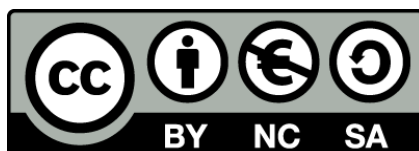




UNIVERSITAT<sub>DE</sub>  
BARCELONA

## Who speaks that language? Assessing speaker-language associations from infancy to childhood

Laia Marcet Jiménez



Aquesta tesi doctoral està subjecta a la llicència **Reconeixement- NoComercial – Compartir Igual 4.0. Espanya de Creative Commons.**

Esta tesis doctoral está sujeta a la licencia **Reconocimiento - NoComercial – Compartir Igual 4.0. España de Creative Commons.**

This doctoral thesis is licensed under the **Creative Commons Attribution-NonCommercial-ShareAlike 4.0. Spain License.**



UNIVERSITAT<sub>DE</sub>  
BARCELONA

# Who speaks that language?

## Assessing speaker-language associations from infancy to childhood

Laia Marcet Jiménez

Doctoral Program in Brain, Cognition and Behavior 2025

Department of Cognition, Development and Educational

Psychology

Supervisors:

**Dr. Ferran Pons**

**Dra. Laura Bosch**



# Acknowledgments

No puc començar d'una altra manera que agraint als meus supervisors, el Ferran i la Laura, per acompanyar-me en aquest camí. Al Ferran per fer-ho tot més fàcil i animar-me a seguir endavant quan les coses no sortien com jo volia. A la Laura per ensenyar-me el valor dels detalls i la feina ben feta. També al Joan Birulés, per ser un referent i obrir-me tantes portes.

I would also like to thank Krista Byers-Heinlein and her amazing team for showing me how fun and collaborative research can be. You have made me fall in love with science again. I am especially grateful to Andrea, Melanie, Miranda, and Nico for helping me build a home in Montreal.

También quiero dar las gracias a todos los amigos que han estado a mi lado de una manera u otra durante estos cuatro años. A lago por estar siempre. Por acordarse de todo, por acordarse de todos. A la gente de Barcelona, por las conversaciones más intensas y también las más absurdas. A Marc, por ser el amigo que necesitas dentro y fuera de la universidad. A la gente de Mundet por mantenerme en el camino cuando las cosas se ponían difíciles y las dudas me asaltaban. A Alex, por siempre empujarme a ser mejor, incluso cuando no lo pongo fácil. A la gente de Terrassa, por ser siempre un refugio donde reír y desconectar.

Por último, quiero dar las gracias a mi familia, por ser así de raros, porqué eso me ha llevado hasta aquí y me ha hecho ser quien soy. También a Zoe y Abril, por venir a cambiarlo todo.



# Table of Contents

Abstract .....	7
Resum .....	11
List of Figures .....	15
List of Tables .....	17
<b>Chapter 1: GENERAL INTRODUCTION .....</b>	<b>19</b>
Overview .....	21
Development of Audiovisual Speech Perception .....	25
Perception of Linguistic Information .....	26
Perception of Linguistic Information in Bilingual Infants .....	29
Perception of Indexical Information .....	31
Perception of Indexical Information in Bilingual Infants .....	37
Integration of Linguistic and Indexical Information .....	38
Integration of Linguistic and Indexical Information in Bilingual Infants .....	40
Development of Language Awareness .....	43
Development of Language Awareness in Bilingual Children .....	44
<b>Chapter 2: RESEARCH AIMS AND QUESTIONS .....</b>	<b>51</b>
Research Aims .....	53
Research Questions .....	54
<b>Chapter 3: SPEAKER-LANGUAGE ASSOCIATIONS BEFORE THE FIRST YEAR OF LIFE .....</b>	<b>57</b>
Study 1: .....	59
Introduction .....	59
Methods .....	62
Results .....	66
Discussion .....	74
Study 2: .....	77
Introduction .....	77

Methods.....	79
Results.....	81
Discussion .....	88
<b>Chapter 4: SPEAKER-LANGUAGE ASSOCIATIONS AFTER THE FIRST YEAR OF LIFE.....</b>	<b>91</b>
Study 3: .....	93
Introduction .....	93
Methods.....	96
Results.....	101
Discussion .....	127
<b>Chapter 5: GENERAL DISCUSSION.....</b>	<b>133</b>
Overview.....	135
Speaker-Language Associations Before the First Year of Life .....	138
Primary Face-Language Associations .....	138
Robust Face-Language Associations .....	143
Speaker-Language Associations After the First Year of Life .....	145
Pragmatic Differentiation .....	145
Language Labeling.....	149
Developmental Trajectory of Speaker-Language Associations .....	152
Limitations and Future Directions .....	155
<b>Chapter 6: CONCLUSIONS .....</b>	<b>159</b>
Conclusions.....	161
<b>Chapter 7: REFERENCES .....</b>	<b>165</b>
References.....	167
<b>Chapter 8: APPENDICES .....</b>	<b>181</b>
Appendix 1: Language Separation and Awareness Questionnaire ....	183
Appendix 2: Items for the Exploratory Factor Analysis .....	196

# Abstract

Knowing what language people in the environment usually speak can be valuable information during language acquisition, especially for infants and children growing up bilingual. In the early stages, associating speakers with their languages may promote language separation and the formation of distinct linguistic systems (Kandhadai et al., 2014). Although infants can track who uses each language within the first six months of life (Colomer et al., 2023; Kinzler et al., 2007), direct assessment of the formation of early speaker-language associations has provided inconclusive results (Schott et al., 2023). These associations remain relevant in later stages of language development by promoting pragmatic differentiation and guiding children's selection of the appropriate language based on their interlocutor (Genesee et al., 1996; Montanari, 2008; Nicoladis, 1998). Moreover, developing explicit knowledge of who speaks each language (Akhtar et al., 2012; Atagi & Sandhofer, 2020) may further enhance children's sociolinguistic awareness.

This thesis explores the emergence and evolution of speaker-language associations across development, with a special focus on infants and children in bilingual contexts. The associations are assessed before and after the first year of life, adapting the methodology to the behavioral repertoire at each developmental stage. Before the first year of life (infants aged 4 to 12 months), the associations are explored using an audiovisual preferential-looking task measuring infants' formation of primary and robust face-language associations (Study 1 and Study 2). After the first year of life (children aged 1 to 5 years), the associations are explored using a newly designed questionnaire collecting parental report measures of children's pragmatic differentiation and language labeling abilities (Study

3). Furthermore, potential linguistic factors that might modulate the formation and expression of speaker-language associations will be examined.

The findings from the present work indicate that primary face-language associations are formed as early as 4 months of age, evidenced by the presence of language-based preferences for speakers. These preferences exhibit an interesting developmental pattern, transitioning from an early familiarity preference for speakers of the native/dominant language to a novelty preference for speakers of the non-native/non-dominant language. However, the formation and retention of robust face-language associations do not appear to occur before age 1, at least for unfamiliar speakers, regardless of the linguistic background of the infant.

Regarding pragmatic differentiation, the results indicate a general establishment around age 2. Notably, children may show an earlier onset, likely dependent on their verbal skills. From age 3 onwards, no significant age-related improvements are observed. Children exposed to one language at home and the other outside demonstrate higher levels of pragmatic differentiation at an earlier age, although the reasons for this advantage require further exploration. Moreover, children are better at adapting their language use to their interlocutor when the language required is their dominant one.

Lastly, regarding language labeling, the findings reveal that this ability emerges around age 3, although some children may develop it earlier. Children frequently exposed to language reflections at home are more likely to label the languages speakers use. However, the impact of language reflections is no longer relevant after age 4, when nearly all children have developed the language labeling ability.

In conclusion, this thesis outlines the developmental trajectory of speaker-language associations. The findings indicate that, early on, infants can track the languages spoken by others, influencing their social interactions and potentially their language acquisition. Although robust speaker-language associations are not evident before the first year of life, they evolve gradually and become more complex throughout development. Once children begin speaking, these associations guide their language use and enable them to explicitly identify the languages present in their sociolinguistic environment. Moreover, this thesis highlights the importance of certain linguistic factors, though not all, in the formation and expression of speaker-language associations.



# Resum

Saber quina llengua parlen habitualment les persones de l'entorn pot ser una informació valuosa durant el procés d'adquisició del llenguatge, especialment per a infants i nens d'entorns bilingües. En les primeres etapes, associar els parlants amb les seves llengües pot afavorir la separació lingüística i la formació de sistemes lingüístics diferenciats (Kandhadai et al., 2014). Tot i que els infants poden fer un seguiment de qui utilitza cada llengua abans dels sis mesos de vida (Colomer et al., 2023; Kinzler et al., 2007), la avaluació directa de la formació primerenca d'associacions entre parlants i llengües ha proporcionat resultats poc concloents (Schott et al., 2023). Aquestes associacions continuen sent rellevants en etapes posteriors del desenvolupament lingüístic, ja que promouen la diferenciació pragmàtica i guien la selecció de la llengua adequada segons l'interlocutor (Genesee et al., 1996; Montanari, 2008; Nicoladis, 1998). A més, desenvolupar un coneixement explícit de qui parla cada llengua (Akhtar et al., 2012; Atagi & Sandhofer, 2020) pot promoure encara més la consciència sociolingüística dels nens.

Aquesta tesi explora l'emergència i l'evolució de les associacions entre parlants i llengües al llarg del desenvolupament, amb una atenció especial en infants i nens en contextos bilingües. Les associacions es valoren abans i després del primer any de vida, adaptant la metodologia al repertori conductual de cada etapa del desenvolupament. Abans del primer any de vida (infants de 4 a 12 mesos), les associacions s'exploren mitjançant una tasca de preferència visual que mesura la formació d'associacions primàries i robustes entre cares i llengües (Estudi 1 i Estudi 2). Després del primer any de vida (nens d'1 a 5 anys), les associacions s'exploren mitjançant un nou qüestionari que recull informes parentals

sobre la diferenciació pragmàtica i la habilitat d'anomenar llengües dels nens (Estudi 3). A més, s'examinen possibles factors lingüístics que podrien modular la formació i expressió de les associacions entre parlants i llengües.

Els resultats d'aquesta tesi indiquen que les associacions primàries entre cares i llengües es formen sobre els 4 mesos d'edat, com ho evidencia la presència de preferències per certs parlants en funció de la seva llengua. Aquestes preferències mostren un patró de desenvolupament interessant, passant d'una preferència primerenca per la familiaritat, cap als parlants de la llengua nativa/dominant, a una preferència per la novetat, cap als parlants de la llengua no nativa/no dominant. No obstant això, la formació i retenció d'associacions robustes entre cares i llengües sembla no produir-se abans del primer any de vida, almenys pel que fa a parlants desconeguts, independentment del context lingüístic de l'infant.

Pel que fa a la diferenciació pragmàtica, els resultats indiquen un establiment general d'aquesta capacitat al voltant dels 2 anys. Notablement, alguns nens poden mostrar un començament més primerenc, probablement depenent de les seves habilitats verbals. A partir dels 3 anys, no s'observen millores significatives relacionades amb l'edat. Els nens exposats a una llengua a casa i una altra fora demostren nivells més alts de diferenciació pragmàtica a una edat més primerenca, encara que les raons d'aquest avantatge requereixen una exploració més profunda. A més, els nens són més eficients a adaptar l'ús de la llengua a l'interlocutor quan la llengua requerida és la seva llengua dominant.

Finalment, pel que fa a la capacitat d'anomenar llengües, els resultats mostren que aquesta emergeix al voltant dels 3 anys, tot i que alguns nens poden desenvolupar-la abans. Els nens que estan freqüentment exposats a reflexions lingüístiques a casa tenen més probabilitats d'anomenar les

llengües que utilitzen els parlants. No obstant això, l'impacte de les reflexions lingüístiques deixa de ser rellevant passats els 4 anys, quan gairebé tots els nens han desenvolupat la capacitat d'anomenar llengües.

En conclusió, aquesta tesi descriu la trajectòria de desenvolupament de les associacions entre parlants i llengües. Els resultats indiquen que, des de ben aviat, els infants poden fer un seguiment de les llengües parlades pels altres, influint en les seves interaccions socials i, potencialment, en l'adquisició del llenguatge. Tot i que les associacions robustes entre parlants i llengües no són evidents abans del primer any de vida, aquestes evolucionen gradualment i es tornen més complexes durant el desenvolupament. Quan els nens comencen a parlar, aquestes associacions guien l'ús de la llengua i els permeten anomenar explícitament les llengües presents en el seu entorn sociolingüístic. A més, aquesta tesi remarca la importància de certs factors lingüístics, encara que no de tots, en la formació i expressió de les associacions entre parlants i llengües.



# List of Figures

- Figure 1.** *Study 1.* Visual representation of the experimental task. /65
- Figure 2.** *Study 1.* PTLT to speakers during the first half of the silent trial. /68
- Figure 3.** *Study 1.* PTLT to speakers during the first half of the silent trial, divided by Linguistic Background. /70
- Figure 4.** *Study 1.* PTLT to speakers during the language trials combined, divided by Linguistic Background. /72
- Figure 5.** *Study 2.* PTLT to speakers during the silent trial. /82
- Figure 6.** *Study 2.* PTLT to speakers during the silent trial, divided by Linguistic Background. /84
- Figure 7.** *Study 2.* PTLT to speakers during the language trials combined. /86
- Figure 8.** *Study 3.* Factor loadings for each item included in the EFA. /103
- Figure 9.** *Study 3.* Scatter plot Pragmatic Differentiation Score x Age. /108
- Figure 10.** *Study 3.* Bar plot Pragmatic Differentiation Score x Age Block. /109
- Figure 11.** *Study 3.* Bar plot Pragmatic Differentiation Score x Language Strategy. /111
- Figure 12.** *Study 3.* Scatter plot Pragmatic Differentiation Score x Age, divided by Language Strategy. /112
- Figure 13.** *Study 3.* Bar plot Pragmatic Differentiation Score x Language Required. /115
- Figure 14.** *Study 3.* Scatter plot Pragmatic Differentiation Score x Age, divided by Language Required. /116
- Figure 15.** *Study 3.* Bar plot Pragmatic Differentiation Score x Age Block, divided by Language Required. /117
- Figure 16.** *Study 3.* Scatter plot Language Labeling x Age. /122

**Figure 17.** *Study 3.* Scatter plot Language Labeling x Age, divided by Speaker. /123

**Figure 18.** *Study 3.* Scatter plot Language Labeling x Language Reflections. /124

**Figure 19.** *Study 3.* Bar plot Proportion of Children Who Label x Language Reflections Groups. /125

**Figure 20.** *Study 3.* Scatter plot Language Labeling x Age, divided by Language Reflections Groups. /126

# List of Tables

**Table 1.** *Study 3.* Age distribution of the participants in each bilingual community. /96

**Table 2.** *Study 3.* Summary of the linear regression model. /105

**Table 3.** *Study 3.* Summary of the linear mixed model. /114

**Table 4.** *Study 3.* Summary of the generalized mixed model. /120



Chapter 1:

**GENERAL**

**INTRODUCTION**



# Overview

The human experience is intrinsically intertwined with language, as it structures our thoughts, emotions, and social interactions. Through language we can communicate with those around us, facilitating the transmission of knowledge and the formation of social bonds. Accordingly, language acquisition is considered a crucial milestone in infants' development and has been extensively studied by researchers all over the world.

According to the Bloom & Lahey (1978) model of language, successful language development requires the acquisition of three different language components: form (phonology, morphology, and syntax), content (semantics), and use (pragmatics). These components are developed gradually over time and through experience, enabling children to effectively comprehend and express themselves.

Language is generally expressed in its spoken form, also known as speech, a complex communication mechanism uniquely found in humans. Exposure to speech starts early on, even before birth, as fetuses already perceive the auditory information contained in their mother's speech (Graven & Browne, 2008). Once they are born, they are rapidly immersed in social situations where they can both hear and see people speaking. Indeed, speech is generally perceived audiovisually, which involves the integration of its auditory and visual cues to create a unified perception.

The development of audiovisual speech perception is essential to successfully extract and process the vast amount of information conveyed in speech. This encompasses not only linguistic elements, such as the

language and the message being transmitted but also non-linguistic or indexical aspects related to the speaker producing the speech (Levi & Pisoni, 2007; Nygaard & Tzeng, 2021). Processing linguistic information enables infants to become familiar with the language spoken around them and to learn its perceptual characteristics. As they gain more experience, they develop richer and more complex linguistic systems that contain more than just the basic perceptual features of the language. They gradually acquire language awareness, defined by a deeper understanding of the form and functions of language (Carter, 2003).

Infants monolingually raised are generally exposed to a single language that is shared and used by all the people in their community. Consequently, they only need to develop one linguistic system to successfully communicate with their peers. However, it is estimated that over half of the world's population is bilingual (Grosjean, 2021). Those raised bilingually encounter a greater challenge, as they must learn two languages, often simultaneously. To succeed in this dual language acquisition, they need to learn the regularities of each language and form two distinct linguistic representations. Additionally, they are exposed to a complex sociolinguistic context, where speakers may use one language, the other, or both.

Tracking which language is spoken by people in the environment can be a highly valuable skill at various stages of bilingual language development. During early language acquisition, associating each speaker with the language they use may further support language differentiation and the formation of separate linguistic systems. Previous research has shown that, even before the first year of life, infants track the languages used by speakers, as evidenced by a preference for speakers of the native language (Colomer et al., 2023; Kinzler et al., 2007). However, the only study directly

assessing infants' formation of speaker-language associations yielded inconclusive results (Schott et al., 2023).

Associating a language to each speaker is also crucial in later stages of bilingual language development. Competent bilinguals must not only understand and speak both languages but also use the appropriate one depending on the language predominantly used by the interlocutor they are interacting with. This ability, known as pragmatic differentiation, has been reported to emerge around the second year of life (Genesee et al., 1995; Nicoladis & Genesee, 1996), or even earlier (Nicoladis, 1998). However, most studies exploring pragmatic differentiation assessed small samples, making it difficult to extract robust conclusions about its developmental trajectory. Later on, around age 3, children become able to explicitly indicate and label the language used by familiar speakers and themselves (Akhtar et al., 2012; Atagi & Sandhofer, 2020). This lays the foundation for engaging in language reflections that might enhance their understanding of their sociolinguistic context and the role of languages in social interactions.

The present thesis aims to assess the emergence and evolution of speaker-language associations across development, with a special focus on infants and children growing up in bilingual contexts. During the early stages of development (infants aged 4 to 12 months), these associations will be explored by measuring audiovisual associations between a speaker's face and the language they spoke. At later developmental stages (children aged 1 to 5 years), the assessed measures will include parental reports of children's pragmatic differentiation as well as their ability to label the languages used by different speakers. Furthermore, potential linguistic factors that might modulate the development of these abilities will be examined.

The following introduction is divided into two main sections. The first section reviews the development of audiovisual speech perception during the first year of life, including infants' perception of linguistic and indexical cues. It also addresses the integration of both types of information leading to the formation of audiovisual speaker-language associations. The second section reviews the development of language awareness, with an emphasis on sociolinguistic awareness. It examines previous research on children's implicit and explicit knowledge of the language used by different speakers in their environment, focusing on children's pragmatic differentiation and language labeling abilities.

# Development of Audiovisual Speech

## Perception

Speech is a complex communication mechanism uniquely found in humans and characterized by the production of spoken language. Speech is inherently audiovisual, as it combines both auditory and visual information. During social interactions, the acoustic properties of speech are accompanied by the visualization of the articulatory movements of the mouth that is producing the sound.

Speech also contains amodal information, cues that are not specific to a particular sensory modality but are redundant across two or more senses, such as temporal synchrony, rhythm, or intensity (Bahrack & Lickliter, 2012). For example, temporal synchrony can be extracted acoustically from the auditory components of speech but also visually from the mouth movements of the speaker who is talking. This audiovisual congruency indicates that the auditory and visual stimuli have a common source which promotes audiovisual integration and the formation of a unified perception (Bahrack & Lickliter, 2012). Moreover, audiovisual redundancy guides infants' early selective attention and perceptual learning (Bahrack & Lickliter, 2000).

The development of audiovisual speech perception is a foundational aspect of early cognitive, linguistic, and social development, enabling infants to extract and process the vast amount of information embedded in spoken language. Speech provides linguistic information about the message intended to be transmitted as well as non-linguistic or indexical information about the speaker producing the speech (Levi & Pisoni, 2007).

Both types of information, as well as their integration, contribute to obtaining a comprehensive understanding of speech and its intended message.

## **Perception of Linguistic Information**

The linguistic information contained in speech includes various aspects related to both the message being transmitted and the language being spoken (i.e., its form and its content). It includes elements regarding how the language sounds (phonology), the structure of words (morphology), the structure of sentences (syntax), and the meaning of the message (semantics) (Levi & Pisoni, 2007). The development of linguistic perception is crucial for language acquisition, as it enables individuals to extract and learn the distinctive characteristics of their native language(s).

The foundation for linguistic speech perception is established early on, even before birth. The auditory system becomes functional by the third trimester of pregnancy, and from that moment the fetus starts to perceive and process the speech produced by the mother (Graven & Browne, 2008). This prenatal experience shapes infants' auditory preferences after birth. One technique that has been extensively used to assess these preferences in newborns is the High Amplitude Sucking (HAS) procedure. This technique measures infants' sucking behavior in response to different auditory stimuli by coupling a pacifier with a pressure transducer. Vouloumanos & Werker (2007) found that newborns increased the amplitude of their suctions when this behavior elicited speech sounds, but not when it was paired with non-speech stimuli. In a similar task, newborns from monolingual mothers showed a preference for listening to the language the mother used during pregnancy, compared to an unfamiliar language (Byers-Heinlein et al., 2010; Mehler et al., 1988; Moon

et al., 1993). These initial perceptual biases are one of the core considerations of the PRIMIR framework (Processing Rich Information from Multidimensional Interactive Representations)(Curtin et al., 2011; Werker & Curtin, 2005) and guide infants' attention to some aspects of speech over others. These biases play a crucial role in early language acquisition by promoting speech processing, especially when produced in the infants' native language, and facilitating the learning of its auditory characteristics.

The native language preference observed in newborns is also indirect evidence of their ability to discriminate their native language from others perceptually. One of the most important aspects for language discrimination is rhythm. Languages can be generally divided into three different rhythmic groups: stress-timed (e.g. English), syllable-timed (e.g. Spanish), and mora-timed languages (e.g. Japanese) (Abercrombie, 1967). Newborns from monolingual mothers are already able to auditorily discriminate between two languages when they belong to different rhythmic groups (Mehler et al., 1988; Nazzi et al., 1998; Ramus et al., 2000), as they increase their sucking amplitude after a language change. The ability to discriminate between languages that belong to the same rhythmic class appears later on, around 4 to 5 months of age for monolingual infants, as long as one of the compared languages is their native one (Bosch & Sebastián-Gallés, 1997; Nazzi et al., 2000).

However, the rhythm of speech can also be conveyed visually, as infants can access the articulatory mouth movements that are producing the sound. Accordingly, visual information can also serve as a valuable cue for language discrimination. In Weikum et al. (2007), 4-, 6- and 8-month-old English monolingual infants were first habituated to silent videos of a speaker using one language (e.g., English), and then, they were tested with

silent videos of the same speakers using a different language (i.e., French). Four- and 6-month-old infants increased their visual attention after the language switch, indicating their ability to visually discriminate the languages. However, this ability declined with age, as it was no longer present in the 8-month-old group. Similar results were obtained when testing same-aged Spanish or Catalan monolingual infants in the same task (Sebastián-Gallés et al., 2012).

