Preschoolers' pragmatic development: How prosody and gesture lend a helping hand

Iris Hübscher

TESI DOCTORAL UPF / 2018

DIRECTOR DE LA TESI

Dra. Pilar Prieto

DEPARTMENT DE TRADUCCIÓ I CIENCIES DEL LLENGUATGE

Universitat Pompeu Fabra Barcelona

Für mini Familie

Language is not a neutral medium that passes freely and easily into the private property of the speaker's intentions; it is populated – overpopulated– with the intentions of others.

Mikhail Bakhtin

Artist: Christine Hübscher

Acknowledgements

Writing a dissertation is not just a matter of getting the work done efficiently and with a good input-output ratio. For me it has been largely about finding my place in the environment of different research traditions and people doing that research. I feel very fortunate to have come across and made friends with a large number of kind, bright and encouraging people during my first years in the world of research. Even though I will try to limit myself, there is still a wide list of people that I would like to express my gratitude to in particular.

Supervisor

First and foremost, all my gratitude goes to my supervisor Dr. Pilar Prieto. This work would never have been possible without her encouragement and support. I have been extremely lucky to have her as my mentor and guide in the process of my predoctoral stage of research. Thank you so very much for all your help, for your enduring positivism and trust in this work and for always being there for me. Right from the start, she pushed and encouraged me to present my work at international conferences and provided me with an inspiring international research environment for doing the research that this dissertation reports. Also, forming part of Pilar's research group GrEP –Prosodic Studies Group, gave me the great opportunity to share my PhD years with colleagues working on similar topics, sharing endless meals, coffee breaks and also conference experiences together and gaining many new friends – this was enriching and invaluable both from an academic and personal point of view.

Previous academic supervisors (MA)

I am also very thankful to previous professors, Jonathan Culpeper, Gila Schauer and Karin Tusting at Lancaster University during my MA. They opened up the world of research to me, especially to the field of pragmatics/interpersonal pragmatics, and gave me valuable guidance and motivated me to continue doing research.

Department, UPF and funding

Additional thanks are owed to the Department of Translation and Language Sciences at the Universitat Pompeu Fabra for their support during my predoctoral stage, as well as for funding my three-month research stay at the Ohio State University in the United States, and for the financial support that allowed me to attend many international conferences. A special thanks also goes to Àlex Alsina, Cristina Gelpí and Toni Badia. Additionally, many thanks to the members of the secretary of the department, especially to Susi Bolós Ginés, Núria Abad, Yolanda Bejarano, Rafa Ordóñez Arcas and Vanessa Alonso. A special thank you to L'Albert Servitje i Riera and his team at the Voluntariat Lingüístic for introducing me in such a delightful way to the Catalan hotspots and to the culture more generally – visca Catalunya!

This PhD was funded by a FPI grant 2014-2018 from the Spanish Ministry of Science and Innovation (BES-2013-065019).

First research stay at Ohio State University

I am very thankful to Professor Laura Wagner from Ohio State University where I carried out my first research stay. She welcomed me with open arms and granted me access to the language lab at the Science Museum COSI in Columbus. These places are of high academic prestige, and I was very fortunate to have been able to spend three months there. Professor Wagner provided me with a great amount of inspiration, skills and knowledge in the field of developmental psychology and linguistics that will also be reflected in the pending projects. Next to being a great mentor academically, she also looked after my private well-being by giving me access to the sports facilities on campus and taking me out for meals on several occasions.

I was also lucky enough to attend many lab meetings held by the psycholinguistic research group and the language acquisition research group (Lacqueys). Thank you to all their members, and special thanks go to Mary Beckman, Kiwako Ito and Shari Speer. Shari also allowed me to attend her graduate class on psycholinguistics where I was able to learn a great deal. Similarly, many thanks go to Scott Swenters for granting me access to his graduate level course on pragmatics.

Thank you to the people I met at the psychology, linguistics, Spanish/Portuguese and German departments at the university. Special thanks go to Katharina Aigner and her boyfriend, Ian Nyre, as well as to Jennifer Zhang. Finally, thanks go to my great housemate Ruth Staveley and her lovely daughters in Arlington, Columbus. Thank you all for this special time.

2nd research stay at the Max Planck Institute for Psycholinguistics

I am also very thankful to Judith Holler and Asli Ozyürek for hosting me during my stay at the renowned Max Planck Institute for Psycholinguistics in Nijmegen, the Netherlands. My stay there was funded by a generous travel stipend from the FPI scholarship. Both Judith and Asli granted me access to their Multimodal Language and Cognition lab, and Judith Holler also introduced me to the Language and Cognition Lab under the supervision of Steven Levinson. Also I feel grateful for having been able to attend the class on 'multimodal context of language: gesture and sign' as well as various lectures, research group meetings and workshops held at the Institute. The Max Planck Institute is a truly remarkable place, and I was able to learn a lot from the world-class research being carried out there. Special thanks also go Judith Holler for having several meetings with me during my stay in which I was able to learn a lot from her. Furthermore, many thanks also go to Marc Swerts (from the University of Tilburg), Sotaro Kita (from the University of Warwick) and Marisa Casillas who I was able to meet during my stay there. The stay in Nijmegen would not have been the same without Lieke Van Maastricht who introduced me to the Dutch culture. I would like to also thank all the members of the two groups at the Institute, with some special thanks to Mathias Bartel for sharing his office space with me and also to Gabriella Garrido, Paul Hömcke and Hannah Lutzenberger, in particular, for the shared lunches.

Feedback to work

Many thanks to everyone who provided feedback during the different stages of this PhD. Special thanks go to Lluís Payrató, Aurora Bell and

Miquel Serra for giving me very valuable feedback after the first year, during the PhD project defense.

Participants and help during the studies

I am extremely grateful to the preschools and parents that collaborated in this research: Escola Sant Martí (Arenys de Munt), Escola La Farigola del Clot (Barcelona), Escola Pública Dr. Estalella Graells (Vilafranca) and Escola Masmitjà (Girona). Many thanks also go to all the adult participants from Universitat Pompeu Fabra who volunteered to take part in the studies or provided me with stimulus materials. Additionally, I would like to thank the adult participants at the linguistics and psychology departments at Ohio State University for providing me with adult data or stimulus materials. I am particularly indebted to Joan Borràs-Comes for his invaluable help with the statistics. Additonally, Judith Llanes Coromina and Anna Massanas helped with part of the Catalan data collection and Faith Stagge aided in carrying out the experiment at COSI in Columbus, Ohio in the United States.

Friends and colleagues

My heartfelt thanks to my friends and colleagues who have supported me throughout the years of my PhD. Special thanks to my "Zippi" from the GLiF, Giorgia Zorzi, for always having been at my side during the past four years; it would not have been the same without you! Many thanks to all the current and previous members of the Prosodic Studies Group, Meghan Armstrong, Florence Baills, Joan Borràs-Comes, Núria Esteve-Gibert, Santiago González Fuente, Alfonso Igualada, Evi Kiagia, Olga Kushch, Cristina Sánchez-Conde, Rafèu Sichel-Bazin, Judith Llanes Coromina, Marússia Pronina, Patrick Rohrer, Ingrid Vila Giménez, Laura Vincze, Yuan Zhang and the visitors in our group, Alice Cravotta and Martina Garufi for sharing the PhD experience with me and, at the same time, becoming esteemed friends with whom I shared many good and intense moments. Also thanks go to the other PhD colleagues in other research groups at the Universitat Pompeu Fabra, with a special thanks to Kata Wohlmut, Toni Bassanages, Sara Cañas and Alexandra Navarrete González (from the GLiF), and some other PhD students from the UPF, Lorena Recalde, Alba Mila and my other Swiss compatriot, Martina Maier who need some special mention. Thanks to all of you for the great times. Furthermore, I would like to extend my thanks to some wonderful people I met several times at conferences and I was able to share great times with at different occasions. In particular, thanks go to Katharina Zahner, Clara Huttenlauch, Yulia Nigmatulina and Brigitta Keij. Some special mention also goes to my housemates here in Barcelona, Teresa Ivars and Minerva Montoya. Finally, I would like to thank all my friends, some of whom have known me for the better part of my life and others who I met in different places in this world – life would be a lot less rich without you. I will not start to list you all - you will know it for yourself when reading this.

Family

I would like to express my special gratitude to my family. My heartfelt thanks to my truly remarkable parents, Christine Hübscher and Walter Seiler, for their constant support, understanding and encouragement. I cannot thank them enough for having kindled my curiosity and eagerness to discover the world and for having given me strong wings and, at the same time, received me again and again with much love in the old abode.

My family would not be complete without my beloved brother, Robin Hübscher, and his great partner, Carola Schabert. And I am truly thankful to all the other members of my Swiss family as well as to my tico/nica family "Pineda-Mendoza" for always receiving me with open arms. Finally, I am especially thankful to my wonderful partner, Yuergen Ramon Pineda Mendoza, who despite the distance has been at my side before and during the time of my PhD. He has surrounded me with his 'pura vida' spirit, which helped me keep things in perspective and has taught me to value things differently.

Many thanks to anyone else with a connection to this thesis or who feels close to this work.

Abstract

While previous research on language development has highlighted the facilitating role of gesture (mainly deictic gesture) in children's early access to meaning, little is known about the possible facilitating role of prosody and also of co-speech gestures in children's pragmatic development in the preschool years. Previous work on developmental pragmatics has focused on the acquisition of morphosyntactic and lexical forms, and there is a need to adopt a more integrative multimodal perspective. The overarching aim of this thesis is to experimentally investigate, through a set of cross-sectional studies with preschool children (3- to 5-year-olds), whether prosodic and gestural cues serve as pragmatic precursors in the development of two key pragmatic phenomena, namely knowledge state (i.e., commitment to the status of information) and politeness (e.g., broadly speaking, adjusting one's language).

The first study uses a forced-choice paradigm to investigate preschool children's understanding of another speaker's knowledge state, presented in audio only, video only and audio-visual formats. Results show that overall children perform significantly better in detecting a speaker's uncertainty when they have gestural cues present, and, importantly, the younger children were significantly better in detecting a speaker's uncertainty when listening to a speaker's intonation contour as compared to uncertainty expressed through a lexical epistemic adverb. The second study analyzes children's (and adults') multimodal expression of their knowledge state through an object guessing game by evaluating their production of prosodic, gestural and lexical cues and additionally assessing

their self-assessment of their knowledge state. Results show that while preschool children are not yet able to self-report on their knowledge state, in the younger group, children encode their knowledge state through prosodic and gestural means only. And only in the older age group do children start to use a few lexical markers to signal their uncertainty. The third study uses a forced-choice paradigm to assess children's understanding of a speaker's politeness presented in audio only, video only and audio-visual formats. Results show that 3-year-old children detect a speaker's polite stance significantly more through facial cues and intonation, highlighting children's early ability to extract meaning from intonation when lexical cues are kept the same. Finally, the fourth study explores children's multimodal production of politeness in semi-spontaneous requests in different sociopragmatic situations. Results show that regardless of the age group, children marked politeness through fine-grained gestural and prosodic means when producing requests to an adult with high social distance as compared to a classmate with low social distance, and this also depends on the cost of the request.

Altogether the results of these studies demonstrate that children's early pragmatic comprehension and expressive abilities rely strongly on prosodic and gestural marking, developing well before children master lexical and morphosyntactic markers. More specifically, the four studies presented in this thesis bring forth evidence that both prosody and co-speech gestures play a precursor role in children's pragmatic development of knowledge state and politeness. Ultimately, the thesis highlights the importance of approaching the study of children's pragmatic development from a multimodal perspective.

Resum

Tot i que les investigacions prèvies sobre el desenvolupament del llenguatge han destacat el paper facilitador del gest (principalment, del gest díctic o d'assenyalament) en l'accés primerenc dels nens al significat del llenguatge, se sap poc sobre els beneficis de la prosòdia i dels gestos de la parla en el desenvolupament pragmàtic dels nens en els anys preescolars. La recerca que s'ha fet sobre el desenvolupament de la pragmàtica s'ha centrat en l'adquisició de formes morfosintàctiques i lèxiques, però és important adoptar una perspectiva multimodal més integradora. L'objectiu general d'aquesta tesi consisteix a investigar experimentalment, a través de quatre estudis transversals amb nens d'edat preescolar (de 3 a 5 anys d'edat), si els senyals prosòdics i gestuals actuen com a precursors pragmàtics en el desenvolupament de dos fenòmens pragmàtics, concretament el posicionament epistèmic (i.e., el grau de certesa del parlant sobre la informació expressada) i la cortesia.

El primer estudi utilitza un paradigma d'elecció forçada per investigar la comprensió dels nens en edat preescolar sobre el posicionament epistèmic d'un altre parlant, presentat en tres condicions (només en àudio, només en vídeo i en format audiovisual). Els resultats mostren que els nens detecten millor la incertesa d'un parlant quan estan exposats a indicis gestuals. Crucialment, els nens de tres anys detecten millor la la incertesa d'un parlant quan la incertesa s'expressa a través de la prosòdia que quan s'expressa a través d'adverbis epistèmics. El segon estudi analitza l'expressió multimodal de l'estat epistèmic dels nens (i dels adults) a través d'un joc d'endevinar objectes. L'anàlisi dels senyals prosòdics, gestuals i

lèxics mostren que, mentre els nens d'edat preescolar encara no poden valorar el seu posicionament epistèmic en canvi sí que codifiquen el seu grau de certesa a través de senyals prosòdics i gestuals. Només en el grup d'edat més gran els nens comencen a utilitzar alguns marcadors lèxics. El tercer estudi utilitza un paradigma d'elecció forçada per avaluar la comprensió dels nens sobre la cortesia d'un parlant, presentada també en tres condicions (només en àudio, només en vídeo i en format audiovisual). Els resultats mostren que els nens de 3 anys poden detectar una actitud educada a través de senyals facials i entonatius quan es mantenen constants les marques lèxiques. Els resultats mostren la capacitat inicial dels nens per extreure el significat de l'entonació, de forma paral·lela al que passa amb la gestualitat. Finalment, el quart estudi explora l'expressió multimodal de la cortesia en peticions emprades en diferents situacions sociopragmàtiques. Els resultats mostren que, independentment del grup d'edat, els nens marquen la cortesia a través de senyals gestuals i prosòdics detallats quan es fan peticions a un adult amb més distància social, en comparació amb un company amb qui presenten una distància social menor. L'ús d'aquestes marques també depèn del cost de la petició.

En resum, els resultats dels quatre estudis de la tesi demostren que les habilitats primerenques de comprensió i d'expressió pragmàtica es basen en el marcatge prosòdic i gestual. i que aquestes es desenvolupen molt abans que els nens controlin els marcadors lèxics i morfosintàctics. Així, els quatre estudis presentats en aquesta tesi posen de manifest que tant la prosòdia com els gestos de la parla actuen com a precursors en el desenvolupament pragmàtic dels llenguatge. En definitiva, la tesi posa de relleu la importància d’aproximar-se a l'estudi del desenvolupament pragmàtic infantil des d'una perspectiva multimodal.

Resumen

A pesar de que las investigaciones previas sobre el desarrollo del lenguaje hayan destacado el papel facilitador del gesto (principalmente, del gesto deíctico o de señalamiento) en el acceso temprano de los niños al significado del lenguaje, poco se sabe aún sobre los beneficios de la prosodia y de los gestos del habla en el desarrollo pragmático de los niños en etapa preescolar. Hasta ahora, las investigaciones sobre el desarrollo de la pragmática se han centrado en la adquisición de formas morfosintácticas y léxicas, haciéndose necesario adoptar una perspectiva multimodal más integradora. El objetivo general de esta tesis consiste en investigar experimentalmente —a través de cuatro estudios transversales con niños de edad preescolar (de 3 a 5 años de edad)— si las señales prosódicas y gestuales actúan como precursores en el desarrollo de dos fenómenos pragmáticos, concretamente del posicionamiento epistémico (i.e., del grado de certeza del hablante sobre la información expresada) y de la cortesía.

El primer estudio utiliza un paradigma de elección forzada para investigar cómo los niños en edad preescolar comprenden el posicionamiento epistémico de otro hablante, presentado los estímulos en tres condiciones (solo en audio, solo en vídeo y en formato audiovisual). Los resultados muestran que los niños detectan mejor la incertidumbre de un hablante cuando están expuestos a indicios gestuales. Crucialmente, los resultados también muestran que los niños de tres años detectan mejor la incertidumbre de un hablante cuando esta viene expresada a través de la prosodia que cuando lo hace a través de adverbios epistémicos. El segundo estudio analiza la expresión multimodal del estado epistémico de los niños

(y de los adultos) a través de un juego de adivinar objetos. El análisis de las señales prosódicas, gestuales y léxicas muestra que, aunque los niños en edad preescolar todavía no pueden valorar su posicionamiento epistémico, los del grupo de menor edad ya son capaces de codificar su grado de certeza a través de señales prosódicas y gestuales, mientras que solo los niños del grupo de mayor edad empiezan a utilizar algunos marcadores léxicos. El tercer estudio utiliza también un paradigma de elección forzada para evaluar la comprensión de los niños sobre la cortesía de un hablante, presentado también los estímulos en tres condiciones (solo en audio, solo en vídeo y en formato audiovisual). Los resultados muestran que los niños de 3 años son capaces de detectar una actitud cortés significativamente mejor a través de señales faciales y entonativas, destacando además la capacidad temprana de los niños para extraer el significado de la entonación cuando las marcas léxicas se mantienen constantes. Finalmente, el cuarto estudio investiga la expresión multimodal de la cortesía en peticiones realizadas en diferentes situaciones socio-pragmáticas. Los resultados muestran, por un lado, que, independientemente del grupo de edad, los niños marcan la cortesía a través de señales gestuales y prosódicas diferenciadas en función de si las peticiones van dirigidas hacia un adulto con el que mantienen una mayor distancia social, o bien hacia un compañero con quien la distancia social es menor. Además, los resultados también muestran que el uso de estas marcas también depende del coste de la petición.

En resumen, los resultados de los cuatro estudios de esta tesis demuestran que las habilidades tempranas de la comprensión y expresión de significados pragmáticos se basan en el marcaje prosódico y gestual, y que, además, estas habilidades se desarrollan mucho antes de que los niños dominen los marcadores léxicos y morfosintácticos. De este modo, los cuatro estudios presentados en esta tesis ponen de manifiesto que tanto la prosodia como los gestos del habla actúan como precursores en el desarrollo pragmático del lenguaje, subrayando así la necesidad de aproximarse al estudio del desarrollo pragmático infantil desde una perspectiva multimodal.

Table of Contents

1

Chapter 1: General Introduction

1.1. Overview and motivation

Early on in development, infants display the desire to interact and communicate as well as to understand and be understood by their interlocutors. They show an early appreciation of the give and take of interaction long before they begin to speak (Rochat, Querido, & Striani, 1999; Trevarthen, 1979). By 9 months, infants are involved in complex triadic interactions between ego, alter and an object in attention (Striano & Tomasello, 2001), e.g., baby, parent and ball. Since these skills arise before children use their first words, the development of pragmatics[1] has been increasingly regarded as a foundation of language acquisition (Matthews, 2014). Work on early developmental pragmatics has shown that infants rely strongly on multimodal aspects of communication, such as prosody, manual gestures, eye gaze, and also on information from the physical, social and affective (emotional) context before using lexical or verbal information (Bates, Camaioni, & Volterra, 1975; Esteve-Gibert, Prieto, & Liszkowski, 2017b; Kita, 2003; McNeill, 1992; Morgenstern, 2014a). In the last few years, arguments have increasingly been brought forth that communication is intrinsically multimodal and that language is expressed through various interlinked channels (see Bavelas & Chovil, 2000; Clark, 1996; Kendon, 2004; Levinson & Holler, 2014; McNeill, 1992; Mondada, 2016). However, this field is still in its early stages of development. Thus, studying human communication and more specifically pragmatic development calls for a multimodal perspective on how children acquire

[1] Pragmatic development is here understood as how children learn to use language in social contexts.

the building blocks for becoming successful members of a language community.

In order to understand and acquire a language, a child must understand what their interlocutor is referring to and what they intend to say. In this context, the use of gesture and prosody has been found to play a pivotal role in early first-language acquisition, working as bootstrapping or scaffolding devices by helping children to identify and acquire the building blocks that language is made up of. Yet, so far beyond the early stages of language development, most research has focused on how children acquire morphosyntactic elements of language. As mentioned, language consists of much more than solely a combination of morphemes and words put together into grammatical constructions. Basically, it is not only *what* is said that matters but also important *how* it is said. Studies looking at adults' pragmatic processing and production of language have found that prosody and gesture are both important carriers of different aspects of meaning such as uncertainty, speech acts and politeness (see e.g. Prieto, 2015). Studies within the field of audio-visual prosody (see Krahmer & Swerts, 2009 for an overview) have helped advance our knowledge on multimodal and prosodic signaling of different pragmatic meanings. Audio-visual cues have been shown to boost the perception of prominence, question intonation, of affective functions as well as social ones. Thus, while it has been established that prosody and gesture are essential in the production and perception of pragmatic meaning in adults, it is not clear when and how very young children start to actively use prosodic and gestural cues to understand another person's communicative intention such as the marking of (un)certainty or a person's politeness. Given the fact that prosody and gesture are important carriers of meaning and have been shown to play in important role in inference making, it is fundamental to understand this developmental process and assess the role prosody and gesture play in the acquisition of pragmatics. In addition to examining children's use and understanding of prosodic and gestural marking of pragmatic meaning, this thesis aims to test whether these cues play a precursor role in children's development of pragmatic skills over the preschool years as they do in infants' early language acquisition.

This thesis follows a sociopragmatic and usage-based approach to language acquisition (see Tomasello, 1992, 2000; Tomasello, 2001). In this account, language usage is consistent with the way we process language and represent it. Crucially, communicative interactions, and later conversations, are the primary way we learn to use language. Despite the great variation in how language is used in conversation, one common trait of human languages is that we learn languages in interaction with other people. Research has shown that communicative interactions are the basis of linguistic development and are acknowledged as conversation. Thus, as children develop, they do not just acquire language, but very importantly, they learn how to use it with other interlocutors in specific communicative contexts. What the usage-based theory proposes is that children approach the process of language acquisition around one year of age and are outfitted with two cognitive skills, namely intention-reading (functional dimension) and pattern-finding (grammatical dimension). While intention-reading, including skills of joint-attention, is the central assumption of the social-pragmatic approach to language acquisition, pattern-finding includes children's going beyond the individual utterances they hear around them in order to be able to create abstract linguistic schemas/ constructions.

Parting from this background, the four studies presented in this thesis will be of interest, on the one hand, to developmental psychologists interested in sociopragmatics and to prosody and gesture researchers who are interested in children's intonational and gestural development, on the other hand. Understanding how gesture and prosody influence early understanding and production of pragmatic attitudes/stances, and whether they can be considered developmental precursors of pragmatics, may also have important practical, as well as theoretical, implications. Next to advancing knowledge on how prosody, gesture and the lexicon interact during the acquisition of pragmatics, therapies and teaching strategies can ideally be informed by a multimodal approach to a child's pragmatic development which is directly linked to a child's social integration and development in their speech community.

The following sections provide an overview of research on language development, focusing on the role of gesture and prosody. First the role of gesture as a precursor in language acquisition is discussed (1.2.). Then an overview of the bootstrapping role of prosody in early language development is provided (1.3.), followed by a discussion of the few studies that have dealt with the role of prosody and gesture in preschool children's pragmatic development (1.4.). Next the previous literature focusing on children's development of knowledge state meaning, reviewing studies on both comprehension and production will be outlined (1.5.). Section 1.6 summarizes the previous literature on children's development of politeness meaning. Finally, we conclude the chapter by presenting the general objectives of the thesis, the research questions and the hypotheses (1.7).

1.2. Gesture in the early development of multimodal communication

1.2.1. An analysis of gesture

If we observe two people in a face-to-face interaction, we will quickly notice that they move their hands, arms and other body parts while they are communicating. These body movements, which are produced while talking, have become known in the literature as co-speech gestures. They are movements that are coordinated with the ongoing speech and are intended to communicate meaning (Kendon, 1972; McNeill, 1998). Regarding form, the most common coding system applied to co-speech gestures is the one by McNeill (1992), who proposed four types of gestures: iconics (also referred to as representational gestures, i.e., gestures that represent the features of the referent in a transparent way, e.g., rapid hand movements up and down in order to mimic the movement of chopping onions), deictics (hand and index finger gestures pointing to objects and locations), emblems (quotable gestures or conventionalized body movements, e.g., the "OK" gesture), beat gestures (rhythmic hand movements reflecting emphatic functions in discourse) and metaphorics (more abstract gestures that have the potential to engage an active cross-domain mapping; that is, the cognitive process of understanding something in terms of something else (Cienki & Müller, 2008)). As Cienki & Müller

(2008) argue, metaphoric gestures are like iconic gestures in that they are pictorial, but the pictorial content presents an abstract idea rather than a concrete object or event.

Studies on co-speech gestures have mainly focused on manual gestures and set aside other articulators such as facial gestures. Yet conversational facial gestures are as important as hand gestures in communication, and they share many of the functions and characteristics of hand gestures (Bavelas, Gerwing, & Healing, 2014). As has been noted by Bavelas et al. (2014), facial gestures are clearly noticeable and abundant in face-to-face communication. Kendon (2004) mentioned the following when writing about kinesic-medium[2] in sign language, “facial gestures, such as eyebrow movements or positionings, movements of the mouth, head postures and sustainments and changes in gaze direction” (p. 310). Facial gestures are thus conversational and include any configuration or movement of the face or also of the head, which is synchronized with speech in timing and meaning, making them similar to manual gestures. Most often facial expressions have been equaled with emotional expression and have mostly been regarded as non-linguistic. Literature exploring the connection between the face and emotions has usually used the term facial expressions (Ekman, Friesen, & Hager, 2002) while studies investigating cues on the face as related to speech employ the term facial gestures (see Bavelas et al. (2014) for an overview). In this thesis, we will integrate the analysis of manual gestures as well as facial gestures when referring to speakers’ use of their body to signal knowledge states or politeness alongside their speech.

---

[2] Communication through body movement

Furthermore, relevant for the present investigation, there is a small body of research that has recently emerged showing how gestural, especially facial gestures, as well as prosodic means are used to index different pragmatic functions in language such as knowledge state-related meanings and also politeness-related meanings among others. Typical examples of conventionalized gestures that perform pragmatic functions are, for example, gestures which mark an utterance as a question and that hold the turn (e.g. Duncan, 1972; Kendon, 1995, 2002, 2004; Mondada, 2006; Müller, 2004). These kinds of gestures have also been referred to as pragmatic gestures (see Kendon, 2004) as they help the speaker express their intentions, mental states and feelings, for example, co-speech gestures that have been shown to encode degrees of speaker knowledge in languages such as Catalan (Borràs-Comes, Roseano, Vanrell, Chen, & Prieto, 2011; Roseano, González, Borràs-Comes, & Prieto, 2016) and Dutch (Swerts & Krahmer, 2005; Visser, Krahmer, & Swerts, 2014). These studies found that different articulators are involved in the marking of ignorance and speaker uncertainty such as the speaker's head, facial cues, such as eyebrows and mouth, and the hands and shoulders. This thesis will integrate the assessment of manual gestures together with facial gestures as interlocutors often use them simultaneously (see Figure 1 for an example of the expression of unknowingness).

Figure 1. Still image of the body and facial gestures used (i.e., raised eyebrows and shoulders accompanied by palm-up gesture) to express unknowingness.

Furthermore, a few studies have investigated adults' multimodal expression of politeness in English (Tree & Manusov, 1998), Korean (Brown & Winter, in press) and Catalan (Nadeu & Prieto, 2011). Similarly to the studies on a speaker's expression of knowledge state, these studies have found that adults use an array of gestural and other bodily signals to mark their politeness in communication, using mitigation gestures such as raised eyebrows, a tilted head and a (slight) smile.

Having outlined the most relevant definitions of gesture for the present thesis, in the following section we will review the literature that has focused on the precursor role of gesture in children's language development.

1.2.2. Precursor role of gesture in language development

Gesture has been found to play a remarkable role in children's early language and communicative development. The ability to interact intentionally with an interlocutor arises very early but also importantly before children use their first words. According to Tomasello (1995), two abilities indicate the start of infants' intentional communication, namely the ability to distinguish means and goals in the infant's own action productions and in the other's action productions and their ability to engage in joint attention frames. And this is where gesture comes into play: deictic gestures have been found to serve as a child's primary means to communicate intentionally and to represent people and objects in the environment (e.g. Tomasello, Carpenter, & Liszkowski, 2007, among many others). This is why gesture has been regarded as a cornerstone for the emergence of language development, as deictic gestures are often clear manifestations of the fact that an infant is becoming an intentional agent (Bates et al., 1975; Kita, 2003; McNeill, 1992). Building on Bates et al.'s proposal, Tomasello et al. (2007) suggested that there are three possible social intentions behind a pointing gesture: (1) a declarative informative intention, providing the interlocutor with information that might be interesting to them; (2) a declarative expressive intention, sharing attention about an object and (3) an imperative intention, that is, requesting an object. That is, producing a pointing gesture implies that the infant has a clear pragmatic and intentional aim while redirecting the interlocutor's attention.

Children have been shown to communicate first by using gesture and only later by using their first words (e.g. Acredolo & Goodwyn, 1985; Bates, 1976; Iverson, Capirci, & Caselli, 1994; Özçalişkan & Goldin-Meadow, 2005). Children, for example, typically first point to a cat before they are able to name it (Bates, Benigni, Bretherton, Camaioni, & Volterra, 1979); they first use iconic gestures to convey information about attributes of an object or actions (e.g., flapping arms to illustrate flying) (Bates et al., 1979) and also first nod their head to convey *yes* and shake it to symbolize *no* (Beaupoil-Hourdel, Morgenstern, & Boutet, 2015; Benazzo & Morgenstern, 2014; Guidetti, 2005).

Next to studies that have indicated that gesture precedes lexical acquisition, it has also been demonstrated that the use of gesture can predict syntactic development (e.g. Capirci, Iverson, Pizzuto, & Volterra, 1996; Iverson & Goldin-Meadow, 2005; Özçalişkan & Goldin-Meadow, 2005). Studies (e.g. Capirci et al., 1996) have shown that there is a relationship between the onset of the production of a deictic gesture and the later appearance of words, and similarly at 16-20 month infants show a relationship between a single-gesture and word combinations and two-word utterances. Capirci et al. (1996) found that the use of gesture and gesture-word combinations is an important facet of the transition to the two-word stage. Furthermore, Iverson and Goldin-Meadow (2005) examined how gesture production corresponds to lexical and syntactic development in the early stages of language development. They observed the gesture and speech development of three children between 10 and 24 months at monthly intervals. They found that referents that were observed initially in children's gestural repertoire appeared some months later in their spoken vocabulary. For example, if children pointed to a ball at 14 months, the word *ball* was highly likely to appear in their vocabulary at 18 months. Also, they found that children who produced first gesture-plus-word combinations that conveyed two elements in a proposition (e.g., they point at a chair and say ‘sit’) were also those who produced first two-word combinations (‘sit chair’). Based on these findings, the authors claim that gesture plays an important role in early language development and can be used to predict it as well. In this case, gestures are helping children to acquire new meanings in their repertoire of communication and to lay the foundation for the appearance of the lexical meanings in speech. Similarly, Özçaliskan and Goldin-Meadow (2005) investigated whether the production of supplementary gesture-speech combinations at 14, 18 and 22 months foreshadowed oncoming changes in children’s speech and served as a forerunner of linguistic advances. Forty children were videotaped while they were interacting with their caregivers. During these interactions, all the meaningful sounds and communicative gestures were transcribed. Unsurprisingly, the results of the study demonstrated that speech and gesture clearly changed with age. More interestingly, however, was that the types of gesture-speech combinations changed with time and indicated changes in speech. Children who produced one word and used a gesture to supplement their speech, such as in "eat" + (point at cookie), and also combinations with an argument and a predicate, like in “touch” + (pointing at a dog) before beginning to produce the same constructions in speech. These results show a strong predictive effect of children’s pointing gestures on later linguistic outcomes. Furthermore, Rowe and Goldin-Meadow (2009) found that children who pointed to more different objects at 14 months had a greater vocabulary at 42 months than children who produced fewer pointing gestures.

Most of the research on children's development of gesture has focused on the role of pointing gestures in infants' early language acquisition. In order to find out whether other types of gestures, such as iconic gestures (e.g., flapping arms to represent a bird flying or holding cupped hands together to depict the roundness of a ball), play a similar role in children's language development, Özçalışkan, Gentner, and Goldin-Meadow (2013) videotaped 40 North American children from the age of 14 to 34 months at home together with their parents. They collected a total of 540 minutes of observation for each child. All the communicative words and gestures that the children produced were transcribed and they then analyzed the children's use of iconic gestures to depict actions. Contrary to pointing gestures, they found that iconic gestures do not predict and precede children's first verbs, but rather that they produced the gestures 6 months after they produced their first verbs. They established that children generally produced their first verbs around 18 months of age, and they produced their first iconic gesture only at around 25 months. The authors suggested several plausible explanations to why children produce few iconic gestures early in development. First of all, they postulate that they might be difficult to produce. Second, it might be related to the parental input, which might be lower for iconic gestures than for pointing gestures. Third, iconic gestures might involve a higher cognitive load compared to pointing gestures. Lastly, iconics might be more difficult to acquire in general since, in language development, object concepts are usually learned before relational concepts. Hopefully, future research can shed more light on this issue.

Moving beyond the early infancy period, one notices that gesture has been investigated much more sparsely in children's later language development, when they are acquiring more complex language tasks. Most studies have centered on the first stages of language development, focusing on how gestures help children with the symbolic power of language and also with gesture-word combinations. However, gestural communication does not disappear with the emergence of word productions. Gestures do not only stay functional over development, but they also diversify in both form and function.

In general, studies on the use of gestures by older children can be divided into two groups: On the one hand, there are those reporting a simultaneous development of gesture and speech. For example, Colletta et al. (2015) investigated multimodal narrative development in 5- and 10-year-old French, American and Italian children's accounts of a wordless cartoon and highlighted that gesture and speech go hand in hand. Specifically, children's gestures related to the narrative organization develop during the same time as children's narratives become more complex. Furthermore, Sekine and Kita (2015) found a co-development of speech and co-speech gestures in elicited narratives in 3-, 5- and 9-year-old children as well as in adults, in terms of sentence and discourse level.

On the other hand, there are those studies that report that gestures continue to have a scaffolding role in language development. For example, reinforcing gestures have been found to work as an effective scaffold for children to comprehend complex spoken messages (McNeill, 2000) since they guide comprehension towards the meaning of the spoken language.

More particularly, research on child development has shown that iconic, metaphoric and deictic gestures help preschool or school-age children to understand the conversation or discourse they are listening to (Clark, Hutcheson, & Buren, 1974; Goldin-Meadow & Wagner, 2005; McNeil, Alibali, & Evans, 2000). Also studies on cognitively demanding tasks have shown that gestures often times display implicit conceptual knowledge which children are not yet able to express verbally. This has been shown to be the case when children completed Piagetian conservation tasks (e.g. Broaders, Cook, Mitchell, & Goldin-Meadow, 2007; Church & Goldin-Meadow, 1986) or also when explaining answers to mathematical problems (e.g. Perry, Breckinridge Church, & Goldin-Meadow, 1988). Furthermore, Kidd and Holler (2009) investigated 3-5-year-old children's use of gestures in order to resolve lexical ambiguity. Children were told three different stories that each contained two homonym senses, such as in bank, and then had to retell the stories to another experimenter. The results showed that at 3 years of age children still struggled to disambiguate the two senses, using pointing gestures while trying to do so. Then at 4 years old, children tried to disambiguate more often by using iconic gestures; finally at 5 years old, children used less iconic gestures, and, in contrast to the 3-year-olds, they were also able to disambiguate the senses verbally. These outlined studies highlight that gesture continues to play an important role in language development beyond the early infancy period. Hostetter (2011) conducted a quantitative meta-analysis of 63 studies on the general communicative role of co-speech gestures (iconic, metaphoric or deictic) in both adults and children. The meta-analysis showed that gestures facilitate comprehension in listeners, and this beneficial role of gesture has also been found in adults. It has been established that iconic and metaphorical gestures have a positive effect on the recall of information (Riseborough, 1981; So, Chen Hui, & L., 2012; Thompson, 1995) as well as on comprehension processes (Cocks, Morgan, & Kita, 2011; Hostetter, 2011; Kelly, Barr, Church, & Lynch, 1999).

On the whole, few studies have systematically assessed the emergence of different pragmatic meanings in gesture. One important exception is the analysis of agreement and refusal/negation gestures (such as nods and side-to-side head shakes), which start to be produced at around one year and one month, and similarly to deictic gestures, which are produced before children learn to express negation and affirmation through corresponding words (e.g. Beaupoil-Hourdel et al., 2015; Guidetti, 2005) (see section 1.4.1 for a more detailed description on the development of negation gestures). However, in general, very little is known of young children's acquisition and perception of other gestural, facial and body cues reflecting pragmatic meanings such as surprise, uncertainty, incredulity amongst others. In addition, much remains to be investigated with respect to their possible precursory role in the developmental process to which this thesis aims to contribute.

1.3. Prosody in the early development of multimodal communication

1.3.1. An analysis of prosody

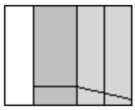
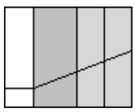
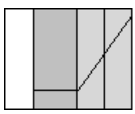
Two key terms of the present thesis are related concepts, namely intonation and prosody, both suprasegmental features of language. The term *intonation* (also called speech melody) will be used in the narrow sense of the word, namely as “the variations in the pitch of the voice” (Ladefoged, 2006 p. 23). The term *prosody* will refer to all sounds of language that supersede its segments, i.e., fundamental frequency (perceived as *pitch*), intensity (perceived as loudness), duration (perceived as length) as well as voice quality.

In the last few decades, a specific framework of intonational analysis has been developed analyze intonational phonology across languages named the Autosegmetal-Metrical (AM) model, which was initially proposed by Pierrehumbert (1980) and later developed by Beckman and Pierrehumbert (1986) among many others. The main focus of this framework is the analysis of language's intonational phonologies. Individual intonational contours are separated into two main tonal categories; edge/boundary tones and pitch accents. The most prominent pitch accent within the prosodic phrase is the nuclear configuration, which typically consists of a pitch accent followed by a boundary tone. In the AM framework, H (high) and L (low) tones are considered to be the main atoms of analysis. Based on this framework, a set of Tones and Break Indices Annotation Conventions (ToBI) have been developed for a variety of typologically different languages (Jun, 2010, 2014). ToBI was originally developed for American

English, and it was later adapted to describe many other languages such as German, Japanese, Italian, Portuguese, Spanish, Catalan (see Jun, 2010 for an overview; 2014). In this thesis, the Catalan proposal (Cat_ToBI, Prieto, 2014) will be used in Chapters 2, 3, and 5, and the American English proposal will be used (AME_ToBI, Beckman, Hirschberg, & Shattuck-Hufnagel, 2010) in Chapter 4.

Intonation is an aspect of the human communication system that serves a variety of communicative functions, ranging from grammatical, semantic-pragmatic ones to the marking of speech acts, knowledge states, emotional states as well as politeness (see Prieto, 2015). Intonation has been shown to encode pragmatic information, such as degrees of a speaker's knowledge state in Catalan (Borràs-Comes et al., 2011; Crespo Sendra, Kaland, Swerts, & Prieto, 2013; Dijkstra, Krahmer, & Swerts, 2006). Table 1 shows three possible nuclear pitch configurations in Catalan to express speaker (un)certainty. Similarly, a body of research has emerged showing how prosodic and gestural means are used to index politeness-related stances (Brown & Winter, in press; Brown, Winter, Idemaru, & Grawunder, 2014; Brown, Winter, Idemaru, & Grawunder, 2015; Grawunder, Oertel, & Schwarze, 2014; Hübscher, Borràs-Comes, & Prieto, 2017a; Kaori, Winter, & Brown, 2016).

Table 1. Schematic, phonetic implementation, Cat_ToBI Labels for possible nuclear configurations to express (un)certainty in Catalan

Schematic	Phonetic Implementation	Cat_ToBI Label	Possible pragmatic meaning
	Low plateau during the last accented syllable and then a fall or a low plateau.	L* L%	Certainty
	Low plateau during the last accented syllable followed by a rise to a high level.	L+H* H%	Uncertainty
	A rise during the last accented syllable that continues into the following syllable(s).	L* H%	Uncertainty

Although the main focus of this thesis is on the use of intonational patterns in development, for which the analysis of pitch is the default, other prosodic correlates will be taken into account. For example, other prosodic cues have been found to be important in adult speech in order to convey knowledge state such as hesitations and elongations (Krahmer & Swerts, 2005). Furthermore, politeness-related meanings are also expressed through a set of prosodic features such as duration, intensity and voice quality (refers to the 'characteristic auditory coloring' of a speaker's voice (Laver, 1980) (see Brown and Prieto (2017), for an overview). So far, however, very little is known about children's use of such cues in the understanding and production and of pragmatic meanings. While Chapter 2 and Chapter 4 mainly focus on children's inference of pragmatic meaning through intonation, Chapter 3 and Chapter 5 lay out a more holistic prosodic analysis is carried out.

1.3.2. Precursor role of prosody in language development

Prosody has been shown to play a key bootstrapping role in early language acquisition. From birth, babies are particularly sensitive to pitch and other prosodic differences in speech. Research indicates that infants are able to use prosodic cues consistently some months after birth, not only in terms of sensitiveness to prosodic acoustic features (Bosch & Sebastián-Gallés, 2001; Fernald & Kuhl, 1987; Jusczyk, Friederic, Wessels, Svenkerud, & Jusczyk, 1993; Nazzi, Bertoncini, & Mehler, 1998; Ramus, 2002), but also in terms of the production of early cries (Mampe, Friederici, Christophe, & Wermke, 2009; Wermke et al., 2016).

Infants prefer the prosodic properties of infant-directed speech to those of adult- directed speech (e.g. Fernald, 1985; Fernald & Kuhl, 1987). In trying to assess which prosodic features children prefer, Fernald and Kuhl (1987) investigated whether this preference was driven by the fundamental frequency (f0) patterns (higher pitch range values), duration patterns (longer time values) or amplitude patterns (higher intensity variability). The results indicated that children were only showing preference to the higher pitch range values typically used in infant directed speech, but not to duration and amplitude features. Other studies demonstrate that at the age of 3 months, infants have the ability to distinguish between two languages belonging to two different rhythmic classes (e.g. Fernald, 1985; Mehler et al., 1988; Nazzi et al., 1998). Next to this, slightly older infants, at 4-5 months, can distinguish languages within the same rhythmic category if these languages have distinct segmental cues (Bosch & Sebastián-Gallés, 2001; Nazzi, Jusczyk, & John, 2000). At 6 months, infants have been shown to be capable of using prosodic information regarding clausal units in their processing of continuous infant-directed speech (Nazzi et al., 2000). Furthermore, they display early sensitivity to the position of prosodic prominence at 6-9 months of age. For example, they prefer the stress patterns of their home- or environment language (Höhle, Bijeljac-Babic, Herold, Weissborn, & Nazzi, 2009; Jusczyk et al., 1993; Pons & Bosch, 2010). All this evidence points to infants' early sensitivity to prosody in language development.

Special focus has been shed on the bootstrapping role of prosody for early speech segmentation into phonemes, syllables and syntactic structure (Cavalho, Dautriche, Millotte, & Christophe, 2018). Gleitman and Wanner (1982) original proposal was that acoustic cues in speech may help infants to detect syntactic boundaries before lexical knowledge is available (Gleitman, Gleitman, Landau, & Wanner, 1988; Morgan & Newport, 1981; Pinker, 1987). This proposal has become known as the "prosodic bootstrapping hypothesis". There are three important elements highlighted in this hypothesis: (1) syntax has to be reliably correlated with acoustic properties; (2) infants are sensitive to the acoustic properties of speech and (3) infants use these cues when they are processing speech.

Prosody has also been found to be an important cue for children's word learning (see Thorson, 2018 for an overview). Infant directed speech, which is characterized by slower speech rates and exaggerated pitch excursions, shows vowels and consonant contrasts to be more pronounced and helps children in their construction of phoneme inventories (Cristià, McGuire, Seidl, & Francis, 2011; Kuhl, Williams, & Meltzoff, 1991; Werker et al., 2007). Furthermore, it has been found that the use of caregiver's slower speaking rate and more emphasized vowel properties help 21-month-olds to better learn and remember new words (Ma, Golinkoff, Houston, & Hirsh-Pasek, 2011). Thus, the exaggerated prosodic properties of infant directed speech make the relationship between prosodic form and function which are uniquely salient to children (Saint-Georges et al., 2013).

When looking at the production of early prosodic patterns, similar configurations can be observed. Studies have shown that human neonates have access to prosodic information in the uterus as infant cries reflect the prosodic patterns of a child's ambient language (Mampe et al., 2009). Mampe et al. investigated the intonation and intensity contours of the first cries of French and German newborns (mean age of 3 days) to see if they differed as a function of the language they heard prenatally. Their results showed that all infants cried following a rising-falling arch-shape, as expected, while the peak of the arches in the melody and intensity contours varied significantly depending on the language they had been exposed to: French infants' cries showed the peak of the melody and intensity contour towards the end of the arch, thus displaying a rising contour. German infants' cries produced the peak of the contours at the beginning of the arch, thus displaying a falling contour. In a more recent study, Wermke et al. (2016) evaluated the variability of the fundamental frequency (f0) in spontaneous cries produced by infants with either a tonal (Cameroon) or a non-tonal (German) ambient language. Results showed that cries produced by Cameroon infants were characterized by a higher f0 variability compared to the cries of the German infants. Mean f0 of the cries were comparably similar. In this respect, the findings in the present study are comparable to the one outlined above on French and German neonates.

The production of stress patterns is developed a bit later on (e.g. Davis, MacNeilage, Matyear, & Powell, 2000; Snow, 2006; Vihman, Nakai, & DePaolis, 2006). Davis et al. (2000) analyzed the acoustic parameters of stress in babbling infants and found that, although they were able to use fundamental frequency, intensity and duration to signal prominence, at that age they did not produce the acoustic cues of stress in an adult-like manner. In fact, it has been found that the specific rhythmic patterns of the ambient language are not developed until later (see Post & Payne, 2018 for an overview). Payne, Post, Astruc, Prieto, and Vanrell (2011) compared the rhythmic patterns of 2-, 3-, and 4-year-old children in Catalan, Spanish and English and found that there is some evidence that at age 2 children use some rhythmic cues that are consistent with the ambient language (particularly interval duration variability), with results improving significantly across the subsequent ages.

All in all, prosody, similarly to gesture, has been found to provide children with cues that stimulate their grammatical development and also word acquisition. In general, the literature on the development of prosody has heavily focused on how infants use their sensitivity to prosody as a tool to discern rhythmic and stress patterns and to help with the task of lexical and syntactic segmentation of the language that surrounds them. However, prosody also plays a very important role in shaping the pragmatic meaning of utterances. Intonation in particular is used by the speaker to express a variety of speaker attitudes towards an object or event, to differentiate among sentence types, to structure information, to organize and maintain communication and interactions and also to convey both epistemic and evidential information (see Barth-Weingarten, Dehé, & Wichmann, 2009 for a review of the prosody-pragmatics interface; Hirschberg, 2017; Prieto, 2015; Prieto & Rigau, 2011). Unfortunately, relatively little is known about how children use prosodic features in their pragmatic development and whether there exists some kind of pragmatic bootstrapping role of prosody to be assessed. The exceptions to this are those studies that are related to the development of emotions, early intentionality and speech act distinctions, which will be reviewed in the upcoming paragraphs.

Early on in infancy, perceptive prosodic skills related to emotion develop. Infants prefer to listen to their mother's voice (Decasper & Fifer, 1980). At 5 months, infants can distinguish between two different emotions based on their interlocutor's facial expressions and the acoustic properties of their speech (Fernald, 1993; Grossmann, Striano, & Friederic, 2006; Vaillant-Molina, Bahrick, & Flom, 2013). Furthermore, results show that, by using continuous speech, young infants rely on the congruency between auditory emotions (happy, angry) and the accompanying facial gestures (Soken & Pick, 1992; Walker-Andrews, 1997). Looking at production of emotional prosody, it has been found that 4-5-month-old infants express emotions, such as sadness or enjoyment through facial expressions signaling emotions, and that at 12 months, they similarly express fear, pain, surprise (Sullivan & Lewis, 2003b). Similar tendencies have been observed for the use of vocal cues to signal emotion (Lindová, Špinka, & Nováková, 2015; Oller et al., 2013).

Furthermore, research suggests that young infants are able to use prosody for early intentional communication well before the one-word stage. Papaeliou and Trevarthen (2006) studied the early production of prosodic cues and how they relate to intentionality. These authors investigated various acoustic and prosodic parameters of 10-month-old English-acquiring children in one of two distinct situations: when playing by themselves or with their mother. The results of their experiment showed that intentional vocalizations, which were directed at their mother, were shorter and displayed higher pitch values than non-intentional vocalizations. In a later study, Esteve-Gibert and Prieto (2013) analyzed a total of 2,701 vocalizations from a longitudinal corpus of four Catalan-babbling infants aged 0-;7 to 0-;11 months[3]. These results also showed that infants use different prosodic patterns to distinguish communicative from investigative vocalizations and to express intentionality. Specifically, they found that requests and expressions of discontent displayed a wider pitch range and longer duration patterns than responses or statements. These results support the hypothesis that babbling children successfully use a set of prosodic patterns to signal intentional speech. Furthermore, Sakkalou and Gattis (2012) examined whether 14- and 18-month-old infants would imitate more intentional or accidental actions purely on the basis of

[3] This is a typical notation to represent age, e.g. in this case 1 year and 3 months

prosody. In their first experiment, intonational actions were accompanied by the word “There” with high amplitude and long duration, and accidental actions were accompanied by the word “Whoops!” with low amplitude and short durations. In their second experiment, they applied the same methodology but removed the lexical information. In both experiments, the results showed that infants imitated the intentional actions more often than the accentual ones, with an age-effect indicating that older infants performed better than younger infants when lexical cues were removed. Thus, the results indicated that infants understand the pragmatic value of prosodic since they are able to link them to intentionality solely basing their imitative behavior on the prosody that accompanies the action. Moreover, Esteve-Gibert et al. (2017b), who focused on the perception of speech act information expressed through prosody, showed that infants use prosodic features (alongside pointing gestures) to infer the motivation for a particular speech act uttered by a communicative partner. In two different experiments, they tested whether 12-month-old infants are able to link specific prosodic patterns (and also hand gestures) to the speaker’s underlying intention of their speech act (declarative, expressive or requestive). While in the first experiment they tested the role of prosody together with lexical and gestural cues, in the second they tested only the role of prosody and gestures. The results of the first experiment showed that infants were able to use the prosody and gestures to understand the intentions behind an attention-directing act while the results of the second experiment illustrated that when there was no lexical information (i.e., this cue was controlled across experimental conditions), infants were able to infer a speaker’s intentions purely on the basis of prosody and manual gestures.

Unfortunately, apart from early intentionality, emotions and speech act distinctions, little is known about children's developing perception and production of prosodic cues reflecting pragmatic meanings and the eventual precursor role these cues have in children's pragmatic development, a gap which this thesis intends to address.

1.4. Development of the prosody-gesture-pragmatics interface. Gesture and prosody as sister systems in pragmatic development?

This thesis starts from the initial hypothesis that prosodic and gesture patterns are sister systems that go hand in hand in pragmatic development (see e.g. Snow, 2017). Therefore, these two modalities will be studied together in the present work. Importantly, prosodic patterns, together with gestural cues, have been shown to be strongly correlated markers of sociopragmatic meaning in adult speech. Adult studies have found that gestural patterns are often as important as prosodic patterns in order to detect pragmatic meaning (see e.g. Borràs-Comes et al., 2011; Goldin-Meadow & Butcher, 2003; Holler & Wilkin, 2009; Krahmer & Swerts, 2005; Prieto, Borràs-Comes, Tubau, & Espinal, 2013). Research over the last few decades has shown that speech and gesture constitute a single communicative system that is tightly integrated semantically, pragmatically and temporally (Kendon, 1980; Levinson & Holler, 2014; McNeill, 1992). Rochet-Capellan, Laboissière, Galván, and Schwartz (2008) identified that there is a temporal relation between jaw and arm movements, due to the synchronous occurrence of pitch accent and the stroke of the gesture (both in trochees and in iambs). Similarly, Loehr (2007) detected a rhythmic relationship between the movements of the hands as well as head and voice, where each articulator peaks synchronously with the other articulators. In the following section, children's understanding and signaling of pragmatic meaning through gesture and prosody is summarized. It will become clear that most research has been carried out with an exclusive focus on either prosody or gesture (with a few exceptions).

1.4.1. Children's ability to understand and signal pragmatic meaning through gesture and prosody

Looking at children's early communicative development as described in sections 1.2 and 1.3 above, it has been demonstrated that infants at 12 months are able to infer and convey intentions, specifically speech act intentions on the basis of prosody and gesture (Esteve-Gibert et al., 2017b; Sakkalou & Gattis, 2012). It is also during this early period that most literature has focused on examining the relevance that gesture and prosody play in children's language development. However, much less is known about the development of co-speech gestures and prosody conveying pragmatic intent during the preschool years, with some exceptions, which will be outlined below. As previously mentioned, most research has been carried out with a focus on either gestural or prosodic patterns, with few studies considering both patterns jointly.

Regarding gesture, there has been some work on the gestural encoding of agreement and disagreement. For example, Guidetti (2005) observed 30

French children aged 1;4, 2;0, and 3;0 at home during an interaction with their mother and found that at 16 months, French children predominantly used gesture alone to convey agreement and refusal messages while interacting with their mothers (72% of the messages). Later, two- and three-year-old children's use of verbal-only agreement and refusal messages outnumbered both isolated head gestures and gesture-speech combinations. In order to obtain a more detailed understanding of children's multimodal pathway into negation, Beaupoil-Hourdel et al. (2015) used longitudinal data of an English monolingual girl interacting with her mother. They recorded the child every month at her home for one hour between 10 months and 4 years. Their analysis revealed a specific pathway with five distinct periods in the acquisition of negation. During period 1 (0;10-1;1), the child solely used non-symbolic, action-based expressions of negation; in period 2 (1;02-1;04), the child moved from non-symbolic to symbolic/conventional gestures; in period 3 (1;05-2;00), she combined gestures with one or two words; in period 4 (2;06-3;00), as the child's speech becomes more complex, fewer gestures were used; finally in period 5 (3;06-4;02), more gestures were used again, but as co-speech gestures accompanying longer and more complex negative utterances. The authors suggest that children undergo a change from embodied negation and then later move on to symbolic negation, with gesture and speech being completely integrated into complex multimodal productions. Furthermore, Dodane, Beaupoil, Del Ré, Boutet, and Morgenstern (2014) recognized the role that prosody can play alongside gesture in children's entry into negation. They analyzed 96 multimodal productions containing the word "non" (no) produced in isolation and on strings of reduplicated *non*'s ("no's") in longitudinal recordings of a monolingual French girl, who they recorded every month for one hour between the age of 1:02 and 2;09 while taking part in spontaneous interactions with her parents (Morgenstern & Parisse, 2012). They coded prosodic properties (direction of the intonation contour, accent range, register, duration, intensity) and nonverbal behavior (hand gestures, joint attention expressed through eye gaze, body movement and facial expression) and also analyzed directional and temporal synchronization patterns comparing gesture and prosody. Prosodically, the first vocal 'no' appeared around 14 months, both as duplication and with exaggerated prosody. Then between 22 and 25 months, mainly rise-fall contours were produced while, from 26-28 months, productions were characterized by flat or falling intonation contours and reduced syllable duration. Together with the multimodal strategies, the child moved from an exaggerated pattern of prosody and body cues to an adult-like intonation contour used to express negation and less frequent use of bodily clues.

