed an increase in word recall when the information was associated with multimodal prominence (i.e., gesture and L+H* pitch accent) and prosodic prominence in isolation (i.e., L+H* pitch accent) in comparison with the no-prominence condition (i.e., no gesture and L* pitch accent). Moreover, the presence of gestural prominence added a beneficial effect on information recall in comparison with prosodic prominence in isolation. In Llanes-Coromina et al. (under revision), which used a similar task to the one in Kushch & Prieto (2016), 5-year-old children remembered more words when they were presented with multimodal prominence compared to either of the other two conditions (prosody with and without prominence: L+H* vs. L* pitch accents). However, prosodic prominence by itself did not differ from the condition without prominence (See experiment 1 by Llanes-Coromina et al. (under revision) in section 5.2).

Regarding neurological evidence, on the one hand, Holle et al. (2012) showed ERP evidence of gesture-specific effects related to beat gestures on a task assessing the ability to solve a syntactic ambiguity. That is, the effect was not facilitated by prosodic

prominence but only for the visual information of the beat gesture. Dimitrova et al. (2016) asked adult participants to identify contrastive focused information. The results of the ERP study showed that, in this particular linguistic context, beat gestures showed a positive activation when associated with the focused target word co-occurring with a pitch accent. However, listeners' integration cost rose when the beat gesture was associated with non-focused information. Hubbard et al. (2009) found with fMRI that the perception of beat gestures during fluent speech was greater when speech and beats were integrated than the sum of the speech alone and beat gesture alone conditions. The authors interpreted this response as a pivotal effect of the multimodal processing of beat gestures. Given these inconsistent results across studies, future research will need to further explore the question of the relative weight of prosodic and gestural prominence. Similarly, it would be important to test whether certain visual markers of prominence (i.e., particular head, eyebrow, or hand movements) are more powerful than others, or act like prosodic markers of prominence (Moubayed, et al., 2010).

5.4. The temporal synchrony rule: A predictor of language acquisition

In this thesis, we suggest that temporally synchronous gestures and speech (i.e., beat dimension) are intentional motor coordination movements signaling language functions and which can thus play a strong role in the production and comprehension processes in language. Previous literature in the early development of gesture-speech coordination has shown that by 6 months of age infants synchronize rhythmic arm movements with babbling (Ejiri & Masataka, 2001; Iverson & Fagan, 2004). Another milestone in the acquisition of temporal multimodal synchronization is the use of the pointing gesture temporally synchronously with speech sounds around their first year of life (Butcher & Goldin-Meadow, 2000; Esteve-Gibert & Prieto, 2014; Murillo & Belinchon, 2013). Importantly, infants' vocalizations have been shown to increase in complexity into more mature patterns when associated with a gesture (Esteve-Gibert & Prieto, 2014; Murillo & Capilla, 2016).

The third study of the thesis provides clear evidence on the relevance of producing synchronous gesture-speech combinations in language development. First, children produced multimodal synchronization patterns differently depending on social context, that is, depending on experimentally controlled joint attention interaction between adult and 12-month-old infants. Thus, infants increased their production of synchronous pointing gestures and

speech in those contexts in which the adult remained attentive by looking into the infant's eyes but did not visually check infant's referent of interest (i.e., available condition). The interpretation of this result is that infants use these pointing-speech combinations to highlight their focus of interest, and guide adults' attention by a multimodal communicative behavior (Wu & Gros-Louis, 2014). Thus, the embodied experience to produce an intentional synchronous multimodal combination might be motivated by intentional pragmatic forces. The fact that the situational context plays a relevant role in the early use of temporally synchronous prominences is a relevant motivation to hypothesize their potential relationship with later linguistic abilities.

The third study in this thesis showed that language abilities assessed at 18 months can be predicted by the early ability to produce vocalizations in synchrony with a pointing gesture. This evidence supports the idea that pointing gestures (temporally) synchronized with speech constitute a powerful joint engagement ability for infants which is related to later language development (see also Murillo & Belinchón, 2012; Wu & Gros-Louis, 2014). This finding relates to a substantial body of literature on how adult-infant joint attention processes affect infants' communication and language abilities (Tomasello & Farrar, 1986; Carpenter et al., 1998; Hoff, 2006). It is well known that infants' sensitiveness to the common conceptual ground of the interlocutor is expressed through multimodal cues (de Ruiter, 2000; Holler & Stevens, 2009; Tomasello, 2008). That is, infants at an early stage of pointing-

speech multimodal development are able to adjust their responses to their interlocutor's knowledge of their shared space.

Expanding on Dynamic System Theory (Iverson & Thelen, 1999), the coordination of the internal pulse based on the temporal entrainment of two motor systems plays a relevant role during social interactions. This intentional ability to express a communicative goal in multimodal ways might shape infants' embodied perception of the world in which one element is given prominence. Grounded cognition proposes that modal simulations, bodily states, and situated action underlie cognition (Barsalou, 2008; Gibbs, 2006; Shapiro, 2007). Multimodal perceptual experience of intentional communication might help children to develop their world of abstract symbols, and thus language. Research has also shown that both adults and infants create optimal learning moments through multimodal interactions (Yu & Smith, 2012). Additionally, infants tend to more frequently respond to adults' interests when they produce synchronized gesture-speech combinations (Miller & Lossia, 2013; Miller & Gros-Louis, 2013).

Finally, the findings of the studies presented in this thesis point to the importance of synchronous gesture-speech combinations in highlighting the communication of intentional prominence, as well as their beneficial effects in language acquisition. There is a need to extend the theoretical models to the role of non-representational

gestures in language production and language processing that integrate the evidence that the temporal synchronicities between gestures and speech are strongly related to linguistic and pragmatic functions.

In sum, future theoretical proposals on language development should explore the interactions of different properties relevant to multimodal communication. Interestingly, a recent proposal offers a reconciling perspective on the foundation of language evolution by proposing that the capacity to use language evolved from the interaction of both oral and gestural systems being used as one integrated system (Levinson, & Holler, 2014; McNeill, 2012; Vigliocco, Perniss, & Vinson, 2014).

