THE VALUE OF GESTURE-SPEECH INTEGRATION IN LANGUAGE DEVELOPMENT.

APPLICATIONS TO SPEECH-LANGUAGE PATHOLOGY

PhD project

Alfonso Igualada

Universitat Pompeu Fabra

Supervisors: Dr. Pilar Prieto

Dr. Laura Bosch

June 13th, 2014

INDEX

0. RESUM

És ben sabut que les modalitats comunicatives de gest i parla es troben integrades en els plans fonològic, semàntic i pragmàtic (McNeill, 1992). En relació amb el desenvolupament del llenguatge, els estudis sobre el valor predictiu de les habilitats comunicatives primerenques generalment s'han interessat per l'anàlisi per separat dels gestos comunicatius (principalment del gest d'assenyalar) o de la parla. D'altra banda, se sap molt poc sobre si la integració simultània entre gest i parla afavoreix el desenvolupament de la comunicació, i si promou el desenvolupament del llenguatge. Aquesta tesi de doctorat té com a objectiu investigar l' impacte de les produccions simultànies de gest i parla en el desenvolupament del llenguatge, tant en població amb un desenvolupament típic, com en nens amb trastorn de l'espectre autista (TEA). La tesi consta de tres estudis experimentals independents sobre la comunicació multimodal durant l'etapa infantil, en una franja que va dels dotze mesos fins als sis anys d'edat.

Els dos primers estudis pretenen proporcionar evidència científica sobre el valor comunicatiu de l'ús simultani de les modalitats de gest i parla en relació amb les habilitats lingüístiques. El primer estudi investiga l'ús infantil de les produccions simultànies de gest i parla, durant interaccions socio-comunicatives controlades, i el seu valor com a predictor del desenvolupament posterior del llenguatge. Dinou infants de 12 mesos d'edat van participar en una tasca que fomentava l'ús del gest d'assenyalar amb funció declarativa, en tres condicions d'interacció conjunta entre infant i adult. Els resultats demostren que les combinacions simultànies de gest i parla es produïen amb més freqüència en la condició en què l'adult mirava al nen i no a l'objecte de referència. A més, l'ús de combinacions simultànies de gest i parla produïdes als 12 mesos durant aquesta condició d'interacció es van correlacionar positivament amb les habilitats de vocabulari i gramaticals als 18 mesos d'edat.

El segon estudi investiga si la presència dels gestos *beat* (també anomenats gestos rítmics, o gestos batuta), utilitzats dins d'un registre de llenguatge dirigit a infants, afavoreix el record d'informació lingüística. La nostra hipòtesi prediu que la presència de gestos *beat* afavorirà el record de la paraula intermèdia en una llista de paraules inclosa dins d'un context discursiu. Seixanta infants de 3, 4 i 5 anys d'edat, i vint adults van ser avaluats amb una tasca de memòria immediata en un context discursiu. L'habilitat de record de paraules de cada participant va ser controlada mitjançant el còmput del nombre màxim d'elements que són capaços de recordar dins d'una llista

oberta de paraules. Els primers resultats extrets del 43% del total de les dades mostren un efecte significatiu de la presència del gest *beat* en la capacitat de recordar la paraula en posició intermèdia. Els resultats també revelen que aquesta habilitat es relaciona directament amb les habilitats de record de paraules més que amb l'edat.

El tercer estudi pretén investigar l'efecte de la presència de gestos *beat* en la comprensió del referent amb focus contrastiu en població amb desenvolupament típic i amb TEA. Investigacions prèvies han demostrat que els nens amb TEA presenten dificultats en la seva capacitat d'integrar els actes comunicatius multimodals (Silverman, Bennetto, Campana, i Tanenhaus, 2010; Grossman, i Tager-Flusber, 2012), i que hi ha una falta de sincronia temporal en les combinacions simultànies entre gest i parla, la qual cosa pot afectar la seva habilitat comunicativa (De Marchena i Eigsti, 2010). Per a aquest estudi compararem deu nens amb autisme i el seu grup control de desenvolupament típic similars en edat, gènere i habilitat verbal mitjançant un paradigma visual amb el procediment de l'*eye tracking visual world paradigm*. També es controlaran les habilitats de teoria de la ment i d'empatia. En aquest sentit, la predicció és que la presència de gestos *beat* tindrà un impacte positiu en la comprensió del llenguatge en ambdues poblacions. També esperem obtenir millors resultats de comprensió en aquells nens, de tots dos grups, amb millors habilitats d'empatia i socio-comunicatives.

En definitiva, aquesta tesi inclou tres estudis independents que pretenen mostrar evidència empírica que a favor de la idea que les habilitats d'integració de gest i parla representen una estratègia comunicativa efectiva durant la infància, la qual promou positivament les habilitats primerenques del llenguatge.

RESUMEN

Las modalidades comunicativas de gesto y habla se encuentran integradas en los planos fonológico, semántico y pragmático (McNeill, 1992, ao). Estudios recientes sobre el valor predictivo de las habilidades comunicativas tempranas generalmente han analizado aisladamente los gestos comunicativos (principalmente el gesto de señalamiento) o el habla. Por otra parte, se sabe muy poco sobre si la integración simultánea de gesto y habla favorece el desarrollo de la comunicación, y si promueve el desarrollo posterior del lenguaje. Esta tesis de doctorado tiene como objetivo investigar el impacto de las producciones simultáneas de gesto y habla sobre el desarrollo del lenguaje, tanto en población con desarrollo típico del lenguaje como en niños con trastorno del espectro autista (TEA). La tesis consta de tres estudios experimentales independientes sobre la comunicación multimodal durante la etapa infantil, concretamente de la franja de doce meses hasta los seis años de edad.

Los dos primeros estudios pretenden proporcionar evidencia científica sobre el valor del uso simultáneo de las modalidades comunicativas de gesto y habla en relación con las habilidades del lenguaje. El primer estudio investiga el valor predictivo del uso de las producciones simultáneas de gesto y habla, durante interacciones socio-comunicativas controladas. Diecinueve niños de 12 meses de edad participaron en una tarea que fomentaba el uso del gesto de señalamiento con función declarativa, en tres condiciones de interacción conjunta entre el niño y el adulto. Los resultados demuestran que las combinaciones simultáneas de gesto y habla, y no las producciones aisladas de habla o señalamiento, se produjeron en mayor frecuencia en la condición en la que el adulto miraba al niño y no al objeto de referencia. Además, el uso de combinaciones simultáneas de gesto y habla producidas a los 12 meses durante esta condición de interacción correlacionaron positivamente con las habilidades de vocabulario y gramaticales a los 18 meses de edad.

El segundo estudio investiga si la presencia de los gestos *beat* (también llamados gestos rítmicos o gestos batuta), utilizados dentro un registro de lenguaje dirigido a niños, favorece el recuerdo de información lingüística. Nuestra hipótesis predice que la presencia de gestos *beat* favorecerá el recuerdo de la palabra intermedia de una lista de palabras incluida en un contexto discursivo relevante. Sesenta niños de 3, 4 y 5 años de edad, y veinte adultos fueron evaluados con una tarea de memoria inmediata que formaba parte de un contexto discursivo. La habilidad de recuerdo de palabras de

cada participante fue controlada mediante el cómputo del número máximo de elementos que son capaces de recordar dentro de una lista abierta de palabras.
Los primeros resultados extraídos del 43% del total de los datos muestran un efecto significativo de la presencia del gesto *beat* en la capacidad de recordar la palabra en posición intermedia de la lista. Los resultados también revelan que esta habilidad se relaciona con las habilidades de recuerdo de palabras más que con la edad.

El tercer estudio pretende investigar el efecto de la presencia de gestos *beat* en la comprensión del referente marcado con foco contrastivo, tanto en población con desarrollo típico del lenguaje como en población con TEA. Investigaciones previas han demostrado que los niños con TEA presentan dificultades en su capacidad para integrar los actos comunicativos multimodales (Silverman, Bennetto, Campana, y Tanenhaus, 2010; Grossman, y Tager-Flusber, 2012), y que existe una falta de sincronía temporal en las combinaciones simultáneas de gesto y habla, lo cual puede afectar a su habilidad comunicativa (De Marchena y Eigsti, 2010). Para este estudio compararemos diez niños con autismo y su grupo control de desarrollo típico similares en edad, género y habilidad verbal mediante el procedimiento *eye tracking visual world paradigm*. También se controlarán las habilidades de teoría de la mente y de empatía de los niños. Nuestra predicción es que la presencia de gestos *beat* tendrá un impacto positivo en la comprensión del lenguaje en ambas poblaciones. También esperamos obtener mejores resultados de comprensión en aquellos niños de ambos grupos con mejores habilidades de empatía y socio-comunicativas.

En definitiva, esta tesis incluye tres estudios independientes que pretenden mostrar evidencia empírica a favor de la idea que las habilidades de integración de gesto y habla representan una estrategia comunicativa efectiva durante la infancia, la cual promueve positivamente las habilidades tempranas del lenguaje.

ABSTRACT

In human communication, gestures and speech are integrated at the phonological, semantic and pragmatic levels (McNeill, 1992, a.o.). Literature on the predictive value of early communicative abilities in language development has typically focused on separate analyses of either gestures (and mainly pointing gestures) or speech vocalizations. Moreover, little is known about whether the simultaneous integration of speech and gestures aids communication in childhood, and whether it also affects and promotes language development. This PhD thesis will investigate the value of gestures integrated together with speech as an important part of the development of language, both in the typical developing population and in children with autism spectrum disorder (ASD). The thesis comprises three independent experimental studies dealing with children in their early and late childhood (twelve months to six years).

The first and second studies were aimed at providing behavioral evidence of the impact of the integrated use of gesture and speech on language abilities. The first study investigated the predictive value of the infants' use of synchronous gesture-speech combinations during controlled socio-communicative interactions for later language development. Nineteen 12-month-old infants participated in a declarative pointing task involving three conditions with different joint engagement patterns. The results revealed that only simultaneous pointing-speech combinations, and not isolated productions of speech or pointing, increased in frequency when the adult looked at the infant but not at the event of reference. Moreover, synchronous pointing-speech combinations in this condition at 12 months positively correlated with expressive vocabulary and grammatical measures obtained from parental reports at 18 months of age.

The second study investigated whether the presence of beat gestures embedded in a child-directed discourse helps children to recall linguistic information. We predicted that the presence of beat gestures would be beneficial in the recall of the target item in a list of words. We tested sixty 3-, 4-, and 5-year-olds and twenty adults with a memory recall activity embedded in a discourse context. Memory span ability was tested to measure the maximum number of items that the child could remember in an open list of words. Pilot results from 43% of the data showed a significant effect of the presence of the beat gesture on the ability to remember the target item. The results also showed that this positive effect is more closely correlated with children's memory span abilities than with their age.

The third study was intended to explore the impact of the presence of beat gestures in the contrastive referent resolution in both ASD and typically developing populations. Previous work has shown that the ASD population present difficulties in their abilities to integrate communicative multimodal cues (Silverman, Bennetto, Campana, & Tanenhaus, 2010; Grossman, & Tager-Flusber, 2012) and that gesture-speech lack of synchrony in ASD individuals may impact their communicative abilities (De Marchena & Eigsti, 2010). We compared ten children with autism and a control group of typically developing children matched for age, gender, and verbal abilities through a visual world eye tracking paradigm. Children were also controlled for theory of mind and empathy abilities. We predicted that the presence of beat gestures would have a positive impact on language understanding in both populations. We also expected better outcomes in children with higher empathy and socio-communicative abilities, in both groups.

All in all, this PhD thesis includes three independent studies which represent an attempt to find empirical evidence in support of the fact that gesture-speech integration abilities represent an effective communicative strategy in childhood which successfully promotes early language abilities.

1. INTRODUCTION

1.1. Object of analysis

This research focuses on the study of multimodal synchronous integration between speech and gesture abilities during the childhood period with three independent studies. This study on the value of gesture-speech integration in development deals with three main research questions: (1) Is the early production of synchronous multimodal communicative strategies related to language abilities? (Study 1); (2) Is the presence of beat gestures beneficial for the recall of linguistic information in children? (Study 2); (3) Is the presence of beat gestures beneficial for the understanding of intonation contrast in a referent resolution task by typically developing children? (Experiment 3.1); and is the presence of beat gestures beneficial for the understanding of intonation contrast in a referent resolution task by ASD children? (Experiment 3.2).

1.2. Prior work

1.2.1. Multimodal gesture-speech modalities and social interaction.

The integration between gesture and speech is an important feature of human communication. As McNeill's (1992) groundbreaking study 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 aligned with the most prominent part of speech), but also at the semantic and pragmatic levels. Yet the literature on the development of these communicative modalities has typically focused separately on the study of speech and gestures (see Rowe, & Goldin-Meadow, 2009; Colonnesi, Stams, Koster & Noom, 2010; Behne, Liszkowski, Carpenter & Tomasello, 2012, as examples of infant gesture studies; and Masataka, 1993; Oller, Buder, Ramsdell, Warlaumont, Chorna & Bakeman, 2013, Vihman, DePaolis & Keren-Portnoy, 2008, as examples of infant speech studies).