Apart from studies on gestures of negation, there have been some studies on children's use of beat gestures, which are considered to serve an emphatic function in discourse. While different types of gestures appear very early in infancy, as outlined in section 1.2., beat gestures tend to appear comparably late. Even though a few studies have reported an earlier appearance of beat gestures around 3 years of age (Mayberry & Nicoladis, 2000; Nicoladis, Mayberry, & Genesee, 1999), most of the studies report a later appearance around 6 years of age (Blake, Myszczyszyn, Jokel, & Bebiroglu, 2008; Colletta, Pellenq, & Guidetti, 2010). Since beats are typically considered as serving as discourse highlighters, they have been studied in relation to children's narrative abilities. Research has found that producing and perceiving beat gestures has a beneficial effect on children's recalling and comprehension of narratives (Llanes-Coromina, Vilà-Giménez, Borràs-Comes, Kushch, & P., 2018, in press). Finally, similarly in relation to narratives, Stites and Özçalışkan (2017) found that children introduce a new character through combinations of an iconic gesture and speech, displaying character viewpoint in gesture before being able to do so in speech (nominal phrases and then pronouns). Therefore, the literature outlined illustrates that gestural cues encoding pragmatic meaning also seem to appear before lexical/verbal encoding of similar meaning in the case of negation and also character viewpoint in narrative. Furthermore, in the case of beats, they seem to have a beneficial effect on children's language understanding and memory recall. However, so far very little is known about the developmental pathway of gestural cues conveying other sociopragmatic meanings such as knowledge state and politeness.

When looking at the role of prosody, and intonation in particular, as a carrier of pragmatic intent, mixed results have been reported. On the one hand, some studies have shown that children have difficulties infering meaning based on intonation and do so relatively late. More specifically, a number of researchers have argued that children do not rely on prosodic information at all when inferring a speaker's pragmatic intention, emotional state or affect (see e.g. Aguert, Laval, Le Bigot, & Bernicot, 2010; Cruttenden, 1985; Morton & Trehub, 2001; Quam & Swingley, 2012; Wells, Peppé, & Goulandris, 2004). Additionally, some studies have found that children can interpret an intonation contour signaling contrast only after the age of 6 (Ito, Jincho, Minai, Yamane, & Mazuka, 2012; Speer & Ito, 2009), and studies on children's irony detection found that children begin to detect certain aspects of ironic intent at 5 years of age (e.g. Milosky

& Ford, 1997), which they do by means of contextual and prosodic cues. However, there are a few recent studies which show that certain pragmatic meanings encoded through intonation can be in place earlier. For example, Armstrong, Esteve-Gibert, Hübscher, Igualada, and Prieto (accepted) assessed 3-5-years-old children's understanding of incredulity when encoded by intonation and facial gestural cues. By presenting them with either audio-only, visual-only or audio-visual cues, they showed that overall children performed significantly better when they had both intonation and gestural cues present, but interestingly children performed equally well when they had both prosody and facial gestural cues present. Similarly, González-Fuente (2017) investigated whether prosodic and gestural cues to emotion facilitated 5-, 8- and 11-year-old children's detection of irony. An irony detection task was presented to three groups. Children were audio-visually presented with six ironic context-utterance pairs produced with prosodic and gestural cues conveying three different type of emotions: one strongly mismatching negative emotion, one slightly mismatching negative emotion and one matching positive emotion. The results showed that when the audio-visual cues to emotion were strongly mismatched, this led to significantly higher irony detection rates in the three age groups. Furthermore, very recently, Kurumada and Clark (2017) have found evidence for earlier sensitivity to intonation contour to interpret changes in encoding contrastive focus. They presented 4-year-old children with utterances containing contrastive prosody, such as *It looks like a ZEBRA (and it is one)* vs. *It LOOKS like a zebra (but it isn't one)*, giving them two pictures, one serving as the target picture and another one as a distractor. They found that children succeed in deriving the conversational implicature of *It LOOKS like a Zebra*, when they are able to access the semantically stronger alternative, that it is a zebra. Thus, by taking into account the alternative utterance, their interpretation of contrast-marking intonation seems bootstrapped.

Studies on children's production of pragmatic uses of prosody and intonation, in particular in the preschool years, are limited, and there has been a primary focus on children's intonational development. Intonation, as mentioned further above, is used to convey a speaker's belief, marking (im)politeness and desires (Armstrong & Prieto, 2015; Escandell-Vidal, 1998; Gunlogson, 2003; Ward & Hirschberg, 1988). A set of studies has shown how children develop an early intonational grammar during their second year of life. Children's early production of the inventory of pitch accents and boundary tones has been studied through longitudinal data analysis and prominently in Ibero-Romance languages. Focusing on Central Catalan, Prieto and Vanrell (2007) showed that by age 2;00, children produce a wide array of language-specific pitch accents and boundary tones, and they showed great adequacy tones-to-text alignment. Furthermore, they reported that Catalan children's emerging intonation is not synchronous with their grammatical development in this language. They analyzed four children and found that they produced a great variety of language-specific pitch accents and boundary tone combinations long before they produced two-word utterances. They concluded that the development of an intonational grammar occurs before grammatical development takes off. Similar findings were made for European Portuguese, wherein children's intonational development occurred five months before the onset of the two-word stage (Frota, Matos, Cruz, & Vigário, 2016). Further studies on Dutch and European Portuguese intonational patterns have found that intonational development is correlated with an increase in vocabulary size (Chen & Fikkert, 2007; Frota et al., 2016). Similarly, Prieto, Estrella, Thorson, and Vanrell (2012b) investigated the development of intonation patterns in four Catalan-speaking and two Spanish-speaking children between 0;11 and 2;4 and found that there was a relationship between lexical and intonational development as the emergence of an intonational grammar was related to the onset of speech and also the existence of a small lexicon. Taken together, these studies for languages like Catalan, Spanish and Portuguese provide evidence that the main intonational categories of the ambient language has been acquired by the age of 2 when this language belongs to the Romance family.

Literature on intonational development in Germanic languages has placed more focus on children's skills in mapping information status to pitch categories and not on the acquisition of pitch category inventories (Chen, 2010, 2011; De Ruiter, 2009, 2014; Grünloh, Lieven, & Tomasello, 2015; Ito et al., 2012; Romøren & Chen, 2015; Thorson, 2015). In general, much less is known in relation to the intonational categories that children use to convey specific pragmatic meanings. Intonational meaning can go far beyond the typical sentence-type distinction as, for example, between an interrogative and a declarative sentence. To our knowledge, Armstrong's (2012) work is one of the few exceptions, which has addressed the role of intonation in belief state development. Armstrong (2012) investigated children's ability to produce intonation contours associated with polar (yes-no) questions, which by nature contain an epistemic gradient (Enfield, Brown, & de Ruiter, 2012). For instance, when a speaker poses a polar question (e.g., *Do you like chocolate)* to an interlocutor, the speaker can simultaneously use prosody and gesture to express varying degrees of certainty regarding the propositional content of the questions (e.g., higher degree of commitment vs. a lower degree of commitment). In many languages, specific intonational categories are used to convey belief about a propositional content, and the epistemic meaning is thus grammaticalized through intonation in polar questions. In her longitudinal corpus data, Armstrong (2015) found that even though intonational contours that contain various layers of meanings (which she referred to as multidimensional meaning) are present in child-directed speech (CDS), they are a rare occurrence in child speech. For example, in cases when the intonation does not mark an utterance as a question and additionally also conveys specific information about the speaker's mental state, this might complicate children's production of those intonation contours of multidimensional nature. Armstrong (2016) looked into the production of mental state intonation in the speech of two Puerto Rican toddlers from 1;7 – 3;6. In the speech of caretakers, she found that the general question-marking contour was used most (in 93% of all polar questions), followed by the disbelief contour (5% of all polar questions) and the belief presence contour (2%). In contrast, in the child data, she found only two instances of each belief-presence contour, and the disbelief contour was never produced felicitously within that age range. The belief presence contour appeared later in the corpus: one child produced only one belief contour at 2;8 and the other child at 3;0, around the age when lexical mental state language is emerging as Armstrong pointed out (Shatz, Wellman, & Silber, 1983).

In sum, as lexical and socio-cognitive skills develop further, children continue gaining mastery in the use of prosodic and gestural features. The review outlined above shows that studies on intonational and gestural development have clearly advanced in recent years, but also that there is a noticeable gap in comprehensive assessments of children's pragmatic development during the preschool years. In this thesis, we will focus on different utterance level meanings that can be expressed through both prosody and gesture. One area that proves to be ideal to investigate is children's pragmatic development of knowledge state and politeness during their preschool years. These two pragmatic skills have often been studied separately. However, it is appropriate to study them together in this thesis since both knowledge state and politeness markers are devices language users employ to position themselves (take a stance) in interactions and are crucial skills that children need to acquire to become competent members of a speech community. While there have been different approaches to stance, in this thesis definitions given by Biber and Finegan (1989) will be followed. Biber and Finegan (1989, 93) use the term stance to describe "the lexical and grammatical expression of attitudes, feelings, judgments, or commitment concerning the propositional content of messages". They then propose a superordinate term for the meaning that language users can communicate in addition to the propositional content. In the following section, the most relevant literature concerning both knowledge state and politeness will be summarized, focusing on both adults and children.

1.5. Children's understanding and signaling of knowledge state

Many of the world's languages signal knowledge state, which are also referred to as epistemic meanings through a varied set of linguistic markers (Aikhenvald, 2004; Chafe, 1986; de Haan, 2001b). Languages use certainty and uncertainty markers to indicate how certain the speaker is about the information they are conveying.

Traditionally, linguists have focused on languages that encode (un)certainty by means of specific morphemes, such as in Japanese and Turkish, or through lexical marking, e.g., by means of modal particles as well as mental state verbs such as *I think*. Even though it has been recognized that stance is not only expressed by means of words, but also very importantly multimodally. There are only a few studies on adult production of knowledge state that take a more integrative approach.

Studies in the audiovisual prosody approach have indicated that prosodic and gestural patterns, such as longer durations and higher pitch, eye squinting or shoulder shrugging, can act as conveyors of speaker uncertainty (Borràs-Comes et al., 2011; Dijkstra et al., 2006; Swerts & Krahmer, 2005). Swerts & Krahmer (2005) provided insights into the topic by showing that speakers' feeling of knowing (epistemicity) is cued by various visual and verbal properties. Their results showed that human observers can distinguish responses particularly well when the stimuli are presented to them in a bimodal way (audiovisual). Relevant for this thesis are those studies that examined children's comprehension and production of uncertainty; this will be dealt with in the following section.

1.5.1. Children's comprehension of a speaker's knowledge state

In a learning environment, it is vital to know whether a speaker is reliable or not since it is better to learn from a reliable source than from an unreliable one. Results on the selective learning literature show that by three years of age children are quite skillful at interpreting signs of reliability or ignorance in adult speech (Koenig & Harris, 2005) and that this can have an effect on the likelihood that children remember the object labels used by those adults (Koenig & Harris, 2005; Sabbagh & Baldwin, 2001). Infants as young as 14 to 18 months understand the basics of the ignorant status of the speaker, i.e., what another person does and does not know. They also understand that another person's knowledge may differ from their own knowledge, depending on previous experience. Carpenter, Akhtar, and Tomasello (1998) investigated whether a child can make use of other people's perspectives when learning new words. In their experiment, two-year-old children played continuously with an experimenter and a parent with three toys. The parent then left the room and a new toy was brought to the child. The child and the experimenter played with this toy for the same amount of time as with the other three toys. When the parent returned, the toys were arranged in a row on a shelf and then the parent looked at all four objects and said, *Oh, a gazzer! Wow, a gazzer! Look at the gazzer!* Even though children played the same amount of time with all four toys, they inferred that the adult was referring to the new toy that the parent saw for the first time. Furthermore, research has shown that young infants are sensitive to important details about the epistemic state of the speaker such as ignorance. Liszkowski, Carpenter, and Tomasello (2008) analyzed infants' understanding of epistemic states of other people, focusing on knowledgeable and ignorant partners. They examined whether 12-month-old infants use pointing gestures appropriately in order to provide uninformed people with information. Their findings showed that infants pointed more often to an object whose location the adult did not know (i.e., had not seen it fall down) and needed help to find the object than in a situation where the adult saw the object fall down and thus did not need help finding it. In essence, these studies show that infants have an initial understanding of the epistemic state of the speaker (knowledgeable vs. ignorant partner) based on contextual evidence, something that has important consequences for language interactions.

In sum, the literature on children's general pragmatic and social-cognitive abilities shows that at an early age, children have sophisticated abilities, which they can use to understand other people's epistemic state. However, it is only later that children are able to access a speaker's knowledge state through lexical epistemic stance markers. Moore, Bryant, and Furrow (1989) tested 3- to 8-year-old children in an experimental setting in which children had to find a hidden object. The object was in one of two boxes, and the children listened to verbal cues from two different puppets telling them about the place where the object was hidden. The utterances each contained a marker with a different degree of certainty, signposting a different box as the place of the hidden object such as *I know it's in the red box* vs. *I think it's in the blue box*. The results showed that while 4-year-olds were able to distinguish different degrees of speaker certainty, 3-year-olds were not. Moore, Pure, and Furrow (1990) carried out the same experiment but with modal expressions such as *It must be in the blue box* and *It might be in the blue box*. The results were the same as in the previous study, thus indicating that the understanding of modal expressions, such as *might*, strongly correlates with the understanding of the mental verbs such as *think*. Furthermore, Matsui, Yamamoto, and McCagg (2006) investigated children's linguistically-encoded speaker knowledge states in Japanese. 3- to 6-year-old children participated in hidden object tasks in which they had to make decisions based on two conflicting utterances, which were each marked with an expression of a different degree of speaker certainty and evidentiality. They chose stimuli sentences including both particles and verbs encoding epistemic information (such as the certainty particle *yo* and the uncertainty particle *kana*). The main findings were that children comprehended speaker knowledge states better when they were conveyed by particles (as early as 3 years of age) rather than verbs.

Surprisingly enough, there is only one study that has included intonation as a cue to knowledge state. Moore, Harris, and Patriquin (1993) compared children's (3 to 6 years old) comprehension of mental state lexicon to their comprehension of mental state prosody. Children had to listen to contrasting pairs of statements by two puppets and guess the location of a hidden object. Each statement pair either differed with respect to the mental state verbs, e.g., *know* vs. *think* or *think* vs. *guess*, or with respect to terminal pitch contour, e.g., falling or rising. While 3-year-olds were not able to use either lexicon or prosody to detect where the object was, 4-year-olds started to do so significantly better in the *know* vs. *think* and falling vs. rising pitch contrast conditions. Furthermore, the *think* vs. *guess* condition was much harder even for the 5-year-old children compared to the *know* vs. *think* condition. In a follow-up experiment with 3- to 5-year-old children, the conditions were presented as either matched or mismatched. While in the matched condition, lexical items of certainty went together with falling intonation and lexical items of uncertainty were matched with rising intonation while in the mismatched condition the opposite was applied. While 4-year-olds performed significantly above chance when *know* vs. *think* was matched with the corresponding prosodic cue, they did not show any significant difference between matched vs. mismatched trials. This time 5-year-olds performed much better on the *think* vs. *guess* distinction in the matching condition. Interestingly, 5-year-old children performed much worse in the mismatched condition, showing a certain awareness of prosodic and lexical integration when speaker (un)certainty is expressed. The authors suggested that prosodic and lexical cues to speaker certainty start to be used around the same time by children. Yet, the authors propose that lexical cues to a speaker's belief state initially seem to be more dominant, with prosody playing a secondary role, modulating the effects of the lexical cues.

To summarize, while the abovementioned studies have concentrated on the role that lexical cues play in the understanding of (un)certainty, little is known about the role of prosody and gesture patterns play in the early detection of uncertainty and whether they act as precursors in children's pragmatic development. The experimental study in Chapter 2 of this dissertation will deal with children's ability to infer a speaker's uncertainty through gesture and prosody in relation to lexical markers.

1.5.2. Children's production of their own knowledge state

The abovementioned literature has shown that infants as young as 12 months old show certain sensitivity to another person's knowledge state, in this case ignorance, when informing them. And while children at age 3 are able to verbally report on their knowledge state when they are in a total ignorance (and also total knowledge) state, they struggle to do so in partial ignorance conditions until the school years (Rohwer, Kloo, & Perner, 2012). Furthermore, children also need to learn to take their own knowledge state into account when informing others. Previous studies have brought forth evidence that children in their preschool years already display some awareness of their own knowledge state (Balcomb & Gerken, 2008; Lyons & Ghetti, 2011, 2013; Nilsen, Graham, Smith, & Chambers, 2008; Paulus, Proust, & Sodian, 2013; Pillow & Anderson, 2006) while other studies found that children do not acquire this before they are 7 years old (e.g. Beck & Robinson, 2001; Flavell, Flavell, & Green, 2001; Pillow & Anderson, 2006). Thus, the findings in the field have been rather contradictory, and there has not been a clear picture as to when the abilities appear. Yet when differentiating the type of evidence and also the type of uncertainty, a clearer pattern seems to emerge. Previous research has shown that until children reach the age of 6, both their verbal and behavioral awareness of uncertainty is undeveloped. However, when looking at children's behavioral sensitivity to uncertainty, studies have shown that children as early as 2 years old display earlier competence (Balcomb & Gerken, 2008). Furthermore, observing children's gestural and other nonverbal cues, Kim, Paulus, Sodian, and Proust (2016) have shown that 4 year olds, when having only partial access to the objects hidden in a box, display their uncertainty by employing different uncertainty/ignorant gestures. The literature outlined above thus shows that there is a change occurring in the preschool years, between children’s early conveying of their uncertainty through behavioral/nonverbal evidence and later through verbal reflection. There is still only fragmentary knowledge about the processes behind these different abilities. Moreover, not only does explicit verbal meta-reflection on a speaker’s uncertainty give reliable information, but also a speaker’s marking of stance while speaking could give a clear indication of a speaker’s epistemic state.

The present thesis starts from the question of how children develop their uncertain knowledge state through multimodal cues. To our knowledge, no study has investigated children’s production of their uncertain knowledge state by examining prosodic, gestural and lexical markers alongside each other over the development of the preschool years. In order to fill this gap, Chapter 3 will present a cross-sectional study in which children’s knowledge of different objects is controlled for when they have to come up with a guess.

1.6. Children’s understanding and signaling of politeness

Politeness is another pragmatic phenomenon that requires children’s ability to infer and encode social intentions. However we have to define what exactly is politeness is. Culpeper, Haugh, and Kádár (2017) started off their introductory chapter in their recently published handbook on (im)politeness by pointing back to the earliest writings on the topic of politeness as a ‘proper’ way of behaving in the civilizations of Ancient Egypt, Greece, India and China more than two and a half thousand years ago. Since then, politeness research has come a long way, and, as the authors also point out, there is no common conceptual framework of politeness, which may not be so surprising given the different research agendas of different fields in social sciences. Yet, undoubtedly the most popular and traditional theoretical framework through which politeness has been defined is the one by Brown and Levinson (1987). These authors defined politeness as attending to the face-wants of interlocutors, i.e., their desire to be approved (so called positive face) and also their desire to be unimpeded in their actions (negative face). The term face is a sociological concept for the individual's public self-image. It is defined by (Goffman, 1967, p. 5) as “the positive social value a person effectively claims for himself by the line others assume he has taken during a particular contact". According to Brown and Levinson (1987), every human being has two notions of face: positive and negative. Positive face represents the individual’s desire to have his/her wants recognized in interaction. Negative face stands for the individual’s desire for freedom of action and also freedom from imposition. In order for successful social interactions to take place, speakers have to pay attention to both the positive and the negative face of the interlocutor.

A request is a speech act that can pose a potential threat to the interlocutor’s face (i.e., a face-threatening act), and politeness behavior demands that it is mitigated to a certain extent. In a request, the speaker imposes on the recipient's freedom. Thus the hearer might perceive this speech act as an intrusion into his/her freedom of action, and the speaker (the requester) might hesitate to utter a speech act which might make the hearer lose face (Blum-Kulka & Olshtain, 1989). A request can thus be face threatening to both the hearer and the speaker. There are various possibilities on how a request can be uttered, varying from more direct (more transparent) to more indirect strategies (minimizing the imposition on the hearer). First, a speaker can formulate a direct request (through an imperative) such as *Clean up the kitchen*. Second, conventionally indirect strategies can be chosen such as the following: *How about cleaning up?* (positive politeness, which aims at supporting or enhancing the addressee's positive face); or *Could you clean up the kitchen?*; or even *Could you maybe do the cleaning up?* (negative politeness, which aims at mitigating the imposition). These strategies are conventionalized and thus refer to contextual preconditions. Third, a request can be made through non-conventionally indirect strategies (hints) as in *You have left the kitchen in a right mess* (off-record politeness). Here the request refers to the object and is highly dependent on contextual clues. The use of these strategies is shaped by contextual, situational and cultural factors (Blum-Kulka & Olshtain, 1989). Furthermore, the appropriate level of directness can vary from language to language. In order to mitigate the face-threatening nature of a request, downgrades can be applied. For example, if someone is asking for a big favor, he/she tends to use more indirect strategies than when asking for a small favor. While Brown and Levinson mentioned both prosody and gesture several times in their influential work on universal principles of politeness, they have not been well-represented in the study of (im)politeness, which has focused mainly on the verbal aspects. However, this has changed more recently and evidence for prosodic and gestural components in adult (im)polite speech will be summarized.

Similar to knowledge state, understanding and producing politeness requires the ability to infer and to signal communicative intent. In recent years, research on adult speech has worked on establishing the role played by both prosody and gesture play in the communication of politeness by focusing on both perception and production. The link between prosody and politeness was established some time ago through Ohala's Frequency Code (Ohala, 1984). The Frequency Code hypothesis proposed a universal relationship between prosody and politeness. Ohala (1984) claimed that high or rising pitch is universally associated with social meanings such as politeness, deference, submission or uncertainty. Low or falling pitch was to be associated with assertiveness, authority, aggression and confidence. Several studies have brought forth evidence for high pitch being related to politeness. For example, in Dutch and English, pitch height has been found to have an effect on the perception of friendliness (Chen, Gussenhoven, & Rietveld, 2004). In Japanese, it has been found that female speakers talking deferentially use a raise in pitch when talking to someone with a higher status than themselves such as a professor (Ohara, 2001). Also in Mexican Spanish, speakers preferred high initial and high final boundary tones in the production of polite requests (Orozco, 2008, 2010). Recently Winter and Grawunder (2012) demonstrated that it is rather low pitch in Korean that is associated with politeness-related meanings, and they highlighted that other prosodic correlates, such as intensity, duration and voice quality (jitter, shimmer and H2-H1), play an important role in the signaling of the (im)politeness of a message. Similarly, Hübscher et al. (2017a) set out to establish a prosodic profile for politeness in formal register speech in Catalan. Analyzing speech data elicited through a discourse elicitation task, they, like Winter & Grawunder (2012), found that low pitch was perceived as more polite. Furthermore, the authors found there was a general mitigation strategy present in adult speech when talking to a higher status person, with lower intensity, slower speech rate and a decrease in jitter, shimmer and more H2-H1. These studies cast doubt on the universal relation of pitch and social meanings, and that it is important to take other prosodic features into consideration and also that there might be language-specific strategies.

The literature on gesture makes it clear that there are different types of multimodal markers that have been employed to signal politeness-related meanings. Tree and Manusov (1998) classified a number of mitigating and aggravating nonverbal behaviors in American English. The following nonverbal markers have been found to have a mitigating function: pleasant facial expression, raised eyebrows, direct body orientation, a tense closed position with small gestures and soft voice, touch and close proximity; on the other hand, other behaviors, such as unpleasant facial expression, lowered eyebrows, no touching, indirect nod orientation, wide gestures and a loud voice, have been found to have an aggravating function. This might be typical for American English, as research has shown that politeness-related gestures are culture-related (Kita, 2009). It is thus unsurprising that different tendencies have been found in Korea. Brown and Winter (in press) analyzed gestures and other nonverbal behaviors used in four types of interactions where different characters interact in a Korean television drama. They coded arm gestures, facial gestures, physical contact, self-touching, head nods and bows and orientation of the speakers. They found that while in casual speech there were more arm gestures, facial gestures, physical contact and self-touching, in polite speech register more head nods and bows were found. Furthermore, forthcoming work on Catalan shows that overall adult participants use fewer gestural cues when talking to a person of higher social status compared to when talking to a classmate, with individual differences depending on the speaker (Hübscher, Sánchez-Conde, Vincze, Borràs-Comes, & Prieto, forthcoming). Another relevant study on speaker's perception of politeness in Catalan, which takes both prosody and gesture into account, found that the expansion of pitch range can play a role in the understanding of politeness, however only if the facial cues match the polite intention signaled (Nadeu & Prieto, 2011). This study again demonstrates the importance of investigating prosodic cues together with gestural cues in order to assess the pragmatic meaning encoded.

While these studies provided evidence that both speakers and listeners integrate information from both auditory and visual sources, much less is known on how children integrate speech and gesture in their polite stance signaling behavior and how this changes over development.

1.6.1. Children's comprehension of a speaker's politeness

To find out whether a child can express his or her own awareness of whether an utterance is polite or not, children's metalinguistic awareness must be addressed. Metalinguistic awareness has been defined as the ability to reflect upon and also manipulate features of a spoken language (Tunmer & Herriman, 1984). Research on the development of children's metalinguistic awareness has demonstrated that 3- to 4-year-old preschool children are able to make metalinguistic judgments for various aspects in their native language (e.g. de Villiers & de Villiers, 1974; Liberman, Shankweiler, Fischer, & Carter, 1974; Smith & Tager-Flusberg, 1982). However, data coming from observation in child language research has shown that even at a much younger age, children show early signs of metalinguistic awareness, for example, through 'self-repairs' (Clancy, 1985; Slobin, 1978). More concretely, research on children's awareness of politeness has brought fourth with varying results as to when it starts to emerge (Andersen, 1977; Axia & Baroni, 1985; Baroni & Axia, 1989; Hollos, 1977; James, 1978; Shatz & Gelman, 1973). Anderson et al. (1999) have suggested that English-speaking children of around 4 years of age start to understand in which way different lexical and non-lexical discourse markers should be used, depending on the status asymmetry between the speaker and the addressee. Tsuji and Doherty (2014) analyzed Japanese-acquiring children and similarly found that at 4 years of age children were able to distinguish appropriately between polite and impolite forms. Bates (1976), to our knowledge, is the only study which has investigated preschool children's production and understanding of politeness through intonation and lexical/morphosyntactic cues (see Shochi, Erickson, Sekiyama, Rilliard, & Aubergé, 2009 for school-aged children's understanding of politeness through intonation and facial expression). Exploring Italian children's (3 to 7 years old) spontaneous requests, Bates (1976) found that while until age 4, children mainly used imperative and direct questions as requests; from age 5 to 6, they are able to produce each different syntactic forms needed, but are not yet versed in modulating their requests, and finally when they are around 7 years old, they manage to be quite polite through modulating the form and content. In order to complement her study on politeness production, Bates (1976) asked Italian-acquiring children to judge how polite a request was. The children were in charge of candies and were asked to give them to the frog that asked in the nicest way. In total, children were exposed to eight pairs of requests. The results were the following: the use of *please* was identified as polite first at 3 years of age, and then at 4 years of age also question intonation was recognized as more polite. However, children struggled with the difference between imperative and interrogative until around 5 years, and later at 6 years, children judged the conditional form to be more polite than the indicative form.

Summing up, while the abovementioned studies have concentrated on the role played by lexical cues in the understanding of politeness-related meanings, little is known about the role of prosody and gesture markers in the early detection of politeness and whether 3-year-old children are able to access a speaker's politeness pragmatic development on the basis of prosody and gesture. These questions will be addressed in Chapter 4 of this dissertation.

1.6.2. Children's production of politeness

In order to be able to use politeness appropriately, children have to master several linguistic and social dimensions. On the one hand, they need to learn those forms that are considered polite, and, on the other hand, they have to identify the different sociopragmatic conditions in which this form is appropriate, and hence take into account the interlocutor's status, age and also the social cost in the case of requests. Children's production of politeness-related meanings has been studied both experientially and also through observations. A great majority of cross-linguistic research on the development of politeness has been concerned with children's acquisition of speech acts, in particular, with directives (requests, commands and orders), for example, in children acquiring English (e.g. Ervin-Tripp & Gordon, 1986; James, 1978; Sealey, 1999), French (Marcos & Bernicot, 1994; Ryckebusch & Marcos, 2004), Swedish (Aronsson & Thorell, 1999), Norwegian and Hungarian (Hollos & Beeman, 1978) and Greek (Georgalidou, 2008). In general, it has been found that while children tend to recognize relatively early that politeness is an important ingredient in achieving their goals, the development of the different forms and conventions take quite some time to develop. For a long time, the focus of developmental research has been on children's development of verbal politeness, but more recently some studies on children's socialization into politeness have highlighted how, for example, in Japan children are socialized at a young age to politeness routines through linguistic and also embodied practice (see Cook & Burdelski, 2017 for an overview). He provides an example of two around 2-year-old children, in which one child encounters himself in a situation wherein he receives a gift. Parents not only model the verbal content of the gratification expression for the boy, but they also place their hands on the boy's back so that he bows along with saying *thank you* when receiving a toy (see Cook & Burdelski, 2017, p. 280).

Focusing on the social rules behind the choices of polite forms, Axia and Baroni (1985) carried out research on whether children varied their requests on the basis of cost. In their experiment, children aged five, seven and nine made repeated requests to adult interlocutors. Whenever the adult judged a request to be insufficiently polite, they did not respond. For all three age groups, the adult intentionally ignored every request that was made for the first time (a refusal to respond increases the 'face' cost to the requester of a subsequent request). In their second, reiterated requests, 5-year-olds rarely knew how to make their request politer while seven-year-olds were somewhat more adept at this. 9-year-olds were much more skillful in this because they knew how to add mitigators like *please* and fall back on question forms or the conditional tense. Furthermore, children's developing awareness to sociocultural factors triggering politeness-related behavior has also been investigated in the social psychology literature, and it has been found that 4- to 7-year-old children, when looking at static pictures that showed people with varying eyebrow positions or facial expressions (smiling vs. not smiling), inferred that people with lowered brows and less positive facial expressions would be socially dominant over those with raised brows or positive expressions (Keating & Bai, 1986). Furthermore, more recently Brey and Shutts (2015) followed up on this and studied whether young children use nonverbal information to make inferences about differences in social power (high power represented by expansive posture, head tilted back, direct gaze toward the other actor, lowered eyebrows and a loud voice vs. low power depicted by hunched posture, head titled down, varied gaze (averted when speaking, but direct when being spoken to), raised eyebrows, and a quiet voice. They found that children by age 5 are able to do so without any accompanying speech. While this points to children's relatively early sensitivity to factors influencing speakers' positioning in conversation, it is not yet known whether children use those cues themselves when interacting with people with varying social factors.

To summarize, while the literature outlined above has shown that children make important strides in their production of politeness-related meanings, the questions that remain unanswered are if and how children express politeness multimodally, and, if so, what the role of prosody and gesture play in this developmental process alongside lexical and grammatical markers of politeness is still unanswered. Thus, following our research on Catalan adults (Hübscher et al., 2017, Hübscher et al., forthcoming), in Chapter 5 we will analyze children's semi-spontaneously elicited speech data in politeness-related situations that are controlled for sociopragmatic factors such as social power/distance and cost. In this cross-sectional study, we will be able to assess how children make progress in their development of politeness-related stances throughout the preschool years.

1.7. Goals and scope of the thesis

This dissertation adopts a comprehensive perspective on the prosody-gesture-pragmatics interface with an emphasis on the role played by prosody and gesture play in preschool children's multimodal pathway into knowledge state and politeness understanding and signaling. Building on previous proposals, this thesis contends that the developmental process of children's development of these crucial pragmatic skills is inherently multimodal.

Before describing the studies of this thesis in more detail, there are some methodological aspects that need to be introduced. Firstly, this thesis includes a combination of both comprehension and production studies. By focusing on both production and comprehension abilities, it can be learned what a child does in a particular situation (production), and through perception it can be tested what the child picks up (comprehension). Studying both comprehension and production together will be useful in assessing how prosody and gesture is used to convey a child's own knowledge state and ability to produce politeness as well as how he or she is sensitive to other people's knowledge states and politeness. Secondly, this thesis consists of a number of cross-sectional studies, comparing different groups at a specific point in development (compared to longitudinal studies, which observe the same subjects over a period of time, sometimes lasting many years). The focus on children's ages between 3 and 5 years has been chosen since it is a period in which important cognitive and linguistic changes are taking place. More particularly, existing experimental studies on children's development of knowledge state and politeness have shown that children make great strides over the preschool years, having acquired certain lexical and certain morphosyntactic strategies usually between 4 and 5 years (e.g. Matsui, 2014 for a review on knowledge state; Zufferey, 2016 for a review on politeness). In order to widen the scope, an age range of 3-5 years seems suitable in order to investigate a possible early development of prosodic and gestural cues to knowledge state and politeness. Thirdly, in all empirical chapters of this thesis, there is a consistent focus on the behavior of one individual, and it is considered as our 'unit of analysis' (Bavelas & Healing, 2013). Also the interaction took place in pragmatically adequate situations. Having introduced some methodological aspects of this thesis, the research questions of the four empirical studies will now be outlined.

The following four main research questions will be assessed, each in a separate chapter:

1. Do preschool children comprehend a speaker's uncertainty earlier through intonational and gestural cues as compared to lexical cues? (Study 1)
2. Do preschoolers express their uncertain knowledge state earlier through intonational and gestural cues than through lexical cues? (Study 2)
3. Do preschoolers rely on prosody and facial gestural cues when assessing a speaker's polite stance in requests? (Study 3)
4. Do preschoolers express their polite stance earlier through prosodic and gestural cues than through lexical cues? Is their polite stance dependent on sociopragmatic factors such as social distance and cost? (Study 4)

The central working hypothesis of this thesis is that both prosody and gesture play an important role in children's acquisition of complex pragmatic meanings and that they have a precursor role in the development of both epistemic and polite stance. The focus is on the Catalan language in 3 out of 4 studies while in the remaining study American English is investigated. This work is the outcome of a research collaboration established with Laura Wagner during a three-month research stay at Ohio State University in Columbus, Ohio (USA).

In order to answer these questions, the current thesis is structured in the following way: Chapters 2 and 3 contain studies centered on knowledge state, and Chapters 4 and 5 are dedicated to politeness. In both these parts, the first chapter focuses on children's ability to understand the epistemic or polite stance of an adult speaker, and the second chapter explores the production of children's own epistemic or polite stance-taking.

Study 1 (Chapter 2) investigates preschoolers' sensitivity to lexical, intonation and gestural information in the comprehension of speaker uncertainty. A total of 102 3- to 5-year-old Catalan-speaking children participated in a comprehension task, which involved the detection of uncertainty in materials that combined lexical, intonational and gestural markers. In a between-subjects design, the children were either exposed to the lexical condition (where they received lexical and gestural cues to uncertainty) or the intonation condition (where they were exposed to intonational and gestural cues to uncertainty). Within each condition, three different presentation formats were used (audio-only, visual-only and audiovisual) as within-subject variables. The hypothesis is that children will understand another speaker's knowledge state earlier and better through gestural cues and also better through prosodic cues than through lexical cues.

In order to have a more complete picture of how children develop knowledge state marking from a developmental perspective, we also explore this issue from a production point of view. In the second cross-sectional study of the thesis (Chapter 3), we investigate how preschool children express their own knowledge state. A total of 40 Catalan preschool children (and 10 adults) took part in a guessing game involving a total of 10 objects (5 easy objects that they had previously seen and touched and 5 difficult ones that they had neither seen nor touched). Next to being asked to report what they think is in the box, children were told afterwards to explicitly assess their knowledge state. The hypothesis is that children would express their uncertain knowledge state first through epistemic prosodic and gestural markers before employing lexical cues and also before being able to explicitly assess level of (un)certainty.

Chapter 4 presents the third study of the thesis, which investigates preschool-age children's sensitivity to intonational and facial cues signaling a speaker's polite stance in requestive speech acts with controlled lexical and contextual materials. Thirty-six 3-year-old American English-speaking children performed a forced-choice decision task (in a between-subject design), which examined whether children at this age use changes in intonational patterns and also facial cues as a marker to a speaker's polite stance in either audio-only, visual-only or audiovisual presentation modalities, when lexical cues are controlled for. The prediction is that children are able to infer speaker intention on the basis of prosody only and also facial cues only, but that they perform better when they have both cues present.

Chapter 5 presents the fourth study of the thesis, which analyzed whether prosody and gesture play a precursor role in children's acquisition of more complex pragmatic skills such as politeness. Through a cross-sectional study, sixty-four 3- to 5-year-old Catalan-dominant children participated in a request production task in four different conditions, which were controlled for sociopragmatic factors such as social power/distance and cost. The hypothesis is that (a) younger children will use an increased range of gestural and prosodic cues to express their polite stance before they are able to express their politeness similarly through lexical markers; and (b) children in general use more politeness markers in specific social conditions (e.g., request towards the adult interlocutor and in high-cost situations).

Finally, Chapter 6 contains a general discussion on the results of this thesis. Although each experimental chapter contains its own discussion, the final discussion aims at pulling all the results from previous chapters together, outlining overarching findings and putting them in the wider context of the literature. It also addresses avenues for future research.

As a final remark, I would like to mention that the four empirical studies presented in this thesis (Chapters 2, 3, 4, 5) have either been published or are under review in peer-reviewed scientific journals. The author of this thesis was the first and leading investigator in all four studies. Each chapter is self-contained and has an introduction and discussion section. Due to this fact, there will be some unavoidable overlap between the different chapters. Furthermore, as the individual studies have been submitted to different journals, there might be small variations in style.

2

Chapter 2: Intonation and gesture as bootstrapping devices in speaker uncertainty

This chapter is adapted from:

Hübscher, I., Esteve-Gibert, N., Igualada, A., & Prieto, P. (2017). Intonation and gesture as bootstrapping devices in speaker uncertainty. *First Language, 37*(1), 24-41.

2.1. Introduction

In everyday conversation, speakers are able to rapidly combine multimodal information during utterance comprehension, including verbal content, prosody and gesture. In particular, in successful social interactions the detection of belief states such as uncertainty (or incredulity, surprise, etc.) is especially important in order to understand the other person's epistemic stance. Epistemic stance refers to the degree of commitment or certainty the speaker has in his or her statements. When inferring uncertainty, listeners can use various cues (depending on the language) such as lexical epistemic markers, morphosyntactic marking, gestures such as head nods and facial expressions, or prosodic features such as delays and final rising intonation (Borràs-Comes et al., 2011; Krahmer & Swerts, 2005). Typical lexical markers in English, for instance, are mental state verbs (such as *think*) and epistemic modal expressions (such as *maybe*). These lexical items convey information about the epistemic stance of individuals. In many languages intonation plays a key role in shaping the pragmatic meaning of utterances and can encode epistemic and evidential information (see Barth-Weingarten et al., 2009, for a review of the literature on the prosody-pragmatics interface; Prieto, 2015). Gesture patterns can also play an important part in conveying epistemic information. For example, Krahmer and Swerts (2005) investigated the role of audio-visual prosody for signalling and detecting epistemic information in question answering. The study showed that there are well-defined visual cues that demarcate a speaker's feeling of knowing and that listeners are more capable of estimating another person's knowledge on the basis of visual and auditory information combined than just auditory input alone.

In the study of language development, one of the interesting questions is how and when children develop the ability to recognise an interlocutor's epistemic stance and feeling of knowing. To date, most research has concentrated on children's acquisition of lexical markers of belief states. Moore et al. (1989) classical study tested 3- to 8-year-old children in an experimental setting where children had to find an object in one of two boxes as they listened to verbal cues from two different puppets telling them about the place where the object was hidden. Each utterance contained a marker with a different degree of certainty, signposting one or the other box as the location of the hidden object, such as *I know it's in the red box* or *I think it's in the blue box*. The results showed that children aged 4 and above were able to find the hidden object based on what they heard but 3-year-olds were not. Furthermore, Moore et al. (1990) also showed that the understanding of modal expressions such as *might* strongly correlates with the understanding of mental verbs such as *think*. Likewise, Noveck, Ho, and Sera (1996) tested 5- to 9-year-old children's understanding of epistemic modals by contrasting *has to* with *might*, etc., and showed that (a) their under- standing of modal expressions develops gradually over time; and (b) by 9 years of age children show an adult-like understanding of these modal expressions. There is one exception, though, by Moore et al. (1993), who compared children's (3 to 6 years old) comprehension of mental state lexicon to their comprehension of mental state prosody. Children had to listen to contrasting pairs of statements by two puppets and guess the location of a hidden object. Each statement pair either differed with respect to the mental state verbs *know* vs *think* or *think* vs *guess* – or with respect to terminal pitch contour – falling or rising. While 3-year-olds were not able to use either lexicon or prosody to detect where the object was, 4-year-olds started to do so significantly in the *know* vs *think* and falling vs rising contrast conditions. Furthermore, the *think* vs *guess* condition was much harder even for the 5-year-old children, compared to the *know* vs *think* condition. In a follow-up experiment with 3- to 5-year-old children, the conditions were presented as either matched or mismatched. While in the matched condition lexical items of certainty went together with falling intonation and lexical items of uncertainty were matched with rising intonation, in the mismatched condition the opposite was applied. While 4-year-olds performed significantly above chance when *know* vs *think* was matched with the corresponding prosodic cue, they did not show any significant difference between matched vs mismatched trials. This time 5-year-olds performed much better on the *think* vs *guess* distinction in the matching condition. But, interestingly, 5-year-old children performed much worse in the mismatched condition, showing a certain awareness of prosodic and lexical integration when speaker (un)certainty is expressed. The authors suggested that prosodic and lexical cues to speaker certainty start to be used around the same time by children. Yet, the authors propose that lexical cues to a speaker's belief state initially seem to be more dominant, with prosody playing a secondary role, modulating the effects of the lexical cues. More recent studies have investigated the acquisition of belief states, focusing on other languages such as Korean (Choi, 1995; Papafragou, Li, Choi, & Han, 2007), Cantonese (Lee & Law, 2001; Tardif, Wellman, & Cheung, 2004), Turkish and Puerto Rican Spanish (Shatz, Diesendruck, Martinez-Beck, & Akar, 2003), Japanese (Matsui et al., 2006), and Japanese and German (Matsui, Rakoczy, Miura, & Tomasello, 2009), yet with a sole focus on lexically encoded mental state information.

There seems to be a general consensus that it is not until age 4 that children are capable of identifying the meaning of modal expressions of uncertainty. Yet Matsui et al. (2006)investigated children's understanding of knowledge states in Japanese, where uncertainty can be encoded through both epistemic particles (*yo* = speaker certainty and *kana* = speaker uncertainty) and mental state verbs (such as *shitteru* = know and *omou* = think). They found that 3-year-old Japanese children already comprehended a speaker's knowledge state, but only when it was conveyed by particles. By contrast, at that age their understanding of mental state verbs was still quite poor (see Matsui, 2014, for a detailed overview of children's understanding of epistemicity and evidentiality).

Studies of children's pragmatic development have claimed to take into account gestural cues in the study of communication and language development (e.g. Furman, Kuntay, & Ozyurek, 2014; Guidetti, 2005; Guidetti, & Nicoladis, 2008; Iverson & Goldin-Meadow, 1998; McNeill, 1998; O'Neill, Bard, Linnell, & Fluck, 2005). There is a growing consensus that gestures act as bootstrapping devices in language development (Butcher & Goldin-Meadow, 2000b; Kelly, 2001). With respect to the acquisition of belief states, some studies seem to suggest an earlier development of uncertainty understanding based on non-linguistic cues. For example, some studies have shown that 3- and 4-year-old children are capable of deciding who to believe based on visual signs of reliability or inference (Koenig, Clément, & Harris, 2004; Koenig & Harris, 2005; Robinson, Mitchell, & Nye, 1995; Robinson & Whitcombe, 2003; Sabbagh & Baldwin, 2001; Whitcombe & Robinson, 2000). There have been two studies that focused on older children (8 to 11 years old) which have investigated the development of their perception and production of facial gestures as cues to uncertainty (Krahmer & Swerts, 2005; Visser et al., 2014). More basic forms of epistemic stance comprehension are also found early on in infancy. It has been shown that 12-month-olds are able to distinguish between knowledgeable and ignorant partners (Liszkowski et al., 2008). The study explored the ability of 12-month-old infants to point appropriately at an object in order to provide uninformed people with information. To signal ignorance, the experimenter raised his/her hands with the palms upturned. Their results showed that infants pointed more often to an object which the adult had (presumably) not seen fall down and thus needed help to find than an object which the adult had seen fall down and thus could find unassisted.

All these studies suggest that children achieve important communicative milestones initially in the realm of gesture before they do so in speech, and gestures can therefore be seen as helping children to access meaning (e.g. Goldin-Meadow, 2007). While the role of prosody as a syntactic bootstrapper has been highlighted in language acquisition research, that is certain types of prosodic features guide children's initial acquisition of word order and syntactic structure (see also Christophe, Nespor, Guasti, & Van Ooyen, 2003; e.g. those related to constituent or prosodic phrasing; for a conceptualisation see Hirsh-Pasek, Tucker, & Golinkoff, 1996), so far very little is known about the role of prosody in early pragmatic development and whether it might have a possible bootstrapping effect on the comprehension of pragmatic meaning.

Recent studies on the prosody–pragmatics interface have shown that 12-month-old infants use prosody (together with pointing gestures) to comprehend an adult's basic communicative intentions like expressive, imperative and informative attention-directing actions (Esteve-Gibert et al., 2017b), and that 14-month-old infants can use prosody to distinguish between intentional and non-intentional acts (Sakkalou & Gattis, 2012). Also, research has shown that infants as young as 2 display a basic inventory of target-like intonation contours with an adult-like intentional meaning (Chen & Fikkert, 2007; Frota et al., 2016; Prieto et al., 2012b). Thus, independent evidence coming from studies investigating the acquisition of pragmatic intonation seems to suggest an initial role for prosody as a bootstrapping mechanism in the early stages of the understanding of pragmatic meaning.

While prosody is a very prominent cue in infancy, studies testing preschool and school-age children's understanding of prosody have yielded conflicting results. On the one hand, research on children's sensitivity to pitch as a cue to emotions has shown that the adult-like ability to judge a speaker's emotional state based on vocal affect is mastered only at 4 years, after children have acquired the lexical semantic meaning of the four basic emotions (happiness, sadness, anger and fear), which happens around age 3 (Morton & Trehub, 2001; Nelson & Russell, 2011; Quam & Swingley, 2012). Yet, when there are cues in competition regarding the relevant emotion conveyed via either the lexical meaning of a sentence (Morton & Trehub, 2001; Waxer & Morton, 2011) or the situational context (Aguert, Laval, Lacroix, Gil, & Bigot, 2013; Aguert et al., 2010), the success of preschoolers at identifying vocal affect seems compromised. For example, if someone utters ‘It’s Christmas time’ with a sad prosody, adults will rely on the prosody and judge the speaker to be sad, whereas 6-year-old children will say the speaker is happy. By the same token, Vernice and Guasti (2014) showed that before age 5 children are not able to use prosodic cues in order to decide which referent to mention next. While all these studies hint at a surprisingly late acquisition of certain prosodic cues at the sentence level, a very recent study by Berman, Chambers, and Graham (2016) discovered that when a more implicit methodology such as eye-tracking is used, young children already at age 3 show themselves able to link speech bearing different acoustic cues to emotion. Unfortunately, overall these studies lack a description of the acoustic characterisation of the prosodic differences between the different emotions described, which makes it hard to track which prosodic cues children learn to attend to (with the exception of Quam & Swingley, 2012). Furthermore, these prosodic cues to emotion might only be subtle cues that do not involve a real change in pragmatic intonation patterns.

On the other hand, however, hardly any research has focused on when and how children understand more complex pragmatic meanings such as epistemicity encoded through prosody and/or gestures. A recent exception is (Armstrong, 2012; Armstrong, 2014), which focused on children’s comprehension of intonationally encoded disbelief in polar questions in Puerto Rican Spanish. Particularly relevant for the current study is the study by Armstrong, Esteve-Gibert, and Prieto (2014), which investigated 3- to 5-year-old understanding of disbelief (or incredulity) through three different modalities: visual-only (facial gesture cues), audio-only (intonation) and audio-visual (facial gestures and intonation). The children were exposed to short discourse reactions such as *Una balena?!* ('A whale?!') produced with either incredulous or credulous intonation, and they had to decide between the two meanings. The results showed that 3-year-old children performed the worst on the audio-only task, while 4-year-olds performed better, but still showed great variability. Also, a great deal of variability was observed for younger children that received the audio- visual condition, arguably because it was difficult for some of them to integrate the two cues. By contrast, 3- and 4-year-olds performed much better in the visual-only condition compared to the other two conditions. Furthermore, 5-year-old children performed equally well in the audio-only condition. The authors suggest that facial gestures seem to provide children with scaffolding for the detection of speaker disbelief. However, one aspect that this study could not explore was the children's sensitivity to prosodic and gestural features relative to lexical cues, which were not included in the study.

The main purpose of the current study is to assess the relative roles of lexical, intonational and gestural cues in preschoolers' understanding of uncertainty and to test the potential bootstrapping role of gestures and intonation in its development. Specifically, we are interested in whether children (1) use gestures as a bootstrapping device in the comprehension of uncertainty and (2) recognise uncertainty more easily through lexical or intonational epistemic markers. To address these questions, we asked 3- to 5-year-olds to select the uncertainty stimuli in a forced-choice task.

A modified version of Armstrong et al.'s (2014) incredulity comprehension task was used here in which children had to decide which speaker was uncertain about something. Importantly, the two experimental conditions (uncertainty/certainty) were tested using stimuli presented in either visual-only, audio-only, or audio-visual modality.

In line with previous studies on the facilitator role of gestures in general (e.g. Guidetti, 2005; Iverson & Goldin-Meadow, 1998; McNeill, 1998; O'Neill, Bard, Linnell, & Fluck, 2005) and in particular in disbelief understanding (Armstrong et al., 2014), we expected that the presence of visual information would bootstrap children's understanding of belief state. Furthermore, contrary to previous studies on the late acquisition of meaning encoded through prosody, it was our position that children younger than 4 years would be sensitive first to intonational cues to uncertainty, then to lexical ones. The results would therefore be important to further our understanding of how pragmatic communication skills develop in children and the role intonation and gesture play in this development.

2.2. Methodology

2.2.1. Participants

A total of 102 3- to 5-year-old children participated in the experiment. Children were divided into a younger group (N = 51, mean age = 3 years and 9 months, SD = 5.50) and an older group (N = 51, mean age = 5 years and 2 months, SD = 5.27). All the participants were preschoolers at three Catalan public schools located in the Barcelona area. In these schools, the main language of instruction is Catalan.[4] Parents were informed about the experiment's goal and signed a participation consent form. Furthermore, language exposure questionnaires (based on Bosch & Sebastián-Gallés, 2001) were administered to the caregivers in order to ensure that the participating children were predominantly exposed to Catalan (as opposed to Spanish) on a daily basis (mean percentage of overall exposure to Catalan: 87%, *SD* = 12.0).

2.2.2. Design

The target materials were video recorded by taking into consideration the results of a general knowledge quiz which was constructed to elicit the spontaneous use of utterances conveying different degrees of certainty in Catalan, based on Krahmer and Swerts (2005). We analysed the lexical, gestural and prosodic expressions of certainty in a total of 180 answers (12 questions × 15 participants). Results showed that participants mainly used two different types of intonation patterns, depending on the certainty condition.

In the certain condition they universally used a falling pitch contour L* L% (100% of the cases) and in the uncertain condition they used two variants of a rising pitch con- tour, L* H% and L+H* H% (which covered 25% of the cases). Furthermore, participants used lexical items (*potser* 'perhaps',

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

crec que ‘I think’, etc.) in 25% of the cases, which went together with a falling intonation (L* L%) when expressing uncertainty (the remaining 50% belonged to lower degrees of uncertainty). Finally, participants produced a head nod when being certain, and a varied group of gestures (e.g. diverted gaze, low/high gaze, raised or furrowed eyebrows, squinted eyes, head tilt) when being very uncertain.

Taking these findings into account, three adult Catalan speakers were videotaped while producing a total of 12 target utterances each (6 trials × 2 epistemic marking conditions; see Appendix A), resulting in a grand total of 36 target stimuli (3 speakers × 12 utterances). The epistemic marking conditions consisted of utterances expressing certainty/ uncertainty through both lexical and gestural markers (this will henceforth be referred to as the lexical condition), and utterances expressing certainty/uncertainty through only intonation and gestural markers (henceforth the intonation condition). A fourth Catalan speaker was recorded for the familiarisation trial.

For the *lexical condition* we used the following lexical epistemic markers[5]: a common adverb signalling uncertainty in Catalan (*potser* ‘maybe’), and a very common epistemic construction which signals certainty (*segur que* ‘[I am] certain that’).[6] Thus, the certainty stimuli consisted of a noun phrase

---

[5] Catalan, like other Romance languages, uses a set of epistemic markers and morphosyntactic resources to mark epistemic commitment such as epistemic adverbs (e.g. *potser* ‘perhaps’), conditional forms (e.g. *vindria* ‘I would come’), verbal tense and subjective mood (e.g. *dubto que vingui* ‘I doubt he’d come-subjunctive’), etc.

[6] The two forms typically appear in sentence-initial position and thus in an especially prominent position for children to acquire them (for more information about epistemic and evidential marking in Catalan, see González, Borràs-Comes, Roseano, & Prieto, 2014).

preceded by the adverb *segur que* '[I am] certain that' (e.g. *Segur que el tomàquet* '[I am] certain that [it's] the tomato') and accompanied by a head nod gesture suggesting certainty (Figure 1, left-hand panels). The uncertainty stimuli consisted of a noun phrase preceded by the adverb *potser* 'maybe' (e.g. *Potser el tomàquet* 'Maybe [it's] the tomato') and accompanied by gestures suggesting uncertainty (squinted eyes, raised eyebrows, head tilt) (Figure 1, right-hand panels). Crucially, both certainty and uncertainty utterances were produced with the same intonation contour (L* H% associated with the adverb plus final falling intonation, L* L%).