6. REFERENCES

- Austin, E. E. & Sweller, N. (2014). Presentation and production: The role of gesture in spatial communication. *Journal of Experimental Child Psychology*, 122, 92-103.
- Barsalou, L. W., Simmons, W.K., Barbey, A. K., and Wilson, C. D. (2003) Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences*, 7(2), pp. 84-91. doi:10.1016/S1364-6613(02)00029-3
- Barsalou, L. W. (2008) Grounded cognition. *Annual Review of Psychology*, 59(1), pp. 617-645.
- Bates, E., Camaioni, L., & Volterra, V. (1975). The acquisition of performatives prior to speech. *Merrill-Palmer Quarterly*, 21, 205–226.
- Bates, E., Benigni, L., Bretherton, I., Camaioni, L. & Volterra, V. (1979). *The emergency of symbols: cognition and communication in infancy*. New York. Academic Press.
- Bavin, E.L., Prior, M., Reilly, S., Bretherton, L., Williams, J., Eadie, P., Barret, Y. & Ukoumunne, O.C. (2008). The Early Language in Victoria Study: predicting vocabulary at age one and two years from gesture and object use. *Journal of Child Language*, 35 (3), 687-701.
- Behne, T. Liskowski, U., Carpenter, M. & Tomasello, M. (2012). Twelve-month-olds' comprehension and production of pointing. *The British Journal of Developmental Psychology*, 30, 359-375.
- Biau, E., Fernández, L. M., Holle, H., Avila, C., & Soto-Faraco, S. (2016). Hand gestures as visual prosody: BOLD responses to audio–visual alignment are modulated by the communicative nature of the stimuli. *NeuroImage*, 132, 129-137.

- Billmyer, K., & Varghese, M. (2000). Investigating instrument-based pragmatic variability: effects of enhancing discourse completion tests. *Applied Linguistics*, 21(4), 517–552.
- Boersma, P., & Weenink, D. (2012). Praat: Doing phonetics by computer (Version 5.3.04). [Computer program] Retrieved from www.praat.org
- Bosch, L., & Sebastián-Galles, N. (2001). Evidence of Early Language Discrimination Abilities in Infants from Bilingual Environments. *Infancy*, 2, 29–49.
- Bott, L. & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*, 51, 437-57.
- Brooks, R. & Meltzoff, A.N. (2008). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: a longitudinal, growth curve modeling study. *Journal of Child Language*, 35 (1), 207-220.
- Büring, D. (2007). Semantics, intonation, and information structure. In: Ramchand G, Reiss C, eds. *The Oxford Handbook of Linguistic Interfaces*. Oxford: Oxford University Press, 445–473.
- Butcher, C. & Goldin-Meadow, S. (2000). Gesture and the transition from one-to-two word speech: when hand and mouth come together. In McNeill, D. (ed.). *Language and gesture*. New York: Cambridge University Press, 235-258.
- Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology*, 96, 31–42.
- Capirci, O., Iverson, J.A., Pizzuto, E. & Volterra, V. (1996). Gestures and words during the transition to two-word speech. *Journal of Child Language*, 23 (3), 645–673.

- Carpenter, R.L., Mastergeorge, A.M., & Coggins, T.E. (1983). The acquisition of communication intentions in infants eight to fifteen months of age. *Language and Speech*, 26, 101-116.
- Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society of Research in Child Development*, 63 (4), 1–143. Serial No. 255.
- Cartmill, E. A., Demir, O. E., & Goldin-Meadow, S. (2012). Studying gesture. In Hoff, E. (Ed.), *Research methods in child language: a practical guide*. Blackwell Publishing.
- Cartmill, E.A., Hunsicker, D., & Goldin-Meadow, S. (2014). Pointing and naming are not redundant: Children use gesture to modify nouns before they modify nouns in speech. *Developmental Psychology*, 50(6), 1660-1666.
- Casasanto, D., & Jasmin, K. (2010). Good and bad in the hands of politicians: Spontaneous gestures during positive and negative speech. *PLoS One*, 5(7), e11805.
- Caselli, M.C., Rinaldi, P., Stefanini, S. & Voterra, V. (2012). Early action and gesture vocabulary and its relation with word comprehension and production. *Child Development*, 83 (2), 526-542.
- Chen, A. (2011). The developmental path to phonological focus-marking in Dutch. In S. Frota, E. Gorka, & P. Prieto (Eds.), *Prosodic categories: Production, perception and comprehension* (pp. 93-109). Dordrecht: Springer.
- Church, R., Kelly, S., & Lynch, K. (2000). Immediate memory for mismatched speech and representational gesture across development. *Journal of Nonverbal Behavior*, 24, 151–174.
- Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational psychology review*, 3(3), 149-210.

- Cole, J. (2015). Prosody in context : a review Prosody in context : a review. *Language, Cognition and Neuroscience*, 30, 1-30.
- Colletta, J. M., Guidetti, M., Capirci, O., Cristilli, C., Demir, O. E., Kunene-Nicolas, R. N., & Levine, S. (2015). Effects of age and language on co-speech gesture production: an investigation of French, American, and Italian children's narratives. *Journal of child language*, 42(01), 122-145.
- Colonnesi, C., Stams, G.J., Koster, I. & Noom, M.J. (2010). The relation between pointing and language development: a meta-analysis. *Developmental Review*, 30 (4), 352-366.
- Cook, S. W., Yip, T. K., & Goldin-Meadow, S. (2010). Gesturing makes memories that last. *Journal of Memory and Language*, 63(4), 465-475.
- Currie, N. K., & Cain, K. (2015). Children's inference generation: The role of vocabulary and working memory. *Journal of Experimental Child Psychology*, 137, 57–75.
- Currie, N. (2014). *Children inference generation: A developmental investigation of the influence of vocabulary knowledge and memory* (Doctoral dissertation). Lancaster University.
- Dahan D., Tannenhaus M.K. & Chambers, C.G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language* 47: 292-314.
- de Diego-Balaguer, R., Martinez-Alvarez, A., & Pons, F. (2016). Temporal attention as a scaffold for language development. *Frontiers in Psychology*, 7(44), 1- 15.
- Demir, Ö. E., Fisher, J. A., Goldin-Meadow, S., & Levine, S. C. (2014). Narrative processing in typically developing children and children with early unilateral brain injury: Seeing gesture matters. *Developmental Psychology*, 50(3), 815–28.