Despite the previously mentioned studies focused on unimodal speech perception, speech is generally perceived audiovisually, involving the integration of both its visual and auditory features. Infants as young as 2.5 months of age show evidence of detecting audiovisual speech synchrony, as they pay more attention to a talking face when it is in sync with the audio recording they are listening to (Dodd, 1979). Furthermore, 2-to-4.5-month-old infants can match a specific phonetic sound with the mouth movement that produces it (Kuhl & Meltzoff, 1982; Patterson & Werker, 2003), indicating they detected the audiovisual phonetic correspondence.

Moreover, when auditory information is compromised, visual information can support and enhance speech perception (Lansing & McConkie, 2003). In line with this, researchers found an interesting attentional developmental pattern when infants were exposed to talking faces. Four-month-old monolingual infants preferentially looked at the eyes of a speaker, but their attention shifted to the speaker's mouth at 8 months of age, coinciding with the beginning of the babbling phase (Lewkowicz & Hansen-Tift, 2012). This attentional shift increased infants' access to visual speech cues, potentially enhancing speech perception and facilitating language acquisition.

As infants gain experience with their native language, they undergo a phenomenon referred to as perceptual narrowing. In early developmental stages, infants possess the ability to discriminate a wide range of speech sound contrasts, irrespective of the language they belong to (Werker & Tees, 1984). However, as they approach the end of their first year of life, this sensitivity begins to narrow favoring the language present in their environment. Infant's sensitivity to non-native sound contrasts declines (Werker & Tees, 1984) while their ability to discriminate contrasts in their native language increases (Kuhl et al., 2006). This shift in perceptual sensitivity makes infants particularly attuned to processing their native language while diminishing their ability to process unfamiliar languages.

In line with this, at 12 months of age, monolingual infants' attention to a talking face shifted back to the eyes, but only if the language used by the speaker was the infant's native language (Lewkowicz & Hansen-Tift, 2012). When processing their native language, infants may no longer need visual support due to their native processing advantage. In contrast, when exposed to an unfamiliar language, they may rely more on visual cues present in the mouth to process it, compensating for their reduced ability to discriminate non-native contrasts.

## **Perception of Linguistic Information in Bilingual Infants**

Bilingual infants are regularly exposed to speech produced in two different languages. To achieve successful dual language acquisition, bilingual infants must process and extract the linguistic information related to each language while keeping them apart.

The first challenge bilingual infants must overcome is discriminating their languages. Evidence from infants from bilingual mothers, who used two languages during their pregnancy, suggests they can auditorily

discriminate the two languages from birth, as long as they belong to different rhythmic classes (Byers-Heinlein et al., 2010). However, infants exposed to two rhythmically similar languages need more experience with their languages to auditorily discriminate them. This ability emerges by 3 months of age for Basque-Spanish bilingual infants (Molnar et al., 2014) and by 4 months of age for infants learning Catalan and Spanish (Bosch & Sebastián-Gallés, 2001). Catalan and Spanish not only belong to the same rhythmic class but also share phonological similarities and a high number of cognate words (Bosch & Ramon-Casas, 2014), which might increase the experience needed to successfully discriminate both languages perceptually.

Bilingual infants have also been observed to discriminate their languages visually. Six- and 8-month-old bilingual infants learning two distant languages (i.e., French and English) detected a speaker switching from one language to the other in silent videos (Weikum et al., 2007). Interestingly, as previously mentioned, visual language discrimination was no longer observed in 8-month-old monolingual infants. Moreover, subsequent research demonstrated that this bilingual advantage was not exclusively related to having experience with the languages, as 8-month-old bilingual infants learning Catalan and Spanish also showed visual discrimination of English and French (Sebastián-Gallés et al., 2012). Taken together, these findings suggest that bilingual infants may rely more on visual information to further support their language acquisition process.

Differences in attention to talking faces were also observed across linguistic backgrounds. Mercure et al. (2019) found that 4-to-8-month-old bilingual infants spend more time looking at talking faces compared to same-aged monolingual infants. Interestingly, the developmental pattern of selective attention to different regions of a talking face is also shaped by

infants' linguistic experience. Pons et al. (2015) found that while 4-month-old monolingual infants looked predominantly at the eyes of a speaker (see also Lewkowicz & Hansen-Tift, 2012), same-aged bilingual infants looked equally at the eyes and the mouth. Moreover, 8- and 12-month-old bilingual infants looked preferentially at the mouth, both when exposed to native and non-native speech (Pons et al., 2015), but 12-month-old monolingual infants looked more to the mouth only when exposed to non-native speech (Lewkowicz & Hansen-Tift, 2012; Pons et al., 2015). Other studies found that the increased attention to the mouth of a talking face is even more prominent in bilingual infants learning two close languages (Birulés et al., 2019). Overall, these findings indicate that bilingual infants shift their attention to a speaker's mouth earlier in development and for a longer period of time in comparison to infants exposed to only one language. This would enable them to take greater advantage of the redundancy of audiovisual speech and use the visual information to further support their language discrimination. This perceptual adaptation seems to be more pronounced in infants learning two close languages, who could benefit even more from exploiting audiovisual information from speech during their language acquisition process.

## **Perception of Indexical Information**

Apart from linguistic information, speech also contains non-linguistic or indexical cues related to the speaker who is producing it. According to Abercrombie (1967), the indexical information can refer to 3 different aspects of the speaker: their group membership (e.g., accent), their emotional state (e.g., fatigue or excitement), or their individual characteristics (e.g., age or gender). Generally, indexical information can be accessed auditorily from the voice of the speaker, or visually from the speaker's physical appearance.

The development of indexical perception is crucial not only for providing context to the linguistic information contained in speech but also for speaker processing. Indexical cues allow for speaker recognition and speaker classification into different social categories, which will guide infants' social interactions.

For example, speakers can be categorized according to their gender. In the case of voice, male and female voices differ in several aspects, such as their fundamental frequency ranges. Initially, researchers found that infants were able to discriminate female voices from male voices around 6 months of age (Miller, 1983; Miller et al., 1982). However, subsequent research found that even near-term fetuses can discriminate voices that differ on gender (Lecanuet et al., 1993).

Newborns are also able to discriminate between two female voices, but only when one of them is from their mother (Decasper & Fifer, 1980; Fifer & Moon, 1989). Moreover, these studies found that newborns preferred to listen to their mother's voice over the voice of the other female speaker. The ability to discriminate two female voices from unfamiliar speakers does not emerge until 4 months of age (Fecher & Johnson, 2019b). When assessing the discrimination of two male voices, different results were observed. Newborns discriminate their father's voice from other male voices, but they do not show a preference for one voice over the other (Decasper & Prescott, 1984). When testing two unfamiliar male voices, 4-month-old infants showed no evidence of discrimination (Yu et al., 2024).

Male and female faces also present differential characteristics that allow for gender-based face discrimination. As early as 3 months of age, infants discriminate between male and female faces with a preference for female faces (Quinn et al., 2002). This preference is no longer present in 9-month-

old infants (Liu, Xiao, Quinn, et al., 2015). Interestingly, 3-month-old infants raised by a male primary caregiver showed a preference towards male faces, indicating the importance of experience in the development of this early face preferences (see exp. 5 in Quinn et al. (2002).

Infants can also discriminate between two faces from the same gender. For example, newborns discriminate the face of their mother from the face of unfamiliar females and show a preference for the mother's face (Field et al., 1984; Walton et al., 1992). However, discrimination of two unfamiliar female faces does not appear until three months of age (Barrera & Maurer, 1981; Quinn et al., 2002). No study to date has explored preference nor discrimination for the father's face over other unfamiliar males, but for unfamiliar male faces, discrimination is not observed until seven months of age (Righi et al., 2014).

The female bias observed both in face and voice perception is likely the reflection of a general increased exposure to female faces and voices during the first year of life, related to a higher number of infants being raised by a female primary caregiver (Rennels & Davis, 2008). The increased experience with faces and voices from that gender makes them more familiar with that type of stimuli, improving their processing abilities.

Faces can also convey information about the race of the speaker. Researchers have also explored infants' ability to discriminate faces considering this social domain. Three-month-old infants show a preference for own-race faces over other-race faces (Fassbender et al., 2016; Liu, Xiao, Xiao, et al., 2015), implying their ability to discriminate between them. Interestingly, this preference fades when they are 6 months old, and reverses when they are 9 months old, with a preference for other-race faces (Fassbender et al., 2016; Liu, Xiao, Xiao, et al., 2015). Similar to

the preference for female faces, this initial preference for own-race faces could be explained by the higher presence of own-race faces in their environment, making them more familiar to these features. When testing discrimination of faces from the same race, 3-month-old infants can discriminate between two own-race faces as well as two other-race faces. The ability to discriminate between other-race faces decreases with age, as 9-month-old infants are no longer able to discriminate two faces that belong to another race (Kelly et al., 2007). These results suggest that face perception also undergoes perceptual narrowing favoring own-race faces, similar to what is observed in the language domain (Krasotkina et al., 2018)

Regarding age, 3-to-6-month-old infants can discriminate adult faces from infant faces, with a preference for adult faces (Heron-Delaney et al., 2017). Once again, this is likely due to a higher number of interactions with adults rather than with other infants in the early months. In line with this differential experience, perceptual narrowing has also been observed to influence discrimination of faces based on age. Three-month-old infants discriminate adult and infant faces, while 9-month-old infants are only able to discriminate adult faces (Macchi Cassia et al., 2014).

However, as previously introduced, indexical information can be conveyed both visually and acoustically. From cumulative experience with the different social categories, infants acquire intermodal knowledge of how the speakers of those categories sound and look. By attending to indexical cues, infants can detect the audiovisual correspondence between faces and voices and match them according to their characteristics and the social category they belong to.

By the second half of the first year of life, infants match faces and voices based on gender (Hillairet de Boisferon et al., 2015; Poulin-Dubois et al.,

1994; Richoz et al., 2017; Walker-Andrews et al., 1991). Interestingly, most of the studies also revealed a gender asymmetry, favoring the face-voice matching for female speakers over male speakers. As previously discussed, this is likely related to infants' higher exposure to females during the early months. Interestingly, robust evidence of matching male voices with male faces is not observed until 18 months of age (Poulin-Dubois et al., 1998).

Similarly, 7-month-old infants exhibit face-voice matching abilities based on the age of the speakers (Bahrick, Netto, & Hernandez-Reif, 1998). The researchers found that those infants that were regularly exposed to other children were better at matching young faces to young voices, highlighting again the importance of experience in acquiring intermodal knowledge of social categories.

During the first year of life, infants are also able to learn the audiovisual attributes of a specific speaker. It is not surprising that one of the first face-voice pairs that infants learn are those from their caregivers, specifically their mother. As previously mentioned, newborns prefer to look at the face of their mother compared to the face of another woman, but interestingly, this only happens after simultaneous exposure to the mother's face-voice combination (Sai, 2005). This indicates that the high familiarity with their mother's voice acquired during pregnancy has prompted the formation of a face-voice association, leading to a preference towards the mother's face. Later on, by 3 months of age, infants start showing evidence of matching their parents' faces with their voices, as they looked preferentially to the face of the corresponding caregiver when their voice was playing in the background (Spelke & Owsley, 1979).

Subsequent research explored the ability to learn the face-voice combination of unfamiliar speakers. The first evidence for the formation of these associations was observed around 3-4 months of age (Bahrick et al., 2005; Brookes et al., 2001). In these studies, infants were first habituated with videos of two different speakers (speaker A and speaker B) and then presented with test trials where the face-voice combinations were switched (speaker A with voice B, and speaker B with voice A). Infants that were 3 to 6 months old, but not 2 months old, increased their looking time after the switch, indicating they had detected the change in the face-voice pairings and therefore learned the associations.

Interestingly, the study by Bahrick et al. (2005) included a subsequent trial after a 10-minute break where they assessed infants' ability to match the faces of the habituation speakers with their corresponding voices in a preferential-looking task. In this task, infants were presented with synchronized side-by-side videos of the speakers from the familiarization while the voice of one of them was playing in the background. If infants were able to retain the previously learned face-voice pairings they would look preferentially to the speaker whose voice was playing in the recording. In line with the increased cognitive demands of these trials, only 6-month-old infants showed evidence of matching the speakers with the corresponding voices.

More recent research has found it difficult to replicate the promising findings obtained with unfamiliar speakers. Fecher et al. (2019) found that 16- to 17-month-old infants were only able to match faces and voices when the pairings being compared differed in gender. Using pairings from different genders simplified the task by increasing the discriminability of the faces and voices. Furthermore, infants at this age are expected to have already acquired intermodal knowledge related to each gender (Hillairet

de Boisferon et al., 2015; Richoz et al., 2017). Accordingly, it is difficult to determine if infants have learned the specific audiovisual face-voice pairings, or if they are matching the faces and the voices based on gender. Conclusive evidence for the formation of face-voice associations of unfamiliar speakers of the same gender has not been observed until 24 months of age (Orena et al., 2022).

Altogether, these results suggest that the development of face-voice associations undergo a gradual maturation (Fecher et al., 2019). Primary associations between the acoustic and visual properties of speakers might be formed at an early age and assessed with tasks with low demands, such as those measuring infants' ability to detect changes in face-voice pairings. In contrast, more robust face-voice associations might be developed later and assessed with tasks with higher demands that require infants to encode, store in memory and retrieve the pairings, such as those measuring infants' ability to match speakers' faces with their corresponding voices.

## **Perception of Indexical Information in Bilingual Infants**

Although bilingual infants are exposed to speakers using different languages, they do not necessarily interact with a greater number or a wider variety of speakers. Consequently, fewer studies have examined potential differences in the perception of indexical information in bilinguals compared to monolingual infants. Studies that addressed this issue mainly used tasks where languages played an active role, and the integration of linguistic and indexical information was involved. Accordingly, these studies will be reviewed in the following section.

## **Integration of Linguistic and Indexical Information**

Linguistic and indexical cues are not always processed independently (Nygaard & Tzeng, 2021), and one might modulate the perception of the other. One aspect that has been extensively explored through different approaches is the influence of language on speaker processing.

As previously mentioned, monolingual infants are able to discriminate two unfamiliar female voices by 4 months of age (Fecher & Johnson, 2019b). However, evidence on 4- and 7-month-old monolingual infants indicates that they are only able to detect the change from one voice to the other when they are speaking in the infant's native language (Fecher & Johnson, 2018, 2019b; Johnson et al., 2011). Similarly, studies revealed that 9-month-old monolingual infants can detect a change in face-voice pairings when the language used is their native language (Fecher & Johnson, 2022), but not when the speakers used a foreign language (Fecher & Johnson, 2019a, 2022).

Language has also been demonstrated to influence speaker face perception. In Boisferon et al. (2021), 9-to-12-month-old monolingual infants were familiarized with an own-race static female face while a voice recording was playing either in the infant's native language or a language unfamiliar to the infant. In the test phase, infants were presented with the face of the female from the familiarization next to the face of a new own-race female. Infants only exhibited face recognition when the face from the familiarization had been paired with their native language, but not when paired with the foreign language. Interestingly, a similar effect was found for other-race faces, but this time face recognition only took place when the other-race face had been paired with a language non-native to the infant (Clerc et al., 2022).

Besides influencing speaker perception, languages themselves can be considered an important feature of a speaker's identity, and be associated to certain physical attributes, like those related to race. Six- and 11-month-old monolingual infants associate certain languages to certain physical appearances, specifically unfamiliar languages with other-race faces (May et al., 2019; Uttley et al., 2013). These findings indicate the potential for languages to be associated with the specific appearance of speakers.

Kinzler et al. (2007) explored for the first time the influence of language on the social perception of specific speakers. Five-to-6-month-old monolingual infants were familiarized with videos of two female speakers, one using the infants' native language (i.e., English) and the other a language unfamiliar to them (i.e., Spanish). Next, infants were tested in a preferential-looking paradigm, with side-by-side static images of the speakers in silence. Infants exhibited a visual preference for the face of the speaker who had previously used the infant's native language compared to speakers who used an unfamiliar language. A recent study by Colomer et al. (2023) replicated those findings by testing a group of 3-to-6-month-old monolingual infants, suggesting that this native preference might be present even earlier. Moreover, the authors also tested 8- to 11-month-old monolingual infants in the same paradigm and found that the preference for the native language speaker was no longer present at these ages.

Although the objective of both studies was to explore how language modulates speakers' social perception, their findings also evidence that early on infants can track the languages used by different speakers. To express these language-based visual preferences, infants must have formed at least primary associations between the auditory aspects of the language and the face of the speaker. However, these studies did not

address the robustness of the association as they did not test the retention of the face-language pairings.

Recent research has further investigated the formation of speaker-language associations using a familiarization-switch procedure in infants aged 5, 12, and 18 months (Schott et al., 2023). The task included two conditions: an auditory-only condition, in which participants heard the speakers without visual input, and an audiovisual condition, where participants both saw and heard the speakers. In both conditions the procedure was similar. Infants were first familiarized with voice recordings or videos of two speakers, one using English and the other French. After familiarization, infants were presented with voice recordings or videos where the speakers used the opposite languages. Infants' attention to the stimuli after the language switch was measured to determine whether they had detected the change in the pairings, indicating they had formed speaker-language associations. Monolingual infants showed no signs of detecting the language switch, regardless of age or condition, providing no evidence of the formation of associations between speakers and the language they used.

## **Integration of Linguistic and Indexical Information in Bilingual Infants**

As previously noted, differences between monolingual and bilingual infants in processing indexical information have only been evaluated in tasks where linguistic information was also involved. For example, Fecher & Johnson (2022) observed that 9-month-old bilingual infants, as well as same-aged monolingual infants, were able to detect a change in face-voice pairings when the language used in the task was their native

language. However, only bilingual infants succeed when tested in a foreign language (Fecher & Johnson, 2019a, 2022).

During early bilingual language acquisition, it has been proposed that associating certain visual characteristics of speakers (i.e., features related to race) with the auditory characteristics of the language they use might be a useful strategy to further support language separation (Kandhadai et al., 2014).

Despite the potential benefits of these associations, only one study has assessed the formation of speaker-language associations in bilingual infants (Schott et al., 2023). As previously introduced, this study consisted of a familiarization-switch paradigm, where infants' detection of changes in speaker-language pairings was assessed. Similar to what was observed in monolingual infants, 5-, 12-, and 18-month-old bilingual infants provided no evidence of the formation of associations between speakers and the language they used, as they showed no sign of detecting a switch in the speaker-language pairings.

It should be considered that the null results could be partially attributed to the use of a familiarization-switch paradigm. This procedure relied on infants' reaction to a speaker switching from one language to another, which may have not been the most suitable method to assess this ability in the tested sample. Given that infants from bilingual communities might be used to speakers changing languages, they might not react in front of this situation, regardless of whether they had formed speaker-language associations and detected the change.

Parallel to this idea, infants' exposure to code-switching at home might influence the development of these associations. Originally, some authors suggested that the best strategy to raise infants bilingually was to follow

the one parent one language strategy, where each caregiver used only one of the languages with the child (Barron-Hauwaert, 2004; Grammont, 1902). This approach considered that associating one language to each speaker could prevent confusion and language mixing. Posterior research demonstrated that this strategy was not essential and that other approaches lead to a successful bilingual acquisition too (De Houwer, 2007). However, exposure to high levels of parental code-switching might hinder the development of speaker-language associations, as in these contexts language might not be perceived as a stable feature of a speaker. In contrast, for bilingual infants whose caregivers stick to the use of only one language, forming speaker-language associations might be a useful tool to navigate their sociolinguistic context. Although Schott et al. (2023) collected data on parental code-switching, they did not include this variable in their analysis.

# Development of Language

## Awareness

The previous section focused on the development of speech perception, which is essential to process the audiovisual characteristics of the native language and form a mental representation including its perceptual regularities. As infants gain experience with the language, these representations become richer and more complex including a wide variety of information, beyond how the language sounds like.

Language awareness has been defined as “the development in learners of an enhanced consciousness of and sensitivity to the forms and functions of language” (Carter, 2003, p. 64). In other words, it refers to an increased knowledge of a language structure and linguistic rules, but also of the use of languages in different contexts. Language awareness develops throughout language acquisition and can be reflected in a wide range of behaviors expressed by children. This expression can be implicit if the knowledge is used unconsciously, or explicit if the knowledge can be consciously accessed and verbalized (Hulstijn, 2005).

Generally, infants are first able to express their language awareness implicitly before they can verbally reflect on their knowledge. A classic example is Berko's (1958) “wug test”, where children were prompted to provide the plural form of new words. Four- and 5-year-old English-speaking children demonstrated implicit knowledge of the pluralization rule in their language by adding the -s suffix to the novel words, even if they were not able to verbally articulate the rule yet. Similarly, 4-year-old children use shorter sentences when interacting with younger children

compared to when they speak with same-aged children or adults (Shatz & Gelman, 1973), indicating implicit awareness of the limited linguistic abilities of younger children.

## **Development of Language Awareness in Bilingual Children**

Language awareness can be especially important in bilingual contexts. Bilingual children must keep their languages separated in two different systems, keeping track of the linguistic rules that belong to each one, while learning when to use one language or the other based on the context or the speakers they encounter.

Initially, researchers believed that bilingual children had one unitary linguistic system that included characteristics from both languages. They argued that the presence of language mixing in children's early speech evidenced their confusion and lack of language differentiation (Volterra & Taeschner, 1978). According to them, infants were only able to form separate representations for each language after gaining sufficient experience with both (Redlinger & Park, 1980). Nowadays, there is a large body of evidence supporting the formation of two separate systems from early on (see Werker & Byers-Heinlein, 2008 for a review), for example, the acquisition of translation equivalents.

During the word-learning process, infants assign labels to the referents they encounter. Generally, this process takes place following the strategy of mutual exclusivity (Markman & Wachtel, 1988), where infants tend to assume that each referent has a unique label and avoid giving two names to the same object. As a result, infants are more likely to associate a new

label with a new referent rather than with an object that has already been named. However, the principle of one word per referent conflicts with the bilingual experience of having to learn two different words for the same object, one for each language. Translational equivalents, or cross-linguistic synonyms, are words from different languages that share the same meaning or the same referent. Several studies have revealed that bilingual infants learn translational equivalents from an early age, around the first year of life (David & Wei, 2008; Legacy et al., 2017; Nicoladis, 1998; Nicoladis & Genesee, 1996). In fact, bilingual infants do not show signs of avoidance when learning a second label corresponding to another language (Tsui et al., 2022). This might be potentially facilitated by their lower likelihood of performing mutual exclusivity compared with their monolingual peers (Byers-Heinlein & Werker, 2009; Kandhadai et al., 2017). Furthermore, Byers-Heinlein et al. (2017) demonstrated that translational equivalents were indeed stored in separate linguistic systems according to the language they belong to. The study showed that 20-month-old bilingual infants were better at recognizing target words when those were presented in a sentence in the same language, but their accuracy decreased when there was a language switch. If translational equivalents were stored in a single shared linguistic system, infants would recognize the target words equally regardless of the language of the sentence.

Bilingual children are generally exposed to complex sociolinguistic environments where one language or the other might be used depending on the specific context or interlocutor. For instance, one language may be spoken primarily at home, while the other is predominantly used outside, or both languages might be spoken within the family but by different members. The acquisition of sociolinguistic awareness provides children

with a greater understanding of how languages are embedded in society and how they might relate to identity and cultural aspects. Previous research has assessed sociolinguistic awareness, particularly regarding the language used by different speakers, by relying on either implicit or explicit measures: pragmatic differentiation and language labeling, respectively.

In addition to forming and maintaining separate linguistic systems for each language, bilingual children must constantly decide which language to use based on the speaker they are interacting with. Pragmatic differentiation refers to the adaptation of the language used according to the context or the interlocutor. It indicates implicit knowledge of what is the language predominantly used by different speakers. This ability is not only important as a marker of sociolinguistic awareness but is also essential for maximizing children's communicative efficacy and improving social interactions.