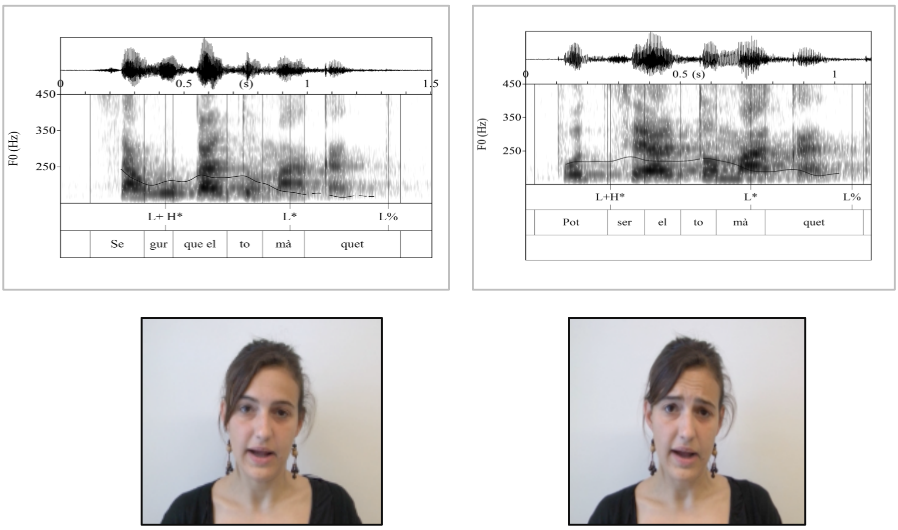


Figure 1. Lexical condition. Upper panels: pitch tracks, spectrograms and waveforms for the certainty utterance (*Segur que el tomàquet* '[I am] certain that [it's] the tomato') (left-hand panel) and uncertainty utterance (*Potser el tomàquet* 'Maybe [it's] the tomato') (right-hand panel). Lower panels: screenshots of facial expressions corresponding to certainty (left-hand panel) and uncertainty utterances (right-hand panel).

For the *intonation condition*, the certainty stimuli (e.g. *El tomàquet* ‘The tomato’) were produced with a falling intonation contour (L* L%) and accompanied by a head nod gesture suggesting certainty (Figure 2, left-hand panels). The uncertainty stimuli (e.g. *El tomàquet?* ‘The tomato?’) were produced with a rising intonation contour (L* H%) and gestures suggestive of uncertainty (squinted eyes, raised eyebrows, head tilt) (Figure 2, right-hand panels). Crucially, in the intonation condition the utterance contained no lexical information such as epistemic adverbs which would help to distinguish certain from uncertain stimuli.

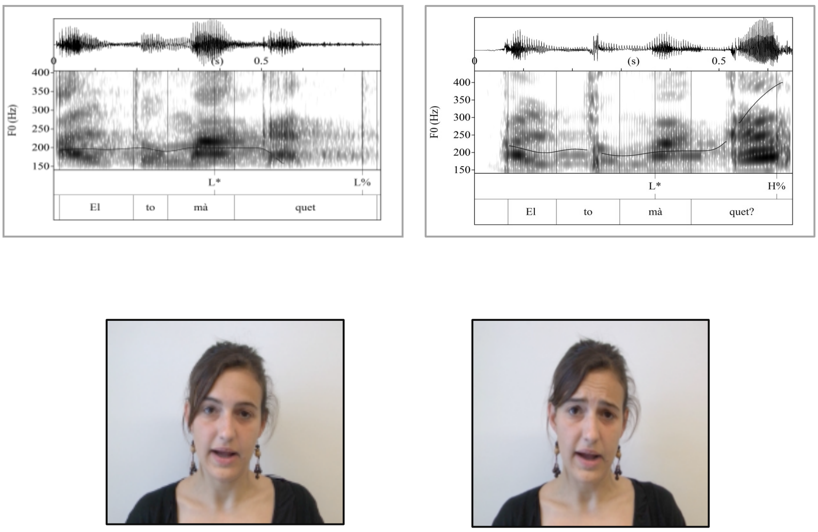


Figure 2. Intonation condition. Upper panels: pitch tracks, spectrograms and waveforms for the certainty utterance (*El tomàquet* ‘The tomato’) (left-hand panel) and uncertainty utterance (*El tomàquet?* ‘The tomato?’) (right-hand panel). Lower panels: screenshots of facial expressions corresponding to certainty (left-hand panel) and uncertainty utterances (right-hand panel).

Each epistemic marking condition was presented in three different modalities: audio-only, visual-only and audio-visual. For the audio-only trials, the audio track was played to subjects and the visual information reduced to a minimum by displaying two still photos of the speakers with a neutral facial expression. For the visual-only trials, the audio track was removed from the original audio-visual stimuli so that only the visual information was available to subjects. For the audio-visual trials, both the audio track and the accompanying video images were presented.

Table 1 summarises how the combination of lexical, intonational and gestural cues differed across the epistemic marking conditions (lexical condition vs intonation condition) and modalities of presentation (audio-only, visual-only, or audio-visual). This design was intended to allow us to assess the role of the visual cues with respect to the speech cues (be they intonational or lexical cues) in (un)certainty detection.

Table 1. Lexical, intonational and gestural cues of the stimuli according to epistemic marking condition (intonation vs lexical condition) and modality of presentation (audio-only, video-only, or audio-visual).

	Audio-only		*Video-only*		*Audio-visual*	
	Certain	*Uncertain*	*Certain*	*Uncertain*	*Certain*	*Uncertain*
Intonation condition	Falling L* L%	Rising L* H%	Head nod	Squinted eyes, raised eyebrows, head tilt	Falling L* L% Head nod	Rising L* H% Squinted eyes, raised eyebrows, head tilt
Lexical condition	*Segur que* '[I am] certain that' L* L%	*Potser* 'Maybe' L* L%	Head nod	Squinted eyes, raised eyebrows, head tilt	*Segur que* 'I'm certain that' Falling L* L% Head nod	*Potser* 'Maybe' Falling L* L% Squinted eyes, raised eyebrows, head tilt

The semantic appropriateness of the stimuli selected was controlled for by running an experiment with the online survey platform SurveyGizmo. Sixty Catalan-speaking adults (30 respondents × 2 epistemic marking conditions) were asked to rate each of the 9 experimental stimuli sets (including both uncertainty and certainty stimuli), yielding a total of 540 tokens. Out of these 540, only two elicited contradictory certainty ratings by respondents.

These two stimuli were subsequently re-recorded, and further testing yielded consistent ratings.

2.2.3. Set-up of the task

This task is an adaptation of the task used in Armstrong et al. (2014). A PowerPoint presentation depicted the story of two twins travelling on a train with their friend Barbara, who plays a game with them to help make the journey pass more quickly.

The game consists of her asking the twins if they know about her favourite things. For example, Barbara asks them, ‘What is my favourite vegetable?’ The answer is then revealed visually as a tomato in a thought bubble (Figure 3, left image), which the experimenter points out to the child. Previous research has shown that 3-year-olds understand thought bubbles as representations of mental contents (Wellman, Hollander, & Schult, 1996).

During the experiment, the child subject was seated in a position to view the screen as a researcher talked and operated the PowerPoint slide show. Once the twins and Barbara had been introduced and the basic guessing game scenario described, the researcher told the child that for each question there was one twin who was sure of the right answer and one who was not, and that the child had to point to the uncertain twin. The child’s response was regarded as ‘correct’ if s/he pointed to the twin who expressed uncertainty (Figure 3, right image).

Figure 3. Sample slides from the PowerPoint presentation used in the comprehension task. Left-hand slide: Barbara is thinking of her favourite vegetable. Right-hand slide: Barbara (top) and the twins (bottom).

2.2.4. Procedure

The children were tested individually in a quiet room at each of the three participating schools. The researcher, a male Catalan-speaking adult (the third author of this article), was seated beside the child in a room at the child's school, so that both faced the computer screen. The children were administered either the lexical or intonation condition (between-subjects), each containing 3 audio-only, 3 visual-only and 3 audio-visual trials (within-subjects) in a randomised order. Prior to performing the comprehension task, each participant first went through a familiarisation trial to make sure that they understood what they were supposed to do. They then performed a total of 9 test trials in two counter- balanced orders, either first 3 audio-only, then 3 visual-only and finally 3 audio-visual, or first 3 visual-only, then 3 audio-only and finally 3 audio-visual. After each set of 3 trials, in order to prepare the child for the change in modality, s/he was shown a filler slide depicting either a photo of an ear (signalling audio-only), an eye (visual-only), or both (audio-visual). In total the procedure lasted at most 10 minutes.

2.3. Results

A total of 918 children's responses were obtained from the comprehension task (9 responses × 102 children) and then analysed through a Generalised Linear Mixed Model (GLMM) using IBM SPSS Statistics 21. The dependent variable was 'child's performance', a numerical measure obtained by calculating the mean proportion of correct to incorrect responses. The fixed factors were epistemic marking condition (two levels: intonation condition, lexical condition), modality of presentation (three levels: audio-only, visual-only, audio-visual), age group (two levels: younger group, older group), and all their possible interactions. The random factor was participants.

Figure 4 shows the mean proportion of correct responses broken down by epistemic marking condition (intonation and lexical) and modality condition (audio-only, visual- only and audio-visual) for the two age groups (younger and older) in the sample.

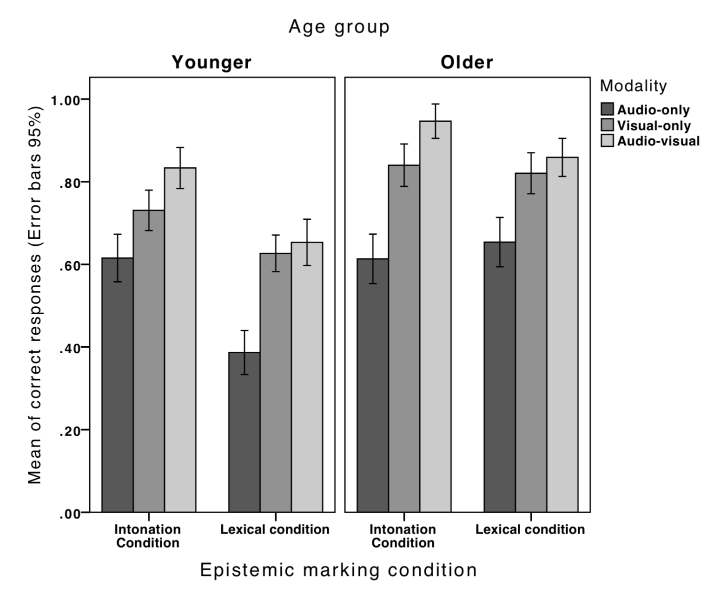


Figure 4. Mean proportion of correct responses to incorrect answers broken down by epistemic marking condition, modality of presentation and age group.

Table 2 displays the relevant means and standard deviation for each epistemic marking condition and modality of presentation in both age groups. The GLMM analysis revealed a main effect of age group, $F(1,294) = 21.215$, $p < .001$, with older children performing significantly better than younger children, and a main effect of epistemic marking condition, $F(1,294) = 10.064$, $p < .01$, indicating that when children were presented with the intonation condition they performed significantly better than when they were presented with the lexical condition.

Table 2. Means and standard deviation (*SD*) of the correct responses

Epistemic marking	Modality	Group					
		Younger		Older		Total	
		Mean	SD	Mean	SD	Mean	SD
Intonation condition	Audio-only	1.85	.88	1.84	.90	1.84	.88
	Visual-only	2.19	.75	2.52	.77	2.35	.77
	Audio-visual	2.50	.76	2.84	.62	2.67	.71
	Total	2.18	.83	2.40	.87	2.29	.86
Lexical condition	Audio-only	1.16	.80	1.96	.92	1.57	.94
	Visual-only	1.88	.67	2.46	.76	2.18	.77
	Audio-visual	1.96	.84	2.58	.70	2.27	.83
	Total	1.67	.84	2.33	.83	2.01	.90
Total	Audio-only	1.51	.90	1.90	.90	1.71	.92
	Visual-only	2.04	.72	2.49	.76	2.26	.77
	Audio-visual	2.24	.84	2.71	.67	2.47	.79
	Total	1.93	.87	2.37	.85	2.15	.89

There was also a main effect of presentation modality, $F(2,294) = 20.314$, $p < .001$. Pairwise contrasts showed that when children were presented with visual modalities (visual-only and audio-visual), they performed significantly better than when presented with the audio-only modality

(p < .001), with no difference between the two visual modalities (p = .052). Having the visual information present clearly helps the children to detect uncertainty better and thus confirms our first hypothesis that gesture has a bootstrapping effect on the child's comprehension of pragmatic meaning.

The model also reported a significant interaction between age group and epistemic marking condition, $F(1,294) = 7.751$, $p < .01$. Pairwise comparisons showed that only the younger children performed significantly better in the intonation condition as com- pared to the lexical condition (p < .001). All the other main effects and possible interactions were not significantly different. This confirms our second hypothesis, namely that younger children are able to detect epistemic meaning first through intonational cues before doing so through lexical cues.

2.4. Discussion and conclusions

The aim of this study was to investigate the role of intonational, lexical and gestural cues in the early development of epistemic understanding. Overall, the results of the comprehension task with 102 3- to 5-year-olds showed that children make great strides in their comprehension of uncertainty between the ages of 3 and 5, as seen by the fact that children in the older age group performed significantly better than those in the younger age group. These results are in line with previous studies that found that lexical understanding of uncertainty is achieved between the ages of 4 and 5. It is not surprising that younger children did not perform well in the lexical condition, since it has been documented across languages that children acquire the difference between different degrees of speaker certainty expressed through modal auxiliaries only around age 4 (e.g. Moore et al., 1990).

Yet the main question addressed by this study was whether younger preschool children attain epistemic understanding earlier through gestural and intonational features as compared with lexical features, and thus whether these features give them their first understanding of others' belief states. In the present study, by comparing three modalities of communication (audio-only, visual-only and audio-visual), it was possible to investigate the relative contributions of gesture, lexical and intonational cues to children's pragmatic comprehension. Our results showed that both younger and older children perform significantly better in both the visual-only and the audio-visual modality than in the audio-only modality. These findings are comparable with those of Armstrong et al. (2014), where facial gestures also seemed to scaffold children's performance in detecting belief state meaning (i.e. incredulity). By the same token, they are compatible with the growing consensus that gestures act as bootstrapping devices in language development in general (e.g. Butcher & Goldin-Meadow, 2000; Kelly, 2001; McNeill et al., 1994).

With respect to the contribution of intonation, the novelty of our study lies in the fact that our experimental methodology allowed for a direct comparison between the children's sensitivity to intonational vs lexical cues to uncertainty. Crucially, our results showed that 3-year-old children were more sensitive to salient intonational cues to uncertainty (in our case, a rising intonation pattern L* H%) than to lexical cues to uncertainty, regardless of whether visual information was also available or not. This result contradicts previously found results by Moore et al. (1993), who regarded prosody as playing a secondary role in children's acquisition of belief state meanings. Furthermore, our results show that 4- and 5-year-olds, by contrast, performed equally well in the audio modality in both epistemic marking conditions, showing that they have acquired an understanding of lexical cues by this age.

These results also seem to point in a different direction than previous studies on children's development of emotional prosody (Morton & Trehub, 2001; Nelson & Russell, 2011; Quam & Swingley, 2012), which have found that children's ability to match the auditory cues to the four basic emotions (happiness, sadness, anger and fear) seems to appear after children have acquired the lexical-semantic meaning of these emotions. Furthermore, when 5-year-old children are confronted with contrasting lexical cues or additional neutral situational cues juxtaposed on prosodic cues, they rely for their judgements on the lexical or situational cues rather than basing their judgement on the vocal cues encoding emotions. Thus, overall, this research would lead to the interpretation that prosodic cues do not seem to be prominent in the preschool years in leading children to detect emotional/attitudinal meaning in speech and that children rather use other cues to guide them (Aguert et al., 2010, 2013; Nelson & Russell, 2011; Waxer & Morton, 2011).

However, these studies deal with emotional prosodic cues (mostly pitch cues of contrasting pitch range) for inferring another person's emotional state, and these are weak prosodic cues not involving distinct pragmatic intonation patterns (Aguert et al., 2013; Quam & Swingley, 2012; Waxer & Morton, 2011). By contrast, our study has shown that 3-year-old children are sensitive to intonational contrasts involving final rise (H%) vs final fall (L%) distinctions for inferring speaker belief. In our data, developmental changes in children's comprehension of belief states become evident first in intonation (and in gesture) and only later in lexical marking. Thus our results seem to suggest that, regardless of the fact that mastering emotional prosody can appear later in development, intonational linguistic contrasts indicating complex pragmatic functions are probably mastered well before children acquire the lexical epistemic markers. Similar to the hypothesis of early *prosodic bootstrapping*, we contend that not only does prosody play a crucial role in the early acquisition of language by helping children to decode syntactic structure but that in later stages of development prosody and gesture both exert a different type of bootstrapping effect, namely, they facilitate the acquisition of pragmatic meaning. While our study shows that young children are able to understand epistemic meaning encoded through intonational and gestural cues earlier than through lexical cues, further steps need to be taken to prove whether prosodic and gestural abilities are predictive of later lexical acquisition, that is, whether there exists a direct correlation between early understanding of prosodic and gestural cues and the subsequently following lexical comprehension.

To summarise, the results of the current study suggest that not only gesture but also pragmatic prosodic patterns act as an integral part of the language-learning process at the intermediate stages of language development. These prosodic and gestural features can probably be claimed to act as bootstrapping devices in which children ground their early pragmatic development. We thus argue that early sensitivity to and acquisition of prosody and gesture patterns should receive more attention in developmental research in order for us to gain a more complete picture of children's pragmatic development.

3

Chapter 3: Children's signaling of their uncertain knowledge state: prosody, face and body cues come first

This chapter is adapted from:

Hübscher, I., Vincze, L., Prieto, P. (under review). Children’s signaling of their uncertain knowledge state: prosody, face and body cues come first. *Journal of Language Learning and Development.*

3.1. Introduction

In social encounters it is important for interlocutors to be able to assess and encode the reliability of transmitted information. Efficiently coding and decoding epistemic states is extremely important in a range of social contexts, from everyday interactions like asking for or giving directions to more formalized encounters like business meetings or courtroom statements. As they acquire language, children have to learn how to assess another person's degree of certainty and how to communicate their own level of certainty. Children's development of knowledge states has been studied from different research perspectives often without any interaction, in both written and conversational contexts, as well as in behavioral contexts. Research carried out within the field of linguistics and communication has predominantly focused on children's development of lexical and morphosyntactic markers of uncertainty, with the exception of a couple of studies which have investigated multimodal cues to uncertainty in school children (Swerts & Krahmer, 2005; Visser et al., 2014). On the other hand, studies in developmental psychology, specifically in the field of metacognition, have been interested in how children develop the ability to monitor their own mental states. While research has found that 2-year-old children are able to signal their ignorance first through behavioral (such as refraining from giving an answer) and gestural cues and shortly after through lexical markers (see Harris, Bartz, & Rowe, 2017 for a review), it is not clear how preschoolers' multimodal expression of their own degree of uncertainty develops and in what order the different cues appear. Drawing on both the metacognition and linguistics literature, the present study explores (a) whether preschool children display awareness of their own knowledge state, (b) whether they multimodally signal their epistemic stance by means of facial, gestural, or prosodic markers even before being able to signal it through lexical cues, and (c) how the multimodal expression of epistemic stance changes in the course of development. Below, we first review the literature on the development of knowledge monitoring, then move on to the role that prosody and gesture play in children's development of epistemic marking, and finally summarize the body of research on how adults and children inform about and signal their epistemic stance.

Children's signaling of their knowledge state

Research on metacognition has addressed the issue of when children become aware of their own level of certainty and are able to report on it (Ghetti, Hembacher, & Coughlin, 2013; Roebers, 2017; Schneider, 2008; Sodian, Thoermer, Kristen, & Perst, 2012; Whitebread et al., 2009 for a review of the different research paradigms). Here we will focus on the paradigm that looks at children's evaluations of their own knowledge states in the classic total/partial ignorance task (Pratt et Bryant, 1990), in which children are asked whether they know what is inside a box under conditions of total, partial, or no previous knowledge. For example, children are either told or shown what is inside (total knowledge) or they are not told or shown anything (no knowledge, or ignorance condition). Results demonstrate that in the certainty condition, children from age 3 onwards are able to accurately express verbally that they know what is in the box (Pillow, 1989; Pratt & Bryant, 1990; Ruffman & Olson, 1989; Tardif, Wellman, Fung, Liu, & Fang, 2005) and correctly signal their ignorance around the same age (Pratt & Bryant, 1990), or even by age 2 (Rohwer et al., 2012). Yet, the picture looks quite different in partial exposure experiments, where children are only exposed to a subset of the items they have to guess. In this case, children struggle until they are about 6 years of age to express their uncertainty verbally. For example, in the study by Rohwer et al. (2012), 4-year-old children were exposed to a set of different objects and then given a closed box containing only one of the objects. The children claimed to know which toy was in the box even though they could not really know for sure which of the several items it was. Only after 5–6 years of age did children correctly deny any knowledge in more than 80% of the trials, and these results remained constant regardless of how many items were included in the set of possibilities. Very young children also seem unable to accurately distinguish between actually knowing and merely guessing. For example, in similar experiments, when asked to indicate which of two boxes contained an object, children could only guess. However, when they guessed correctly, they usually claimed that they *knew* where the object was, and only when they did not guess correctly did they confess ignorance (Moore et al., 1990; Perner & Ruffman, 1995). This tendency to equate guessing correctly with knowing is observable until age 6. This immature understanding of partial knowledge can also be seen when children receive deliberately unclear or ambiguous instructions, such as “Hand me the tall box” when there are two tall boxes, yet judge them to be clear (Beal & Flavell, 1982; Robinson & Robinson, 1982).

Nonetheless, other studies argue that metacognitive abilities do not necessarily have to involve awareness of mental and cognitive processes and thus suggest that these skills appear much earlier than previously believed (see Ghetti et al., 2013 for a review). Studies using tasks that do not require verbal responses have found evidence of sensitivity to their own knowledge state in preschool-aged children and infants (Balcomb & Gerken, 2008; Bernard, Proust, & Clément, 2014; Call & Carpenter, 2000; Gerken, Balcomb, & Minton, 2011; Lyons & Ghetti, 2011; Paulus, Tsalas, Proust, & Sodian, 2014), and even in animals (Kornell, Son, & Terrace, 2007; Neldner, Collier-Baker, & Nielsen, 2015). For example, Call and Carpenter (2000) tested 2- and 3-year-old children using two tubes in one of which food or stickers were hidden. In the total knowledge condition the children saw the food or sticker being hidden, whereas in the partially ignorant condition they did not. When asked whether one or the other tube contained something, it took children much longer to make a decision in the partially ignorant condition. Other studies employing the 'opt-out' paradigm (in which child participants are allowed to skip trials when they are uncertain about answers, thus avoiding inaccurate responses) have also detected evidence of early metacognitive abilities (Balcomb & Gerken, 2008 among others). Further evidence of children's early competence has been found in ambiguous reference experiments. In such studies, even children who cannot determine whether an instruction they receive is ambiguous or has an indeterminate referent nevertheless react differently to such instructions, as seen in more frequent eye contact, puzzled expressions, or prolonged reaction times (Patterson, Cosgrove, & O'Brien, 1980; Plummert, 1996; Sekerina, Stromswold, & Hestvik, 2004). Also, it has been shown that 3-5-year-olds signal their knowledge state through behavioral cues, such as selectively skipping trials (Lyons & Ghetti, 2013) or seeking help when reporting they are less certain in a perceptual discrimination task, thus showing the ability to introspect on their confidence and use this introspection to guide their decision (Coughlin, Hembacher, Lyons, & Ghetti, 2015). Side by side with this, researchers have also started to study children's use of nonverbal cues to signal their ignorance. For example, it has been found that when 2-year-old children lack knowledge, they look towards the adult with whom they are interacting. They also convey ignorance via gestures like shrugs and open-palm hand gestures and only slightly later are able to verbally acknowledge their unknowingness (see Harris et al., 2017 for an overview). Similarly, Kim et al. (2016) measured how 3- and 4-year-olds monitor their knowledge state through giving verbal judgments and/or by using different gestural cues. In their experiment, they manipulated children's access to the contents of a box by granting them full access, partial access, or no access to the objects in order to see whether the children would display sensitivity to their own ignorance by refraining to answer when asked to inform another ignorant person. Afterwards, the children were asked to verbally give a judgment on their knowledge state. Importantly, the researchers investigated not only the children's decisions about informing a third person but also their use of nonverbal signals of uncertainty (such as tilting their head to one side, shaking their head, shrugging their shoulders, or looking away). The results showed that although children in both age groups were accurately reflecting their knowledge state in the partial and complete ignorance conditions by refraining from informing, they overestimated their knowledge state when asked to verbally report on it. However, interestingly, the 4-year-olds produced significantly more nonverbal signs of uncertainty than the 3-year-olds, thus displaying early monitoring skills, and gestured more in inverse relation to their degree of certainty.

Early production of gesture and prosody

Most literature from the field of metacognition indicates that children are not able to verbally express their uncertainty (in contrast to certainty and ignorance) until rather late (around 6 years), which raises the question of how children's lexical, prosodic, and nonverbal signaling of their (un)certainty develops over time. A number of studies report evidence that gesture comes first in the development of language. Children produce their first deictic gestures between 9 to 12 months, pointing to indicate objects in the environment, for example, well before they start producing their first words (Behne, Liszkowski, Carpenter, & Tomasello, 2012; Camaioni, Perucchini, Bellagamba, & Colonnesi, 2004; Liszkowski, Carpenter, Striano, & Tomasello, 2006). Also, once children start to speak, they produce gestures in combination with words (e.g., they point at a cup while saying "mine") and these gesture-word combinations generally precede the production of two-word combinations (e.g., "my cup") (Butcher, 2003; Capirci et al., 1996; Iverson & Goldin-Meadow, 2005; Özçalişkan & Goldin-Meadow, 2005). During developmental changes, gesture has also been shown to be an important tool to convey implicit knowledge of a concept that is just emerging. In other words, children who are on the edge of acquiring a new concept use gesture to convey information that clarifies or extends the information expressed in speech, such as when they are learning their first sentences (Butcher & Goldin-Meadow, 2000a) or solving difficult cognitive problems (Church & Goldin-Meadow, 1986; Perry et al., 1988).

Yet language development clearly does not end with the acquisition of the first words or the production of sentences. Children learn to manage more complex language skills such as communicating discourse functions and pragmatic messages. While most research has focused on children's early production of various types of gestures (whether representational, deictic, or conventional) and their precursor role in lexical acquisition, only a small number of studies have focused on the role of gestures in the acquisition of pragmatic functions. Some of these exceptions have investigated how children learn to express agreement, refusal, and negation (Beaupoil-Hourdel et al., 2015; Benazzo & Morgenstern, 2014; Guidetti, 2005). They found that the gestural modality is operational before the verbal modality, with children using conventional gestures such as head shakes and head nods to convey negation and affirmation before they learn to use the corresponding lexical strategies. In relation to knowledge state monitoring and in particular to children's monitoring of ignorance, as mentioned above, it has been found that 2-year-old children first signal their ignorance through gestures like shrugs and open-palm hand gestures before doing so verbally (see Harris et al., 2017 for an overview). However, little is known about whether children can use different types of gestures and other body cues to signal a partial knowledge state, and whether those cues act as precursors in children's development of uncertainty signaling skills.

There is increasing evidence that, side by side with gesture, prosody can act as a bootstrapping device in early language development (for an overview, see Esteve-Gibert & Guellaï, 2018). That is, certain types of prosodic features have been shown to guide children's initial acquisition of word order and syntactic structure (for a conceptualization, see Hirsh-Pasek, Tucker, & Golinkoff, 1996; see also Christophe, Nespor, Guasti, & Van Ooyen, 2003). In addition, there is also evidence that adult use of infant-directed speech, which is characterized by a slower speech rate and exaggerated pitch excursions, helps infants to build the phoneme inventories necessary in order to produce words (Cristia, 2011; Kuhl, Williams, & Meltzoff, 1991; Werker, Pons, Dietrich, Kajikawa, Kais, & Amano, 2007). Furthermore, in relation to intonation patterns, it has been found that early intonation patterns appear before the onset of combinatorial speech (Frota et al., 2016; Prieto, Estrella, Thorson, & Vanrell, 2012a). Yet, much less is known about children's production of complex pragmatic meanings expressed through prosody, and in particular it is not clear how children's signaling of their uncertain knowledge state through prosody develops.

Adults' and children's production of certainty and uncertainty markers

The study of epistemic stance, that is, a speaker's commitment to the truth value of the proposition communicated, has been predominately concerned with the *linguistic* strategies language users employ to position themselves towards knowledge, with traditional studies mainly focusing on the use of lexical or morphosyntactic features (Conrad & Biber, 2000; Cornillie, 2010; de Haan, 2001a; Dendale & Tasmowski, 2001; Heritage, 2012a, 2012b, 2013; Kärkkäinen, 2003; Marín-Arrese, 2011; Zuczkowski, Bongelli, & Riccioni, 2017 among others). Far fewer studies have adopted a *multimodal* perspective, whereby verbal resources are analyzed alongside voice, facial, and body signals to measure a speaker's commitment (Borràs-Comes et al., 2011; Krahmer & Swerts, 2005; Mondada, 2013; Roseano et al., 2016).

When it comes to children, studies on the communication of epistemic stance are even sparser and, again, have mainly looked at the development of lexical/grammatical markers. Previous research has shown that young children start using mental state verbs such as *I know* vs. *I think* to express certainty around age 3 (Diessel & Tomasello, 2001; Shatz et al., 1983) and that they start using modal auxiliaries such as *might* to express epistemic modality around 3;6 - 4 years of age (Papafragou, 1998; Shatz & Wilcox, 1991 for a review). These studies are based on naturalistic data mostly coming from English. Though one study (Lee & Law, 2001) analyzed naturalistic data from Cantonese children, the number of participants was very small (only three children in total). The authors of this study concluded that the direct evidence/certainty particle (*lo*) occurred much earlier than the uncertainty/inference particle *(gwaa)*. However, they also noted that while the participants' mothers used the certainty particle more than 1000 times, they only used the uncertainty particle five times, suggesting a clear influence of exposure. However, findings stemming from bigger and also more controlled data sets are still lacking.

As with the research on adults, the *nonverbal* expression of epistemic stance by children has received little attention, with a few exceptions. Krahmer and Swerts (2005) investigated how Dutch-speaking children (ages 7-8) and adults perceived and produced audiovisual cues to uncertainty. In a first experiment, they applied the Feeling of Knowing paradigm (Hart, 1965) to assess participants' display of their degree of knowledge when answering factual questions. While adults mainly used pauses, fillers (pauses filled with prolonged sounds like *uhh* or *mm*), rising intonation, raised eyebrows, head tilts, and funny faces, children mainly relied on pauses and rising intonation. In a second experiment, adults and children watched the responses that had been recorded in the first experiment and had to judge the speaker's level of uncertainty. In general, the adults' judgments about the recordings were more reliable than the children's. Furthermore, the adults also judged the older children's level of certainty more accurately than the younger children's based on their nonverbal signals. Also, overall, children in the experiment were better at detecting adults' expressed degree of certainty compared to other children's. Later, Visser et al. (2014) investigated the gestural and verbal expressions of uncertainty by children (8-11) using the Feeling of Knowing paradigm as a quiz game in either a collaborative or a competitive setting (two children working as a team vs. two children competing against each other). They found that uncertainty production increased between ages 8 and 11 and that only the older children were affected by the social setting, expressing their confidence more in competitive than collaborative contexts. Finally, Hübscher et al. (2017) tested the ability of 3- to 5-year-old children to detect speaker uncertainty from intonational, gestural, and lexical cues. The children, who were divided into two groups by age were asked to indicate by pointing which speaker was uncertain when answering questions. In a between-subjects design, the children were able to rely on audio-only (lexical or intonational cues), visual-only (gestural cues), or audio-visual (combinations of gestural + lexical or gestural + intonational cues). These results confirmed that overall, children are able to decode uncertainty states earlier in development when gestural cues are present. Additionally, younger children are better able to infer a speaker's uncertainty through intonational cues than through lexical marking (Hübscher, Esteve-Gibert, Igualada, & Prieto, 2017b). Taking this further, the present study aims to test the hypothesis that both prosody and gesture act as forerunners of pragmatic change in children's signaling of a partial knowledge state. This hypothesis will be tested by looking at children's production of uncertainty markers and then by performing a cross-sectional analysis.

The current study

Based on the findings by Hübscher et al. (2017b) and Kim et al. (2016), our hypothesis is that pragmatic gestures and also pragmatic prosody will act as precursors of children's expression of an uncertain knowledge state. To get a fine-grained picture of how uncertainty signaling develops over the preschool years, we designed an experimental guessing task based on Phan, Meza, Littlewort, Barlett, and Reilly (2010), whereby we manipulated children's epistemic access to items hidden in a box. Out of 10 objects in total, the children were allowed to see and touch five, while they were not allowed to either see nor touch the other five. The two groups of children (3-4.5 year-olds and 4.5-5 year-olds) were first asked "What is in the box?" Then, after coming up with an answer, they had to state the degree of certainty they felt about the belief they had communicated. By way of control, a group of adults also carried out the same experimental task.

The aim of the study was threefold. First, we wished to test the preschool children's monitoring patterns of knowledge state by asking them to self-assess their own (un)certainty. Second, we wanted to see whether the children could express their uncertain knowledge state earlier through prosodic and gestural means compared to lexical marking. And finally, we sought to determine whether the adults' marking of uncertainty through multimodal cues would differ from that of the children. We predicted that the children would (1) be more accurate in multimodally signaling their degree of (un)certainty than by self-reporting on it and (2) produce their epistemic stance first and more clearly through gestural and prosodic cues and only later through lexical cues. We also predicted that (3) the adults' multimodal expression of epistemic stance would be overall more complex than that of the children.

3.2. Methodology

3.2.1. Participants

A total of forty children (20 male and 20 female) were recruited at three Catalan public preschools in the metropolitan areas of Barcelona and Girona. Twenty children were 3-4.5 years old (*M* = 4.0, *SD* = 0.47) and 20 were 4.5-5 (*M* = 5.1, *SD* = 0.53). Parents were informed about the experiment's goal and signed a participation consent form prior to their children's participation in the study. Furthermore, language exposure questionnaires (based on Bosch & Sebastián-Gallés, 2001) were administered to the caregivers in order to ensure that the participating children were predominantly exposed to Catalan (as opposed to Spanish) on a daily basis (mean percentage of overall exposure to Catalan: *M* = 88%, *SD* = 0.128). Even though in all public schools in Catalonia the main language of instruction is Catalan, the target schools were chosen based on the high use of Catalan in the school's catchment area. An additional group of 10 students aged between 19 and 24 (*M* = 21.3%, *SD* = 1.72) from the Universitat Pompeu Fabra in Barcelona were recruited to serve as a control group. They were Catalan-dominant, reporting a mean daily usage of Catalan of 89% (*SD* = 0.09%). These adults signed a written consent prior to their participation in the study and were paid a small amount for participating. This study, including the consent procedure, was approved by the Ethics Board of the Universitat Pompeu Fabra.

3.2.2. Materials

Two sets of objects were used for the guessing game (Figure 1). Set 1 consisted of five common objects (top row in Figure 1: book, spoon, keys, eraser, pen) with which the participants were familiarized before the experiment, and Set 2 consisted of five less common objects (bottom row in Figure 1: tea bag, candle, tape, hazelnut, piece of bark) which the participants did not previously see.

A simple cardboard box measuring 15 cm × 15 cm × 25 cm was covered in wrapping paper for decoration and then two slots were cut at each end, each measuring about 5 cm × 10 cm. The slot on the opposite side, where the participant would sit, was covered with paper strips, so that the participant could easily put his/her hand inside the box to touch the object, but could not see what was inside.

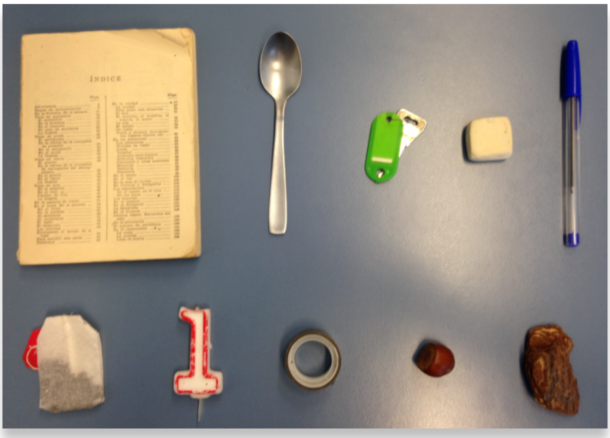


Figure 1. Picture of the objects used in the guessing game.
Upper row: Set 1. Lower row: Set 2.

3.2.3. Experimental Procedure

The basic procedure was as follows. A researcher and participant sat at a small table, facing each other, with the magic box in between. A video camera mounted on a tripod was placed behind the researcher and set to record the face and upper body of the participant throughout the experiment. On her lap out of sight of the participant, the researcher held a box containing the ten objects of Set 1 and Set 2. First, the researcher laid the five objects from Set 1 on the table and told the participant to handle and name each one. That done, the five objects were removed from sight once more. Before starting with the trials, the box was situated on the table and the children were informed that some of the objects that they would be touching inside the magic box would be new whereas other would bey the same as they had touched and seen before. The researcher then took one of the ten objects and placed it in the magic box in such a way that the participant could not see what she was doing. The researcher then asked the participant to reach their hand through the strip-concealed slot on their side of the box, feel the object, and say what they thought it was. This procedure was repeated for each of the ten objects, with objects being selected randomly by the researcher.

To pilot-test the suitability of the target materials, the experiment was carried out first on the ten adult control participants, one at a time, in a quiet room at the Universitat Pompeu Fabra. While these participants were able to guess the objects of Set 1, which they had previously seen, touched, and named, in 100% of the cases, they were able to correctly guess the unfamiliar objects of Set 2 only 50% of the time. Thus while all participants were certain about the Set 1 objects (book, spoon, keys, eraser and pen), they displayed various degrees of certainty when guessing the other five objects. This pilot study seeming to validate the procedure, the experiment proper involving children was then carried out.

The experiment involving child participants took place in a quiet room at their respective preschools, with each child tested individually. The procedure was identical to that followed in the pilot study. In each trial, immediately upon touching an object the child was asked by the researcher *Què és això?* ('What is this?'). Adult participants were asked the same question. However, in the case of the child participants, after supplying an answer to the first question they were additionally asked *Com de segur n'estàs? Molt, mig, o poc?* ('How sure are you: very, somewhat, or not very?). They were told that they could reply to this question either verbally or gesturally by holding out their arm high, at mid-level, or low. The adults were not asked this additional question because it was felt that they signaled their level of certainty reliably enough through their verbal answers, prosody, facial expressions, and gestures.

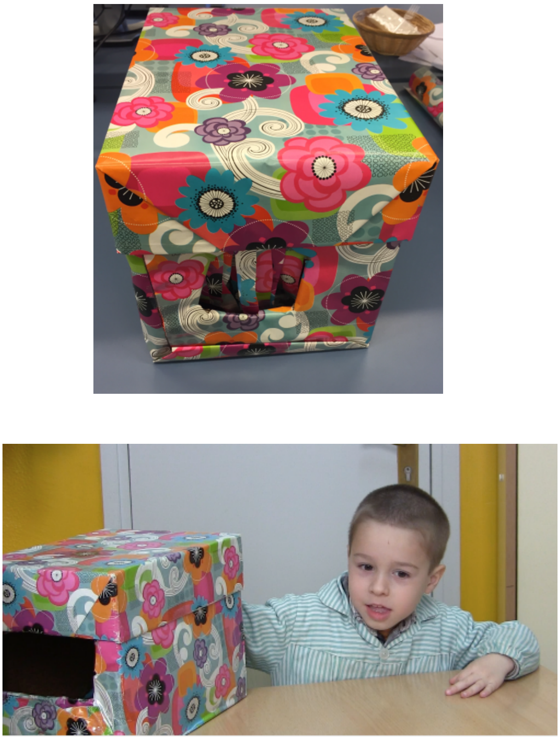


Figure 2. Pictures of the concealed slot on the participant's end of the "magic box" (left panel) and the open slot on the opposite end through which the researcher placed objects in the box (right panel).

The experiment lasted around 10 minutes. A total of 100 recorded responses were obtained from the adult control participants (10 adults × 10 items) and 400 recorded responses were obtained from the child participants (40 children × 10 items).

3.2.4. Data Coding

The resulting total of 500 responses were first given a binary score according to whether the participant had correctly guessed the object at hand (1 = correct, 0 = not correct). Next, in the case of the 400 responses by children, the child's self-reported degree of certainty (or epistemic stance) was given a binary score, with 'very certain' (*molt segur*) assigned a value of 1, while the two degrees of uncertainty ('somewhat certain' and 'not very certain') were subsumed under the same cover category of 'uncertainty' and assigned a value of 0. Note that when the children signaled complete ignorance of the object (i.e., expressed unknowingness), nonverbally and/or verbally, we did not ask them to report on their degree of certainty since they had not come up with a hypothesis. There were 33 such instances (16.5%) for the group of younger children but only two (1%) for the group of older children. These cases were excluded from further analysis since our study focuses on partial knowledge rather than ignorance.

Finally, the contents of the 500 audiovisual recordings were orthographically transcribed and labeled for their lexical, prosodic, and gestural information by means of ELAN (Lausberg & Sloetjes, 2009) by the first author for the children, and by the second author for the adults. The goal here was to have an external assessment of the participants' epistemic stance independent of their own self-reported input. The two coders met on several occasions during the initial stages of the coding process in order to look at examples, clarify doubts, and refine the coding system. In the final stages the third author was brought into the discussions as well.

Since one of the three goals of this study was to assess the contribution of lexical marking to epistemic stance relative to prosodic or gestural marking, the data were labeled lexically, prosodically, and gesturally, as follows:

Lexical coding. Here, the category 'lexical' comprises all lexical and morphosyntactic elements, as well as instances of 'thinking aloud', where participants appeared to be verbalizing their thoughts as they touched the object in the box. Thus, in the orthographic transcriptions of the recordings, note was made of all occurrences of epistemic adverbs (e.g., *potser* 'maybe'), epistemic verbs (e.g., *crec que* 'I think'), morphosyntactic cues like the conditional mood (*com una xinxeta podríem dir* 'like a drawing pin, we *could* say'), references to the thinking process (*ai no, està dur* 'oh no, it's hard'), and vague language (*una cosa així* 'a thing like'). Vague language has been defined as language which is fuzzy, general, and imprecise, has a low semantic content, and is heavily dependent on shared contextual knowledge for its meaning (Channell, 1994). People are vague either because they lack precise knowledge or because they lack the goal of being precise; in our experiment, the motivation was clearly the former. Such lexical markers were scored as either present (= 1) or absent (= 0).

Prosodic coding. The oral data were prosodically labeled following the Cat_ToBI system (Prieto, 2014). Only nuclear configurations were labeled because it is this part of the contour that typically conveys the pragmatic meaning of the utterance (see Ladd, 1996, among others). The rising and rising-falling nuclear pitch contours (L+H* L%, L* H%, L+H* !H%) were labeled as uncertainty pitch contours, while the falling pitch contours (L* L% , H+L* L%) were labeled as certainty pitch contours. Two other prosodic features were also labeled as uncertainty markers: fillers like *uhh* or *mm* and final vowel elongations such as in *un boliii* ('a pe-e-e-en'). The presence of one or more markers was scored as present (= 1) or absent (= 0).

Gesture coding. The coding scheme for face and body signals was agreed among the authors after a series of exploratory analyses of the two datasets (children and adults) (see Figure 3). It is based on Allwood's (2007) MUMIN, with some modifications, and on Ekman et al.'s (2002) Facial Action Coding System (FACS). To decide which of the array of facial and body cues that participants produced during the guessing task were specifically aimed at signaling uncertainty, we conducted a literature review. Note that when deciding which signals to regard as signals of uncertainty we did not restrict ourselves to solely intentional signals but also included signals that were most likely completely unconscious but which nevertheless offered information about the person's cognitive state. This is consistent with Morris (2002) view that:

> … what matters with gesturing is not what signals we think we are sending out, but what signals are being received. The observers of our acts will make no distinction between our intentional and our unintentional, incidental gestures. In some way, our incidental gestures are the more illuminating of the two, if only for the very fact that we do not think of them as gestures, and therefore we do not censor and manipulate them so strictly. (p. 21)

It is well known that the upper part of the face, namely eyebrows and eyelids, play a significant part in conveying various cognitive processes such as attention, reflection, concentration, or mnemonic effort (Ekman, 1979). Besides these meanings, the *eyebrows* can also communicate uncertainty and doubt. For instance, according to Eibl-Eibesfeldt (1974) raised eyebrows can convey meanings associated with doubt, questioning, and emphasis during conversation. Frowning is another eyebrow movement associated with uncertainty. According to Maatman, Gratch, and Marsella (2005), frowning and averted gaze are often linked to a speaker's communication of uncertainty, while Givens (2001) reports a variety of cues associated with uncertainty or doubt, including facial expressions (eyebrow frowns, eye movements, lip-pouting, lip-pursing), head movements (headshakes, head tilts), and gestures like palm-up open hand gestures, shoulder shrugs, and adaptors. These last, also called self-manipulators, are largely unconscious hand movements typically involving scratching, touching, or covering a part of the face or body which are performed without communicative intention and denote psychological discomfort and anxiety.

With regard to the *eyelids*, De Sanctis (1902) and later on Bitti, Bonfiglioli, Melani, and Garotti (2014) observed a marked reduction of the eye aperture and a tightening of the eyelids during mnemonic effort. The study conducted by Bitti et al. (2014) offered further proof that eyebrow raising and squinted (or tightened) eyelids are reliable behavioral cues signaling speakers' uncertainty. In a question-answer task, the authors differentiate between answers where speakers communicate their lack of knowledge; answers where speakers communicate their uncertainty; and answers where speakers, although uncertain, try to retrieve the information requested by the speaker. In the first case, there is no eyebrow-raising, in the second case the verbal answer is accompanied by eyebrow-raising, while in the third case together with the eyebrow-raising there also occurs a squinting of the eyelids and gaze aversion.

Although not an uncertainty marker *per se*, a co-speech signal that can sometimes contribute to conveying uncertainty is the *nose wrinkle*. Usually known in the literature as a cue of disgust in situations related to food and bad odors in general (Chapman, Kim, Susskind, & Anderson, 2009; Ekman & Friesen, 2003; Rozin, Haidt, & McCauley, 1999), this gesture typically implies a distancing from the disgusting situation or object. If we take certainty and knowledge as the ideals we all tend to in communication, wrinkling one's nose while distancing oneself from an unpleasant, uncertain situation can be interpreted as an uncertainty cue.

In general, the lower part of the face has received less attention in the literature focusing on the nonverbal communication of uncertainty. Among the few authors who investigated mouth configurations during uncertain answers, Krahmer and Swerts (2005) mention that both adults and children, when uncertain or unknowing, produced 'funny faces', that is, a kind of "marked facial expression". According to Krahmer and Swerts (2005), a face is perceived as funny if in its composition there is a 'funny' configuration of the lips, such as lip corner depression (AU 15), lip stretching (AU 20) or lip pressing (AU 24), manifested in combination with eyes opened wide (AU 5) and possibly some brow movement as well. In our data set, when uncertain, both children and adults pressed and/or stretched their lips, and when they had no idea at all (i.e., were 'unknowing'), they lowered the angles of their lips.

As far as head movements are concerned, *head tilts* have often been found to correlate with expressions of uncertainty and lexical repairs (Heylen, 2005; Lee & Marsella, 2006; Marsi & Van Rooden, 2007), indicating insecurity, shyness, or lack of knowledge (Allwood, 1998). *Headshakes* have been typically correlated with negation (Ekman & Friesen, 1969; Kendon, 2004; Robinson & Heritage, 2016), intensification, and inclusivity (Goodwin, 1980; McClave, 2001), and more recently with high commitment to beliefs (Vincze & Poggi, 2017). Here we noted that negation headshakes co-occurred with the formulation of a hypothesis and therefore interpreted such headshakes as indicating a high degree of uncertainty in the speaker.

In our corpus we also noticed a large number of *shoulder shrugs* performed by both children and adults. This signal—shoulders first raised and then going back down to their initial position—is a polysemous item which may assume, depending on the context, quite diverse meanings, such as obviousness, lack of knowledge, or non-commitment (because of either carelessness or powerlessness) (Debras, 2015; Debras & Cienki, 2012; Jokinen & Allwood, 2010). Sometimes, participants (both children and adults) shrugged their shoulders in the initial phase of the guessing, and then came up with a hypothesis in the final phase. Participants hence moved from an initial unknowing epistemic stance to an uncertain one.

In browsing our adult corpus, we noticed that, along with facial and head cues, adults also performed two types of hand gestures: metaphoric and iconic. The Palm Up Open Hand (PUOH) gesture (Cienki & Müller, 2008; Kendon, 2004; Müller, 2004) is a metaphoric gesture performed with palm open and turned upwards, and with fingers extended more or less loosely. The PUOH is a polysemous gesture as well which may take on different meanings derived from two basic domains of action: (1) giving, showing, or offering an object by presenting it on the open hand, or (2) displaying an empty hand to indicate the fact of not having something (Müller 2004). As pointed out by Calbris (2003) and Müller (2004), actions performed with the open hand are not limited to representing the manipulation of *concrete* objects but may also illustrate the handling of *abstract* concepts. Specifically, PUOH may represent the idea that an abstract concept is either *visible* (i.e., patently true) or *missing* (i.e., one is showing one's empty hands to the interlocutor). In other words, depending on the context, this gesture can signal either *obviousness*, or *lack of knowledge*. In the context of our experiment, participants' use of PUOH while guessing objects from the unseen category was clearly intended to communicate lack of knowledge.

A second type of hand gesture found in our adult data was iconic gestures. Participants used these gestures to represent the characteristics of the objects they were trying to guess. In the illustration of an iconic gesture in Figure 3 below, a participant mimics touching the pointed tip of a candle. As widely noted in the literature, iconic gestures seem to play a functional role in lexical access or word retrieval (Beattie & Coughlan, 1999; Krauss, 1998), so the participant may be performing the gesture here to help recall the word he is seeking.

As in the case of prosodic and lexical coding, the presence of one or more markers was scored as present (= 1) or absent (= 0).

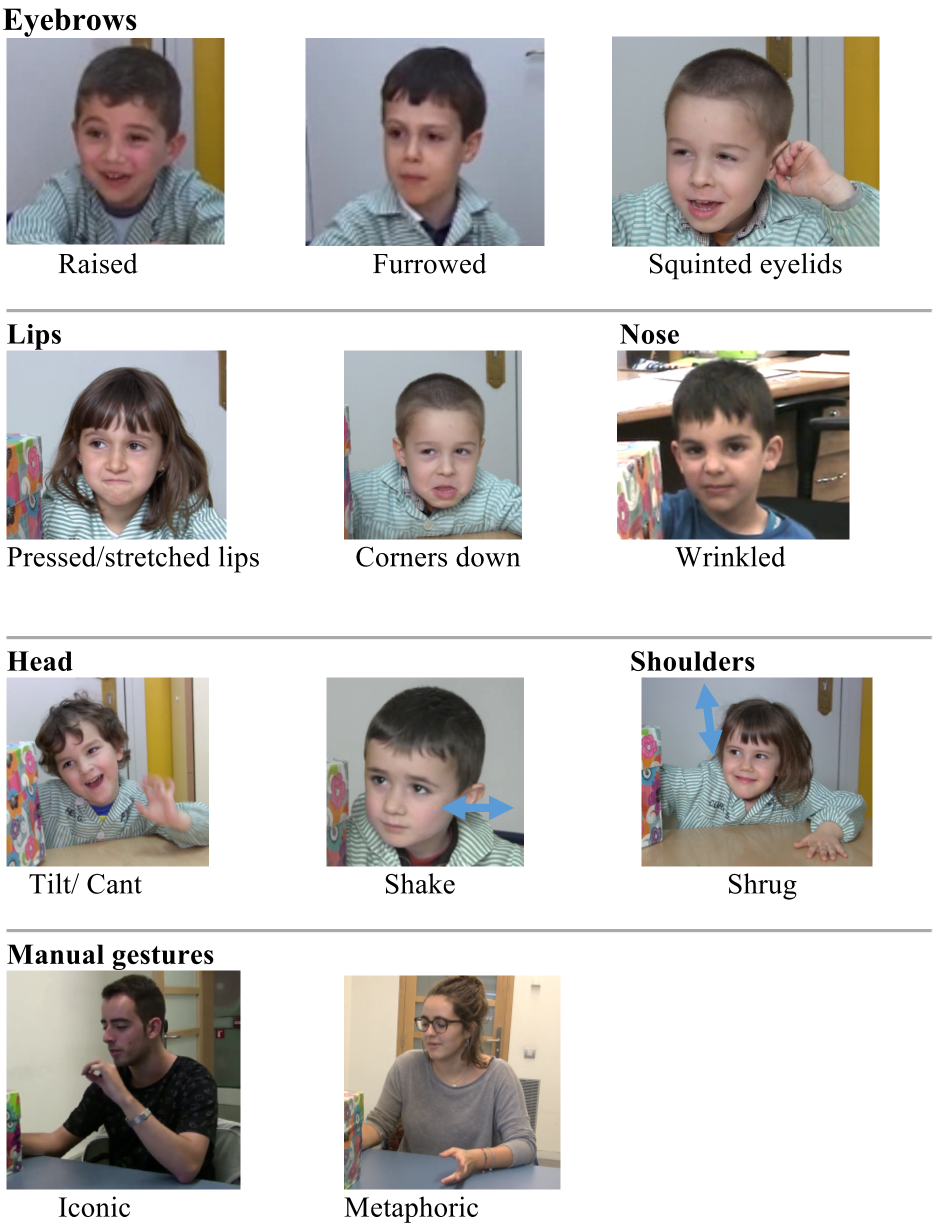


Figure 3. Categories of face and body signals annotated in the data set

Overall external epistemic stance score. The epistemic stance of the participant was assessed for each trial (10 trials per participant) and assigned one of three values: 'certain', 'uncertain', or 'unknowing'. This external judgment was based on all the information that had been conveyed multimodally, that is, it took into account the all lexical, prosodic, and gestural markers captured by the labeling processes described above. The 'unknowing' label was applied when a child participant failed to self-report a degree of certainty about his/her guess. As noted, these responses were excluded from further analysis.

Thus, for the 466 fully coded video-recorded responses that remained, a total of 13 tiers were annotated: (1) the *orthographic tier* containing the transcription of the verbal message and annotation of the lexical/morphosyntactic markers; (2) the *coder assessment tier* containing the coders' interpretation of the speaker's degree of certainty (certain, uncertain, unknowing); (3) the *self-assessment tier* (in the case of children only) containing the child's own reflection on his/her degree of certainty (high, mid-level, low); (4) the *guessed correctly* tier (yes or no); (5) the 'ToBI' tier containing the intonational transcription; (6) the *other prosodic markers* tier, where filled pauses and final vowel elongations were noted; (7) the '*eyebrow tier*' (raised, furrowed); (8) the *eyelid tier* (squinted/squeezed shut); (9) the *lips tier* (Pressed/stretched lips, lower lip forward, lip corners down); (10) the *nose tier* (wrinkled); (11) the *head tier* (tilted, shaking); (12) the *shoulder tier* (shrug); and (13) the *manual gesture tier* (iconic, metaphoric).

3.2.5. Reliability Test

An inter-rater reliability test was carried out to make sure that the two coders were consistent in terms of their epistemic coding of data as 'certain' or 'uncertain'. Twenty percent of the database (i.e., 80 items guessed by the children and 20 items guessed by the adults) was randomly selected, with care taken to ensure that objects from Sets 1 and 2 were uniformly represented across speakers. Three external raters (all members of the Prosodic Studies Group at the Universitat Pompeu Fabra) were asked to independently annotate this subset of the audiovisual recordings. The sections of video they were exposed to were edited so that they saw the children after they had come up with a hypothesis but before they self-evaluated the certainty of their guess, the idea being that the child's reckoning of their own certainty might bias the rater's judgment. The The Kappa statistic for rater annotations was obtained. Since four raters in total were involved (for each data set there was one official coder + three external raters), the Fleiss fixed marginal statistical measure was used. Fleiss equally arbitrary guidelines characterize kappas over 0.75 as excellent, 0.40-0.75 as fair to good, and below 0.40 as poor. The fixed marginal kappa statistic obtained for external epistemic assessment was 0.79 for the coding of data from the children and 0.86 for the coding of data from the adults. These scores thus reveal high agreement among internal and external raters for epistemic assessment.