- De Ruiter, J. P. (2000). The production of gesture and speech. In D. McNeill (Ed.), *Language and Gesture* (pp. 284-311). Cambridge University Press.
- Dimitrova, D., Chu, M., Wang, L., Özyürek, A., & Hagoort, P. (2016). Beat that word: How listeners integrate beat gesture and focus in multimodal speech discourse. *Journal of cognitive neuroscience*, 28(9), 1255-1269.
- Dohen, M. (2009). Speech through the ear, the eye, the mouth and the hand. In Esposito, A., Hussain, A., & Marinaro, M. (Eds.), *Multimodal signals: Cognitive and algorithmic issues* (pp. 24-39). Berlin/Heidelberg: Springer.
- Ejiri, K., & Masataka, N. (2001). Co-occurrences of preverbal vocal behavior and motor action in early infancy. *Developmental Science*, 4(1), 40-48.
- Esteve-Gibert, N. & Prieto, P. (2013). Prosodic structure shapes the temporal realization of intonation and manual gesture movements. *Journal of Speech, Language, and Hearing Research*, 56 (3), 850-864.
- Esteve-Gibert, N. & Prieto, P. (2014). Infants temporally coordinate gesture-speech combinations before they produce their first words. *Speech Communication*, 57, 301-316.
- Esteve-Gibert, N., Prieto, P., & Pons, F. (2015). Nine-month-old infants are sensitive to the temporal alignment of prosodic and gesture prominences. *Infant Behavior & Development*, 38, 126-9.
- Ekman, P. (1979). About brows: emotional and conversational signals. In von Cranach, M., Forra, K., Lepinies, W., & Ploog, D. (Eds.) *Human ethology: Claims and limits of a new discipline: Contribution to the Colloquium* (pp. 169-248). Cambridge: Cambridge University Press.

- Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology*, *27*, 209–221.
- Fraundorf, S. H., Watson, D. G., & Benjamin, A. S. (2010). Recognition memory reveals just how contrastive accenting really is. *Journal of Memory and Language*, *63*(3),367–386.
- Gibbs, R. W., Jr. (2006). *Embodiment and cognitive science*. Cambridge: Cambridge University Press.
- Goldin-Meadow, S., & Alibali, M. W. (2013). Gesture's role in speaking, learning, and creating language. *Annual Review of Psychology*, *64*, 257–83.
- Goldin-Meadow, S. & Butcher, C. (2003). Pointing toward two-word speech in young children. In Kita, S. (Ed.) *Pointing: Where Language, Culture, and Cognition Meet*. (pp. 85–107). Mahwah, NJ: Erlbaum.
- Goldin-Meadow, S., Cook, S. W., & Mitchell, Z. A. (2009). Gesturing gives children new ideas about math. *Psychological Science*, *20* (3), 267–272.
- Goldin-Meadow, S., Goodrich, W. G. & Iverson, J. (2007). Young children use their hands to tell their mothers what to say. *Developmental Science*, *10* (6), 778-785.
- Goldstein, M.H., Schwade, J., Briesch, J., & Syal, S. (2010). Learning while babbling: prelinguistic object-directed vocalizations indicate a readiness to learn. *Infancy*, *15* (4), 362-391.
- Grassmann, S., & Tomasello. M. (2007). Two-year-olds use primary sentence accent to learn new words. *Journal of Child Language*, *34*,.677–87.
- Gros-Louis, J., West, M.J. & King, A.P. (2014). Maternal responsiveness and the development of directed vocalizing in social interactions. *Infancy*, *19* (4), 385-408.

- Gros-Louis, J., & Wu, Z. (2012). Twelve-month-olds' vocal production during pointing in naturalistic interactions: Sensitivity to parents' attention and responses. *Infant Behavior & Development, 35* (4), 773–778.
- Guasch, M., Boada, R., Ferré, P. & Sánchez-Casas, R. (2013). NIM: A Web-based Swiss army knife to select stimuli for psycholinguistic studies. *Behavior Research Methods, 45*, 765-771.
- Guellai, B., Langus, A., & Nespors, M. (2014). Prosody in the hands of the speaker. In *Language by mouth and by hand* (Vol. 5, p. 700). Frontiers in Psychology.
- Gullberg, M., deBot, K., & Volterra, V. (2008). Gestures and some key issues in the study of language development. *Gesture, 8*(2), 149–179.
- Gullberg, M., & Holmqvist, K. (2006). What speakers do and what addressees look at: Visual attention to gestures in human interaction live and on video. *Pragmatics & Cognition, 14*(1), 53-82.
- Gullberg, M., & Kita, S. (2009). Attention to speech-accompanying gestures: Eye movements and information uptake. *Journal of nonverbal behavior, 33*(4), 251-277.
- Henry, L.A., Messer, D., Luger-Klein, S., & Crane, L. (2012). Phonological, visual, and semantic coding strategies and children's short-term picture memory span. *The Quarterly Journal of Experimental Psychology, 65*(10), 2033-2053.
- Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review, 26* (1), 55-88.
- Holle, H., Obermeier, C., Schmidt-Kassow, M., Friederici, A.D., Ward, J., & Gunter, T. C. (2012). Gesture facilitates the syntactic analysis of speech. *Frontiers in Psychology, 3*(74), 1-12.

- Holler, J., & Stevens, R., 2009. The effect of common ground on how speakers use gesture. *Journal of Language and Social Psychology*, 26 (1), 4-27.
- Hostetter, A. B., & Alibali, M. W. (2007). Raise your hand if you're spatial: relations between verbal and spatial skills and gesture production. *Gesture*, 7(1), 73-95.
- Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, 15, 495-514.
- Hostetter, A. B., & Alibali, M. W. (2010). Language, gesture, action! A test of the Gesture as Simulated Action framework. *Journal of Memory and Language*, 63(2), 245-257.
- Hostetter, A.B., Alibali, M.W., & Kita, S. (2007). I see it in my hand's eye: Representational gestures reflect conceptual demands. *Language and Cognitive Processes*, 22, 313-336.
- Huang, Y., & Snedeker, J. (2009). On-line interpretation of scalar quantifiers: Insight into the semantics-pragmatics interface. *Cognitive Psychology*, 58, 376-415.
- Hubbard, A. L., Wilson, S. M., Callan, D. E., & Dapretto, M. (2009). Giving speech a hand: gesture modulates activity in auditory cortex during speech perception. *Human Brain Mapping*, 30(3), 1028-37.
- Igualada, A., Bosch, L., Prieto, P. (2015). Language development at 18 months is related to multimodal communicative strategies at 12 months. *Infant Behavior and Development*, 39, 42-52.
- Igualada, A., Esteve-Gibert, N. & Prieto, P. (2017). Beat gestures improve word recall in 3- to 5-year-old children. *Journal of Experimental Child Psychology*, 156, 99-112.
- Ito, K. (2014). Children's pragmatic use of prosodic prominence. In McNeill, D. (ed.). *Pragmatic development in first language acquisition*. John Benjamins Publishing Company, 199-218.