Research on pragmatic differentiation has primarily examined children's language use with their caregivers during free play sessions, either at home or in the lab. Most studies have focused on children from families using the one-parent-one-language strategy, as this provides an ideal context for assessing children's ability to adapt their language use with familiar speakers while requiring them to use both languages. For instance, Lanza (1992) explored pragmatic differentiation in a 2-year-old bilingual child learning English (from the mother) and Norwegian (from the father) at home. Results suggest that by age 2 children can use their languages in context-sensitive ways, as the child used English predominantly with her mother and Norwegian with her father. Similar results were obtained in a subsequent study assessing 5 French-English bilingual children whose ages ranged from 1;10 to 2;2 (Genesee et al.,

1995). Children used more of each language with the corresponding caregiver, even when both parents were interacting together with the child. This was a demanding situation where children had to constantly switch languages depending on the parent they were addressing. A longitudinal study of 4 French-English bilingual children found that pragmatic differentiation began to emerge between the ages of 1;9 and 2;4, when the children started to use more of each caregiver's language than would be expected by their language dominance (Nicoladis & Genesee, 1996). A subsequent study by the same author found evidence of pragmatic differentiation even earlier, at age 1;5, when assessing an English-Portuguese bilingual child (Nicoladis, 1998). Research on trilingual children revealed a similar age of emergence, with pragmatic differentiation being established around age 2 or even before (Montanari, 2008; Quay, 2008). Other studies focused on children's ability to adapt their language use to new speakers. In line with previous results, around age 2 bilingual children exhibited pragmatic differentiation when interacting with unfamiliar speakers (Genesee et al., 1995, 1996).

Despite children's clear ability to adapt their language to their caregivers and new speakers by the time they approach their second birthday, children in all the studies still exhibited some degree of language mixing. Genesee (1989) proposed some factors that might modulate pragmatic differentiation and account for the language mixing presented, such as the child's language dominance and their exposure to language mixing.

Regarding language dominance, several studies noted less language adaptation and more language mixing when the child interacted with a caregiver or a new speaker who used their non-dominant language (Genesee et al., 1995, 1996; Lanza, 1992). Nicoladis & Secco (2000) conducted a longitudinal study on a Portuguese-English bilingual child

from age 1;0 to age 1;6 further describing this pattern. They found that the child's language mixing could be largely attributed to limited lexical knowledge, with nearly 90% of the mixing resulting from lexical gaps. This was more commonly observed when the child was required to use their non-dominant language, where fewer translational equivalents were available (see also Montanari, 2008). As suggested in Nicoladis & Genesee (1996), language mixing could be a strategy used by bilingual children "to extend their limited communicative abilities in their less proficient language", rather than a sign of language confusion.

Exposure to language mixing has also been suggested to influence pragmatic differentiation (Genesee, 1989). Goodz (1989) observed that the children's frequency of language mixing correlated with the parental rates of mixing. However, subsequent research failed to replicate these results (Genesee et al., 1995, 1996), likely because all caregivers showed very low levels of language mixing. As previously mentioned, most studies on pragmatic differentiation have focused on children from families following the one-parent-one-language strategy. This approach is characterized by low levels of language mixing, as each caregiver predominantly uses only one of the languages with the child. Examining how pragmatic differentiation develops in other family language strategies would offer a more comprehensive exploration of the effects of different levels of language mixing while also providing a better understanding of how children adapt their language use in different contexts.

Despite the valuable contributions, the downside of most pragmatic differentiation studies is their small samples, ranging from case studies to studies with 5 participants. Moreover, children's ability to adapt their language use varied significantly both across and within studies, indicating significant individual differences in its development. This makes

it challenging to draw robust and generalizable conclusions about the emergence of pragmatic differentiation and the factors that might influence it. Additionally, less attention has been paid to how this ability develops beyond its initial emergence or how the impact of its predictors may vary over time. Interestingly, a study conducted with bilingual children aged 3 and 4 suggests that by age 3, children consistently use the appropriate language, with no significant age-related improvements observed (Tare & Gelman, 2010).

Another way to evaluate bilingual children's sociolinguistic awareness is by assessing their ability to verbally express which languages are used by different speakers in their environment. This ability has been experimentally evaluated by directly asking children what languages were used by their caregivers and themselves (Akhtar et al., 2012; Atagi & Sandhofer, 2020). Akhtar et al. (2012) tested 3- and 4-year-old children with and without exposure to a second language. Participants were tested in a word learning procedure where new words were presented by a speaker that used their native language (i.e., English) or a speaker using an invented language (i.e., "Nordish"). After the word learning task, children were asked to indicate which speaker used their native language and what languages the child was able to speak. Nearly 70% of the participants successfully identified the speaker who had previously used their language, regardless of exposure to a second language. However, when children were asked which language(s) they spoke, linguistic background predicted children's ability to label their language(s). Over 50% of the children exposed to a second language were able to label at least one of the languages they spoke, while only 19% of the non-exposed participants were able to answer the question. The researchers attributed this advantage to an increased frequency of linguistic conversations in homes

where two languages were spoken, compared to homes where only one language was present.

In a subsequent study, Atagi & Sandhofer (2020) tested language labeling in 3-to-5-year-old children. In this case, all participants were functional monolinguals but had different levels of exposure to a second language: no exposure, exposure in the community, or exposure at home. The authors asked the children to label their own language, as well as the languages used by their caregivers. Regardless of exposure, children were more likely to label their caregivers' languages rather than their own. Their results also revealed that children exposed to a second language at home were better at language labeling compared to those with no exposure or exposure in the community. As in the previous study, the authors attributed the differences in performance to a higher frequency of language reflections at homes where two languages are present. In fact, 40% of caregivers of the children exposed to a second language at home reported engaging in this type of conversation, compared to 0% and 9% of the caregivers of non-exposed and exposed in the community children, respectively. However, no direct comparison was conducted between exposure to language reflections and the child's language labeling performance.

Chapter 2:

# **RESEARCH AIMS AND QUESTIONS**



# Research Aims

The findings reviewed in the introduction highlight the importance of keeping track of the language used by the speakers in the environment, specifically for infants and children simultaneously acquiring two languages. During the early stages of dual language acquisition, associating a language to each speaker might facilitate language separation. In later stages, bilingual children might use these associations to select what is the appropriate language to use with different speakers. In addition, this sociolinguistic knowledge will enable them to reflect on who uses each language and how they are integrated into society. Despite the importance of speaker-language associations, the developmental trajectory of this ability is yet to be well described.

This thesis aims to explore the formation of speaker-language associations across two developmental stages: before the first year of life (infants aged 4 to 12 months) and afterward (children aged 1 to 5 years). Additionally, it seeks to assess how different linguistic factors, including bilingualism and other language-related experiences, shape and modulate the developmental trajectory of these associations.

# Research Questions

## Regarding infants **before** the first year of life:

1. Does the ability to form face-language associations emerge within the first year of life?
  - a. Does infants' linguistic background or exposure to parental code-switching influence the development of speaker-language associations?

## Regarding children **after** the first year of life:

2. What is the general developmental trajectory of pragmatic differentiation?
  - a. How do children's language dominance and the family language strategy influence pragmatic differentiation over development?
3. What is the general developmental trajectory of language labeling?
  - a. Does exposure to language reflections at home influence the development of language labeling?

Research question 1 will be addressed in Study 1 and Study 2. These studies focus on the formation of both primary and robust audiovisual face-language associations before the first year of life by testing 4-, 6-, and 10-month-old monolingual and bilingual infants in an audiovisual preferential-looking task. In Study 1 participants are exposed to two speakers using close languages (Catalan and Spanish), and in Study 2, to two speakers using distant languages (Catalan/Spanish and English). Besides linguistic background, the role of exposure to parental code-switching will be also examined as a potential modulator. The prediction is

that infants will be able to form primary face-language associations from early on, but that more robust associations will not appear until the end of the first year of life. It is also expected for bilingual infants to outperform monolingual infants, at least regarding the formation of robust face-language associations, as they are more frequently exposed to different speakers using different languages and might benefit more from this type of association. Moreover, bilingual infants with little to no exposure to parental code-switching are expected to outperform those with higher exposure, as they are more likely to perceive language as a stable trait linked to a specific speaker.

Research questions 2 and 3 will be addressed in Study 3. This study focuses on the development of sociolinguistic awareness regarding the languages used by different speakers after the first year of life by collecting parental report measures of pragmatic differentiation and language labeling in 1- to 5-year-old bilingual children from two bilingual communities (i.e., Quebec and Catalonia). Furthermore, different language-related factors will be evaluated as predictors: children's language dominance and family language strategy for pragmatic differentiation, and exposure to language reflections at home for language labeling. Based on previous research, the prediction is that pragmatic differentiation will emerge around age 2 and that children will be better at adapting their language use to their interlocutor when they are required to use their dominant language. Moreover, those language strategies that expose children to less language mixing at home, such as the one-parent-one-language, are expected to facilitate the development of pragmatic differentiation. Regarding language labeling, the prediction is that it will emerge around age 3, with a potential early emergence in those children exposed to high levels of language reflections at home.



Chapter 3:

**SPEAKER-LANGUAGE  
ASSOCIATIONS  
BEFORE THE FIRST  
YEAR OF LIFE**



# Study 1:

## Formation of Audiovisual Face-Language Associations for Speakers Using Close Languages (Catalan and Spanish)

### Introduction

Numerous studies have explored the development of speaker recognition by examining infants' ability to associate the specific auditory characteristics of a speaker's voice with the distinct visual features of their face (Bahrick et al., 2005; Brookes et al., 2001; Fecher et al., 2019; Orena et al., 2022). Although some researchers have explored the effect of language on speaker perception (Boisferon et al., 2021; Clerc et al., 2022; May et al., 2019; Uttley et al., 2013), much less attention has been paid to infants' ability to form associations between the specific physical characteristics of speakers and the language they use. Associating a speaker with a language may be particularly beneficial for bilingual infants during their dual language acquisition process, as they may use visual cues to separate their languages promoting the formation of two distinct linguistic systems (Kandhadai et al., 2014).

The first evidence of these associations can be indirectly extracted from a seminal study by Kinzler et al. (2007). Using a preferential-looking paradigm, the researchers found that 5-to-6-month-old monolingual infants showed a visual preference for speakers that had previously used their native language. More recently, Colomer et al. (2023) replicated and

### Chapter 3: Speaker-Language Associations Before the First Year of Life

extended these findings by testing a group of 3-to-6-month-old monolingual infants, suggesting that this preference might be present even earlier. Although in both studies researchers were mainly interested in how languages guide social interactions early on, the preferences found indicate that infants were able to track the languages used by the speakers. They must have formed at least primary face-language associations to express a visual preference for speakers who had previously used their native language. Despite their significant contribution, these studies do not address the robustness of the language-face associations or infants' ability to retain them. Furthermore, how these language-based preferences, and the underlying associations, are expressed in bilingual infants remains unexplored.

Subsequent research directly explored the formation of face-language associations by testing 5-, 12-, and 18-month-old English and/or French monolingual and bilingual infants in a familiarization-switch procedure (Schott et al., 2023). After infants were familiarized with two speakers, each one using one language, researchers assessed face-language associations by testing whether infants detected a change when the speakers used the opposite language. Regardless of age or linguistic background, they showed no sign of change detection, providing no evidence of the formation of associations between speakers and the language they used. One possibility for these inconclusive results is that infants growing up in bilingual communities are used to speakers changing from one language to another, so they did not react to the change even though they might have formed face-language associations.

Furthermore, it is possible that not all types of bilingual experience equally promote the formation of these associations. For example, bilingual infants learning close languages, such as Spanish and Catalan, may

### Chapter 3: Speaker-Language Associations Before the First Year of Life

benefit more from associating a language to each speaker, as the similarities between the languages could make it harder for them to keep them separate. Previous research suggests that bilingual infants learning close languages may use extra visual cues to help language separation (Birulés et al., 2019). Exposure to code-switching could also influence the development of face-language associations. Infants exposed to high levels of code-switching are used to see speakers changing from one language to another. Consequently, they may not perceive languages as a stable trait of a speaker, which would prevent these infants from forming these associations.

This study explored the formation of both primary and more robust face-language associations in monolingual and bilingual infants during the first year of life, as well as potential linguistic factors that might modulate the development of this ability.

## Methods

### Participants

The final sample for this study consisted of 156 infants aged 4, 6, or 10 months from the metropolitan area of Barcelona (Catalonia, Spain). All infants were full-term babies ( $\geq 37$  weeks) and were born with normal birth weight ( $\geq 2500$ gr). The caregivers reported no developmental delays or hearing or vision problems.

The Language Exposure Assessment Tool (LEAT) (DeAnda et al., 2016) was used to measure infant's linguistic exposure. All infants were learning Catalan and/or Spanish and were classified as monolingual or bilingual based on their language exposure to each language, following previous recommendations (Bosch, 2024; Byers-Heinlein, 2015). Infants were considered monolingual if they had been exposed predominantly to only one of the languages and had had less than 75% exposure to the other. Infants were considered bilingual if their relative exposure to the languages ranged between 75-25% and 50–50. Infants with more than 10% exposure to a language other than Catalan or Spanish were excluded from the analyses.

For the bilingual participants, measures of their exposure to parental code-switching were also collected using the Language Mixing Questionnaire (Byers-Heinlein, 2013). Scores on the questionnaire range from 0 (no code-switching at all) to 30 (high levels of code-switching) and were obtained for each caregiver.

Fifty-six 4-month-old infants were tested (age range = 3.4–4.6 months, mean age = 4.2 months, SD = 0.3), 32 of whom were monolingual (mean L1 = 94%, SD = 7.0, 12 Catalan dominant) and 24 were bilingual (mean L1 = 61.8%, SD = 8.3, 10 Catalan dominant). Fifty 6-month-old infants were

### Chapter 3: Speaker-Language Associations Before the First Year of Life

also tested (age range = 5.5–7.8 months, mean age = 6.5 months, SD = 0.5), 27 of whom were monolingual (mean L1 = 96.7%, SD = 6.2, 8 Catalan dominant) and 23 were bilingual (mean L1 = 63.5%, SD = 8.5, 10 Catalan dominant). Lastly, 50 10-month-old infants were tested (age range = 8.7–11.9 months, mean age = 10.3 months, SD = 0.9), 27 of whom were monolingual (mean L1 = 89%, SD = 8.4, 17 Catalan dominant) and 23 were bilingual (mean L1 = 59%, SD = 7.5, 12 Catalan dominant).

Sixty-two additional infants were tested but excluded from the analyses due to preterm birth ( $n = 9$ ), fussiness or excessive crying ( $n = 27$ ), exposure to a language other than Spanish or Catalan ( $n = 12$ ), or because they did not provide a minimum of 20% of data in the two language test trials ( $n = 14$ , following Birulés et al. (2019) and Frank et al. (2012)).

#### Stimuli

The stimuli for the familiarization phase consisted of 10-s videos of a female uttering a monologue using infant-directed speech (IDS). In half of the videos, a native Spanish speaker was recorded, and in the other half, a native speaker of Catalan. Both females were Caucasian with dark hair and similar hairstyles.

The silent test trial consisted of a single 8-s trial displaying side-by-side static images of the speakers that appeared in the familiarization in silence.

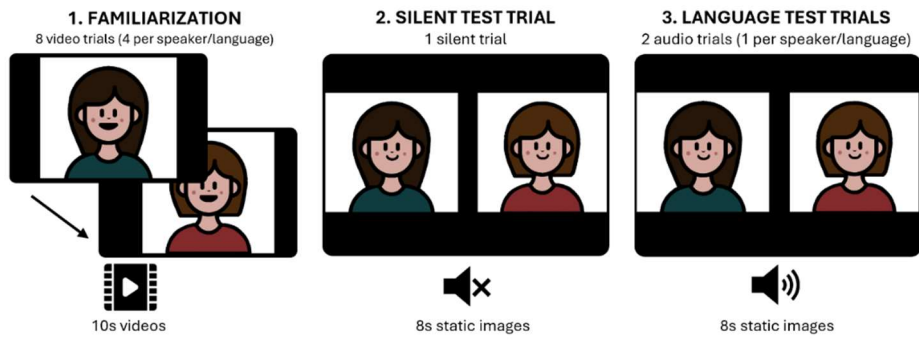
The language test trials consisted of two 8-s trials presenting the same static images but with a voice recording playing simultaneously in the background. These audio clips were extracted from the familiarization stimuli, one trial in Catalan and the other in Spanish.

## Procedure

Infants were tested in a sound-attenuated and dimly lit room, seated in an infant seat approximately 60 cm in front of a 17" computer monitor. The Tobii X120 standalone eye tracker was used at a sampling rate of 60 Hz to record infants' eye movements during the experimental task. The stimuli were presented on the computer monitor using the Tobii Studio software. Before starting the task, each participant's gaze was calibrated with the Tobii eye tracker's five-point calibration routine.

Once the calibration was completed, the familiarization phase started. Infants were exposed to eight 10-s videos, half of them showed the Catalan speaker, and the other half the Spanish speaker. The videos were presented alternating the languages, and the language that was presented first was counterbalanced across participants. After familiarization, infants were presented with the test phase. Infants first saw the silent test trial where they watched side-by-side pictures of the speakers in silence for 8 seconds. Right after, they were presented with two language test trials. Infants watched 8-s trials with the same pictures while a voice recording played in the background, one trial in Spanish and the other in Catalan. The order of the trials was counterbalanced across participants. The side in which the speakers appeared was consistent during the test phase for each participant but counterbalanced across participants (see **Figure 1** for a visual representation of the experimental task). During test trials, the eye-tracker monitored the infants' gaze directed at three areas of interest (AOI), one for the face of each speaker and one for the entire screen.

## Chapter 3: Speaker-Language Associations Before the First Year of Life



**Figure 1.** Visual representation of the experimental task. The images included in the figure are illustrative and do not correspond to the actual stimuli used in the task.

## Results

### Silent test trial

To evaluate the formation of primary face-language associations, infants' visual preferences for the speakers were analyzed during the silent test trial, as in Colomer et al. (2023) and Kinzler et al. (2007). Before the analysis, the proportion of total looking time (PTLT) to the speakers was computed by dividing the time infants spent looking at the face of each speaker by the time they spent looking at the screen. Each speaker was labeled according to infants' language exposure to the language they used during the familiarization. For monolingual infants, the speakers were labeled as "native language speaker" or "non-native language speaker", while for bilingual infants, for whom both languages were native, the speakers were labeled as "dominant-language speaker" or "non-dominant-language speaker" depending on whether the speakers used the language infants were most or least exposed to.

For the analysis, a mixed-effects analysis of variance (ANOVA) was performed including PTLT as the dependent variable, Speaker (Native/Dominant language vs. Non-native/Non-dominant language) as a within-subjects variable, and Linguistic Background (Monolingual vs. Bilingual) and Age (4, 6, and 10 months) as between-subjects variables.

Results from the ANOVA revealed that none of the main effects or interactions were significant. Most of them fell far from significance ( $ps > .1$ ) except for a trend toward a significant Speaker main effect ( $F(1, 150) = 3.11, p = .08$ ) and a marginally significant Speaker  $\times$  Age interaction ( $F(2, 150) = 2.72, p = .07$ ).

According to theoretical expectations and the marginally significant Speaker  $\times$  Age interaction, the PTLT to the speakers was compared for

each age group separately. Four-month-old infants looked preferentially at the speaker of their native/dominant language ( $M = 0.55$ ,  $SD = 0.28$ ) over the speaker of their non-native/non-dominant language ( $M = 0.39$ ,  $SD = 0.29$ ;  $t(55) = 2.05$ ,  $p < .05$ ). Six- and ten-month-old infants showed no preference for any of the speakers ( $t(49) = 1.37$ ,  $p = 0.18$  and  $t(49) = -0.97$ ,  $p = .34$ , respectively).

### **Silent test trial: exploring differences throughout the trial**

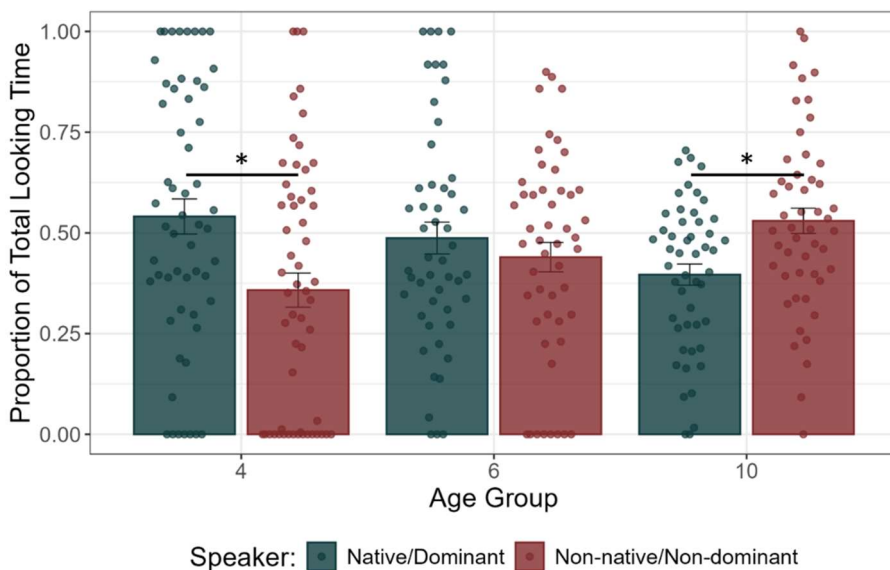
The duration of the silent test trial was selected to match the duration of the subsequent language test trials to maintain consistency across the task. However, in the language test trials infants need to process the faces of the speakers, the language used in the recording, and then show a preference for the speaker that previously used that language during familiarization, while in the silent test trial there is no auditory stimuli to process, simplifying the task. Accordingly, infants may require less time to visually process the two faces and show a preference in the silent test trial compared to the language test trials. In fact, recent research used shorter trials (i.e. 6-s) when measuring this type of visual preferences (Colomer et al., 2023). To explore if infants' visual behavior differed throughout the duration of the silent test trial, the previous analysis was repeated for each half trial separately.

### **Silent test trial: First half**

In the first half of the silent test trial, the ANOVA revealed a significant Speaker  $\times$  Age interaction ( $F(2,150) = 4.78$ ,  $p < .01$ ), while none of the other main effects or interactions approached significance (all  $ps > .1$ ).

To further explore the interaction, the PTLT to each speaker was compared for each age group separately. Four-month-old infants looked preferentially to the speaker of native/dominant language ( $M = 0.54$ ,  $SD =$

0.33), compared to the speaker of their non-native/non-dominant language ( $M = 0.36$ ,  $SD = 0.32$ ;  $t(55) = 2.18$ ,  $p < .05$ ). Six-month-old infants looked equally at the native/dominant ( $M = 0.49$ ,  $SD = 0.28$ ) and the non-native/non-dominant language speaker ( $M = 0.44$ ,  $SD = 0.26$ ;  $t(49) = 0.63$ ,  $p = .53$ ). Ten-month-old infants looked significantly longer at the speaker of their non-native/non-dominant language ( $M = 0.53$ ,  $SD = 0.22$ ), in comparison to the speaker of their native/dominant language ( $M = 0.40$ ,  $SD = 0.19$ ;  $t(49) = -2.40$ ,  $p < .05$ ). See **Figure 2**.



**Figure 2.** PTLT to the speaker of the native/dominant and the non-native/non-dominant language during the first half of the silent test trial in 4-, 6-, and 10-month-old infants. Dots represent each individual's PTLT values and error bars the standard error (SE) of the group mean.