The classification of the types of gestures originally proposed by McNeill (1992), and followed by Gullberg, DeBot & Volterra (2008) and Cartmill, Demir & Goldin-Meadow (2012), is one the most commonly used in gesture studies. McNeill classifies communicative gestures into the following four main categories, according to their use and form:

a) Deictic gestures, such as pointing gestures, 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).
b) Representational gestures (including iconic and metaphoric gestures) reference objects, actions, or relations by creating an aspect of their referent's shape or movement. Iconic gestures express more tangible information and metaphoric gestures are related to more abstract concepts.
c) Conventional gestures are culturally shared symbols, with an arbitrary form and meaning within a given community.
d) Beat gestures are movements (typically of the hand or head) that correspond to and highlight prosodic prominence. Beats have been typically described as not having a clear semantic meaning, and they have been usually studied in relation to their function within a narrative or discourse structure.

Though McNeill (1992) and Cartmill, Demir & Goldin-Meadow (2012) state that gestures can be classified into one of the latter four dimensions, many gestures involve more than one dimension. That is, communicative gestures may direct attention to a referent (deictic gestures), may be employed for some representational purpose (representational gestures), may highlight prosodic prominence in speech (beat gestures) and/or may express different degrees of social arbitrariness (conventional gestures). For example, a pointing gesture may be temporally aligned with prosodic patterns and act as a beat gesture too.

> Important questions within this field are the dependency between the gesture modality on the speech modality, to what degree are gestures and speech co-expressing semantic information, how this gesture is conventionalized by a group of people, etc. (McNeill, 2012).

The development of intentionality in communication has been strongly related with the use of communicative gestures, and specifically pointing gestures. Tomasello, Carpenter & Liszkowski (2007), following previous work by Austin (1962), Searle (1969) and Bates, Camaioni & Volterra (1975), proposed the following three basic communicative intentions as expressed by pointing gestures:

a) In the *request intention* the communicator engages with the recipient to make him/her do what he or she wants or needs.
b) In the *expressive declarative intention* the communicator engages with the recipient to express and share emotions.
c) In the *informative declarative intention* the communicator engages with the recipient to inform about something. This intention has been interpreted as an ability to know that something is relevant to the interlocutor (Tomasello, 2008). For example, an infant may employ a pointing gesture together with speech to inform about an object appearing behind the recipient's back, which was jointly seen by both communicators in advance (Liszkowski, Albrecht, Carpenter & Tomasello, 2008).

Despite the fact that the study of the development of early communicative intentions has concentrated on the development of pointing gestures, it is well known that the construction of intentional communication is multimodal. For example, the development of expressive intentionality is initiated during the rich emotional exchanges that start in early infancy. This affective attunement between infant and adult is multimodal, so that both communicators produce signals from the very first months through facial, body, vocal or hand/arm intentional actions (Tomasello, 2008).

The early uses of gesture and speech are built on the basis of social interactions between the child and his or her caregiver (e.g., Soltis, 2004; Tomasello, Carpenter & Liszkowski, 2007). Some studies have linked the infant's ability to follow an adult's gaze, which typically emerges around 6 months of age, to the infant's ability to direct others' attention by means of a pointing gesture (Matthews, Behne, Lieven, & Tomasello, 2012), and infants who attain gaze following abilities together with pointing abilities have better later language abilities (Brooks & Meltzoff, 2008). Liszkowski, Albrecht, Carpenter & Tomasello (2008) showed that communicative behaviors of 12- and 18-month-olds were affected by adults' joint attention patterns. The authors measured the children's communicative responses (speech-only and pointing-only behaviors) produced in two experimental social interaction conditions involving differences in the 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 controlling the adult's attention patterns (Carpenter, Nagell & Tomasello, 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 available and unavailable conditions, the adult looked at the infant but not at the object of reference, or was not visually attentive to either of them, respectively. The results of the study revealed that infants pointed significantly more, and produced more vocalizations during and after pointing, in the available and unavailable conditions than in the baseline condition. Moreover, adult social interaction patterns during the unavailable condition triggered less pointing behavior than during the available condition. Therefore, infant communicative responses change depending on the adult's attention behaviors.

Some studies have provided evidence that infants' multimodal combinations of pointing gestures and vocalizations are more likely to trigger contingent responses from caregivers (e.g., Gros-Louis & Wu, 2012). This means that caregivers are more sensitive to their infant's communicative multimodal communicative initiatives. In return, caregivers' contingent interactions tend to elicit more pointing and speech combinations by infants (e.g., Miller & Lossia, 2013). Gros-Louis & Wu's (2012) analysis of 10- to 13-month-old infants' interactions with their mothers showed that children were more likely to combine vocalizations with pointing when mothers were not looking at the target event. In fact, a great bulk of evidence indicates that infants' joint attention abilities are linked to later language development (e.g., Tomasello & Farrar, 1986; Mundy, 1998; Laakso, Poikkeus, Katajamäki, & Lyytinen, 1999; Kristen, Sodian, Thoermer, & Perst, 2011). Finally, some studies have correlated infants' early multimodal production abilities in spontaneous interactions with better outcomes in later language development (Murillo & Belinchón, 2012;Wu & Gros-Louis, 2014; see section 1.2.3 for a review of the literature on this topic).

Despite the fact that the latter studies have shown the developmental value of the use of simultaneous gesture-speech combinations, little is known about the use of this strategy in specific socio-communicative contexts and also its potential predictive value regarding language development. Nonetheless, some of these studies already 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.

1.2.2. Synchronous gestures-speech integration in development

Gesture and speech modalities has been studied with respect to whether they convey an equivalent meaning, or if instead there is a mismatch in meaning expressed by the two modalities (Goldin-Meadow, 2003). In this section we will focus on the temporal relations between gestures and speech. Gesture synchronicity with speech is a key point in the arguments in favor of the study of communicative modalities as an integrated system (see section 3.1. for a review of the theoretical framework). The temporal realization of intonation and pointing gestures was investigated in Esteve-Gibert & Prieto, (2013a) with fifteen adults using a pointing-naming task. They tested whether prominent parts of the gesture (i.e., the apex) and intonation pattern (i.e., the fundamental frequency peak) were synchronized. Temporal alignment differed if the word was situated in a word-final position or in a non-word-final position. Importantly, in both conditions the peak of the contrastive focus pitch accent was the most stable anchoring point for the onset of the stroke of the pointing gesture. This precise temporal alignment constitutes evidence of the linkage between manual and vocal articulators. Leonard & Cummins (2011) found asymmetries in the sensitivity of listeners to the relative timing of beats and speech. In their perception study, 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. In general, it seems that adults' intonation and gestures may be synchronized with different anchoring points, though the apex of the gesture and the intonation peak seem to be among the most stable anchoring points (Leonard & Cummins, 2011; Esteve-Gibert & Prieto, 2013a).

Young infants can use hand and mouth synchronously. Around 6 months of age infants tend to coordinate hand/arm actions and vocalizations, and 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, probably due to the fact that speech motor abilities develop more slowly and are more effortful than gestures (Iverson & Thelen, 1999; Cartmill, Demir & Goldin-Meadow, 2012). An important developmental period for gesture-speech interplay occurs when children begin to use synchronous gesture-speech combinations intentionally. These 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).

The study of word and gesture productions in relation to the meanings conveyed in each modality was the goal in Butcher & Goldin-Meadow (2000). The procedure consisted of longitudinal observations in spontaneous play situations in the interval from the one-word period to the two-word period. They found that at the beginning of one-word period gestures were not combined with speech, but at the beginning of the two-word period, children produced synchronized gesture-speech combinations referring to only one meaning. That is, children produced gestures that conveyed the same information as the accompanying word, e.g., *cookie+point at cookie*. Later studies showed that the use of supplementary gesture+word combinations, e.g., saying "want" and pointing to a cookie, is correlated with the ability to produce verbal two-word combinations (Iverson & Goldin-Meadow, 2005). However, so far the development and benefits of synchronous gesture-speech combinations has been scarcely explored in literature.

The temporal coordination between gesture-speech combinations in infancy was assessed in Esteve-Gibert & Prieto (2014). The authors analyzed communicative productions from a longitudinal sample of 4 infants recorded in naturalistic interactions with their caregivers. The results showed that 11-month-old infants already produced temporally synchronous gesture-speech combinations, but isolated pointing was still more frequently produced at this age. Moreover, the sample of gesture-speech productions increased significantly at 15 months of age. These multimodal productions were mostly pointing and reaching gestures with a declarative communicative purpose. Synchronous gesture-speech productions were precisely temporally aligned at this early age, and the prominent parts of the gesture were produced before the prominent parts of the speech. Moreover, temporal alignment variation between the onset of the speech and the onset of the gestures decreased in time from 11 to 19 months. Interestingly, these results about precise temporal alignments between both modalities in early infancy contrast with those seen for the autistic population (Marchena & Eigsti, 2010; see section 1.2.4). To our knowledge, this early ability to temporally synchronize pointing gestures with prosody might be one of the first instances of beat dimension co-occurring with a referential dimension (deictic gesture).

Yet little is known about the relation between the development described above in relation to the acquisition of beat gestures in childhood. McNeill (1992) mentioned that

children begin to produce beat gestures when early discourse abilities appear, i.e., around the second year of life. Cartmill et al. (2011) pointed out that children use beat gestures to emphasize important aspects of a narrative sometime between 4 and 5 years. Beat gestures are extremely relevant to the study of synchronous gesture-speech integration, since by definition these gestures are rhythmic markers of prosodic prominence in speech. Thus, it is important to briefly review what we know about beat gestures so far. As we have mentioned before, beat gestures have been found to co-occur with prosodic prominences, as they follow similar rhythmic and intonation temporal alignments (Loehr, 2012; McNeill, 2012). The linguistic function of beat gestures has been compared to a yellow highlighter, that is, they serve to emphasize information in the speech stream. In adult studies, beat gestures have been mainly related to the prominence function of the gesture in the discourse. 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 intonation prominences, beat gestures can be aligned with and highlight the new information of the discourse context, e.g., when enumerating the features of a newly introduced character in the story: “his GIRfriend, ALice, Alice WHIte” (example from McNeill, 2012). Thus we would expect that beat gestures may impact on language processing abilities when the prominent element is embedded in a significant context which allows the contrast (see section 1.2.3 about the benefits of gesture-speech combinations). Thus, it seems that prosody and beat gestures may both function as a yellow highlighter when embedded in an informational shared context.

The development of prosodic prominence of information in children was reviewed by Speer & Ito (2009). They reviewed that the use of intonational prominence in childhood serves two main functions, including novelty marking and contrastive focus. Intonation prominence can mark novel information with respect to previously known information (topic), and this ability may favor language development. Speer & 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 accents casts a contrastive relation between the entity with a prominence and its alternatives, making the information about alternatives more accessible. This contrastive relation helps word resolution, as the scope of the contrast gains importance. Fernald & 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 & Tomasello, (2007) investigated the effects of intonation prominence on word learning in

early childhood. The behavioral procedure that they applied consisted of interaction situations with conditions differing in the intonation prominence of the novel word at 2 years of age. Their results indicated that the learning of nouns related to novel objects were favored by the use of prosodic prominence.

And we know that children with ASD, prosody has been singled out as atypical (McCann & Peppé, 2003; Peppé, McCann, Gibbon, O'Hare & Rutherford, 2007; Kehoe, 2007). Peppé et al. (2007) found that ASD children show difficulties with intonation abilities which reflect a pragmatic function, such as contrastive pitch, but not with lexical stress, which reflects grammatical function (Grossman, Bernis, Skweker & Targer-Flusberg, 2010). On the other hand, Kehoe (2013) pointed that ASD children have difficulties with prosody expressing emotions and more pragmatic abilities, such us contrastive focus. In fact, Speer & Ito highlight the importance of the link between prosodic prominence and its pragmatic context. In this study we are interested in this shared function to focus and highlight a piece of linguistic information with beat gestures and prosodic prominence marking. Our hypothesis is that prosody abilities may be linked to similar functions in gesture modality, such as in the use of prosodic prominence marking with beat gestures. Yet much work remains to be done to fully understand the functional interplays between verbal and gesture modalities.

1.2.3 The benefits of gesture-speech integration in language and cognitive development

Gesture studies have found that gestures expressing referential information, such as deictic and representational gestures, are related to the early acquisition of language (Colonnesi, et al. 2010; Iverson & Goldin-Meadow, 2005), as well as to the acquisition of other cognitive abilities (Goldin-Meadow, Cook & Mitchell, 2009; Tellier, 2008). The use of the pointing gestures has been highlighted as an important milestone in communicative and language development. Research on gesture production and its relationship to language acquisition has shown that communicative gestures (e.g., iconic and pointing gestures) signal intentional communication and predict the emergence of verbal language (Bates, Benigni, Bretherton, Camaioni, & Volterra, 1979; Bavin, Prior, Reilly, Bretherton, Williams, Eadie, Barret, & Ukoumunne, 2008; Caselli, Rinaldi, Stefanini, & Volterra, 2012). Colonnesi, Stams, Koster, & Nooms (2010) meta-analysis of the longitudinal relation between the early use of communicative gestures and the emergence of language indicates that the early appearance of pointing gestures, and more specifically, those associated with a declarative intention, has been

systematically related to language development. This early use of pointing gestures to comment about something (declarative intention) is correlated with language development to a greater extent than the appearance of pointing used with an imperative intention.