3.3. Results

First, we analyzed the percentage of objects that were guessed correctly across the two conditions (objects from Set 1 vs. objects from Set 2), for all three age groups (children aged 3-4.5, children aged 4.5-5, adults). Results showed that for Set 1 the 3-year-olds guessed the objects correctly in 96% of cases, the older children in 94% of cases, and the adults in 100% of cases. For Set 2, on the other hand, both the younger and the older children only correctly guessed the objects in 22% of cases, whereas the adults guessed correctly in 44% of cases. The results clearly show that while both groups of children and the adults guessed the Set 1 objects in most cases, they both performed much more poorly with Set 2, even though, unsurprisingly, the adults performed better in this latter task.

Self and external assessment of children's knowledge state

Next, we analyzed the children's self-reported epistemic stance in relation to their correct and incorrect guesses of the target object, as well as the coders' external assessments in relation to the children's self-assessments. The goal here was to establish whether the children's being sure about their guesses correlated with guessing correctly. We also measured their correct guesses against the external coders' rating of their being certain or not. The data set from the children was also broken down by age group (younger or older) to see the effect of age on their performance.

As noted above, for both external and self-reported epistemic assessments certainty was scored as 1 and uncertainty was scored as 0. All the cases where the children did not come up with a hypothesis (when they were completely ignorant) were excluded from the data set. Figure 4 shows the mean proportion of certainty scores as rated by children (self-assessment) or coders (external assessment) across guessing success (0 guessed incorrectly; 1 guessed correctly), broken down by age group. The graphs show that while preschool children were highly accurate in self-assessing their own state (see the correctly guessed items), they are not accurate in self-assessing their own uncertain epistemic state (see the incorrectly guessed items), and that they tend to overstate their own knowledge in this case. Interestingly, in the case of older children's incorrectly guessed items there is a contrast between the less accurate self-assessment and the more accurate external assessment.

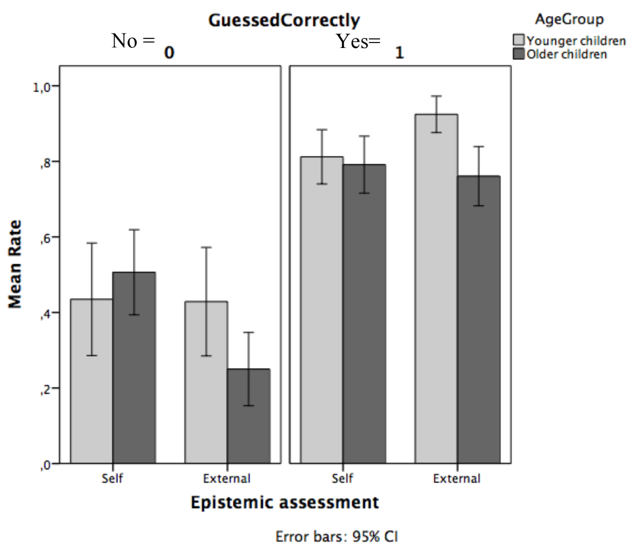


Figure 4. Mean certainty scores as rated by children (self-assessment) or by coders (external assessment), across younger and older children groups, and across guessing success

In a further step, a Generalized Linear Mixed Model (GLMM) with Mean Rate (Certainty rate) as a dependent variable (Binomial distribution, Logit link) was carried out using IBM SPSS Statistics v24 software. A random intercept was set for Participant. The following variables were set as fixed factors: AgeGroup (younger children, older children), GuessedCorrectly (0, 1), and EpistemicAssessment (self, external). The results revealed a significant main effect of GuessedCorrectly (F (1, 714) = 98,253, $p < .001$). Both self-reported and external ratings indicated significantly more certainty when the children guessed the object correctly than when they did not guess it correctly. Furthermore, there was a significant interaction between AgeGroup and EpistemicAssessment (F (1, 714) = 3,905, p = .049). Interestingly, in incorrectly guessed items, older children self-reported a higher degree of certainty than the external rater perceived (p < .022). This apparently contradictory result indicates that, at this age, children are not able to reflect and verbally report on their uncertainty state: they verbally state that they are certain of their answer, but at the same time their face and prosody convey uncertainty.

Children's use of uncertainty lexicon, gesture, and prosody

We then examined whether children's and adults' marking of uncertainty varied in relation to age and cue typology (lexical, prosodic, and body and facial cues). To do so we included all the responses which were assessed as uncertain by both children and raters coders, or by either one or the other. Figure 5 presents the mean proportion of lexical, prosodic, and gestural cues of uncertainty present in the data.

A GLMM with Mean Presence (or Proportion rate) as a dependent variable (Binomial distribution, Logit link) was carried out. The following effects were set as fixed factors: AgeGroup (younger children, older children, adults) and Cue (lexical, prosodic, gestural). A significant main effect of AgeGroup was found (F (1, 612) = 30,884, p < .001). Adults produced significantly more cues than the younger and older children, and the older children produced significantly more cues than the younger children. Furthermore, there was a main effect of Cue (F (1, 612) = 24,302, p < .001). There were significantly more prosodic cues than lexical cues (p < .001), significantly more prosodic than gestural cues (p < .001), and significantly more gestural than lexical cues (p < .001). No significant interactions were found between AgeGroup and Cue.

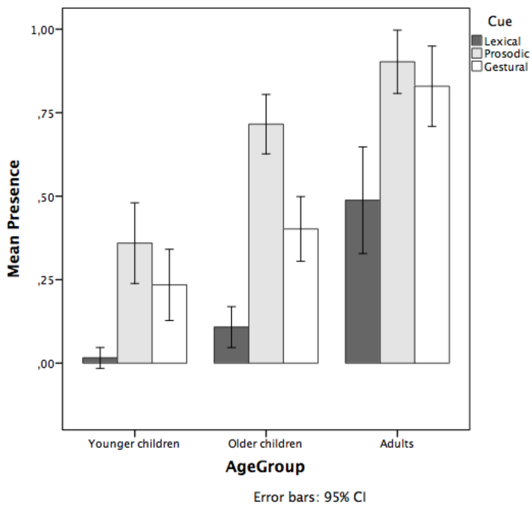


Figure 5. Mean presence of lexical, prosodic and gestural cues of uncertainty across the three age groups

A close inspection of the results shows, not surprisingly, that adults' multimodal communication of uncertainty—both lexical and gestural—is richer than children's. Adults not only employ more epistemic markers, whether lexical, prosodic, or gestural, but they also use a greater variety of cues. First, adults employ an array of lexical and morphosyntactic markers such as epistemic verbs or 'thinking aloud' talk which do not yet occur in the children's repertoire. For example, while younger children are not yet producing any lexical markers of uncertainty, in the older children's group only one occurrence of *sembla* 'it seems' (1% of the uncertain trials) and two occurrences of *crec que* 'I think' (2%) were found. By contrast, in the adult data 16 occurrences of lexical markers (44.44%) were found. In other words, adults employed a total of 16 lexical cues: besides the two mentioned above, they employed other epistemic verbs such as *he pensat* 'I thought' and *m'imagino* 'I imagine' and many cases of *no ho sé* 'I don't know'. Adults used the thinking aloud strategy in eight cases (22%) and the morphosyntactic strategy in 14 guesses (39%). For their part, children in the older age group employed the thinking-aloud strategy in five cases (6%), and one child in this group even employed the morphosyntactic strategy *podria ser una goma* 'it might be an eraser' (1%).

With regard to prosody, children in the younger age group extensively used the rising and rising-falling uncertainty pitch contours (22 occurrences, 67% of the trials) and the older age group used them even more (65 occurrences, 73% of the trials). Additionally, the younger children started to use fillers (6 occurrences, 19%) and vowel lengthening (3 occurrences, 9%), strategies increasingly put to use over the preschool years given that the older children used fillers in 29 trials (33%) and vowel lengthening in nine trials (10%). In contrast, the adults used uncertainty pitch contours 27 times (75%), fillers 17 times (47%), and vowel lengthening 22 times (61%).

However, it is important to highlight that the younger group of children (3-4.5 years old) successfully used target-like prosodic expressions of uncertainty such as rising and rising-falling pitch contours.

Regarding body and facial cues, children in the younger age group typically used their head and eyebrows to signal uncertainty. While younger children produced raised eyebrows in ten guesses (31%), older children produced them in 21 guesses (24%) and additionally produced furrowed eyebrows in two cases (2%). Also, both younger and older children produced head tilts (3 (9%) and 4 (5%) times respectively). The adults' use of eyebrows and head was more pronounced than the children's. The adults produced raised eyebrows in 16 guesses (44%) and furrowed eyebrows in 16 guesses (44%), while they tilted and shook their head in eight (22%) and seven guesses (19%) respectively. Furthermore, younger children produced shoulder shrugs in five guesses (16%), older children produced them in one guess (1%), while adults shrugged their shoulders in eight guesses (22%).

There are other facial signals such as nose wrinkles, squinted eyelids, and various lip movements (pressed/stretched lips and corners down) which were only encountered in the older children's group (but to a very limited extent: three occurrences of nose wrinkles, seven occurrences of squinted eyelids, seven occurrences of stretched lips, and six occurrences of lips corners down. Adults, on the other hand, wrinkled their nose in 11 guesses (31%), squinted their eyelids in 6 (17%), pressed/stretched their lips in 9 cases (26%) and pulled the lip corners down in 7 cases (19%).

Furthermore, adults produced a variety of hand gestures (four iconic gestures (11%) and three metaphoric gestures (8%)), which were not encountered in the child data. It is worth mentioning that the younger children did successfully use some facial/gestural markers of knowledge state, such as raised and furrowed eyebrows, head tilts, and shoulder shrugs, but only the older children are able to employ all the nonverbal signals that are in the adult data, like wrinkled nose, squinted eyelids and various lip configurations (pressed/stretched lips and corners down).
To sum up, while children start to signal their uncertainty at age 3 through prosodic and gestural cues before using lexical signals, over the preschool years their signaling of epistemic stance through multimodal signals clearly increases, however it is far from being fully developed at age 5.

3.4. Discussion and conclusions

To our knowledge, this is the first study to explore experimentally whether and how preschool children aged 3-5, as compared to adults, express uncertainty multimodally. The study presents a fine-grained analysis of both children's and adults' gestural, prosodic, and lexical patterns while controlling for awareness of epistemic stance. The three questions that the present study aimed to address were whether preschool children are able to multimodally signal their epistemic stance before are they able to reflect and report on the strength of their beliefs; whether children exploit prosody and gesture to signal their knowledge state before they use lexical items; and finally, how children's multimodal marking of uncertainty develops over the preschool years and how it differs from adult patterns.

First of all, the results of the children's self-reporting seen here confirm that young children tend to overestimate their own knowledge state, which is in line with previous research on children's difficulties in expressing their uncertainty in words (Kim et al., 2016; Rohwer et al., 2012). In cases where children assessed themselves as certain and guessed the object correctly, this corresponded with external assessments by adult researchers. In the cases where children did not guess the object correctly, both younger and older children reported themselves as being significantly less certain. However, whereas the older children were perceived as significantly less certain by the external raters as compared to their own self-assessment, this was not the case for the younger children. This might be an indication that younger children's uncertainty signaling is only starting to develop, as adult raters were not able to perceive any uncertainty markers, whether prosodic, gestural or lexical. The finding that older children self-reported a higher degree of certainty than was perceived by the raters is similar to that reported by Kim et al. (2016). While both studies reveal that 4–5-year-old children signal their uncertainty through gestural cues, our study shows that, alongside gestural cues, children also make use of specific prosodic cues before they are capable of verbally reporting on their uncertainty. This combination of gestural and prosodic cues allowed raters in our study to perceive the children as uncertain even when the children assessed themselves as certain.

This leads to the second research question, namely the relative amount of lexical, prosodic, and gestural marking in the expression of the uncertain stance across age groups. Our analysis of the utterances produced by 3- and 5-year-old children showed that 3-year-old children used prosodic and gestural cues to express uncertainty, but not yet lexical markers such as epistemic uncertainty adverbs. Comparing these results with previous studies on children's acquisition of uncertainty markers, the current study shows similar results to those in Swerts and Krahmer's (2005) study with 7- and 8-year-old children, but at a much earlier age than previously reported. Similar to previous studies, gesture also comes first in children's signaling of uncertainty, and the present results provide further evidence that gesture is also exploited early by children to mark different pragmatic strategies, as has been found for agreement, refusal, negation, and total ignorance (Beaupoil-Hourdel et al., 2015; Benazzo & Morgenstern, 2014; Guidetti, 2005; Morgenstern, 2014a) and also in children's narrative development (Demir, Levine, & Goldin-Meadow, 2015). Our study confirms and extends this general finding, revealing that multimodal signs of partial knowledge state precede the development of the ability to verbally express this state. Crucially, gesture in our study clearly goes hand in hand with prosody, which is used as an even stronger marker of uncertainty. Prosodic cues of uncertainty such as rising intonation, final vowel elongation, and filled pauses were all found to be widely used in children's guesses. Interestingly, based on a qualitative analysis of our data, we found that children seem to mark their uncertainty by manipulating their speech phonetically. In a future study it would be interesting to also take into account phonetic factors in children's marking of uncertainty, such as syllable duration and intensity, and also measures of voice quality. Doing so might lead to more robust results with regard to how children mark their epistemic stance through prosodic cues.

Comparing our results to previous studies on children's lexical cues to uncertainty, the current study shows a later appearance of such cues than pointed out by other previous studies (see Papafragou, 1998; Shatz & Wilcox, 1991 for modal auxiliaries); and (O'Neill & Atance, 2000 for modal adverbs). There are several possible explanations for this result. First of all, most previous studies have investigated children's lexical signaling of uncertainty by using naturalistic observations and often base their conclusions on a very small sample of children. As pointed out by previous studies (see e.g. Lee & Law, 2001), the input provided by caregiver speech might play a big role in children's production of uncertainty cues, something which has not been investigated in naturalistic investigations which are uncontrolled in nature and where children are in interaction with other adults. In our experiment, this factor was controlled for because the children were not exposed to adults' use of markers of uncertainty whatsoever. Second, since the current study tested children in one particular experimental setting, it cannot be ruled out that children might produce lexical cues to uncertainty earlier in other settings. On the other hand, the present findings suggest possible directions for developmental and more applied research, because they show that prosodic and nonverbal cues seem to be used as early markers of knowledge state, appearing before children are able to use lexical cues such as adverbs and modal auxiliaries. Comparing children's versus adults' use of gesture during speech, it would seem that gestural cues steadily develop over the preschool years and thereafter into adulthood. So while gesture together with prosody comes first in children's signaling of their uncertainty, not only does the use of gestural and body cues seem to increase significantly over the preschool years, but so does the variety of articulators used. There seems to be a trend in 3-year-old children to mainly use their head, shoulders, and eyebrows to mark their uncertainty, while other articulators, such as eyelids, lips, nose, and hands seem to be used only later on starting at 5 years of age and even more so in adulthood. In general, these findings are in line with the two previous studies on children's nonverbal signaling of uncertainty (Krahmer & Swerts, 2005; Visser et al., 2014). Those studies showed that children's expressiveness continues to develop between 7 and 8 years and between 8 and 11 and is characterized by the following verbal or visual features: filled pauses ('hmm' or 'err') and pauses in speech, rising intonation (typical of questioning), eyebrow movements, smiles, and marked facial expressions. Thus it does not seem surprising that children's multimodal expression of uncertainty increases significantly over the preschool years. Nevertheless, our results are novel in the sense that they take a detailed look at the individual gestural articulators to assess children's early multimodal expression of uncertainty. We have found that besides the eyebrows and an overall marked facial expression, the head too is employed early by children as an articulator to express uncertainty (by tilting it to the side). Also compared to Krahmer and Swerts (2005) study, which analyzed 'funny' faces created by a combination of facial movements, in the present study we analyzed facial cues separately, showing that while certain articulators such as the eyebrows were employed earlier by children, uncertainty facial marking involving the nose, lips, and eyelids increased in frequency over the preschool years. Finally, regarding prosodic cues, while Krahmer and Swerts (2005) found that fillers only play a marginal role in children's signaling of uncertainty, our data show that children employ them already very early and in 30 percent of the uncertain guesses in the older age group.

To sum up, the ability to signal uncertainty to an interlocutor is crucial in communication. Speakers signal their belief state by using multimodal strategies, and interlocutors need to take all of them into account when inferring a speaker's level of certainty. The gaining of such abilities is therefore an important step in children's pragmatic development The current study is the first to investigate the verbal and nonverbal development of epistemic marking over the preschool years, and it has revealed that both prosodic and gestural patterns are actively exploited in the first stages of the acquisition of the communication of epistemic stance. Children make use of prosody and gesture before they employ lexical cues to signal their stance, and also before they can accurately report on their feeling of knowing.

The results of the present study give rise to further interesting questions. While in the present study we were able to demonstrate that the ability to signal uncertainty through prosody and gesture appears earlier in children, it is not yet clear whether there exists any causal relation between these earlier appearing cues and the later appearing lexical cues. In a further study it would thus be of interest to test the facilitating role that prosody and gesture may play in children's uncertainty signaling. Furthermore, widening the analysis of uncertainty behavior to more natural and spontaneous contexts taking into account other social behaviors such as requesting for help and asking questions could offer a more comprehensive view of how children acquire the ability to signal uncertainty. Finally, it would also be of interest to study the social aspect of (un)certainty signaling through observing, for example, when children start to use uncertainty or hedging devices as a face-saving strategy in particular social contexts. One method to explore this might be to expose children to different interlocutors and see whether their multimodal signaling of uncertainty changed depending on the person they were talking to.

To summarize, our results suggest that multimodal competence plays an integral part in children's early development of epistemic stance marking, and might signal important upcoming changes in children's emerging pragmatic abilities

4

Chapter 4: Three-year-olds' infer polite stance from intonation and facial cues

This chapter is adapted from:

Hübscher, I., Wagner, L., & Prieto, P. (accepted). Three-year-olds' infer polite stance from intonation and facial cues. *Journal of Politeness Research Language Behaviour and Culture.*

4.1. Introduction

The ability to infer another person's interpersonal stance is crucial and central to our communicative interactions. In day-to-day interactions listeners are confronted with a complex set of cues happening simultaneously. Consider, for example, the utterance 'Pass me the bread'. The sociopragmatic meaning of this utterance can vary greatly depending on whether the speaker is being polite, friendly or impolite, that is, the stance that the speaker is taking. In order to infer a speaker's polite stance not only what is said but also how it is said needs to be decoded. Recent studies have highlighted the importance of the multimodal encoding of (im)politeness and the relevance of investigating prosodic, facial and body cues alongside the traditionally analysed lexical structure of communicative acts (Brown & Prieto, 2017; Culpeper, 2011; Culpeper, Bousfield, & Wichmann, 2003; Langlotz & Locher, 2017). In an investigation of a sketch staged in Monty Python's Flying Circus: Live from the Grill-O-Mat, Langlotz and Locher (2017) demonstrated how an English-speaking shop assistant switches from taking a positive and polite stance to taking an aggressive and impolite one by changing intonation and facial expression: "While negative stance is expressed through more stress and vocal force, positive stance is mediated through the higher pitched voice and rising intonation. This is usually accompanied by a smiling vs. a grim facial expression" (Langlotz & Locher, 2017: p. 313). Even though the importance of emotional cues in politeness research was noted already in Brown and Levinson (1987) seminal work on politeness, it is only recently that this relationship has been given serious attention within interpersonal pragmatics (for an overview see Culpeper, 2011; Langlotz &

Locher, 2013; Langlotz & Locher, 2017; Locher & Koenig, 2014; Locher & Langlotz, 2008; Spencer-Oatey, 2011). More specifically, as Langlotz and Locher (2017) observe, "emotions thus play an important part in arriving at emic judgments on relational work so that they have a clear place in the theoretical arguments of interpersonal pragmatics" (p. 315). Interpersonal relationships constitute the most important source for emotions. Andersen and Guerrero (1998, p. 64) argue that "the primary elicitor of emotions is interpersonal interaction". From an interactional perspective, the focus lies on external emotional 'representations' in communication, compared to the 'internal' psychological research perspective on people's emotions (Bänziger & Scherer, 2010). People often strategically induce emotional states in others as a way of achieving interpersonal goals (Andersen and Guerrero, 1998). This implies that when analyzing communicative acts of relational work, as Langlotz and Locher (2017) point out, emotional signals accompanying the messages should be considered. In the current study we set out to investigate whether and how preschool-aged children infer a speaker's polite affective stance (from now on referred to as polite stance) by considering the role played by intonational and facial cues. We will start out by reviewing previous research that focusses on children's ability to infer a speaker's emotional state through prosody and gesture, as well as the few existing studies that look at how children perceive politeness through intonation and prosody.

Early sensitivity to prosody and facial cues in the detection of emotion

While the present study focusses on how preschool children (aged 3) interpret affective meanings conveyed through prosody (specifically intonation) and facial cues, there is clear evidence that infants gain access to emotional meaning through various types of prosodic and gestural encoding at very young ages. Previous studies have shown that 3.5-month-old infants have a sophisticated ability to recognise emotion in facial expressions (Farroni et al., 2005; Kahana-Kalman & Walker-Andrews, 2001). Yet while infants in their first year of life seem to be able to distinguish emotional expressions, the ability to explicitly name such emotions only appears much later. There is evidence that already by two years of age children can appropriately match the name of at least some of these emotions with the corresponding facial expressions (Izard, 1971; Markham & Adams, 1992; Nelson & Russell, 2011; Widen & Russell, 2003, 2008). Furthermore, preschool children can recognize happiness, sadness, anger, fear and disgust on the basis of facial expressions (e.g. Denham, McKinley, Couchoud, & Holt, 1990; Gagnon, Gosselin, Hudon-ven der Buhs, Larocque, & Milliard, 2010; Gagnon, Gosselin, & Maassarani, 2014; Harrigan, 1984; Kayyal & Widen, 2015; McClure, 2000; Nelson & Russell, 2015; Székely et al., 2011; Tracy, Robins, & Lagattuta, 2005). However, other research suggests that learning to recognize expressions is a protracted process which lasts into the school years (see amongst others, Gao & Maurer, 2010; Herba, Landau, Russell, Ecker, & Phillips, 2006; Nelson & Russell, 2012; Roberson, Damjanovic, & Kikutani, 2010; Widen, 2013).

Though research on infants' and young children's sensitivity to emotions has largely concentrated on the identification of facial expressions, several studies have also shown infants' early sensitivity to (a) the emotional valences of speech as expressed through prosodic cues, and (b) the correspondences between prosodic and gestural cues in the expression of emotions. Mastropieri and Turkewitz (1999) showed that newborns react differently when presented with vocal cues expressing different emotions in their first language. Furthermore, 5-month-old infants have been found to react differently depending on the accompanying affective facial expression such as smiling in response to approval vocalisations (Fernald, 1993). They are also able to distinguish between sad and angry vocalisations when they are accompanied by matching facial expressions (Vaillant-Molina et al., 2013) and can match both positive and negative affective vocalisations with the corresponding face (Walker-Andrews & Grolnick, 1983). Results obtained by both electrophysiological and brain imaging studies have confirmed these results (e.g. Grossmann, Oberecker, Koch, & Friederici, 2010; Zhang et al., 2014). Nonetheless, the above-mentioned studies have either used prosody discrimination tasks or facial expression-prosody matching tasks, neither of which really show that infants are attributing affective stance to the speaker. More helpful in this respect are studies that include behavior regulation tasks with toddlers. For example, the results of Mumme, Fernald, and Herrera (1996) study using a novel toy paradigm with 12-month-old children showed that fearful emotional prosody conveyed by meaningful utterances was sufficient to elicit an appropriate behaviorial reaction in children. Similarly, Vaish and Striano (2004) demonstrated that 12-month-old infants could be persuaded to cross a visual cliff purely on the basis of positive vocalisations. However, the target utterances used also contained semantic meaning and in fact there are few studies that allow us to conclude that prosody was the only factor at work in emotion recognition tasks. One example is Hoicka and Wang (2011), which showed that 15-month-olds were able to make prosody-meaning associations independently of situational and lexical context. In this study, the experimenter produced either humorous or sweet vocal cues, the only difference between the two utterances being the mean values for pitch, with higher values in the humorous utterances. After infants were exposed to the vocal cues, they observed the experimenter performing either matching or mismatching behavior (i.e., either a humorous action or a 'sweet' action), with a neutral facial expression in all conditions. The infants looked longer at the mismatching behavior than at the matching behavior. These results showed that 15-month-olds are able to develop expectations about an adult's behavior exclusively on the basis of prosody. Finally, some studies have also assessed infants' early ability to express their own emotions through differentiated prosodic patterns and facial expressions (Jhang & Oller, 2017; Oller et al., 2013; Papaeliou, Minadakis, & Cavouras, 2002; Scheiner, Hammerschmidt, Jürgens, & Zwirner, 2003). Thus, infants not only are able to perceive emotions early, they also learn to express affective cues themselves through through protophones and facial affect valences within the first 3 months of life (Jhang & Oller, 2017).

Nonlinear development of emotional and pragmatic comprehension in preschool years

Though the above-mentioned results point to an early (and parallel) development of infants' understanding of emotional meaning encoded through both prosody and facial cues, this seems to change as children get older and start acquiring the lexical and grammatical features of language. There is some evidence that infants' and toddlers' use of prosodic and facial cues begins to change once they start to develop a lexicon. Friend (2001) investigated 15- and 16-month-olds' sensitivity to prosody as well as facial cues vs. lexical content. Before the children were given the opportunity to play with a novel object, they saw videos of a speaker who conveyed either an approving or disapproving message. The behavior of those children who understood the lexical meaning of the message was better regulated through lexical content than through prosody or facial cues, making receptive vocabulary a significant predictor of the children's behavior. This led Friend to conclude that there is a transition stage from affective to linguistic meaning around 15 months of age. A similar tendency was found by Lawrence and Fernald (1993). In their study, while 9-month-olds were better regulated through tone of voice than through lexical content, 18-month-olds were better regulated through lexical content.

Furthermore, it has been found that when presented with multiple cues to emotion, young children tend to rely on the lexical cues when prosodic cues and the lexical message conflict (Friend, 2001; Friend & Bryant, 2000; Morton, Trehub, & Zelazo, 2003). For example, Morton et al. (2003) carried out three experiments with 4- to 10-year-old children and adults. In experiment 1, children were exposed to 20 sentences with matching or mismatching propositional information (such as *My mommy gave me a treat* vs. *I lost my sticker collection*) and paralinguistic cues (happy vs. sad prosody). 'Happy' prosody sentences were produced with higher pitch level, greater pitch and loudness variation and a faster speaking rate, while 'sad' prosody was produced with lower average pitch, attenuated pitch and loudness variation and slower speaking rate relative to the happy paralanguage. In experiment 2, in order to remove any possible interference from the propositional content, the sentences were presented in a foreign language (Italian), and in experiment 3 the sentences were low-pass filtered to remove potential distraction from the semantic content. The overall results of these experiments showed that 9- to 10-year-old children and adults judged the speaker's feelings by *how* s/he spoke whereas children aged 8 or younger judged the speaker by *what* s/he said. However, results from experiment 2 and 3 showed that even 4-year-old children were able to attribute an emotion to the affective paralanguage of both the foreign language and the low-passed filtered speech, although accuracy improved with age. Waxer and Morton (2011) confirmed this lexical bias in 6-year-olds, since the children in their study proved to be inflexible in their interpretations of conflicting speech cues when they had to decide between emotions based on words and emotions based on prosody (happy/neutral/sad paralanguage). Later, Aguert et al. (2013) suggested that it is not that lexicon or context invokes a bias but rather that prosody plays a 'subordinate' role when it is in competition with situational context, a phenomenon that persists even into children's early teenage years. In their study, Aguert et al. (2013) presented 5- to 13-year-old children with a judgment task in which they were exposed to animal figures located in neutral situations and then asked them to judge the emotional state (happy or sad) of the animal on the basis of prosody alone since both lexical and contextual cues were devoid of emotional valence. The utterances the children heard while viewing the animals were five syllables long and deliberately made unintelligible, with the syllables being randomly mixed. The prosody employed was described as either positive (happy) or negative (sad). The children were then asked to judge the animal's emotional state by pointing to a drawing with either a happy face or a sad face with which the children had been previously familiarised during a pretest. They were also asked to verbally explain their judgment. The results of the experiment showed that the youngest children in the study (the 5-year-olds) struggled to infer the speaker's emotional state on the basis of prosody alone and only performed at chance level. The authors concluded from this that there are no specific biases involved but that prosody is simply a difficult cue for preschool children.

These above-mentioned results stand in stark contrast to other developmental findings showing that from very early in development (and also later) children are extremely sensitive to prosodic cues, which in fact act as bootstrapping mechanisms for language development (see Hoehle, 2009 for a review). In addition, a number of studies have shown that toddlers can regulate their behavior depending on prosodic cues alone (see Hoicka & Wang, 2011, amongst others). Why do infants clearly detect and respond to emotional cues in infancy and early childhood (and in a variety of tasks), but then struggle to detect prosody in the preschool years? Apart from the fact that at later ages children need to integrate information coming from prosody with other contextual and lexical cues, some authors have suggested that while happy and sad contours may be accessible to babies in infancy, they may lose their iconicity through reinterpretation during the language acquisition process, and the late learning of connections between pitch and emotion could be due to the 'complexity of pitch-contour patterning in the language as a whole' (Quam & Swingley, 2012). Also, as Aguert et al. (2013) note, the fact that the few studies available reflect a variety of experimental designs is not ideal and has really yielded only fragmentary knowledge about the ability of children to detect emotional stance from prosody.

Regarding the comprehension of meanings through facial cues, there are number of papers that have examined children's ability to recognise emotion through both facial cues and affective prosody (amongst others McCluskey & Albas, 1981; Nelson & Russell, 2011; Nowichi & Duke, 1994; Quam & Swingley, 2012). In particular, several studies have demonstrated an advantage in the preschool years compared to vocal expression. For example, Nelson and Russell (2011) tested preschool children's (3-5 years old) ability to label emotions (happiness, sadness, anger and fear) based on video clips which were produced in four different conditions: face-only, body posture-only, voice-only and multi-cue (i.e., face + body + voice). Results showed that preschoolers were equally able to provide labels in the face-only and multi-cue conditions, yet they were significantly less accurate in the body posture condition. However, the children performed worst in the voice-only condition. Furthermore, Quam and Swingley (2012) found that while 4- and 5-year-olds were consistent in detecting happy or sad prosody (created by manipulating the low-pass filter on the audio recording) to decide whether a puppet had succeeded or failed at a task, 2- and 3-year-olds depended more on facial-gestural and body language cues. Clearly there remains considerable room for further exploration of young children's sensitivity to affective stance on the basis of prosody, facial cues and lexical information, and the interaction and relative importance of these factors.

Children's sensitivity to polite stance

Children's multimodal acquisition of politeness in requests is greatly under-researched and little is known about children's ability to infer a speaker's polite stance on the basis of intonation and/or facial cues only. Also, at this moment still surprisingly little is known about children's sensitivity to a speaker's polite stance and the role affect plays therein. Focussing on first grade children's and adults' understanding of the sociality of emotions, Camras, Pristo, and Brown (1985) study showed that children expect there to be a relationship between speaker affect and directive choice. Particularly, angry speakers are expected to be less polite than happy or neutral speakers. The authors concluded from this that adults and children sometimes make inferences about a speaker's emotions based on his or her directive choice. In another study, Shochi et al. (2009) focussed on children's understanding of politeness meaning from a multimodal point of view. Investigating 9- and 10-year-old Japanese children, they found that facial cues were beneficial for the processing of politeness and impoliteness meanings (Shochi et al., 2009). They also found that, in comparison to adults, children relied on facial cues earlier than auditory information to understand politeness meanings. One of the only and most comprehensive studies with a focus on preschool children's developing perception of politeness by including intonation as a cue is Bates' (1976) study. Bates (1976) experimentally tested whether 60 Italian children aged 3-6 could perceive politeness as encoded through lexical cues or through prosody only. The children were asked to judge which frog puppet made the most polite request as the experimenter varied the puppet's use of different lexical cues and tones of voice. Tone of voice varied from harsh to gentle and the verbal message *dammi un dolce* 'give me a candy' was either accompanied by *per favore* 'please' or not. The results showed that the children had acquired *per favore* 'please' as a politeness marker by age 3 but the use of gentle intonation as a strategy only reached significance after 4 years of age. Nonetheless, Bates argued that the younger children may have judged the harsh tone as 'nicer' because they found it amusing, hinting at an earlier sensitivity to intonation after all.

While at this stage it is not clear whether preschool-aged children can infer a speaker's polite stance from prosodic and/ or facial cues only, our hypothesis is that by age 3 children will be able to successfully detect a change in polite stance by only accessing changes in these specific cues. Despite contradictory results in the literature on preschool children's sensitivity to emotional states encoded through prosody and facial gesture, we claim that the use of a child-directed pragmatically relevant task will allow us to adequately assess this issue.

The current study

The current study intends to determine whether 3-year-old children would be able to distinguish a polite stance from a non-polite stance in requestive speech acts exclusively on the basis of prosodic (in particular intonation) cues, solely on facial cues and on the basis of the two combined, removing any possible lexical or contextual bias. To do so, we compared children's behavior on a between-subjects polite stance comprehension task where a set of requests were presented to the child which always contained the same polite stance lexical cue (the word *please*, e.g., '*Can you give me the ball, please*') and the same neutral speech act situation as a situational prompt. We adapted Bates (1976) experimental procedure whereby the subject is asked to give an object to the person who seems to be asking more nicely. Crucially, in order to investigate children's sensitivity to intonational patterns (and facial cues), the experimental materials were presented in three different between-subject modalities: (a) audio-only (AO), with just verbal and prosodic cues available; (b) visual-only (VO), with exclusively nonverbal cues available; and (c) audio-visual (AV), with both verbal and nonverbal cues available. This is an adaptation of the methodology used in Hübscher et al. (2017b) and originally created by Armstrong et al. (2014). Importantly, we chose to use intelligible speech but to keep the lexical cues constant in both the polite affect and the non-polite affect conditions, thus controlling for lexical content. Both intonational and facial cues were varied, with a falling nuclear configuration (L* L%) and frown signaling the non-polite affect condition, and a rising nuclear configuration (L+H* H%) and smile signaling the polite affect condition.

Based on the previous literature, we had several hypothesis regarding children's ability to access a speaker's polite stance through intonational and gestural information. Despite incongruent findings in the previous literature, we hypothesized that the use of a child-directed and pragmatically relevant task would allow us to adequately assess children's ability to infer meaning based on facial cues and intonation only and the combination of both. First, we hypothesized that, in line with Quam and Swingley (2012) and Nelson and Russell (2011) children should be able to detect a speaker's polite stance best in the AV condition, where both intonational and facial-gestural are cues present. However, contrary to some other previous research, and more consistent with the early exploitation of pitch and facial cues for communicative functions, we hypothesized that both intonation and facial cues are strong cues in the age under investigation, just as in earlier ages, and children should be able to infer to a certain degree another person's polite stance through those cues in the AO and VO conditions.

4.2. Methodology

4.2.1. Participants

Thirty-six (18 female and 18 male) American English-speaking children participated in a between-subjects polite stance comprehension task. Five additional children had to be excluded for various reasons (they were bilingual, had developmental problems or dropped out of the experiment). Their ages ranged from 2;10,22 to 4;0,16 (mean age: 3.51 years, *SD* = 0.31). The participants were mostly high socioeconomic status white children visiting the Center of Science and Industry, a science education center in Columbus, Ohio (Wagner et al., 2015). Parental consent was obtained before the experiment. Children were given an animal-shaped stamp on the hand as a reward for participation. All research practices and consent forms were reviewed and approved by the Ohio State University Social and Behavioral Sciences Institutional Review Board.

4.2.2. Preliminary elicitation study with adults

In order to create the stimulus materials to be used in the study, it was first necessary to determine the prosodic and facial cues characteristics that are used in American English to encode polite and non-polite stances. To this end, a free discourse elicitation task was carried out with ten US English native speakers who were students or researchers at Ohio State University. The participants were asked to imagine themselves in the company of a child and read a set of context prompts in which they had to make a request to that child. To give them a better idea of the target age group, next to the text describing the context the participants were provided with a picture of a 3-year-old child. Two examples of these contextual prompts are given in (1) and (2) (see Appendix B for all prompts). In (1) the context prompts the speaker to ask for help in a polite way (i.e., a polite affect request) while context (2) prompts the speaker to produce a command rather than a question (i.e., a non-polite affect request).

(1) **Polite stance condition**: You have both your hands full of plates on your way to the kitchen and you have just dropped a fork. Ask the child nicely to give you the fork.
(2) **Non-polite stance condition**: The child is very excited and plays continuously with a noisy toy but you want him/her to be quiet. You're quite annoyed. Tell the child to give you the toy.

Video recordings were made of the participants as they responded to the prompts, yielding recordings of a total of 100 requests (10 requests × 10 speakers). Analysis of the recordings showed that the lexical structures used most often by speakers were the conventional request questions *Can you give me xy?* (with or without *please*) and *Could you give me xy?* (with or without *please*) in the polite affect condition and the imperative *Give me xy* (with or without *please*) in the non-polite affect condition. The recordings were then analysed for prosodic content using MAE_ToBI (Beckman, Hirschberg, & Shattuck-Hufnagel, 2005). MAE_ToBI is an annotation system used for labeling intonation and prosody in databases of spoken Mainstream American English (for a consensus account of English intonation and prosody on which the MAE_ToBI system is based (see Beckman et al., 2005). The frequency distribution of the types of intonation contours displayed in the non-polite vs. polite affect requests can be seen in Table 1.

Table 1. Prosodic cues displayed in the requests

	Non-polite stance	Polite stance
L* L%	49 (98%)	7 (14%)
L* H%	0	5 (10%)
L+H* H%	1 (2%)	38 (76%)

The results of the prosodic analysis showed that there was a clear preference for a falling (L* L%) nuclear configuration in the non-polite affect condition and a rising (L+H* H%) nuclear configuration in the polite affect condition. Finally, the recordings were analyzed in terms of the facial cues that speakers had used. It was found that while in the non-polite affect condition participants displayed a stern facial expression, in the polite affect condition they consistently displayed a polite smile.

On the basis of this analysis, it was decided that in the subsequently prepared experimental stimulus materials the non-postive affect condition would be characterised by the L* L% nuclear configuration prosodically and a stern facial expression, while the polite affect condition would be characterised by the L+H* H% nuclear configuration and a smiling facial expression. Regarding the lexicon, we decided to add *please* at the end of the conventional request structure *'Can you give me xy'* since it is a politeness cue that is usually taught to children early on in order to make a request sound acceptable. In order to control for lexical input, the actual utterance would be the same in both polite and non-polite affect conditions.

4.2.3. Experimental materials

Six female students or researchers from Ohio State University volunteered to take part in preparing the stimulus materials. They were individually videotaped while producing a set of sentences with the same target structure *'Can you give me the [toy] please'*. The name of one of six toy items, namely *ball*, *shark*, *frog*, *bear*, *horse* and *duck*, was inserted in the sentence with a different toy per speaker. The speakers each produced one request in two conditions, namely non-polite and polite stance, in which they replicated the main prosodic and facial-gestural features of the non-polite and polite stance conditions described in section 2.2.1. A total of 12 sentences were obtained (1 target structure × 1 toy × 2 affect conditions × 6 speakers).

The four panels in Figure 1 illustrate the pitch contours and facial expressions used in the two AV conditions.

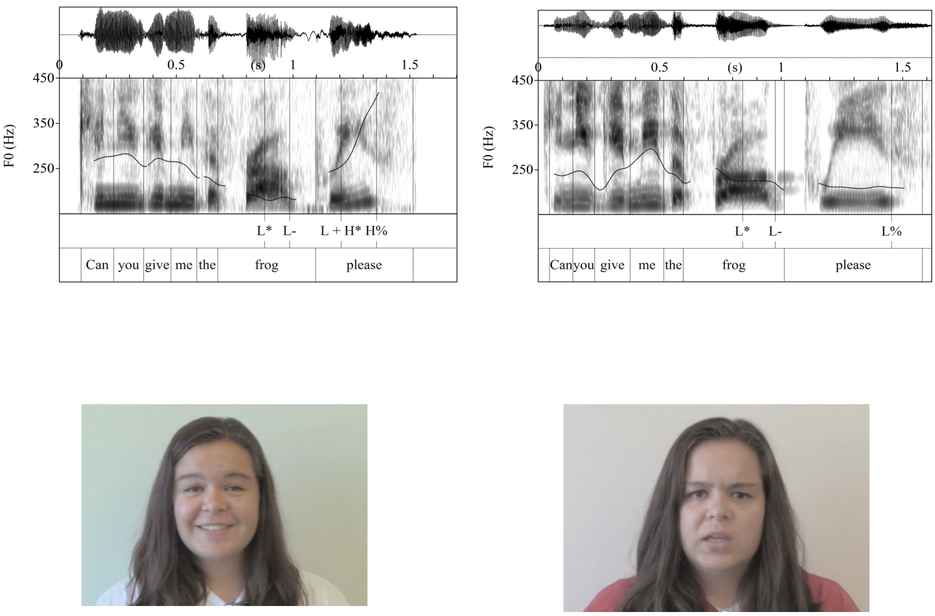


Figure 1. Pitch tracks, spectrograms and waveforms (top panels) and facial-gestural cues (bottom panels) for the polite affect request (left panels) and non-polite affect request (right panels) stimuli used in the AV condition.

In order to be sure that our stimuli were clearly distinguished as portraying polite stance vs. non-polite stance, we performed a preliminary control survey with the online survey platform SurveyGizmo. Ten adult American English speakers were recruited and were asked to judge each of the six paired video recordings in terms of which one of each pair depicted the more polite request. Their responses were 100% consistent across raters.

PowerPoint presentations (PPTs) were then created for each of the three between-subjects experimental conditions (AO, VO and AV). Each PPT contained eight slides in the following order: a familiarization slide, three test trials, the familiarization slide again, and finally three more test trials. In the AV version of the PPT, the full audiovisual recordings of the speakers were embedded in the trial slides. The PPT presentation for the VO condition was identical except that the audio track was eliminated from each of the six videos. Thus, the children would only be exposed to facial cues without any audio cues. Finally, the PPT presentation for the AO condition included the audio track in the trial slides but instead of video content subjects were shown two grey squares under which they could merely see the outline of a face, this was done to remove any possible information from facial cues but enable participants to associate the sound with a person. In order to counterbalance the presentation order of the stimuli within each trial, we created four different PPTs for each modality of presentation.

4.2.4. Procedure

First, each child was randomly assigned to one of the three conditions, namely AO (12 children), VO (12) and AV (12). The child was seated next to the experimenter, a native American English speaker, on the floor in a quiet room in the Center of Science and Industry Museum. As a warm up, the child was asked to name each of the six toy items that would be seen in the experimental trials, namely a ball, shark, frog, bear, horse and duck. For the experiment proper, the child faced a laptop on which the polite stance comprehension task was presented with two empty buckets placed between the child and the computer.

To familiarize him/her with the test procedure, the child was told that s/he was going to play a game. Two animated stars appeared at the left and right of the screen. After the star on the left said 'I'm the blue star', the experimenter passed a fuzzy ball to the child and asked *Can you put this fuzzy ball in the bucket right in front of the blue star?* The same procedure was then followed for the red star on the right. Then the child was presented with the test trials. In each trial, the screen showed two embedded videos (in the AV and VO condition PPTs) or still pictures (in the AO condition PPT) of two speakers, one to the left, the other to the right. Because it was the same speaker in both videos or pictures, the experimenter referred to them as 'sisters that look very much alike' (see Figure 2). Each trial slide also depicted one of the target toy items in the top right or left corner of the screen (e.g., in Figure 2, a ball at top right).

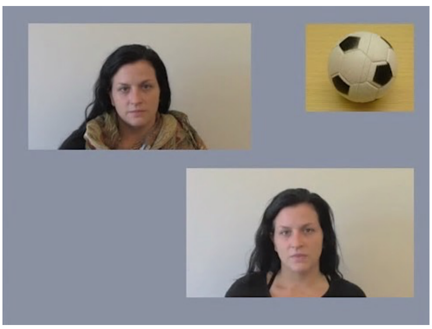


Figure 2. Example of a test trial PPT slide, showing position of stimulus recordings or photos and one of the toy items. In this case, the 'sisters' both portray a neutral expression because the slide is from the audio-only PPT.

The experimenter then directed the child's attention to the toy item by saying *What do you think they want? Both of them are going to ask you to* *give them the ball. You have to listen/watch very carefully. Can you give the ball to the person who asks more nicely?* The two embedded videos (or audio tracks in the AO condition) were then played consecutively, one in which the speaker made a request while displaying the polite affect cues and the other in which the speaker made the same request while displaying the non-polite affect cues. The child then indicated which one of the two speakers s/he thought had asked more nicely by placing the named object (e.g., the ball) in the bucket that was in front of that person. In each trial a new set of 'sisters' and a different toy was displayed.

After the first three test trials, the child was again shown the familiarization activity with the stars and then performed the remaining three test trials. In half of the trials, the polite affect stimulus recording was presented first and was located on the left side of the screen; in the other half, the polite affect choice was presented second and was located on the right side of the screen. Thus, as noted, in total each child performed six test trials and two familiarization trials. The whole experiment lasted about 5 minutes and was videotaped.

4.2.5. Coding

For the coding of responses, a trial was coded as 'correct' and scored '1' if the child put the toy into the bucket in front of the speaker who had produced the 'polite affect' cues, in other words, the speaker who had seemed nicer in making their request. In their responses, children quite often pointed first towards the person they thought seemed nicer and only afterwards put the object into the bucket in front of that person. However, those children who pointed first always put the object in the corresponding bucket. The combination of 'polite affect' cues varied according to the modality, in the AV condition consisting of a verbal request (e.g., *Can you give me the ball please?*) with a rising nuclear configuration (L+H* H%) and a smile; in the AO condition a rising nuclear configuration in the audio playback but a neutral facial expression in a still photo; and in the VO condition a smile in the video playback but no audio track. If the child put the bucket in front of the speaker who displayed 'non-polite affect' cues, this was counted as an 'incorrect' response and scored as '0'. A total of 216 trials were analysed (6 test trials × 36 participants).

4.3. Results

In order to test overall performance in each condition, a binomial test was applied to the data. A bar graph illustrating the mean ratios of correct responses for each modality is shown in Figure 3.

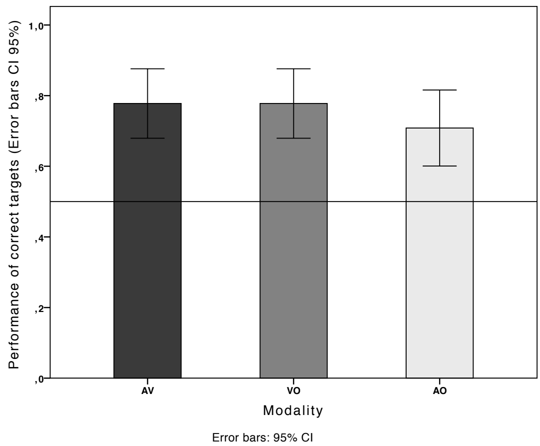


Figure 3. Mean ratio of correct responses by modality (AV = audio-visual, VO = visual-only, AO = audio-only)

The results of the binomial test indicated that the ratio of correct interpretations of polite stance was .78 in conditions AV and VO, thus greater than the chance ratio of .50 ($p < .001$, 1-sided), as predicted. Contrary to our predictions, however, the ratio of correct interpretations of polite affect in the AO condition was .71, also greater than chance ($p < .001$, 1-sided). Thus, the children performed significantly better than chance in all three conditions, indicating that they were sensitive to both prosodic and visual characteristics of polite stance. To assess potential differences between modalities, a second Generalized Linear Mixed Model (GLMM) was performed in which Choice was set as the dependent variable, Modality was set as fixed factor and Subject and intercept were set as random factors. Again, modality in the GLMM analysis was not found to be significant, with $F(2, 213) = 0.315$, $p = .730$. In other words, the children performed as well in the AO condition as in the other two conditions. Odd ratios (Exp(β)) were calculated with the AV condition as baseline in order to analyze effect sizes with logistic regressions.

AV vs. AO: (β =-.485, *SE* = .596, *p* = .416, *Exp(β) = .616)*

AV vs. VO: (β =-.066, *SE* = .609, *p* = .914, *Exp(β) = .936)*

Odd ratios represent the odds that an outcome will occur given a particular exposure and in comparison to the odds of the outcome occurring in the absence of that exposure (Szumilas, 2010). Odd ratios superior to 1 are associated with higher odds of outcome compared with the baseline category. In the two comparisons, the odd ratios were inferior to 1. In the first case (AV-VO) the β regression coefficient indicates that the probability of a child in the VO condition detecting a speaker's polite stance is about .066 times higher than for a child in the AV condition. In the second case (AV-AO) the β regression coefficient indicates that the probability of a child in the AO condition detecting a speaker's polite stance is about .485 times higher than for a child in the AV condition. Even though the difference between AV and AO compared to AV and VO is bigger based on the effect sizes, there is no statistically significant difference between the conditions, implying that this tendency might only be found in the present data sample and not in the total American English population in the USA which is represented by the sample.

4.4. Discussion and conclusions

This study tested 3-year-old children's understanding of a speaker's polite stance in requests by analyzing their ability to exploit audio cues (intonation), visual cues (facial expression) and the combination of both in a situation where lexical cues were controlled for. The results of the study showed that children performed significantly above chance in all three modalities of presentation, namely audio-visual, visual-only and audio-only. In the experimental task, children had to choose from two possible answers and the only cues they could access in order to detect a difference between the two requests were either intonation or facial cues, or both, since lexical content and context were controlled for. The results of the experiment add interesting new insights to research on children's ability to infer politeness from both intonational and facial cues at a very early age.

First, the results confirm that facial expression is a cue that children can broadly use to access polite meanings conveyed by speakers. These results are consistent with other studies suggesting that preschoolers have the ability to access both emotional information and more complex pragmatic meanings which might require them to understand the other person's pragmatic perspective through gestural and facial cues (Armstrong et al., 2014; Armstrong & Hübscher, 2018; Butcher & Goldin-Meadow, 2000b; Hübscher et al., 2017b; Kelly, 2001). Furthermore, interestingly, children performed similarily well in the visual only as in the audio-visual condition and there was no additive effect in the audio-visual condition.

Second, the children performed as well in the audio-only condition as in the other two conditions, suggesting that children's sociopragmatic understanding of intonation is as good at this age as is their understanding of facial cues encoding the same meaning. This is the first study showing that 3-year-olds can infer a speaker's polite affective meaning just as well from intonation as they can from facial cues. While most previous research has neglected to investigate children's ability in infering politeness meaning from intonation and facial/gestural cues, those studies that did find later sensitivity to these cues (Bates, 1976) or found that facial cues play an important scaffolding role in older children's understanding of politeness meanings (Shochi et al., 2009).

Thus the present results differ from the results of other studies that have found a lack of sensitivity in pre-school and primary school children to prosodic cues to a speaker's emotional state relative to lexical or contextual cues (Morton & Trehub, 2001; Morton et al., 2003; Waxer & Morton, 2011; Quam & Swingley, 2012; Aguert et al., 2013). The question that arises is why the present findings are so discrepant with prior studies on children's ability to infer emotional/pragmatic meaning from intonational cues. There are several plausible explanations which might account for this difference. First of all, it might be due to the different nature of the experimental speech stimuli employed. In research on emotional state encoded through intonation, the experimental stimuli have often been pseudo-utterances or low-pass filtered speech, with children being told that the animal characters displayed were speaking another language. While this strategy certainly controls for any possible lexical interference or bias, it may have significantly affected the children's ability to interpret the pragmatic intonation of the message. In the present study, in order to avoid these issues, children were exposed to a pragmatically relevant task involving a request situation whereby they were asked to judge a real person's intelligible polite stance as encoded through either lexical and prosodic cues alone or facial expressions alone or all of these things at once. Regarding the materials, they were exposed to actual video recordings of child-directed speech produced by a set of speakers.

Second, in most studies on children's sensitivity to (emotional) meaning encoded through prosody, prosody is actually poorly defined and is often referred to as a 'paralinguistic cue'. For example, Morton and Trehub (2001) characterise the target prosodic cues used in their experiment as 'affective paralanguage', without giving any further information. Aguert et al. (2013) likewise fail to offer information about the prosodic cues that their child participants were exposed to beyond the observation that the prosody was either happy or sad and thus emotionally salient. Even when such information is provided, it may be that the prosodic features selected to cue emotion have been excessively subtle. For instance, though Aguert et al. (2013) define the prosodic cues they employed with sufficient detail, informing us that their positive-affect prosody cues had a high F0 mean accompanied by a large F0 range while their negative-affect cues had a lower, more monotonous F0 pattern, the question remains as to whether these differences were sufficiently salient as cues to emotion. In order to avoid this issue, we defined and controlled for the specific contours of the intonational cues employed in our experiment (basically rising and falling interrogative pitch contours) in accordance with the Autosegmental Model (Jun, 2010; Pierrehumbert, 1980).

Third, the experimental design could have played a role as well. In the current study children were tested in a forced-choice task in which they were exposed to two video stimuli in sequence. It may be that this experimental set-up actually helped children to perform better compared to other designs where children were exposed to one stimulus at a time and then had to make a decision between two possible answers. Though at this point it is not possible to draw any firm conclusions about the potential effects of this difference in experimental design, it might be worthwhile to explore this issue further. Futhermore, when analyzing the understanding of emotions/polite stance it must be borne in mind that the affects' valence depends on context, which is itself partially generated by societal factors. Thus, in order to understand the intentionality of affect, an appropriate contextualization and social context must be provided (see e.g. Khu, Chambers, Graham, & submitted; Wilutzky, 2015). Finally, there might be a difference in the way preschool children access the various meanings encoded through prosody. It thus remains to be tested whether with the current experimental paradigm children could just as easily access a person's emotional state as they do someones' polite stance.

While our results differ from those found in some prior studies, we would like to highlight the fact that the sensitivity to intonation by 3-year-old children we report here is consistent with reports in the literature. Many studies in fact suggest that prosody helps infants detect emotional information, but there are at least three studies with children aged 12 to 36 months showing that prosodic information in isolation was sufficient to regulate their behavior. For example, Mumme et al. (1996) showed that 12-month-old children regulated their behavior when hearing emotional prosody that made them afraid. Similarly, Hoicka and Wang (2011) showed that 15-month-olds responded appropriately when the experimenter produced either humorous or sweet vocal cues combined with mismatching and matching behavior. These results also seem to be in line with Hübscher et al. (2017b) study which showed that 3-year-old children are able to attribute mental states (in this case, epistemic stance) to another person on the basis of intonation only, and crucially that they are better at detecting this epistemic stance through intonation than through lexical marking. Together, these results suggest that between 1 and 3 years of age children are able to make the appropriate semantic links between prosody and pragmatic and emotional information (see also Armstrong & Hübscher, 2018 for a review; Esteve-Gibert & Prieto, 2018).