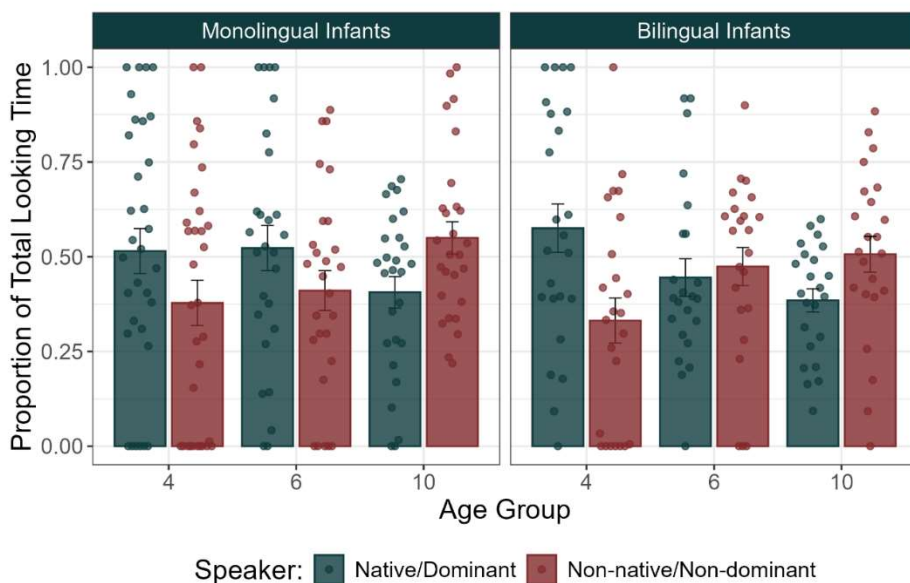
### Silent test trial: Second half

For the second half of the silent test trial, the same mixed-effects ANOVA was performed as in the analyses for the whole trial and for the first half. The ANOVA did not reveal any significant main effect or interaction (all  $ps > .1$ ).

### Silent test trial: exploring the bilingual experience

As the language-based preferences were mainly found during the first half of the silent test trial, this data was selected to explore the potential effect of different linguistic experiences on these preferences.

The lack of a significant Linguistic Background main effect or interaction in the ANOVA for the first half of the silent trial suggests that the results obtained were consistent between monolingual and bilingual infants. Accordingly, both linguistic backgrounds showed a very similar developmental pattern, transitioning from a preference for the native/dominant language speaker in younger infants to a preference for the non-native/non-dominant language speaker in older infants. Anecdotally, the preferences expressed by the two groups appear to differ slightly at 6 months of age. While for bilingual infants the native/dominant language speaker preference was no longer observable at that age, same-aged monolinguals seemed to still retain a slight preference for that speaker (see **Figure 3**).



**Figure 3.** PTLT to the speaker of the native/dominant and the non-native/non-dominant language during the first half of the silent test trial in 4-, 6-, and 10-month-old monolingual and bilingual infants. Dots represent each individual's PTLT values and error bars the SE of the group mean.

Next, the influence of exposure to Parental Code-switching was assessed. Parental Code-switching was computed by averaging the Language Mixing Questionnaire scores of both caregivers. As the obtained scores deviated towards the lower end ( $M = 10.31$ ,  $SD = 8.2$ ), the sample was divided performing a median split, providing two groups: those with Low Code-switching Exposure (scores  $<8.5$ ) and those with High Code-switching Exposure (scores  $>8.5$ ).

A mixed-effects ANOVA was performed only for the bilingual participants, including PTLT as the dependent variable, Speaker (Native/Dominant language vs. Non-native/Non-dominant language) as a within-subjects variable, and Parental Code-switching (Low Code-switching Exposure vs. High Code-switching Exposure) as a between-subjects variable.

The ANOVA revealed a marginally significant Speaker  $\times$  Age interaction ( $F(2,64) = 2.86, p = .06$ ), but none of the other main effects or interactions approached significance (all  $ps > .1$ ).

As the Speaker  $\times$  Age interaction in bilingual infants had already been described (see **Figure 3**), no further analyses were performed.

### Language test trials

To evaluate the formation of robust face-language associations, infants' visual preferences for the speakers were analyzed during the language test trials. In these trials, if infants had associated each speaker with the language they used, they should look preferentially at the corresponding speaker when that language was playing in the background. As before, the PTLT to each speaker was computed. In the language test trials, the speakers were labeled as "match speaker" and "mismatch speaker," depending on whether the voice recording matched the language they had used in the familiarization. Furthermore, the trials were labeled as "native language" or "non-native language" for monolingual infants, and "dominant trial" or "non-dominant trial" for bilingual infants, according to the language exposure of each infant.

For the analysis, a mixed-effects ANOVA was performed including PTLT as the dependent variable, Speaker (Match vs. Mismatch) and Trial Language (Native language vs. Non-native language) as within-subjects variables, and Linguistic Background (Monolingual vs. Bilingual) and Age (4, 6, and 10 months) as between-subjects variables.

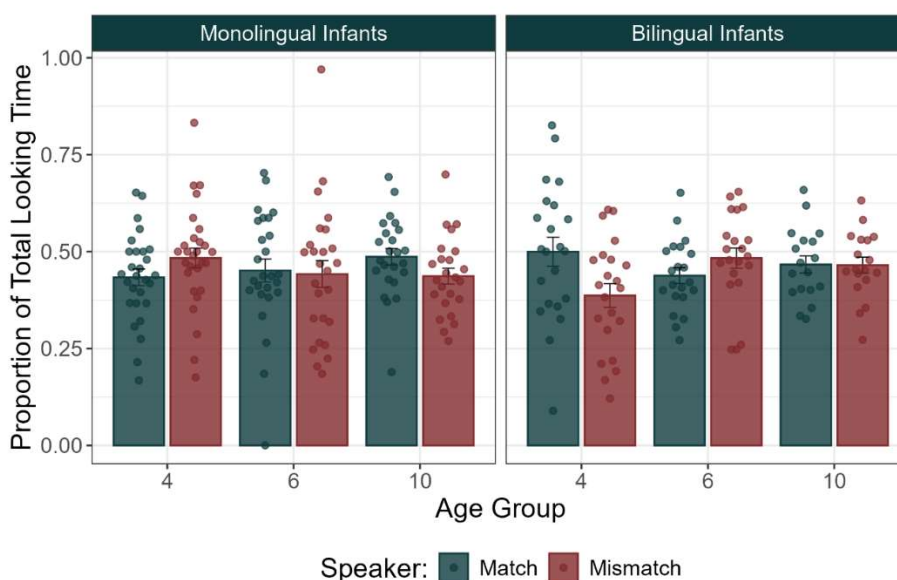
The ANOVA revealed a significant Speaker  $\times$  Linguistic Background  $\times$  Age triple interaction ( $F(2,133) = 3.24, p < 0.05$ ). None of the other main effects or interactions were significant (all  $ps > 0.1$ ). The results of the significant interaction are developed in the next section.

### Language test trials: exploring the bilingual experience

The potential effect of different linguistic experiences on the formation of robust face-language associations was explored with the two language test trials combined.

The previously mentioned Speaker x Linguistic Background x Age interaction was assessed by comparing the PTLT to the speakers between the linguistic backgrounds at each age separately. Four-month-old bilingual infants showed a marginally significant preference for the matching speaker ( $M = 0.50$ ,  $SD = 0.18$ ) compared to the mismatching speaker ( $M = 0.39$ ,  $SD = 0.14$ ;  $t(21) = 1.89$ ,  $p = .07$ ). None of the other groups showed any preference that approached significance (all  $ps > .1$ ). See

**Figure 4.**



**Figure 4.** PTLT to the match and mismatch speaker in the language test trials combined, in 4-, 6-, and 10-month-old monolingual and bilingual infants. Dots represent each individual's PTLT values and error bars the SE of the group mean.

### Chapter 3: Speaker-Language Associations Before the First Year of Life

Next, the influence of exposure to Parental Code-switching was assessed. A mixed-effects ANOVA was performed only for the bilingual participants, including PTLT as the dependent variable, Speaker (Match vs. Mismatch) as a within-subjects variable, and Age (4, 6, and 10 months) and Parental Code-switching (Low Code-switching Exposure vs. High Code-switching Exposure) as between-subjects variables.

The ANOVA revealed a nearly significant Speaker  $\times$  Age interaction ( $F(2,55) = 3.09, p = .053$ ), but none of the other main effects or interactions approached significance (all  $ps > .1$ ).

As the Speaker  $\times$  Age interaction in bilingual infants had already been described (see **Figure 4**), no further analyses were performed.

## Discussion

The main goal of the current study was to explore the formation of both primary and robust face-language associations during the first year of life, as well as potential linguistic factors that might modulate the development of this ability. After a familiarization phase with videos of a Catalan and a Spanish speaker, infants were tested in a preferential-looking paradigm with two types of test trials that had different task demands. In the Silent test trial, primary face-language associations were assessed by measuring infants' language-based preference for the speakers in silence. In the language test trials, robust face-language associations were assessed by measuring infants' ability to match the faces of the speakers with the language they used during familiarization.

In line with previous research (Colomer et al., 2023; Kinzler et al., 2007), results from the silent test trial revealed that during the first year of life infants form primary face-language associations, as they showed visual preferences for speakers depending on the language they previously used.

Interestingly, these language-based preferences followed a developmental pattern transitioning from a familiarity to a novelty preference. Four-month-old infants showed a visual preference for the speaker who used their native/dominant language. By 6 months of age, this preference was no longer present, and infants looked equally at both speakers. Lastly, 10-month-old infants preferred to look at the speaker who used their non-native/non-dominant language. A similar familiarity to novelty preference transition has been previously observed when measuring the visual attention to own- and other-race faces during the first year of life (Fassbender et al., 2016; Liu, Xiao, Xiao, et al., 2015), but it is

### Chapter 3: Speaker-Language Associations Before the First Year of Life

the first time it has been reported regarding languages since Colomer et al. (2023) found no preference in 8-to-11-month-old infants.

It should be noted that the developmental pattern was more robust when considering only the first half of the trial, indicating infants mainly expressed their language-based preferences at the beginning. The preferences were less pronounced during the second half, suggesting that towards the end of the trial, infants shifted to a more exploratory behavior, attending more to the face of the other speaker.

When assessing potential differences across linguistic backgrounds, the analyses revealed that monolingual and bilingual infants exhibited similar patterns regarding language-based preferences, both showing a familiarity to novelty preference transition. Exposure to parental code-switching did not affect these preferences either.

Results from the language test trials do not provide conclusive evidence of the formation of robust face-language associations, regardless of age or linguistic background. The level of exposure to parental code-switching did not appear to influence the outcome either. In general, infants did not show consistent signs of matching the speakers with the language they used during the familiarization. Only 4-month-old bilingual infants tended to look preferentially to the speaker matching the language of the recording, but these results were not replicated at any of the other ages tested. In fact, 6-month-old bilingual infants looked slightly longer at the speaker who did not match the language of the recording. Although the results obtained at 4 months of age could suggest a bilingual advantage in the formation of face-language associations, it is very unlikely that the ability is developed at that age, but it is no longer present in older infants.

### Chapter 3: Speaker-Language Associations Before the First Year of Life

In sum, the above-mentioned results provide evidence for the formation of primary face-language associations during the first year of life, both in monolingual and bilingual infants. However, conclusive evidence for more robust face-language associations was not found, as infants did not show signs of retaining the previously learned face-language pairings regardless of age or linguistic background.

However, it should be considered that the stimuli used in this task may have not been easy to differentiate. Although the results from the silent test trial indicate that infants were able to discriminate the two close and similar languages, the similarities between them might have increased the task's cognitive demands, hindering infants' ability to form more robust face-language associations. Testing two distant languages might make the task easier enhancing the formation of these face-language associations.

# Study 2:

## Formation of Audiovisual Face-Language Associations for Speakers Using Distant Languages (Catalan/Spanish and English)

### Introduction

Findings from Study 1, in line with previous research (Colomer et al., 2023; Kinzler et al., 2007), suggest that infants during the first year of life are able to form primary face-language associations which allows them to express language-based preferences for speakers. However, when measuring more robust face-language associations the results are not as clear, as infants did not show consistent signs of retaining the language used by each speaker, regardless of their age or linguistic background.

As previously stated, the stimuli in Study 1 might have been difficult to discriminate, as Spanish and Catalan are rhythmically close and phonologically similar languages. This could have increased the cognitive demands of the task hindering infants' ability to form and retain robust face-language associations. To rule out this possibility, the previous experimental task was modified in the current study to increase the discriminability of the stimuli. In this new version, infants were presented with speakers using two distant languages: one speaker used Spanish or Catalan (which were native and dominant to the tested infants), and the other used English (which was unfamiliar to all of them).

### Chapter 3: Speaker-Language Associations Before the First Year of Life

This study explored the formation of both primary and more robust face-language associations in a less demanding task with easier-to-discriminate stimuli. As in Study 1, this ability was assessed in both monolingual and bilingual infants during their first year of life, and potential linguistic factors modulating its development were evaluated.

## Methods

### Participants

The final sample consisted of 83 infants aged 4, 6, or 10 months from the metropolitan area of Barcelona (Catalonia, Spain). All infants were born at term ( $\geq 37$  weeks) and with normal weight ( $\geq 2500$  gr), and caregivers reported no developmental delays or hearing or vision problems. None of the infants had prior exposure to English, except for occasional songs or videos. The Language Exposure Assessment Tool (LEAT) (DeAnda et al., 2016) was used to measure infant's linguistic exposure. Infants were classified as monolingual or bilingual following previous recommendations (Bosch, 2024; Byers-Heinlein, 2015). Monolingual participants were predominantly exposed to either Catalan or Spanish, with an exposure higher than 75%. Bilingual infants were also predominantly exposed to Catalan or Spanish, but their exposure ranged from 50% to 75%. For most of them, their second language was also Catalan or Spanish. However, for 6 participants, their second language was different from Catalan or Spanish but belonged to the same rhythmic class (syllable-timed languages): French (1), Polish<sup>1</sup> (1), Italian (3), and Portuguese (1). These participants were all part of the 10-month-old group, but this distribution was entirely coincidental. Measures of exposure to parental code-switching were collected for all the bilingual participants via the Language Mixing Questionnaire (Byers-Heinlein, 2013).

Twenty-seven 4-month-old infants were tested (age range = 4.03–5.5 months, mean age = 4.7 months, SD = 0.4), 13 of whom were monolingual

---

<sup>1</sup> Polish is generally considered to be an intermediate language, with characteristics from syllable-timed languages as well as stress-time languages (Ramus et al., 2003).

## Chapter 3: Speaker-Language Associations Before the First Year of Life

(mean L1 = 91.4%, SD = 7.7, 10 Catalan dominant) and 14 were bilingual (mean L1 = 61.9%, SD = 7.5, 8 Catalan dominant). Twenty-eight 6-month-old infants were also tested (age range = 5.6–7.8 months, mean age = 6.6 months, SD = 0.7), 17 of whom were monolingual (mean L1 = 95.1%, SD = 5.3, 12 Catalan dominant) and 11 were bilingual (mean L1 = 65.1%, SD = 9.1, 8 Catalan dominant). Lastly, 28 10-month-old infants were tested (age range = 8.4–12.1 months, mean age = 9.9 months, SD = 0.9), 13 of whom were monolingual (mean L1 = 96.2%, SD = 5.6, 11 Catalan dominant) and 15 were bilingual (mean L1 = 60.3%, SD = 7.9, 9 Catalan dominant).

Thirty-four additional infants were tested but excluded from the analyses due to preterm birth ( $n = 7$ ), fussiness or excessive crying ( $n = 6$ ), exposure to English ( $n = 13$ ), having an L1 other than Spanish or Catalan ( $n = 2$ ), or because they did not provide a minimum of 20% of data in the two language test trials ( $n = 6$ , following Birulés et al. (2019) and Frank et al. (2012)).

### Stimuli

The stimuli were the same as in Study 1, with the addition of a third female who was a native English speaker.

### Procedure

The procedure was the same as in Study 1. However, infants were presented with either the Catalan condition (Catalan and English speakers) or the Spanish condition (Spanish and English speakers) depending on their language exposure and which was the most predominant language in their environment.

## Results

### Silent test trial

To evaluate the formation of primary face-language associations of easier-to-discriminate stimuli, infants' visual preferences for the speakers were analyzed during the modified silent test trial. First, the PTLT to the speakers was computed by dividing the time infants spent looking at the face of each speaker by the time they spent looking at the screen. The speakers were labeled according to the language they used in the familiarization. The Spanish and the Catalan speakers were labeled as “native language speaker” and the English speaker as “non-native language speaker”.

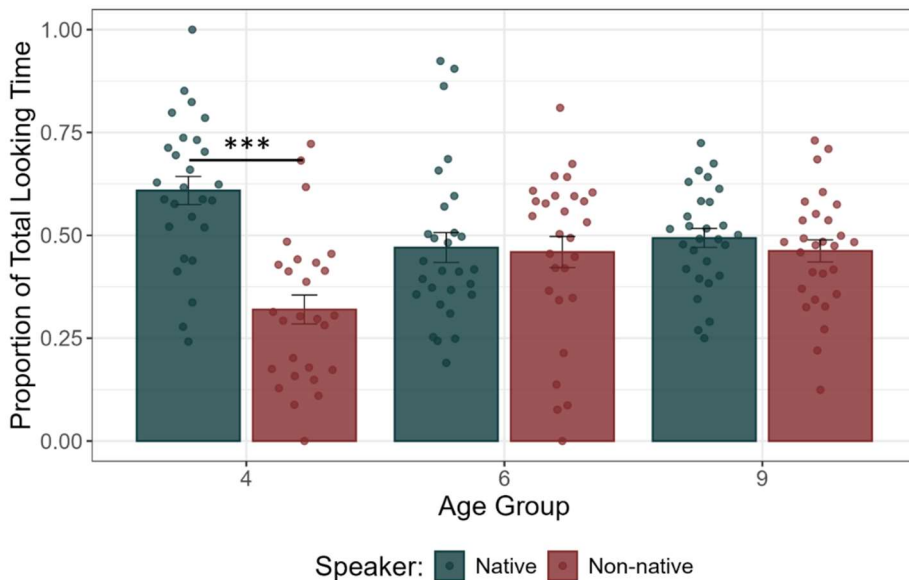
Then, a mixed-effects ANOVA was performed including PTLT as the dependent variable, Speaker (Native language vs. Non-native language) as a within-subjects variable, and Linguistic Background (Monolingual vs. Bilingual) and Age (4, 6, and 10 months) as between-subjects variables.

The ANOVA revealed a significant Speaker main effect ( $F(1, 77) = 8.54, p < .01$ ), as well as a significant Speaker  $\times$  Age interaction ( $F(2, 77) = 5.69, p < .01$ ). None of the other main effects or interactions approached significance (all  $ps > .1$ ).

When collapsing across ages and linguistic backgrounds, infants showed a general preference for the speaker of their native language ( $M = 0.52, SD = 0.18$ ) over the non-native language speaker ( $M = 0.41, SD = 0.19; t(82) = 2.8, p < .01$ ).

To explore the Speaker  $\times$  Age interaction the PTLT to the speakers was compared for each age group separately. Four-month-old infants looked preferentially at the speaker of their native language ( $M = 0.61, SD = 0.18$ ) over the speaker of their non-native language ( $M = 0.32, SD = 0.18; t(26) =$

4.25,  $p < .001$ ). Six- and ten-month-old infants showed no preference for any of the speakers ( $t(27) = 0.15$ ,  $p = .88$  and  $t(27) = 0.64$ ,  $p = .52$ , respectively). See **Figure 5**.



**Figure 5.** PTLT to the speaker of the native and the non-native language during the silent test trial in 4-, 6-, and 10-month-old infants. Dots represent each individual's PTLT values and error bars the SE of the group mean.

### Silent test trial: exploring differences throughout the trial

As in Study 1, the previous analysis was repeated for each half-trial separately to explore whether infants' visual behavior differed throughout the silent test trial.

#### Silent test trial: First half

In the first half of the silent test trial, the ANOVA revealed a significant Speaker main effect ( $F(1,75) = 15.01$ ,  $p < .001$ ) and a significant Speaker  $\times$  Age interaction ( $F(2,75) = 4.42$ ,  $p < .05$ ). None of the other main effects or interactions approached significance (all  $ps > .1$ ).

## Chapter 3: Speaker-Language Associations Before the First Year of Life

The Speaker main effect reflected a general preference for the native language speaker ( $M = 0.54$ ,  $SD = 0.19$ ) over the speaker of the non-native language ( $M = 0.38$ ,  $SD = 0.21$ ;  $t(80) = 3.71$ ,  $p < .001$ ).

To explore the Speaker x Age interaction, the PTLT to each speaker was compared for each age group separately. As in the whole trial, 4-month-old infants showed a preference for the native language speaker ( $M = 0.62$ ,  $SD = 0.21$ ) over the non-native language speaker ( $M = 0.28$ ,  $SD = 0.22$ ;  $t(25) = 4.06$ ,  $p < .001$ ), while 6- and 10-month-old infants showed no preference ( $t(27) = 1.34$ ,  $p = .19$  and  $t(26) = 0.94$ ,  $p = .36$ , respectively).

### Silent test trial: Second half

The same mixed-effects ANOVA was performed for the second half of the silent test trial. The ANOVA revealed a significant Age x Speaker interaction ( $F(2,75) = 4.35$ ,  $p < .05$ ), while none of the other main effects or interactions approached significance (all  $ps > .1$ ).

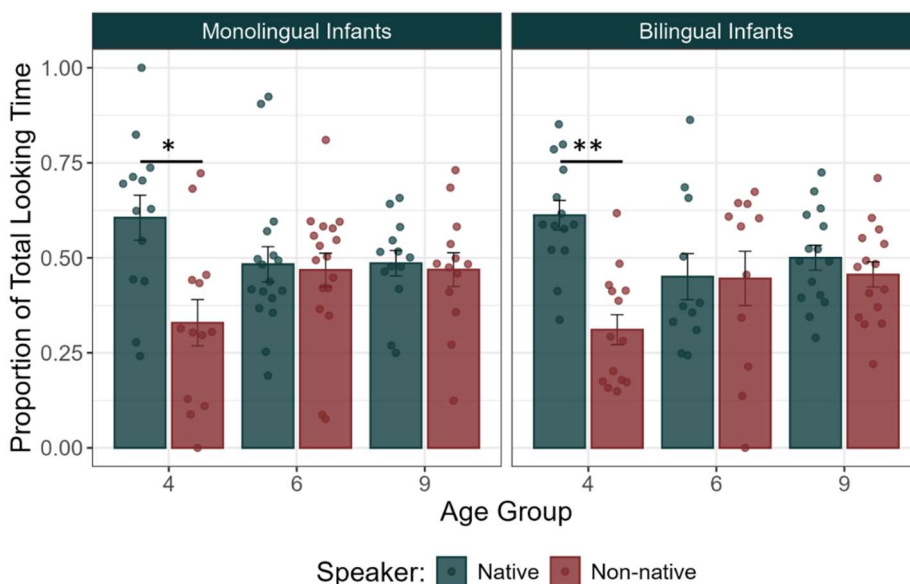
As before, 4-month-old infants looked preferentially at the native ( $M = 0.63$ ,  $SD = 0.27$ ) over the non-native language speaker ( $M = 0.35$ ,  $SD = 0.2$ ;  $t(24) = 2.58$ ,  $p < .05$ ), and 6- and 10-month-old infants showed no preference for any speaker ( $t(27) = -1.36$ ,  $p = .18$  and  $t(27) = 0.63$ ,  $p = .53$ , respectively).

### Silent test trial: exploring the bilingual experience

As the analyses for the whole trial and both halves revealed very similar results, the effect of different linguistic experiences was examined using the data from the whole trial.