Research on the relations between gesture and speech modalities has shown that mismatches in the meaning conveyed in each modality can predict readiness to learn (Golding-Meadow, 20013; Gullberg et al. 2008). A number of studies have related the use of supplementary gesture-speech combinations with the emergence of two-word combinations (Capirci, Iverson, Pizzuto & Volterra, 1996; Iverson & Goldin-Meadow, 2005; Özçaliskan & Goldin-Meadow, 2005). In Iverson & Goldin-Meadow (2005), supplementary combinations of speech and gesture expressing two meanings are of particular interest, as children may express two semantic meanings using two different modalities (i.e., pointing to a toy and saying "give"). The authors found that while the onset of non-supplementary (complementary) combinations, which convey only a semantic meaning (i.e., saying "no" and gesturing with the head), was not related to the later abilities to combine two words, the ability to produce supplementary combinations was related to the onset of two-word combinations. To our knowledge only two studies have explored the predictive role of early synchronous gesture-speech combinations on later language development, such as synchronized pointing to an object and speech with a declarative intention. In Murillo & Belinchón (2012), a sample of 11 parent-infant dyadic interactions were recorded in a semi-structured play context at three longitudinal moments, namely 9, 12 and 15 months. Their results showed that the use of pointing gestures at 12 months, especially when accompanied with vocalizations and the infant's gaze, correlated with vocabulary development at 15 months of age. In a recent study, Wu & Gros-Louis (2014) observed fifty-one dyadic spontaneous interactions of 10- to 13-month-old infants with their mothers and showed that infants' combinations of vocalization and pointing, especially those produced when mothers were not attentive to the target event, correlated with infants' comprehension skills at 15 months. The two studies above are based on the analysis of spontaneous interactions, which makes it difficult to behaviorally control for two important aspects of early pointing patterns, namely, (a) the pragmatic intention or motive behind children's use of pointing gestures to comment about an event/object; and (b) the social interaction gaze patterns used by the adult during the communication.

Other studies have explored the impact of representational gestures in cognitive processing abilities. In Goldin-Meadow, Cook & Mitchell (2009), 8-9 year-old children

were taught how to solve a math 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 math problems. Children performed better in the correct gesture condition, but the gesture information was beneficial to 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. Gestures also seem to help memorization in second language learning (Tellier, 2008). In this study, Tellier asked twenty 4 to 5 year-old children to remember a number of words in two conditions, namely, a verbal utterance was accompanied by a representational (i.e., iconic) gesture of the word, or else it was accompanied by a picture image. Results showed that gestures had a positive effect on word memorization 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 clear semantic information, such as beat gestures, have an effect on language processing.

The impact of gestures on language and cognitive abilities has typically concentrated on gestures expressing some kind of information about the referent or event, i.e., representational gestures. Much less is known about how beat gestures, which do not have a clearly associated meaning, impact positively on early cognitive or language processing abilities. Recent studies with an adult population have related beat gestures to semantic, syntactic and phonological linguistic abilities through neurological techniques (Holle et al. 2012; Biau & Soto-Faraco, 2013; Wang & Chu, 2013). These studies have shown the neural activation of different linguistic information when beat gestures are presented to highlight a piece of language. Yet, some differences of activation has been also shown in an ASD population that contrast with a typically developing population (see the following section 1.2.4), though very few behavioral studies have investigated how this communicative behavior affects language processing. Some studies have shown that beat gestures have an important impact on prosodic acoustic cues. For example, Krahmer & Swerts (2007) showed that adults perceived prosody cues as more prominent when they were accompanied by a beat gesture than when they were not. These results were consistent even when the participants only perceived the auditory information, and the visual information was not

presented. In fact, it seems that this effect of increasing prominence found for beat gestures could enhance auditory processing of speech (Hubbard, Wilson, Callan & Dapretto, 2009), which could be an important cue to increase the audiovisual signal during language processing.

Very few studies have investigated how language processing may be affected by the use of prosody and beat gestures functioning as cues to prominence. A series of studies by Ravizza (2003) investigated the effect of hand movements, such as tapping fingers, in a task of word retrieval. In the first study, the participants underwent a test to select words in a tip-of-the-tongue situation, that is, the participants were sure that they knew the word but they couldn't retrieve it. While performing this task, they were asked either to keep completely still or to tap the index fingers of both hands at their own pace. The results revealed that participants who could produce hand actions, which do not convey semantic information, had a positive effect on word retrieval. However, a third study in Ravizza (2003) obtained results that seemed to contradict the results of the first study. Moving hands did not help word retrieval in a task in which participants had to name as many words that started with a specified letter (i.e., a letter fluency task), so that the lexical retrieval task was influenced by switches in the phonemic information. The authors explained the conflicting results as a consequence of the abilities elicited in the two tasks, in that the task in the third study may have been related to executive function abilities. The authors concluded that these results may show links between the use "beats" and lexical processing, which may interfere phonemic processing in the second task.

To our knowledge, only one article has explored the effects of beat gestures on recall abilities in childhood. Chen-Hui & Wei-Shan (2012) tested thirty-six children aged 4-5 and thirty adults to investigate the effects of both iconic and beat gestures on word recall abilities. The authors used a list of verbs to test three conditions, namely (a) iconic gesture, (b) beat gesture and (c) no gesture. Children were presented with a closed list of 5 words and adults with a list of closed list of 10 words, which they had to remember without a given context. The results found that while adults benefited from the presence of both iconic and beat gestures to recall words, children only benefited from the presence of iconic gestures, but not from the presence of beat gestures. In this study, however, every word on the list was presented with a beat gesture, and thus the children could not perceive beat gestures as a prominent cue that contrasted the target element another neighboring element. Moreover, the list of words was presented

without a pragmatic context, which has been claimed to be an important factor in the testing of prosody in development (Ito, Jincho, Minai, Yamane & Mazuka, 2012).

1.2.4 Gesture-speech integration in the Autism Spectrum Disorder (ASD)

Autism spectrum disorder (ASD) is a neurological developmental disorder characterized by atypical development and impairments with persistent deficits in (a) communication and social interactions, and (b) repetitive behaviors and restricted behaviors and interests (McPartland, Reichow & Volkmar, 2012; Greaves-Lord, Eussen, Verhulst, Minderaa, Mandy, Hudziak, Steenhuis, de Nijs & Hartman, 2012). The American Psychiatric Association's new DSM-5 diagnostic criteria for ASD has included for the first time the difficulty in the processing of sensory information as a diagnose signal. In fact, children with ASD present impairment in the integration of multimodal cues, and especially when undertaking linguistic tasks. Bebko, Weiss, Demark & Gomez (2006) found that 4-6 year-old children with ASD were able to discriminate temporal asynchronies in non-linguistic audiovisual scenes, but not in simple linguistic events (i.e., counting) or with linguistically complex information (i.e., telling a story). And Stevenson, Siemann, Schneider, Eberly, Woynaroski, Camarata & Wallace (2014) showed that high-functioning individuals with ASD (from 6 to 18 years of age) had more difficulty integrating speech stimuli (i.e., audiovisual syllables) than non-speech information (i.e., an object synchronized with a sound). Other studies have demonstrated multimodal integration deficits with different types of stimuli used in speech, such as deficits in the integration of audiovisual material and lip-reading in adolescence (Smith & Bennetto, 2007), and deficits in the integration of prosodic information with facial expressions (Grossman & Tager-Flusberg, 2012).

Silverman, Bennetto, Campana & Tanenhaus (2010) investigated the ability of ASD adolescents to integrate iconic gestures produced simultaneously with speech as compared to a matching typically developing (TP) population. These two groups were tested with an eye-tracking paradigm procedure in a comprehension task. The participants were exposed to naturally elicited speech-gesture descriptions of an image that could be easily gestured (i.e., a key) and a shape that was difficult to elicit by gesture (i.e., a curved line with a loop). Target descriptions were presented in two conditions, according to the modality of exposure, (a) a speech-only condition and (b) an iconic gesture plus speech condition. The results showed that TP participants benefited from the use of gestures, as they identified the image corresponding to the verbal utterance when it was accompanied by a gesture. By contrast, ASD participants

were slower to match the spoken word with the picture when the gesture was present. More importantly for our purpose, Hubbard, McNealy, Scott-Van Zeeland, Callan, Bookheimer & Dapretto (2012) investigated neural activity in ASD subjects during speech-synchronized gestures which served to provide prominence to a piece of information, namely, beat gestures. The authors found that the severity of their socio-communicative impairment correlated with different patterns of neural activation. These results suggest that ASD individuals show different processing mechanisms for gesture-speech integration, which are related to impairments in the integration of multisensory temporal information, which in turn may impact negatively on their language and communication abilities.

The production of gesture-speech combinations may also be impaired in populations with ASD. Iverson & Wozniak (2007) observed longitudinal changes in the multimodal productions of the younger siblings of children diagnosed with ASD (ages 5 to 18 months), a group that is considered to be at high risk (HR) of developing autism, in comparison to a control low risk (LR) population. The patterns of rhythmic activity during the babbling period were more attenuated in the HR population. Later in development, both HR and LR groups of children showed similar proportions of gesture-speech combinations. Around 13 months of age, HR children showed a more restricted repertoire of gesture-speech combinations. And specifically, three of the infants who were later diagnosed with ASD displayed atypical patterns of vocal and motor development, and exhibited extreme delays in production of gesture-speech multimodal production. Following this line of research, De Marchena & Eigsti (2010) assessed the synchronized gesture-speech combinations, mostly focusing on iconic gestures, of ASD and TD adolescents. The frequency of use was similar in both populations, but ASD gestures were less synchronized with speech than in the control population. ASD adolescents showed a greater delay between prominent parts of both modalities, and this had a negative effect on comprehensibility ratings and interest of the story. Nonetheless, to date few studies have focused on the study of temporal gesture-speech integration in an ASD population and its implications for their socio-communicative impairments. And, to our knowledge, no previous study has tested the correlations between production and comprehension in the use of beat gestures by this population.

2 GOALS OF THE DISSERTATION

The goal of this PhD thesis is to investigate the acquisition of synchronous gesture-speech integration by typically developing and autistic children. Three main research questions will be addressed: (1) whether the capacity to produce pointing-speech combinations represents an early pragmatic ability which relates to later language development; (2) whether the presence of beat gestures favors an early ability to recall linguistic information; and (3) whether the presence of beat gestures is beneficial for the understanding of intonation contrasts in a referent resolution task by typically developing children (study 3.1) and autistic children (study 3.2).

Our general hypothesis is that synchronous gesture-speech abilities represent an effective pragmatic strategy which successfully promotes the acquisition of the target linguistic system. This work will try to offer a general perspective on the benefits of the use of gestures that are synchronous with speech from a production and perception point of view during early and late childhood development.

3. THEORETICAL FRAMEWORK

In the last few decades, gesture studies have started to investigate the interplay between gestures, speech and the linguistic systems. Within this field, gesture and speech have been argued to be closely linked, forming an integrated system (McNeill, 1992; McNeill, 2005). According to this author, gestures are defined as communicative acts which relate together with the speech stream at different levels of linguistic analysis.

a) The phonological level, in which prominences of gesture and speech are temporally aligned.
b) The semantic level, in which gestures and speech occurring together express the same idea.
c) The pragmatic level, in which both modalities serve for the same intentional function.

Two of the most important arguments in favor of an integrated communicative system are given by studies investigating the co-occurrence between gestures and speech at the semantic and temporal levels. We know that a communicative gesture may take the linguistic functions of the speech modality. Colonnesi et al.'s (2010) review study shows that the use of iconic and pointing gestures referring to an object or event is related to the appearance of lexical development. Iverson & Goldin-Meadow (2005) and Özçaliskan & Goldin-Meadow (2005) showed that the use of the combination of a gesture together with a word expressing a different meaning is correlated to the onset of two-word combinations (see section 1.2.2. for a summary). Thus, gestures (and specifically pointing gestures) can be used with the same or a different linguistic function than that expressed with words. The connection between the two modalities can be also seen in neurophysiological evidence of shared neural substrates in the use of both communicative modalities (Bates & Dick, 2002; Willems & Hagoort, 2007; Hubbard, Wilson, Callan & Dapretto, 2009). Similarly, through work on populations with language disabilities we know that congenitally blind children naturally gesture when speaking to an interlocutor (see Gullberg, DeBot & Volterra, 2008).

Studies on the temporal alignment of gesture and speech have shown that the prominent part of the gesture, i.e., the stroke phase,[1] is temporally aligned with the prominent part in speech (e.g., Esteve-Gibert & Prieto 2013b, among others). In relation to developmental data we know that infants temporally align gestures and speech intentionally very early in childhood (see summary in section 1.2.2.). This precise ability to integrate gesture and speech productions seems to be impaired in the ASD population, which shows temporal delays between prominences of gesture with respect to speech (Marchena & Eigsti, 2010; see section 1.2.4). Whether this lack of temporal alignment is caused by a dynamic adaptation of speech and gesture in relation to specific communicative goals remains an unresolved question.

The formalization of a unified system of gesture and speech has been modeled by different authors, though some authors have regarded speech as a primary element and gesture as a complement (Kendon, 2007). For instance, the Growth Point Theory in McNeill (1992) and (2005), and the Interface Hypothesis in Kita & Özyürek (2003) consider both modalities to be a unit integrated in communicative acts (to see Gullberg et al., 2008, for a review). Moreover, the study of the development of communicative modalities shows important milestones in multimodal development. Thelen & Iverson's (1999) dynamic model provides a developmental framework in which gestures and speech develop together during the process of language acquisition. This model aims at explaining 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). Findings such as those in Ejiri & Masataka (2001) and Iverson & 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 gesture-speech combinations. Iverson & Wozniak's (2007) study of populations at high risk of developing ASD found that those infants who had less frequent early vocal-motor coordination also developed less gesture-speech

[1] According to McNeill (1992) the stroke phase of the gesture coincides with the interval of maximum effort in the gesture.

coordination. In fact, those children who received an ASD diagnosis were impaired in the production of gesture-speech productions. (see a summary in section 1.2.4).