To sum up, clearly intonation and facial cues have been grossly overlooked in most studies on children's developing understanding of politeness, and the current results provide new evidence that those cues might pay a much bigger role in this process. This is the first study of its kind to show that very young (3-year-old) children are sensitive to sociopragmatic meaning conveyed through intonation and facial cues and that they are able to metacognitively react to it. This has implications for parents, caregivers and preschool teachers because it suggests a means of scaffolding children's sociopragmatic awareness, which often solely focusses on verbal content. However, judging from our findings, it seems clear that neither prosodic (in particular intonational) nor facial cues should be forgotten.

One possible limitation of the present study is that the data set includes a relatively small number of children in each condition and future research should address whether the trends observed in the present study are applicable to a greater number of children. In this context it would also be interesting to compare American English speaking children with children with other language backgrounds, to see whether the currently obtained results are comparable, or whether there are cross-cultural differences in children's development of politeness understanding. For example, it might well be that languages which do not display an intonational contrast between polite and non-polite requests offer a greater difficulty for children to grasp the prosodic cues of politeness. Research should also consider testing children on perspective-taking tasks, such as for example theory of mind and emotion detection to investigate more closely individual differences. Finally, it should be noted that this study has examined only one speech act, e.g., requests, at a specific age and it would be interesting in future research to widen up the scope and investigate whether children's development of politeness behaves similarly across different speech acts and also across a wider age span. In general, more studies with accurate descriptions of intonational patterns are needed (and which include pragmatically relevant situations) to start assessing how children develop their prosodic awareness in relation to social and pragmatic meanings.

5

Chapter 5: Preschooler's development of polite stance: how prosody, face and body cues pave the way

This chapter is adapted from:

Hübscher, I., Garufi, M., Prieto, P. (minor revision). The development of polite stance in preschoolers: how prosody, gesture and body cues pave the way. *Journal of Child Language.*

5.1. Introduction

People constantly negotiate different social meanings such as their identities, roles and interpersonal relationships. These social meanings are expressed not solely through words and syntactic and grammatical choices, but also through prosodic, gestural and body signals, which are considered to be key features of pragmatic communication in adults (see e.g. Kelly, Barr, Breckinridge Church, & Lynch, 1999 for gesture; Prieto, 2015 for prosody). For example, a seemingly polite request such as 'Can you open the window' can be intended or interpreted as impolite depending on the tone of voice and the gestural features accompanying it. However, according to Brown and Levinson (1987) widely accepted sociolinguistic analysis, politeness is a function of not only the linguistic and nonverbal message conveyed but also in the relationship between the speaker and hearer. Thus Brown and Levinson's theory predicts a certain interaction between the dimensions of the social context and the use of face redress strategies, whereby more face-saving strategies will need to be used to make requests of people who have more power or greater social distance, and the same will be true of requests intended to show a higher degree of imposition or cost to the 'face' of the person who receives the request. According to this framework, children in the process of acquiring their native language need to learn not only what certain speech cues signal about the social context but also how to combine various linguistic elements into coherent styles. This sort of attunement to social relations is a cognitive trait of humans which guides children's understanding of communication from an early age. The ability to recognise and then communicate politeness is therefore going to be a clear manifestation of a child's socio-cognitive development.

In this regard, surprisingly little is known about the role gesture and prosody play in children's developing use of politeness. Research on the acquisition of politeness phenomena has usually focused only on verbal means, such as polite words, mitigating lexical cues and honorific markers (see e.g. Aronsson & Thorell, 1999; Axia & Baroni, 1985; Bates, 1976; Bernicot & Legros, 1987; Ervin-Tripp & Gordon, 1986; Georgalidou, 2008; Hollos & Beeman, 1978; James, 1978; Nakamura, 1999; Nippold, Leonard, & Anastopoulos, 1982; Read & Cherry, 1978; Ryckebusch & Marcos, 2004), while ignoring gestural and prosodic strategies. By contrast, studies on the role of gesture and prosody in language development have hardly ever explored the signalling of polite stance (see Goodwin & Goodwin, 2001; Goodwin, Goodwin, & Yaeger-Dror, 2002; Matoesian, 2005 for exeptions). Clearly, observations that fail to analyse verbal and nonverbal cues together will provide an incomplete picture of polite stance. Furthermore, as we will see below, there are sound empirical reasons to hypothesize that non-propositional strategies involving gesture and prosody pave the way for propositional pragmatic marking in children's sociopragmatic development.

To test this hypothesis, the present study investigates preschool children's use of multimodal indexing (i.e., prosodic, gestural, body and lexical/morphosyntactic markers) of the expression of politeness and the role that such indexing plays in their developmental process. First, however, let us review the literature on the scaffolding role of gesture and prosody in the acquisition of language and then summarize what previous research has said about children's acquisition of politeness.

Previous work on the impact of gestures and prosody on language acquisition

Gesture has been shown to play a scaffolding role in early language acquisition. Before uttering their first words, infants use a repertoire of gestural signals with communicative functions. For example, a child may express his/her emotions through different facial expressions, direct attention to objects through the use of eye gaze and pointing gestures, wave in order to greet or negate by shaking his/her head (Bates et al., 1979; Bates et al., 1975; Guidetti, 2002, 2005). This repertoire expands over the second year of life as more gestures appear with different representational and pragmatic properties, such as using an empty hand to mean 'open' or 'give'. A number of studies have also demonstrated that children's gestures can serve as predictors of their subsequent language acquisition, mainly focusing on the facilitating role of gesture in acquiring the symbolic nature of language and syntax. For example, children's progress in the use of pointing gestures typically anticipates progress in their spoken language, thereby predicting the size of their lexicon (Acredolo & Goodwyn, 1988; Brooks & Meltzoff, 2008; Carpenter, Nagell, & Tomasello, 1998; see Colonnesi, Stams, Koster, & Noom, 2010 for a review) as well as its content (Iverson & Goldin-Meadow, 2005). Sansavini et al. (2010) investigated longitudinally the early development of gestures, object-related actions, word comprehension and word production in Italian-acquiring infants from 10 to 17 months and found that they all increased significantly from 10 to 12 months, with gesture developing earlier than object-related actions and word production developing from age 1. They also detected that gestures supported the emergence of verbal abilities, while object-related actions developed in parallel with word comprehension. By the same token, gesture has been found to be a predictor of the transition to multiword speech. At around 17 months children start to produce supplementary deictic gesture-speech combinations (in which word and gesture convey different but related concepts), and two-word combinations emerge about four months later. Gesture-word 'sentences' such as saying 'Mummy' (the argument of the sentence) while pointing to a chair (the predicate) to ask their mother to sit down appear several months before children can form the same construction entirely in speech ('mummy sit') (Capirci, Contaldo, Caselli, & Volterra, 2005; Iverson & Goldin-Meadow, 2005; Özçalişkan & Goldin-Meadow, 2009; Özçalışkan & S., 2005). The number of representational gesture-word or two-word combinations that a child can use at 18 months predicts the complexity of the sentences they produce two years later, at 42 months (Rowe & Goldin-Meadow, 2009).

While most research has primarily focused on children's early production of representational, deictic and conventional gestures and their precursor role in lexical acquisition, less is known about the role of gestures in the acquisition of pragmatic functions and politeness in particular. There are some relevant studies, mostly of longitudinal nature, which have explored the interplay between gesture and the ways children establish meaning in language. Investigating children's agreement and refusal messages, Guidetti (2005), Beaupoil-Hourdel et al. (2015); Benazzo and Morgenstern

(2014) found that the gestural modality was operational before the verbal modality, with children using conventional gestures such as a head shake and a head nod to convey negation and affirmation before they learned to use the corresponding lexical items. Also, a number of studies have looked at the relationship between gesture and speech in narrative development, investigating the new types of gestures which appear as children grow (Colletta & Pellenq, 2004; McNeill, 1992), such as beats (rhythmic gestures), metaphoric gestures (gestures that express abstract concepts) and discourse cohesion gestures (gestures that accompany connectives; see Kendon (2004)). These studies argue that the pragmatic function of such gestures is to help children negotiate aspects of situational embedding by transmitting attitudes, different levels of attention and agreement between participants in an interaction, or to break apart speech into different information packages or turns, thereby directing the organisation of a discourse. Surprisingly little is known, however, about if and how children use gestures and other body cues to signal their (polite) stance and whether those signals act as a predictor of the onset milestones in children's sociopragmatic development.

What is the role of prosody in children's language acquisition process? First and foremost, language acquisition research has shown that prosody acts as a kind of syntactic bootstrapper. That is, certain types of prosodic features have been shown to guide children's initial acquisition of word order and syntactic structures (for a conceptualisation, see Hirsh-Pasek, Tucker, & Golinkoff, 1996; see also Christophe, Nespor, Guasti, & Van Ooyen, 2003). Furthermore, recent evidence stemming from an experimental study looking at the ways children develop an understanding of another speaker's uncertain knowledge state suggests that prosody and gesture might play similar bootstrapping roles in giving children access to pragmatic meaning before they comprehend the relevant lexical cues (Hübscher et al., 2017).

With regard to production, prosody has also been shown to play an important role in a child's pragmatic language development. Between the ages of 7 and 11 months infants make prosodic distinctions in their communicative versus investigative vocalizations (Papaeliou & Trevarthen, 2006), and are also able to make prosodic (and gestural) distinctions in their intentions at 12 months (Esteve-Gibert, Liszkowski, & Prieto, 2016). Specifically, for vocalizations that require something of their caretaker (e.g., requests and expressions of discontent), infants produce vocalizations with a wider pitch range and longer durations (Esteve-Gibert & Prieto, 2013). A series of case studies investigating longitudinally the emergence of prosody in pragmatic contexts (Dodane & Martel, 2009a, 2009b; Martel & Dodane, 2012) have reported a similar tendency. For example, Dodane and Martel (2009b) found that an infant's prosodic production is affected by the particular communicative situation. Specifically, infants vary intonation depending on whether they are addressing an interlocutor or not and do so even before they produce their first words, illustrating the role of prosody as a precursor in children's early communication. Other research has shown that between the one- and two-word stage infants can produce adult-like intonation contours that are pragmatically appropriate for different situations (Chen & Fikkert, 2007; Frota et al., 2016; Prieto et al., 2012a), pointing to an early acquisition of the speech act meaning of intonational contours, along with the above outlined early ability to convey intentionality through prosodic cues.

Yet, depending on the pragmatic meaning encoded through intonation, the production of certain contours may take longer to develop than others. While the production of acoustic correlates of stress seems to be acquired between ages 2 and 3 (see, e.g., Furrow, 1984; Hornby & Hass, 1970), according to the literature other pragmatic uses of intonation are in place only later in development, due to cognitive and social constraints. For example, expression of belief states or politeness involve more complex cognitive skills and thus have been found to develop only after age 3 (Esteve-Gibert & Prieto, 2018). However, thus far no research has explored whether and how children use prosody as they develop an ability to express politeness and also whether prosody might work as a facilitating device in that process.

Children's acquisition of politeness

In language socialization theory, politeness is generally considered a type of affective stance, with 'polite' affective stances conveying notions like formality, respect and deference. Although research on how children learn to produce gestural and prosodic cues has started to investigate the role of such cues in the acquisition of pragmatic meanings, there is a clear lack of research on the development of politeness stance. Politeness is a complex issue in language use, especially for children. In order to make appropriate use of politeness, children must not only know what forms are used but also take into consideration pragmatic conditions such as social distance, unequal power and the cost to interlocutors' 'face' of the interaction. It is thus not surprising that the literature on children's acquisition of language reports that the acquisition of politeness signalling is a protracted process and the general ability to employ polite speech which involves more conventionalized adult targets is not acquired until age 5 or older (Baroni & Axia, 1989; Nippold et al., 1982; Pedlow, Sanson, & Wales, 2004). Be that as it may, children are usually socialised early into politeness routines. Adults generally support children's conversational contributions by providing a model and by scaffolding children's performance (see e.g. Morgenstern, 2014b). Interactional routines have been shown to play an especially important role in managing face-work and are often specifically modelled for children by adult interlocutors, who have more sociopragmatic expertise (e.g. Gleason, Perlmann, & Greif, 1984). Research on the acquisition of politeness has studied intensively how children are socialised in politeness routines at an early age, by exposure in English, for example, to verbal forms such as *thank you*, *please* and *I'm sorry* (Gleason & Weintraub, 1976; Greif & Gleason, 1980). Similarly, Kaluli-speaking children in Papua New Guinea are taught to use appropriate forms of address (Schieffelin, 1990), Cakchiquel-speaking children in Guatemala are taught to perform end-of-the-meal routines and Japanese children are taught how to bow (Nakamura, 2002). As a result, by age 3 children in all culture groups already have a good grasp of the sociolinguistic function of greetings, polite expressions and formal language (Nakamura, 1999; Nakamura, 2006a). In a separate study looking at children's ability to understand linguistic register, Wagner, Vega-Mendoza, and Horn (2014) found that children already at age 3 were able to access formal/polite speech in Spanish by linking the register to the corresponding addressee when they had sufficiently strong cues, whether social or linguistic (such as pronouns). These results similarly showed that when such cues to register were produced consistently children were able to access this knowledge earlier.

Other studies exploring children's acquisition of politeness have focused on speech acts and more particularly on directives like requests, commands and orders. The ability to handle requests is of key importance in conversational competence, especially in a developmental context in which the child speaker needs to interact with interlocutors that represent different conditions of social distance and power. Previous research has examined children's use and understanding of politeness in requests in various languages, including English (Axia & Baroni, 1985; Bates, 1976; Bernicot & Legros, 1987; Ervin-Tripp & Gordon, 1986; James, 1978; Nippold et al., 1982; Read & Cherry, 1978), French (Ryckebusch & Marcos, 2004), Greek (Georgalidou, 2008); Japanese (Nakamura, 1999); Swedish (Aronsson & Thorell, 1999), Norwegian and Hungarian (Hollos & Beeman, 1978) and Turkish (Uçar & Bal, 2015). Generally, results show that children use mainly direct request strategies in early childhood and that the ability to tailor their language in order to take into account a listener's age and status and the cost of the exchange starts around ages 4 or 5.

One of the most complete studies of this issue is Bates (1976). First looking at the spontaneous production of requests in Italian-speaking children, she found that there were three main phases. Until about age 4, children mainly used direct questions and imperatives as requests. Then, from ages 5 to 6 they acquired all the syntactic forms needed to produce requests but were not yet very skilled at modulating them. Later, by age 7, they were able to vary both the form and the content of their requests using expressions such as *please* or a softer tone of voice to make their requests more polite. Furthermore, Bates was also interested in the children's ability to judge how polite a request was. However, the children struggled to recognize the difference between interrogative and imperative forms till age 5. Similar results were also found by Nippold et al. (1982) for English-speaking children.

Focusing on the social rules behind the choices of polite forms, Axia and Baroni (1985) investigated whether children varied their requests on the basis of cost. In their experiment, children aged 5, 7 or 9 made repeated requests to adult interlocutors. Whenever the adult judged a request to be insufficiently polite, they did not respond. For all three age groups, the adult intentionally ignored every request that was made for the first time (a refusal to respond increases the 'face' cost to the requester of a subsequent request). In their second, reiterated requests, 5-year-olds rarely knew how to make their request more polite, while 7-year-olds were somewhat more adept at this and 9-year-olds much more so, because they knew how to add mitigators like *please* and fall back on question forms or the conditional tense.

Using a similar approach, Read and Cherry (1978) instructed English-speaking 4-year-olds to make requests of the Cookie Monster until their request was accepted. They observed that after being turned down twice the children produced more indirect requests and politeness markers. However, the politeness marker *please* often conflicted with the tone of voice of the request (shouting). Thus the children seemed to be aware that they needed to change tactics, but were not able to match their morphosyntactic strategy with the intonation they applied. Taken together, these studies suggest that children start to recognize relatively early (between 3 and 4 years) that certain strategies such as *please* are used in order to be polite, but the ability to produce appropriate politeness strategies spontaneously seems to develop only slowly.

However, in a recent study we investigated children's sensitivity to non-propositional cues to politeness (Hübscher, Wagner, & Prieto, 2016). Thirty-six 3-year-old American English-speaking children performed a forced-choice decision task which investigated whether they were are able to interpret pitch and facial expression as cues to a speaker's polite stance in audio-only, visual-only or audio-visual presentation modalities, when lexical cues were controlled for. The results showed that at that age children could infer a speaker's polite stance equally well in all three conditions, suggesting that intonation and facial cues do indeed serve children as strong cues to a speaker's polite stance in requests.

Summarizing, the previous research on children's acquisition of politeness is characterized by a main focus on the lexical/morphosyntactic indexes of politeness meaning, and in very few cases is tone of voice/intonation analysed (see Bates, 1976; Hübscher et al., 2016; Read & Cherry, 1978). There is also a surprising gap in the literature regarding how children learn to produce gestural and postural signals of politeness, the exception being Goodwin et al. (2002). Using a conversation analysis approach, these authors investigated children's multimodal expressions of disagreement in disputes and found that turn shape, intonation and body positioning were all critical to the construction of stance. The use of prosody by children to express politeness has likewise been relatively neglected in previous studies. Importantly, the results of Hübscher, Wagner, and Prieto (accepted) study point to a potentially important role for prosody and nonverbal cues in the marking of early politeness stance in children, and it is this issue which constitutes the central research question of the present study.

Prosody and nonverbal signals in the expression of politeness in adult speech

Research on adults gives some insight into both prosodic and nonverbal cues in producing politeness-related meanings. Though early studies claimed that the perception of politeness increased with pitch range and pitch height (see Ohala (1984) Frequency Code hypothesis), this evidence has been recently contested. While it is true that in certain languages a high pitch range tends to be perceived as more polite (Chen et al., 2004 for Dutch and English), in other languages such as Korean (Winter & Grawunder, 2012) and also Catalan politeness seems to be associated with a somewhat lower, not higher, mean pitch (Hübscher et al., 2017a). In addition to pitch, the importance of prosodic/acoustic correlates such as speech rate (Hübscher et al., 2017a; Lin, Kwock-Ping, & Fon, 2006; Ofuka, McKeown, Waterman, & Roach, 2000; Ruiz Santabalbina, 2013; Winter & Grawunder, 2012), intensity and voice quality (Hübscher et al., 2017a; Ito, 2004; Winter & Grawunder, 2012) has also been pointed out. Most importantly for the present study, adult Catalan speakers have been found to display a prosodic mitigation strategy that involves decreasing the rate and intensity of their speech and displaying less jitter and shimmer to communicate politeness in situations where the power distance is high between speakers (Hübscher et al., 2017a). Furthermore, in their investigation of how politeness is encoded through intonation in adult Catalan-speakers, Astruc, Vanrell, and Prieto (2016) found that both the cost of the action and social distance had significant effects on intonation choices. While high cost situations triggered more rising pitch patterns than did low cost situations, power distance did not have a significant effect on the choice of intonation contour.

Regarding nonverbal behavior, a small number of studies have shown that a range of mitigating gestural behaviors are often employed to express politeness-related meanings. Mitigation has been associated with face-preserving efforts, which can also involve politeness (Briz, 2002, p. 21). In American English, it has been shown that a range of mitigating gestural behaviors are often employed to signal politeness, such as a pleasant facial expression, raised eyebrows, a direct body orientation or a tense, closed posture with small hand gestures, accompanied by a softer voice, touch and close proximity. In contrast, aggravating behaviors include greater distance, an indirect body orientation, unpleasant facial expressions, lowered eyebrows, a loud voice and wide gestures (Tree and Manusov (1998). Since lexical epistemic (uncertainty) markers have often been mentioned in the context of hedging in order to lessen the face-treat (see, e.g., Barrios Sabador, 2016), the detailed description of the nonverbal correlates of uncertainty/doubt markers offered by Givens (2001) is of considerable value. His report includes facial expressions (eyebrow frowns, eye movements, lip-pouting, lip-pursing), head movements (headshakes, head tilts), and cues like adaptors,[7] palm-up open hand gestures and shoulder shrugs.

Research in social psychology has revealed that power is communicated nonverbally through behaviors implying strength, comfort-relaxation and fearlessness, whereas submissiveness is communicated through behaviors implying weakness, smallness, discomfort, tension and fearfulness (Mehrabian, 1981). When the social difference is high between interlocutors, powerful individuals typically adopt an expansive posture, speak loudly, lower their eyebrows, gaze directly at their social partners when speaking, nod more, use fewer self-touches, make more arm and hand gestures, shift their position more frequently, thereby show less body relaxation, and stand closer. In comparison, less powerful individuals typically have a hunched posture, speak quietly, raise their eyebrows and vary their gaze (for reviews, see Ellyson & Dovidio, 1985; Hall, Coats, & LeBeau, 2005). Similarly, in high power distance situations involving Korean speakers the interlocutor of inferior status shows deference by a direct orientation of the body and constrained posture, and by suppressing gestures and touching. By contrast, in situations of low power distance in Korean body positioning is more relaxed and there are more gestures and touching (Brown and Winter (in press).

While there is not much literature on the cultural norms regulating the use of manual gesture, in many cultures a pointing gesture realized with an

[7] Adaptors, also called self-manipulators, are largely unconscious hand movements typically involving scratching, touching or covering a part of the face or body which are performed without communicative intention and are generally assumed to denote psychological discomfort or anxiety.

index finger extended is considered rude in anyone other than a small infant. A study carried out in Poland confirmed that, among Poles, as language acquisition continues, the pointing gesture particularly when directed at people begins to be perceived as inconsistent with Polish cultural norms and is often suppressed (Jarmołowicz-Nowikow, 2014). Similar observations have been made in relation to perceptions about certain ways of pointing among the Yoruba (Ola Orie, 2009). Taken together, these results suggest that given that not only prosody but also gestural and body cues are important markers of interpersonal positioning in speech between adults, they merit careful scrutiny in any investigation of how children develop an awareness of polite stance marking.

Present study

To our knowledge, no study so far has explored the early production of lexical and morphosyntactic strategies concurrently with the emergence of prosody, gesture and body signals in the expression of polite stance from a developmental perspective. In the current cross-sectional study we proposed to fill this gap by exploring the temporal link between gestural, body, prosodic and lexical/morphosyntactic representation in young (aged 3-5) Catalan-speaking children's multimodal indexing of polite stance in request situations. In particular, we wished to examine (1) if and how young children mitigate their requests depending on the social parameters of social distance and cost, (2) whether children use prosodic and gestural and other body strategies earlier and more predominantly than lexical and morphosyntactic strategies and (3) whether differences between a younger age group (3;0-4;6 years) and an older age group (4;6-5;0 years) reflect different stages of sociopragmatic development. To do so, we conducted a request production task in which children were pragmatically induced to ask for a certain object and in which we varied the variables of social distance and cost. In the social distance dimension, in two of the situations the children were prompted to request an object from an experimenter (high social distance) and in the other two situations they were prompted to request something from a peer (low social distance). We also manipulated the context so that in one situation the cost of the request to the child would be low and in the other it would be high. Basing ourselves on previous literature which report an early facilitation role for gesture (see Colonnesi, Stams, Koster, & Noom, 2010 for a meta-analysis) and prosody in the expression of intentionality (Sakkalou & Gattis, 2012), speech acts (Esteve-Gibert et al., 2017b) and contrast resolution (Ito et al., 2012), we hypothesised as follows: in answer to our first research question above, these children would mitigate their requests in different ways depending on the degree of social distance between interlocutors and degrees of cost; in answer to our second question, the children would more predominantly mitigate by means of prosodic, gestural and other body signals compared to lexical and morphosyntactic markers to mark politeness; and the children's repertory of mitigation strategies to signal politeness would expand over the preschool years in terms of not only lexical and morphosyntactic but also prosodic, gestural and body markers.

5.2. Methodology

5.2.1. Participants

An initial group of ninety-two 3- to 5-year-old children were recruited for participation in the cross-sectional study from four Catalan public preschools in the Barcelona metropolitan area, where the population is largely Catalan-Spanish bilingual. However, because the experimental materials were prepared in Catalan, it was felt important to ensure that all participating children were predominantly Catalan users and would therefore be fully comfortable. As a result, prior to the experiment, the parents of all those children recruited completed a questionnaire (based on Bosch & Sebastián-Gallés, 2001) to determine the degree to which their child was exposed to Catalan on a daily basis. The 20 children whose parents reported a daily exposure to Catalan of less than 50% were excluded from the study, as were an additional eight whose parents failed to complete the questionnaire. This left a final participant population of 64, for whom the mean percentage of daily exposure to Catalan was 85% (SD = 0.158) This population was then divided into two groups by age, with younger age group made up of 32 children (mean age 3;8, SD = 0.464) and older age group made up of 32 (mean age 5;1, SD = 0.495). Both groups were balanced for gender, with 17 girls and 17 boys in each. The children's parents were informed about the experiment's goal and signed a participation consent form prior to the commencement of the study. This study, including the consent procedure, was approved by the Ethics Board of the Universitat Pompeu Fabra.

5.2.2. Materials

In order to elicit request speech acts from the children which would be as close as possible to natural speech, a semi-spontaneous discourse elicitation task was designed which included four controlled and pre-planned situations (an adaptation of Uçar & Bal, 2015 experimental design). Importantly, in contrast to the Discourse Completion Task (DCT) that is usually carried out with adults and often also with older children (see e.g. Vanrell, Feldhausen, & Astruc, in press), where subjects respond to hypothetical or fictional discourse contexts, the present data collection method actually places children in real situations, thus removing the metacognitive layer of the DCT task which may cause participant to produce unnatural, prototypical responses. However, the virtue of both the DCT and the method employed here is that they allow the researcher to control for contextual variables and demographic information, facilitating the possibility of drawing legitimate cross-cultural comparisons. Similarly to the DCT, the current method constitutes an excellent means to elicit speech and nonverbal nonverbal cues by controlling the contextual variables and demographic information, opening up the possibility to make cross-cultural comparisons. However, due to its controlled nature, the DCT, might elicit a more prototypical response as compared to natural spoken interaction due to the lack of a more interactional nature, something which should be born in mind when analysing the data.

The pre-planned target situations in the experiment were designed to allow us to modulate two variables that intermediate in the expression of politeness, as follows:

a) Social distance[8] between Hearer and Speaker. This factor had two levels: low and high.
b) Cost of the face-threatening act, meaning the degree of imposition by the Speaker on the Hearer that the request implied. This factor likewise had two levels: low and high.

We constructed four different conditions, varying along two variables, social distance (low/high) and degree of imposition (low/high) (see Figure 1). In the first condition (low social distance, low imposition) children worked in pairs to put together a jigsaw puzzle. Whenever a child discovered that one of his/her partners had a piece s/he needed, s/he would ask for it. In the second condition (low social distance, high imposition) children were grouped into pairs and then asked to guess the number the researcher was thinking. The child who first guessed the right number received a small bubble blower, while the other one received plain white dough. It was expected that the child who did not guess the number would ask the other child to share his/her coloured dough with them. In the third condition (high social distance, low imposition), the children were shown some stickers and were told that if they wanted to have them that they could ask for them individually. Finally, in the fourth condition (high social distance, high imposition) one of the experimenters was looking through a kaleidoscope. The children had to ask for permission to look through the kaleidoscope.

[8] We use the term *social distance* as shorthand for a more complex dimension which includes the power differential between participants.

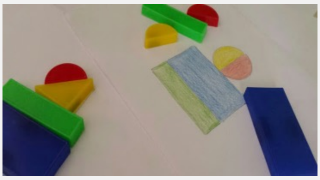
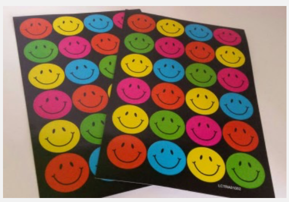
	Low cost	High cost
Low social distance (addressing to a familiar child)	Situation 1: Requesting a missing puzzle piece from a classmate 	Situation 2: Requesting a newly won bubble blower from a classmate 
High social distance (addressing to an unfamiliar adult)	Situation 3: Requesting a sticker that the researcher was handing out freely 	Situation 4: Requesting a kaleidoscope from the researcher with personal value to her 

Figure 1. Description of the four request situations and picture of the objects to be requested

5.2.3. Procedure

The children were tested in pairs in a quiet room at each of their respective preschools. We ascertained beforehand by consultation with their teacher that the children in each pair were compatible, that is, that they would feel comfortable interacting. The two children were seated next to each other along one side of a table (see Figure 2). Experimenter 1, a native speaker of Catalan sat adjacent to the children and gave them instructions to guide them through the four situations. Experimenter 2 sat opposite the children and operated the two video cameras which recorded each session, with one camera centred on the child to the left and the other centred on the child to the right. Experimenter 2 also participated as an interlocutor during the high social distance situations.

Figure 2. Video stills showing the four request situations

Experimenter 1 accompanied each pair of children from their classroom to the experimental setting and then chatted with them briefly to put them at their ease, asking them questions about what they had been doing in class. She also introduced them to Experimenter 2 by name but did not draw her further into the conversation. She then explained that they were going to play some games. Then the four experimental request-elicitation contexts were created, always starting with the two low social distance situations (i.e., interaction between the two children) and then moving on to the high social distance situations (i.e., interaction between the children and the experimenter), with the cost variable alternating with each pair of contexts. Descriptions of each of the four situations follow (instructions given in Catalan during the experiment are rendered here in English translation).

Situation 1: Low social distance - Low cost. This involved giving each child a simple line drawing and several coloured wooden pieces whose shapes exactly matched the shapes depicted in the drawing. The children received an equal number of pieces, but each received one piece that the other child lacked. Experimenter 1 then told them, 'Here is a drawing and some wooden pieces. Put each wooden piece on top of the shape that looks the same in the drawing. You might not have all pieces you need because the children who were here before you might have misplaced them, so you might have to ask each other for a piece.' If the experimenter saw that the children failed to follow this last instruction, it was reiterated.

Situation 2: Low social distance - High cost. Children were each given a piece of modelling dough to play with. Experimenter 1 then said, 'Now let's play a guessing game. Think of a number between 1 and 5. Whoever guesses the number I am thinking of wins and can play with bubbles." When one of the children guessed correctly, they were told 'You win! Great! Now you can play with the bubble blower.' After a little while, Experimenter 1 told the child who had lost, 'I think you can ask your classmate if you can take a turn blowing bubbles.' If necessary, the losing child was again urged to ask for a turn. The 'winning' number to be guessed was manipulated to ensure that both children had a turn being the winner and loser and thus both produced requests. Also, the prize for winning alternated between the bubble blower and a windup toy in order to keep the object of the request desirable.

Situation 3: High social distance - Low cost. Experimenter 1 showed the children that Experimenter 2 had a set of smiley face stickers and then told them, 'Look, the children who were here before you got to stick some of these stickers on paper. But to get a sticker then had to ask her [indicating Experimenter 2 by name].' This was followed by a five-second pause to see if they children would ask Experimenter 2 for stickers. If they did not, they were told 'Wouldn't you like to stick a sticker? I asked her for one and she gave it to me. But you have to ask her because they're her stickers.' If this also failed to elicit requests from the children, after a second pause the instruction was repeated. If another five seconds elapsed without a request being produced, the experimenters moved on to the next situation.

Situation 4: High social distance - High cost. Experimenter 2 held up a kaleidoscope and started looking through, meanwhile

> exclaiming happily about the colours and shapes that she was seeing. She then passed it to Experimenter 1, who looked through it and made similar comments. Then addressing the children, Experimenter said, ‘Are you interested? This kaleidoscope is Iris’ favourite object and it is very important to her since it was a gift from her brother. So if you want to look through it you have to ask her.’ If five seconds then elapsed without the children making a request, Experimenter 1 continued, ‘The colours and shapes that you can see are really cool! Come on! Go ahead and ask her.” If this too failed to elicit a request, after five seconds the children were again urged with ‘Come on! Try and ask her! She’s very nice!’

The full experimental session lasted about ten minutes. After the experiment, the children were accompanied back to their classroom. Scrutiny of the video material collected in the total of the 32 sessions showed a total of 231 verbal requests being made by the participating children. This was somewhat short of the 256 requests that could potentially have been produced (64 children × 4 situations) but there were 25 instances in which a child failed to produce a request either because they were too shy or did not really want the object in question. In addition to these 231 verbalised requests, the video recordings showed 11 occasions where a request was made by nonverbal means only.

5.2.3. Data coding

Both verbal and nonverbal content of all requests was coded. Coding of the lexical and morphosyntactic content of each request was assigned manually and recorded on an Excel spreadsheet. PRAAT (Boersma & Weenink, 2017) was used for the prosodic coding and ELAN (Lausberg & Sloetjes, 2009) for the gestural and nonverbal coding. The specific procedure used for the coding of the different levels was as follows:

Modality. First we coded whether a request was verbal (with or without accompanying nonverbal cues) or exclusively nonverbal. (Although some these latter strictly nonverbal requests were pointing/reaching gestures it was decided not to include pointing/reaching gestures in the overall analysis of politeness-related cues because the specific set-up of the experimental situations did not allow for a strict comparison of the appearance of pointing gestures. For example, the situation requesting a puzzle piece out of several pieces triggered most of the pointing gestures, while the other situations with no object location ambiguity triggered almost no pointing behavior.)

Lexical and morphosyntactic coding. From the literature on the expression of politeness (see Fivero, 1976 for Catalan; Hübscher et al., 2017a; Payrató & Cots, 2011), we know that there are a number of ways requests can be modified in order to make them less face-threatening and that adults do so depending on the degree of imposition implied by the request (i.e., its cost) and the social distance between speakers. To such modification strategies belong speech act types (direct: *Vull això* ‘I want this' vs. indirect speech acts: *Em dones això?* 'Can you give me this?'), downgraders (e.g., diminishing the force of the request through the epistemic modal *potser* 'maybe'), terms of address (informal *tu* vs. formal *vostè*), change of mood in the verb forms (indicative *pots* 'can you' vs. conditional *podries* 'could you') and the use of lexical politeness cues such as *si us plau* 'please'. The combination of these cues heavily influence the degree to which a request is perceived as more or less polite. In our data set only the following two types of lexical/morphosyntactic cues occurred:

a) direct vs. indirect requests
b) presence vs. absence of *please*

These were therefore the only two variables that were coded in this regard.

Prosodic coding. The prosodic coding was carried out based on Hübscher et al. (2017a) study on prosodic correlates of mitigation in Catalan formal register speech. The phonological and phonetic cues listed below have been found relevant in distinguishing formal polite speech from informal speech in various languages (e.g. Catalan: Hübscher et al., 2017a; Korean: Winter & Grawunder, 2012). Some items were coded manually within the Praat interface while Praat registered others automatically, as follows.

Tier 1, Orthographic transcription of the target requests, separated by words.

Tier 2, Syllables, manually segmented. They were marked as (s) and were used to analyse duration patterns.

Tier 3, Final intonation, roughly classified into falling and rising pitch contour.

Tier 4, Intonation patterns, labelled in accordance with the Cat_ToBI framework (Prieto, 2014). The two graphs in Figure 3 illustrate the two most frequent intonation patterns found in the data, as well as the labelling procedure used.

Tier 5, F0 marks, with the following measures indicated manually for each Intonational Phrase: reference line (R, start of the pitch contour of each IP), the baseline (L, lowest F0 point in the nuclear pitch contour) and the top line (H, highest F0 point within the nuclear contour).

The two graphs in Figure 3 show the annotated Praat output for the two requests *Vull això* 'I want this' (left graph) and *Me'l deixes?* 'Can you give it to me?' taken from our data set. The five tiers described above can be seen below the waveform, spectrogram and F0 contour in each graph.

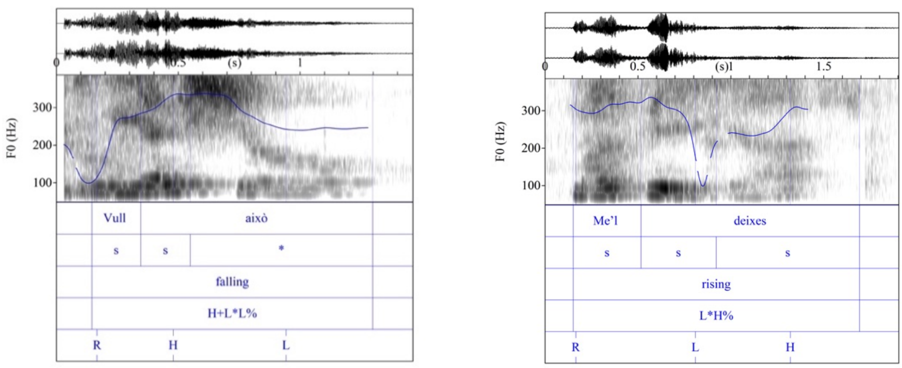


Figure 3. Waveforms, spectrograms and F0 contours of the requests *Vull això* 'I want this' (falling pitch contour, left panel) vs. *Me'l deixes?* 'Can you give it to me?' (rising pitch contour, right panel). The orthographic and prosodic annotation tiers (tiers 1–5) are explained above in this section.

Finally, a series of phonetic measures were automatically extracted within each annotated syllable, namely pitch (mean F0), duration, voice quality (mean jitter, shimmer, H1-H2 as a correlate of breathiness) and intensity. Additionally, the amplitude difference between the first and second harmonics (H1-H2) was automatically obtained in order to assess the breathiness of the participants.

Body and facial coding.

As mentioned above, our nonverbal coding was carried out in ELAN (see Table 3 for an example). Our coding system is primarily based on the MUMIN multimodal scheme developed by Allwood, Cerrato, Jokinen, Navarretta, and Paggio (2007). For those facial gestural cues that are not included in the MUMIN, we used elements from the FACS coding system by Ekman et al. (2002).

In order to create a comprehensive profile of politeness-related behaviors, we assessed an extensive range of facial and body cues all identified in previous studies as interacting with negative politeness (strategy used by the speaker to show that he cares and respects the hearers' negative face), namely mitigation, power and submission (see introduction section for literature review). We then coded these cues primarily and importantly by adapting the codding system proposed in Brown and Winter (in press). Furthermore, similar to Brown and Winter (in press), we took into consideration all body and face signals accompanying the verbal request. We refer especially to body postures which have been documented to arise in states of uncertainty or submission, and to adaptors, i.e. hand movements which denote psychological discomfort and anxiety. As far as the former type of signals are concerned – signals arising in states of uncertainty or submission – it is well-known that uncertainty *lexical* markers are used as hedges in order to lessen the face-treat (e.g. Caffi, 1999). Hence it seemed relevant in this context to consider the nonverbal correlates of uncertainty markers. A detailed description of body and facial cues associated with uncertainty/doubt has been given by (Bitti et al., 2014; Givens, 2001; Krahmer & Swerts, 2005). Their reports include facial expressions (eyebrow frowns, eye movements, lip-pouting, lip-pursing), head movements (headshakes, head tilts), and cues like palm-up open hand gestures, and shoulder shrugs. As far as the latter signals are concerned – adaptors – although they do not communicate negative politeness in a narrow/strict sense, but rather a speaker's discomfort, we decided to include them in the analysis as they signal the speaker's hard time in making a request that does not sound as an imposition on the hearer, but which nonetheless determines the hearer to act conformingly and grant the speaker's request. This is especially complicated in asymmetrical contexts, where the speaker has a lower status than the hearer. Although adaptors are performed with little awareness and no intention to communicate, we decided to take them into account as they nonetheless allow convey information and allow observers to make inferences on the speaker's emotional state (Morris 2002). The resulting set of body and facial cues are illustrated in Figure 4 below, labelled with the terms we used for coding purposes.

Figure 4. Annotated gestural and nonverbal categories

In order to illustrate children's multimodal signalling of politeness-related meanings, we will describe one example in more detail. In Figure 5, the boy on the left is requesting a sticker from the experimenter. About a second before the child starts to voice the request he produces several nonverbal cues, namely a lowered head, slumped shoulders and an averted gaze. Then while he produces the request *Em deixes enganxar un gomet?* 'Can I stick a sticker (on a piece of paper)?', marked with a rising intonation, the boy tilts his head and adopts a sidelong gaze towards the experimenter. Towards the end of the verbal request he produces a smile, lightly shrugs his shoulders before reverted to his initial slumped posture. He also bites his lower lip, a typical adaptor.

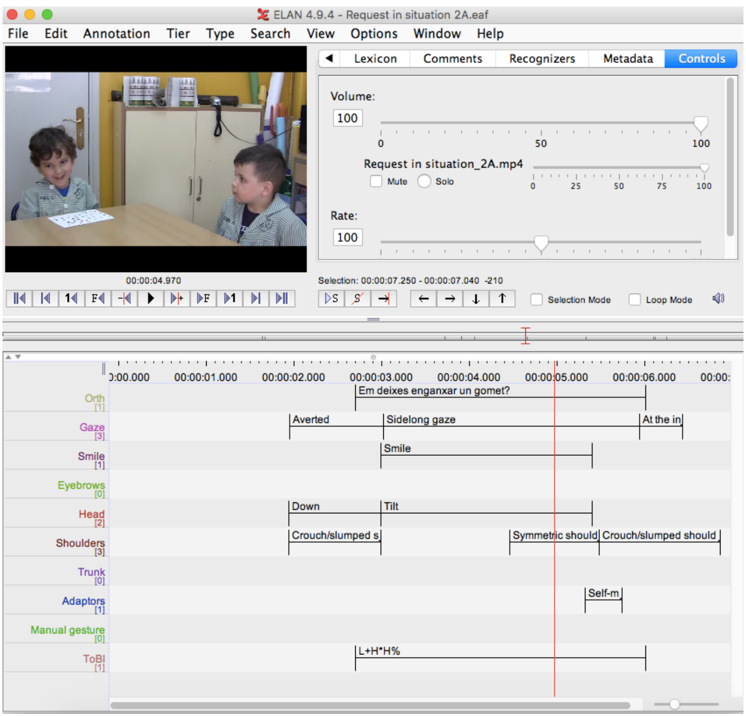


Figure 5. Example of labelling with the request *Em deixes enganxar un gomet?* ('Can I stick a sticker (on a piece of paper)?')

5.2.4. Reliability of the coding

Because coding was carried out by two researchers (the first and second author), steps were taken to standardise the coding process and strengthen inter-rater reliability. The two researchers first worked together to analyse a subset of video recordings to discuss and mutually agree on coding schemes for the gestural/body, prosodic and lexical/morphosyntactic markers. Then, after the full data set was coded by one researcher using ELAN for the gestural/body coding and Praat for the prosodic coding, the other researcher checked the prosodic coding of all examples and 40% of the gestural/body coding. Any differences of opinion regarding instances of coding were discussed until agreement was reached.

5.2.5. Data Extraction and Statistical Analyses

All statistical analyses were carried out using IBM SPSS Statistics v24 software. More specifically, a series of Generalized Linear Mixed Models (GLMMs) were undertaken to compare the values found for the two levels of each of our three fixed factors SOCIAL DISTANCE (two levels: low vs. high), COST (two levels: low vs. high) and AGE (2 levels: younger vs. older). Random intercept was specified for Subject.

The dependent variables were (1) the modality of the request (verbal vs. nonverbal); (2) the presence of morphosyntactic or syntactic cues (*indirect request* or *please*); (3) prosodic features (rising intonation, mean speech duration, pitch, jitter, shimmer, F1-F0 and intensity); and (4) the presence of gestures and other body signals, broken down into three broad categories: gaze, facial cues and body cues. The lexical and morphosyntactic cues as well as some of the prosodic, gestural and body cues were count in nature (i.e., they might occur zero, one or several times during a single request unit). On the other hand, the phonetic correlates of prosody such as duration, pitch, intensity and voice quality were all gradient in nature and measured in their respective units.

5.3. Results

In this section, we first analyse the data in relation to verbal vs. nonverbal modality (Section 3.1). We then analyse the data related to morphosyntactic and lexical marking (Section 3.2). In Section 3.3 we report on the prosodic features of the children's requests. Finally, in Section 3.4 we provide an analysis of the gestural data.

5.3.1. Modality of requests

The vast majority of the requests made by children in our data set were executed verbally (with or without accompanying nonverbal cues). However, participants occasionally simply pointed at or reached for an object in order to express their desire that it should be given to them, and these communicative acts were thus coded as nonverbal requests. Results of the GLMM showed a significant main effect of SOCIAL DISTANCE **(**F (1, 220) = 10,804, p = .001) and COST, (F (1, 220) = 8,300, p = .004) on the modality of requests, such that there were more verbal than nonverbal requests in high social distance and high cost situations (p = 0.023), compared to low social distance and low cost situations (See Table A in Appendix C for a table of mean occurrences of verbal requests per situation, broken down by age group and experimental parameter).

5.3.2. Morphosyntactic and lexical strategies cues

In the total of 220 verbal requests in our data set, the only morphosyntactic cue to politeness observable was the use of indirect question structures, and the only lexical cue sometimes deployed by the children was the mitigator *si us plau* 'please'. The mean occurrence of indirect requests and si us plau is shown respectively in the two graphs in Figure 6, broken down by age group and social distance and cost parameters. Statistical analysis of the data showed a main effect of COST on children's production of indirect requests (F, (1, 220) = 13,260, p < .001), with significantly more indirect requests in high cost requests (see Table 2). Furthermore, there was a statistically significant interaction between the effects of SOCIAL DISTANCE and AGE on the production of indirect requests **(**F (1, 220) = 12,434, p = .001). While in the younger age group there were significantly more indirect requests in low social situations (p = .007), in the older age group there were significantly more indirect requests in high social distance situations (p = .035). This suggests that while the children in the younger age group had not yet assimilated the relationship between indirectness and politeness in Catalan, while the older group had. Furthermore, there was a main effect of both SOCIAL DISTANCE (F (1, 220) = 12,875, p = .001) and COST (F (1, 220) = 6,331, p = .013) on the presence of *si us plau*. In other words, the children tended to produce more *si us plau*s in situations involving higher social distance or higher cost. (See Table B in Appendix C for full results.)

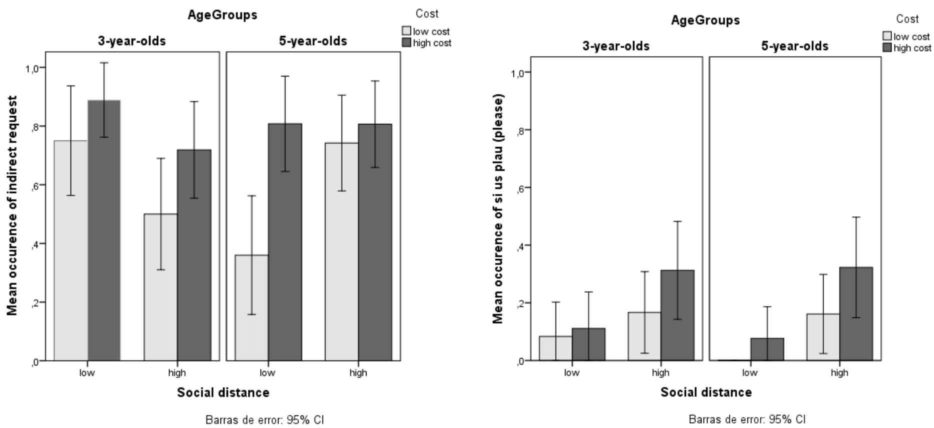


Figure 6. Mean occurrence of indirect requests (left panel) and *si us plau* 'please' (right panel), broken down by social distance, cost and age group. Error bars indicate standard error.

5.3.3. Prosodic features

Intonation contour. As noted, because rising intonation has been identified as a marker of politeness, all instances of rising intonation in the data set were noted. A GLMM analysis showed a significant effect of COST on the production of rising intonation contours (F (1, 201) = 8,906, p = .003), with rising tunes being used more often in high cost requests than in low cost requests, as can be seen in Figure 7 (See Table C in Appendix C for full results).

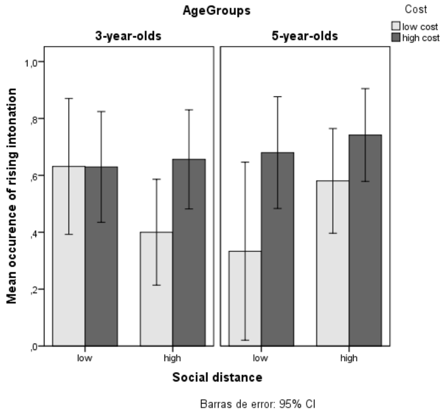


Figure 7. Mean occurrence of rising intonation across the three conditions, namely social distance, cost and age. Error bars indicate standard errors.

Mean syllable duration. Mean syllable duration was extracted automatically from all regular syllables produced in the requests (results for this and all other phonetic parameters are given in Table 1). Statistical analysis showed a main effect of SOCIAL/POWER DISTANCE (F (1, 1245) = 22,660, p < .001) and a main effect of COST (F (1, 1243) = 5,691, p = .017) on the mean duration of syllables. In other words, the duration of syllables tended to be significantly higher in high social distance situations and also in high cost situations. This suggests that, independently of their age, the children produced significantly longer requests when they had to request something from someone with higher social distance or when their request implied a higher degree of imposition.

Average pitch. Average pitch was extracted automatically from all syllables produced in the requests. Moreover, three other pitch measures were extracted by using manually placed specific points in each intonational phrase, namely the reference line, the top line and the baseline.

Table 1 shows the results of the various GLMMs applied plus the means estimated by the models. There was a significant interaction between AGE and COST in pitch height (calculated from the mean of all syllables) (F (1, 1218) = 10, 951, $p < .001$). In the younger age group, average pitch height was significantly higher in high cost requests ($p = .006$). By contrast the effect of COST on pitch was slightly under significance (p = .58) for the older age group, with higher pitch more frequent in low cost requests.

Voice quality and intensity. The following measures of voice quality were automatically extracted for each syllable in our recordings: intensity (in dBs), perturbation by amplitude (shimmer), perturbation by F0 period (jitter) and the harmonic differential (the difference in amplitude between the first and second harmonics, F1–F0, in Hz) (see Table 1).

Statistical analysis showed a main effect of COST on jitter (F (1, 1211) = 10,117, $p = 0.002$), with significantly less jitter in high cost requests (p = 0.002) compared to low cost requests. There was a main effect of AGE on shimmer (F (1, 1207) = 8,244, $p = .004$), with more shimmer in the older group and a main effect of COST (F (1, 1207) = 7,390, $p = .007$), with more shimmer in high cost requests. However, there was a significant interaction between AGE and COST (F (1, 1207) = 11,977, $p < .001$). In the older group, cost had a significant effect on the production of shimmer (p < .001), with more shimmer in high cost situations compared to low cost situations, but this difference was not seen in the younger group (p = .600). There was a significant interaction between AGE and COST in relation to the production of F1-F0, which can be taken as an index of breathiness (F (1, 1218) = 8,743, $p = .003$). In the older age group, cost had a significant effect on the production of F1-F0 ($p < .001$), with more breathiness in high cost requests, there was no effect in the younger group ($p = .368$). Finally, regarding syllable dB (intensity) there was a significant interaction between AGE and COST (F (1, 1243) = 8,701, $p = .003$). In the older group, COST had a significant effect on the intensity rate, with higher intensity in low cost requests ($p = 0.003$). There was no effect of COST in the younger group ($p = 0.224$).

Table 1. Mean (and standard deviation) phonetic values for all the syllables in the data set of verbal requests uttered by children broken down by age group and social distance and cost parameters. Units of each values are given in parentheses in the left-hand column.

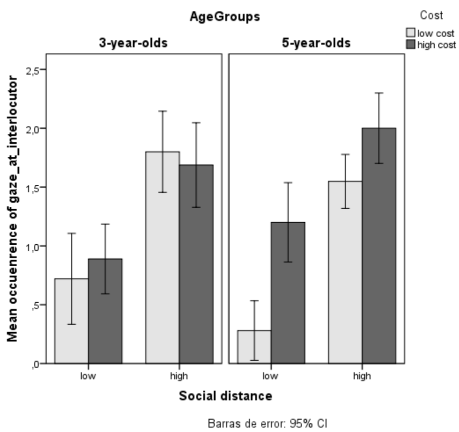
	Age Group	Low Social distance	High Social distance	Low Cost	High Cost
Mean Syll Dur (ms)	**1**	.3067 (.1590)	.3595 (.1948)	.3313 (.1915)	.3391 (.1684)
	2	.2755 (.1520)	.3357 (.2291)	.2880 (.1878)	.3350 (.2168)
Pitch (st)	**1**	17.38 (4.20)	17.19 (3.69)	16.88 (4.25)	17.71 (3.51)
	2	17.95 (4.72)	17.44 (4.37)	18.12 (4.21)	17.18 (4.75)
Intensity (dB)	**1**	61.68 (9.77)	60.88 (10.40)	61.99 (10.02)	60.43 (10.17)
	2	63.38 (9.12)	61.13 (16.03)	63.98 (8.87)	60.15 (16.99)
Jitter	**1**	.0245 (.0148)	.0246 (.0130)	.0232 (.0124)	.0261 (.0152)
	2	.0245 (.0143)	.0250 (.0192)	.0233 (.0195)	.0262 (.0152)

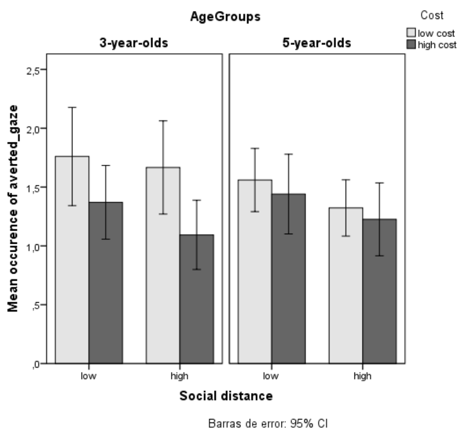
Shimmer	**1**	.1764 (.0517)	.1685 (.0406)	.1725 (.0512)	.1718 (.0401)
	2	.1555 (.0505)	.1566 (.0423)	.1475 (.0471)	.1646 (.0427)
f1–f0 (Hz)	**1**	662.48 (275.13)	673.37 (283.18)	674.60 (285.44)	661.34 (272.65)
	2	673.24 (366.70)	659.44 (334.41)	613.64 (323.32)	713.65 (362.45)
Topline pitch (st)	**1**	19.48 (4.47)	18.45 (4.48)	18.70 (3.73)	19.10 (5.07)
	2	19.03 (7.14)	19.90 (3.45)	19.62 (4.77)	19.60 (5.21)
Bottom-line pitch (st)	**1**	15.57 (4.66)	15.78 (2.53)	15.49 (3.42)	15.87 (3.70)
	2	15.40 (5.56)	15.99 (4.29)	16.41 (4.94)	15.16 (4.69)
Reference line pitch (st)	**1**	18.40 (3.56)	17.81 (5.86)	17.90 (5.69)	18.21 (4.28)
	2	18.00 (4.44)	19.21 (3.18)	18.85 (4.02)	18.68 (3.50)

5.3.4. Gestures and other body signals

As noted above, the set of 11 gestural or body signals was divided into three categories, gaze, facial cues and body signals, with this last further separated by part of the body into head, shoulders and trunk (See Table D in Appendix C for full results.)

Gaze. With regard to the direction of the speaker's gaze while making a request, the children in our study displayed three different behaviors, with gaze either directed at one of the experimenters, averted, or directed side to the side. The distribution of these three behaviors varied according to whether the child was interacting with an adult or with a peer, and also according to the social distance and cost dimensions of the situation, as can be seen in Figure 8. GLMM analysis revealed a main effect of SOCIAL DISTANCE (F (1, 220) = 46,492, p < .001), with significantly more gazes directed at the interlocutor in high social distance contexts. Complimentarily, there was a main effect of COST (F (1, 220) = 6,996, p = .009), with children averted their eyes significantly more when making low cost requests. A near significant effect of social distance was also revealed (F (1, 220) = 3,842, p = .051), with a non-significant tendency for the child's gaze to be averted in low social distance situations. Furthermore, there was a main effect of SOCIAL DISTANCE on the occurrence of sidelong gazes (F (1, 220) = 8,537, p = .004), with significantly more sidelong gazes in situations of high social distance.