- Ito, K., Bibyk, S. a, Wagner, L., & Speer, S. R. (2014). Interpretation of contrastive pitch accent in six- to eleven-year-old English-speaking children (and adults). *Journal of Child Language*, 41(1), 84–110.
- Ito, K., Jincho, N., Minai, U., Yamane, N., & Mazuka, R. (2012). Intonation facilitates contrast resolution: Evidence from Japanese adults & 6-year olds. *Journal of Memory and Language*, 66 (1), 265-284.
- Iverson, J. M. (2010). Developing language in a developing body: The relationship between motor development and language development. *Journal of child language*, 37(02), 229-261.
- Iverson, J. M., Capirci, O., Volterra, V. & Goldin-Meadow, S. (2008). Learning to talk in a gesture-rich world: early communication of Italian versus American children. *First Language*, 28,164–81.
- Iverson, J. M., & Fagan, M. K. (2004). Infant vocal-motor coordination: precursor to the gesture-speech system? *Child Development*, 75(4), 1053–66.
- Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science*, 16(5), 367–71.
- Iverson, J. M., & Thelen, E. (1999). Hand, Mouth and Brain. *Journal of Consciousness Studies*, 6(11–12), 19–40.
- Jannedy, S., & Mendoza-Denton, N. (2005). Structuring information through gesture and intonation. *Interdisciplinary Studies on Information Structure*, 3, 199–244.
- Järvikivi, J., Pyykkönen-Klauck, P., Schimke, S., Colonna, S., & Hemforth, B. (2015). Information structure cues for 4-year-old and adults: Tracking eye movements to visually presented anaphoric referents. *Language, Cognition and Neuroscience*, 29(7), 877-892.

- Kelly, S. D. (2001). Broadening the units of analysis in communication: speech and nonverbal behaviours in pragmatic comprehension. *Journal of Child Language*, 28, 325-349.
- Kelly, S. D., Ozyürek, A., & Maris, E. (2010). Two sides of the same coin: speech and gesture mutually interact to enhance comprehension. *Psychological Science*, 21(2), 260–7.
- Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In M. R. Key (Ed.), *The relationship of verbal and nonverbal communication* (pp. 207–227). The Hague, The Netherlands: Walter de Gruyter.
- Kendon, A. (1995). Gestures as illocutionary and discourse structure markers in Southern Italian conversation. *Journal of pragmatics*, 23(3), 247-279.
- Kendon, A. (2007). Some topics in gesture studies. In Esposito, A., Bratanić, M., Keller, E., Marinaro, M., *Fundamentals of Verbal and Nonverbal Communication and the Biometric Issue*. 2007. Amsterdam: IOS Press.
- Kirk, E., Pine, K. J., & Ryder, N. (2011). I hear what you say but I see what you mean: The role of gestures in children's pragmatic comprehension. *Language and Cognitive Processes*, 26(2), 149-170.
- Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and gesture* (pp. 162–185). Cambridge: Cambridge University Press.
- Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, 48(1), 16-32.
- Krahmer, E., & Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language*, 57(3), 396–414.

- Kristen, S., Sodian, B., Thoermer, C. & Perst, H. (2011). Infants' joint attention skills predict toddlers' emerging mental state language. *Developmental Psychology*, 47 (5), 1207-1219.
- Kushch, O., Igalada, A., & Prieto, P. (under revision). Prominence in speech and gesture favor second language novel word learning. *Language, Cognition & Neuroscience*.
- Kushch, O., & Prieto, P. (2016). The effects of pitch accentuation and beat gestures on information recall in contrastive discourse. *Proceedings of Speech Prosody*. Boston, USA, May 31 – June 3.
- Laakso, M.L., Poikkeus, A.M., Katajamäki, J. & Lyytinen, P. (1999). Early intentional communication as predictor of language development in young toddlers. *First Language*, 19 (56), 207-231.
- Ladd, D. R. (2008). *Intonational phonology* (2nd ed.). Cambridge and New York, NY: Cambridge University Press.
- Lausberg, H. & Sloetjes, H. (2009). Coding gestural behavior with the NEUROGES-ELAN system. *Behavior Research Methods, Instruments, & Computers*, 41 (3), 841-849.
- Leonard, T., & Cummins, F. (2011). The temporal relation between beat gestures and speech. *Language and Cognitive Processes*, 26, 1295-1309.
- Levinson, S. C., & Holler, J. (2014). The origin of human multi-modal communication. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 369(August), 20130302.
- Lewkowicz, D. J., & Hansen-Tift, A. M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences*, 109(5), 1431-1436.