This PhD thesis is interested in explaining the empirical questions above within an approach that views gestures and speech as parts of an integrated system in human language and communication. A better comprehension of the development of the production and perception of gesture-speech integration during childhood will provide a better understanding of the potential benefits of this ability.

4. HYPOTHESES

Based on the research reviewed in the previous sections, three experimental studies will be conducted:

1. **Experimental study 1.** This longitudinal study investigates the effects of the ability of children aged 12 months to produce synchronic gesture-speech combinations on their later language development. The general aim of the study is to test whether the use of synchronous multimodal cues to attract an adult's attention in a communicative context is an important ability related to later language development. This study will address two main goals:
    a. The first goal will be to replicate and extend prior findings on the role of different social conditions in triggering multimodal communicative strategies, such as synchronous pointing-speech combinations. Liszkowski et al. (2008) reported measures of the vocalizations produced during and after the pointing behavior, but not straight measures of synchronous multimodal productions. We expect a greater use of early multimodal productions during the available condition compared to the baseline condition, but also differences in the use of gesture-speech combinations between the available and unavailable conditions.
    b. The second goal will be to explore the predictive value of the early ability to use multimodal, synchronous pointing-speech combinations at 12 months of age on language acquisition. We predict that the early ability to use synchronous multimodal combinations at 12 months (Esteve-Gibert & Prieto, 2014) in communicatively demanding social interaction situations is related to language development and will be positively correlated with measures of language growth at 18 months of age. We will also extend previous results obtained in naturalistic settings by Murillo & Belinchón (2012) and Wu and Gros-Louis (2014) with experimentally controlled data.

2. **Experimental study 2.** This study will investigate the potential benefits of the presence of beat gestures on children's ability to recall information. Previous studies suggested that children do not rely on beat gestures to recall information in an adult-like way (So et al., 2012). We expect children to understand beat gestures embedded in a child-significant discourse as a

relevant cue to prominence. Thus it is expected that children will benefit from the presence of beat gestures in a word recall activity. Children will therefore recall the intermediate target item in a list better in the beat condition than in the no-beat condition.

3. **Experimental study 3.** The third study will examine the on-line understanding of audiovisual prominence in TD and ASD children (3- and 6-year-olds) with a series of two experiments.

   Experiment 3.1. The first experiment will explore the effects of the presence of focal intonation and beat gesture cues on the understanding of contrastive focus by 4- and 6-year-olds. A reference resolution task will be used, in two conditions, namely, (a) synchronized beat gesture-speech combinations and (b) speech-only. We expect that the use of a beat gesture accompanying intonation prominence will cause an earlier reference disambiguation and thus faster eye fixations to the target word. We thus hypothesize that synchronous gesture-speech integration will favor reference resolution tasks.

   *Experiment 3.*2. The second experiment aims at investigating the use of audiovisual prominence cues in an ASD population, and determining whether they can make use of visual cues to understand focal prominence as a cue for referent resolution (in a contrastive focus situation) before and after training their ability to understand this intonation contrastive cue. Children around 6 with ASD will be tested with the same reference resolution task and conditions as in Experiment 1, and will be compared with a matched TD population. In this experiment we will use a three-duration test-treatment-test design to test the potential benefits of explicit instructions on the use of audiovisual prominence cues and its impact on reference resolution. We expect difficulties in the understanding of prominences in ASD children, with slower fixations to the target object marked with prosodic and gestural prominence (Kehoe, 2013; Peppé, McCann, Gibbon, O'Hare & Rutherford, 2007). We also expect an improvement in the understanding of contrastive reference resolutions abilities after explicit training. Finally, in both experiments, we also expect children with better pragmatic abilities to perform with faster looking fixations to gesture and prosody cues of pragmatic linguistic knowledge such as prominence (Silverman et al. 2010; Hubbard et al., 2012).

5. EXPERIMENTAL STUDIES

5.1. Study 1

5.1.1. Methods

Participants and materials

A sample of nineteen infants (N = 19; 12 boys and 7 girls) born at term, healthy and with normal hearing participated in this research. They were all growing up in middle-class monolingual Spanish families and they were tested at 12 months of age (mean age: 12;12; range: 11;23-12;27) and contacted again at 18 months of age to obtain expressive vocabulary measures from parental report.

The experimental setting was based on Liszkowski et al. (2008). The experimental sessions took place in a 2 x 3.5 m distracter-free testing room where the portable set-up could be easily placed (see Figure 1). The caregiver wore a pair of earphones with music to distract him or her from the activity, thus avoiding interference with the child's spontaneous behavior. The setting includes a curtain with six holes, three on each side, where six of the objects manipulated by an assistant hidden behind it.

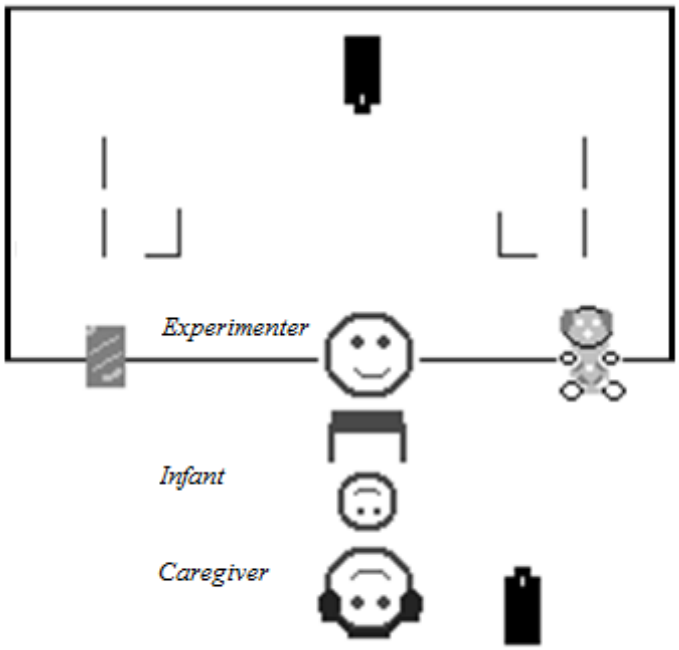

***Figure 1**. Schematic representation of the central area within the testing room.*

Through a large opening in the upper center of the curtain a camera recorded the child's reactions, and a second camera, placed at the back of the room, recorded the sequence of events from the child's perspective (presentation of the stimuli and experimenter's visual and verbal behavior).
(see Figure 2).

Figure 2. *Views of the back (left side) and front (right side) cameras.*

A total of ten stimuli were manipulated by an assistant hidden behind the screen: 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) (see Figure 3), which always were visible to the child, located on the floor at approximately 30° to the infant's left and right, and also activated by the assistant. 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, 2005). A moveable bead toy and a pair of books were used between conditions to return the infant's attention to the experimenter.

Figure 3. Experimental stimuli consisting of eight puppets, a light and an electronically activated dancing pig.

Procedure

Liszkowski et al.'s (2008) procedure was adapted to elicit the targeted communicative behaviors (i.e., pointing-speech combinations, pointing-only and speech-only productions) through an enjoyable event, by presenting puppets/toys from behind the experimenter. The procedure involved different social interaction in two experimental conditions (i.e. available and unavailable) and a baseline condition differing on the adult's joint attention patterns (see Figure 4). The declarative pointing task is especially suited to our purposes for two main reasons: (a) it controls for the infant's communicative intention to comment about an interesting event happening behind the adult; and (b) it also controls for the adult's joint attention patterns.

Social conditions were defined as in the original 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., the experimenter repeatedly looked back and forth between the stimulus and the infant's face), pointed to the object, and used speech to say, "Oh..., look, it's a cat! Look! It's saying hi to you! Oh…, a cat!" (see Figure 2, top row).
- In the ***available condition*****,** the experimenter looked to the infant but did not look to 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 continued looking to the infant's eyes, but did not look to the stimulus, while saying, "Hmm…? What…? What's there…? Hmm…?" (see Figure 2, middle row).
- In the ***unavailable condition****,* the experimenter was not visually attending to the infant neither to the object. First, while the stimulus was activated the experimenter's attention was directed at neither the infant nor the stimulus, but to the book. After the first infant's pointing to the stimuli, the experimenter continued looking at the book while saying "Hmm…? What…? What's there…? Hmm…?" (see Figure 2, bottom row).

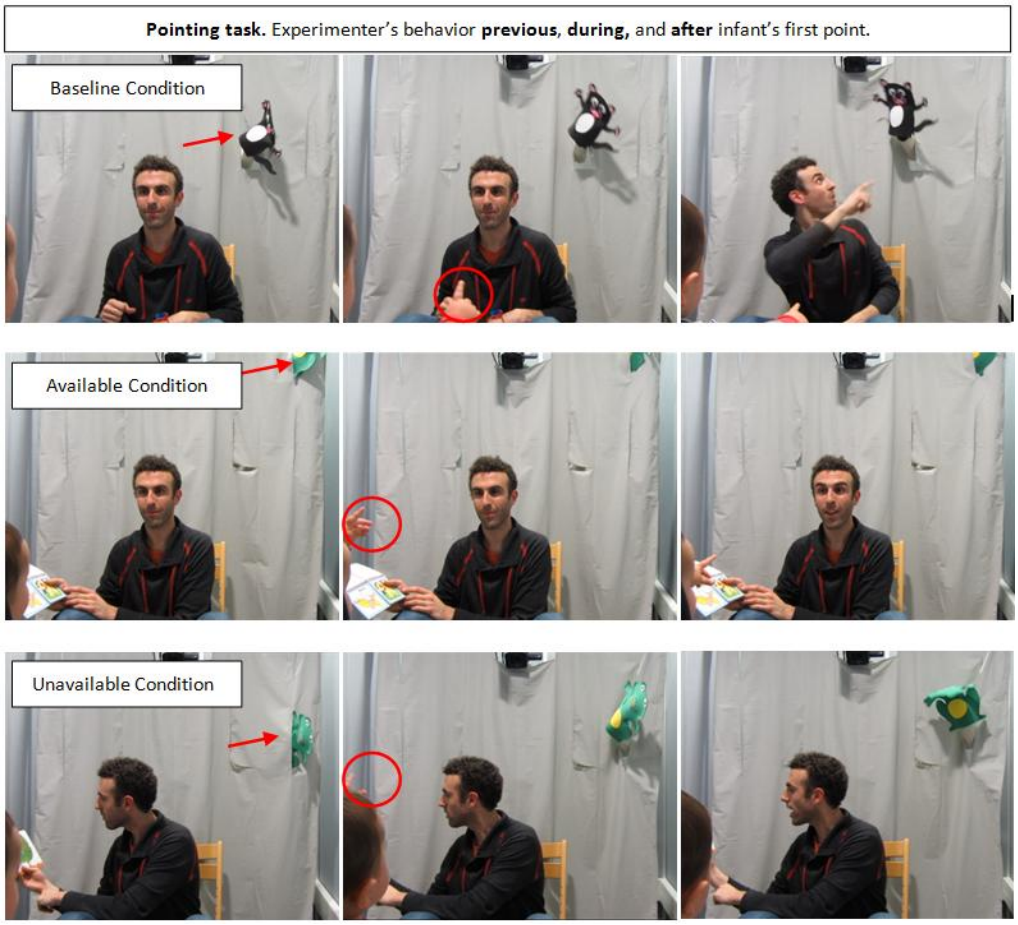

Figure 4. Images showing the experimenter's behavior in the baseline (first row) and the two experimental conditions (available and unavailable, second and third rows respectively) before, during and after infants' first pointing (circled). The arrow in the pictures of the first column shows the time of stimulus activation.

The warm-up period began consisted of an enjoyable play (which ranged from 5 to 20 minutes) between the experimenter and the infant. Then they moved to the testing room where the pointing task was developed. 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 he/she did so, the stimulus continued to be visible for another 20 seconds or until the infant was no longer interested. If no pointing gesture was produced in reaction to the stimuli (i.e.

no communicative productions or vocalizations without a pointing), 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, was controlling the duration of the trial and by clucking her tongue she indicated to the experimenter the end of the trial. Before experimental trials, the experimenter and the infant shared a book activity, until the infant was relaxed and attentive, then gradually withdrew the activity, indicated to the assistant by means of a finger snap to activate the next stimulus, and looked at the book which he hold at the opposite side of the stimulus side of appearance.

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 between subjects (i.e., Baseline–Available–Unavailable or Baseline–Available–Unavailable) and two orders counterbalanced in terms of the side of appearance of the first stimulus (right or left side). Each sequence was repeated 5 times, so that every child completed a total of 15 trials. The order of presentation of the stimuli was randomly chosen by the assistant. The whole experimental session lasted approximately 18 minutes. 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 & Gallego, 2005). They were contacted again and requested to fill out and send back the form when the child was 18 months of age.

Coding

Coding was performed with ELAN software (Lausberg & Sloetjes, 2009), especially 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. Only instances of pointing that signaled at the target stimulus of the trial were coded, while other communicative gestures (e.g.., waving the hand to say “hello”) were not taken into account.

Figure 5. Snapshot of the ELAN coding scheme.

In what follows, specific criteria used for coding the infants' behavior are described.