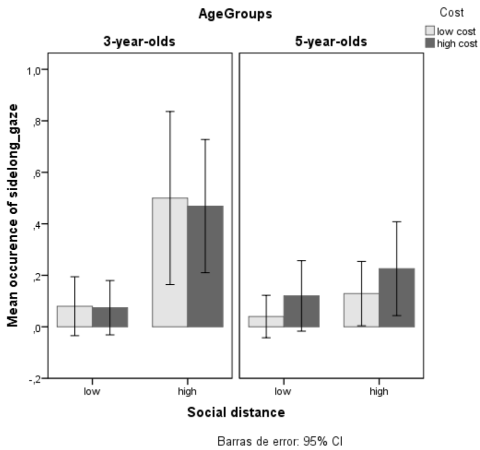


Figure 8. Mean occurrences of 'gaze at interlocutor' (top left panel), 'averted gaze' (top right panel) and 'sidelong gaze' (bottom mid panel) in each request across the three conditions, namely social distance, cost and age. Error bars indicate standard error.

Facial cues. The two facial cues to politeness that appeared in our data set were raised eyebrows and smiling. As can be seen in Figure 9, both raised eyebrows and smiling occurred more frequently in high social distance contexts. This difference was confirmed to be significant by GLMM analysis, which showed a main effect of SOCIAL DISTANCE on eyebrow raising (F (1, 220) = 15,134, p < .001) and smiles (F (1, 220) = 11,353, p < .001), with significantly more raised eyebrows and smiles in the high social distance situations.

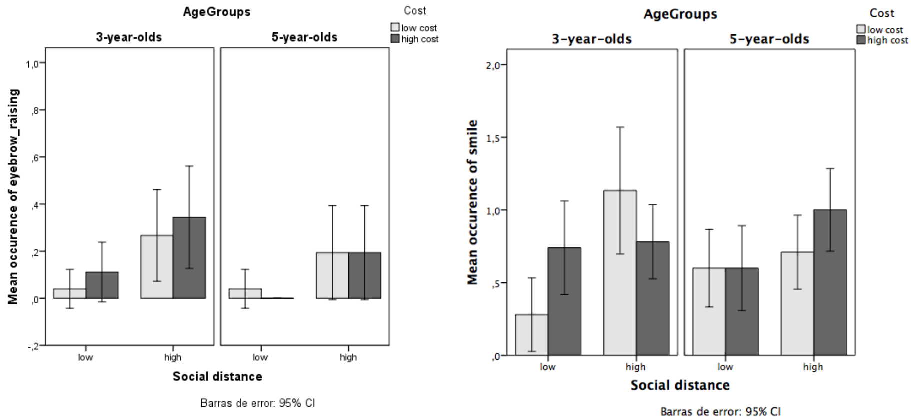


Figure 9. Mean occurrence of 'raised eyebrows' (left panel) and 'smile' (right panel) during requests across the three conditions, namely social distance, cost and age. Error bars indicate standard error.

Body cues. As noted above, body cues were broken down by body part. During requests, the children sometimes tilted their head to the side and inclined it forwards so that they faced down (see Figure 10). Analysis of our results detected a main effect of SOCIAL DISTANCE on both head 'tilt' (F (1, 220) = 7,122, p = .008) and head down (F (1, 220) = 5,032, p = .026).

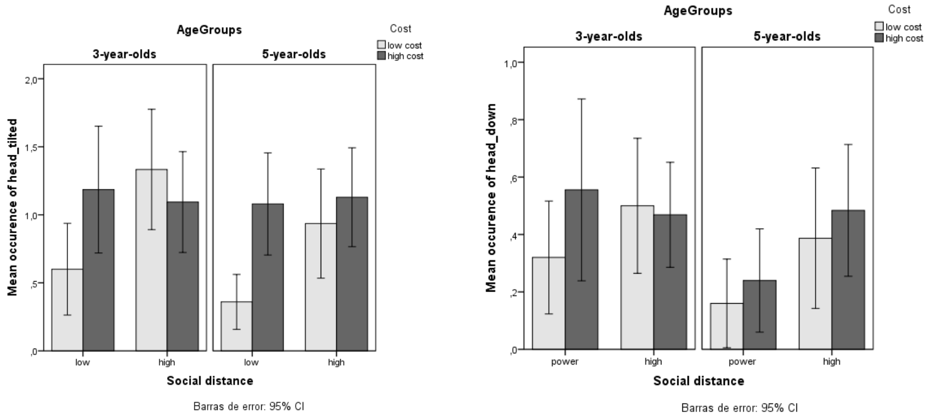


Figure 10. Mean occurrence of 'head tilted' (left panel) and 'head down' (right panel) during requests across the three conditions, namely social distance, cost and age. Error bars indicate standard error.

The distribution of shoulder movements made during requests, which were categorized as either shrugs or slouches, were analysed. No significant effects were found for slouched shoulders. The results for shrugs can be seen in Figure 11. Analysis of this data revealed a main effect of SOCIAL DISTANCE (F (1, 220) = 13,537, p < .001) on shoulder shrugs with more shrugs in high social difference situations and a main effect of AGE (F (1, 220) = 7,657, p = .006), with the younger age group producing more shoulder shrugs. However, there was a significant interaction between AGE and SOCIAL DISTANCE (F (1, 220) = 9,173, p = .003). Only in the older age group were there significantly more shrugs in high social difference situations (p = .001), this effect being completely absent in the younger group (p = .500).

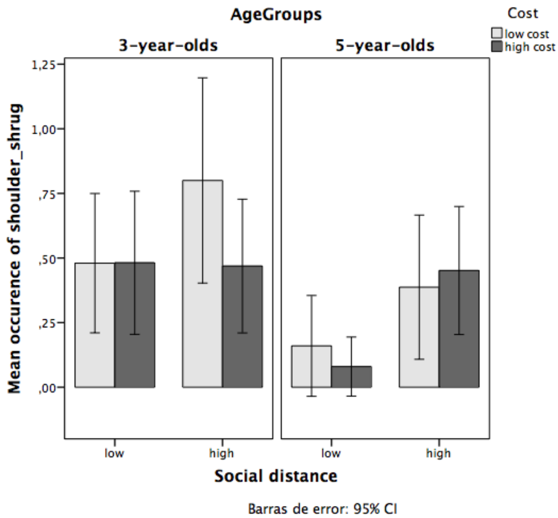


Figure 11. Mean occurrence of 'shoulder shrug' made by children while making requests across the three conditions. Error bars indicate standard errors.

The children were seen to hold their bodies in one of two ways while making requests, either forward or to the side. The results for these two trunk movements are shown in Figure 12 below. GLMM analysis showed a main effect of SOCIAL DISTANCE on both 'forward leaning' (F (1, 220) = 6,710, p = 0,010) and lateral leaning (F (1, 220) = 5,160, p = .024), with more occurrences in high social distance situations and a main effect of age (F (1, 220) = 6,298, p = .013), with more lateral leanings produced by younger children ($p < .008$). There was, however, an interaction between the effects of SOCIAL DISTANCE and AGE, (F (1, 220) = 6,810, p = .010). In the older group, high social/power distance caused more occurrences of laterally leaned trunks ($p < 0.003$), while there was no significant effect of SOCIAL DISTANCE in the younger group ($p = .779$).

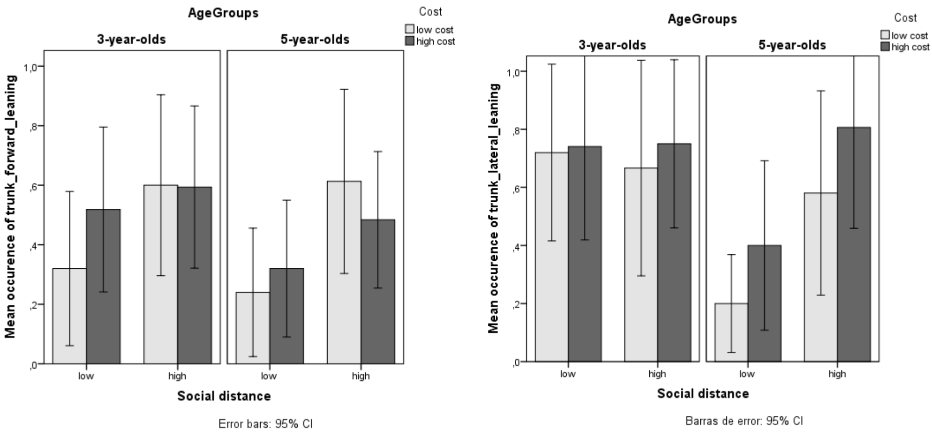


Figure 12. Mean occurrence of body postures ‘forward leaning’ (left panel) and ‘lateral leaning’ (right panel) adopted by children while making requests across the three conditions. Error bars indicate standard errors.

Finally, regarding the children’s use of ‘adaptors’ like touching their face or mouth, GLMM analysis showed a main effect of SOCIAL DISTANCE (F (1, 220) = 20,472, p < .001). As can be seen in Figure 13, significantly more adaptors were produced in the high social distance condition, meaning that both groups the children touched themselves or other objects significantly more when talking to a person of greater social distance (i.e., an adult).

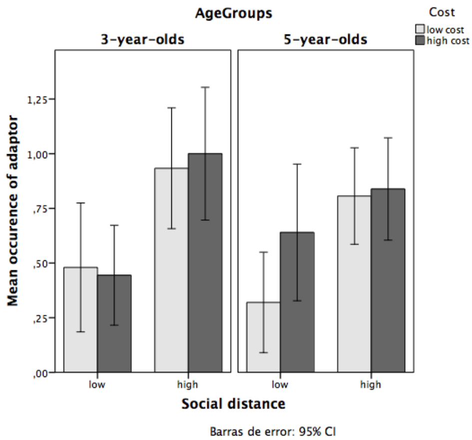


Figure 13. Mean occurrence of 'adaptor' gestures made by children while making requests across the three conditions. Error bars indicate standard errors.

5.4. Discussion and conclusions

The cross-sectional research presented here constitutes the first study to systematically document the multiple cues, both verbal and nonverbal, that 3- to 5-year-old children use to make a request more polite. We have also analysed the interaction between those cues and two parameters, social/power distance and the cost to face, as well as age. We hypothesized that (1) preschool children—like adults—would mitigate their requests in different ways depending on the degree of social distance between interlocutors and degrees of cost; (2) they would more be more likely to show politeness by means of prosodic, gestural and other body signals than by the use lexical or morphosyntactic markers; and (3) their repertoire of mitigation strategies to signal politeness would increase as they got older. Overall, our three hypotheses have been borne out by our results, with several of our findings being of particular interest.

With regard to our first hypothesis, our findings confirm that preschool children use a wide set of prosodic mitigation strategies, including rising intonation, slower speech rates, less jitter and more breathiness, to render requests appropriately more polite in contexts where either their interlocutor is socially distant from them, or their request implies a high cost in face to their interlocutor. Interestingly, in contexts involving a high power distance between the child and their interlocutor (an adult researcher, for example) the favoured strategy observed in our sample was a reduced speech rate, whereas rising intonation tended to be deployed more often in high cost contexts. Comparing these prosodic findings to results found for Catalan-speaking adults by Hübscher et al. (2017), it would seem that 5-year-old children can make use of much the same prosodic cues as adults. Nonetheless, because Hübscher et al. (2017) analysed adult politeness in interaction with only one social parameter, power distance, the results of the two studies are not strictly comparable. The present findings regarding intonation are more readily comparable to those made by Astruc et al. (2016), who showed that for adult Catalan-speakers high cost situations triggered more rising pitch patterns than low cost situations, but social distance did not have a significant effect on the choice of intonation contour. This means that, with regard to intonation and syllable duration, already at the age of three children use a phonological mitigation strategy similar to that seen in adults and by age five they can deploy most of the other phonetic cues to politeness in an adult-like way.

With respect to nonverbal signals, our study has shown that preschool children use a wide array of gestural and body signals to mark their polite stance towards an adult with higher social distance and/or to request an object with more cost to the interlocutor's face. Children produce significantly more eyebrow raises, smiles, adaptors, head tilts and head downs, raised shoulders and trunk lateral leanings in high social distance conditions than in low social distance conditions. Most of these cues have been found to be submission cues displayed towards a person with more power (Ellyson & Dovidio, 1985; Hall et al., 2005; Tree & Manusov, 1998). Two of the body postures or movements we analysed in this study—the head tilted to one side or facing down and shrugging shoulders—have been found elsewhere to be important cues in the expression of uncertainty (Roseano et al., 2016) and it is therefore reasonable to suppose that they could serve a mitigation function in requests. And indeed our results show that preschool children display significantly more tilted or lowered heads and raised shoulders (among other cues) when making a request in high social distance contexts.

In this context it is interesting to note that while prosody was mainly adapted in regard to increased cost, mitigating gestural and body cues were used predominantly in requests with high social distance. Clearly there was a strong social distance between the child participants and the adult experimenter. This possibly intimidating situation, next to being face-threatening, might have elicited such a high number of gestural and body cues on part of the children when requesting something from an unknown adult. On the same token, perception experiments would be necessary to delve into the question of the individual weight of each of these cues in the production of politeness cues.

Our second main finding relates to the use of lexical and morphosyntactic strategies used by preschool children to convey politeness relative to prosodic and gestural ones. While both age groups used the lexical mitigator *si us plau* ‘please’ significantly more in high social distance contexts, only the older children were able to vary their requests morphosyntactically by using indirect constructions. These results are comparable to previous studies that have found that children up to the age of five mainly use direct request strategies (see e.g. Aronsson & Thorell, 1999; Axia & Baroni, 1985; Bates, 1976; Bernicot & Legros, 1987; Ervin-Tripp & Gordon, 1986; Georgalidou, 2008; Hollos & Beeman, 1978; James, 1978; Nakamura, 1999; Nippold et al., 1982; Read & Cherry, 1978; Ryckebusch & Marcos, 2004). Furthermore, *please* as mitigation strategy appears relatively early in childhood, which could be explained through the heavy emphasis that parents and caregivers place on this lexical item (Gleason & Weintraub, 1976; Greif & Gleason, 1980; Nakamura, 2006b). Yet other request internal mitigation strategies which can be found in adult Catalan speech, such as the polite form of address *vostè* vs. the more informal *tu*, the choice of verbal forms (conditional vs. indicative) and other lexical hedges (see Fivero, 1976; Hübscher et al., 2017a; Payrató & Cots, 2011), are clearly lacking in preschool children’s requests. Taken as a whole, our results thus provide tentative confirmation that the number of prosodic and nonverbal politeness markers available to young children greatly outweighs their lexical and morphosyntactic repertoire.

This ties in with our final hypothesis regarding the expansion of this repertoire of politeness markers over time. Our findings are consistent with previous studies showing that children’s ability to deploy lexical and morphosyntactic politeness markers takes place slowly in and also after the preschool years (Baroni & Axia, 1989; Nippold et al., 1982; Pedlow et al., 2004). Comparing these results to adults' systems of politeness, children at age five clearly still have a long way to go. Most notably, mitigation strategies such as the use of conditionals or the use of *vostè* are completely absent from our data set. Furthermore, the younger children in our study produced more indirect requests in low social distance situations, contrary to what one would expect, suggesting that they had not fully grasped the mitigating value of this structure. By contrast, the children in the older age group produced more indirect requests in social distance situations, as one would see in adult discourse.

In fact, the present data shows that children actually already at age 3 exploit a wide range of gestural, body and prosodic cues to express politeness. Indeed, when looking at the development of those strategies, there is very little variation in children's use of gestural and other body markers of politeness-related meanings over the preschool years. Only lateral leaning and shoulder shrugs are used more by the older age group in high social distance situations. Regarding prosodic cues, although the younger children already use intonation and duration as mitigation cues in an adult-like way, the older children can manipulate a much broader and adult-like arsenal of phonetic features such as intensity, jitter and breathiness to convey politeness. Thus, our results as a whole make it obvious that preschool-aged children are adjusting their requests depending on who they are talking to and the degree of imposition that the request implies by employing a rich system of non-propositional markers before they are able to express similar meanings through lexical/morphosyntactic cues.

It is worth here reflecting on the general functions of gestures and other body signals and prosody in the marking of politeness-related meanings, their utility in communication and thus the desirability of studying them. There has been a strong focus on the role that gestures play in children's language development in the recent years. It has been proposed that 'gesture can serve as an additional window to the mind of the developing child' (Butcher & Goldin-Meadow, 2000a, p. 231). Also, the hypothesis has been formulated that gesture can play a causal role in language learning by providing children with the opportunity to practice communicating the ideas that underlie words and constructions that they are not yet able to express in speech (Butcher, 2003; Carpenter, Nagell, & Tomasello, 1998; Iverson & Goldin-Meadow, 2005; Özçalişkan & Goldin-Meadow, 2005, 2006a, 2006b). Gestures are indeed very prevalent in the data we have presented here; yet no less so is prosody able to shed light on children's awareness of and ability to signal sociopragmatic meanings. Thus the present findings provide further proof of the previously found scaffolding role that gestures (Bates et al., 1979; Bates et al., 1975; Beaupoil-Hourdel et al., 2015; Benazzo & Morgenstern, 2014; Guidetti, 2002, 2005) and prosody (Esteve-Gibert & Prieto, 2013; Papaeliou & Trevarthen, 2006) both provide children, allowing them to express pragmatic meanings first through those non-propositional cues before they have access to lexical cues. The children in the present study clearly used both gestural and prosodic cues more often and earlier than propositional cues to mitigate the possible face-threat of a request. Much as previous work has pointed to the tight temporal coordination of babbling with the first use of gestures (Esteve-Gibert & Prieto, 2014) and 1-year-old infants' ability to signal pragmatic meaning through both prosodic and gestural means (Esteve-

Gibert et al., 2017b), the present data has shed light on children's later pragmatic development by describing a similar parallel use of prosodic and nonverbal markers to signal social meanings.

Furthermore, the current results have direct implications for our understanding of children's socio-cognitive development. Given that gestural and prosodic cues appear earlier as markers of children's ability to signal social positioning, this would open up the possibility of intervening with young children at risk for delays in language and cognitive development.

To sum up, the current results support previous studies which maintained that in order to gain more insight into children's communicative development it is essential to take a multimodal approach since 'it is not only the verbal facet but also the nonverbal facet which together codetermine the meaning of an utterance' (Kelly, 2001: 345) and enable the child to better assimilate pragmatic aspects of conversation. Prosodic and nonverbal gestural cues are communicative strategies which are exploited early on in order to signal sociopragmatic meanings which so far have hardly been taken into account in studies of how children acquire sociopragmatic skills. The results of this paper have clear implications for our understanding of the development of sociopragmatic competence in children, particularly their acquisition of the social significance of politeness marking.

There are several future avenues of investigation. While it is assumed that children will apply the same culturally specific behavioral patterns obtained in the semi-elicited contexts used in the present study to more natural speaking contexts, there may well be differences between the two. It might therefore make sense to cross-check experimental data like this against completely spontaneous data in order to achieve a more comprehensive analysis of politeness in requests. Apart from that, while in the present study only one speech act has been elicited, future research should investigate whether the trends in children's multimodal signalling are equally manifest across different speech acts. Also, the present study is of a certain relevance for the fields of intercultural and second language acquisition learning, since it opens up several interesting questions. For example, it would be of interest to obtain a more comprehensive picture of the mitigating gestural and prosodic cues that are used to express politeness intra-linguistically and crosslinguistically, and use the present experimental paradigm to test the results from a cross-cultural perspective, fleshing out whether there are similarities between children's development of indexing politeness-related meanings in different languages and cultures.

6

Chapter 6: General discussion and conclusions

This thesis reported on four empirical studies that each discuss the relevance of prosody and gesture in children's perception and production of knowledge state and politeness. In this concluding chapter, first a brief summary of all four studies is provided (6.1.) followed by a comparison of the findings with the previous literature, by fleshing out various novelties encountered in this thesis. Specifically, the subsequent highlights of the thesis will be discussed: prosody and gesture are precursors in pragmatic meanings (6.2.), prosody and gesture act as sister systems (6.3.) and pragmatic development is crucially multimodal (6.4.). Then a more tentative discussion on the universality/language-specific nature of the pragmatic meanings examined follows (6.5.), and this thesis will end with suggestions for future work (6.6.) and with a general conclusion (6.7).

6.1. Summary of findings

The general aim of this thesis was to assess the role played by prosody, facial and body gestures in children's pragmatic development. In fact, this is the first time that prosodic and gestural development are assessed in a joint fashion and are compared with lexical development by focusing on two related pragmatic meanings, knowledge state and politeness. By taking this multimodal perspective, different methodological approaches in four independent but complementary studies were adapted to assess how preschool children access these meanings both through comprehension and production. While (a) Chapters 2 and 3 focused on preschool children's development of knowledge state (b) Chapters 4 and 5 concentrated on the development of politeness, with always one chapter focusing on comprehension and the following on production. Subsequently, the main findings will be briefly summarized before moving on to the more general discussion along with the relevance of the present findings.

In the first empirical chapter, entitled ‘Intonation and gesture as bootstrapping devices in speaker uncertainty’ (Chapter 2), the goal was to study children’s ability to understand another speaker’s uncertain knowledge state on the basis of different cues, uncertainty encoded through intonation (rising intonation), gesture (raised eyebrows, tilted head and raised shoulders) and lexical cues (epistemic adverb *potser* ‘maybe’). This study was inspired by the work of Armstrong et al. (accepted), which demonstrates a parallel development of intonation and facial gesture in children’s development of incredulity. The main aim of this research was to understand more the interplay between the three different epistemic markers (lexicon, intonation and gesture) of another type of belief state, namely knowledge state, in a more in-depth manner. The study in Chapter 2 thus aimed to find out how preschool (3- to 5-year-old) children use audio, visual or audiovisual cues in the comprehension of a speaker’s knowledge state and how their understanding develops. In order to investigate the weight of the different cues, two different conditions were constructed: one intonation only and another lexicon only, and each of these conditions contained three different modes of presentation (audio-only, visual-only and audiovisual). The results showed first that both the younger and the older children performed better when visual information was present (i.e., the visual-only and the audiovisual conditions). Second, children were more sensitive to intonational cues encoding a speaker’s epistemic state than compared to lexical cues. Children who received the intonation condition performed significantly better than those who received the lexicon condition (both in the audio-only presentation modality). This study demonstrates for the first time the strong role played by intonation in 3-year-olds' access to uncertainty meanings, and it also adds further evidence to the importance of gesture in children's acquisition of pragmatic meanings such as uncertainty.

The second study entitled 'Children's signaling of uncertainty: prosody, face and body cues come first' (Chapter 3) explored, from a production point of view, how 3- to 5-year-old preschoolers signal their own knowledge state and how it develops multimodally in a guessing game. In order to control their epistemic access to different objects, two conditions were created; one contained easy objects with which the children were previously familiarized (they had to touch and name the objects) while the other was made up of more difficult objects (the children did not see these prior to the experiment). Apart from eliciting children's expression of their knowledge state in the guessing game, children also were asked to report on their own knowledge state. The results demonstrated that 3– to 5-year-old children accurately use multimodal and prosodic markers before being able to explicitly reflect on their own certainty. When venturing a guess, 3-year-old children signal their uncertainty prosodically by employing fillers, lengthenings and also intonational marking. Furthermore, they signal their uncertainty gesturally by employing different body gestures, such as raised eyebrows, head tilt, and, to a more limited extent, other facial cues including wrinkled nose and stretched lips. While lexical marking was absent in the younger age group, at 5 years children started to employ few lexical markers of uncertainty, such as epistemic verbs (*crec que* 'I think') or epistemic adverbs (*potser* 'maybe'), and the use of the prosodic and gestural cues further increased in the older age group as well as in the adult control group. Furthermore, children multimodally signaled their uncertainty significantly earlier than they were able to self-assess their own knowledge state. This study was therefore able to demonstrate for the first time the rich interplay of children's employment of prosodic and gestural markers, the whole depth of children's use of different epistemic makers as well as how both prosody and gesture play a precursory role in this multimodal pathway into knowledge state marking.

In our third study, 'Three-year-olds infer polite stance from intonation and facial cues (Chapter 4), children's early sensitivity to another type of pragmatic meaning that can be encoded through intonation and gesture was tested, namely a speaker's polite stance in a request speech act. Similar to the first study, children were assessed in three different conditions: audio-only, visual-only and audiovisual (this time in a between subject design). Importantly, the lexicon remained the same during the two requests that children were exposed to at the same time. The results showed that in all three conditions children at three years of age could infer a speaker's polite stance solely on the basis of intonation and facial cues (polite smile). In the present task, both intonation and facial cues were equally strong in children's understanding of a speaker's polite stance expressed in requests. This study demonstrated, for the first time, the simultaneous pragmatic force of intonation and gesture for children to infer a speaker's polite stance at an age when they still struggle to report the difference between lexically and morphosyntactically encoded politeness markers.

In our fourth study entitled ‘Preschooler's development of polite stance: how prosody, face and body cues pave the way’ (Chapter 5), preschool-aged 3- to 5-year-old children’s ability to signal politeness-related meanings was examined from a multimodal point of view. In order to test whether children would vary in formulating their requests depending on social conditions, the sociopragmatic parameters of the requestive task were controlled for (e.g., requesting an object from a classmate vs. an experimenter and requesting something with low cost vs. high cost). The results of the study showed that 3- to 5-year-old children use a set of mitigating prosodic and gestural strategies to encode politeness-related meanings earlier and more often than they use lexical or morphosyntactic markers. While children use significantly more gestural cues in high social distance situations, they mostly manipulate their prosody (lower intensity, longer duration, lower mean pitch in older children) when the cost of the action is high. Our results suggest that prosody, gesture and other body signals show an essential first step in the development of children’s expression of sociopragmatic competence. The study demonstrates, for the first time, the strong and joint role played by pragmatic prosody and gesture in preschool children’s multimodal pathway into politeness marking.

In sum, the results of the three cross-sectional studies (Chapters 2, 3 and 5) revealed that both prosody and gesture play a precursory role in children’s development of the complex pragmatic skills under investigation. Furthermore, it was also established that when lexical content was controlled for (Chapter 4), children showed an early ability to infer politeness-related meanings through prosody and facial cues. Therefore, the overall results point to the common core that prosody and gesture are in pole position in children's pragmatic development.

6.2 Prosodic and gestural cues as precursors to children's comprehension and production of pragmatic meanings

Taking the studies together, a general picture emerges that preschool-aged children first comprehend and produce knowledge state (uncertainty) and politeness-related meanings through gestural and prosodic cues before doing so through lexical markers. Importantly, while previous research illustrated how, on the one hand, pointing gestures play a precursory (e.g. Acredolo & Goodwyn, 1985; Bates, 1976; Bates et al., 1979; Iverson et al., 1994; Özçalişkan & Goldin-Meadow, 2005) and facilitating role (Capirci et al., 1996; Iverson & Goldin-Meadow, 2005) in language development and, on the other hand, research on children's speech development showed that different prosodic cues play a boostrapping role (Gleitman et al., 1988; Morgan & Newport, 1981; Pinker, 1987), these two different phenomena were usually studied separately in the literature and the main focus was on the role these cues play in infants‘ early development.

The results reported in this thesis provide clear evidence that, similar to children's early ability to understand and produce certain language aspects, such as early speech act differentiations first through pointing gestures and prosody (Esteve-Gibert et al., 2017b), prosody and gesture patterns also play a precursory role later during pragmatic development in the preschool years (specifically in the case of more complex pragmatic meanings, namely knowledge state and politeness). More precisely, all four studies presented in this thesis have produced converging evidence that prosody and gesture appear as precursors in children's pragmatic development. In a nutshell, the results show that by three years old children are able to infer a speaker's uncertainty through gestural cues and significantly better through intonational cues compared to lexical cues (Chapter 2) and are also better able to signal their own uncertainty earlier through prosodic and gestural markers than through lexical markers (Chapter 3). Similar findings were obtained for children's development of politeness-related meanings. 3-year-old children were able to understand a speaker's polite stance well through intonation and gestural cues (Chapter 4). They were similarly using those markers in order to express their own interpersonal positioning in relation to other people, better and much more dominantly, compared to lexical and morphosyntactic markers which convey politeness (Chapter 5).

With studies from both the comprehension and production side, those results can be compared in the present thesis, facilitating a good approximation to the development of the two abilities. Especially remarkable is the consistency of the results across comprehension and production abilities as prosody and gesture emerge as pragmatic precursors in both the development of knowledge state marking (particular uncertainty) and politeness-related meanings. Interestingly, it can be observed that both pragmatic meanings mature parallel to one another. It seems that children first understand and produce both uncertainty and politeness- related meanings through prosody and gesture before children have acquired the capacity to detect and produce these meanings through lexical/morphosyntactic marking in speech. Therefore, there seems to be a parallel development of the pragmatic meanings whose relevance will be further discussed in the next section (6.3).

A general view in language acquisition exists that perception comes before production (Clark, 1993, p. 246); however, there have been cases where the reverse is true, e.g., the acquisition of pronouns (Hendriks & Koster, 2010) and evidential morphology in Turkish (Ünal & Papafragou, 2016). Ünal & Papafragou suggested that evidential comprehension is delayed by the development of mental perspective-taking abilities needed to compute another person's knowledge sources compared to their own. As evidence they state that while the children in their experiments have difficulties assessing another person's evidence in non-linguistic tasks, this difficulty disappears when the task involves assessing their own evidential source. In the case of knowledge state in Chapters 2 and 3, it seems that while children were neither able to understand nor to produce the lexical epistemic marker *maybe* in the younger age group (3-4.5. years), they were able to do so at 5 years. Consequently, this is also the age at which they start to produce lexical epistemic cues to knowledge state. However, when looking at the prosodic and gestural markers of knowledge state, it becomes clear that early preschool-aged children are able to use facial cues and intonation to understand another person's intention as well as express their own intention early on. However, a strict comparison between children's perception and production abilities is not possible since, in the case of production, children signal their own knowledge state; in the case of most comprehension studies, they have to detect another person's knowledge state. Similarly, in dealing with politeness, children themselves produce a polite stance, and they have to take another person's perspective in order to assess their politeness. Taken together, children dispose of a rich set of strategies to communicate and understand uncertainty or politeness well before they are able to express this through lexical devices; we thus propose that both prosody and gesture play a precursory role in children's pragmatic development of understanding and signaling knowledge states and politeness-related meanings.

However, we would like to take this one step further and propose that both prosody and gesture most probably have a bootstrapping function in children's pragmatic development. Evidence for this can be found in Chapter 3, as children were found to be able to signal their own knowledge state long before they were able to verbally report on it (see also, e.g., Kim et al., 2016). Future work will need to test this issue further to be able to make a more conclusive statement (see Section 6.6. for suggestions of how to assess this topic). The bootstrapping claim would go hand in hand with one of the main principles of the grounded or embodied cognition theory (Barsalou, 2008), which advocates that cognition is grounded in multiple ways. Recent work on embodied cognition states that language and body movements are supported by the same neural substrates (e.g. Glenberg & Kaschak, 2002; Pulvermüller, Hauk, Nikulin, & Ilmoniemi, 2005). The environment and the body as external informational structure is used by the cognitive system to support internal representations (Barsalou, Simmons, Barbey, & Wilson, 2003; Niedenthal, Barsalou, Winkielman, Krauth-Gruber, & Ric, 2005). Evidence from neurophysiology has yielded evidence that the self-performance of a gesture when learning verbal information leads to the formation of sensorimotor networks which represent and store words in one's native language (Masumoto et al., 2006) as well as in foreign languages (Macedonia, Müller, & Friederici, 2011). Furthermore, while prosody has not to our knowledge been included in embodied theories of cognition, it would be reasonable to integrate it. As is known that gesture can depict and represent different semantic meanings, it can also structure information in the discourse and also indicate pragmatic implicatures. Gesture has a great deal in common with prosodic properties in speech (see also the next section, 6.3. for a further discussion), and prosody might work in parallel with gesture. Following this theory, it could be expected that providing children with embodied representations of a speaker's uncertainty or politeness would help them to improve their ability to access another person's knowledge or interpersonal state, as compared to solely exposing them to lexical marking of these pragmatic meanings (see Section 6.6. for an explanation of current ongoing work on this topic).

6.3. Prosody and gesture as sister systems: developmental evidence

In addition to working as precursors in children's pragmatic development, prosody and gesture can seem more closely related than previously assumed and to be so-called sister systems; this may have important consequences for the literature on the gesture-primacy hypothesis and on language development more generally. First of all, as indicated in the introduction (see Section 1.4), research has shown that speech and gesture constitute a single communicative system which is well integrated semantically, pragmatically and temporally (Kendon, 1980; Levinson & Holler, 2014; McNeill, 1992). As one particular element of speech, prosody shares a great deal with gesture, namely that they are related on different levels. At the temporal level, gesture movements of speakers have been proven to be tightly coordinated with the prosodic structure in speech; pitch accents function as anchoring places for the prominent phases in body movements (e.g., Ambrazaitis & House, 2017; De Ruiter, 1998; Esteve-Gibert, Borràs-Comes, Asor, Swerts, & Prieto, 2017a; Esteve-Gibert & Prieto, 2013; Hadar, Steiner, Grant, & Rose, 1983; Ishi, Ishiguro, & Hagita, 2014; Leonard & Cummins, 2011). Moreover, prosodic patterns, together with hand gestures, have been demonstrated to possess a beat component through which speakers highlight prominent positions in speech (Krahmer & Swerts, 2007), and they can disambiguate syntactic constituents (Guellaï, Langus, & Nespor, 2014; Krivokapic, Tiede, Tyrone, & Goldenberg, 2016). In relation to the expression of pragmatic meaning, several studies have documented how pragmatic prosody and gesture serve similar functions in communication (e.g. Goldin-Meadow & Iverson, 1998; Kendon, 1980; McNeill, 1992). A certain consensus prevails in the literature on adults that prosody and gesture (taking a more holistic approach to gesture, including other articulators next to manual gestures) simultaneously convey intentional meaning in adult speech and gestures have been regarded as the visual correlate of prosody (see overview on audiovisual prosody (Krahmer & Swerts, 2009). For example, in the case of prominence and focus, speakers apply eyebrow flashes, head nods and beat gesture, tightly aligned with speech (Dohen & Loevenbruck, 2009; Krahmer & Swerts, 2007; Swerts & Krahmer, 2008); in the case of question intonation, similar observations have been made (Borràs-Comes & Prieto, 2011; Srinivasan & Massaro, 2003). Moreover, some studies have confirmed that prosodic and gestural patterns mutually influence the processing of speaker's emotions, beliefs and attitudes (Ekman, 1979;

Kendon, 2004; Poggi, D'Errico, & Vincze, 2013).

The majority of language studies on pragmatic development have focused heavily on language proper, leaving out both context and multimodal expression, which are known to contribute to the construction of utterances and meaning. As mentioned in the introduction, there has been limited output on the intersection between gesture and pragmatics as well as on the intersection of prosody and pragmatics. In general, it remains a relatively unexplored field, which will hopefully attract more research in the future. Subsequently, numerous reasons will be outlined as to why more intersection between the two fields has not occurred, and this will be complemented with research providing clear evidence that prosody and gesture have much more in common than previously assumed. Evidence from development has demonstrated that the temporal alignment of gesture and prosodic prominence mature early in language and cognitive development. For example, 6-month-old infants have been shown to couple rhythmic arm movements with babbling (Ejiri & Masataka, 2001; Iverson & Fagan, 2004). Similarly, during the first year of life, infants synchronize their pointing gestures temporally with the speech sounds produced simultaneously (Butcher & Goldin-Meadow, 2000b; Esteve-Gibert & Prieto, 2014; Murillo & Belinchón, 2013); and for a summary of the role of gesture-prosody integration in language development, see Igualada (2017). Furthermore, in relation to learning, infants tend to respond more frequently to adults' interests when they produce synchronized gesture-speech combinations (Miller & Lossia, 2013; Miller & Gros-Louis, 2013). Additionally, Esteve-Gibert and Guellaï (2018) have shown that prosody and gesture overlap in relation to the linguistic functions they are used for in development. They review how infants use visual correlates of prosody to do the following things on the production side: to segment the speech stream (e.g. Guellaï, Streri, Chopin, Rider, & Kitamura, 2016; Kitamura, Guellaï, & Kim, 2014), to organize information at the discourse level (Capone & McGregor, 2004; Mathew, Yuen, Shattuck-Hufnagel, Ren, & Demuth, 2014; Nicoladis et al., 1999) and to express emotions, intentions, and beliefs (Aureli et al., 2017; Berman et al., 2016; Esteve-Gibert & Prieto, 2014; González-Fuente, 2017; Sullivan & Lewis, 2003a). Recent research by Armstrong et al. (accepted) actually shows that children's development of incredulity understanding through facial gesture and intonation does seem to run in parallel. The results of this thesis clearly contribute more evidence to these findings. The empirical evidence obtained by the present investigation has consistently shown that gestural and prosodic cues jointly lead the way to pragmatic development. From a developmental perspective, the research presented here has shown that children simultaneously exploit prosodic and gestural means to understand and signal a speaker's knowledge state and politeness through a rich set of different cues.

While we are not the first to make a claim regarding the close relationship between prosody and gesture, we will briefly outline how we would like to expand and expound previous proposals. Certain early studies maintained that gestures and intonation are closely related parts of a single gestural complex. To back up this claim, Bolinger (1986) pointed out that if intonation is affected by acquired language disorders, gestures are usually similarly affected, suggesting a very close connection between intonation and gesture (Cruttenden, 1997, p. 177). Recently, Snow (2017) proposed that intonation and gesture are sister systems, i.e., they share the same pragmatic function; in order to support his hypothesis, he focused on the late "prelinguistic" and early "linguistic" age frame. Snow (2017) (based on (Bruner, 1974/1975) suggests that intonation is a hybrid system which does not fit the traditional classifying schemas nor the schemas based on descriptors like nonverbal vs. verbal. He put forth that intonation works as a bridge between gesture and language. As an example, Snow (2017) discusses the protodeclarative and how the pitch prominence of an utterance vocally designates or "points to" the information focus of the utterance, similar to a pointing gesture. Building on Snow (2017), this thesis expands the proposal of prosody and gesture as a sister system in two important ways. First, next to the main focus on speech act meaning, the results show that other pragmatic meanings are also simultaneously conveyed through prosody and gesture, such as knowledge state and politeness meaning, and their development appears to run in a similar fashion. Second, while Snow (2017) focused on manual gestures and intonation, the findings of this thesis make it clear that the scope of investigation should be widened by looking at other non-manual gestures (head, face, shoulders) and also taking into account other prosodic features in addition to pitch, i.e., intensity, duration and voice quality.

While the results of this thesis provide further evidence on the importance of gesture in language development, there is also converging evidence found in prosody; thus, a more reconciling perspective on the foundation of language evolution and language development as recently proposed by several researchers (Levinson & Holler, 2014; Vigliocco, Perniss, & Vinson, 2014) seems more suitable in order to interpret our results.

Looking at language from an evolutionary perspective, researchers have tried to explain how language has developed and the role played by gesture in this process. One dominant theory on the origins of human language is the gesture-primacy hypothesis (see Hewes, 1999 for a historic overview). According to the *gesture-primacy* hypothesis, humans first communicated in a symbolic way using gesture (e.g., movement of the hands and body to express meaning). The gesture-primacy hypothesis suggests that spoken language emerged through adaptation of gestural communication (Corballis, 2002; Hewes, 1999). Fundamental to this view is the idea that gesture and speech emerged *sequentially*. Most evidence collected in favor of this hypothesis comes from studies on non-human primates and great apes. It has been established that, within each monkey and ape species, individuals tend to have the same basic vocal repertoire; when raised in isolation, they still produce calls that are typical of their own species (Tomasello, 2008). However, the gesture repertoire of apes has been found to vary much more in each individual than the vocal repertoire (Pollick & De Waal, 2007), and researchers have been successful in teaching chimpanzees manual actions with the help of symbolic gestures derived from American Sign Language (Gardner & Gardner, 1969). Shifting from phylogenetic development to ontogenetic development, studies on the emergence of speech and gesture in human infants have demonstrated that infants produce pointing gestures before producing the first words (Behne et al., 2012; Camaioni et al., 2004; Liszkowski et al., 2006). Also after their first birthday, when children start to produce their first words, they make pointing gestures (e.g., point at elephant) alongside one-word utterances (‘eat’) up to three months before producing two-word utterances (e.g., ‘elephant eats’). Thus, referentiality arises first in gesture and only later in speech. This has led many researchers to conclude that gesture paves the way for early language development (Butterworth, 2003; Iverson & Goldin-Meadow, 2005).

Proposing a more integrative view, Levinson & Holler (2014) suggested that the capacity to use language evolved from the interaction of both oral and gestural systems being used as one integrated system; they convincingly outlined that speech and gesture go hand in hand and that there is a tight integration of different modalities in human communication. They put forth evidence, which both behavioral and neurobiological findings show, that processing of language is inherently multimodal. There is clear evidence for a tight integration between vocal and visual channels in processing. It has been shown that the brain's neural responses are related for the processing of speech and iconic gestures (Ozyurek, 2014). Skipper (2014) adds that hearing is deeply multimodal. He analyzed the auditory cortex in meaningful linguistic versus non-meaningful auditory contexts as well as in speech-only versus speech and gesture contexts. He found that the auditory cortex is less active in multimodal and more meaningful contexts. This suggests that our brain constructs meaning primarily predictively by using information from context – auditory or visual – in order to generate predictions. All this evidence points to an argument in favor of a possibly simultaneously development of speech/prosody and gesture.

6.4. Widening the lens of language: a multimodal approach to pragmatic development

So what do the findings of this thesis tell us about the relationship between gesture, prosody, pragmatics and language development more generally? Most importantly the results provide further evidence as to why a multimodal approach to language is crucial. As has been recognized by the usage-based approach to language (Tomasello, 1992, 2000; Tomasello, 2001) in regular circumstances, it is in multimodal settings that children acquire language and develop form-meaning mappings. However, research on children's acquisition of pragmatics has been mainly unimodal, focusing primarily on children's acquisition of language without taking a multimodal approach. Recent research has yielded clear evidence for gesture preceding and predicting the acquisition of structure in speech (Goldin-Meadow, 2014). Furthermore, Esteve-Gibert et al. (2017) determined that infants are able to infer meaning from multimodal cues in the communicative act, such as prosody and gesture, before they are able to use spoken language forms. Similarly, in relation to children's pathway into the expression of negation, Beaupoil-Hourdel et al. (2015) illustrated how children undergo a change from expressing negation through embodiment then later move on to symbolic negation, where gesture and speech are completely integrated into composite multimodal productions.

Thus, building upon previous research on audiovisual prosody, speech processing and production, this thesis adds evidence of the inherently multimodal developmental process of children's pragmatic skills. The findings demonstrate that children understand and produce pragmatic meanings, such as knowledge state and politeness, earlier through prosody and gesture. It also provides evidence that pragmatic meanings are communicated and understood in a multimodal fashion. Moreover, these cues are seen to afford children early access to meanings and probably help them gain access to the form-meaning mapping of lexical pragmatic items acquired later. Furthermore, it is crucial to discuss the nature of prosodic and gestural cues, which, in the present case, encode knowledge state and politeness. Regarding gesture, most studies have focused on the role of manual gestures (deictic and partially iconic) in children's language acquisition process; however, in the present thesis, a more holistic approach was adopted, focusing on different gestural cues displayed by different articulators such as head and facial cues. The results in all four empirical chapters show that children are not only able to understand and produce knowledge state and politeness-related meanings earlier, but they also clearly benefit from having multimodal cues present when they have already acquired the lexical cues (Chapter 2).

In previous research with respect to prosody, the primary focus has been on the relationship between pitch and pragmatic meanings, leaving aside other prosodic cues, which can contribute to the meaning-making process when inferring or expressing a certain intention in conversation. In the case of knowledge state, and in even more detail in the case of politeness in Chapters 4 and 5, it has been demonstrated that children (and also adults in Chapter 3) mark pragmatic meanings not only through pitch, but also through other prosodic cues such as duration, intensity and voice quality. All in all, evidence shows that taking these phonetic cues into account yields a much more complete picture of the acquisition process.

In sum, there is sufficient evidence for language being intrinsically multimodal. Future theoretical proposals on language development and, more particularly on pragmatic development, should explore the interactions of different properties relevant to multimodal communication.

6.5. Language specific vs. universally applicable prosodic and gestural cues to knowledge state and politeness?

One further consideration, which this chapter would like to address, is to what extent the pragmatic meanings investigated in this thesis are universally encoded through similar prosodic and gestural cues in different languages. This final section is of a more speculative nature, as the current results do not allow for the inference of any firm conclusions.

Under the assumption that words are arbitrarily linked to objects and actions in the world (de Saussure, 1983; Hockett, 1960), learning new vocabulary can be seen as a difficult task. It has been claimed that language, in addition to being arbitrary, is also crucially iconic (maintaining transparent links between spoken form and meaning). Iconicity would be especially helpful in cases where objects or events are not visually present, by providing 'sensory-motor' properties such as long vowels to refer to a tall person. There are several parts of language, which seem to be targeted by iconicity: in the phonology of words (onomatopoeia), in co-speech gestures that are iconic or even metaphoric (e.g., manual gestures evoking the shape or movement of the objects under discussion) or prosodic properties such as pitch or duration (e.g., through elongation of a vowel).

The role of iconicity in language development and language evolution has been addressed in recent works. Looking at spoken language acquisition, Imai, Kita, Nagumo, and Okada (2008) found that 3-year-old Japanese children were better at learning novel action words when the sound of the word matched the action as compared to when it did not. Furthermore, Imai and Kita (2014) provided evidence as to why sound-meaning mappings (sound-symbolism) play a crucial role in language development. Sound symbolic words are those that have an inherent non-arbitrary link between sound and meaning. One systematic sound meaning correspondence in English exists, for example, in the following cluster of words ‘glitter’, ‘glare’, ‘glow’ and ‘glistening’; they all have 'gl’ at the beginning, and their meanings are related to light (another example would be English sh-words). The sound-symbolism bootstrapping hypothesis proposes that sound symbolism helps children understand that perceived sounds refer to things in the world, and, in this way, it would help children to focus on specific form-meaning mappings (Imai & Kita, 2014).

The element of iconicity has also been addressed in relation to prosody. Perlman, Dale, and Lupyan (2015) tested the idea that not all sounds in language are arbitrary but that there might be some underlying iconicity. To investigate this, they asked pairs of students to invent new words for 18 contrasting ideas (such as *big, small, slow, fast, rough, smooth*). The partner was then given ten seconds to guess the concept the person was describing. They performed really well, since the person describing the concept included prosodic information, e.g., *up* was represented with a rising pitch while *down* had a falling pitch. *Slow* on the other hand had long duration and low pitch while *fast* had short duration and high pitch. *Smooth* had a high degree of harmonicity while *rough* had a high degree of noise. Furthermore, Ohala's Frequency Code (1984) has been regarded as universal and iconic since dominance is signaled by low frequency and submission is signaled by high frequency. According to this, one 'biological code' (see Gussenhoven, 2002 for the other two proposed biological codes, Effort Code and Production Code) a voice with low pitch signals dominance, and a high-pitched voice signals subdominance through the projection of body size. Considerable evidence has been accumulated which supports the idea that both animals and humans interpret pitch in relation to dominance (Puts, Gaulin, & Verdolini, 2006; Puts, Hodges, Cárdenas, & Gaulin, 2007). Looking at the nature of the prosodic and gestural cues that convey politeness-related meanings, there may be something inherently symbolic about them, which could be one possible reason why children are able to infer and encode knowledge state and politeness-related meanings earlier through prosody and gesture as compared to lexical cues. In a recent study conducted on Catalan adults (Hübscher et al., 2017a) as well as in Chapter 5 in a study on children, evidence was offered that the relation between pitch and politeness might be less straightforward than the one suggested by the Frequency Code; there might also be an overall prosodic attenuation strategy which is used more cross-linguistically. In line with recent findings in Korean, German and Russian (Grawunder et al., 2014; Kaori et al., 2016; Winter & Grawunder, 2012), but not as predicted by the Frequency Code (Ohala, 1984), Catalan speakers lowered their overall pitch height when applying formal polite register instead of using a higher overall pitch range. Rather than pitch that seems to be more variable, the different languages investigated encode politeness-related meanings through a set of more attenuated and softened prosodic correlates such as slower speech rate, lower intensity and more breathy voice quality.

Iconicity has also been extensively addressed in gesture and sign language research. Sign languages are particularly rich in iconic form-meaning patterns. Recent work has addressed the question of whether iconicity helps sing language acquisition (Thompson, Vinson, Woll, & Vigliocco, 2012). They found that iconicity predicts both early sign comprehension and production, taking into account familiarity, phonological complexity, imageability and concreteness. Iconic gestures in particular have also been hypothesized to be crucial to language evolution (e.g. Arbib, 2012; Armstrong & Wilcox, 2007; Tomasello, 2008). Airbib et al. (2009) argued that it is the coupling of gestural communication with enhanced capacities for imitation that made the emergence of protosign possible in providing the important scaffolding for protospeech in the evolution of protolanguage. Therefore, iconicity could strengthen the link between the linguistic form and the human experience; at the same time, it might help the learnability by providing the scaffolding for meaningful communication, as suggested by Thompson et al. (2012). However, more work needs investigate this as there is contradictory evidence by Özçalışkan et al. (2013) which shows that iconic gestures, unlike pointing gestures, do not precede the acquisition of first verbs, and generally few iconic gestures are found early in development.

In Chapter 5, similarly to the prosodic mitigation strategy used crosslinguistically to mark politeness in adults (see description above), it was found that children apply a range of mitigating gestural cues such as raised eyebrows, smiles, tilted head and forward leaning. Based on these findings, we propose that prosodic mitigation and gestural mitigation strategies are used crosslinguistically and can be considered important universal characteristics of polite register. It may be that a certain iconic element encountered in prosody and gesture constitutes the basis of a universal code that people tap into in order to express these concepts. Moreover, these codes might help bootstrap children's development of the form-meaning mapping of the later acquired lexical meanings. In this thesis, we examined Central Catalan in Chapter 2, 3 and 5 and American English in Chapter 4; we hope that future studies will address this topic in relation to additional languages in order to find more converging results.

6.6. Caveats, predictions and future directions

The work presented in this thesis opens up new avenues for future work, which will help to provide a more detailed picture of some of the various issues that have been analyzed here. In the future, it will be interesting to test the bootstrapping role of prosody and gesture in relation to pragmatic language development and also to other linguistic abilities. This could be addressed by a longitudinal study testing whether earlier abilities in the expression of pragmatic meanings conveyed through prosody and gesture can bootstrap children's later pragmatic abilities by using lexical strategies.

Also, future work should investigate in more depth individual differences along with the relation between children's pragmatic development and their Theory of Mind development. For example, previous results on uncertainty comprehension (see Moore et al., 1993), have found that children are able to begin to infer another speaker's uncertainty at four years. Moore et al. (1993) hypothesized that these results may be due to children's augmented ability to take perspective, and access their Theory of Mind. Theory of Mind refers to the ability to make predictions about another person's behavior and to infer one's own and others' mental states, i.e., the others' intentions, beliefs and desires (Premack & Woodruff, 1978 and many other thereafter). Theory of Mind abilities have been typically assessed by means of false-belief tasks, such as the Sally-Ann task (Baron-Cohen, Leslie, & Frith, 1985; Wimmer & Perner, 1983), and found that children are able to pass it by four years of age. However, since the false-belief task requires complex mental abilities, several studies have created less cognitively demanding tasks that investigate whether younger children show evidence of mind-reading abilities (Baillargeon, Scott, & He, 2010; Buttelmann, Carpenter, & Tomasello, 2009; Kovács, Téglás, & Endress, 2010; Onishi, 2005). Taken together, these studies show that infants younger than 3 or 4 years of age show mind-reading abilities when tested with cognitively easier tasks. Since, in our studies in Chapter 2 and 4, children at 3 years of age are able to begin to infer a speaker's knowledge state and polite stance on the basis of facial cues only and on the basis of intonation, future work should compare children's cognitive development and the interplay with linguistic abilities. This topic has been hardly addressed so far (see Armstrong et al., accepted, for an exception).

This thesis also has implications for future work in the area of language assessment and intervention. While this work has focused on children's use and understanding of two pragmatic meanings through prosody and gesture and alongside their lexicon, looking at language assessments, it becomes obvious that there is no standardized measure to test children's prosodic and gestural development. In regards to pragmatics, there is the Comprehensive Assessment of Spoken Language (CASL) (Carrow-Woolfolk, 2017), which is a comprehensive language test battery including one part dedicated to pragmatics. Generally speaking, future pragmatic tests should include prosodic and gestural aspects as well.

Furthermore, from a pedagogical perspective, our findings support the use of teaching and learning methods that implement more active multimodal strategies in the language classroom. A number of studies on second language acquisition, for example, provide evidence of the beneficial role of the use of pitch gestures (a specific type of metaphoric gesture that mimics melody in speech) in order to improve the production of intonation in Spanish (Yuan, González-Fuente, Baills, & Prieto, 2018, in press) or the acquisition of Mandarin Chinese tones (Baills, Suárez-González-N., González-Fuente, & Prieto, 2018, in press). Moreover, it has been shown that for L2-French learning children, hand clapping is beneficial to improve their pronunciation (Baills & Prieto, in preparation). Likewise it would be interesting to test whether the embodied use of gestural and prosodic markers of pragmatic meanings help in second language acquisition to facilitate pragmatic acquisition. On a related note, it will be useful to address the question of whether perceiving and enacting multimodal expressions of internal states and emotions can contribute to enhancing perspective taking skills. While previous training studies have shown the beneficial effect of language games and conversations about internal state terms on children's cognitive development (Ornaghi, Brockmeier, &

Gavazzi, 2011), they have not taken a multimodal perspective and have not considered the role of enacting internal states. In an ongoing study, we are carrying out a between-subjects training experiment with a subsequent comparison of pre-test and post-test scores related to mental state and pragmatic skills (see Pronina, Hübscher, Holler, & Prieto, in preparation).

Finally, in order to delve more deeply into the question of iconicity as well as universality, it will be of interest to compare the present findings with the development of those pragmatic skills under investigation with typologically different languages, such as Japanese or Turkish, which, in the case of knowledge state, have grammaticalized markers apparently acquired before lexical markers of (un)certainty (Matsui et al., 2006). Similarly, languages vary in the way they signal politeness, and it will be interesting to explore to what extent similar or different prosodic and gestural cues are used in comparable situations. In relation to this, another area worth examining more closely concerns the comparison of children's acquisition of pragmatic skills of taking a stance in oral languages with how these meanings are communicated through non-manual cues in sign languages. To sum up, future work on these topics will help us deepen our understanding of the interaction between multimodal pragmatic language development, cognition and culture.

6.7. General conclusions

The four studies in this dissertation show from various angles how children develop their multimodal pathway in the comprehension and expression of knowledge state and politeness-related meanings. First, the studies show that prosody and gesture act as precursors in children's pragmatic development of stance-related meanings. It is through prosody and co-speech gestures that preschool-age children are able to both understand and produce knowledge states and also politeness earlier than through lexical or morphosyntactic pragmatic markers. Second, prosody and gesture behave very similarly in children's pragmatic language development and seem to be sister systems working in parallel fashion. Third, as a direct implication of the first two points, pragmatic development has been shown to be a multimodal venture, and in order to understand the developmental process, more holistic approaches should be undertaken in the future. Thus, prosody and gesture are crucial pieces of the puzzle that children have to solve when developing (socio)pragmatic skills, and it is therefore indispensable to include them in any enquiry in this field.