- Liszkowski, U., Albrecht, K., Carpenter, K. & Tomasello, M. (2008). Infants' visual and auditory communication when a partner is or is not visually attending. *Infant, Behavior & Development*, 31 (2), 157-167.
- Liszkowski, U., Carpenter, M., Henning, A., Striano, T. & Tomasello, M. (2004). Twelve-month-olds point to share attention and interest. *Developmental Science*, 7 (3), 297-307.
- Llanes-Coromina, J., Vilà-Giménez, I., Kushch, O., & Prieto, P. (under revision). Beat gestures help preschoolers recall and comprehend discourse information. *Journal of Experimental Child Psychology*.
- Loehr, D. P. (2012). Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory Phonology*, 3, 71–89.
- López-Ornat, S., Gallego, C., Gallo, P., Karousou, A., Mariscal, S. & Martínez, M. (2005). *Inventario del desarrollo MacArthur: versión española*. Madrid. TEA.
- Mathew, M. M., Yuen, I., Shattuck-Hufnagel, S., Ren, A., & Demuth, K. (2014). The use of prosodic gesture during parent-child discourse interactions. *Australian Linguistic Society Annual Conference*. The University of Newcastle, Australia.
- Matthews, D., Behne, T., Lieven, E., & Tomasello, M. (2012). Origins of the human pointing gesture: a training study. *Developmental Science*, 15 (6), 1–14.
- McGillion, M.L., Herbert, J.S., Pine, J.M., Keren-Portnoy, T., Vihman, M.M. & Matthews, D. (2013). *IEEE Transactions on Autonomous Mental Development*, 5 (3), 240-248.
- McNeill, D. (1992). *Hand and Mind*. Chicago. The Chicago University Press.
- McNeill, D. (2005). *Gesture and Thought*. Chicago: University of Chicago Press.

- McNeill, D. (2012). *How language began: Gesture and speech in human evolution*. Cambridge University Press.
- McNeill, D. & Duncan, S. (2000). Growth points in thinking-for-speaking. In D. McNeill (Ed.), *Language and Gesture* (pp. 141-161). Cambridge, United Kingdom: Cambridge University Press.
- McNeill, D., Duncan, S.D., Cole, J., Gallagher, S. & Bertenthal, B. (2008). Growth points from the very beginning. *Interaction Studies*, 9 (1), 117-132.
- Miller, J.L. & Gros-Louis, J. (2013) Socially guided attention influences infants' communicative behavior. *Infant Behavior & Development*, 36 (4), 627-634.
- Miller, J. L., & Lossia, A. K. (2013). Prelinguistic infants' communicative system: Role of caregiver social feedback. *First Language*, 33 (5), 433–448.
- Morford, M., & Goldin-Meadow, S. (1992). Comprehension and production of gesture in combination with speech in one-word speakers. *Journal of Child Language*, 19(3), 559–580.
- Moubayed, S.A., Beskow, J. & Granström, B. (2010). Auditory visual prominence. *Journal of Multimodal User Interfaces*, 3, 299-309.
- Mundy, P. & Gomes, A. (1998). Individual differences in joint attention skill development in the second year. *Infant Behavior & Development*, 21 (3), 469-482.
- Murillo, E., & Belinchón, M. (2012). Gestural-vocal coordination. Longitudinal changes and predictive value on early lexical development. *Gesture*, 12 (1), 1, 16-39.
- Murillo, E., & Belinchon, M. (2013). Multimodal communicative patterns on the transition to first words: Changes in the

coordination of gesture and vocalization. *Infancia y Aprendizaje*, 36(4), 473-487.

Murillo, E., & Capilla, A. (2016). Properties of vocalization-and gesture-combinations in the transition to first words. *Journal of child language*, 43(4), 890-913.

Murillo, E., Galera, N., & Casla, M. (2015). Gesture and speech combinations beyond two-word stage in an experimental task. *Language, Cognition and Neuroscience*, 30(10), 1291–1305.

Nathani, S. & Oller, D.K. (2001). Beyond ba-ba and gu-gu: Challenges and strategies in coding infant vocalization. *Behavior Research Methods, Instruments, & Computers*, 33 (3), 321-330.

Nouri, M. (2010). *IBM Statistics Advanced Statistics 19*. Prentice Hall.

Oakhill, J., & Cain, K. (2012). The precursors of reading ability in young readers: Evidence from a four-year longitudinal study. *Scientific Studies of Reading*, 16, 91–121.

Oller, D. K., Niyogi, P., Gray, S., Richards, J. A., Gilkerson, J., Xu, D., Yapanel, U., & Warren, S. F. (2010). Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development. *Proceedings of the National Academy of Sciences*, 107(30), 13354-13359.

Özçalışkan, S., & Goldin-Meadow, S. (2005). Gesture is at the cutting edge of early language development. *Cognition*, 96(3), B101–113.

Paivio, A. (1990). *Mental representations: A dual coding approach*. New York, NY: Oxford University Press.

Parladé, M. V., & Iverson, J. M. (2011). The interplay between language, gesture, and affect during communicative transition: a dynamic systems approach. *Developmental psychology*, 47(3), 820.

- Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in communication* (pp. 271–311). Cambridge, MA: MIT Press.
- Pizzuto, E., Capobianco, M., & Devescovi, A. (2005). Gestural–vocal deixis and representational skills in early language development. *Interaction Studies*, 6 (2), 223–252.
- Pons, F., Andreu, L., Sanz-Torrent, M., Buil-Legaz, L., & Lewkowicz, D. J. (2013). Perception of audio-visual speech synchrony in Spanish-speaking children with and without specific language impairment. *Journal of child language*, 40(03), 687-700.
- Prieto, P., & Roseano, P. (2010). *Transcription of Intonation of the Spanish Language*. München: Lincom.
- Ramscar, M., & Gitcho, N. (2007). Developmental change and the nature of learning in childhood. *Trends in Cognitive Science*, 11, 274–279.
- Rochet-Capellan, A., Laboissière, R., Galván, A., & Schwartz, J. (2008). The Speech Focus Position Effect on Jaw-Finger Coordination in a Pointing Task. *Journal of Speech Language and Hearing Research*, 51(6), 1507-1521.
- Rollins, P. R. (2003). Caregivers' contingent comments to 9-month infants: Relationships with later language. *Applied Psycholinguistics*, 24 (2), 221–234.
- Rowe, M.L. & Goldin-Meadow, S. (2009). Early gesture selectively predicts later language development. *Developmental Science*, 12 (1), 182-187.
- Rusiewicz, H. L. (2010). *The Role of Prosodic Stress and Speech Perturbation on the Temporal Synchronization of Speech and Deictic Gestures*. Doctoral Dissertation, University of Pittsburgh.