- *Pointing-only:* Isolated pointing gestures were coded when the infant signaled with the whole arm or with the hand to the stimulus of the trial. Extended finger and extended palm downwards hand configurations were considered as pointing gestures (as in Brooks & Meltzoff, 2008; Liszkowski et al., 2008; Cartmill et al., 2012).
- *Speech-only*: Isolated speech productions were coded when the infant uttered speech sounds consistently to communicate about the target stimulus presented in the trial. These communicative vocalizations were coded when they were directed at the experimenter or clearly intended as a comment about the stimulus (e.g., immediately after the infant had looked at the object). Speech sounds were excluded if the infant was not looking at the object but towards any other place/object in the room. Infants' shouts, laughs, groans, or vegetative sounds were excluded. In order to consider two sounds as independent vocalizations they had to be separated by a pause of at least one second in duration.
- *Pointing-speech combinations.* Synchronous pointing-speech combinations are defined to share all pragmatic function, semantic content and phonological temporal cues (McNeill, 1992; Butcher & Goldin-Meadow, 2000). According to

McNeill (1992), the stroke phase of the gesture coincides with the interval of maximum effort in the gesture. Therefore, we regarded communicative productions as synchronous pointing-speech combinations by looking to their temporal alignment, so that vocalizations which overlapped at least with some portion of the stroke of the pointing gesture were coded as synchronous (McNeill, 1992). 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 on the coding procedure. Observers assessed 18% of the data across conditions, which corresponded to a total number of 52 trials. Agreement for presence/absence of communicative productions in each trial was good: overall agreement was 90% and the fixed-marginal kappa statistic was 0.73. The overall agreement for the classification of communicative acts (61 items) into one of the three categories (namely, *speech-only*, *gesture-only* or *pointing-speech combination*) was 86% and the fixed-marginal kappa was 0.77, indicating that there was substantial agreement among independent coders.

5.1.3. Summary 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 synchronous pointing-speech combinations at 12 months of age and (2) the predictive value of early pointing-speech combinations on language development measures at 18 months. A total of 341 communicative behaviors were coded, including speech-only productions (N = 192), pointing-only productions (N = 75) and pointing-speech combinations (N = 74).

(1) 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 conditions) we conducted a Generalized Linear Mix Model (GLMM) with number of communicative production as a dependent variable (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), and 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.

The results of the GLMM analysis showed a main effect of communicative modalities (F (2, 828) = 26,418, 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, 828) = 22,667, p < .001). Figure 6 shows the distribution of the number of occurrences per trial (expressed in means) of speech-only, pointing-only and pointing-speech combinations grouped by social conditions produced by the 19 children in our sample.

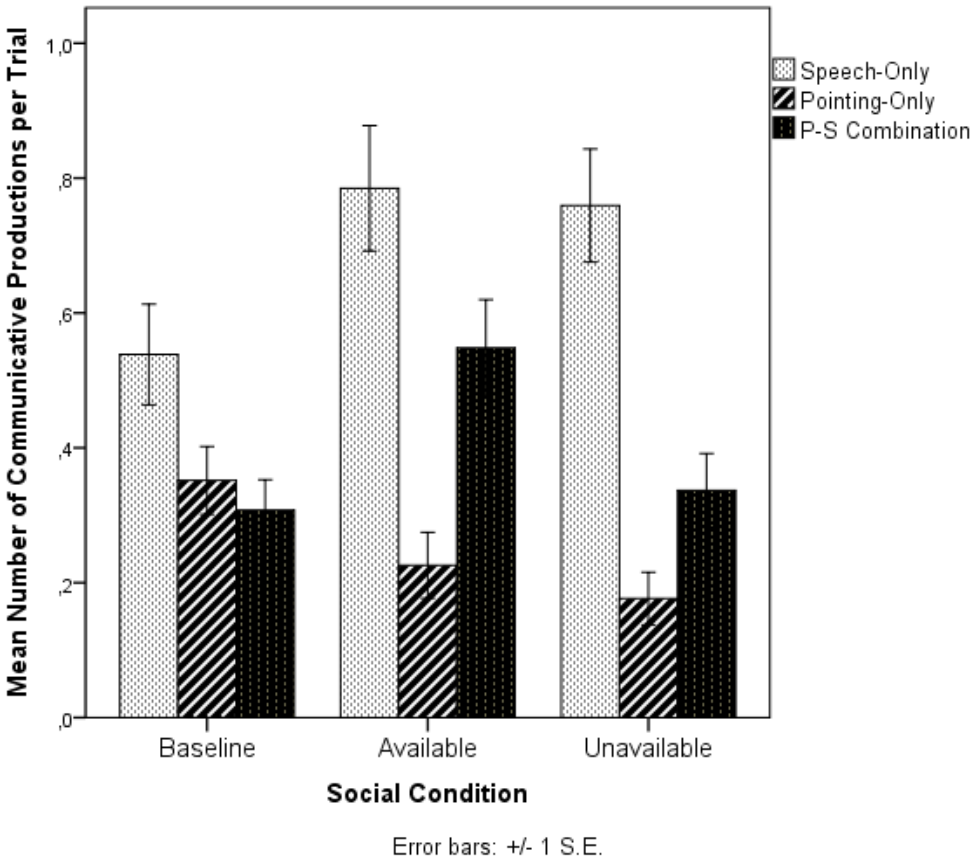

Figure 6. Mean number of occurrences of each type of communicative productions (speech-only, pointing-only and pointing-speech combinations) per trial as a function of social condition (baseline, available and unavailable conditions). Error bars: +/- 1 S.E.

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 modalities, (F (4, 828) = 3,597, p < .01), indicating that communicative productions behave differently depending on the social condition. More interestingly for our purposes, post-hoc analysis of the effect of social condition on the distribution of communicative modalities showed main significant effects of synchronous pointing-speech combinations with a greater number of productions in the available condition than in the baseline condition or the unavailable condition, (F (2, 828) = 5,915, p < .01). By contrast, speech-only productions and pointing-only productions did not significantly differ between social conditions, respectively, (F (2, 828) = 1,165, p = .312) and (F (2, 828) = 1,392, p = .249). Only synchronous pointing-speech combinations thus increased in number in the available condition with respect to the baseline condition. As expected, post-hoc analyses revealed again that speech-only productions were more frequently produced than the other communicative productions in every social condition (baseline F (2, 828) = 5,262, p < .01; available F (2, 828) = 13,859, p < .001; and unavailable (F (2, 828) = 12,258, p < .001), but they did not differ between conditions.

(2) *Predictive value of synchronous pointing-speech combinations for the development of expressive language.*

In this subsection we first provide the results of the Spanish version of the MacArthur-Bates CDI that were obtained at 18 months of age. Expressive vocabulary and sentence (word endings and morphosyntactic section) CDI sections were obtained at 18 months as dependent variables. The expressive vocabulary section measures the infant's productive outcome, and sentence sections measure the ability to form correct word endings (endings section) and to form correct morphosyntactic utterances (morphosyntactic section). Table 1 shows the median, average, standard deviation (SD), minimum and maximum CDI scores on vocabulary and grammar sections at 18 months.

	Median	Mean	SD	Min	Max
Expressive vocabulary section	44	65,42	82,823	3	367
Endings section	0	1,63	2,409	0	8
Morphosyntax section	6	11,21	15,072	0	67

Table 1. CDI scores at 18 months.

In this section, the independent variables are defined as the total number of communicative productions (i.e., separated by speech-only, pointing-only and pointing-speech combination productions) in each of the social conditions (i.e., baseline, available and unavailable conditions), as follows: speech-only productions, pointing-only productions and synchronous pointing-speech productions in the baseline condition; speech-only productions, pointing-only productions and synchronous pointing-speech productions in the available condition; and speech-only productions, pointing-only productions and synchronous pointing-speech productions in the unavailable condition.

We calculated Pearson's bivariate correlations between language measures of different language outcomes at 18 months and communicative measures elicited in each social condition at 12 months. The results of the correlations revealed that only synchronous pointing-speech productions uttered in available condition were positively correlated with later language development. Pointing-speech combinations expressed at 12 months in the available condition were positively correlated with expressive vocabulary (r = .537, p <.050) and morphosyntactic (r = .521, p <.050) measures, but not with the endings measure (r = .521, p <.050).

In order to test the predictive value of earlier pointing-speech combinations in different social conditions, a total of nine multiple regression analysis were run with expressive vocabulary, word endings and morphosyntactic measures by using a backward elimination method. We ran a multiple regression analysis for the three types of communicative production in each social condition, since the results in the previous section (Section 3.1) already showed significant differences depending on social condition in terms of the infant's communicative behavior. For example, the first regression analysis contained three independent variables, namely the number of speech-only productions, the number of pointing-only productions and the number of pointing-speech combinations in the baseline condition as independent variables, and

expressive vocabulary measures as the dependent variable. The second regression analysis was run with the same variables but in the available condition, and the third analysis included communicative productions elicited in the unavailable condition. Independent variables were included similarly in the consecutive regression analysis, but with word endings and morphosyntactic measures as dependent variables, from fourth to sixth and from seventh to ninth regression models, respectively. Table 2 shows the results of the most effective set of independent variables (the number of communicative productions uttered in trials with different social conditions) in the prediction of different language measures at 18 months.

Model number	*Social condition tested in the model*	*Communicative productions included in the model*	*R^2 statistic (%)*	*p-value*
Expressive vocabulary measure at 18 months (dependent variable)				
1	Baseline variables			$p = .083$
2	**Available variables**	**P-S combinations**	**28.8**	$\mathbf{p < .05}$
3	Unavailable variables			$p = .431$
Endings measure at 18 months (dependent variable)				
4	Baseline variables			$p = .151$
5	Available variables			$p = .067$
6	Unavailable variables			$p = .312$
Morphosyntactic measure at 18 months (dependent variable)				
7	Baseline variables			$p = .098$
8	**Available variables**	**P-S combinations**	**27.2**	$\mathbf{p < .05}$
9	Unavailable variables			$p = .466$

Table 2. Multiple regression analyses of the most effective models predicting infants' vocabulary, endings and morphosyntactic 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.

The results of the first multiple regression analysis, which included communicative productions elicited in the baseline condition in relation to expressive vocabulary measures, did not show any significant differences (F (3, 18) = 2,688, p = .084). The results of the second regression model, which included communicative productions

elicited in the available condition, showed significant differences and indicated that the number of pointing-speech combinations in available condition at 12 months explains 28.8% of the expressive vocabulary variance (R^2 =.288, *F* (1, 18) = 6,877, *p* = .018). It was found that synchronous pointing-speech combinations uttered during the available condition was the best model which predicted vocabulary tendencies (β = .54, *p* <.05). The third multiple regression analysis did not indicate significant relations between communicative productions in the unavailable condition at 12 months and vocabulary measures at 18 months (*F* (1, 18) = .650, *p* = .431). In relation to the word endings measure, the most effective model in the multiple regression analysis, which included this measure as the dependent variable (models numbered from forth to sixth in table 2), did not find significant differences for any of the groups of dependent variables in the baseline (*F* (1, 18) = 2.265, *p* = .151), available (*F* (1, 18) = 3.822, *p* = .067) or unavailable conditions (*F* (1, 18) = 1.088, *p* = .312). Finally, the dependent variable related to morphosyntactic measure was tested in models numbered from seventh to ninth in table 2. The best model, which related communicative productions and morphosyntactic measure, reported significant differences in the available condition (*F* (1, 18) = 6.340, *p* <.05), but did not report significant differences in either the baseline condition (*F* (1, 18) = 3.069, *p* = .098) or the unavailable condition (*F* (1, 18) = .556, *p* = .466). The results of the eighth regression model, which included communicative productions elicited in the available condition, indicated that number of pointing-speech combinations in available condition at 12 months explains 27.2% of the morphosyntactic variance (R^2 =.272). It was found that synchronous pointing-speech combinations uttered during the available condition was the model which best predicted morphosyntactic tendencies (β = .52, *p* <.05).

5.2. Study 2

5.2.1. Methods

For this study we tested sixty 3-, 4- and 5-year old Catalan-dominant children, as well as 20 adults. Preliminary results presented in this document are from 43% of the data (26 children out of 60, 3- (*N* = 8), 4- (*N* = 6), and 5-year-olds (*N* = 12)). These participants were recruited at a primary school in Arenys de Munt in the province of Barcelona.

The children participated in a live story-telling activity, which was presented by the experimenter, featuring an elephant named Elmer that enjoys travelling but has a tendency to forget things. The story-telling was accompanied by a PowerPoint presentation shown by means of a laptop (see Figure 7). Prior to each trial the experimenter told the child that he or she would need to help the elephant remember the list of target words, which would be uttered by a female adult, who was introduced as "a good friend of the elephant" (see Figure 7, left). The list of words of the exposure phase were presented in an audiovisual recording (see Figure 7, center). A distractor image was presented after each exposure phase (see Figure 7, right).

Figure 7. (1) Slide from PowerPoint presentation showing Elmer the elephant and his friend (left), (2) screenshot from audiovisual recording of one of used in the exposure trials (center) and (3) distractor image (right).

Each exposure trial consisted of an audiovisual recording of the female adult (the "elephant's friend") uttering first the introductory sentence "Elmer, remember! Before leaving..."; and then information on the duties of the elephant, which consisted of four sets of 5 target nouns, each set related to a specific semantic context. Each trial consisted of five different disyllabic nouns with word-initial stress (i.e., trochaic words) from the same semantic category, for a total of 4 trials. The complete set of 4 lists with 5 targets words each is presented in Table 3.