References

Acredolo, L. P., & Goodwyn, S. W. (1985). Symbolic gesturing in language development. *Human Development, 28*, 40-49.

Aguert, M., Laval, V., Lacroix, A., Gil, S., & Bigot, L. L. (2013). Inferring emotions from speech prosody: Not so easy at age five. *PLoS ONE, 8*(12), 1-9. doi:10.1371/journal.pone.0083657

Aguert, M., Laval, V., Le Bigot, L., & Bernicot, J. (2010). Understanding expressive speech acts: The role of prosody and situational context in French-speaking 5- to 9-year-olds. *Journal of Speech, Language, and Hearing Research, 53*(6), 1629-1641. doi:10.1044/1092-4388(2010/08-0078)

Aikhenvald, A. (2004). *Evidentiality*. Oxford, UK: Oxford University Press.

Allwood, J. (1998). Cooperation and flexibility in multimodal communication. *Proceedings of International Conference on Cooperative Multimodal Communication*, Tilburg, The Netherlands, 113-124.

Allwood, J., Cerrato, L., Jokinen, K., Navarretta, C., & Paggio, P. (2007). The MUMIN coding scheme for the annotation of feedback, turn management and sequencing phenomena. *Language Resources and Evaluation, 41*(3-4), 273-287. doi:10.1007/s10579-007-9061-5

Ambrazaitis, G., & House, D. (2017). Multimodal prominences: Exploring the patterning and usage of focal pitch accents, head beats and eyebrow beats in Swedish television news readings. *Speech Communication, 95*, 100-113. doi:10.1016/j.specom.2017.08.008

Andersen, E. (1977). Young children’s knowledge of role-related speech difference: A mommy is not a daddy is not a baby. *Papers and Reports on Child Language Development, 13*, 83-90.

Andersen, P. A., & Guerrero, L. K. (1998). *Handbook of communication and emotion: Research, theory, applications, and contexts*. San Diego: Academic Press.

Arbib, M. A. (2012). *How the brain got language*. Oxford: OUP.

Armstrong, D. F., & Wilcox, S. E. (2007). *The gestural origin of language*. New York, NY Oxford University Press.

Armstrong, M. E. (2012). *The development of yes-no question intonation in Puerto Rican Spanish* (Doctoral dissertation). Ohio State University. Available from https://etd.ohiolink.edu/ñ

Armstrong, M. E. (2014). Child comprehension of intonationally-encoded disbelief. *Proceedings of the Boston University Conference on Language Development*, Boston, 25–38.

Armstrong, M. E. (2015). Accounting for intonational form and function in Puerto Rican Spanish polar questions. *Probus*. doi:10.1515/probus-2014-0016

Armstrong, M. E. (2016). Production of mental state intonation in the speech of toddlers and their caretakers. *Language Acquisition*. doi:10.1080/10489223.2016.1242588

Armstrong, M. E., Esteve-Gibert, N., Hübscher, I., Igualada, A., & Prieto, P. (accepted). Developmental and cognitive aspects of children's disbelief comprehension through intonation and facial gesture. *Submitted to First Language*.

Armstrong, M. E., Esteve-Gibert, N., & Prieto, P. (2014). The acquisition of multimodal cues to disbelief. *Proceedings of Speech Prosody*, 1134-1138.

Armstrong, M. E., & Hübscher, I. (2018). Children's development of internal state prosody. In P. Prieto & N. Esteve-Gibert (Eds.), *Prosodic Development in First Language Acquisition* (pp. 272-293): John Benjamins.

Armstrong, M. E., & Prieto, P. (2015). The contribution of context and contour to perceived belief in polar questions. *Journal of Pragmatics, 81*, 77-92. doi:10.1016/j.pragma.2015.03.006

Aronsson, K., & Thorell, M. (1999). Family politics in children's play directives. *Journal of Pragmatics, 31*(1), 25-47.

Astruc, L., Vanrell, M., & Prieto, P. (2016). Cost of the action and social distance affect the selection of question intonation in Catalan. In M. E. Armstrong, N. Henriksen, & M. M. Vanrell (Eds.), *Interdisciplinary approaches to intonational grammar in Ibero-Romance* (Vol. 93–114). Amsterdam: John Benjamins.

Aureli, T., Spinelli, M., Fasolo, M., Garito, M. C., Perucchini, P., & D'Odorico, L. (2017). The Pointing-Vocal Coupling Progression in the First Half of the Second Year of Life. *Infancy (Online First View)*.

Axia, G., & Baroni, M. R. (1985). Linguistic politeness at different age levels. *Child Development, 56*(4), 918-927.

Baillargeon, R., Scott, R. M., & He, Z. (2010). False belief understanding in infant. *Trends in Cognitive Sciences, 14*, 110–118.

Baills, F., & Prieto, P. (in preparation). Embodying rhythmic properties of a foreign language through hand-clapping help children to better pronounce words in that language.

Baills, F., Suárez-González-N., González-Fuente, S., & Prieto, P. (2018, in press). Observing and producing pitch gestures facilitates the learning of mandarin chinese tones and words. *Studies in Second Language Acquisition*.

Balcomb, F. K., & Gerken, L. (2008). Three-year-old children can access their own memory to guide responses on a visual matching task. *Developmental Science, 11*(5), 750-760. doi:10.1111/j.1467-7687.2008.00725.x

Bänziger, T., & Scherer, K. R. (2010). Introducing the Geneva Multimodal Emotion Portrayal (GEMEP) corpus. In K. R. Scherer, T. Bänziger, & E. B. Roesch (Eds.), *Blueprint for affective computing: A sourcebook* (pp. 271–294). Oxford: Oxford University Press.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition, 21*, 37-46.

Baroni, M. R., & Axia, G. (1989). Children's meta-pragmatic abilities and the identification of polite and impolite requests. *First Language, 9*(27), 285-297. doi:10.1177/014272378900902703

Barsalou, L. W. (2008). Grounded cognition. *Annual review of psychology, 59*(1), 617-645.

Barsalou, L. W., Simmons, W. K., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences, 7*(2), 84-91. doi:10.1016/S1364-6613(02)00029-3

Barth-Weingarten, D., Dehé, N., & Wichmann, A. (2009). *Where prosody meets pragmatics*.

Bates, E. (1976). *Language and context: the acquisition of pragmatics*. New York: Academic Press.

Bates, E., Benigni, L., Bretherton, I., Camaioni, L., & Volterra, V. (1979). *The emergence of symbols: Cognition and communication in infancy*. New York: Academic Press.

Bates, E., Camaioni, L., & Volterra, V. (1975). The acquisition of performatives prior to speech. *Merrill-Palmer Quarterly, 21*(3), 205-226.

Bavelas, J., Gerwing, J., & Healing, S. (2014). Including facial gestures in gesture-speech ensembles. 15-34. doi:10.1075/z.188.02bav

Bavelas, J., & Healing, S. (2013). Reconciling the effects of mutual visibility on gesturing: A review. *Gesture, 13*(1), 63-92. doi:10.1075/gest.13.1.03bav

Bavelas, J. B., & Chovil, N. (2000). Visible acts of meaning: An integrated message model of language in face-to-face dialogue. *Journal of*

Language and Social Psychology, 19(2), 163-194. doi:10.1177/0261927X00019002001
Beal, C. R., & Flavell, J. H. (1982). Effect of increasing the salience of message ambiguities on kindergartners' evaluations of communicative success and message adequacy. *Developmental Psychology, 18*(1), 43-48. doi:10.1037/0012-1649.18.1.43
Beattie, G., & Coughlan, J. (1999). An experimental investigation of the role of iconic gestures in lexical access using the tip-of-the-tongue phenomenon. *British Journal of Psychology, 90*(1), 35-56.
Beaupoil-Hourdel, P., Morgenstern, A., & Boutet, D. (2015). A child's multimodal negations from 1 to 4: The interplay between modalities *Negation and Polarity: Experimental Perspectives* (Vol. 1, pp. 95-123).
Beck, S. R., & Robinson, E. J. (2001). Children's ability to make tentative interpretations of ambiguous messages. *Journal of Experimental Child Psychology, 79*(1), 95-114. doi:10.1006/jecp.2000.2583
Beckman, M. E., Hirschberg, J., & Shattuck-Hufnagel, S. (2005). The Original ToBi System and the Evolution of the ToBi Framework. In S.-A. Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing* (pp. 9-54). Oxford: Oxford University Press.
Beckman, M. E., Hirschberg, J., & Shattuck-Hufnagel, S. (2010). The Original ToBi System and the Evolution of the ToBi Framework *Prosodic Typology: The Phonology of Intonation and Phrasing*.
Beckman, M. E., & Pierrehumbert, J. (1986). Intonation structure in Japanese and English. *Phonology Yearbook, 3*, 255-309.
Behne, T., Liszkowski, U., Carpenter, M., & Tomasello, M. (2012). Twelve-month-olds' comprehension and production of pointing. *British Journal of Developmental Psychology, 30*(3), 359-375. doi:10.1111/j.2044-835X.2011.02043.x
Benazzo, S., & Morgenstern, A. (2014). A bilingual child's multimodal path into negation. *Gesture, 14*(2), 171-202. doi:10.1075/gest.14.2.03ben
Berman, J. M. J., Chambers, C. G., & Graham, S. A. (2016). Preschoolers' real-time coordination of vocal and facial emotional information. *Journal of Experimental Child Psychology, 142*, 391-399.
Bernard, S., Proust, J., & Clément, F. (2014). The medium helps the message: Early sensitivity to auditory fluency in children's endorsement of statements. *Frontiers in Psychology, 5*. doi:10.3389/fpsyg.2014.01412

Bernicot, J., & Legros, S. (1987). Direct and indirect directives: What do young children understand? *Journal of Experimental Child Psychology, 43*(3), 346-358. doi:10.1016/0022-0965(87)90012-9

Biber, D., & Finegan, E. (1989). Styles of stance in English: Lexical and grammatical marking of evidentiality and affect. *Text-interdisciplinary journal for the study of discourse, 9*(1), 93-124.

Bitti, P. E. R., Bonfiglioli, L., Melani, P., Caterina, R., & Garotti, P. (2014). Expression and communication of doubt/uncertainty through facial expression. *Journal of Theories and Research in Education, 9*(1), 159-177.

Blake, J., Myszczyszyn, D., Jokel, A., & Bebiroglu, N. (2008). Gestures accompanying speech in specifically language-impaired children and their timing with speech. *First Language, 28*(2), 237-253. doi:10.1177/0142723707087583

Blum-Kulka, S., & Olshtain, E. (1989). Requests and Apologies: A Cross-Cultural Study of Speech Act Realization Pattern. *Applied Linguistics, 5*(3), 196-213.

Boersma, P., & Weenink, D. (2017). Praat: doing phonetics by computer [Computer program]. Version 6.0.30. Retrieved 22 July 2017 from http://www.praat.org/.

Bolinger, D. (1986). *Intonation and Its Parts: Melody in Spoken English*. Stanford: Stanford University Press.

Borràs-Comes, J., & Prieto, P. (2011). 'Seeing tunes'. The role of visual gestures in tune interpretation. *Laboratory Phonology, 2*(2), 355-380.

Borràs-Comes, J., Roseano, P., Vanrell, M. D. M., Chen, A., & Prieto, P. (2011). Perceiving uncertainty: facial gestures, intonation, and lexical choice. . *Proceedings of GESPIN.*

Bosch, L., & Sebastián-Gallés, N. (2001). Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy, 2*(1), 29-49.

Brey, E., & Shutts, K. (2015). Children use nonverbal cues to make inferences about social power. *Child Development, 86*(1), 276-286. doi:10.1111/cdev.12334

Briz, A. (2002). La estrategia atenuadora en la conversación cotidiana española. *Proceedings of Actas del Primer Coloquio del Programa EDICE*, Estocolmo, 17–46.

Broaders, S. C., Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2007). Making children gesture brings out implicit knowledge and leads to learning. *Journal of Experimental Psychology: General, 136*, 539-550.

Brown, L., & Prieto, P. (2017). (Im)politeness: prosody and gesture. In C. J., H. M., & K. D. (Eds.), *The Palgrave Handbook of Linguistic (Im)politeness*. London: Palgrave Macmillan.

Brown, L., & Winter, B. (in press). Multimodal Indexicality in Korean: "Doing Deference" and "Performing Intimacy" through Nonverbal Behavior. *Journal of Politeness Research: Language, Behaviour, Culture*.

Brown, L., Winter, B., Idemaru, K., & Grawunder, S. (2014). Phonetics and politeness: Perceiving Korean honorific and non-honorific speech through phonetic cues. *Journal of Pragmatics, 66*, 45-60. doi:10.1016/j.pragma.2014.02.011

Brown, L., Winter, B., Idemaru, K., & Grawunder, S. (2015). *The sound of honorific language: How speech indexes social meaning in Korean, Japanese, German and Russian*. Paper presented at the IPrA, Antwerp, Belgium.

Brown, P., & Levinson, S. (1987). *Politeness: Some universals in language use*. Cambridge, UK: Cambridge University Press.

Bruner, J. (1974/1975). From communication to language: a psychological perspective. *Cognition, 3*, 255–287.

Butcher, C. (2003). Pointing toward two-word speech in young children susan goldin-meadow *Pointing: Where Language, Culture, and Cognition Meet* (pp. 85-107).

Butcher, C., & Goldin-Meadow, S. (2000a). Gesture and the transition from one- to two-word speech: When hand and mouth come together. In D. McNeill (Ed.), *Language and Gesture* (pp. 235–257). New York: Cambridge University Press.

Butcher, C., & Goldin-Meadow, S. (2000b). Gesture and the transition from one- to two-word speech: When hand and mouth come together. In D. McNeill (Ed.), *Language and Gesture* (pp. 235–257). New York: Cambridge University Press.

Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen- month-old infants show false belief understanding in an active helping paradigm. *Cognition, 112*, 337–342.

Butterworth, G. (2003). Pointing is the royal road to language for babies. In S. Kita (Ed.), *Pointing: Where Language, Culture, and Cognition Meet* (pp. 9-34). Mahwah, NJ: Lawrence Erlbaum Associates.

Caffi, C. (1999). On mitigation. *Journal of Pragmatics, 31*, 881-909.

Calbris, G. (2003). From cutting an object to a clear cut analysis: Gesture as the representation of a preconceptual schema linking concrete actions to abstract notions. *Gesture, 3*(1), 19-46.

Call, J., & Carpenter, M. (2000). Do apes and children know what they have seen? *Animal Cognition, 3*(4), 207-220. doi:10.1007/s100710100078

Camaioni, L., Perucchini, P., Bellagamba, F., & Colonnesi, C. (2004). The role of declarative pointing in developing a theory of mind. *Infancy, 5*(3), 291-308.

Camras, L. A., Pristo, T. M., & Brown, M. J. K. (1985). Directive choice by children and adults: Affect, situation and linguistic politeness. *Merrill-Palmer Quarterly, 31*, 19-31.

Capirci, O., Iverson, J. M., Pizzuto, E., & Volterra, V. (1996). Gestures and words during the transition to two-word speech. *Journal of Child Language, 23*(3), 645-673.

Capirci, P., Contaldo, A., Caselli, M. C., & Volterra, V. (2005). From action to language through gesture: A longitudinal perspective. *Gesture, 5*(1-2), 155-177.

Capone, N. C., & McGregor, K. K. (2004). Gesture Development: A Review for Clinical and Research Practices. *Journal of Speech, Language, and Hearing Research, 47*(1), 173-186. doi:10.1044/1092-4388(2004/015)

Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen- through 18-month-old infants differentially imitate intentional and accidental actions. *Infant Behavior and Development, 21*(2), 315-330.

Carrow-Woolfolk, E. (2017). *Comprehensive assessment of spoken language, Second Edition (CASL-2).* . Torrance, CA: Western Psychological Services.

Cavalho, A., Dautriche, I., Millotte, S., & Christophe, A. (2018). Early perception of phrasal prosody and its role in syntactic and lexical acquisition In P. Prieto & N. Esteve-Gibert (Eds.), *The development of prosody in first language acquisition*. Amsterdam: John Benjamins.

Chafe, W. (1986). Evidentiality in English conversation and academic writing. In W. Chafe & J. Nichols (Eds.), *Evidentiality: The linguistic coding of epistemology* (pp. 261–272). Norwood, NJ Ablex.

Channell, J. (1994). *Vague language*. Oxford: Oxford University Press.

Chapman, H. A., Kim, D. A., Susskind, J. M., & Anderson, A. K. (2009). In bad taste: Evidence for the oral origins of moral disgust. *Science, 323(5918)*, 1222-1226.

Chen, A. (2010). Is there really an asymmetry in the acquisition of the focus-to-accentuation mapping? *Lingua, 120*(8), 1926-1939. doi:10.1016/j.lingua.2010.02.012

Chen, A. (2011). Tuning information packaging: intonational realization of topic and focus in child Dutch. *Journal of Child Language, 38*(5), 1055-1083. doi:10.1017/S0305000910000541

Chen, A., & Fikkert, P. (2007). Intonation of early two-word utterances in Dutch. *Proceedings of the XVIth ICPhS*, Dudweiler, 315-320.

Chen, A., Gussenhoven, C., & Rietveld, T. (2004). Language-specificity in the perception of paralinguistic intonational meaning. *Language and Speech, 47*(4), 311-349. doi:10.1177/00238309040470040101

Choi, S. (1995). The development of epistemic sentence-ending modal forms and functions in Korean children. In J. Bybee & S. Fleischman (Eds.), *Modality in Grammar and Discourse* (pp. 165-204). Amsterdam: John Benjamins.

Christophe, A., Nespor, M., Guasti, M. T., & Van Ooyen, B. (2003). Prosodic structure and syntactic acquisition: The case of the head-direction parameter. *Developmental Science, 6*(2), 211-220. doi:10.1111/1467-7687.00273

Church, R. B., & Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition, 23*(1), 43-71. doi:10.1016/0010-0277(86)90053-3

Cienki, A., & Müller, C. (2008). Metaphor, gesture, and thought. In R. W. Gibbs (Ed.), *The Cambridge handbook of metaphor and thought*. Cambridge: Cambridge University Press.

Cienki, A., & Müller, C. (2008). Metaphor, gesture, and thought. In R. W. Gibbs (Ed.), *Cambridge handbook of metaphor and thought* (pp. 483-501).

Clancy, P. (1985). The acquisition of Japanese. In D. Slobin (Ed.), *The cross-linguistic study of language* (pp. 373–524). Hillsdale, NJ: Lawrence Erlbaum.

Clark, E. T. (1993). *The lexicon in acquisition*. Cambridge: Cambridge University Press.

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

Clark, R., Hutcheson, S., & Buren, P. V. (1974). Comprehension and production in language acquisition. *Journal of Linguistics, 10*(1), 39-54. doi:10.1017/S0022226700003996

Cocks, N., Morgan, G., & Kita, S. (2011). Iconic gesture and speech integration in younger and older adults. *Gesture, 11*(1), 24-39. doi:10.1075/gest.11.1.02coc

Colletta, J. M., Guidetti, M., Capirci, O., Cristilli, C., Demir, O. E., Kunene-Nicolas, R. N., & Levine, S. (2015). Effects of age and language on co-speech gesture production: An investigation of

French, American, and Italian children's narratives. *Journal of Child Language, 42*(1), 122-145. doi:10.1017/S0305000913000585
Colletta, J. M., & Pellenq, C. (2004). Multimodal explanations in French children aged from 3 to 11 years. In N. Nippold & C. Scott (Eds.), *Expository discourse in children, adolescents, and adults: development and disorders* (pp. 63-97). New York: Psychology Press, Erlbaum, Taylor & Francis.
Colletta, J. M., Pellenq, C., & Guidetti, M. (2010). Age-related changes in co-speech gesture and narrative: Evidence from French children and adults. *Speech Communication, 52*(6), 565-576. doi:10.1016/j.specom.2010.02.009
Colonnesi, C., Stams, G. J. M., Koster, I., & Noom, M. J. (2010). The relation between pointing and language development: A meta-analysis. *Developmental Review, 30*(4), 352-366. doi:10.1016/j.dr.2010.10.001
Conrad, S., & Biber, D. (2000). Adverbial Marking of Stance in Speech and Writing. In S. Hunston & G. Thompson (Eds.), *Evaluation in Text. Authorial Stance and the Construction of Discourse* (pp. 56-73). Oxford: Oxford Linguistics.
Cook, H. M., & Burdelski, M. (2017). (Im)politeness: Language Socialization. 461-488. doi:10.1057/978-1-137-37508-7_18
Corballis, M. C. (2002). *From hand to mouth: The origins of language*. Princeton, NJ: Princeton University Press.
Cornillie, B. (2010). An interactional approach to epistemic and evidential adverbs. In G. S. Diewald, E. (Ed.), *Linguistic Realization of Evidentiality in European Languages* (pp. 309-330). Berlin/ New York: Walter de Gruyter.
Coughlin, C., Hembacher, E., Lyons, K. E., & Ghetti, S. (2015). Introspection on uncertainty and judicious help-seeking during the preschool years. *Dev Sci, 18*(6), 957-971. doi:10.1111/desc.12271
Crespo Sendra, V., Kaland, C., Swerts, M., & Prieto, P. (2013). Perceiving incredulity: The role of intonation and facial gestures. *Journal of Pragmatics, 47*(1), 1-13.
Cristià, A., McGuire, G. L., Seidl, A., & Francis, A. L. (2011). Effects of the distribution of acoustic cues on infants' perception of sibilants. *Journal of Phonetics, 39*(3), 388-402. doi:10.1016/j.wocn.2011.02.004
Cruttenden, A. (1985). Intonation comprehension in ten-year-olds. *Journal of Child Language, 12*(3), 643-661.

Cruttenden, A. (1997). *Intonation, second ed.* . Cambridge: Cambridge University Press.

Culpeper, J. (2011). "It's not what you said, it's how you said it!": prosody and impoliteness. In L. P. R. Group (Ed.), *Discursive approaches to politeness* (Vol. 8, pp. 57-83). Berlin: Mouton de Gruyter.

Culpeper, J., Bousfield, D., & Wichmann, A. (2003). Impoliteness revisited: With special reference to dynamic and prosodic aspects. *Journal of Pragmatics, 35*(10-11), 1545-1579. doi:10.1016/S0378-2166(02)00118-2

Culpeper, J., Haugh, M., & Kádár, D. Z. (2017). *The Palgrave handbook of linguistic (im)politeness*.

Davis, B. L., MacNeilage, P. F., Matyear, C. L., & Powell, J. K. (2000). Prosodic correlates of stress in babbling: an acoustical study. *Child Development, 71*, 1258–1270.

de Haan, F. (2001a). The relation between modality and evidentiality. In R. Müller & M. Reis (Eds.), *Modalität und Modalverben im Deutschen* (Vol. Linguistische Berichte, pp. 201-216). Hamburg: H. Buske.

de Haan, F. (2001b). The relation between modality and evidentiality. In R. Müller & M. Reis (Eds.), *Modalität und Modalverben im Deutschen* (Vol. Linguistische Berichte, Sonderheft 9. , pp. 201–216). Hamburg, Germany: H. Buske.

De Ruiter, J.-P. (1998). *Gesture and speech production*. University of Nijmegen., Unpublished PhD diss.

De Ruiter, L. E. (2009). The prosodic marking of topical referents in the German "vorfeld" by children and adults. *Linguistic Review, 26*(2-3), 329-354. doi:10.1515/tlir.2009.012

De Ruiter, L. E. (2014). How German children use intonation to signal information status in narrative discourse. *Journal of Child Language, 41*(5), 1051-1061. doi:10.1017/S0305000913000251

De Sanctis, S. (1902). *"La Mimica del Pensiero." [The facial expression of thought]*. Palermo: Sandron.

de Saussure, F. (1983). *Course in general linguistics [Transl. R. Harris].* La Salle, IL: Open Court.

de Villiers, J. G., & de Villiers, P. A. (1974). Competence and performance in child language. *Journal of Child Language, 1*, 11-12.

Debras, C. (2015). *Visual stance markers: is shrugging lexical or grammatical?* Paper presented at the 13th International Cognitive Linguistics Conference, Newcastle.

Debras, C., & Cienki, A. (2012). Some uses of head tilts and shoulder shrugs during human interaction, and their relation to stancetaking. *Proceedings of ASE/IEEE International Conference on Privacy,* *Security, Risk and Trust and 2012 ASE/IEEE International Conference on Social Computing, SocialCom/PASSAT 2012*, 932-937.

Decasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science, 208*(4448), 1174-1176.

Demir, O. E., Levine, S. C., & Goldin-Meadow, S. (2015). A tale of two hands: children's early gesture use in narrative production predicts later narrative structure in speech. *Journal of Child Language, 42*(3), 662-681. doi:10.1017/S0305000914000415

Dendale, P., & Tasmowski, L. (2001). Introduction: Evidentiality and related notions. *Journal of Pragmatics, 33*(3), 339-348. doi:10.1016/S0378-2166(00)00005-9

Denham, S. A., McKinley, M., Couchoud, E. A., & Holt, R. (1990). Emotional and behavioral predictors of peer status in young preschoolers. *Child Development, 61*, 1145–1152.

Diessel, H., & Tomasello, M. (2001). The acquisition of finite complement clauses in English: A corpus-based analysis. *Cognitive Linguistics, 12*(2), 97-142. doi:10.1515/cogl.12.2.97

Dijkstra, C., Krahmer, E., & Swerts, M. (2006). Manipulating uncertainty: The contribution of different audiovisual prosodic cues to the perception of confidence. *Proceedings of Speech Prosody*, Dresden.

Dodane, C., Beaupoil, P., Del Ré, A., Boutet, D., & Morgenstern, A. (2014). *The role of gesture and prosody in children’s multimodal pathway into negation* Paper presented at the Sound to Gesture conference (S2G), Padova, Italy.

Dodane, C., & Martel, K. (2009a). Emergence de formats prosodiques stables en français chez des enfants français de 10 à 12 mois. . *In Enfance, numéro spécial La prosodie et l’émergence du langage, 3*.

Dodane, C., & Martel, K. (2009b). Évolution de l’inventaire de contours de Fo chez deux enfants français de 10 à 12 mois : l’importance du contexte pour décrire le stade pré-linguistique. *Enfance, 2009*(03), 305. doi:10.4074/s001375450900305x

Dohen, M., & Loevenbruck, H. (2009). Interaction of audition and vision for the perception of prosodic contrastive focus. *Lang Speech, 52*(Pt 2-3), 177-206. doi:10.1177/0023830909103166

Duncan, S. (1972). Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology, 23*(2), 283-292. doi:10.1037/h0033031

Eibl-Eibesfeldt, I. (1974). Somiglianze interculturali tra movimenti espressivi. In R. A. Hinde (Ed.), *La comunicazione non verbale nell'uomo* (pp. 395-405). Bari: Laterza.

Ejiri, K., & Masataka, N. (2001). Co-occurrence of preverbal vocal behavior and motor action in early infancy. *Developmental Science, 4*(1), 40-48. doi:10.1111/1467-7687.00147

Ekman, P. (1979). About brows: Emotional and conversational signals. *Human ethology*, 163-202.

Ekman, P., & Friesen, W. V. (1969). The Repertoire of Nonverbal Behavior: Categories, Origins, Usage, and Coding. *Semiotica, 1*(1), 49-98.

Ekman, P., & Friesen, W. V. (2003). *Unmasking the face: A guide to recognizing emotions from facial clues*. Los Altos, CA: Malor Books.

Ekman, P., Friesen, W. V., & Hager, J. C. (2002). *Facial Action Coding System. Manual and Investigator's Guide*. Salt Lake City, UT: : Research Nexus.

Ellyson, S. L., & Dovidio, J. F. (1985). Power, dominance, and nonverbal behavior: Basic concepts and issues. In S. L. Ellyson & J. F. Dovidio (Eds.), *Power, Dominance, and Nonverbal Behavior* (pp. 1-27). New York: Springer.

Enfield, N. J., Brown, P., & de Ruiter, J. P. (2012). Epistemic dimensions of polar questions: Sentence-final particles in comparative perspective. In J. P. D. Ruiter (Ed.), *Questions: Formal, functional and interactional perspectives* (pp. 193-221). New York: Cambridge University Press.

Ervin-Tripp, S., & Gordon, D. (1986). The development of children's requests. In R. Schiefelbusch (Ed.), *Communicative Competence: Assessment and intervention* (pp. 61–96). San Diego: College Hill Press.

Escandell-Vidal, V. (1998). Intonation and procedural encoding: the case of Spanish interrogatives. In V. Rouchota & A. Jucker (Eds.), *Current Issues in Relevance Theory* (pp. 163–203). Amsterdam: John Benjamins.

Esteve-Gibert, N., Borràs-Comes, J., Asor, E., Swerts, M., & Prieto, P. (2017a). The timing of head movements: The role of prosodic heads and edges. *Journal of the Acoustical Society of America, 141*(6), 4727-4739. doi:10.1121/1.4986649

Esteve-Gibert, N., & Guellaï, B. (2018). The links between prosody and gestures: A developmental perspective. *Frontiers in Psychology*.

Esteve-Gibert, N., Liszkowski, U., & Prieto, P. (2016). Prosodic and gesture features distinguish the pragmatic meaning of pointing gestures in child-directed communication. In M. E. Armstrong, N. Henriksen, & M. M. Vanrell (Eds.), *Intonational Grammar in Ibero-Romance. Approaches across linguistic subfields* (pp. 249-276). Amsterdam: John Benjamins.

Esteve-Gibert, N., & Prieto, P. (2013). Prosody signals the emergence of intentional communication in the first year of life: Evidence from Catalan-babbling infants. *Journal of Child Language, 40*(5), 919-944.

Esteve-Gibert, N., & Prieto, P. (2014). Infants temporally coordinate gesture-speech combinations before they produce their first words. *Speech Communication, 57*, 301-316. doi:10.1016/j.specom.2013.06.006

Esteve-Gibert, N., & Prieto, P. (2018). Early development of prosody-meaning interface. In N. Esteve-Gibert & P. Prieto (Eds.), *The Development of Prosody in First Language Acquisition* (pp. 228-246). Amsterdam: John Benjamins.

Esteve-Gibert, N., Prieto, P., & Liszkowski, U. (2017b). Twelve-month-olds understand social intentions based on prosody and gesture shape. *Infancy, 22*(1), 108-129. doi:10.1111/infa.12146

Farroni, T., Johnson, M. H., Menon, E., Zulian, L., Faraguna, D., & Csibra, G. (2005). Newborns' preference for face-relevant stimuli: Effects of contrast polarity. *Proceedings of the National Academy of Sciences of the United States of America, 102*(47), 17245-17250. doi:10.1073/pnas.0502205102

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development, 8*(2), 181-195. doi:10.1016/S0163-6383(85)80005-9

Fernald, A. (1993). Approval and disapproval: Infant responses to vocal affect in familiar and unfamiliar languages. *Child Development, 64*, 657–674.

Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development, 10*(3), 279-293. doi:10.1016/0163-6383(87)90017-8

Fivero, M. (1976). A Speech Act Analysis of Polite Verb Forms in Romance. *Proceedings of 6th Linguistic Symposium on Romance Languages*, Ottawa, Canada.

Flavell, J. H., Flavell, E. R., & Green, F. L. (2001). Development of children's understanding of connections between thinking and feeling. *Psychological Science, 12*(5), 430-432. doi:10.1111/1467-9280.00379

Friend, M. (2001). The transition from affective to linguistic meaning. *First Language, 21*(63), 219-243. doi:10.1177/014272370102106302

Friend, M., & Bryant, J. B. (2000). A developmental lexical bias in the interpretation of discrepant messages. *Merrill-Palmer Quarterly, 46*(2), 342-369.

Frota, S., Matos, N., Cruz, M., & Vigário, M. (2016). Early prosodic development: Emerging intonation and phrasing in European Portuguese. In M. Armstrong, M. M. Vanrell, & N. C. Henriksen (Eds.), *Issues in Hispanic and lusophone linguistics: Interdisciplinary approaches to intonational grammar in Ibero-Romance* (pp. 295–324). Philadelphia, PA: Benjamins.

Gagnon, M., Gosselin, P., Hudon-ven der Buhs, I., Larocque, K., & Milliard, K. (2010). Children's recognition and discrimination of fear and disgust facial expressions. *Journal of Nonverbal Behavior, 34*(1), 27-42. doi:10.1007/s10919-009-0076-z

Gagnon, M., Gosselin, P., & Maassarani, R. (2014). Children's ability to recognize emotions from partial and complete facial expressions. *Journal of Genetic Psychology, 175*(5), 416-430. doi:10.1080/00221325.2014.941322

Gao, X., & Maurer, D. (2010). A happy story: Developmental changes in children's sensitivity to facial expressions of varying intensities. *Journal of Experimental Child Psychology, 107*(2), 67-86. doi:10.1016/j.jecp.2010.05.003

Gardner, R. A., & Gardner, B. T. (1969). Teaching sign language to a chimpanzee. *Science, 165*(3894), 664-672.

Georgalidou, M. (2008). The contextual parameters of linguistic choice: Greek children's preferences for the formation of directive speech acts. *Journal of Pragmatics, 40*(1), 72-94. doi:10.1016/j.pragma.2007.08.009

Gerken, L., Balcomb, F. K., & Minton, J. L. (2011). Infants avoid 'labouring in vain' by attending more to learnable than unlearnable linguistic patterns. *Developmental Science, 14*(5), 972-979. doi:10.1111/j.1467-7687.2011.01046.x

Ghetti, S., Hembacher, E., & Coughlin, C. A. (2013). Feeling uncertain and acting on it during the preschool years: A metacognitive approach. *Child Development Perspectives, 7*(3), 160-165. doi:10.1111/cdep.12035

Givens, D. B. (2001). *The Nonverbal Dictionary of Gestures, Signs & Body Language Cues, Nonverbal Communication*. Spokane, Washington: Center for Nonverbal Studies Press.

Gleason, J. B., Perlmann, R. Y., & Greif, E. B. (1984). What's the Magic Word: Learning Language through Politeness Routines. *Discourse Processes, 7*(4), 493-502. doi:10.1080/01638538409544603

Gleason, J. B., & Weintraub, S. (1976). The acquisition of routines in child language. *Language in Society, 5*(2), 129-136. doi:10.1017/S0047404500006977

Gleitman, L. R., Gleitman, H., Landau, B., & Wanner, E. (1988). Where learning begins: Initial representations for language learning. In F. J. Newmeyer (Ed.), *Language: Psychological and biological aspects* (pp. 150-193). Cambridge, England: Cambridge University Pres.

Gleitman, L. R., & Wanner, E. (1982). Language acquisition: the state of the art. In W. E. & G. L. R. (Eds.), *Language acquisition: the state of the art*. New York: C.U.P.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin and Review, 9*(3), 558-565. doi:10.3758/BF03196313

Goffman, E. (1967). *Interaction ritual: Essays on face-to-face behavior*. New York: Anchor Books.

Goldin-Meadow, S. (2007). The challenge: Some properties of language can be learned without linguistic input. *The Linguistic Review, 24*(4). doi:10.1515/tlr.2007.016

Goldin-Meadow, S. (2014). Widening the lens: what the manual modality reveals about language, learning and cognition. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 369*(1651), 20130295. doi:10.1098/rstb.2013.0295

Goldin-Meadow, S., & Butcher, C. (2003). Pointing toward two-word speech in young children. In S. Kita (Ed.), *Pointing: Where Language, Culture, and Cognition Meet* (pp. 85-107). Erlbaum: Mahwah, NJ.

Goldin-Meadow, S., & Iverson, J. M. (1998). The Nature and Functions of Gestures in Children's Communication. *New Directions for Child Development, 79*, 1-7.

Goldin-Meadow, S., & Wagner, S. M. (2005). How our hands help us learn. *Trends Cogn Sci, 9*(5), 234-241. doi:10.1016/j.tics.2005.03.006

González-Fuente, S. (2017). *Audiovisual Prosody and Verbal Irony* (Doctoral dissertation). Universitat Pompeu Fabra.

Goodwin, M. H. (1980). Processes of mutual monitoring implicated in the production of description sequences. *Sociological inquiry, 50*((3-4)), 303-317.

Goodwin, M. H., & Goodwin, C. (2001). Emotion within situated activity. In A. Duranti (Ed.), *Linguistic Anthropology: A Reader* (pp. 239-257). Malden: Blackwell.

Goodwin, M. H., Goodwin, C., & Yaeger-Dror, M. (2002). Multimodality in girls' game disputes. *Journal of Pragmatics, 34*(10/11), 1621-1649.

Grawunder, S., Oertel, M., & Schwarze, C. (2014). Politeness, culture, and speaking task - Paralinguistic prosodic behavior of speakers from Austria and Germany. *Proceedings of Proceedings of the international conference on Speech Prosody*, Dublin, 159-163.

Greif, E. B., & Gleason, J. B. (1980). Hi, thanks, and goodbye: More routine information. *Language in Society, 9*(2), 159-166. doi:10.1017/S0047404500008034

Grossmann, T., Oberecker, R., Koch, S. P., & Friederici, A. D. (2010). The Developmental Origins of Voice Processing in the Human Brain. *Neuron, 65*(6), 852-858. doi:10.1016/j.neuron.2010.03.001

Grossmann, T., Striano, T., & Friederic, A. D. (2006). Crossmodal integration of emotional information from face and voice in the infant brain. *Developmental Science, 9*(3), 309-315. doi:10.1111/j.1467-7687.2006.00494.x

Grünloh, T., Lieven, E., & Tomasello, M. (2015). Young Children's Intonational Marking of New, Given and Contrastive Referents. *Language Learning and Development, 11*(2), 95-127. doi:10.1080/15475441.2014.889530

Guellaï, B., Langus, A., & Nespor, M. (2014). Prosody is in the hands of the speaker. *Proceedings of Proceedings of the international conference on Speech Prosody*, 544-547.

Guellaï, B., Streri, A., Chopin, A., Rider, D., & Kitamura, C. (2016). Newborns' sensitivity to the visual aspects of infant-directed speech: Evidence from point-line displays of talking faces. *Journal of Experimental Psychology: Human Perception and Performance, 42*(9), 1275-1281. doi:10.1037/xhp0000208

Guidetti, M. (2002). The emergence of pragmatics: Forms and functions of conventional gestures in young French children. *First Language, 22*(3), 265-285. doi:10.1177/014272370202206603

Guidetti, M. (2005). Yes or no? How young French children combine gestures and speech to agree and refuse. *Journal of Child Language, 32*, 911-924. doi:10.1017/S0305000905007038

Gunlogson, C. (2003). *True to Form: Rising and Falling Declaratives as Questions in English*. New York: Routledge.
Gussenhoven, C. (2002). Intonation and interpretation: Phonetics and phonology. *Proceedings of Speech Prosody,* Aix-en-Provence, 47-57.
Hadar, U., Steiner, T. J., Grant, E. C., & Rose, F. C. (1983). Kinematics of head movements accompanying speech during conversation. *Human Movement Science, 2*(1-2), 35-46. doi:10.1016/0167-9457(83)90004-0
Hall, J. A., Coats, E. J., & LeBeau, L. S. (2005). Nonverbal behavior and the vertical dimension of social relations: A meta-analysis. *Psychological Bulletin, 131*(6), 898-924. doi:10.1037/0033-2909.131.6.898
Harrigan, J. A. (1984). The effects of task order on children's identification of facial expressions. *Motivation and Emotion, 8*(2), 157-169. doi:10.1007/BF00993071
Harris, P. L., Bartz, D. T., & Rowe, M. L. (2017). Young children communicate their ignorance
and ask questions. *Proceedings of the National Academy of Sciences of the United States of America, 114*(38), E8128. doi:10.1073/pnas.1715210114
Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology, 56*(4), 208-216.
Hendriks, P., & Koster, C. (2010). Production/comprehension asymmetries in language acquisition. *Lingua, 120*(8), 1887-1897. doi:10.1016/j.lingua.2010.02.002
Herba, C. M., Landau, S., Russell, T., Ecker, C., & Phillips, M. L. (2006). The development of emotion-processing in children: Effects of age, emotion, and intensity. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 47*(11), 1098-1106. doi:10.1111/j.1469-7610.2006.01652.x
Heritage, J. (2012a). The Epistemic Engine: Sequence Organization and Territories of Knowledge. *Research on Language and Social Interaction, 45*(1), 30-52. doi:10.1080/08351813.2012.646685
Heritage, J. (2012b). Epistemics in Action: Action Formation and Territories of Knowledge. *Research on Language and Social Interaction, 45*(1), 1-29. doi:10.1080/08351813.2012.646684
Heritage, J. (2013). Action formation and its epistemic (and other) backgrounds. *Discourse Studies, 15*(5), 551-578. doi:10.1177/1461445613501449

Hewes, G. (1999). A history of the study of language origins and the gestural primacy hypothesis. In A. Lock & C. R. Peters (Eds.), *Handbook of human symbolic evolution* (pp. 571-595). Oxford, UK: Oxford University Press, Clarendon Press.

Heylen, D. (2005). Challenges ahead head movements and other social acts in conversations. *Proceedings of AISB'05 Convention:Proceedings of the Joint Symposium on Virtual Social Agents: Social Presence Cues for Virtual Humanoids Empathic Interaction with Synthetic Characters Mind Minding Agents*, 45-52.

Hirschberg, J. (2017). Pragmatics and Prosody. In Y. Huang (Ed.), *The Oxford Handbook of Pragmatics* (pp. 532-549). Oxford: Oxford University Press.

Hirsh-Pasek, K., Tucker, M., & Golinkoff, R. M. (1996). Dynamic systems theory: Reinterpreting "prosodic bootstrapping" and its role in language acquisition. *Signal to syntax Bootstrapping from speech to grammar in early acquisition*, 449-466 ST - Dynamic systems theory: Reinterpreti.

Hockett, C. F. (1960). The origin of speech. *Scientific American*.

Hoehle, B. (2009). Bootstrapping mechanisms in first language acquisition. *Linguistics, 47*(2), 359-382. doi:10.1515/LING.2009.013

Höhle, B., Bijeljac-Babic, R., Herold, B., Weissborn, J., & Nazzi, T. (2009). Language specific prosodic preferences during the first half year of life- Evidence from German and French infants. *Infant Behaviour and Development, 32*(3), 262-274. doi:10.1016/j.infbeh.2009.03.004

Hoicka, E., & Wang, S. H. (2011). Fifteen-month-old infants match vocal cues to intentional actions. *Journal of Cognition and Development, 12*(3), 299-314. doi:10.1080/15248372.2010.542215

Holler, J., & Wilkin, K. (2009). Communicating common ground: How mutually shared knowledge influences speech and gesture in a narrative task. *Language and Cognitive Processes, 24*(2), 267-289. doi:10.1080/01690960802095545

Hollos, M. (1977). Comprehension and use of social rules in pronoun selection by Hungarian children. In S. Ervin-Tripp & C. Mitchell-Kernan (Eds.), *Child discourse* (pp. 211–223). York, NY: Academic Press.

Hollos, M., & Beeman, W. (1978). The development of directives among Norwegian and Hungarian children: An example of communicative style in culture. *Language in Society, 7*(3), 345-355. doi:10.1017/S0047404500005765

Hostetter, A. B. (2011). When do gestures communicate? A meta-analysis. *Psychological Bulletin, 137*, 297–315.

Hübscher, I., Borràs-Comes, J., & Prieto, P. (2017a). Prosodic mitigation characterizes Catalan formal speech: The Frequency Code reassessed. *Journal of Phonetics, 65*, 145-159. doi:10.1016/j.wocn.2017.07.001

Hübscher, I., Esteve-Gibert, N., Igualada, A., & Prieto, P. (2017b). Intonation and gesture as bootstrapping devices in speaker uncertainty. *First Language, 37*(1), 24-41.

Hübscher, I., Sánchez-Conde, C., Vincze, L., Borràs-Comes, J., & Prieto, P. (forthcoming). Gestural mitigation strategies characterize formal register in Catalan requests.

Hübscher, I., Wagner, L., & Prieto, P. (2016). Young children's sensitivity to polite stance expressed through audiovisual prosody in requests. *Proceedings of Speech Prosody*, Boston, 897-901.

Hübscher, I., Wagner, L., & Prieto, P. (accepted). Three-year-olds' infer polite stance from intonation and facial expression. *Journal of Politeness Research Language Behaviour and Culture*.

Igualada, A. (2017). *Gesture-speech temporal integration in language development* (Doctoral dissertation). Universitat Pompeu Fabra.

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philos Trans R Soc Lond B Biol Sci, 369*(1651), 20130298. doi:10.1098/rstb.2013.0298

Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition, 109*(1), 54-65. doi:10.1016/j.cognition.2008.07.015

Ishi, C. T., Ishiguro, H., & Hagita, N. (2014). Analysis of relationship between head motion events and speech in dialogue conversations. *Speech Communication, 57*, 233-243. doi:10.1016/j.specom.2013.06.008

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.

Ito, M. (2004). Politeness and voice quality – the alternative method to measure aspiration noise. *Proceedings of 2nd International Conference on Speech Prosody*, Nara, Japan.

Iverson, J. M., Capirci, O., & Caselli, M. C. (1994). From communication to language in two modalities. *Cognitive Development, 9*(1), 23-43. doi:10.1016/0885-2014(94)90018-3

Iverson, J. M., & Fagan, M. K. (2004). Infant vocal-motor coordination: precursor to the gesture-speech system? . *Child Development, 75*(4), 1053-1066.

Iverson, J. M., & Goldin-Meadow, S. (1998). Why people gesture when they speak (Vol. 396, pp. 228-228).

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science, 16*(5), 367-371. doi:10.1111/j.0956-7976.2005.01542.x

Izard, C. E. (1971). *The face of emotion*. New York: Appleton–Century–Crofts.

James, S. L. (1978). Effect of listener age and situation on the politeness of children's directives. *Journal of Psycholinguistic Research, 7*(4), 307-317. doi:10.1007/BF01068112

Jarmołowicz-Nowikow, E. (2014). Pointing by hand: Types of reference and their influence on gestural form. In C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, & J. Bressem (Eds.), *Body - Language - Communication: An international Handbook on Multimodality in Human interaction (Handbooks of Linguistics and Communication Science 38.2)* (pp. 1824-1833). Berlin/Boston: De Gruyter Mouton.

Jhang, Y., & Oller, D. K. (2017). Emergence of functional flexibility in infant vocalizations of the first 3 months. *Frontiers in Psychology, 8*(MAR). doi:10.3389/fpsyg.2017.00300

Jokinen, K., & Allwood, J. (2010) Hesitation in intercultural communication: Some observations and analyses on interpreting shoulder shrugging. *Vol. 6259 LNCS. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (pp. 55-70).

Jun, S. A. (2010). *Prosodic typology: The phonology of intonation and phrasing*. Oxford: Oxford University Press.

Jun, S. A. (2014). *Prosodic Typology II: the Phonology of Intonation and Phrasing*. Oxford/ New York: Oxford University Press.

Jusczyk, P. W., Friederic, A. D., Wessels, J. M. I., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants' Sensitivity to the Sound Patterns of Native Language Words. *Journal of Memory and Language, 32*, 402-420.

Kahana-Kalman, R., & Walker-Andrews, A. S. (2001). The role of person familiarity in young infants' perception of emotional expressions. *Child Development, 72*(2), 352-369. doi:10.1111/1467-8624.00283

Kaori, I., Winter, B., & Brown, L. (2016). *The polite voice in Japanese: Acoustic correlates of polite and casual speech styles with*

comparison to Korean. Paper presented at the Japanese/Korean Linguistics Conference, Tokyo, Japan.
Kärkkäinen, E. (2003). *Epistemic Stance in English Conversations. A Description of its Interactional Functions, with a Focus on I Think*. Amsterdam: Benjamins.
Kayyal, M. H., & Widen, S. C. (2015). What made sahar scared?: Imaginary and realistic causes in palestinian and American children's concept for fear. *Journal of Cognition and Culture, 15*(1-2), 32-44. doi:10.1163/15685373-12342139
Keating, C. F., & Bai, D. L. (1986). Children's attributions of social dominance from facial cues. *Child Development, 57*, 1269-1276.
Kelly, S. D. (2001). Broadening the units of analysis in communication: Speech and nonverbal behaviours in pragmatic comprehension. *Journal of Child Language, 28*(2), 325-349.
Kelly, S. D., Barr, D. J., Breckinridge Church, R., & Lynch, K. (1999). Offering a Hand to Pragmatic Understanding: The Role of Speech and Gesture in Comprehension and Memory. *Journal of Memory and Language, 40*, 577–592.
Kelly, S. D., Barr, D. J., Church, R. B., & Lynch, K. (1999). Offering a Hand to Pragmatic Understanding: The Role of Speech and Gesture in Comprehension and Memory. *Journal of Memory and Language, 40*(4), 577-592. doi:10.1006/jmla.1999.2634
Kendon, A. (1972). Chapter 9 - Some Relationships Between Body Motion and Speech: An Analysis of an Example1 A2 - Siegman, Aron Wolfe. In B. Pope (Ed.), *Studies in Dyadic Communication* (Vol. 7, pp. 177-210). Amsterdam: Pergamon.
Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In M. Ritchie Key (Ed.), *The Relationship of Verbal and Nonverbal Communication* (pp. 207-227). The Hague: Mouton.
Kendon, A. (1995). Gestures as illocutionary and discourse structure markers in Southern Italian conversation. *Journal of Pragmatics, 23*(3), 247-279. doi:10.1016/0378-2166(94)00037-F
Kendon, A. (2002). Some uses of the head shake. *Gesture, 2*(2), 147-182. doi:10.1075/gest.2.2.03ken
Kendon, A. (2004). *Gesture: visible action as utterance*. Cambridge: Cambridge University Press.
Khu, M., Chambers, C., Graham, S. A., & (submitted). When you're happy and I know it: Four year-olds' emotional perspective-taking during on-line language comprehension. *Manuscript revised and resubmitted for publication*.

Kidd, E., & Holler, J. (2009). Children's use of gesture to resolve lexical ambiguity. *Dev Sci, 12*(6), 903-913. doi:10.1111/j.1467-7687.2009.00830.x

Kim, S., Paulus, M., Sodian, B., & Proust, J. (2016). Young children's sensitivity to their own ignorance in informing others. *PLoS ONE, 11*(3). doi:10.1371/journal.pone.0152595

Kita, S. (2003). *Pointing: Where language, culture, and cognition meet*.

Kita, S. (2009). Cross-cultural variation of speech-accompanying gesture: A review. *Language and Cognitive Processes, 24*(2), 145-167. doi:10.1080/01690960802586188

Kitamura, C., Guellaï, B., & Kim, J. (2014). Motherese by eye and ear: infants perceive visual prosody in point-line displays of talking heads. *PLoS ONE, 9*(10), e111467. doi:10.1371/journal.pone.0111467

Koenig, M. A., Clément, F., & Harris, P. L. (2004). Trust in testimony: Children's use of true and false statements. *Psychological Science, 15*(10), 694-698.

Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development, 76*(6), 1261-1277. doi:10.1111/j.1467-8624.2005.00849.x

Kornell, N., Son, L. K., & Terrace, H. S. (2007). Transfer of metacognitive skills and hint seeking in monkeys: Research article. *Psychological Science, 18*(1), 64-71. doi:10.1111/j.1467-9280.2007.01850.x

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: susceptibility to others' beliefs in human infants and adults. *Science (New York, N.Y.), 330*(6012), 1830-1834.

Krahmer, E., & Swerts, M. (2005). How children and adults produce and perceive uncertainty in audiovisual speech. *Language and Speech, 48*(1), 29-53.

Krahmer, E., & Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language, 57*(3), 396-414. doi:10.1016/j.jml.2007.06.005

Krahmer, E., & Swerts, M. (2009). Audiovisual Prosody--Introduction to the Special Issue. *Language and Speech, 52*(2-3), 129-133. doi:10.1177/0023830909103164

Krauss, R. M. (1998). Why do we gesture when we speak? . *Current Directions in Psychological Science, 7*, 54-59.

Krivokapic, J., Tiede, M. K., Tyrone, M. E., & Goldenberg, D. (2016). Speech and manual gesture coordination in a pointing task.

Proceedings of Proceedings of the international conference on Speech Prosody, 1240-1244.

Kuhl, P. K., Williams, K. A., & Meltzoff, A. N. (1991). Cross-Modal Speech Perception in Adults and Infants Using Nonspeech Auditory Stimuli. *Journal of Experimental Psychology: Human Perception and Performance, 17*(3), 829-840. doi:10.1037/0096-1523.17.3.829

Kurumada, C., & Clark, E. V. (2017). Pragmatic inferences in context: learning to interpret contrastive prosody. *J Child Lang, 44*(4), 850-880. doi:10.1017/S0305000916000246

Ladefoged, P. (2006). *A Course in Phonetics*. Boston, Massachusetts: Thomson Wadsworth.

Langlotz, A., & Locher, M. A. (2013). The role of emotions in relational work. *Journal of Pragmatics, 58*, 87-107. doi:10.1016/j.pragma.2013.05.014

Langlotz, A., & Locher, M. A. (2017). (Im)politeness and emotion *Palgrave Handbook of Linguistic (Im)Politenes* (pp. 287-322). London.

Lausberg, H., & Sloetjes, H. (2009). Coding gestural behavior with the NEUROGES-ELAN system. *Behavior Research Methods, 41*(3), 841-849. doi:10.3758/BRM.41.3.841

Laver, J. (1980). *The Phonetic Description of Voice Quality*. Cambridge: Cambridge University Press.

Lawrence, L. L., & Fernald, A. (1993). *When prosody and semantics conflict: infants' sensitivity to discrepancies between tone of voice and verbal content.* . Poster session presented at the biennial meeting of the Society for Research in Child Development, . New Orleans, LA.

Lee, J., & Marsella, S. (2006) Nonverbal behavior generator for embodied conversational agents. *Vol. 4133 LNAI. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (pp. 243-255).

Lee, T. H., & Law, A. (2001). Epistemic modality and the acquisition of Cantonese final particles. In M. Nakayama (Ed.), *Issues in East Asian Language Acquisition* (pp. 67-128). Tokyo: Kurosio Publishers.

Leonard, T., & Cummins, F. (2011). The temporal relation between beat gestures and speech. *Language and Cognitive Processes, 26*(10), 1457-1471. doi:10.1080/01690965.2010.500218

Levinson, S. C., & Holler, J. (2014). The origin of human multi-modal communication. *Philos Trans R Soc Lond B Biol Sci, 369*(1651), 20130302. doi:10.1098/rstb.2013.0302

Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology, 18*(2), 201-212. doi:10.1016/0022-0965(74)90101-5

Lin, H., Kwock-Ping, J. T., & Fon, J. (2006). An acoustic study on the paralinguistic prosody in the politeness talk in Taiwan Mandarin. *Proceedings of ISCA Tutorial and Research Workshop on Experimental Linguistics*, Athens, Greece, 173-176.

Lindová, J., Špinka, M., & Nováková, L. (2015). Decoding of baby calls: Can adult humans identify the eliciting situation from emotional vocalizations of preverbal infants? *PLoS ONE, 10*(4). doi:10.1371/journal.pone.0124317

Liszkowski, U., Carpenter, M., Striano, T., & Tomasello, M. (2006). 12- and 18-month-olds point to provide information for others. *Journal of Cognition and Development, 7*(2), 173-187. doi:10.1207/s15327647jcd0702_2

Liszkowski, U., Carpenter, M., & Tomasello, M. (2008). Twelve-month-olds communicate helpfully and appropriately for knowledgeable and ignorant partners. *108*, 732-739. doi:10.1016/j.cognition.2008.06.013

Llanes-Coromina, J., Vilà-Giménez, I., Borràs-Comes, J., Kushch, O., & P., P. (2018, in press). Beat gestures help prechoolers recall and comprehend discourse information. *Journal of Experimental Child Psychology*.