The lack of a significant Linguistic Background main effect or interaction in the ANOVA for the whole silent trial suggests that the results obtained were consistent between monolingual and bilingual infants. Accordingly, both linguistic backgrounds showed a very similar developmental pattern,

with a preference for the native language speaker in 4-month-old monolingual ( $t(12) = 2.32, p < .05$ ) and bilingual infants ( $t(13) = 3.99, p < .01$ ), but no preference at later ages (see **Figure 6**).



**Figure 6.** PTLT to the speaker of the native and the non-native language during the silent test trial in 4-, 6-, and 10-month-old monolingual and bilingual infants. Dots represent each individual’s PTLT values and error bars the SE of the group mean.

As noted in the Participants section, 60% of the 10-month-old bilinguals were exposed to Catalan and Spanish (Catalan-Spanish bilinguals), while the other 40% were exposed to Catalan or Spanish as well as a second language other than those but from the same rhythmic class (Catalan/Spanish-Other bilinguals). To rule out potential differences in bilingual infants learning a second language other than Catalan or Spanish, the language-based preferences were assessed for each type of bilingual separately. Due to the small samples, non-parametric tests were used. Wilcoxon Signed-Rank Test revealed no preference for any of the

speakers in the 10-month-old Catalan-Spanish bilinguals ( $W = 35, p = .16$ ) or Catalan/Spanish-Other bilinguals ( $W = 5, p = .31$ ).

Lastly, the influence of exposure to Parental Code-switching was assessed. Parental Code-switching was computed as described in Study 1 and classified bilingual participants as having Low Code-switching Exposure (scores  $<10.25$ ) or High Code-switching Exposure (scores  $>10.25$ ), after performing a median split.

A mixed-effects ANOVA was performed only for the bilingual participants, including PTLT as the dependent variable, Speaker (Native language vs. Non-native language) as a within-subjects variable, and Age (4, 6, and 10 months) and Parental Code-switching (Low Code-switching Exposure vs. High Code-switching Exposure) as between-subjects variables.

The ANOVA revealed a marginally significant Speaker main effect ( $F(1,34) = 3.51, p = .07$ ) and a significant Speaker  $\times$  Age interaction ( $F(2,34) = 3.95, p < .05$ ). None of the other main effects or interactions approached significance (all  $ps > .1$ ).

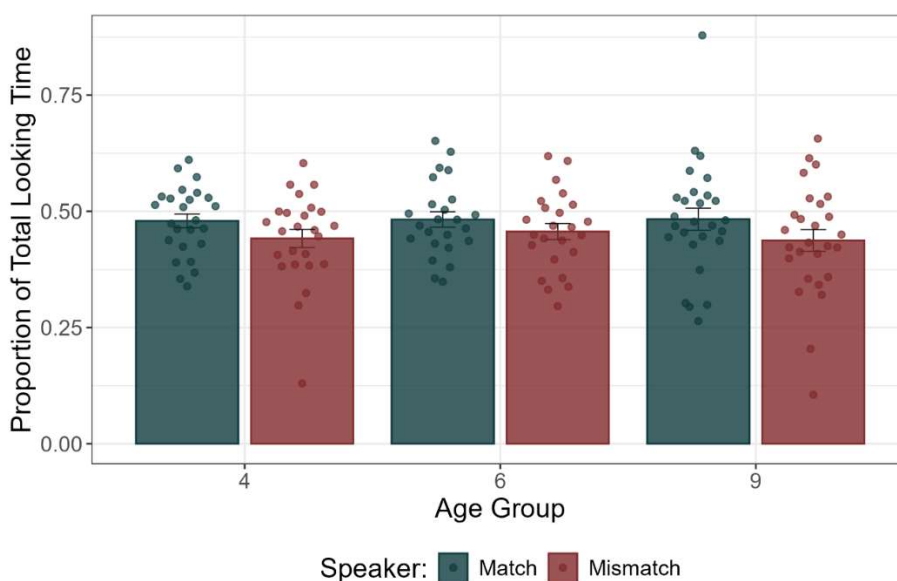
As the Speaker  $\times$  Age interaction in bilingual infants had already been described (see **Figure 6**), no further analyses were performed.

### Language test trials

To evaluate the formation of robust face-language associations of easier-to-discriminate stimuli, infants' visual preferences for the speakers were analyzed during the modified language test trials. The PTLT to each speaker was computed as before. The speakers were labeled as "match speaker" and "mismatch speaker," depending on whether the voice recording matched the language they had used in the familiarization. Furthermore, the trials in Catalan or Spanish were labeled as "native trial" and the English trials were labeled as "non-native trial".

A mixed-effects ANOVA was performed including PTLT as the dependent variable, Speaker (Match vs. Mismatch) and Trial Language (Native language vs. Non-native language) as within-subjects variables, and Linguistic Background (Monolingual vs. Bilingual) and Age (4, 6, and 10 months) as between-subjects variables.

The ANOVA revealed none of the main effects or interactions were significant (all  $ps > .1$ ). No significant visual preference for any speaker was observed at any age (all  $ps > .1$ ). See **Figure 7**.



**Figure 7.** PTLT to the match and mismatch speaker in the language test trials combined, in 4-, 6-, and 10-month-old infants. Dots represent each individual's PTLT values and error bars the SE of the group mean.

### Language test trials: exploring the bilingual experience

The potential effect of different linguistic experiences on the formation of robust face-language associations was explored with the two language test trials combined.

### Chapter 3: Speaker-Language Associations Before the First Year of Life

The lack of a significant Linguistic Background main effect or interaction in the ANOVA for the language test trials indicates that the results obtained were consistent between monolingual and bilingual infants.

To rule out potential differences in 10-month-old bilingual infants learning a second language other than Catalan or Spanish, the results were compared across Catalan-Spanish bilinguals and Catalan/Spanish-Other bilinguals separately. As in the previous analysis, non-parametric tests were used due to the small samples. Wilcoxon Signed-Rank Test revealed that 10-month-old Catalan-Spanish bilinguals ( $W = 11, p = .20$ ) and Catalan/Spanish-Other bilinguals ( $W = 18, p = .16$ ) looked equally to both speakers.

Lastly, the influence of exposure to Parental Code-switching was assessed. A mixed-effects ANOVA was performed only for the bilingual participants, including PTLT as the dependent variable, Speaker (Native language vs. Non-native language) as a within-subjects variable, and Age (4, 6, and 10 months) and Parental Code-switching (Low Code-switching Exposure vs. High Code-switching Exposure) as between-subjects variables.

Results from the ANOVA revealed that none of the main effects or interactions were significant (all  $ps > .1$ , except for the Age x Parental Code-switching interaction where  $p = .09$ ).

## Discussion

As in Study 1, the main goal of the current study was to explore the formation of primary and robust face-language associations during the first year of life, as well as potential linguistic factors that might modulate the development of this ability. However, in this study, the demands of the experimental task were decreased by using easier-to-discriminate stimuli. Infants were familiarized with two speakers using distant languages, one using Spanish or Catalan, and the other using English. While Catalan/Spanish was infants' native and dominant language, none of them had been previously exposed to English. After familiarization, infants were tested in a preferential-looking paradigm with two types of test trials. The silent test trial assessed primary face-language associations by measuring infants' language-based preference for the speakers in silence, and the language test trials assessed robust face-language associations by measuring infants' ability to match the faces of the speakers with their corresponding languages.

Results from the silent test trial further support that early in the first year of life infants form primary face-language associations, as 4-month-old infants exhibited language-based preferences for speakers. These results add to the existing literature (Colomer et al., 2023; Kinzler et al., 2007) and build upon the findings of the previous study.

Interestingly, the developmental pattern of these language-based preferences differs from that observed in Study 1 yet aligns with the findings of Colomer et al. (2023). In the present study, the preferences were only significant in 4-month-old infants, who preferred to look at the speaker of their native language compared to the speaker of the non-

### Chapter 3: Speaker-Language Associations Before the First Year of Life

native language. Six- and 10-month-old infants showed no significant preference for any speaker.

The differences in the developmental patterns of language-based preferences observed across studies may be attributed to two distinct, yet potentially complementary factors: the distance between the languages tested or infants' relative familiarity with each one. These considerations will be further addressed in the general discussion.

When assessing potential differences across linguistic backgrounds, the analyses revealed that language-based preferences were not modulated by bilingualism in general, nor by the level of exposure to parental code-switching. Monolingual and bilingual infants, regardless of the specific characteristics of the bilingual experience, showed the same preference pattern across development, with an early preference for speakers of their native language but no preference later in the first year of life.

As in Study 1, results from the language test trials do not provide strong evidence for the formation of robust face-language associations during the first year of life, despite testing infants with two distant languages to facilitate the task. The analyses indicate that infants looked equally at both speakers despite what language was playing in the background, and regardless of their age, linguistic background, or level of exposure to parental code-switching.

In sum, this study provides additional evidence for the early formation of primary face-language associations before the first year of life in both monolingual and bilingual infants. However, no evidence of face-language matching was observed, suggesting that robust face-language associations may not be formed before the first year of life.

## Chapter 3: Speaker-Language Associations Before the First Year of Life

Chapter 4:

**SPEAKER-LANGUAGE  
ASSOCIATIONS  
AFTER THE FIRST  
YEAR OF LIFE**



# Study 3:

## Development of Pragmatic Differentiation and Language Labeling

### Introduction

Language awareness is especially important in bilingual contexts, where children must form a separate linguistic system for each language, while constantly deciding which language to use based on the context or the speakers they encounter. The acquisition of sociolinguistic awareness, particularly regarding the language used by different speakers, has been explored in previous research by relying on either implicit or explicit measures.

Pragmatic differentiation refers to the adaptation of the language used according to the context or the interlocutor and is an implicit indicator that children know what language is commonly used by that speaker. Extensive research has focused on investigating this ability, primarily by examining bilingual children's language use during interactions with their caregivers and evaluating their level of language adaptation (Genesee et al., 1995; Goodz, 1989; Lanza, 1992; Montanari, 2008; Nicoladis, 1998; Nicoladis & Genesee, 1996; Quay, 2008). Most studies reported pragmatic differentiation to typically emerge around age 2, although other researchers found evidence of this ability as early as age 1;5 (Nicoladis, 1998). In general, pragmatic differentiation studies found that nearly all participants adjusted their language use to align with that of their

caregivers. However, language mixing was observed in most of the cases. Similar results were obtained in trilingual children (Montanari, 2008; Quay, 2008) or when pragmatic differentiation was assessed with unfamiliar speakers (Genesee et al., 1995, 1996). Two factors have been repeatedly proposed to account for differences in children's ability to adapt their language use to their interlocutor: a) exposure to language mixing (Genesee, 1989; Goodz, 1989), and b) language dominance (Genesee, 1989; Genesee et al., 1995, 1996; Montanari, 2008; Nicoladis, 1998; Nicoladis & Genesee, 1996). Children performed better in their dominant language compared to their non-dominant language, and those exposed to language mixing at home exhibited higher levels of mixing themselves.

Despite their valuable contributions, pragmatic differentiation studies have several limitations. One major limitation is the small sample sizes, typically ranging from 1 to 5 participants, which limits the generalizability of the findings. Additionally, less attention has been paid to how this ability develops beyond its initial emergence (see Tare & Gelman, 2010) or how the impact of its predictors may vary over time. Lastly, most of the studies focused on children from families following the one-parent-one-language strategy where the language mixing is generally very limited. This restricts the potential to evaluate the impact of parental language mixing and provides little insight into how pragmatic differentiation might evolve in other family language contexts.

Another way to evaluate bilingual children's sociolinguistic awareness is by assessing their ability to verbally express which languages are used by different speakers in their environment. This ability has been experimentally evaluated by directly asking the children what languages were used by their caregivers and themselves (Akhtar et al., 2012; Atagi & Sandhofer, 2020). Both studies revealed that the language labeling ability

was already present in 3-year-old children, although not in all the participants. Interestingly, the children tested were better at labeling the languages used by their caregivers rather than their own (Atagi & Sandhofer, 2020). Furthermore, children exposed to two languages (at home or in the community) were more likely to label the languages of their caregivers compared to children who were exposed to only one language (Akhtar et al., 2012; Atagi & Sandhofer, 2020). The authors attributed this advantage to the potentially greater levels of language-related conversations and reflections in bilingual contexts, compared to families where only one language is spoken. While this theory is plausible, the studies did not collect data on how frequently the families engaged in language reflections at home or compare it with children's performance.

The current study aims to explore the development of both implicit and explicit markers of sociolinguistic awareness in bilingual children aged 1 to 5 years, with a particular focus on children's understanding of which languages are spoken by different speakers. To do so, a new online questionnaire was created to collect parental reports on pragmatic differentiation and language labeling, as well as other aspects of general linguistic awareness. This approach enabled the recruitment of a large sample of children across a broad age range, which is essential for accurately assessing the development of pragmatic differentiation and language labeling. Both abilities were examined across the age range, along with potential predictors. Family language strategies and Language dominance were examined in relation to pragmatic differentiation, while language reflections at home were considered for language labeling. To increase the generalizability of the results of the study children from two bilingual communities (i.e., Quebec, Canada; and Catalonia, Spain) were tested.

## Methods

### Participants

The final sample included 473 caregivers who were raising a bilingual child aged between 1 and 5 years (up to their 6th birthday), with no developmental or health issues that could impact their language development. Two hundred ninety-seven participants were recruited from the autonomous community of Catalonia (Spain), primarily from the Barcelona region. One hundred seventy-six participants were recruited from the province of Quebec (Canada), predominantly from the Montreal region. See **Table 1** for the age distribution in each bilingual community.

**Table 1**

*Age distribution of the participants in each bilingual community*

	1-year-old	2-year-old	3-year-old	4-year-old	5-year-old
Catalonia	54	63	60	60	60
Quebec	42	52	32	24	26

Children were learning Catalan and Spanish in Catalonia, and French and English in Quebec, the most common languages in each community. Exposure to both languages must have started before the 6<sup>th</sup> month of life and children with more than 10% of cumulative exposure to other languages were excluded. To ensure the sample primarily consisted of functional bilinguals, a minimum proficiency in both languages was required for inclusion. Children must have at least a 6 out of 10 reported proficiency in understanding and speaking each of the languages, relative to other children their age. Since assessing proficiency in younger children can be challenging, especially for caregivers, the inclusion criteria for children under three years of age was based on their relative exposure to

the languages. Children must have at least a 25% reported cumulative exposure to each language, with relative exposures ranging from 75-25% to 50-50%.

## Language Separation and Awareness Questionnaire

The Language Separation and Awareness Questionnaire was designed as part of a project aiming to assess the development and the underlying structure of language separation and awareness in young bilingual children. The questionnaire consisted of 58 questions divided into two sections. The first focused on collecting descriptive data and information regarding potential predictors, while the second assessed children's language separation and awareness by collecting parental reports on diverse behaviors related to these dimensions of language development. The questionnaire was developed in Catalan, Spanish, English and French. The English version of the questionnaire can be found in **Appendix 1**.

### **Section 1: Descriptive data**

The first section was designed to gather demographic data (e.g., age, sex, or parental education), as well as information about the child's linguistic background (e.g., cumulative language exposure, language proficiency, or parental language use). Data from this section was also used to compute different variables to be included in the analyses as predictors for the development of language separation and awareness.

The Dominant Language for each child was determined by considering both their proficiency and exposure to each language. Catalan-dominant children represented 38.7% of the sample, 18.0% were French-dominant, 15.8% were Spanish-dominant, and 11.2% were English-dominant. No dominant language could be determined for the remaining 16.3% of the

## Chapter 4: Speaker-Language Associations After the First Year of Life

children, as they had identical proficiency and exposure to both languages.

The Language Strategy variable assessed families' patterns of language use and was determined from the reported percentage of time each caregiver used each language with the child. Participants were classified into 4 language strategies. For each strategy, the mean language mixing was computed considering the percentage of time the caregivers spoke to the child in the language they typically do not use with them.

- One language at home one outside (20.9%). Both caregivers predominantly used the same language with the child, and the child was exposed to the second language outside of the home. [mean language mixing: 8.25%]
- One parent one language (36.2%). Each caregiver predominantly used only one of the languages with the child. [mean language mixing: 6.25%]
- One parent bilingual (31.5%). One caregiver predominantly used only one language with the child, but the other caregiver used both. [mean language mixing: 8.25% for the monolingual caregiver and 37.8% for the bilingual caregiver]
- Two bilingual parents (11.4%). Both caregivers used both languages with the child. [mean language mixing: 37.4%]

The Language Reflections variable assessed children's exposure to language-related conversations at home. Four different types of discussions were measured: bilingualism in general, the language strategy used by the family, the languages used by different speakers, and the languages the child should use with different speakers. Caregivers were asked to report the frequency of each type of discussion using a 5-point Likert scale [1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Often, 5 = Very often].

A Language Reflections Score was computed by summing the answers provided in the 4 types of reflections.

## **Section 2: Language separation and awareness items**

The second section of the questionnaire contained 28 items designed to assess children's level of language separation and awareness. To perform a comprehensive exploration of this construct, the items tapped into 1 of 6 different aspects related to language separation and awareness. The included aspects were selected based on previous experimental research: language discrimination (e.g., Bosch & Sebastián-Gallés, 2001; Byers-Heinlein et al., 2010), lexical differentiation (e.g., David & Wei, 2008; Nicoladis, 1998), phonological differentiation (e.g., Bosch & Ramon-Casas, 2014; Mitchell et al., 2022), pragmatic differentiation (e.g., Genesee et al., 1995; Lanza, 1992), language preference and affiliation (e.g., Kinzler et al., 2007, 2012), and explicit language awareness (e.g., Akhtar et al., 2012; Atagi & Sandhofer, 2020). The items were partially derived from experimental measures used in previous research but were adapted to target behaviors that occur spontaneously in children and could be easily identified and reported by caregivers. See the full list of items in **Appendix 2**.

As the present thesis is focused on the development of speaker-language associations, particular attention will be paid to those items related to children's awareness of the languages used by different speakers. This type of sociolinguistic awareness was measured implicitly by assessing pragmatic differentiation, and explicitly by assessing children's ability to label the languages used by different speakers.

Ten items were developed to measure pragmatic differentiation, answered on a 5-point Likert scale [1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Often, 5 = Always].

## Chapter 4: Speaker-Language Associations After the First Year of Life

- Child's frequency of initiating conversation in the caregivers' most used language (items 10 and 14)
- Child's frequency of initiating conversation in the caregivers' least used language (items 11 and 15)
- Child's frequency of responding in the caregivers' most used language (items 12 and 16)
- Child's frequency of responding in the caregivers' least used language (items 13 and 17)
- Child's frequency of speaking in the language used by new speakers (items 18 and 19)

Four items were developed to measure language labeling.

- Child's ability to label own language(s) (item 34) [None - Only one - Both].
- Child's ability to label caregivers' language(s) (items 32 and 33) [Yes - No].
- Child's ability to label new speakers' language(s) (item 34) [Yes - No].

### Procedure

The Language Separation and Awareness Questionnaire was created and distributed online via the Qualtrics platform. Recruitment was initially conducted by reaching out to caregivers listed in the lab databases of families who had previously participated in, or expressed interest in participating in, studies from the Concordia Infant Research Lab (Quebec) or the APAL Lab (Catalonia). The study was also promoted on the social media pages of the aforementioned labs. Furthermore, participants were encouraged to share the questionnaire with further potential participants (snowball sampling).

## Results

### Exploratory Factor Analysis

To reveal the underlying structure of language separation and awareness an Exploratory Factor Analysis (EFA) was performed. Based on preliminary analyses, 4 items were excluded from the final analysis: item 28, which showed minimal variability across participants, items 13 and 17, which contained a high proportion of missing data, and item 5, which exhibited very low loading scores across all factors. The remaining 24 items were included in the final EFA. As the data from Quebec and Catalonia yielded similar results regarding the underlying structure of this construct, both datasets were combined, and a joint EFA was conducted.

First, the correlations across the included items were estimated. Given that the items were ordinal and did not follow a normal distribution, a polychoric correlation matrix was generated, in line with EFA recommendations (Baglin, 2014; Watkins, 2018). The appropriateness of the data was assessed by the Bartlett's test of sphericity ( $\chi^2(276, N = 473) = 2590.004, p < .001$ ) and the Kaiser-Meyer-Olkin measure of sampling adequacy (KMO = .40). The low overall KMO indicates that the correlations among items are not as strong as desired, suggesting the need for further item refinement. Consequently, the factors extracted from this data should be viewed as a preliminary insight into the structure of language separation and awareness, serving as a foundational step for further investigation.

Several strategies were employed to determine the optimal number of factors to extract. According to the Kaiser's criterion, combined with a scree plot, only factors with eigenvalues greater than 1 were retained. Additionally, a parallel analysis was performed. The results obtained from

both strategies were contrasted with theoretical expectations, ensuring the factors were not only statistically significant but also theoretically meaningful.

Finally, the EFA was conducted with a 3-factor solution. The factor loading of each item was computed using the minimal residual method with an oblique rotation, as it was anticipated that the factors would exhibit some degree of correlation with each other. The three extracted factors were labeled based on the items that loaded onto them: Language Separation, Language Preference, and Pragmatic Differentiation. All items designed to measure pragmatic differentiation loaded onto the Pragmatic Differentiation factor, except for the item "Speaking in a new person's language (Spanish/English)", which instead loaded onto the Language Separation factor. The items assessing children's ability to label the language used by different speakers also loaded onto the Language Separation factor. See **Figure 8** for a detailed representation of the factors and the factor loadings of the items.

Chapter 4: Speaker-Language Associations After the First Year of Life

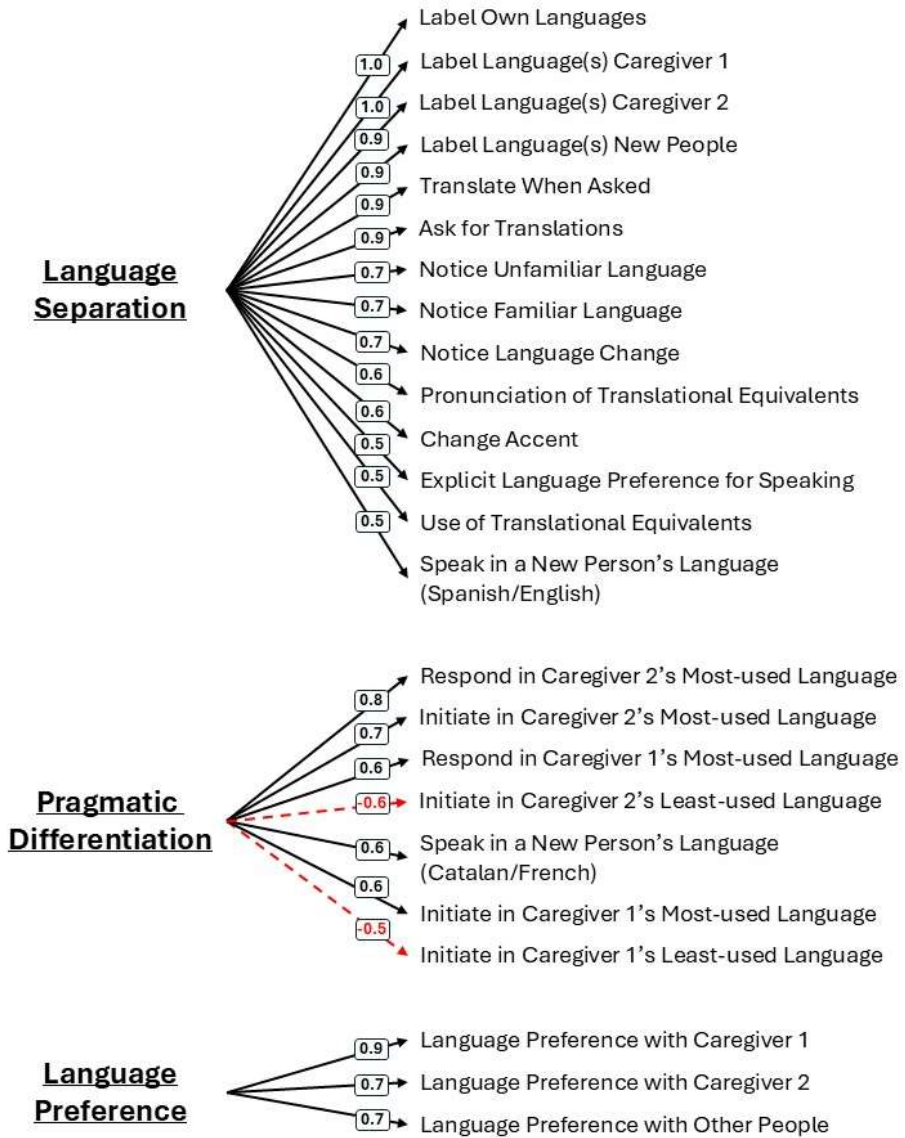


Figure 8. Factor Loadings for each item included in the EFA.