Introductory sentence	
"Elmer, recorda! Abans de marxar…" (Catalan)	
"Elmer, remember! Before leaving..." (English)	
Semantic contexts	***Target nouns***
Drawing the school	
"Fés un dibuix de l'escola que tingui:..."	*"... portes, arbres, classes, taules, plantes."*
"Draw a picture of the school with…"	*"... doors, trees, rooms, desks, plants."*
Tidying up the room	
"Endreça l'habitació i posa a l'armari:..."	*"…cotxes, nines, bosses, papers, llibres."*
"Tidy up your room and put in the wardrobe...	*"…cars, dolls, bags, papers, books."*
Buying food	
"Ves al mercat i compra:..."	*"...pomes, iogurts, cebes, raïm, aigua."*
"Go to the market and buy:..."	*"... apples, yogurts, onions, grapes, water."*
Saying goodbye to animals in the zoo	
"Ves al zoo i acomiada't dels:..."	*"...ossos, lloros, ànecs, pardals, cavalls."*
"Go to the zoo and say good-bye to the:..."	*"... bears, parrots, ducks, birds, horses."*

Table 3. Linguistic materials presented during the word exposure phase.

Importantly, the design of the experiment controlled for potential serial sequential effects. A serial sequential effect is a domain-general ability to remember items when presented in initial (primacy effect) and final positions (recency effect) in a list of successive items. This ability has been reported from very early in infancy (Lewkowicz, 2013). We controlled for these effects in our list of nouns by only showing the beat gesture only when the intermediate (target) item (third word out of five) was uttered, since by implication this word would benefit from neither primacy nor recency effect.

The experimental conditions were defined as follows:

a) In the *beat condition* the target item was uttered with clear hand and head beat gestures. This contrasted with the preceding and following items, as no specific gesture was coupled with them. The beat gesture consisted of an up-down head nod movement produced together with an up-down movement of both hands with palms up, and synchronized together with the target word in speech (see Figure 8).
b) In the *no-beat condition* the adult produced the five words in the list without hand or head gestures (see Figure 9).

Figure 8. Picture of the visual information during the beat condition

Figure 9. Picture of the visual information during the no-beat condition

The materials in Table 3 were audiovisually recorded to be apt for presentation in the beat condition (i.e., with the intermediate target word produced with a beat gesture), and in the no-beat condition (i.e., with the intermediate target word produced with no beat gesture).

Procedure

The experimental session was organized in the following way. First, the experimenter tested children's memory span, so that we could control for the maximum number of items that the child could remember in an open list of disyllabic words. The experimenter told the story about the elephant, and showed the audiovisual recordings of the exposure phase. The exposure phase began with two familiarization trials which included a list of three words which were different from those in the exposure phase. The task continued with the four trials presented during the exposure phase. After seeing the recording of each exposure trial the child was asked by the experimenter to recall as many items as he/she could. Each test trial was followed by a distracting image of the elephant, so that children could relax and engage again with the storytelling (see right-hand picture in Figure 7). The within-subjects experimental design was organized as a sequence of two trials per condition (a total of 4 trials) in a counterbalanced order (i.e,. beat/no-beat x 2 trials and no-beat/beat x 2 trials) between subjects. Each sequence was repeated twice, so that every child completed a total of 4 testing trials. The experiment lasted a total of approximately 8-10 minutes.

5.2.3. Preliminary results

In order to assess the effects of gesture condition (i.e., beat vs. no-beat condition) on the proportion of the number of items remembered in each position (namely, first, second, target, fourth and fifth positions) we conducted a T-test analysis with proportion of the number of communicative production as a dependent variable. Preliminary results with 26 children showed that infants recalled the target item in the beat condition significantly better than in the no-beat condition when their memory span ability was similar to the length of the testing list (4-5 words), $t(23) = -2.713$, $p < .001$ (see right panel, Figure 10). But infants who had a more limited memory span (2-3 words) remembered the target item significantly better in the no-beat condition than in the beat condition, $t(24) = 4.371$, $p < .01$ (see left panel, Figure 10). Children behaved differently depending on their memory span ability and not according to their age, and remembered the target item better in beat condition when their memory was similar to the number of items. On the other hand, if their memory span was lower (2-3 words) they remembered the target items better in the no-beat condition than in the beat condition. That is, there was a significant effect of condition on the ability to remember the target item, but this effect was more closely correlated with their memory span than with their age.

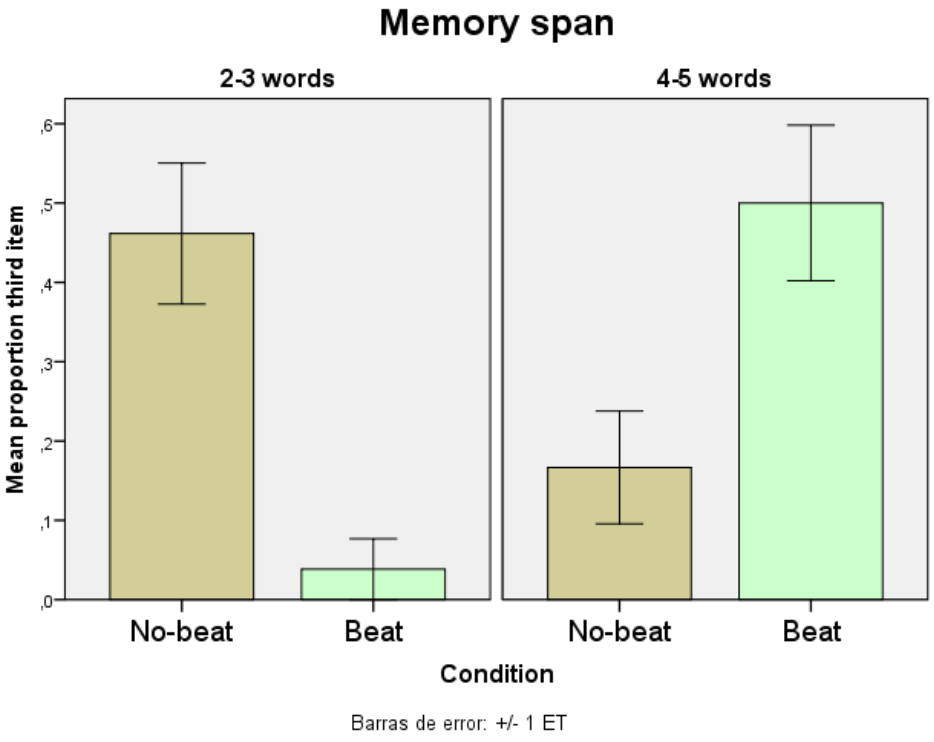

Figure 10. Mean proportion of items remembered in the third (target) position of the word list depending on condition (beat and no-beat conditions) and

separated by memory span abilities (2-3 word memory span = left panel; 4-5 word memory span = right panel).

To explore these results further, we tested the behavior of all items in the list. First, the second, fourth and fifth items in the list behaved similarly in the beat and the no-beat condition. The right-hand panels in Figure 7 show a clear increase in the children's ability to recall the third (target) item when presented with a beat gestures. Children relied positively on the presence of the beat gesture when their memory abilities were similar to the demands of the task. By contrast, children who were tested with a longer list in relation to their memory ability tended to forget the first items in the control no-beat condition. This result may be explained by an overriding of the primacy effect when the task is too demanding (see top left panel in Figure 11).

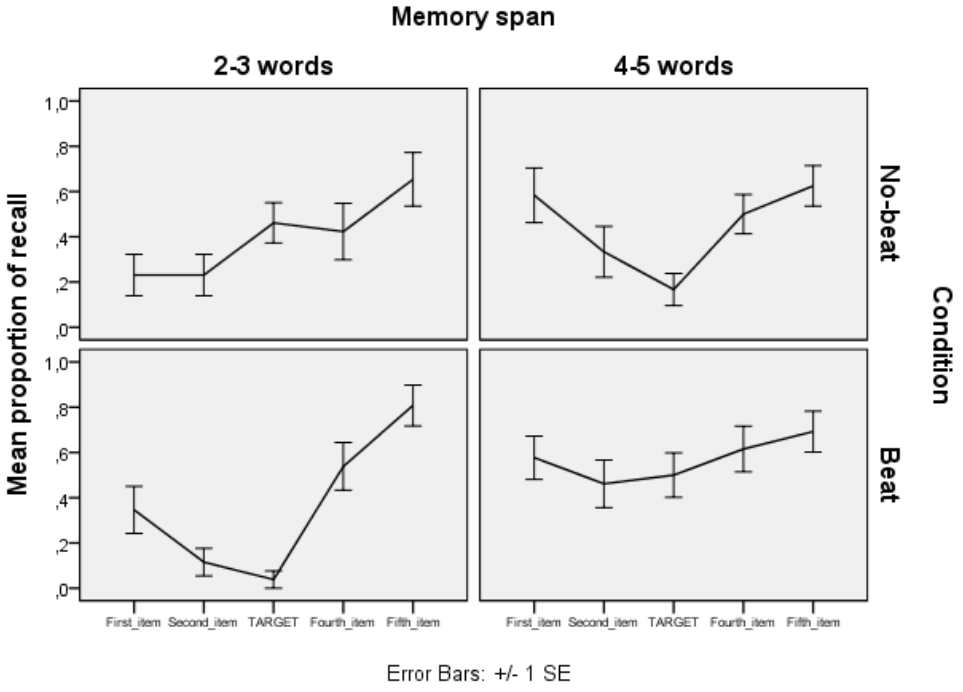

Figure 10. Mean proportion of remembered items in the five possible positions of the list separated by condition (beat and no-beat conditions) and memory span abilities.

Thus it would appear that children's ability to extract information about prominence from a beat gestural cue in a memory task depends on their memory span ability. Infants whose memory span accorded with the number of items included in the test list benefited from the beat gestural cue. By contrast, infants who had shorter memory spans did not benefit from the presence of the beat gesture and tended to forget the

target item. Therefore, we can tentatively conclude that beat gestures help children to recall information when they function as a prominent cue in a discourse context depending on the memory ability of the child. We expect to further investigate whether the drop in recall of the target items in children with smaller memory span capacities may be explained by (a) the length and complexity of the testing list, or (b) potential interference due to the presence of gestural cues.[2]

[2] In Kelly & Lee (2012) the use of communicative gestures, specifically iconic gestures, actually disrupted word learning when the phonetic demands were too high with a group of second-language learners. If the diminished recall could be explained by the demands of the task, we expect that this question could be solved by adapting the task to children who are able to remember a maximum of 2-3 words in a memory span task , namely by using lists of three items in the beat and no-beat conditions. If on the other hand the diminished recall was due to a potential interference of the presence of gestural cues, we should find the same kind of diminishing effects on serial recall abilities with iconic gestures. We follow the first hypothesis; the second one would contradict the first part of our pilot results showing the benefits for recall when abilities were commensurate with the demands of the task, and also inconsistent with the results in So et al (2012) regarding reference recall when presented with an iconic gesture.

5.3. Study 3

5.3.1. Methods

The third study examines the on-line understanding of contrastive focus in TD and ASD Catalan-dominant (or Spanish-dominant) children with a series of two experiments. It is important to note here that the experimental design of this series of studies will be reconsidered once collaboration with a research group specialized that specializes in ASD is underway, and we anticipate that some of the key points of the study will need to be reconsidered at that time.

Experiment 3.1.

In experiment 3.1 we will test forty 4- and 6-year-old children with a referent resolution task (see a detailed description of the task in Ito, et al., 2012; Ito et al., 2014) in two different conditions, namely, (a) beat gesture-speech combinations and (b) no-beat gesture. Contrastive focus prosodic prominence will be marked with a L+H* focus marking versus a flat intonation contour. The beat gesture would be an up-down head nod movement produced together with an up-down movement of both hands with palms up, and synchronized with the target word in speech.

In the original task, children's eye fixations are measured with an eye tracking procedure to test their attention to images after a verbal command. We will adapt this task to audiovisual information in a visual world paradigm similar to Silverman et al. (2010), in which the audiovisual input is placed in the middle surrounded the target images (See Figure 11).

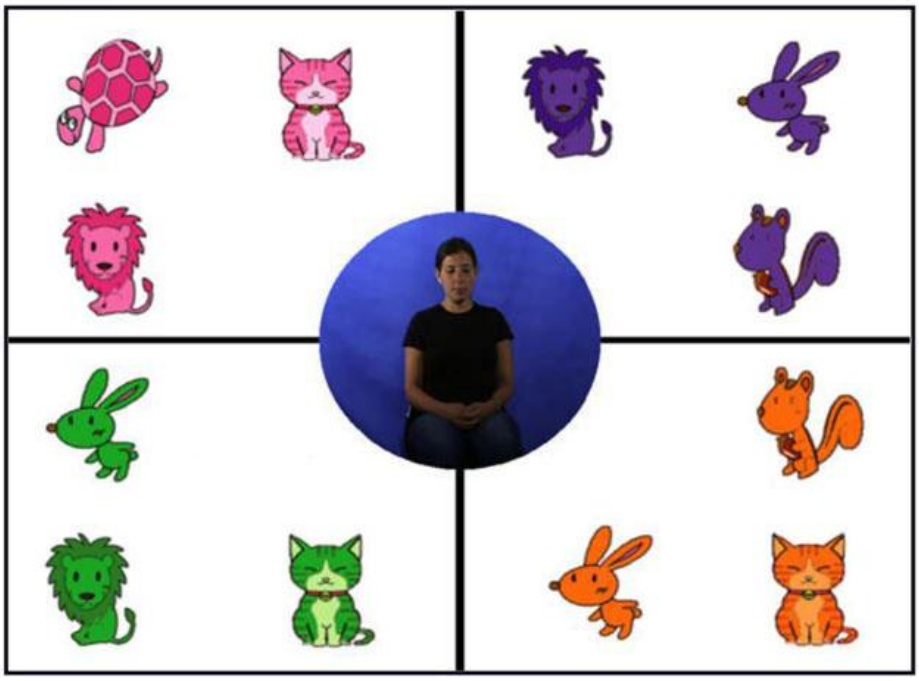

Figure 11. Adaptation of the visual world paradigm in Silverman et al. (2010) and in Ito et al. (2012).