- Rusiewicz, H. L. & Esteve-Gibert, N. (accepted). Set in time: Temporal coordination of prosodic stress and gesture in the development of spoken language production. In P. Prieto & N. Esteve-Gibert (Eds.), *The development of prosody in first language acquisition*. John Benjamins.
- Rusiewicz, H. L., Shaiman, S., Iverson, J. M., & Szuminsky, N. (2013). Effects of prosody and position on the timing of deictic gestures. *Journal of Speech, Language, and Hearing Research*, *56*(2), 458-470.
- Rusiewicz, L. H., Shaiman, S., Iverson, J. M., & Szuminsky, N. (2014). Effects of perturbation and prosody on the coordination of speech and gesture. *Speech Communication*, *57*, 283–300.
- Schmidt, C. R., & Paris, S. G. (1983). Children's use of successive clues to generate and monitor inferences. *Child Development*, *54*, 742-759.
- Sekine, K., Sowden, H., & Kita, S. (2015). The Development of the Ability to Semantically Integrate Information in Speech and Iconic Gesture in Comprehension. *Cognitive Science*, *39*, 1855–1880.
- Sekine, K., & Kita, S. (2015). Development of multimodal discourse comprehension: cohesive use of space by gestures. *Language, Cognition and Neuroscience*, *30*(10), 1245-1258.
- Serrat, E.; Sanz-Torrent, M.; Badia, I., Aguilar, E., Olmo, R., Lara, M.F., Andreu. L. y Serra, M. (2010). La relación entre el aprendizaje léxico y el desarrollo gramatical. *Infancia y Aprendizaje*, *33*(4), 435-448.
- Shapiro, L. (2007). The embodied cognition research programme. *Philosophy Compass*, *2*, 338–346.
- Shattuck-Hufnagel, S., Ren, A., Mathew, M., Yen, I., & Demuth, K. (2016). Non-referential gestures in adult and child speech: Are

they prosodic? *Proceedings of Speech Prosody*. Boston, USA, May 31-June 3.

- Silverman, L. B., Bennetto, L., Campana, E., & Tanenhaus, M. K. (2010). Speech-and-gesture integration in high functioning autism. *Cognition*, *115*(3), 380–93.
- So, W. C., Chen-Hui, C. S., & Wei-Shan J. L. (2012). Mnemonic effect of iconic gesture and beat gesture in adults and children: Is meaning in gesture important for memory recall? *Language and Cognitive Processes*, *27*(5), 665-681.
- Speer, S. R., & Ito, K. (2009). Prosody in First Language Acquisition - Acquiring Intonation as a Tool to Organize Information in Conversation. *Language and Linguistics Compass*, *3*(1), 90–110.
- Stefanini, S., Bello, A., Caselli, M. C., Iverson, J. M., & Volterra, V. (2009). Co-speech gestures in a naming task: Developmental data. *Language and Cognitive Processes*, *24*(2), 168–189.
- Swerts, M., & Kraemer, E. (2008). Facial expression and prosodic prominence: Effects of modality and facial area. *Journal of Phonetics*, *36*(2), 219-238.
- Tamis-LeMonda, C. S., Bornstein, M. H., & Baumwell, L. (2001). Maternal responsiveness and children's achievement of language milestones. *Child Development*, *72* (3), 748–767.
- Tellier, M. (2008). The effect of gestures on second language memorization by young children. *Gesture*, *8*(2), 219-235.
- Terken, J. (1991). Fundamental frequency and perceived prominence. *Journal of the Acoustical Society of America* *89*(4), 1768–1776.
- Theakston, A. L., Coates, A., & Holler, J. (2014). Handling agents and patients: Representational co-speech gestures help children

- comprehend complex syntactic constructions. *Developmental Psychology*, 50(7), 1973-1984.
- Thompson, L., Driscoll, D., & Markson, L. (1998). Memory for visual-spoken language in children and adults. *Journal of Nonverbal Behavior*, 22, 167-187.
- Tomasello, M. (1988). The role of joint attentional processes in early language development. *Language sciences*, 10 (1), 69-88.
- Tomasello, M. (2008). *The origins of human communication*. MIT press.
- Tomasello, M., Carpenter, M. & Liszkowski, U. (2007). A new look at infant pointing. *Child Development*, 78 (3), 705-722.
- Tomasello, M. & Farrar, M.J. (1986). Joint attention and early language. *Child development* 57 (6), 1454-1463.
- Tomlinson, J., Gotzner, N., & Lewis, B. (in press). Intonation and pragmatic enrichment: how intonation constrains ad-hoc scalar inferences. *Language and Speech*.
- Treffner, P., Peter, M., & Kleidon, M. (2008). Gestures and phases: The dynamics of speech-hand communication. *Ecological Psychology*, 20(1), 32-64.
- Tribushinina, E., (2014). Comprehension of degree modifiers by pre-school children: What does it mean to be 'a bit cold'? *Folia Linguistica*, 48(1), 255-276.
- Van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*, 45(3), 598-607.
- Vigliocco, G., Perniss, P., & Vinson, D. (2014). Language as a multimodal phenomenon : implications for language learning, processing and evolution. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 369, 1-7.

- Vilà-Giménez, I, Igualada, A., & Prieto, P. (under revision). Training with rhythmic beat gestures improves children's narrative discourse skills. *Developmental Psychology*.
- Wagner, P., Malisz, Z., & Kopp, S. (2014). Gesture and speech in interaction: An overview. *Speech Communication*, 57, 209–232.
- Wagner, P., Origlia, A., Avesani, C., Christodoulides, G., Cutugno, F., D'Imperio, M., Escudero Mancebo, D., Gili Fivela, B., Lacheret, A., Ludusan, B., Moniz, H., Ní Chasaide, A., Niebuhr, O., Rousier-Vercruyssen, L., Simon, A.C., Šimko, J., Tesser, F. & Vainio, M. (2015). Different parts of the same elephant: a roadmap to disentangle and connect different perspectives on prosodic prominence. *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, Scotland, August 10-14.
- Wang, L., & Chu, M. (2013). The role of beat gesture in pitch accent in semantic processing: An ERP study. *Neuropsychologia*, 51, 2847-2855.
- Wetherby, A.M. & Prizant, B.M. (1989). The expression of communicative intent: assessment guidelines. *Seminars in Speech and Language*, 10 (1), 77-91.
- Wu, Z. & Gros-Louis, J. (2014). Infants' prelinguistic communicative acts and maternal responses: Relations to linguistic development. *First Language*, 34 (1), 72-90.
- Yap, D.F., & Casasanto, D. (2017). The Semantics of Beat Gestures. *ISGS7 conference*, Paris, July, 18-22.
- Yasinnik, Y., Renwick, M., & Shattuck-Hufnagel, S. (2004). The timing of speech-accompanying gestures with respect to prosody. *Proceedings of From Sound to Sense* (pp. 97–102). Cambridge, MA: Massachusetts Institute of Technology.

Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition*, *125*(2), 244–62.

APPENDIX 1. EXPERIMENT 1 (CHAPTER 2)

Scripts for narration of the two parts in the word recall task.

1 Presentation of the characters and plot of the story

1.1 Presentation of the main character and his friends by the experimenter.

1.1.1 Slide with a drawing of the main character hidden under a colored rectangle to draw children's attention to the story.

(Catalan) *Saps qui és aquest?* (English) *Do you know who this is?*

1.1.2 Slide appears with a picture of the main character, Elmer the elephant.

(Catalan) *És un elefant que es diu Elmer! L'Elmer és un elefant de colors i li agrada molt viatjar. Viatjar és el que més li agrada!* (English) *This is an elephant called Elmer! Elmer is a colorful elephant and he enjoys traveling very much. Traveling is what he likes most!*

1.1.3 Slide appears with a picture of other elephants and among them a photo of the actor, who is about to appear in the video.

Presentation of the main character's friends.

(Catalan) *Mira! Aquests són els amics de l'Elmer. Els seus amics són... (pausa) elefants! I aquesta noia? Qui és?*

(pausa). *També és la seva amiga, i es diu Núria.* (English)
Look! These are Elmer's friends. His friends are... (pause)
elephants! And this girl? Who is she? (pause). She's his
friend too, and her name is Núria.

1.2 **Presentation of the plot of the story.**

Same as previous slide. (Catalan) *L'Elmer i els seus amics*
volen marxar de viatge, però saps què passa? (pausa).
L'Elmer sempre s'oblida de les coses. Té molta mala
memòria, i sempre s'oblida de les coses que ha de fer. Sort
que hi ha la Núria, que és molt maca i l'ajuda a recordar.
Escolta. Que l'ajudem nosaltres també? (pausa). Va, anem
a ajudar-lo a recordar el que ha de fer l'Elmer abans de
marxar de viatge. Para molta atenció al què diu. (English)
Elmer and his friends want to go on a trip but do you know
what happens? (pause). Elmer always forgets things. He
has a very bad memory and he always forgets the things
that he has to do. Luckily Núria is really nice and helps
him to remember. Listen. Shall we help him too? (pause).
Let's help Elmer to remember what he needs to do before
going on the trip. Pay close attention to what Núria is
saying.

2 **Repeating sequence of the plot consisting of the following three phases:**

2.1 **Word list exposure phase.** Slide with a video recording of the actor. In the first part of the introductory sentence, the actor gazes straight at the camera and says “Hi!” to engage the child’s attention. Then, when a drawing of Elmer the elephant appears in the bottom left part of the screen, she shifts her gaze so that she appears to be looking at Elmer. She addresses Elmer by name and keeps her gaze directed at him for the remainder of this phase as she presents a particular context and reminds Elmer of the list of things he must do here.

Market context. (Catalan) *Hola! Elmer! Abans de marxar, has de anar al mercat i comprar pomes, iogurt, cebes, aigua, raïm.* (English) *Hi, Elmer! Before leaving, you have to go to the market and buy apples, yogurt, onions, water, grapes.*

Zoo context. (Catalan) *Hola! Elmer! Abans de marxar, has de anar al zoo has d'acomiar-te dels óssos, lloros, ànecs, cavalls, pardals.* (English) *Hi, Elmer! Before leaving, you have to go to the zoo and say goodbye to the bears, parrots, ducks, horses, birds.*

Room context. (Catalan) *Hola! Elmer! Abans de marxar, a l'habitació has d'endreçar llibres, nines, cotxes, papers, globus.* (English) *Hi, Elmer! Before leaving, in your room*

you have to tidy up your books, dolls, cars, papers, balloons.

Drawing context. (Catalan) *Hola! Elmer! Abans de marxar, has de fer un dibuix de l'escola i que tingui portes, arbres, classes, taules, coixins.* (English) *Hi, Elmer! Before leaving, you have to make a drawing of the school with doors, trees, classrooms, desks, pillows.*

2.2 Word list recall phase prompted by the experimenter.

Experimenter: (Catalan) *Què ha de fer l'Elmer?* (English) *What does Elmer have to do?*

2.3 Story resumption or concluding scene.

2.3.1 Distractor phase and motivation to link to next sequence.

Slide with a picture of the elephant in different scenes.

2.3.1.1 Interaction between child and experimenter. (Catalan)

Mira! Ha anat a... Allà farà... (English) *Look! He has gone to... There he will...*

2.3.1.2 Link to the next sequence of the plot

(Catalan) *Però la Núria li diu: "Elmer, encara has de fer més coses". L'Elmer s'ha oblidat de altres coses que ha de fer. Anem a ajudar-lo a recordar un altre cop. A veure què diu ara, la Núria. Para molta atenció.* (English) *But Núria is saying: "Elmer, you still have things to do". Elmer has forgotten about other things that he has to do. Let's help*

him again to remember. Let's see what Núria is saying now. Pay close attention.

2.3.2 End of the story. Slide with a picture of the elephant playing with his elephant friends.

(Catalan) Finalment, aquesta és l'última cosa que l'elefant havia de fer! Mira que content que està després de treballar tant. L'Elmer ara ja pot anar de viatge amb els seus amics. (English) Finally, that was the last thing that the elephant had to do! Look how happy he is after all this hard work. Elmer can now go on a trip with his friends.

APPENDIX 2. EXPERIMENT 2 (CHAPTER 3)

Six sentence stories followed by the inference tapping question, lexical items and pictures of the concepts (i.e., target, competitor, literal and distractor).

Familiarization story

Story 0.

En Pau esperava a la vora de la PORTA.

Va veure al seu pare que corria





i va pensar què més podia fer.

El pare intentava agafar un ANIMAL.