## Exploring Pragmatic Differentiation

According to the results provided in the EFA, the items related to pragmatic differentiation were analyzed combined as a factor. For each child, the responses to each item were multiplied by their corresponding factor loading, and a Pragmatic Differentiation Score was computed by averaging across items. The score ranged from 0 to 2.53, with a mean of 1.42 and a standard deviation of 0.6.

To explore the development of pragmatic differentiation, the Pragmatic Differentiation Score was introduced as the dependent variable in a linear regression model (R package ‘stats’ version 4.0.3, function `lm`), with Age, Language Strategy (One language at home one outside, One parent one language, One bilingual parent or Two bilingual parents), and Community (Catalonia or Quebec) as predictors. The dominant language effect will be examined later in a separate analysis. The model included the fixed effects of the three predictors, as well as their interactions, and was specified as:

*Pragmatic Differentiation Score ~ Age \*Language Strategy\* Community*

‘One language at home one outside’ for the Language Strategy variable and ‘Catalonia’ for the Community variable were used as the baseline in the model. The results obtained from the model are summarized in **Table 2**.

**Table 2**

*Summary of the linear regression model*

<i>Predictors</i>	<i>Estimates</i>	<i>std. Error</i>	<i>t value</i>	<i>p</i>	<i>p&lt;.05</i>
(Intercept)	1.15	0.17	6.62	<b>&lt;.001</b>	<b>***</b>
Age	0.14	0.04	3.56	<b>&lt;.001</b>	<b>***</b>
Language Strategy (One parent one language)	-0.82	0.20	-3.93	<b>&lt;.001</b>	<b>***</b>
Language Strategy (One bilingual parent)	-0.76	0.22	-3.44	<b>&lt;.001</b>	<b>***</b>
Language Strategy (Two bilingual parents)	-0.83	0.29	-2.85	<b>.004</b>	<b>**</b>
Community (Quebec)	0.36	0.37	0.98	.328	
Age * Language Strategy (One parent one language)	0.13	0.05	2.64	<b>.008</b>	<b>**</b>
Age * Language Strategy (One bilingual parent)	0.14	0.05	2.46	<b>.014</b>	
Age * Language Strategy (Two bilingual parents)	0.08	0.07	1.02	.308	
Age * Community (Quebec)	-0.10	0.09	-1.14	.252	

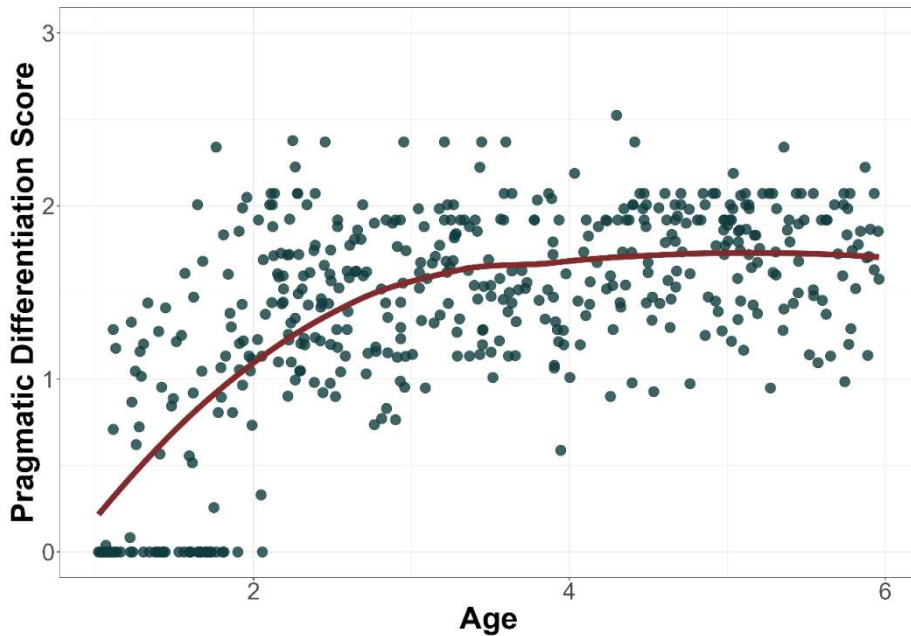
**Table 2 (continued)**

<i>Predictors</i>	<i>Estimates</i>	<i>std. Error</i>	<i>t value</i>	<i>p</i>	<i>p&lt;.05</i>
Language Strategy (One parent one language) * Community (Quebec)	0.21	0.42	0.50	.617	
Language Strategy (One bilingual parent) * Community (Quebec)	-0.12	0.41	-0.31	.757	
Language Strategy (Two bilingual parents) * Community (Quebec)	0.01	0.48	0.010	.992	
(Age * Language Strategy (One parent one language)) * Community (Quebec)	0.01	0.11	0.100	.920	
(Age * Language Strategy (One bilingual parent)) * Community (Quebec)	0.07	0.10	0.656	.512	
(Age * Language Strategy (Two bilingual parents)) * Community (Quebec)	0.097	0.13	0.735	.463	

*Note.* '\*\*' < 0.05; '\*\*\*' < 0.01; '\*\*\*\*' < 0.001

As no significant Community main effect or interaction was found, the subsequent analyses were performed using the combined datasets from Catalonia and Quebec.

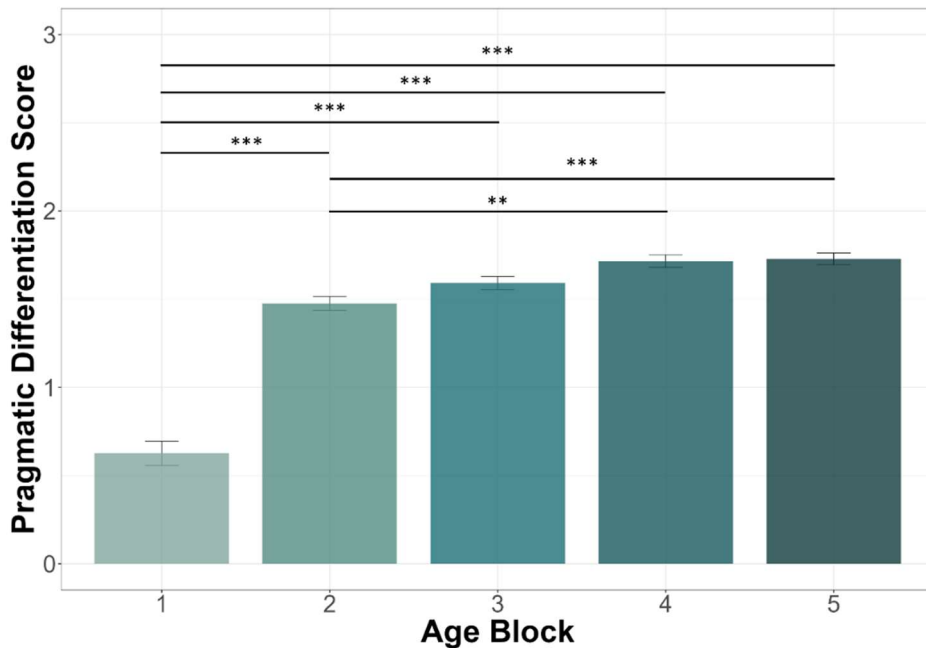
In line with the significant Age main effect [Estimate = 0.146 (0.041),  $t = 3.565$ ,  $p < .001$ ], the Pragmatic Differentiation Score showed a moderate positive correlation with age ( $r(471) = .57$ ,  $p < .001$ ). However, visual inspection of the data (**Figure 9**) suggests that the age effect is not linear throughout the whole age range tested. To better assess variations in the age effect throughout development, the sample was divided into 1-year age blocks and correlations were computed between age and pragmatic differentiation for each block. The Pragmatic Differentiation Score was significantly and positively correlated with age in the 1-year-old block ( $r(94) = .41$ ,  $p < .001$ ), but no correlation was found in any of the other ages (all  $ps > .1$ ).



**Figure 9.** Scatter plot showing the relationship between age and the Pragmatic Differentiation Score. Dots represent individual scores, and the solid line, the LOWESS-smoothed trend line.

To further explore the developmental trajectory of pragmatic differentiation, a one-way ANOVA was performed with Pragmatic Differentiation Score as the dependent variable and Age Block (1-year-old to 5-year-old) as the between-subjects variable. The ANOVA revealed a significant Age Block main effect ( $F(4,468) = 102.7, p < .001$ ). A Tukey HSD post hoc test was performed to determine which age groups differed significantly in their mean Pragmatic Differentiation Score. The 1-year-old group provided significantly lower scores ( $M = 0.63, SD = 0.67$ ) compared to the 2-year-old ( $M = 1.48, SD = 0.41$ ), the 3-year-old ( $M = 1.59, SD = 0.35$ ), the 4-year-old ( $M = 1.72, SD = 0.33$ ) and the 5-year-old groups ( $M = 1.73, SD = 0.31$ ), with all comparisons showing  $p < .001$ . Further exploration of the 1-year-old group revealed that 45% of the children in this group were not yet able to produce speech in either of their two languages, preventing

them from expressing pragmatic differentiation. The 2-year-old group had lower scores than the 4-year-old ( $p < .01$ ) and 5-year-old groups ( $p < .001$ ), but not when compared with the 3-year-old group ( $p = .33$ ). Finally, no significant differences were found between the 3-, 4- and 5-year-old groups, with all comparisons showing  $p > .1$ . These results are depicted in **Figure 10**.

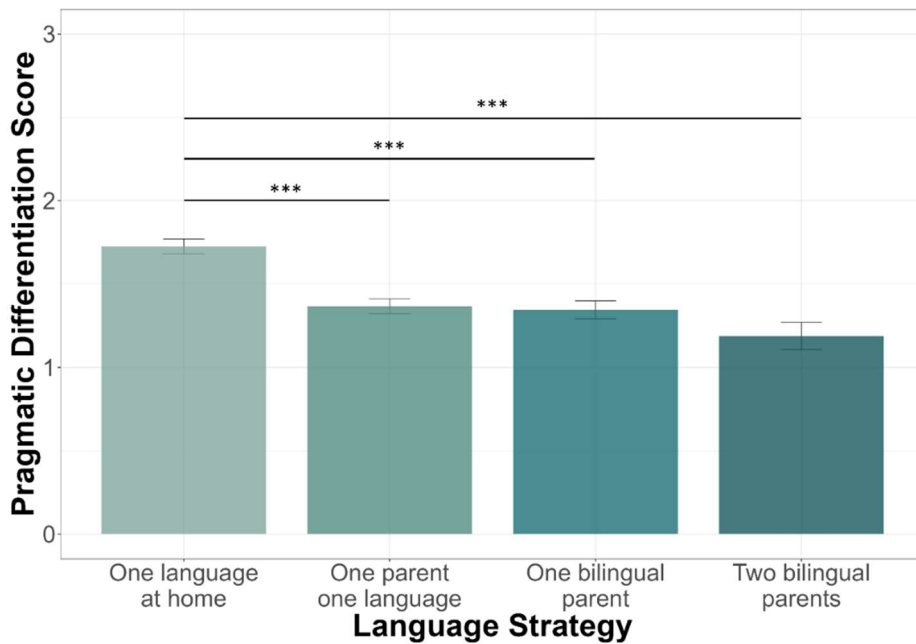


**Figure 10.** Bar plot depicting the mean Pragmatic Differentiation Score for each 1-year Age Block. Error bars represent the SE of the group mean.

The linear regression model also revealed differences in pragmatic differentiation across the different Language Strategies at home. Taking ‘One language at home one outside’ as the reference, all the other strategies presented negative estimates, suggesting lower Pragmatic Differentiation Scores: ‘One parent one language’ [Estimate = -0.823 (0.209),  $t = -3.935$ ,  $p < .001$ ], ‘One bilingual parent’ [Estimate = -0.765

(0.222),  $t = -3.440$ ,  $p < .001$ ] and 'Two bilingual parents' [Estimate = -0.837 (0.293),  $t = -2.856$ ,  $p < .01$ ].

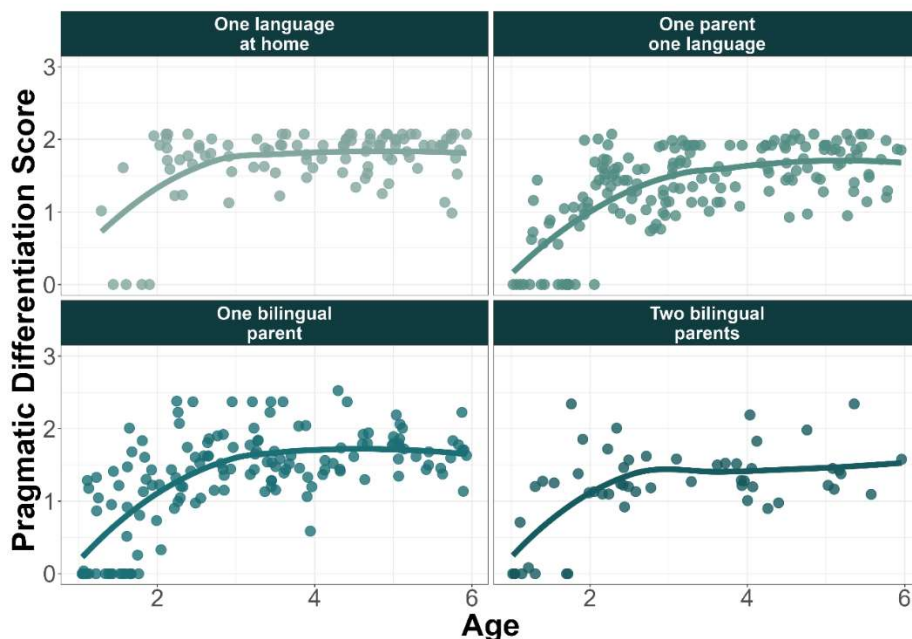
To further explore the differences across strategies a one-way ANOVA was performed with Pragmatic Differentiation Score as the dependent variable and Language Strategy (One language at home one outside, One parent one language, One bilingual parent or Two bilingual parents), as the between-subjects variable. The ANOVA revealed a significant effect of Language Strategy ( $F(3,469) = 13.52$ ,  $p < .001$ ). A Tukey HSD post hoc test was used to perform multiple comparisons across the different strategies. The test revealed higher scores for 'One language at home one outside' ( $M = 1.72$ ,  $SD = 0.44$ ) compared to 'One parent one language' ( $M = 1.36$ ,  $SD = 0.57$ ), 'One bilingual parent' ( $M = 1.34$ ,  $SD = 0.65$ ) and 'Two bilingual parents' ( $M = 1.18$ ,  $SD = 0.60$ ), with all comparisons showing  $p < .001$ . No significant differences were found across the other strategies (all  $ps > .1$ ). These results are depicted in **Figure 11**.



**Figure 11.** Bar plot depicting the mean Pragmatic Differentiation Score for each Language Strategy. Error bars represent the SE of the group mean.

Lastly, the significant interaction observed in the model between Age and Language Strategy: ‘One parent one language’ [Estimate = 0.137 (0.052),  $t = 2.648$ ,  $p < .01$ ] suggests that the age effect is stronger in this strategy compared to the reference level ‘One language at home one outside’. Further assessment of this interaction revealed that, while age was a significant predictor for all strategies, the strength of the correlation varied across them. In the ‘One parent one language’ strategy, age strongly positively correlated with the Pragmatic Differentiation Score ( $r(169) = 0.6$ ,  $p < .001$ ). Moderate positive correlations were observed in the ‘One parent bilingual’ ( $r(147) = 0.57$ ,  $p < .001$ ) and ‘Two parents bilingual’ ( $r(52) = 0.49$ ,  $p < .001$ ) strategies. Lastly, in the ‘One language at home one outside’ strategy, the observed correlation was weak ( $r(97) = 0.37$ ,  $p < .001$ ). Visual examination of the data (see **Figure 12**) suggests that children from families following the ‘One language at home one outside’ strategy

had higher Pragmatic Differentiation Scores from early on, while for the rest of the strategies, this score gradually increased with age.



**Figure 12.** Scatter plots showing the relationship between the Pragmatic Differentiation Score and Age for each Language Strategy. Dots represent individual scores, and the solid lines, the LOWESS-smoothed trend line.

Another factor that has been repeatedly reported to influence pragmatic differentiation is language dominance. To explore the language dominance effect, children’s ability to adapt their language use to their interlocutors was assessed when the required language was their dominant language, as well as when it was their non-dominant language.

For this analysis, the pragmatic differentiation items with negative factor loadings were excluded (items 11 and 15) to contemplate only those items that were direct and positive indicators of language adaptation. For this same reason, item 19 was included despite loading into the Language Separation factor. The items were classified depending on whether

children were required to use their dominant language or their non-dominant language in order to adapt their language use to the interlocutor. The classification was made for each child, according to their language dominance (see Participants section). Children who did not show clear dominance in one language over the other were excluded from this analysis (N = 77). Then, a Pragmatic Differentiation Score was computed for the dominant and the non-dominant language by averaging the raw answers for the items that required each one of the languages.

The Pragmatic Differentiation Score was introduced as the dependent variable in a linear mixed model (R package 'lme4' version 1.1-26, function `lmer`), with Age and Language Required (Dominant vs. Non-dominant) as predictors. The model included the fixed effects of the two predictors, as well as their interaction. Participant was added as a random effect. The model was specified as:

$$\textit{Pragmatic Differentiation Score} \sim \textit{Age} * \textit{Language Required} + (1 \mid \textit{Participant})$$

For the Language Required variable, 'Dominant' was used as the baseline in the model. The results obtained from the model are summarized in **Table 3**.

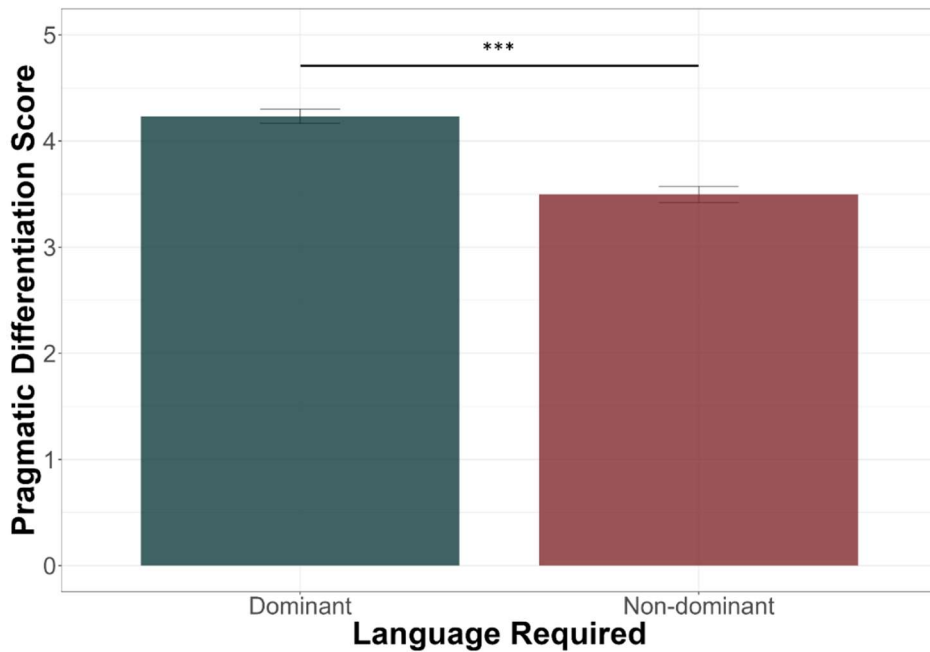
**Table 3***Summary of the linear mixed model*

<i>Predictors</i>	<i>Estimates</i>	<i>std. Error</i>	<i>t value</i>	<i>p</i>	<i>p &lt; .05</i>
(Intercept)	2.42	0.15	15.42	<b>&lt;.001</b>	<b>***</b>
Age	0.51	0.04	12.43	<b>&lt;.001</b>	<b>***</b>
Language Required (Non-dominant)	-1.30	0.15	-8.32	<b>&lt;.001</b>	<b>***</b>
Age * Language Required (Non- dominant)	0.16	0.04	3.90	<b>&lt;.001</b>	<b>***</b>

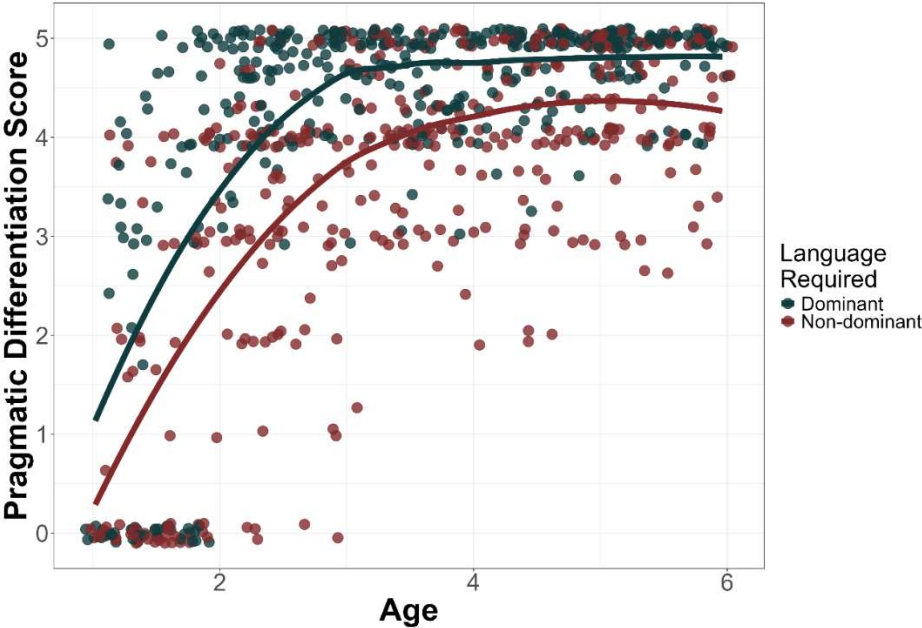
*Note.* '\*\*' < 0.05; '\*\*\*' < 0.01; '\*\*\*\*' < 0.001

The significant fixed effect of Age [Estimate = 0.517 (0.042),  $t = 12.434$ ,  $p < .001$ ] implies that pragmatic differentiation increased with age, as already described in the previous analysis. The model also revealed differences in pragmatic differentiation depending on the Language Required [Estimate = -1.306 (0.157),  $t = -8.328$ ,  $p < .001$ ]. As represented in **Figure 13**, children obtained higher Pragmatic Differentiation Scores when the language they had to use was their dominant language ( $M = 4.23$ ,  $SD = 1.33$ ) compared to when they had to use their non-dominant language ( $M = 3.50$ ,  $SD = 1.53$ ;  $t(395) = 12.46$ ,  $p < .001$ ). Furthermore, the significant interaction between Age and Language Required [Estimate = 0.162 (0.042),  $t = 3.900$ ,  $p < .001$ ] indicates that the development of pragmatic differentiation through age differed depending on the language required to express it (see **Figure 14**). To further explore this interaction, the mean Pragmatic Differentiation Score was compared between Dominant and Non-dominant language for each 1-year age block separately. Multiple comparisons were adjusted

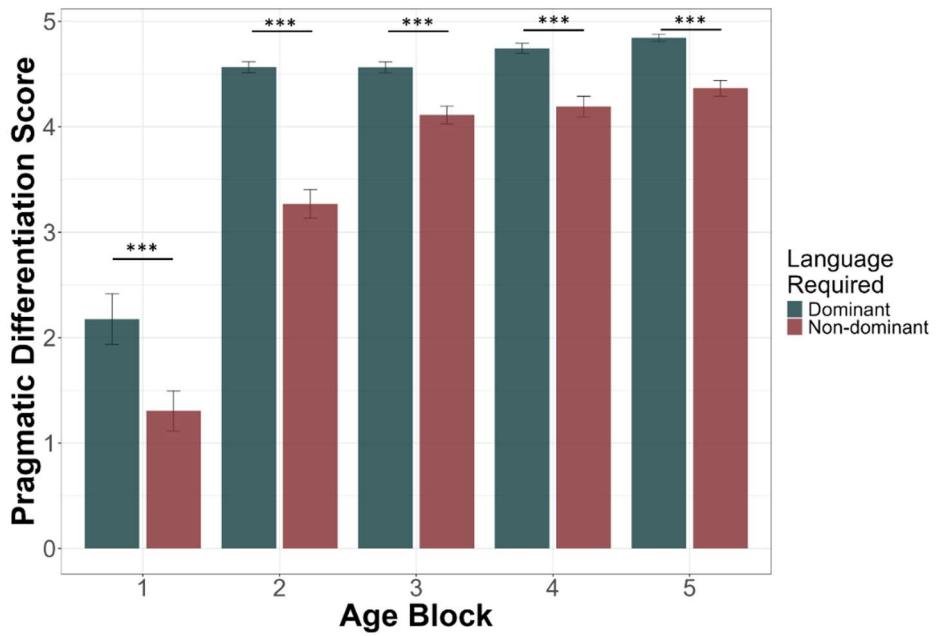
using Tukey's method. The analysis revealed significant differences in pragmatic differentiation between languages in all age blocks (all comparisons  $p < .001$ ). However, the differences across languages were smaller in older children (see **Figure 15**)



**Figure 13.** Bar plot depicting the mean Pragmatic Difference Score for the Dominant and the Non-dominant language. Error bars represent the SE of the group mean.