The reference resolution task will be adapted from Ito et al. (2012) using their images of colored animals. Ito et al.'s (2012) experiment revealed faster fixations with the use of pitch contrast prominence on the adjective in a sequence like “Point to the orange cat”→”Now, point to the GREEN cat” than when it was not marked with prosodic prominence (i.e., “Point to the orange cat”→”Now, point to the green cat”). Children perceived the use of intonation contrast signaling contrastive focus as a relevant cue in the selection of the noun referent. In our case, because of the different order in the adjectival phrase in Spanish, the context of the game will be adapted to point to animals colored in different colors. One possible adaptation of this task to Spanish would be as follows, placing the contrastive prominence on the target noun rather than the adjective (1).The prominence on “dog” will signal that the adjective will be the same as in the preceding sentence. Remember that in the original task, the prominence was associated with the pronominal adjective.

(1) “Señala al gato rosa” → “ahora, señala el PERRO rosa”,
“Point to the pink cat” → “now point to the pink DOG” (prominence associated with the noun)

A second option for the adaptation of the linguistic materials (perhaps more appropriate from a pragmatic point of view) would be to consider animals performing different actions (i.e., eating, drinking or dancing), as follows in (2) and (3).[3]

(2) "Señala al gato que come" → "ahora, señala al PERRO que come"
"Point to the cat which is eating" → "now, point to the DOG which is eating"
(3) "la mosca vuela" → "el MOSQUITO vuela"
"the fly is flying" → "the MOSQUITO is flying"

Experiment 3.2.

The second experiment aims at investigating the use of audiovisual prominence cues in an ASD population and determining whether they can make use of visual cues to understand prominence as a cue to referent/contrast resolution (Ito et al., 2012; Ito et al., 2014). Children around 6 years of age with ASD will be tested with the same reference resolution task described above. They will also be compared with a control TD population matched for age, gender, and verbal abilities (Silverman et al. 2010). Both groups will be tested with the same two conditions as in our first study, namely, (a) beat gesture-speech combinations and (b) no-beat gesture. In this study we will use a test-treatment-test design to test the potential benefits of explicit instructions on the audiovisual integration of prominence cues and their impact on reference resolution. The ASD population will be separated into two treatment groups according to (a) an explicit training session about the use of beat gestures as a cue of to prominence and (b) a control training session. This study will address some questions about gesture-speech integration in populations with socio-communicative impairments such as ASD.

To examine ASD symptom severity, the Autism Diagnostic Observation Schedule - ADOS (Lord, Rutter, DiLavore & Risi, 2002) will be administered by a qualified

[3] Another possible adaptation of the materials to test reference resolution is based on Dahan, Tanenhaus & Chambers (2002), who found similar results on word resolution. When the target word was repeated (e.g., candy-candy) a prominent accent (e.g., CAN...) induced higher fixations to the cohort competitor (e.g., candle). In the non-repeated sequences (e.g., candy-candle), a lack of accent (e.g., can...) induced higher fixations to the just-mentioned object (e.g., candy).

(1) "Pon la cama encima del triángulo" →"ahora, pon la CASA encima del triángulo"

research clinician. This is a well-known standardized semi-structured play-based assessment (Gotham, Pickles & Lord, 2012; Lord & Jones 2013) which provides researchers with quantitative and qualitative information relevant to ASD symptoms, by combining measures of social, communication and repetitive behaviors for different age and communication levels. We will take into account the use of the Spanish adaptation of Peppé's prosodic ability test (PEPS-C: Profiling Elements of Prosodic Systems-Children) (Peppé & McCann, 2003) in Martínez-Castilla & Peppé (2008), which will allow comparative analysis.

We will control for empathy abilities with an adaptation of the task applied in Ruffman, Slade & Crowe (2002), as gesture production abilities have been related to children's empathy abilities (Chu, Meyer, Foulkes & Kita, 2014). In this task, first, children will be shown four faces personifying different emotions (happy, sad, scared, and angry), and then will be asked to identify each of them. If a child does not respond with the correct label, the experimenter will tell the child to find the face with a particular emotion, until all four emotions are clearly associated with the correct label. Children will then be given 8 stereotypical situations which can cause happy, sad, scared or angry emotions, in which the characters' faces are not shown, so that children cannot infer an emotion from their face (e.g., the character is offered a present by her mother (happy), a dog wants to bite the character (afraid), a friend of the character does not want to give him back his toy (angry, sad)). For each situation, children will be asked, "How does the [name of the character] feel?" Moreover, theory of mind abilities will be assessed with an adaptation of the classical Sally/Ann task (de Villiers, 2007) following the adaptation in Armstrong, Esteve-Gibert & Prieto, (2014). The material for this task will consist of a short video featuring two puppets in which Puppet 1 (a princess) puts a ball in a bucket and leaves for the school, then Puppet 2 (a lion) transfers the ball to a different bucket, covers the ball, laughs in a sneaky way and leaves. Once Puppet 2 reappears, children will be asked to point to (a) where Puppet 1 will look for the ball and (b) where the ball really is.

5.3.2. Expected results

Experiment 3.1 aims at studying the impact of beat gesture on the understanding of intonation contrasts in a referent resolution task by TD children at 4 and 6 years old. We expect that reference resolution will be performed by children with faster fixations when presented the word with an intonation prominence, as in Ito et al. (2014). This result will replicate Ito et al.'s (2014) results and additionally provide more evidence

about the understanding of in-situ prominence in Romance languages like Spanish or Catalan. To our knowledge no study to date has tested whether the use of a beat gesture accompanying an intonation prominence will cause faster eye fixations to the target word. We hypothesize that synchronous gesture-speech integration will favor reference resolution since the gesture's onset occurs before speech (Silverman et al. 2010) and beat gestures enhance auditory perception (Krahmer & Swerts, 2007).

In experiment 3.2. we will seek to find out whether the ASD population is sensitive to intonational/gestural contrast in a referent resolution task and whether this ability can be trained. We expect difficulties in the understanding of prominences in ASD with slower fixations to the target object marked with a prominence. Previous studies on the understanding of intonation prominence in an ASD population have shown difficulties in this ability (Kehoe, 2013; Peppé, McCann, Gibbon, O'Hare & Rutherford, 2007). On the other hand, some literature has found processing difficulties in the integration of representational gestures synchronized with speech in ASD (Silverman et al. 2010) and different neural activity during exposure to beat-gesture synchronized with speech in an ASD population was found to be related to their socio-communicative abilities, such that the better their socio-communicative abilities, the more similar the neural activity was to that seen in a matched control population (Hubbard, et al., 2012). We will expect children with better pragmatic abilities to perform with faster looking fixations to gesture and prosody cues to pragmatic linguistic knowledge such as prominence. We also expect an improvement in gesture-speech integration abilities after explicit training in the understanding of prominence expressed through gesture and speech.

6. WORKING SCHEDULE

June 2014	○ Submission of the Ph.D. project ○ Corrections of the article to be submitted in the *Journal of Infant Behavior & Development* (Study 1) ○ Oral presentation at VI Workshop of Catalan prosody
July 2014 to October 2014	○ Data collection and analysis for Study 2 (beat gesture and memory)
November 2014 to December 2015	○ Writing of Study 2 ○ Autism Diagnostic Observation Schedule – ADOS five days of training
January 2015 to February 2014	○ Preparation of the linguistic materials for Study 3
March 2015 to May 2015	○ Potential 3-month stay abroad months to collaborate in study 3.
June 2015 to October 2016	○ Data collection and analysis for Study 3
November 2016 to January 2016	○ Writing of Study 3
February 2016 to May 2016	○ Writing of the Ph.D. dissertation
July 2016	○ Dissertation defence

Conferences:

Academic year 2014-15

- IASCL, International Association of the Study of Child Language. July 2015, 14-18th. Amsterdam.
- LARP7: Laboratory Approaches to Romance Phonology. September 2014, 3-5th. Aix-en-Provence, France.
- AETAPI Congress, Spanish Association of Autism Professionals. November 2015, 13-15th.
- PaPE: Phonetics and Phonology in Europe. June 2015. Oxford, England.

Academic year 2015-16

- 16th ICPLA Conference. International Clinical Phonetics and Linguistics Association.
- ASHA association congress. American Speech-Language Pathology Association.
- BCCCD16 Budapest CEU Conference on Cognitive Development.
- ISGS International Society of Gesture Studies Conference.

7. SELECTED REFERENCES

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 & Prieto (2014) analyzed four 11- to 19-month-old Catalan-speaking children during daily activities in order to test (1) when and how children combine communicative gestures synchronously with speech, (2) the early temporal synchronizations between pointing and speech combinations. Results showed that infants already produced temporally synchronous gesture-speech combinations in the babbling period (11-month-olds) but isolated pointing was still produced more frequently around this period. The sample of gesture-speech productions increased significantly by 15 months of age. These multimodal productions were mostly pointing and reaching gestures with a declarative communicative purpose. When combined, the two modalities were temporally synchronized, i.e., the onset of the stroke coincided with the onset of the prominent syllable in speech, and the gesture apex was produced before the end of the accented syllable.

Matthews, D., Behne, T., Lieven, E., & Tomasello, M. (2012). Origins of the human pointing gesture: a training study. *Developmental Science*, *15*(6), 817–29.

Matthews et al. (2012) investigated one hundred and two typically developing infants of 9, 10 and 11 months of age on the effects of experience of gesture training on the later use of pointing gestures. The procedure consisted of a training condition, which favoured pointing activities during daily home activities with the caregivers and a control condition which consisted of musical activities. Pointing measures of the infant and the caregiver were taken into account, as well as infants' gaze following abilities. The results revealed that infants who had better gaze following abilities perforemed better at pointing in later sessions. There were no significant effects of training condition on the infant's ability to produce communicative gestures, although caregivers' frequency of pointing gesture production was correlated with their infant's abilities to produce pointing gestures.

Liszkowski, U., Albrecht, K., Carpenter, M., & Tomasello, M. (2008). Infants' visual and auditory communication when a partner is or is not visually attending. *Infant Behavior & Development*, *31*(2), 157–67.

Liszkowski et al. (2008) tested the effects of adults' shared attention on infants' pointing productions in sixty-four 12- and 18-month-old infants. The authors measured the childrens' communicative responses (speech-only and pointing-only behaviors) during an experimental procedure consisting of a declarative pointing task with an unknown adult in one of the two experimental conditions; (a) available experimental condition in which the adult shared attention with the infant but not with the object, and (b) unavailable experimental condition in which the adult did not share attention with the infant or the object; and in comparison to the baseline condition in which the adult jointly shared attention between the infant and the object of reference. The results of the study revealed that infants pointed significantly more and produced more vocalizations during and after pointing in the available and unavailable condition compared to the baseline condition. Moreover, adults' social interaction patterns during the unavailable condition triggered less pointing behavior than during the available condition.

Silverman, L. B., Bennetto, L., Campana, E., & Tanenhaus, M. K. (2010). Speech-and-gesture integration in high functioning autism. *Cognition*, *115*(3), 380–93.

Silverman, Bennetto, Campana, & Tanenhaus (2010) tested nineteen adolescents with high functioning autism with a comprehension task with a visual world paradigm. They aimed at investigating the visual attention responses to gesture-speech integrated modalities in this population in comparison to their matching typical developing population (TP). The task assessed visual fixations during a video-base task of a person describing one of four shapes (e.g. curved line with a loop), using speech-and-gesture or speech only descriptions. Since gesture naturally precedes speech, earlier visual fixations to the target shape during speech-and-gesture compared to speech only trials were interpreted as immediate integration of audiovisual information. The results showed that TP benefited from the use of gestures, as they identified the image of the corresponding speech when it was accompanied by a gesture. Contrary to this, the high functioning autism population was slower to match the spoken word with the picture when the gesture was present.

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, Bibyk, Wagner & Speer (2014) investigated how English-speaking children from 6- to 11- years old (*N*=121) understand intonation contrast marking in relation to adults (*N*=38). The authors measured gaze fixation with an eye-tracking visual world paradigm to test the effect of pitch accent on referential solution. The materials consisted of a picture including a total of 18 drawings of animals: 6 animals (rabbit, lion, squirrel, turtle, monkey, and cat) each repeated 3 times, only differing in colour (purple, orange, green and pink). Referent resolution to the target image was faster when presented with a pitch prominence than without a prominence. Younger infants looked to the target item at the onset on the noun, and older children showed faster gaze fixation more similar to adult pattern, which started during the pronominal adjective.

8. GENERAL REFERENCES

Alibali, M.W., Heath, D.C. & Myers, H.J. (2001). Effects of visibility between speaker and listener on gesture production: some gestures are meant to be seen. *Journal of Memory and Language 44,* 169–188.

Armstrong, M., Esteve-Gibert, N. & Prieto, P. (2014). The acquisition of multimodal cues to disbelief. *Proceedings of the Speech Prosody 2014.*

Austin, J. L. (1962). *How to Do Things With Words*. Oxford University Press: Oxford, England.

Bates, E., Camaioni, L. & Volterra, V. (1975). The acquisition of performatives prior to speech. *Merrill-Palmer Quarterly, 21*, 205-224.