Volia agafar-lo amb una mica de FORMATGE.

Ell sempre té bones idees!

Inference tapping question: El pare d'en Pau, quin animal volia agafar?

Target: mouse	Competitor: duck	Literal: door	Distractor: scissors
			

Experimental stories

Story 1.

L'Enric sempre aprofitava per sortir de casa els dies sense NÚVOLS.

Aquell dia la seva mare també l'havia acompanyat





ja que feia molts dies que no sortien junts.

L'Enric es va quedar una estona parat mirant uns INSECTES.

Volia saber si estaven fent MEL.

Va passar un dia bonic amb la seva mare!

Inference tapping question: L'Enric, què va veure?

Target: bee	Competitor: butterfly	Literal: cloud	Distractor: shoe
			

Story 2.

La Núria va seure amb un amic al BANC

El seu pare l'havia ajudat a preparar la bossa



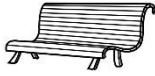

però no havia tingut massa temps de veure què li havia posat dins.

Es va trobar que portava una FRUITA

Va veure que era VERDA.

Van parlar molt mentre menjaven i s'ho van passar molt bé!

Inference tapping question: La Núria, què va portar de fruita?

Target: pear	Competitor: cherry	Literal: bench	Distractor: drum
			

Story 3.

En Carles va mirar a fora per la FINESTRA.

Es va quedar una estona parat mentre li tocava l'aire a la cara.



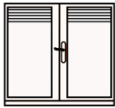

No tenia gaire deures per fer.

Mirava atentament una cosa al CEL.

Va veure que tenia CRÀTERS

Es va quedar embadalit!

Inference tapping question: En Carles, què va veure al cel?

Target: moon	Competitor: star	Literal: window	Distractor: horse
			

Story 4.

L'Anna va deixar a l'entrada el seu ABRIC.

Va entrar ràpid per a anar directament cap al menjador





ja que la mare li havia dit que havia d'arribar aviat.

De sobte es va adonar que a casa hi havia una MASCOTA.

Va veure que portava un OS.

Es va sentir molt afortunada!

Inference tapping question: L'Anna, quina mascota va veure?

Target: dog	Competitor: fish	Literal: coat	Distractor: nose
			

Story 5.

La Judith estava esperant que vinguéss la seva amiga a CASA.

Va arreglar l'habitació amb molta cura.

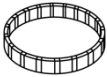



Sempre s'ho passaven bé juntes.

A la Judith li agradava jugar a disfressar-se amb JOIES

Li encantava posar-se'n moltes als BRAÇOS

Al pare li va fer gràcia trobar-les disfressades!

Inference tapping question: A la Judith, quines joies li encanta posar-se?

Target: bracelet	Competitor: earrings	Literal: house	Distractor: barrel
			

Story 6.

En Toni havia anat a l'hort de l'avi a recollir VERDURES

L'avi ja era gran i li havia demanat que l'ajudés de tant en tant.


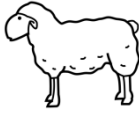


A ell li agradava anar-hi perquè s'ho passava bé

En Toni de sobte va veure un ANIMAL.

Va dir que havia posat un OU

L'avi els va arreplegar tots abans d'acabar el dia!

Inference tapping question: En Toni, quin animal va veure?

Target: chicken	Competitor: sheep	Literal: vegetables	Distractor: square
			

Story 7.

Els amics d'en Manel van venir a jugar a CASA.

Tots plegats van decidir jugar a un joc secret





i necessitaven estar dins de casa.

En Manel va agafar una LLUM

Es va quedar embadalit mirant la CERA

Al final tots van riure molt!

Inference tapping question: En Manel, quina llum va necessitar?

Target: candle	Competitor: flashlight	Literal: house	Distractor: orange
			

Story 8.

En Miquel va anar amb la seva mare a la MUNTANYA.

La seva mare s'ho mirava tot embadalida





Ell estava molt content aquell dia.

En Miquel es va fixar en una cosa que baixava del CEL.

Mai havia vist de tan a prop unes PLOMES.

Li agradava tant passejar amb la seva mare!

Inference tapping question: En Miquel, què va veure al cel?

Target: bird	Competitor: cloud	Literal: mountain	Distractor: hat
			

Story 9.

Aquell dia la Maria es va deixar la MOTXILLA.

Va pensar que podia tornar a buscar-la a l'escola.





I va arribar a casa després d'una estona.

Havia d'acabar un treball sobre un tipus de PLANTES

Li va encantar parlar sobre les formes de les BRANQUES

Al final del dia es va quedar ben tranquil·la perquè havia acabat el treball!

Inference tapping question: La Maria, sobre quin tipus de vegetació va fer la redacció?

Target: tree	Competitor: flower	Literal: rucksack	Distractor: eye
			

Story 10.

Era diumenge, i abans de marxar tots de casa la Marta va agafar el seu LLIBRE

No s'esperava la sorpresa que li havien preparat els pares.

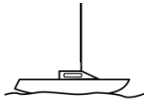



Estava contenta però a la vegada una mica nerviosa.

La Marta mai no havia estat en aquest tipus de TRANSPORT.

Es va espantar quan es va obrir la VELA.

El final del dia ho van celebrar amb una paella!

Inference tapping question: La Marta, quin tipus de transport va agafar?

Target: boat	Competitor: train	Literal: book	Distractor: cinema
			

Story 11.

La Carlota estava a la cuina menjant un ENTREPÀ

Va sentir un soroll.





Es va apropar al menjador.

Va veure que el pare havia comprat un APARELL.

Va dir que feia molt d'AIRE.

Després va tornar a la cuina

Inference tapping question: La Carlota, quin aparell va veure?

Target: fan	Competitor: coffee machine	Literal: sandwich	Distractor: candy
			

Story 12.

L'Estefania va passar tot el dia enfeïnada per de CASA.

Va trigar una bona estona per fer-ho.

ARA, va ser molt complicat.

Va arreglar una cosa de la CUINA

Finalment va aconseguir que sortís l'AIGUA.

Es va quedar tranquil·la i molt contenta!

Inference tapping question: L'Estefania, quin aparell de la cuina va arreglar?

Target: tap	Competitor: fridge	Literal: house	Distractor: snail