**Figure 14.** Scatter plots depicting the relationship between the Pragmatic Differentiation Score and Age for each Language Required. Dots represent individual scores (2 per participant, 1 per each language), and the solid lines, the LOWESS-smoothed trend line.



**Figure 15.** Bar plot depicting the mean Pragmatic Difference Score for the Dominant and the Non-dominant language at each 1-year Age Block. Error bars represent the SE of the group mean.

## Exploring Language Labeling

Unlike the pragmatic differentiation items, the items designed to measure language labeling did not load onto a distinct, separate factor but instead loaded onto the broader Language Separation factor. As a result, these items will be analyzed separately rather than as part of a single factor.

Four items were created to reflect children's ability to explicitly label the languages used by different speakers: themselves, their caregivers, and new people. Some items had to be re-coded or combined to provide binary Language Labeling measures (Yes-No) for each of the speakers assessed. Item 26 measured children's ability to label their own languages. Most of the participants could either label both of their languages or none of them. Only 3.4% of the participants were reported to label one of their languages but not the other. This measure was re-coded to divide the sample into children who could not label any of their languages and children who could label at least one of them. Items 23 and 24 assessed children's ability to label the language(s) used by their caregivers. Both items were collapsed to obtain a combined measure. In almost every case, the answer to both items was the same. Generally, children were either able to label the language(s) used by both of their caregivers or unable to label the language(s) used by either caregiver. Only 2% of children were reported to label the languages of one of the caregivers but not the other. The combined measure divided the sample into children unable to label the language(s) of any of their caregivers and children who were able to label at least the language(s) of one of them. Lastly, item 25 measured children's ability to label the language(s) used by new people.

## Chapter 4: Speaker-Language Associations After the First Year of Life

As the dependent variable was binary and each participant provided data for the 3 different measures of Language Labeling, a generalized linear mixed model (R package 'lme4' version 1.1-26, function glmer) was used to assess the development of children's language labeling ability. Language Labeling (Yes or No) was included as the dependent variable in the model, with Age, Speaker (Themselves, Caregivers, or New people), Language Reflections, and Community (Catalonia or Quebec) as predictors. The model included the fixed effects of the four predictors, as well as the Age x Speaker, Age x Language Reflections, and Age x Community interactions. Participant was added as a random effect. The model was specified as:

*Language Labeling ~ Age + Speaker + Language Reflections + Community + Age\*Speaker + Age\*Language Reflections + Age\*Community + (1|Participant)*

'Themselves' for the Speaker variable and 'Catalonia' for the Community variable were used as the baseline in the model. The results obtained from the model are summarized in **Table 4**.

**Table 4***Summary of the generalized linear mixed model*

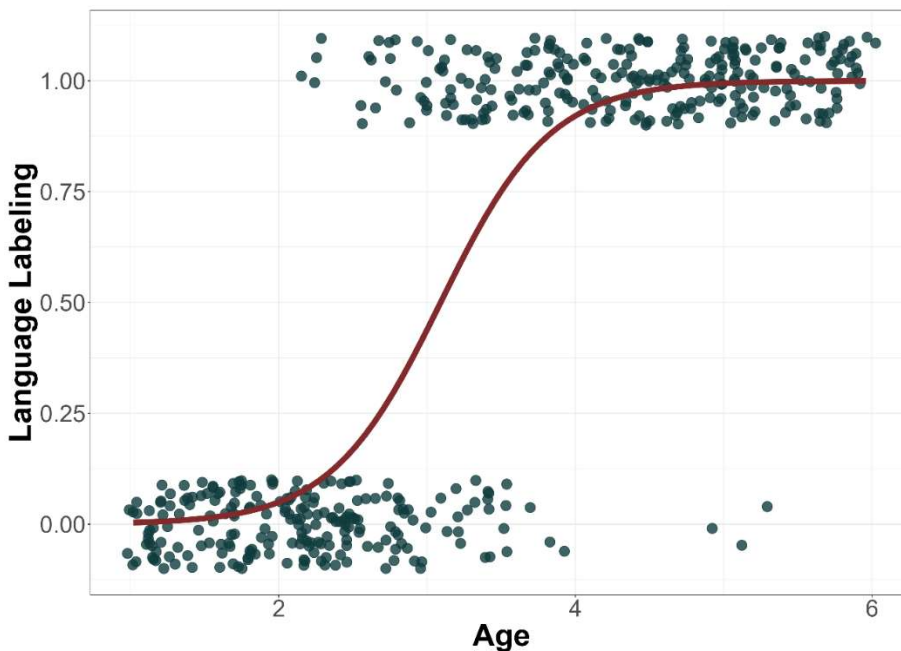
<i>Predictors</i>	<i>Estimates</i>	<i>std. Error</i>	<i>z value</i>	<i>p</i>	<i>p&lt;.05</i>
(Intercept)	-52.94	10.97	-4.82	<b>&lt;.001</b>	<b>***</b>
Age	14.55	2.95	4.91	<b>&lt;.001</b>	<b>***</b>
Speaker (Caregivers)	-0.42	2.43	-0.17	.861	
Speaker (New person)	4.59	2.67	1.71	.086	
Language Reflections	4.32	1.37	3.13	<b>.002</b>	<b>**</b>
Community (Quebec)	-9.15	5.77	-1.58	.113	
Age * Speaker (Caregivers)	-0.41	0.69	-0.59	.552	
Age * Speaker (New person)	-2.49	0.83	-2.99	<b>.003</b>	<b>**</b>
Age * Language Reflections	-0.89	0.32	-2.73	<b>.006</b>	<b>**</b>
Age * Community (Quebec)	2.53	1.62	1.56	.118	

---

*Note.* **'\*** <0.05; **'\*\*'** <0.01; **'\*\*\*'** <0.001

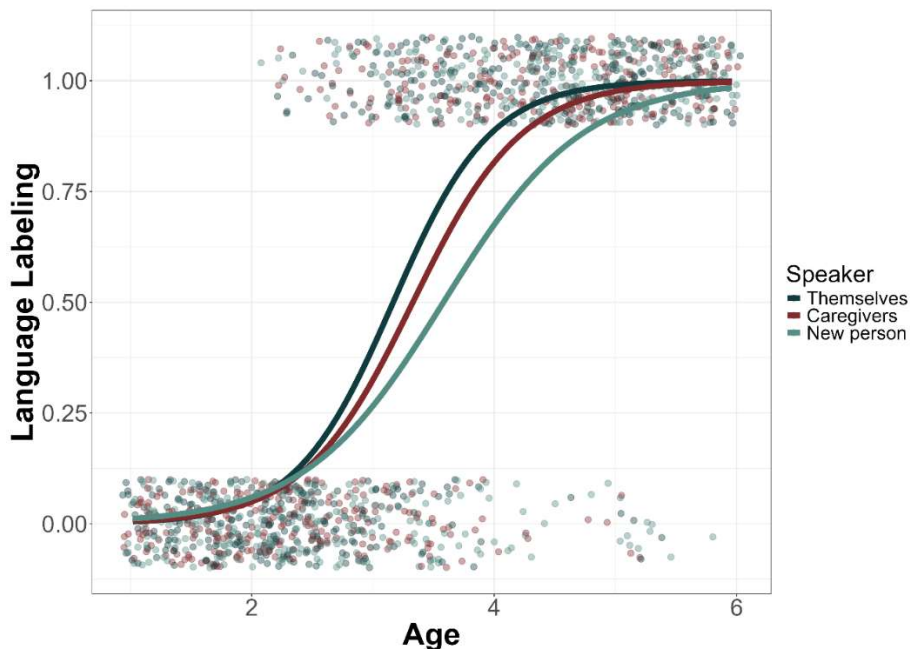
As no significant Community main effect or interaction was found, the subsequent analyses were performed using the combined datasets from Catalonia and Quebec.

The model revealed that Age had a significantly positive effect on Language Labeling [Estimate = 14.550 (2.959),  $z = 4.917$ ,  $p < .001$ ], indicating that the proportion of participants able to label the languages used by speakers increased with age. To further explore the main Age effect, the 3 measures were collapsed into a combined variable that classified the sample into children who exhibit labeling in at least one of the measures, and children who did not label in any of the situations. This new Language Labeling variable was contrasted with Age. Visual inspection of the data (**Figure 16**) shows that no child before the age of 2 was able to label the languages used by speakers and that after the 4<sup>th</sup> birthday, almost all participants (only 4 exceptions) could label the language of at least one of the speakers assessed. To determine the age at which the probability of Language Labeling was significantly higher than the chance level (0.5), a logistic regression model was fitted with Age as the predictor and Language Labeling as the binary outcome. Predicted probabilities and 95% confidence intervals were calculated across the age range, revealing that from age 3.22 children were significantly more likely to label the languages used by speakers compared to chance.



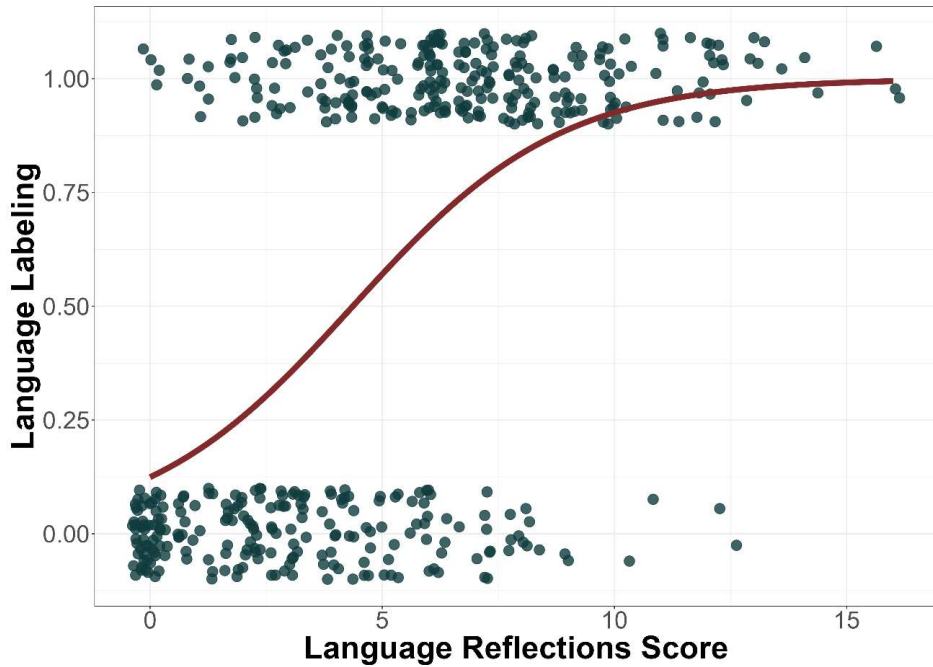
**Figure 16.** Scatter plot showing the relationship between the binary variable Language Labeling and Age. Dots represent individual observations, with a slight jitter added to prevent overlapping. The solid line is the predicted probability of Language Labeling as a function of Age, derived from a logistic regression model.

The model revealed no significant Speaker main effect. However, the significant interaction between Speaker: New person and Age [Estimate = -2.496 (0.835),  $z = -2.990$ ,  $p < .01$ ] suggests that the development of the language labeling ability differs depending on the measure being assessed (see **Figure 17**). To further explore this interaction, the age at which the probability of language labeling was significantly higher than the chance level (0.5) was calculated for each measure separately, following the method previously described. According to the results, children were first able to label their own languages (age 3.32) followed by the language(s) of their caregivers (age 3.52), and lastly, the language(s) of new speakers (3.82).

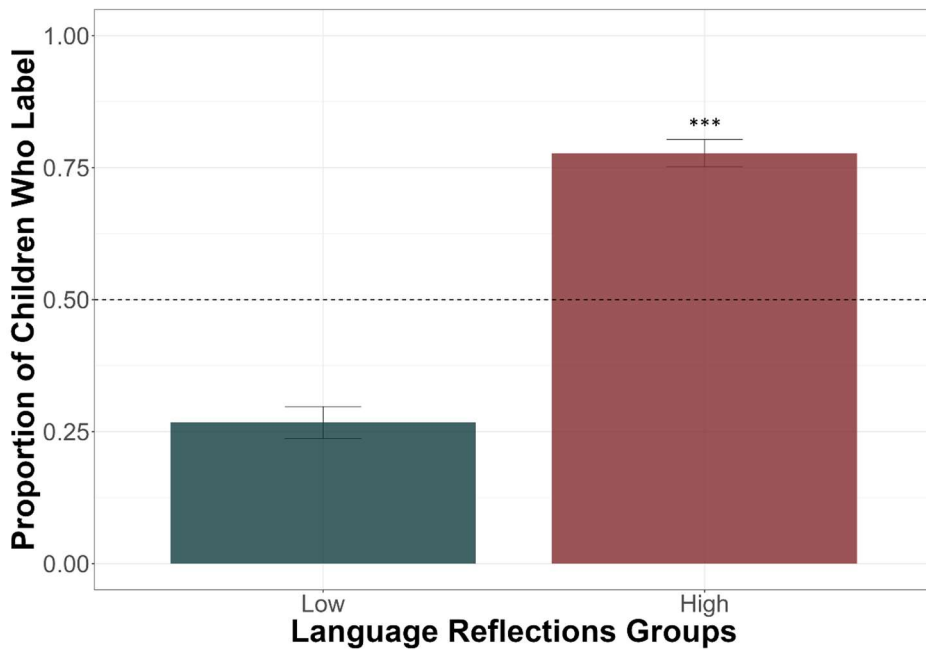


**Figure 17.** Scatter plot comparing the relationship between the binary variable Language Labeling and Age across the 3 different measures. Dots represent individual observations (3 per participant, 1 per each measure) with a slight jitter and transparency added to reduce overlapping. The solid lines are the predicted probability of Language Labeling for each Speaker as a function of Age, derived from logistic regression models.

The model also suggests that the proportion of Language Labeling was higher as the Language Reflections at home increased [Estimate = 4.324 (1.378),  $z = 3.138$ ,  $p < .01$ ]. However, this relationship was not as strong as the one observed with age (see **Figure 18**). To further explore the Language Reflections effect, a median split was performed to divide participants into children with a high frequency of exposure to language reflections at home and those with a low frequency. Language labeling was present in 78% of the children frequently exposed to language reflections, in contrast with only 27% of the children who had low frequency of exposure to those reflections. See **Figure 19**.



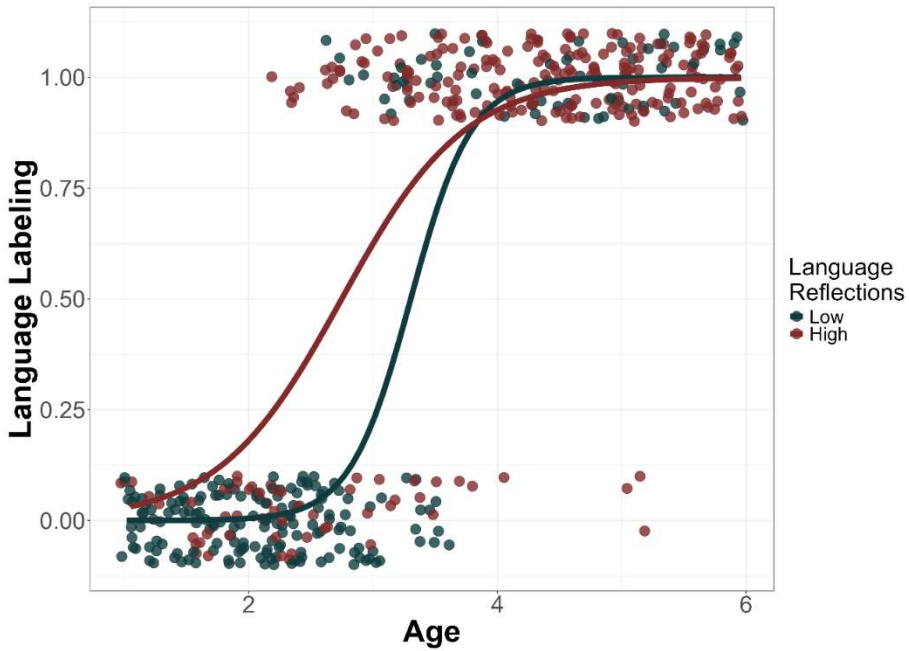
**Figure 18.** Scatter plot showing the relationship between the binary variable Language Labeling and Language Reflections. Dots represent individual observations, with a slight jitter added to prevent overlapping. The solid line is the predicted probability of Language Labeling as a function of Language Reflections, derived from a logistic regression model.



**Figure 19.** Bar plot depicting the Proportion of Children Who Label for the Low vs High Language Reflections Groups. Error bars represent the SE of the group proportion. The dashed line at 0.5 represents 50% of the sample.

It is important to note that Age and the Language Reflections Score were moderately correlated ( $r(471) = 0.51, p < .001$ ). Older children were more likely to be frequently exposed to language reflections at home compared to younger children. To separate the effect of these two variables on Language Labeling, and to explore the Age x Language Reflections significant interaction observed in the model [Estimate = -0.898 (0.328),  $z = -2.734, p < .01$ ], the Age effect was assessed for the children with low and high frequency of exposure to language reflections at home separately. The age at which the probability of language labeling was significantly higher than the chance level (0.5) was calculated for each group as described in the previous analyses. Children with high frequency of exposure to language reflections at home showed above-chance language

labeling earlier (age 3.05) than children with a low frequency of exposure to language reflections at home (age 3.52). See **Figure 20**.



**Figure 20.** Scatter plot comparing the relationship between the binary variable Language Labeling and Age between children exposed to Low and High levels of Language Reflections at Home Dots represent individual observations with a slight jitter added to avoid overlapping. The solid lines are the predicted probability of Language Labeling for each Language Reflections group as a function of Age, derived from logistic regression models.

## Discussion

This study aimed to explore the development of both implicit and explicit markers of sociolinguistic awareness in 1- to 5-year-old bilingual children from two bilingual communities (i.e., Quebec, Canada and Catalonia, Spain). More specifically, it focused on children's awareness of the languages spoken by different speakers. For this purpose, a new online questionnaire was created to collect parental reports on pragmatic differentiation and language labeling. Pragmatic differentiation refers to children's ability to adapt the language they use based on the speaker they are interacting with and is an implicit indicator of their awareness of what language that speaker primarily uses. In turn, language labeling refers to children's ability to verbally and explicitly specify the languages used by different speakers. This study examined the development of both abilities across the tested age range while also evaluating the impact of potential predictors: family language strategies and language dominance for pragmatic differentiation, and language reflections at home for language labeling.

The first finding in this study is the absence of significant differences between the communities tested in terms of pragmatic differentiation or language labeling. This suggests that the development of sociolinguistic awareness is similar in Catalonia and Quebec. Accordingly, the following discussion refers to the results obtained from the combined data from both communities.

The analysis of the pragmatic differentiation data revealed that, as expected, this ability improved with age. However, the age effect was only relevant during the early stages of development, especially before age 2. In the 1-year-old group, a big part of the children (45%) were not yet able to

produce speech in either of their two languages. This prevented them from expressing pragmatic differentiation regardless of whether they had developed two linguistic systems or knew which language corresponded to each speaker. The performance of the remaining children in this group, those who were already able to speak, showed considerable variability that could be explained by age. As children approached their second birthday, their pragmatic differentiation scores were higher. After the second birthday, the effect of age diminished drastically becoming negligible by age 3. No differences were observed in the level of pragmatic differentiation among 3-, 4-, and 5-year-old children, suggesting a ceiling effect.

Taken together, these results indicate that pragmatic differentiation is largely established during age 2, in line with previous findings (Genesee et al., 1995; Goodz, 1989; Lanza, 1992; Montanari, 2008; Nicoladis & Genesee, 1996; Quay, 2008). Moreover, the results also provide evidence for the potential earlier emergence of pragmatic differentiation, before age 2, as several 1-year-old children who were already able to speak exhibited levels of pragmatic differentiation comparable to those of older children (see also Nicoladis, 1998). Lastly, results from Study 3 also suggest that after the third birthday, there is no further improvement in pragmatic differentiation that could be attributed to age or the developmental stage of the children (see also Tare & Gelman, 2010).

Despite the importance of age as a predictor for pragmatic differentiation, children's ability to adapt their language to their interlocutor still exhibited notable variability that could not be accounted for by age alone. This implies that other factors must be contributing to these individual differences, such as the language strategy followed by the family and the dominant language of the child.