Bates, E., & Dick, F. (2002). Language, gesture, and the developing brain. *Developmental Psychobiology*, *40*(3), 293–310.

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,* 687-701.

Bebko, J. M., Weiss, J. a, Demark, J. L., & Gomez, P. (2006). Discrimination of temporal synchrony in intermodal events by children with autism and children with developmental disabilities without autism. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *47*(1), 88–98.

Behne, T., Liszkowski, U., Carpenter, M., & Tomasello, M. (2012). Twelve-month-olds' comprehension and production of pointing. *British Journal of Developmental Psychology, 30*, 359–375.

Biau, E., & Soto-Faraco, S. (2013). Beat gestures modulate auditory integration in speech perception. *Brain and Language*, *124*(2), 143–52.

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.

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*(01), 207–220.

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, 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.

Cartmill, E. A., Demir, Ö. E., & Goldin-meadow, S. (2011). Studying Gesture. In E. Hoff (Ed.), *The guide to research methods in child language* (pp. 208–225). Blackwell.

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.

Colonnesi, C., Stams, G.J., Koster, I. & Noom, M.J. (2010). The relation between pointing and language development: a meta-analysis. *Developmental Review, 30*, 352-366.

Chu, M., Meyer, A., Foulkes, L., & Kita, S. (2014). Individual differences in frequency and saliency of speech-accompanying gestures: the role of cognitive abilities and empathy. *Journal of Experimental Psychology. General*, *143*(2), 694–709.

Dahan, D., Tanenhaus, M. K. & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. Journal of Memory and Language 47, 292–314.

De Marchena, A., & Eigsti, I.-M. (2010). Conversational gestures in autism spectrum disorders: asynchrony but not decreased frequency. *Autism Research : Official Journal of the International Society for Autism Research*, *3*(6), 311–22.

de Villiers, J. (2007). The interface of language and Theory of Mind, *Lingua. 117*(11), 1858-1878.

Dick, A. S., Goldin-Meadow, S., Solodkin, A., & Small, S. L. (2012). Gesture in the developing brain. *Developmental Science*, *15*(2), 165–80.

D'Imperio, M., Elordieta, G., Frota, S., Prieto, P., Vigário, M., 2005. Intonational phrasing in Romance: the role of syntactic and prosodic structure. In: Frota, S. Viga S. Vigário, M., Freitas, M.J. (Eds.), *Prosodies*, Mouton de Gruyter, The Hague, pp. 59–98.

D'Odorico, L. (1991). Selective production of vocalization types in different communciation contexts. *Journal of Child Language*, *18*, 475–499.

Ejiri, K., & Masataka, N. (2001). Co-occurrence of preverbal vocal behavior and motor action in early infancy, *1*, 6–11.

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. (2013a). Prosody signals the emergence of intentional communication in the first year of life: evidence from Catalan-babbling infants. *Journal of Child Language*, *40*(5), 919–44.

Esteve-Gibert, N. & Prieto, P. (2013b). Prosodic structure shapes the temporal realization of intonation and manual gesture movements. *Journal of Speech, Language, and Hearing Research, 56* (3), 850-864.

Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. Developmental Psychology, 27, 209–221.

Goldin-Meadow, S. (2003). Hearing gesture: How our hands help us think. Cambridge, MA: The Belknap Press.

Goldin-Meadow, S., Cook, S. W., & Mitchell, Z. a. (2009). Gesturing gives children new ideas about math. *Psychological Science*, *20*(3), 267–72.

Gotham, K., Pickles, A. & Lord, C. (2012). Trajectories of autism severity in children using standardized ADOS scores. *Pediatrics, 130*, 1278–1284.

Gullberg, M., DeBot, K., & Volterra, V. (2008). Gestures and some key issues in the study of language development. *Gesture*, *8*(2), 149–179.

Grassmann, S., and M. Tomasello. 2007. Two-year-olds use primary sentence accent to learn new words. Journal of Child Language 34.677–87.

Greaves-Lord, K., Eussen, M.L., Verhulst, F.C., Minderaa, R.B., Mandy, W., Hudziak, J.J., Steenhuis, M.P., de Nijs, P.F., Hartman, C.A. (2012). Empirically Based Phenotypic Profiles of Children with Pervasive Developmental Disorders: Interpretation in the Light of the DSM-5. *Journal of Autism and Developmental Disorders, 43* (8), 1784-1797.

Gredebäck, G., Melinder, A., & Daum, M. (2010). The development and neural basis of pointing comprehension. Social Neuroscience, 5 (5–6), 441–450.

Grice, H.P. (1957). Meaning. *The Philosophical Review, 66*, 377-388.

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*, 773–778.

Grossman, R.B., Bernis, R., Skweker, D., & Targer-Flusberg, H. (2010). Lexical and affective prosody in children with high functioning autism. *Journal of Speech, Language, and Hearing Research, 53*, 778-793.

Grossman, R. B., & Tager-Flusberg, H. (2012). "Who said that?" Matching of low- and high-intensity emotional prosody to facial expressions by adolescents with ASD. *Journal of Autism and Developmental Disorders*, *42*(12), 2546–57.

Herold, D. S., Nygaard, L. C., & Namy, L. L. (2011). Say It Like You Mean It: Mothers' Use of Prosody to Convey Word Meaning. *Language and Speech*, *55*(3), 423–436.

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*(March), 74.

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.

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 and 6-year olds. *Journal of Memory and Language*, *66*(1), 265–284.

Iverson, J. M. (2010). Multimodality in infancy: vocal-motor and speech-gesture coordinations in typical and atypical development. *Enfance*, *3*, 257–274.

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), 367–371.

Iverson, J. M., & Goldin-Meadow, S. (1997). What's communication got to do with it? Gesture in children blind from birth. *Developmental Psychology*, *33*(3), 453–67.

Iverson, J.M. & Thelen, E. (1999). Hand, mouth and brain. The dynamic emergence of speech and gesture. *Journal of Consciousness Studies,6* (11-12), 19-40.

Iverson, J. M., & Wozniak, R. H. (2007). Variation in vocal-motor development in infant siblings of children with autism. *Journal of Autism and Developmental Disorders*, *37*(1), 158–70.

Kehoe, M. Atypical prosody. In Kehoe, M. (2013) *The Development of Prosody and Prosodic Structure*, Language and Linguistics, New York.páginas164-188.

Kelly, S. D., & Lee, A. L. (2012). When actions speak too much louder than words: Hand gestures disrupt word learning when phonetic demands are high. *Language and Cognitive Processes*, *27*(6), 793–807.

Kendon, Adam (2007). Some topics in gesture studies. In Anna Esposito, Maja Bratanic, Eric Keller, & Maria Marinaro (Eds.), Fundamentals of verbal and nonverbal communication and the biometric issue (pp. 3–19). Amsterdam: IOS 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). Memory and Language The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception, *57*, 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), 1207–19.

Laakso, M.-L., Poikkeus, a.-M., Katajamaki, J., & Lyytinen, P. (1999). Early intentional communication as a predictor of language development in young toddlers. *First Language*, *19*(56), 207–231.

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*(10), 1457–1471.

Liszkowski, U., Albrecht, K., Carpenter, M., & Tomasello, M. (2008). Infants' visual and auditory communication when a partner is or is not visually attending. *Infant Behavior & Development*, *31*(2), 157–67.

Liszkowski, U., Schäfer, M., Carpenter, M., and Tomasello, M. (2009) Prelinguistic infants, but not chimpanzees, communicate about absent entities. *Psychological Science, 20*, 654–660.

Loehr, D.P. (2012). Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory Phonology, 3* (1), 71–89.

López-Ornat, S. & Gallego, C. (2005*). Inventario del desarrollo MacArthur: versión española.* Madrid. TEA.

Lord, C., & Jones, R. M. (2013). New strategies and findings for behavioral interventions in autism spectrum disorders. *Annals of the New York Academy of Sciences*, *1304*, 70–6.

Lord C, Rutter M, DiLavore PC, Risi S. (2002). *Autism Diagnostic Observation Schedule ADOS*. Los Angeles, CA: Western Psychological Services;

Mampe, B., Friederici, A.D., Christophe, A. & Wermke, K. (2009). Newborns' cry melody is shaped by their native language. *Current Biology, 19*, 1994-1997.

Matthews, D., Behne, T., Lieven, E., & Tomasello, M. (2012). Origins of the human pointing gesture: a training study. *Developmental Science*, *15*(6), 817–29.

Masataka, N. (2003). From index-finger extension to index- finger pointing: ontogenesis of pointing in preverbal infants. In S. Kita (Ed.), Pointing: Where language, culture, and cognition meet (pp. 69–84). Mahwah, NJ: Erlbaum.

Masataka (1993). Effects of contingent and noncontingent maternal stimulation on the vocal behavior of three-to four-month-old Japanese infants. *Jounal of Child Language*

McCann, J., Peppé, S., 2003. Prosody in autistic spectrum disorders: a critical review. *International Journal of Language Communication Disorders, 38* (4), 325–350.

McLeod, S. (2009). Speech sound acquisition. In J. E. Bernthal, N. W. Bankson, & P. Flipsen (Eds.), *Articulation and phonological disorders: Speech sound disorders in children* (pp. 385–405). Boston: Pearson Education.

McNeill, D. (2012). *How Language Began. Gesture and speech in human evolution.* Cambridge University Press.

McNeill, D. (2005). *Gesture and Thought*. Chicago: University of Chicago Press.

McNeill, D. (1992). *Hand and Mind*. Chicago.The Chicago University Press.

McPartland, J.C., Reichow, B., Volkmar, F.R. (2012). Sensitivity and specificity of proposed DSM-5 diagnostic criteria for autism spectrum disorder. *Journal of American Academy of Child and Adolescent Psychiatry, 51*(4):368-83.

Miller, J. L., & Lossia, A. K. (2013). Prelinguistic infants' communicative system: Role of caregiver social feedback. *First Language*, *33*, 433–448.

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., & Belinchon, M. (2012). Gestural-vocal coordination. Longitudinal changes and predictive value on early lexical development. *Gesture 12*, 1, 16-39.

Oller, D.K., Buder, E.H., Ramsdell, H.L., Warlaumont, A.S., Chorna, L. & Bakeman, R. (2013). Functional flexibility of infant vocalization and the emergence of language. *PNAS*, 1-6.

Özçaliskan, S. & Goldin-Meadow, S. (2005). Gesture is at the cutting edge of early language development. *Cognition 96*, 101-113.

Peppé, S., McCann, J., Gibbon, F., O'Hare, A., & Rutherford, M. (2007). Receptive and expressive prosodic ability in children with high-functioning autism. *Journal of Speech, Language, and Hearing Research, 50*, 1015-1028.

Pizzuto, E., Capobianco, M., & Devescovi, A. (2005). Gestural–vocal deixis and representational skills in early language development. *Interaction Studies, 6*, 223–252.

Ravizza, S. (2003). Movement and lexical access: do noniconic gestures aid in retrieval? *Psychonomic Bulletin & Review*, *10*(3), 610–5.

Rowe, M.L. & Goldin-Meadow, S. (2009). Early gesture selectively predicts later language development. *Developmental Science 12*, 182-187.

Ruffman, T., Slade, L., & Crowe, E. (2002). The Relation between Children's and Mothers? Mental State Language and Theory-of-Mind Understanding. *Child Development*, *73*(3), 734–751.

Searle, J.R. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press: Cambridge, England.

Silverman, L. B., Bennetto, L., Campana, E., & Tanenhaus, M. K. (2010). Speech-and-gesture integration in high functioning autism. *Cognition*, *115*(3), 380–93.

Smith, E. G., & Bennetto, L. (2007). Audiovisual speech integration and lipreading in autism. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *48*(8), 813–21.

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.

Soltis, J. (2004). The signal functions of early infant crying. *The Behavioral and Brain Sciences*, *27*(4), 443–58; discussion 459–90.

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.

Stevenson, R. a, Siemann, J. K., Schneider, B. C., Eberly, H. E., Woynaroski, T. G., Camarata, S. M., & Wallace, M. T. (2014). Multisensory temporal integration in autism spectrum disorders. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *34*(3), 691–7.

Swerts, M., 2007. Contrast and accent in Dutch and Romanian. *Journal of Phonetics 35*, 380–397.

Tellier, M. (2008). The effect of gestures on second language memorisation by young children. *Gesture*, *8*(2), 219–235.

Tomasello, M. (2008). Origins of Human Communication. Cambridge, MA: MIT press.

Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A new look at infant pointing. *Child development78 (3)*:705-722.

Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development*, *57*(6), 1454–63.

Vihman, M.M. (1996). *Phonological development. The origins of language in the child*. Oxford: Blackwell.

Vihman, M.M., DePaolis, R.A. & Keren-Portnoy, T. (2008). A dynamic systems approach to babbling and words. In Bavin, L. (ed), *The Cambridge handbook of child language.* Cambridge.

Wagner, P., Malisz, Z., & Kopp, S. (2014). Gesture and speech in interaction: An overview. *Speech Communication*, *57*, 209–232.

Wang, L., & Chu, M. (2013). The role of beat gesture and pitch accent in semantic processing: an ERP study. *Neuropsychologia*, *51*(13), 2847–55.

Willems, R. M., & Hagoort, P. (2007). Neural evidence for the interplay between language, gesture, and action: a review. *Brain and Language*, *101*(3), 278–89.

Wu, Z. & Gros-Louis, J. (2014). Infants' prelinguistic communicative acts and maternal responses: Relations to linguistic development. *First Language, 34* (1), 72-90.