Locher, M. A., & Koenig, R. (2014). 'All I could do was hand her another tissue' - Handling emotions as challange in reflective writing texts by medical students. In A. Langlotz & A. Soltysk Monnet (Eds.), *Emotion, affect, sentiment: The language and aesthetics of feeling (Swiss paper in English Language and Literature 30* (pp. 121-141). Tübingen: Narr.

Locher, M. A., & Langlotz, A. (2008). Relational work: at the intersection of cognition, interaction and emotion. *Swiss Association of Applied Linguistics, 88*, 165-191.

Loehr, D. (2007). Aspects of rhythm in gesture and speech. *Gesture, 7*(2), 179-214.

Lyons, K. E., & Ghetti, S. (2011). The Development of Uncertainty Monitoring in Early Childhood. *Child Development, 82*(6), 1778-1787. doi:10.1111/j.1467-8624.2011.01649.x

Lyons, K. E., & Ghetti, S. (2013). I Don't Want to Pick! Introspection on Uncertainty Supports Early Strategic Behavior. *Child Development, 84*(2), 726-736. doi:10.1111/cdev.12004

Ma, W., Golinkoff, R. M., Houston, D. M., & Hirsh-Pasek, K. (2011). Word learning in infant-and adult-directed speech. *Language Learning and Development, 7*(3), 185-201. doi:10.1080/15475441.2011.579839

Maatman, M., Gratch, J., & Marsella, S. (2005). Natural behavior of a listening agent. *Proceedings of 5th International Working Conference on Intelligent Virtual Agents*, Kos, Greece, 25-36.

Macedonia, M., Müller, K., & Friederici, A. D. (2011). The impact of iconic gestures on foreign language word learning and its neural substrate. *Human Brain Mapping, 32*(6), 982-998. doi:10.1002/hbm.21084

Mampe, B., Friederici, A. D., Christophe, A., & Wermke, K. (2009). Newborns' cry melody is shaped by their native language. *Current Biology, 19*(23), 1994-1997.

Marcos, H., & Bernicot, J. (1994). Addressee co-operation and request reformulation in young children. *Journal of Child Language, 21*(3), 677-692. doi:10.1017/S0305000900009508

Marín-Arrese, J. I. (2011). Epistemic legitimizing strategies, commitment and accountability in discourse. *Discourse Studies, 13*(6), 789-797. doi:10.1177/1461445611421360c

Markham, R., & Adams, K. (1992). The effect of type of task on children's identification of facial expressions. *Journal of Nonverbal Behavior, 16*(1), 21-39. doi:10.1007/BF00986877

Marsi, E., & Van Rooden, F. (2007). Expressing uncertainty with a talking head in a multimodal question-answering system. *Proceedings of Workshop on Multimodal Output Generation*, At Enschede, The Netherlands, 105.

Martel, K., & Dodane, C. (2012). Le rôle de la prosodie dans les premières constructions grammaticales: étude de cas d'un enfant français monolingue. *Journal of French Language Studies, 22*(01), 13-35. doi:10.1017/s0959269511000561

Mastropieri, D., & Turkewitz, G. (1999). Prenatal experience and neonatal responsiveness to vocal expressions of emotion. *Developmental Psychobiology, 35*(3), 204-214. doi:10.1002/(SICI)1098-2302(199911)35:3<204::AID-DEV5>3.0.CO;2-V

Masumoto, K., Yamaguchi, M., Sutani, K., Tsuneto, S., Fujita, A., & Tonoike, M. (2006). Reactivation of physical motor information in the memory of action events. *Brain research, 1101*(1), 102-109. doi:10.1016/j.brainres.2006.05.033

Mathew, M., Yuen, I., Shattuck-Hufnagel, S., Ren, A., & Demuth, K. (2014). The development of prosodic gestures during parent-child discourse interactions. *Proceedings of LabPhon Conference*, Munich, Germany.

Matoesian, G. (2005). Struck by speech revisited: Embodied stance in jurisdictional discourse. *Journal of Sociolinguistics, 9*(2), 167-193. doi:10.1111/j.1360-6441.2005.00289.x

Matsui, T. (2014). Children's understanding of linguistic expressions of certainty and evidentuality. In D. Matthews (Ed.), *Pragmatic Development in First Language Acquisition* (pp. 295–316). Amsterdam: John Benjamins.

Matsui, T., Rakoczy, H., Miura, Y., & Tomasello, M. (2009). Understanding of speaker certainty and false-belief reasoning : a comparison of Japanese and German preschoolers. *4*, 602-613. doi:10.1111/j.1467-7687.2008.00812.x

Matsui, T., Yamamoto, T., & McCagg, P. (2006). On the role of language in children's early understanding of others as epistemic beings. *Cognitive Development, 21*, 158-173. doi:10.1016/j.cogdev.2005.10.001

Matthews, D. (2014). *Pragmatic development in first language acquisition*. Amsterdam/Philadelphia John Benjamins.

Mayberry, R. I., & Nicoladis, E. (2000). Gesture reflects language development: Evidence from bilingual children. *Current Directions in Psychological Science, 9*(6), 192-196. doi:10.1111/1467-8721.00092

McClave, E. Z. (2001). The relationship between spontaneous gestures of the hearing and American Sign Language. *Gesture 1.1*, 51-72.

McClure, E. B. (2000). A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. *Psychological Bulletin, 126*(3), 424-453.

McCluskey, K. W., & Albas, D. C. (1981). Perception of the emotional content of speech by Canadian and Mexican children, adolescents, and adults. *International Journal of Psychology, 16*, 119−132.

McNeil, N. M., Alibali, M. W., & Evans, J. L. (2000). The role of gesture in children's comprehension of spoken language: now they need it, now they don't. *Journal of Nonverbal Behavior, 24*(2), 131-150.

McNeill, D. (1992). *Hand and mind – What gestures reveal about thought*. Chicago, IL: University of Chicago Press.

McNeill, D. (1998). Speech and gesture integration. In J. M. Iverson & S. Goldin-Meadow (Eds.), (pp. 11-27): Jossey Bass.

McNeill, D. (2000). *Language and Gesture: Window into Thought and Action*. Cambridge: Cambridge University Press.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition, 29*(2), 143-178. doi:10.1016/0010-0277(88)90035-2

Mehrabian, A. (1981). *Silent messages: Implicit communication of emotions and attitudes (2nd ed.)*. Belmont, CA: Wadsworth.

Milosky, L. M., & Ford, J. A. (1997). The role of prosody in children's inferences of ironic intent. *Discourse Processes, 23*(1), 47-61. doi:10.1080/01638539709544981

Mondada, L. (2006). Participants' online analysis and multimodal pratices: Projecting the end of the turn and the closing of the sequence. *Discourse Studies, 8*(1), 117–129.

Mondada, L. (2013). Displaying, contesting and negotiating epistemic authority in social interaction: Descriptions and questions in guided visits. *Discourse Studies 15, 5*, 597–626.

Mondada, L. (2016). Challenges of multimodality: Language and the body in social interaction. *Journal of Sociolinguistics, 20*, 336-366.

Moore, C., Bryant, D., & Furrow, D. (1989). Mental terms and the development of certainty. *Child Development, 60*(1), 167-171.

Moore, C., Harris, L., & Patriquin, M. (1993). Lexical and prosodic cues in the comprehension of relative certainty. *Journal of Child Language, 20*(1), 153-167.

Moore, C., Pure, K., & Furrow, D. (1990). Children's understanding of the modal expression of speaker certainty and uncertainty and its relation to the development of a representational Theory of Mind. *Child Development, 61*(3), 722-730.

Morgan, J. L., & Newport, E. L. (1981). The role of constituent structure in the induction of an artificial language. *Journal of Verbal Learning and Verbal Behavior, 20*(1), 67-85. doi:10.1016/S0022-5371(81)90312-1

Morgenstern, A. (2014a). The blossoming of children's multimodal skills from 1 to 4 years old. In C. Muller, A. Cienki, & E. Frickee (Eds.), *Body – Language – Communication. An international Handbook on Multimodality in Human Interaction.* (pp. 1848-1857). Berlin: De Gruyter Mouton.

Morgenstern, A. (2014b). Children's multimodal language development. In C. Fäcke (Ed.), *Manual of language acquisition* (pp. 123-142). Berlin/ Boston: De Gruyter.
Morgenstern, A., & Parisse, C. (2012). The Paris Corpus. *Journal of French Language Studies, 22*(1). doi:DOI10.1017/S095926951100055X
Morris, D. (2002). *People watching*. London: Vintage.
Morton, J. B., & Trehub, S. E. (2001). Children's understanding of emotion in speech. *Child Development, 72*(3), 834-843.
Morton, J. B., Trehub, S. E., & Zelazo, P. D. (2003). Sources of Inflexibility in 6-Year-Olds' Understanding of Emotion in Speech. *Child Development, 74*(6), 1857-1868. doi:10.1046/j.1467-8624.2003.00642.x
Müller, C. (2004). The palm-up-open-hand. A case of a gesture family? In C. Müller & R. Posner (Eds.), *The semantics and pragmatics of everyday gestures* (pp. 233-256). Berlin: Weidler Verlag.
Müller, C. (2004). Forms and uses of the Palm Up Open Hand: A case of a gesture family. In C. Müller & R. R. Posner (Eds.), *The semantics and pragmatics of everyday gesture, proceedings of the Berlin conference April 1998* (pp. 233–256). Berlin: Weidler Buchverlag.
Mumme, D. L., Fernald, A., & Herrera, C. (1996). Infants' Responses to Facial and Vocal Emotional Signals in a Social Referencing Paradigm. *Child Development, 67*(6), 3219-3237.
Murillo, E., & Belinchón, M. (2013). Multimodal communicative patterns on the transition to first words: Changes in the coordination of gesture and vocalization. *Infancia y Aprendizaje, 36*(4), 473-487. doi:10.1174/021037013808200258
Nadeu, M., & Prieto, P. (2011). Pitch range, gestural information, and perceived politeness in Catalan. *Journal of Pragmatics, 43*(3), 841-854. doi:10.1016/j.pragma.2010.09.015
Nakamura, K. (1999). The use of polite language by Japanese childrne's narratives. *Japanese/Korean Linguistics, 3*, 84-99.
Nakamura, K. (2002). Polite language usage in mother-infant interactions: A look at language socialization. In Y. Shirai, S. Kobayashi, S. Miyata, & K. Nakamura (Eds.), *Studies in Language Sciences II* (pp. 175-191). Tokyo: Kurosio.
Nakamura, K. (2006a). The acquisition of linguistic politeness in Japanese. In M. Nakayama, R. Mazuka, & Y. Shirai (Eds.), *Handbook of Japanese Psycholinguistics* (pp. 110-115). Cambridge: CUP.
Nakamura, K. (2006b). The acquisition of linguistic politeness in Japanese. In M. N., R. Mazuka, & Y. Shirai (Eds.), *The handbook of East*

Asian psycholinguistics: volume II Japanese (pp. 110–115). Cambridge: Cambridge University Press.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance, 24*(3), 756-766.

Nazzi, T., Jusczyk, A. M., & John, E. K. (2000). Language Discrimination of English Learning 5-month-olds: Effects of Rhythm and Familiarity. *Journal of Memory and Language, 43*, 1-19. doi:10.1006/jmla.2000.2698

Neldner, K., Collier-Baker, E., & Nielsen, M. (2015). Chimpanzees (Pan troglodytes) and human children (Homo sapiens) know when they are ignorant about the location of food. *Animal Cognition, 18*(3), 683-699. doi:10.1007/s10071-015-0836-6

Nelson, N. L., & Russell, J. A. (2011). Preschoolers' use of dynamic facial, bodily, and vocal cues to emotion. *Journal of Experimental Child Psychology, 110*(1), 52-61.

Nelson, N. L., & Russell, J. A. (2012). Children's understanding of nonverbal expressions of pride. *Journal of Experimental Child Psychology, 111*(3), 379-385. doi:10.1016/j.jecp.2011.09.004

Nelson, N. L., & Russell, J. A. (2015). Children distinguish between positive pride and hubris. *Developmental Psychology, 51*(11), 1609-1614. doi:10.1037/dev0000048

Nicoladis, E., Mayberry, R. I., & Genesee, F. (1999). Gesture and early bilingual development. *Developmental Psychology, 35*(2), 514-526. doi:10.1037/0012-1649.35.2.514

Niedenthal, P. M., Barsalou, L. W., Winkielman, P., Krauth-Gruber, S., & Ric, F. (2005). Embodiment in attitudes, social perception, and emotion. *Personality and Social Psychology Review, 9*(3), 184-211. doi:10.1207/s15327957pspr0903_1

Nilsen, E. S., Graham, S. A., Smith, S., & Chambers, C. G. (2008). Preschoolers' sensitivity to referential ambiguity: Evidence for a dissociation between implicit understanding and explicit behavior. *Developmental Science, 11*(4), 556-562. doi:10.1111/j.1467-7687.2008.00701.x

Nippold, M. A., Leonard, L. B., & Anastopoulos, A. (1982). Development in the use and understanding of polite forms in children. *Journal of speech and hearing research, 25*(2), 193-202.

Noveck, I. A., Ho, S., & Sera, M. (1996). Children's understanding of epistemic modals. *Journal of Child Language, 23*(3), 621-643.

Nowichi, S., & Duke, M. P. (1994). Individual differences in the nonverbal communication of affect: The diagnostic analysis of nonverbal accuracy scale. *Journal of Nonverbal Behavior, 18*, 9−35.

O'Neill, D. K., & Atance, C. M. (2000). Maybe my dadd give me a big piano: the development of children's use of modals to express uncertainty. *First Language, 20*, 29-52.

O'Neill, M., Bard, K., Linnell, M., & Fluck, M. (2005). Maternal gestures with 20-month-old infants in two contexts. *Developmental Science, 8*(4), 352-359. doi:10.1111/j.1467-7687.2005.00423.x

Ofuka, E., McKeown, J. D., Waterman, M. G., & Roach, P. J. (2000). Prosodic cues for rated politeness in Japanese speech. *Speech Communication, 32*(3), 199-217. doi:10.1016/S0167-6393(00)00009-1

Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica, 41*, 1-16. doi:10.1159/000261706

Ohara, Y. (2001). Finding one's voice in Japanese: A study of the pitch levels of L2 users. In A. Pavlenko, A. Brackledge, I. Pillar, & M. Teiutsch-Dwye (Eds.), *Multilingualism, second language learning, and gender*. New York: Mouton de Gruyter.

Ola Orie, O. (2009). Pointing the Yoruba way. . *Gesture, 9*(2), 237-261.

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. *Proceedings of the National Academy of Sciences of the United States of America, 110*(16), 6318-6323. doi:10.1073/pnas.1300337110

Onishi, K. H. (2005). Do 15-Month-Old Infants Understand False Beliefs ? *, 308*(April), 255-258.

Ornaghi, V., Brockmeier, J., & Gavazzi, I. G. (2011). The role of language games in children's understanding of mental states: A training study. *Journal of Cognition and Development, 12*(2), 239-259. doi:10.1080/15248372.2011.563487

Orozco, L. (2008). Peticiones corteses y factores prosódicos. In Z. E. Z. E. Herrera & P. M. Butragueño (Eds.), *Fonología instrumental. Patrones fónicos y variación*. México D. F.: El Colegio de México.

Orozco, L. (2010). *Estudio sociolingüístico de la cortesía en tratamientos y peticiones. Datos de Guadalajara.* México D. F.: El Colegio de México.

Özçalışkan, S., Gentner, D., & Goldin-Meadow, S. (2013). Do iconic gestures pave the way for children’s early verbs? *Applied Psycholinguistics*, 1-20.

Özçalişkan, Ş., & Goldin-Meadow, S. (2005). Gesture is at the cutting edge of early language development. *Cognition, 96*(3), B101-B113. doi:10.1016/j.cognition.2005.01.001

Özçalişkan, Ş., & Goldin-Meadow, S. (2009). When gesture-speech combinations do and do not index linguistic change. *Language and Cognitive Processes, 24*(2), 190. doi:10.1080/01690960801956911

Özçalışkan, S., & S., G.-M. (2005). Do parents lead their children by the hand? . *Journal of Child Language, 32*(3), 481-505.

Ozyurek, A. (2014). Hearing and seeing meaning in speech and gesture: insights from brain and behaviour. *Philos Trans R Soc Lond B Biol Sci, 369*(1651), 20130296. doi:10.1098/rstb.2013.0296

Papaeliou, C., Minadakis, G., & Cavouras, D. (2002). Acoustic patterns of infant vocalizations expressing emotions and communicative functions. *Journal of Speech, Language, and Hearing Research, 45*(2), 311-317.

Papaeliou, C. F., & Trevarthen, C. (2006). Prelinguistic pitch patterns expressing 'communication' and 'apprehension'. *Journal of Child Language, 33*(1), 163-178.

Papafragou, A. (1998). The Acquisition of Modality : Implications for Theories of Semantic Representation. *13*(3), 370-399.

Papafragou, A., Li, P., Choi, Y., & Han, C. (2007). Evidentiality in language and cognition. *103*, 253-299. doi:10.1016/j.cognition.2006.04.001

Patterson, C. J., Cosgrove, J. M., & O'Brien, R. G. (1980). Non-verbal indicants of comprehension and noncomprehension in children. *Developmental Psychology, 16*, 38-48.

Paulus, M., Proust, J., & Sodian, B. (2013). Examining implicit metacognition in 3.5-year-old children: an eye-tracking and pupillometric study. *Front Psychol, 4*, 145. doi:10.3389/fpsyg.2013.00145

Paulus, M., Tsalas, N., Proust, J., & Sodian, B. (2014). Metacognitive monitoring of oneself and others: Developmental changes during childhood and adolescence. *Journal of Experimental Child Psychology, 122*(1), 153-165. doi:10.1016/j.jecp.2013.12.011

Payne, E., Post, B., Astruc, L., Prieto, P., & Vanrell, M. M. (2011). Measuring Child Rythm. *Language and Speech, 55*(2), 203-229.

Payrató, L., & Cots, J. M. (2011). *The pragmatics of Catalan*. Berlin/ Boston: Walter de Gruyter Gmbh.

Pedlow, R., Sanson, A., & Wales, R. (2004). Children's production and comprehension of politeness in requests: Relationships to behavioural adjustment, temperament and empathy. *First Language, 24*(3), 347-367. doi:10.1177/0142723704046188

Perlman, M., Dale, R., & Lupyan, G. (2015). Iconicity can ground the creation of vocal symbols. *R Soc Open Sci, 2*(8), 150152. doi:10.1098/rsos.150152

Perner, J., & Ruffman, T. (1995). Episodic memory and autonoetic conciousness: Developmental evidence and a theory of childhood amnesia. *Journal of Experimental Child Psychology, 59*(3), 516-548. doi:10.1006/jecp.1995.1024

Perry, M., Breckinridge Church, R., & Goldin-Meadow, S. (1988). Transitional knowledge in the acquisition of concepts. *Cognitive Development, 3*(4), 359-400. doi:10.1016/0885-2014(88)90021-4

Phan, L., Meza, R., Littlewort, G., Barlett, M., & Reilly, J. (2010). *Measuring behavioural indeces of cognitive processing in children*. Paper presented at the Measuring Behavior, Eindhoven, The Netherlands.

Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation*. MIT.

Pillow, B. H. (1989). Early understanding of perception as a source of knowledge. *Journal of Experimental Child Psychology, 47*(1), 116-129. doi:10.1016/0022-0965(89)90066-0

Pillow, B. H., & Anderson, K. L. (2006). Children's awareness of their own certainty and understanding of deduction and guessing. *British Journal of Developmental Psychology, 24*(4), 823-849. doi:10.1348/026151005X83360

Pinker, S. (1987). The bootstrapping problem in language acquisition. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. (pp. 399-441).). Hillsdale, NJ: Lawrence Erlbaum Associates.

Plummert, J. M. (1996). Young children's ability to detect ambiuity in descriptions of location. *Cognitive Development, 11*, 375-396.

Poggi, I., D'Errico, F., & Vincze, L. (2013). Comments by words, face and body. *Journal on Multimodal User Interfaces, 7*(1-2), 67-78. doi:10.1007/s12193-012-0102-z

Pollick, A. S., & De Waal, F. B. M. (2007). Ape gestures and language evolution. *Proceedings of the National Academy of Sciences of the United States of America, 104*(19), 8184-8189. doi:10.1073/pnas.0702624104

Pons, F., & Bosch, L. (2010). Stress Pattern Preference in Spanish-Learning Infants: The Role of Syllable Weight. *Infancy, 15*(3), 223-245. doi:10.1111/j.1532-7078.2009.00016.x

Post, B., & Payne, E. (2018). Speech rhythm in development: What is the child acquiring? In P. Prieto & N. Esteve-Gibert (Eds.), *The Development of Prosody in First Language Acquisition* (pp. 125-144). Amsterdam: John Benjamins.

Pratt, C., & Bryant, P. (1990). Young Children Understand That Looking Leads to Knowing (So Long as They Are Looking into a Single Barrel). *Child Development, 61*(4), 973-982. doi:10.1111/j.1467-8624.1990.tb02835.x

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

Prieto, P. (2014). Intonational phonology of Catalan. In S.-A. Jun (Ed.), *Prosodic typology II: The phonology of intonation and phrasing* (pp. 43–80). Oxford: Oxford University Press.

Prieto, P. (2015). Intonational meaning. *Wiley Interdisciplinary Reviews: Cognitive Science, 6*(4), 371-381. doi:10.1002/wcs.1352

Prieto, P., Borràs-Comes, J., Tubau, S., & Espinal, M. T. (2013). Prosody and gesture constrain the interpretation of double negation. *Lingua, 131*, 136-150. doi:10.1016/j.lingua.2013.02.008

Prieto, P., Estrella, A., Thorson, J., & Vanrell, M. D. M. (2012a). Is prosodic development correlated with grammatical and lexical development? Evidence from emerging intonation in Catalan and Spanish. *Journal of Child Language, 39*(2), 221-257.

Prieto, P., Estrella, A., Thorson, J., & Vanrell, M. M. (2012b). Is prosodic development correlated with grammatical and lexical development? Evidence from emerging intonation in Catalan and Spanish. *Journal of Child Language, 39*(02), 221-257. doi:10.1017/S030500091100002X

Prieto, P., & Rigau, G. (2011). Prosody and pragmatics. In L. Payrató & J. M. Cots (Eds.), *The Pragmatics of Catalan,* (pp. 17-48). The Hague: Mouton de Gruyter.

Prieto, P., & Vanrell, M. M. (2007). Early intonational development in Catalan. *Proceedings of XVIth International Congress of Phonetic Sciences*, Dudweiler, 309-314.

Pronina, M., Hübscher, I., Holler, J., & Prieto, P. (in preparation). Does enacting mental states and emotions improve children's perspective taking and pragmatic skills? A training study.

Pulvermüller, F., Hauk, O., Nikulin, V. V., & Ilmoniemi, R. J. (2005). Functional links between motor and language systems. *European Journal of Neuroscience, 21*(3), 793-797. doi:10.1111/j.1460-9568.2005.03900.x

Puts, D. A., Gaulin, S. J. C., & Verdolini, K. (2006). Dominance and the evolution of sexual dimorphism in human voice pitch. *Evolution and Human Behavior, 27*(4), 283-296. doi:10.1016/j.evolhumbehav.2005.11.003

Puts, D. A., Hodges, C. R., Cárdenas, R. A., & Gaulin, S. J. C. (2007). Men's voices as dominance signals: vocal fundamental and formant frequencies influence dominance attributions among men. *Evolution and Human Behavior, 28*(5), 340-344. doi:10.1016/j.evolhumbehav.2007.05.002

Quam, C., & Swingley, D. (2012). Development in children's interpretation of pitch cues to emotions. *Child Development, 83*(1), 236-250.

Ramus, F. (2002). Language discrimination by newborns. *Annual Review of Language Acquisition, 2*, 85-115.

Read, B., & Cherry, L. (1978). Preschool Children's Production of Directive Forms. *Discourse Processes, 1*, 233-245.

Riseborough, M. G. (1981). Physiographic gestures as decoding facilitators: Three experiments exploring a neglected facet of communication. *Journal of Nonverbal Behavior, 5*(3), 172-183.

Roberson, D., Damjanovic, L., & Kikutani, M. (2010). Show and tell: The role of language in categorizing facial expression of emotion. *Emotion Review, 2*(3), 255-260. doi:10.1177/1754073910361979

Robinson, E. J., Mitchell, P., & Nye, R. (1995). *Young children's treating of utterances as unreliable sources of knowledge*. Retrieved from

Robinson, E. J., & Robinson, W. P. (1982). Knowing when you don't know enough: Children's judgements about ambiguous information. *Cognition, 12*(3), 267-280. doi:10.1016/0010-0277(82)90034-8

Robinson, E. J., & Whitcombe, E. L. (2003). Children's Suggestibility in Relation to their Understanding about Sources of Knowledge. *Child Development, 74*(1), 48-62.

Robinson, J. D., & Heritage, J. (2016). How patients understand physicians' solicitations of additional concerns: implications for up-front agenda setting in primary care. *Health Communication, 31*(4), 434-444. doi:10.1080/10410236.2014.960060

Rochat, P., Querido, J. G., & Striani, T. (1999). Emerging sensitivity to the timing and structure of protoconversation in early infancy. *Developmental Psychology, 35*(4), 950-957.

Rochet-Capellan, A., Laboissière, R., Galván, A., & Schwartz, J. L. (2008). The speech focus position effect on jaw-finger coordination in a pointing task. *Journal of Speech, Language, and Hearing Research, 51*(6), 1507-1521. doi:10.1044/1092-4388(2008/07-0173)

Roebers, C. M. (2017). Executive function and metacognition: Towards a unifying framework of cognitive self-regulation. *Developmental Review, 45*, 31-51. doi:10.1016/j.dr.2017.04.001

Rohwer, M., Kloo, D., & Perner, J. (2012). Escape From Metaignorance: How Children Develop an Understanding of Their Own Lack of Knowledge. *Child Development, 83*(6), 1869-1883. doi:10.1111/j.1467-8624.2012.01830.x

Romøren, A. S. H., & Chen, A. (2015). Quiet is the New Loud: Pausing and Focus in Child and Adult Dutch. *Language and Speech, 58*(1), 8-23. doi:10.1177/0023830914563589

Roseano, P., González, M., Borràs-Comes, J., & Prieto, P. (2016). Communicating Epistemic Stance: How Speech and Gesture Patterns Reflect Epistemicity and Evidentiality. *Discourse Processes, 53*(3), 135-174. doi:10.1080/0163853X.2014.969137

Rowe, M. L., & Goldin-Meadow, S. (2009). Early gesture selectively predicts later language learning. *Developmental Science, 12*(1), 182-187. doi:10.1111/j.1467-7687.2008.00764.x

Rozin, P., Haidt, J., & McCauley, C. R. (1999). Disgust: The body and soul emotion. In T. Dalgleish & M. J. Power (Eds.), *Handbook of cognition and emotion* (pp. 429-445). Chichester, UK: John Wiley & Sons, Ltd.

Ruffman, T. K., & Olson, D. R. (1989). Children's Ascriptions of Knowledge to Others. *Developmental Psychology, 25*(4), 601-606.

Ruiz Santabalbina, M. (2013). Duración y percepción de la cortesía en español (experimento piloto). In A. Cabedo Nebot, M. J. Aguilar Ruiz, & E. López-Navarro (Eds.), *Estudios de lingüística: Investigaciones, propuestas y aplicaciones* (pp. 411–425). Valencia: Universitat de València.

Ryckebusch, C., & Marcos, H. (2004). Speech acts, social context and parent-toddler play between the ages of 1;5 and 2;3. *Journal of Pragmatics, 36*(5), 883-897. doi:10.1016/j.pragma.2003.10.019

Sabbagh, M. A., & Baldwin, D. A. (2001). Learning Words from Knowledgeable versus Ignorant Speakers: Links between Preschoolers' Theory of Mind and Semantic Development. *Child Development, 72*(4), 1054-1070.

Saint-Georges, C., Chetouani, M., Cassel, R., Apicella, F., Mahdhaoui, A., Muratori, F., . . . Cohen, D. (2013). Motherese in Interaction: At the Cross-Road of Emotion and Cognition? (A Systematic Review). *PLoS ONE, 8*(10), 1-18. doi:10.1371/journal.pone.0078103

Sakkalou, E., & Gattis, M. (2012). Infants infer intentions from prosody. *Cognitive Development, 27*(1), 1-16.

Sansavini, A., Bello, A., Guarini, A., Savini, S., Stefanini, S., & Caselli, M. C. (2010). Early development of gestures, object-related-actions, word comprehension and word production, and their relationships in Italian infants: A longitudinal study. *Gesture, 10*(1), 52-85. doi:10.1075/gest.10.1.04san

Scheiner, E., Hammerschmidt, K., Jürgens, U., & Zwirner, P. (2003). Unterschiede im vokalen Ausdruck von Emotionen bei hörenden und hochgradig schwerhörigen Säuglingen. In M. Gross & E. Kruse (Eds.), *Aktuelle phoniatrisch-pädaudiologische Aspekte 2003/2004* (Vol. 11, pp. 282–286). Heidelberg: Median.

Schieffelin, B. B. (1990). *The Give and Take of Everyday Life: Language Socialization of Kaluli Children* Cambridge: CUP.

Schneider, W. (2008). The development of metacognitive knowledge in children and adolescents: Major trends and implications for education. *Mind, Brain, and Education, 2*(3), 114-121. doi:10.1111/j.1751-228X.2008.00041.x

Sealey, A. (1999). 'Don't be cheeky': Requests, directives and being a child. *Journal of Sociolinguistics, 3*(1), 24-40.

Sekerina, I., Stromswold, K., & Hestvik, A. (2004). How do adults and children process referentiality ambiguous pronouns. *Journal of Child Language, 31*, 123-152.

Sekine, K., & Kita, S. (2015). The parallel development of the form and meaning of two-handed gestures and linguistic information packaging within a clause in narrative. . *Open Linguistics, 1*(1), 490-502.

Shatz, M., Diesendruck, G., Martinez-Beck, I., & Akar, D. (2003). The influence of language and socioeconomic status on children's understanding of false belief. *Developmental Psychology, 39*(4), 717-729. doi:10.1037/0012-1649.39.4.717

Shatz, M., & Gelman, R. (1973). The development of communication skills: modifications in the speech of young children as a function of listener. *Monographs of the Society for Research in Child Development, 38*(5), 1-38.

Shatz, M., Wellman, H. M., & Silber, S. (1983). The acquisition of mental verbs: A systematic investigation of the first reference to mental state. *Cognition, 14*(3), 301-321.

Shatz, M., & Wilcox, S. (1991). Constraints on the acquisition of English modals. In S. Gelman & J. Byrnes (Eds.), *Perspectives on language and thought* (pp. 319-353). Cambridge: Cambridge University Press.

Shochi, T., Erickson, D., Sekiyama, K., Rilliard, A., & Aubergé, V. (2009). Japanese children's acquisition of prosodic politeness expressions. *Proceedings of Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 1743-1746.

Skipper, J. I. (2014). Echoes of the spoken past: How auditory cortex hears context during speech perception. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1651). doi:10.1098/rstb.2013.0297

Slobin, D. (1978). A case study of early language awareness. In A. Sinclair, R. Jarvella, & W. J. M. Levelt (Eds.), *The child's conception of language* (pp. 45–54). Heidelberg, Germany: Springer-Verlag.

Smith, C. L., & Tager-Flusberg, H. (1982). Metalinguistic awareness and language development. *Journal of Experimental Child Psychology, 34*(3), 449-468. doi:10.1016/0022-0965(82)90071-6

Snow, D. P. (2006). Regression and reorganization of intonation between 6 and 23 months. *Child Development, 77*(2), 281-296. doi:10.1111/j.1467-8624.2006.00870.x

Snow, D. P. (2017). Gesture and intonation are "sister systems" of infant communication: Evidence from regression patterns of language development. *Lang Sci, 59*, 180-191. doi:10.1016/j.langsci.2016.10.005

So, W. C., Chen Hui, C. S., & L., W.-S. J. (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.

Sodian, B., Thoermer, C., Kristen, S., & Perst, H. (2012). Metacognition in infants and young children. In M. J. Berran (Ed.), *Foundations of Metacognition* (pp. 119–133). Oxford: Oxford University Press.

Soken, N. H., & Pick, A. D. (1992). Infants' perception of dynamic affective expressions: Do infants distinguish specific expressions? *Child Development, 70*, 1275–1282.

Speer, S. R., & Ito, K. (2009). Prosody in first language acquisition - Acquiring intonation as a tool to organize information in conversation. *Linguistics and Language Compass, 3*(1), 90-110. doi:10.1111/j.1749-818X.2008.00103.x

Spencer-Oatey, H. (2011). Conceptualising 'the relational' in pragmatics: Insights from metapragmatic emotion and (im)politeness comments. *Journal of Pragmatics, 43*(14), 3565-3578. doi:10.1016/j.pragma.2011.08.009

Srinivasan, R. J., & Massaro, D. W. (2003). Perceiving Prosody from the Face and Voice: Distinguishing Statements from Echoic Questions in English. *Language and Speech, 46*(1), 1-22.

Stites, L. J., & Özçalışkan, Ş. (2017). Who Did What to Whom? Children Track Story Referents First in Gesture. *Journal of Psycholinguistic Research, 46*(4), 1019-1032. doi:10.1007/s10936-017-9476-0

Striano, T., & Tomasello, M. (2001). Infant development: Physical and Social Cognition. In N. J. Smelser & P. Baltes (Eds.), *International Encyclopedia of the Social and Behavioral Sciences* (pp. 7410-7414). Oxford: Pergamon.

Sullivan, M. W., & Lewis, M. (2003a). Contextual determinants of anger and other negative expressions in young infants. *Developmental Psychology, 39* 693–705.

Sullivan, M. W., & Lewis, M. (2003b). Emotional expressions of young infants and children. *Infants and Young Children, 16*(2), 120-142.

Swerts, M., & Krahmer, E. (2005). Audiovisual prosody and feeling of knowing. *Journal of Memory and Language, 53*, 81-94. doi:10.1016/j.jml.2005.02.003

Swerts, M., & Krahmer, E. (2008). Facial expression and prosodic prominence: Effects of modality and facial area. *Journal of Phonetics, 36*(2), 219-238. doi:10.1016/j.wocn.2007.05.001

Székely, E., Tiemeier, H., Arends, L. R., Jaddoe, V. W. V., Hofman, A., Verhulst, F. C., & Herba, C. M. (2011). Recognition of facial expressions of emotions by 3-year-olds. *Emotion, 11*(2), 425-435. doi:10.1037/a0022587

Szumilas, M. (2010). Explaining Odds Ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry, 19*, 227.

Tardif, T., Wellman, H. M., & Cheung, K. M. (2004). False belief understanding in Cantonese-speaking children. *Journal of Child Language, 31*(4), 779-800. doi:10.1017/S0305000904006580

Tardif, T., Wellman, H. M., Fung, K. Y. F., Liu, D., & Fang, F. (2005). Preschoolers' understanding of knowing-that and knowing-how in the United States and Hong Kong. *Developmental Psychology, 41*(3), 562-573. doi:10.1037/0012-1649.41.3.562

Thompson, L. A. (1995). Encoding and memory for visible speech and gestures a comparison between young and older adults. *Psychology and Aging, 10*(2), 215-228. doi:10.1037/0882-7974.10.2.215

Thompson, R. L., Vinson, D. P., Woll, B., & Vigliocco, G. (2012). Language learning is iconic. *Psychological Science, 23*(12). doi:10.1177/0956797612459763

Thorson, J. (2018). The role of prosody in early word learning: Behavioral evidence. In P. Prieto & N. Esteve-Gibert (Eds.), *The Development of Prosody in First Language Acquisition* (pp. 59-78). Amsterdam: John Benjamins.

Thorson, J. M., J. L. (2015). Acoustic correlates of information structure in child and adult speech. *Proceedings of 29th annual Boston University Conference on Language Development [BUCLD 39]*, Boston, 411–423.

Tomasello, M. (1992). *First Verbs: A Case Study of Early Grammatical Development*. Cambridge: Cambridge University Press.

Tomasello, M. (1995). Joint attention as social cognition (pp. 103-130).

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition, 74*(3), 209-253.

Tomasello, M. (2001). Perceiving intentions and learning words in the second year of life. In M. Bowerman & S. Levinson (Eds.), *Language Acquisition and Conceptual Development*. Cambridge: Cambridge University Press.

Tomasello, M. (2008). *Origins of Human Communication*.

Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A New Look at Infant Pointing. *78*(3), 705-722.

Tracy, J. L., Robins, R. W., & Lagattuta, K. H. (2005). Can children recognize pride? *Emotion, 5*(3), 251-257. doi:10.1037/1528-3542.5.3.251

Tree, A. R., & Manusov, V. (1998). Managing face concerns in criticism integrating nonverbal behaviors as a dimension of politeness in female friendship dyads. *Human Communication Research, 24*(4), 564-583.

Trevarthen, C. (1979). Instincts for human understanding and for cultural cooperation: their development in infancy. In M. von Cranach, K. Foppa, W. Lepenies, & D. Ploog (Eds.), *Human ethology: claims and limits of a new discipline* (pp. 530-571). Cambridge: Cambridge University Press.

Tsuji, H., & Doherty, M. J. (2014). Early development of metalinguistic awareness in Japanese: Evidence from pragmatic and phonological aspects of language. *First Language, 34*(3), 273-290. doi:10.1177/0142723714538003

Tunmer, W., & Herriman, M. (1984). The development of metalinguistic awareness: A conceptual overview. In W. W. Tunmer, C. Pratt, & M. Herriman (Eds.), *Metalinguistic awareness in children* (pp. 12-35). New York: Springer-Verlag.

Uçar, E., & Bal, Ö. A. (2015). Preschoolers' use of requests. *Dilbilim Araştırmaları Dergisi, 2*, 25-43.

Ünal, E., & Papafragou, A. (2016). Production-comprehension asymmetries and the acquisition of evidential morphology. *Journal of Memory and Language, 89*, 179-199.

Vaillant-Molina, M., Bahrick, L. E., & Flom, R. (2013). Young infants match facial and vocal emotional expressions of other infants. *Infancy, 18*(SUPPL.1). doi:10.1111/infa.12017

Vaish, A., & Striano, T. (2004). Is visual reference necessary? Contributions of facial versus vocal cues in 12-month-olds' social referencing behavior. *Developmental Science, 7*(3), 261-269. doi:10.1111/j.1467-7687.2004.00344.x

Vanrell, M. M., Feldhausen, I., & Astruc, L. (in press). The Discourse Completion Task in Romance prosody research: status quo and outlook. In I. Feldhausen, U. Reich, & M. M. Vanrell (Eds.), *Methods in prosody: A Romance language perspective*. Berlin: Language Science Press.

Vernice, M., & Guasti, M. T. (2014). Effects of prosodic cues on topic continuity in child language production. *First Language, 34*(5), 406-427. doi:10.1177/0142723714550111

Vigliocco, G., Perniss, P., & Vinson, D. (2014). Language as a multimodal phenomenon: implications for language learning, processing and evolution. *Philos Trans R Soc Lond B Biol Sci, 369*(1651), 20130292. doi:10.1098/rstb.2013.0292

Vihman, M. M., Nakai, S., & DePaolis, R. A. (2006). Getting the rhythm right: A cross-linguistic study of segmental duration in babbling and first words. In L. Goldstein, D. Whalen, & C. Best (Eds.), *Laboratory Phonology 8* (pp. 341-366). New York: Mouton de Gruyter.

Vincze, L., & Poggi, I. (2017). I am really certain of this! Towards a multimodal repertoire of signals communicating a high degree of certainty. *Proceedings of Linköping Electronic Conference Proceedings 141*, 102-110.

Visser, M., Krahmer, E., & Swerts, M. (2014). Children's expression of uncertainty in collaborative and competitive contexts. *Language and Speech, 57*(1), 86-107.

Wagner, L., Speer, S. R., Moore, L. C., McCullough, E. A., Ito, K., Clopper, C. G., & Campbell-Kibler, K. (2015). Linguistics in a Science Museum: Integrating Research, Teaching, and Outreach at the Language Sciences Research Lab. *Linguistics and Language Compass, 9*(10), 420-431. doi:10.1111/lnc3.12164

Wagner, L., Vega-Mendoza, M., & Horn, S. V. (2014). First Language comprehension. doi:10.1177/0142723714544413

Walker-Andrews, A. S. (1997). Infants' perception of expressive behaviors: Differentiation of multimodal information. *Psychological Bulletin, 121*, 437–456.

Walker-Andrews, A. S., & Grolnick, W. (1983). Discrimination of vocal expressions by young infants. *Infant Behavior and Development, 6*(4), 491-498. doi:10.1016/S0163-6383(83)90331-4

Ward, G., & Hirschberg, J. (1988). Intonation and propositional attitude the pragmatics of L* + H L H%. . *Proceedings of the Fifth Eastern States Conference on Linguistics*, Philadelphia, 512–522.

Waxer, M., & Morton, J. B. (2011). Children's judgments of emotion from conflicting cues in speech: Why 6-year-olds are so inflexible. *Child Development, 82*(5), 1648-1660. doi:10.1111/j.1467-8624.2011.01624.x

Wellman, H. M., Hollander, M., & Schult, C. A. (1996). Young Children's Understanding of Thought Bubbles and of Thoughts. *Child Development, 67*(3), 768-788. doi:10.1111/j.1467-8624.1996.tb01763.x

Wells, B., Peppé, S., & Goulandris, N. (2004). Intonation development from five to thirteen. *Journal of Child Language, 31*(4), 749-778.

Werker, J. F., Pons, F., Dietrich, C., Kajikawa, S., Fais, L., & Amano, S. (2007). Infant-directed speech supports phonetic category learning in English and Japanese. *Cognition, 103*(1), 147-162. doi:10.1016/j.cognition.2006.03.006

Wermke, K., Teiser, J., Yovsi, E., Kohlenberg, P. J., Wermke, P., Robb, M., . . . Lamm, B. (2016). Fundamental frequency variation within neonatal crying: Does ambient language matter? *Speech, Language and Hearing, 19*(4), 211-217. doi:10.1080/2050571X.2016.1187903

Whitcombe, E. L., & Robinson, E. J. (2000). Children's decisions about what to believe and their ability to report the source of their belief. *Cognitive Development, 15*(3), 329-346. doi:10.1016/S0885-2014(00)00033-2

Whitebread, D., Coltman, P., Pasternak, D. P., Sangster, C., Grau, V., Bingham, S., . . . Demetriou, D. (2009). The development of two observational tools for assessing metacognition and self-regulated learning in young children. *Metacognition and Learning, 4*(1), 63-85.

Widen, S. C. (2013). Children's interpretation of facial expressions: The long path from valence-based to specific discrete categories. *Emotion Review, 5*(1), 72-77. doi:10.1177/1754073912451492

Widen, S. C., & Russell, J. A. (2003). A Closer Look at Preschoolers' Freely Produced Labels for Facial Expressions. *Developmental Psychology, 39*(1), 114-128. doi:10.1037//0012-1649.39.1.114

Widen, S. C., & Russell, J. A. (2008). Children acquire emotion categories gradually. *Cognitive Development, 23*(2), 291-312. doi:10.1016/j.cogdev.2008.01.002

Wilutzky, W. (2015). Emotions as pragmatic and epistemic actions. *Frontiers in Psychology, 6*(OCT). doi:10.3389/fpsyg.2015.01593

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*(1), 103-128.

Winter, B., & Grawunder, S. (2012). The phonetic profile of Korean formal and informal speech registers. *Journal of Phonetics, 40*(6), 808-815. doi:10.1016/j.wocn.2012.08.006

Yuan, C., González-Fuente, S., Baills, F., & Prieto, P. (2018, in press). Observing pitch gestures favors the learning of Spanish intonation by Mandarin speakers. *Studies in Second Language Acquisition*.

Zhang, D., Liu, Y., Hou, X., Sun, G., Cheng, Y., & Luo, Y. (2014). Discrimination of fearful and angry emotional voices in sleeping human neonates: A study of the mismatch brain responses. *Frontiers in behavioral neuroscience, 8*(DEC), 1-10. doi:10.3389/fnbeh.2014.00422

Zuczkowski, A., Bongelli, R., & Riccioni, I. (2017). *Epistemic Stance in Dialogue: Knowing, Unknowing, Believing*. Amsterdam: John Benjamins.

Zufferey, S. (2016). Pragmatic acquisition. In J.-O. Östman & J. Verschueren (Eds.), *Handbook of Pragmatics*. Amsterdam: John Benjamins.

Appendices

Appendix A: Chapter 2

Target words and phrases used for the comprehension task with their English translation in the audio-only condition.

Intonation Condition

Certain		**Uncertain**	
Catalan	English	Catalan	English
El gos (L* L%)	Dog	El gos (L* H%)	Dog
El futbol (L* L%)	Soccer	El futbol (L* H%)	Soccer
La poma (L* L%)	Apple	La poma (L* H%)	Apple
La guitarra (L* L%)	Guitar	La guitarra (L* H%)	Guitar
El pernil (L* L%)	Ham	El pernil (L* H%)	Ham
La pizza (L* L%)	Pizza	La pizza (L* H%)	Pizza
El tomàquet (L* L%)	Tomato	El tomàquet(L* H%)	Tomato
La platja (L* L%)	Beach	La platja (L* H%)	Beach
El blau (L* L%)	Blue	El blau (L* H%)	Blue

Lexical Condition

Certain		**Uncertain**	
Catalan	English	Catalan	English
Segur que el gos (L* L%)	[I am] certain that [it's] the dog.	Potser el gos (L* L%)	Maybe [it's] the dog.
Segur que el futbol (L* L%)	[I am] certain that [it's] soccer.	Potser el futbol (L* L%)	Maybe [it's] soccer.
Segur que la poma (L* L%)	[I am] certain that [it's] the apple.	Potser la poma (L* L%)	Maybe [it's] the apple.
Segur que la guitarra (L* L%)	[I am] certain that [it's] the guitar.	Potser la guitarra (L* L%)	Maybe [it's] the guitar.
Segur que el pernil (L* L%)	[I am] certain that [it's] ham.	Potser el pernil (L* L%)	Maybe [it's] ham.
Segur que la pizza (L* L%)	[I am] certain that [it's] the pizza	Potser la pizza (L* L%)	Maybe [it's] the pizza.
Segur que el tomàquet (L* L%)	[I am] certain that [it's] the tomato	Potser el tomàquet (L* L%)	Maybe [it's] the tomato.
Segur que la platja (L* L%)	[I am] certain that [it's] the beach	Potser la platja (L* L%)	Maybe [it's] the beach.
Segur que el blau (L* L%)	[I am] certain that [it's] blue	Potser el blau (L* L%)	Maybe [it's] blue.

Appendix B: Chapter 4

DCT

Instructions:

Imagine yourself being a pre-school teacher, in charge of a small group of three to four-year-old children. The task consists of 10 situations. Imagine yourself being in each of these situations and then respond to them as spontaneously as possible.

Non-polite stance condition	Polite stance condition
Scene 1a. The child is playing with your smartphone and dropped it. You have already asked him/her to stop playing with your smartphone before. You're getting impatient. Tell the child to give you the smartphone.	**Scene 1b.** You have both your hands full of plates on your way to the kitchen and you have just dropped a fork. Ask the child nicely to give you the fork.
Scene 2a. The child is scribbling in a book from the class. You have told the child many times before to not scribble in the picture books. You're getting annoyed and tell the child to give you the book.	**Scene 2b.** The child is reading a picture book with a pen close by. You need to write something down. Ask the child nicely to give you the pen.
Scene 4a. The child is very excited and plays continuously with a noisy toy but you want him/her to be quiet. You're quite annoyed by the time. Tell the child to give you the toy.	**Scene 4b.** You bring some toys to the table. A huge frog is still in front of the child. In order to have room for the new toys you ask the child nicely to give you the frog.

Scene 5a. The child is playing with a ball in the kitchen and you have told the child several times before not to play with the ball there. You get annoyed and tell the child to give you the ball.	**Scene 5b.** You're sitting at a table next to the children. You can't reach the bread and you ask the child who sits closest to the bread nicely to give you the bread.

Appendix C: Chapter 5

Table A. Mean occurrence (and standard deviation) of verbal requests across social distance, cost and age conditions.

	Age Group	Low Social distance	High Social distance	Low Cost	High Cost
Verbal	**1**	.9020 (.3003)	10000 (.0000)	.9074 (.2926)	10000 (.0000)
	2	.7451 (.4401)	10000 (.0000)	.7857 (.4140)	.9825 (.1325)

Table B. Mean occurrence (and standard deviation) morphosyntactic and lexical marking in verbal requests across social distance, cost and age conditions.

	Age Group	Low social distance	High social distance	Low Cost	High Cost
Indirect request	**1**	.8235 (.3850)	.6129 (.4911)	.6111 (.4921)	.7966 (.4060)
	2	.5882 (.4971)	.7742 (.4215)	.5714 (.4994)	.8070 (.3981)
Si us plau	**1**	.0980 (.3003)	.2419 (.4318)	.1296 (.3390)	.2203 (.4180)
	2	.0392 (.1960)	.2419 (.4318)	.0893 (.2877)	.2105 (.4113)

Table C. Mean occurrence (and standard deviation) of rising intonation in verbal requests across social distance, cost and age conditions.

	Age Group	Low Social distance	High Social distance	Low Cost	High Cost
Rising intonation	**1**	.6304 (.4880)	.5323 (.5030)	.4898 (.5051)	.6441 (.4829)
	2	.5676 (.5022)	.6613 (.4771)	.5116 (.5058)	.7143 (.4558)

Table D. Mean occurrence (and standard deviation) for the 12 gestural/body signals occurring during requests across social distance, cost and age conditions.

	Age Group	Low Social distance	High Social distance	Low Cost	High Cost
Gaze at the interlocutor	**1**	.808 (.841)	1742 (.957)	1309 (1069)	1322 (.973)
	2	.740 (.853)	1774 (.756)	.982 (.884)	1643 (.903)
Averted gaze	**1**	1558 (.916)	1371 (.979)	1709 (1031)	1220 (.811)
	2	1500 (.735)	1274 (.750)	1429 (.657)	1321 (.834)
Sidelong gaze	**1**	.077 (.269)	.484 (.805)	.309 (.717)	.288 (.589)
	2	.080 (.274)	.177 (.426)	.089 (.288)	.179 (.431)
Eyebrow raising	**1**	.077 (.269)	.306 (.561)	.164 (.420)	.237 (.503)
	2	.020 (.141)	.194 (.538)	.125 (.429)	.107 (.412)
Smile	**1**	.519 (.754)	.952 (.965)	.745 (1040)	.763 (.751)
	2	.600 (.670)	.855 (.743)	.661 (.668)	.821 (.765)
Manipulation adaptor	**1**	.462 (.641)	.968 (.789)	.727 (.757)	.746 (.779)
	2	.480 (.677)	.823 (.615)	.589 (.626)	.750 (.694)
Head tilt	**1**	.904 (1053)	1210 (1103)	1000 (1089)	1136 (1090)
	2	.720 (.809)	1032 (1040)	.679 (.917)	1107 (.947)
Head down	**1**	.442 (.669)	.484 (.565)	.418 (.567)	.508 (.653)
	2	.200 (.404)	.435 (.643)	.286 (.563)	.375 (.558)
Shoulder shrug	**1**	.481 (.671)	.629 (.910)	.655 (.907)	.475 (.704)
	2	.120 (.385)	.419 (.714)	.286 (.653)	.286 (.563)

Crouch/slumped shoulders	**1**	.327 (.585)	.661 (.676)	.455 (.603)	.559 (.702)
	2	.160 (.468)	.323 (.566)	.161 (.371)	.339 (.640)
Trunk forward	**1**	.423 (.667)	.597 (.778)	.473 (.742)	.559 (.726)
	2	.280 (.536)	.548 (.739)	.446 (.737)	.411 (.596)
Trunk lateral leaning	**1**	.731 (.770)	.710 (.894)	.691 (.879)	.746 (.801)
	2	.300 (.580)	.694 (.951)	.411 (.781)	.625 (.865)

Appendix D: Publication List

Journal papers

Armstrong, M., Esteve-Gibert, N., **Hübscher, I.**, Igualada, A., & Prieto, P. (accepted). Developmental and cognitive aspects of children's disbelief comprehension through intonation and facial gesture. *First Language.*

Hübscher, I., Borràs-Comes, J., & Prieto, P. (2017). Prosodic mitigation characterizes Catalan formal speech: The Frequency Code reassessed. *Journal of Phonetics 65*, 145-159.

Hübscher, I., Esteve-Gibert, N., Igualada, A., & Prieto, P. (2017). Intonation and gesture as bootstrapping devices in speaker uncertainty. *First Language, 37*(1), 24-41.

Hübscher, I., Garufi, M., Prieto, P. (minor revision). Preschooler's development of polite stance: how gesture and prosody pave the way. *Journal of Child Language.*

Hübscher, I., Vincze, L., & Prieto, P. (under review). Children's signalling of their uncertain knowledge state: prosody and body cues come first. *Journal of Language Learning and Development.*

Hübscher, I., Wagner, L., & Prieto, P. (accepted). Three-year-olds' infer polite stance from intonation and facial cues. *Journal of Politeness Research Language and Culture.*

Papers in conference proceedings

Hübscher, I., Wagner, L., & Prieto, P. (2016). Young children's sensitivity to polite stance expressed through audiovisual prosody in requests. *Proceedings of the Speech Prosody 2016*. Boston, MA (USA).

Hübscher, I., Garufi, M., & Prieto, P. (2018). Preschoolers use prosodic mitigation strategies to encode polite stance. *Proceedings of the Speech Prosody 2018*. Poznań (Poland).

Book chapter

Armstrong, M. & **Hübscher, I.** (2018). Children's development of internal state prosody. In P. Prieto & N. Esteve-Gibert (Eds.), *Prosodic Development in First Language Acquisition* (pp. 272-293). John Benjamins.

Appendix E: Biographical note

Iris Carmela Hübscher was born in Zurich, Switzerland in 1987. She studied English, history and anthropology for her BA at the University of Zürich, Switzerland and also completed an exchange semester with a focus on linguistics at the University of New England in Armidale, Australia. She received her bachelor's degree in 2011. She did her master's degree in general linguistics at Lancaster University in Lancaster, UK, with a focus on variational pragmatics, investigating the production of requests in three German varieties: German German, Austrian German and Swiss German. Before embarking on a PhD she spent time in Latin America where she taught English at a public primary school in Ecuador and at a Waldorf school in Costa Rica. This led her to apply for a four year PhD position (combined with a scholarship from the Spanish Ministry) on language acquisition at the Group of Prosodic Studies at the Universitat Pompeu Fabra in Barcelona, Spain which she started in 2014. For her doctoral thesis, she conducted her doctoral research in Catalunya at several public preschools and at the Center of Science and Industry (COSI) in Columbus, Ohio for which she was awarded an additional stipend by the University Pompeu Fabra for a further three-month research stay. Furthermore, she was awarded another grant by the Spanish Ministry to spend three months as a visiting scholar at the Max Planck for Psycholinguistics in Nijmegen (Netherlands).