Differences in pragmatic differentiation were observed across different language strategies. Children from families that followed a 'One language at home one outside' strategy exhibited higher levels of pragmatic differentiation compared to those from families using any of the other strategies. The results for the 'One parent one language,' 'One bilingual parent,' and 'Two bilingual parents' strategies were similar, with no significant differences observed among them. Furthermore, the developmental trajectory of pragmatic differentiation was also different across the strategies. The age effect was moderate to strong in all the strategies, except for the 'One language at home one outside' strategy where the effect was small. Visual exploration of the data indicated that the results were driven by children from families who followed this strategy exhibiting higher levels of pragmatic differentiation from an earlier age. In contrast, the development of the ability was more progressive and gradual in children from families following any of the other strategies. The potential reasons for the advantage of the 'One language at home one outside' strategy will be further addressed in the general discussion.

Another important factor proposed to modulate pragmatic differentiation is children's language dominance (Genesee, 1989; Genesee et al., 1995, 1996; Nicoladis, 1998; Nicoladis & Genesee, 1996). Results from the present study revealed that, in line with previous research, children were better at adapting their language use to their interlocutor when the required language was their dominant language, in comparison to when the required language was their non-dominant. Differences in pragmatic differentiation between their languages were present along the whole tested age range, but the magnitude of the difference diminished with age.

The analysis of the language labeling data revealed that this ability also improved with age. Although some 2-year-old children were reported to

label the language of at least one of the speakers assessed, the language labeling ability was not robustly present until age 3 (see also Akhtar et al., 2012 and Atagi & Sandhofer, 2020). Moreover, after their fourth birthday, almost every child exhibited language labeling. Results also revealed that children's ability to label the language used by a speaker slightly differed depending on the speaker assessed. Children were reported to label their own languages(s) earlier than the languages of their caregivers and the languages of new people, despite all of them started to occur also between ages 3 and 4. These results contrast with previous research that found that children were better at labeling their caregivers' language compared to their own (Atagi & Sandhofer, 2020). However, the contrasting findings can be attributed to differences in the linguistic backgrounds of the samples included in each study. In the present study all children were functional bilinguals, while in Atagi & Sandhofer's study, the participants were primarily children exposed to a second language, but unable to speak it.

Children's language labeling ability was also found to be modulated by the presence of language reflections at home. Children frequently exposed to language-related reflections were more likely to label the languages used by speakers. However, language reflections at home were positively correlated with age. Older children were more frequently exposed to these conversations than younger children. It is possible that as children progress in their language development, caregivers engage more in language reflections with the child. To separate the effect of the two predictors, the development of children's language labeling ability was assessed across the age range separately for children exposed to low and high levels of language reflections at home. Results revealed that exposure to linguistic conversations was associated with an earlier onset of this

ability, increasing the likelihood of its appearance before age 3. These findings support the theory proposed by Akhtar et al. (2012) and Atagi & Sandhofer (2020), who attributed the bilingual advantage they found to greater exposure to language-related conversations at home.

In sum, this study broadens the understanding of how sociolinguistic awareness evolves in bilingual children. The findings indicate that the foundations for pragmatic differentiation are generally established by age 2 and that one year later, by age 3, children can explicitly label the languages used by different speakers. However, several predictors can modulate the development of these abilities. Pragmatic differentiation is affected by children's language dominance and the language strategy followed at home. Children are better at adapting their language use to the interlocutor when the language required is their dominant language, especially in early developmental stages. Moreover, children who were exposed to one of their languages at home and the other outside exhibited higher levels of pragmatic differentiation at an earlier age. Children's language labeling ability is modulated by the frequency of language reflections at home. Children from families who regularly engage in language-related conversations can label the languages used by different speakers earlier. However, after their 4<sup>th</sup> birthday, nearly all children exhibited language labeling, regardless of their exposure to these reflections.



Chapter 5:  
**GENERAL  
DISCUSSION**



## Overview

The studies presented in this thesis were designed to explore the evolution of speaker-language associations across development, with a particular focus on bilingual infants and children. The associations were assessed before the first year of life (infants aged 4 to 12 months) and afterward (children aged 1 to 5 years), with different methodological approaches based on the behavioral repertoire available at each developmental stage. Furthermore, potential linguistic factors that could modulate the formation and expression of speaker-language associations were examined.

Study 1 and Study 2 focused on the formation of audiovisual face-language associations before the first year of life by testing 4-, 6-, and 10-month-old monolingual and bilingual infants. Two types of associations were assessed using a preferential-looking paradigm: primary face-language associations by measuring infants' language-based preferences for speakers and robust face-language associations by measuring infants' ability to match the face of a speaker with their corresponding language. Besides linguistic background, the role of exposure to parental code-switching was also examined as a potential predictor. To better assess the formation of speaker-language associations under different conditions, Study 1 included two speakers using two close languages (Catalan and Spanish), and Study 2, two speakers using two distant languages (Catalan/Spanish and English). Moreover, in Study 1, infants were relatively familiar with the non-native/non-dominant language (bilinguals at home, but monolinguals in the community), while in Study 2, none of the infants had been exposed to the non-native language (i.e., English).

## Chapter 5: General Discussion

Results from Study 1 and Study 2 revealed that primary face-language associations can be formed as early as 4 months of age, evidenced by the observed preferences for speakers based on the language they previously used. Interestingly, these language-based preferences showed a developmental pattern, that differed depending on the languages included in each study. In Study 1, 4-month-old infants looked preferentially at the speaker of the native/dominant language, a preference that faded by 6 months of age. In contrast, 10-month-old infants showed a preference for the speaker of the non-native/non-dominant language. In Study 2, this familiarity to novelty preference transition was not observed. Four-month-old infants showed a preference for the speaker of the native language, but no preferences were observed at the other ages. Regarding robust face-language associations, Study 1 and Study 2 suggest that these associations may not develop before the first year of life, as infants showed no signs of matching the speakers with their corresponding languages in any of the studies. The results from both studies were consistent across monolingual and bilingual infants and were not influenced by exposure to parental code-switching.

Study 3 focused on the development of sociolinguistic awareness after the first year of life by testing 1- to 5-year-old bilingual children from two bilingual communities (i.e., Quebec and Catalonia). A new questionnaire was designed to collect parental reports on two markers of children's formation of speaker-language associations, pragmatic differentiation and language labeling. Pragmatic differentiation assesses children's ability to adjust their language use according to the language predominantly spoken by their interlocutor, and language labeling, their ability to explicitly verbalize the language(s) used by different speakers. Furthermore, the language strategy followed at home, the child's language dominance, and

## Chapter 5: General Discussion

their exposure to language reflections were considered as potential predictors that could modulate the development of these abilities.

Results from Study 3 revealed a general establishment of pragmatic differentiation around age 2, with several children showing signs of a potential earlier emergence. From age 3 onwards, no significant age-related improvements in pragmatic differentiation were observed. Interestingly, children who were exposed to one language at home and one outside showed higher levels of pragmatic differentiation from an earlier age. Moreover, children were better at adapting their language use to their interlocutor when they were required to use their dominant language. Study 3 also revealed that the language labeling ability appeared consistently around age 3, although several children exhibited this ability earlier. Children exposed to high levels of language reflections at home were more likely to label the languages used by speakers. However, the impact of language reflections was no longer relevant after age 4, when nearly all children had developed this ability. The findings related to pragmatic differentiation and language labeling were consistent across bilingual communities.

# Speaker-Language Associations

## Before the First Year of Life

During early language acquisition, infants process and extract the perceptual characteristics of their native language and form a mental representation containing its regularities. Infants growing up in bilingual contexts face the challenge of learning two languages, often simultaneously. They must extract the regularities of each language while keeping them separate to form a distinct linguistic representation for each one. The higher demands of this dual language acquisition lead bilingual infants to develop different audiovisual speech processing strategies to enhance and facilitate the process. For example, bilingual infants attend more to talking faces compared to their monolingual peers (Mercure et al., 2019). Moreover, bilingual infants pay more attention to the mouth of a speaker (Pons et al., 2015), especially those infants learning two rhythmically close languages (Birulés et al., 2019). The formation of associations between the appearance of speakers and the language they use could also be an additional cue to further support early bilingual language acquisition and separation.

## Primary Face-Language Associations

The findings from Study 1 and Study 2 regarding the formation of primary face-language associations are consistent with previous research (Colomer et al., 2023; Kinzler et al., 2007) and further support the idea that these associations are formed early in the first year of life. In both studies, infants as young as 4 months of age showed visual preferences for speakers based on the language they had previously used, meaning

## Chapter 5: General Discussion

infants were able to track who used each language. Moreover, these studies provide the first evidence of primary face-language associations in bilingual infants. Interestingly, the associations were equally observed in monolingual and bilingual infants, and the results from the bilingual infants were not influenced by their exposure to high or low levels of parental code-switching. Taken together, these results indicate that the ability to form primary face-language associations develops before the first year of life, independently of the linguistic background of the infant.

Although not the primary focus of this thesis, these studies also offer valuable insights into how language-based preferences for speakers evolve over time, and how their developmental patterns may vary depending on the languages used by the speaker being compared.

The familiarity preference observed in 4-month-old infants in Study 1 and Study 2, followed by its decline at 6 months of age, is consistent with previous research on language-based preferences (Colomer et al., 2023; Kinzler et al., 2007). Similar patterns have been reported in visual preferences for faces based on their gender (Liu, Xiao, Quinn, et al., 2015) or their race (Fassbender et al., 2016; Liu, Xiao, Xiao, et al., 2015). Young infants exhibit a preference for female and own-race faces, likely due to greater visual exposure to these social categories during early development (Rennels & Davis, 2008). However, as they accumulate more experience with these faces, the initial familiarity preference tends to decline. A similar phenomenon may arise with languages.

The results from the 10-month-old infants in Study 1 indicate a subsequent developmental shift toward a novelty preference, as they increased their attention to the speaker of the non-native/non-dominant language. This is the first time this reversed preference has been observed regarding languages, as Colomer et al. (2023) reported no preference for

any speaker in a group of 8-to-11-month-old infants. However, the novelty preference has previously been reported regarding race, with 9-month-old infants showing a visual preference for other-race faces (Fassbender et al., 2016; Liu, Xiao, Xiao, et al., 2015). This suggests that as infants become more proficient at processing familiar stimuli (i.e., own-race faces or native/dominant language), they may redirect their cognitive resources to process stimuli that are less present in their environment and are less familiar to them (i.e., other-race faces or non-native/non-dominant languages). Interestingly, the transition from familiarity to novelty preference was not observed in Study 2, as 10-month-old infants looked similarly at either speaker, consistent with the findings from Colomer et al. (2023).

The differences in the developmental patterns of language-based preferences observed across studies can be explained by two distinct, though not mutually exclusive, causes directly related to the languages included in each study.

One possibility is that the differences across studies are related to the distance between the languages tested. In Study 1, infants were tested with two rhythmically close languages: Catalan and Spanish. In contrast, in Study 2 as well as in Colomer et al. (2023), infants were tested in two distant languages: Catalan/Spanish and English or English and French, respectively. The novelty preference for non-native/non-dominant language speakers described in older infants might be only expressed if that language shares some similarities with the native/dominant language of the infant (i.e., belonging to the same rhythmic class). Speakers who use a language markedly different from the infant's native/dominant language may be more easily classified as outsiders, not attracting infants' attention. In fact, linguistic information has been reported to be a crucial

social marker for ingroup membership, sometimes even more influential than race (Kinzler et al., 2009).

Another possibility is that the differences are related to the relative familiarity with the non-native/non-dominant language. In Study 1, monolingual infants were primarily exposed to one of the languages but likely had some familiarity with the other, as both languages are widely spoken in Catalonia. Bilingual infants were learning both languages simultaneously and were therefore familiar with both. In contrast, in Study 2 as well as in Colomer et al. (2023), all infants were mainly exposed to only one of the languages and had no experience with the other. As infants become more proficient at processing their native/dominant language, they may begin to shift their attention to other languages but only if they are present in their environment (i.e., at home or in the community). A speaker using a language infants' have never heard may not be considered relevant and might not be prioritized over a speaker of the native/dominant language. Moreover, speakers who use a language completely unfamiliar to the infant may also be perceived as outsiders.

The findings from Study 1 and Study 2 not only extend previous research by describing diverse developmental patterns of language-based preferences but also represent the first exploration of these preferences in bilingual infants. Interestingly, no differences were observed between monolingual and bilingual infants, as the previously described results were consistent across both linguistic backgrounds.

The language-based preferences for speakers observed in infants before the first year of life highlight the significant influence of language on speaker perception and how it might modulate infants' social interactions from an early age. Moreover, these biases might play an important role in supporting early language acquisition. According to the PRIMIR

## Chapter 5: General Discussion

framework, infants possess early perceptual biases that guide their attention to different aspects of speech (Curtin et al., 2011; Werker & Curtin, 2005). The early preference for speakers of the native/dominant language could be considered one of these biases, leading to increased attention to the linguistic information related to that language. This would facilitate processing it and extracting and learning its characteristics. As infants gain experience with their native/dominant language, this preference starts to fade. This suggests a developmental transition, with attention gradually shifting away from speakers of the native/dominant language and toward speakers of other languages, at least when that language is present in the environment or shares some similarities with the infants' native/dominant language. According to the results of the present thesis, the preference transition is completed toward the end of the first year of life. This coincides with the development of perceptual narrowing (Kuhl et al., 2006; Werker & Tees, 1984), which further supports the idea that the preference shift might be driven by infants' increased proficiency in their native/dominant language. The attention shift favoring speakers of a non-native/non-dominant language might be especially important for bilingual language acquisition. This bias might increase infants' attention to their less dominant language, enhancing their opportunities to process and learn its linguistic characteristics, thereby supporting a more balanced language development. However, this developmental pattern may also be relevant for monolingual infants growing up in bilingual communities who are indirectly exposed to a second language. Although these infants may not be acquiring both languages at that moment, increased attention to speakers of that second language could enhance their processing abilities promoting greater sensitivity to the sounds and structures of the language, potentially supporting future language learning.

## **Robust Face-Language Associations**

The findings from Study 1 and Study 2 regarding robust face-language associations provide no evidence of their formation and retention before the first year of life. Infants showed no signs of matching speakers' faces with their corresponding languages, regardless of age or linguistic background. Additionally, for bilingual infants, no effect of exposure to parental code-switching was observed. These results are consistent even when the experimental task included easier-to-discriminate languages from different rhythmic classes (i.e., Study 2).

The obtained results are in line with previous research by Schott et al. (2023). This study explored speaker-language associations in 5-, 12-, and 18-month-old monolingual and bilingual infants using a familiarization-switch paradigm. The authors did not observe evidence of infants detecting a change in previously familiarized speaker-language pairings, not even when assessing 18-month-old infants. However, the results could have been partially caused by how the associations were tested, as infants from bilingual environments may be used to speakers switching languages and might not react to such changes, even if they know which language corresponds to each speaker. To avoid relying on infants' reactions to a speaker changing from one language to another, the studies included in the present thesis used a different experimental approach (i.e., a preferential-looking paradigm). The absence of robust face-language associations regardless of the task used suggests that forming and retaining these associations may be too cognitively demanding for such young infants. This hypothesis is consistent with the findings obtained for face-voice associations. Infants are unable to match the face of a familiarized speaker with their corresponding voice until 17 months of age, and only if there is gender information available (Fecher et al., 2019). When

## Chapter 5: General Discussion

assessing speakers of the same gender, these associations are not observed until 24 months of age (Orena et al., 2022).

Study 1 and Study 2 also extend previous research by assessing infants from another bilingual community learning a different language pair. Schott et al. (2023) tested infants from Montreal exposed to two distant languages: French and English. In contrast, the studies from the present thesis focused on infants from Catalonia exposed to two close languages: Catalan and Spanish. However, no bilingual advantage was observed regardless of the community, or the pair of languages being learned. Robust face-language associations were not observed even in bilingual infants learning two close languages, despite the potential benefits the associations might offer as additional visual cues to further support the separation of two similar languages (see also Birulés et al., 2019). Although no bilingual advantage was observed, it remains possible that such a benefit may emerge later in life. The associations may have been assessed too early in development to observe their formation regardless of the potential effect an infant's linguistic background may have.

In summary, the results from Study 1 and Study 2 combined with previous research indicate that, although infants before the first year of life can track the languages spoken by different speakers, they are still unable to form robust face-language associations and retain them.

# Speaker-Language Associations

## After the First Year of Life

After the initial stages of language acquisition, knowing the language used by speakers in the environment is still crucial for bilingual children. This sociolinguistic awareness will allow them to select the most appropriate language to use depending on the context or the interlocutor they encounter, while also giving them a deeper understanding of the social significance of language and its role in shaping social interactions.

### Pragmatic Differentiation

First, Study 3 revealed no significant differences across communities regarding pragmatic differentiation, indicating that the development of this ability was similar in Catalonia and Quebec.

The findings from Study 3 support previous research describing the general establishment of pragmatic differentiation around age 2 (Genesee et al., 1995; Goodz, 1989; Lanza, 1992; Montanari, 2008; Nicoladis & Genesee, 1996; Quay, 2008). This indicates that at that age most children are already aware of the languages used by different speakers and have the required cognitive and linguistic skills to adapt their language use. Before the second birthday, many of the tested children were not yet able to use both of their languages, preventing them from expressing pragmatic differentiation regardless of whether they knew which language corresponds to each speaker. The performance of the children who were already able to speak showed considerable variability, with several of them exhibiting levels of pragmatic differentiation comparable to those of older

children, in line with Nicoladis (1998). These findings suggest that the main limitation for the emergence of pragmatic differentiation is more likely to be related to children's language proficiency rather than to language confusion or not knowing the corresponding language for each speaker. From age 3 onwards, no further improvement in pragmatic differentiation was observed that could be attributed to age. This developmental pattern was also suggested by the findings from Tare & Gelman (2010), where no differences in pragmatic differentiation were observed between 3- and 4-year-old bilingual children.

Regarding the effect of language strategies, Study 3 is the first to assess the development of pragmatic differentiation in children from families following diverse strategies other than the one parent one language. The 'One parent one language', 'One bilingual parent' and 'Two bilingual parents' strategies provided very similar results, with a relatively gradual increase of pragmatic differentiation throughout children's development. This finding could suggest that, in general, the developmental trajectory of pragmatic differentiation is not strongly influenced by the patterns of language use children are exposed to at home. However, significant differences were observed across these strategies and the 'One language at home one outside' strategy. Results revealed that children who were exposed to this strategy showed higher levels of pragmatic differentiation from an earlier age. This pragmatic differentiation advantage could be related to different aspects associated with the strategy.

One potential explanation could be the low levels of language mixing observed in caregivers following the 'One language at home one outside' strategy. Previous research has observed that children exposed to higher levels of parental language mixing exhibited more mixing themselves, reducing their pragmatic differentiation (Genesee, 1989; Goodz, 1989). In

contrast, strategies providing lower levels of language mixing would be expected to enhance children's pragmatic differentiation. However, similar levels of language mixing were reported from caregivers following the 'One parent one language' strategy but no advantage in pragmatic differentiation was observed for this strategy. Additionally, children whose caregivers followed the 'Two bilingual parents' strategy, characterized by constant language mixing at home, would be expected to exhibit lower levels of pragmatic differentiation. Nonetheless, the results for these children were comparable to those in the 'One parent one language' and 'One bilingual parent' strategies. Taken together, these results suggest it is unlikely that the pragmatic differentiation advantage associated with the 'One language at home one outside' strategy is exclusively related to the caregivers' level of language mixing.

An alternative explanation for the 'One language at home one outside' strategy advantage could be related to the fact that in this approach languages are separated not only by the speakers who use them but also by the broader contexts in which they are used. This contextual separation may enhance children's sociolinguistic awareness, potentially fostering their ability to adapt their language use to different interlocutors.

Lastly, it should also be considered that the differences found between the different family language strategies could be partially due to how pragmatic differentiation is measured in young bilingual children. This study, as many others before, mainly assessed pragmatic differentiation by exploring how children interact and use their languages with their caregivers. This approach provides valuable insights into children's pragmatic differentiation abilities when each caregiver uses a different language, as it allows for an examination of how children use their two languages with two familiar speakers. For this reason, most of the studies

## Chapter 5: General Discussion

exploring pragmatic differentiation assessed children being raised in families following the 'One parent one language' strategy. However, measuring infants' language adaptation with the caregivers is less informative when both use the same language, as in the 'One language at home one outside' strategy. In that case, children only need to use one of their languages to successfully communicate with both their caregivers, presenting a much less challenging situation compared to having to actively select and switch between languages depending on the caregiver they are addressing. Furthermore, it cannot be determined whether children are using the appropriate language because they know it is the language their caregivers usually use, or if they are using it by default because it is the language to which they have been more frequently exposed. These factors, or a combination of them, could have influenced the results, potentially leading to an overestimation of the pragmatic differentiation in children from families following the 'One language at home one outside' strategy.

The children's dominant language was also found to be an important predictor of their ability to adapt the language they use to the interlocutor. In line with previous research (Genesee, 1989; Genesee et al., 1995, 1996; Nicoladis, 1998; Nicoladis & Genesee, 1996), children exhibited higher levels of pragmatic differentiation when the required language was their dominant language, in comparison to when the required language was their non-dominant. Interestingly, the differences across languages were present but less pronounced in older children. As children grow up and gain more experience with their languages, it becomes more likely that they acquire sufficient proficiency in their non-dominant language to use it with similar ease as their dominant language. This would make them more capable of adjusting their language use effectively, regardless of the language required. These findings highlight again the importance of

language proficiency in the emergence and expression of pragmatic differentiation.

The use of parental report measures in Study 3 allowed for the exploration of a large sample of children across a wide age range. This increased the chances of capturing general developmental trends rather than anecdotal data, enabling the extraction of more robust and generalizable conclusions. This had been a significant limitation in previous research on pragmatic differentiation, which relied on small samples due to the lengthy process of collecting experimental data on the development of this ability. Crucially, the findings obtained in Study 3 are in line with previous experimental research, suggesting that parental reports can be a good alternative to assess pragmatic differentiation in a much less time-consuming manner.

### **Language Labeling**

As before, Study 3 revealed no significant differences across communities regarding language labeling, indicating that the development of this ability was similar in Catalonia and Quebec.

The findings from Study 3 were consistent with previous research describing the emergence of language labeling around age 3 (Akhtar et al., 2012; Atagi & Sandhofer, 2020). Nevertheless, the results also indicated that, in some children, language labeling can appear earlier, months before the third birthday. Moreover, after age 4, nearly all children exhibited this ability.

Children's language labeling ability seemed to differ depending on the speaker whose language they were asked to label. While in Atagi & Sandhofer (2020) the authors observed that children were better at

## Chapter 5: General Discussion

labeling their caregivers' language compared to their own, the results from Study 3 revealed the opposite pattern. These contrasting findings can be attributed to differences in the linguistic backgrounds of the samples included in each study. In Study 3 all children were functional bilinguals, while in Atagi & Sandhofer's study, the participants were primarily monolingual children with different degrees of exposure to a second language. Actively using and switching between two languages is an integral part of bilingual children's communicative process, which might enhance their language awareness about their own experience promoting language labeling. In contrast, children who speak only one language are less likely to reflect on it, and caregivers may not emphasize its label as much. However, if these monolingual children are exposed to people using more than one language at home or in the community, this experience may draw their attention increasing their language awareness regarding their sociolinguistic context. Taken together, these findings suggest that different linguistic experiences can enhance the development of different aspects of language awareness and that speaking two languages is not always necessary to achieve these benefits.

Language labeling was also observed to be modulated by children's exposure to language reflections at home. Those children whose families engaged more in conversations regarding languages and bilingualism were more likely to label the languages used by different speakers compared to children who were not exposed to these conversations. Interestingly, the presence of language reflections at home was also associated with an earlier emergence of the language labeling ability. The impact of language reflections was no longer relevant after age 4, when nearly all children had developed the ability regardless of their exposure to these reflections. These findings are in line with the theory proposed by Akhtar et al. (2012) and Atagi & Sandhofer (2020) to explain the bilingual advantage they found

## Chapter 5: General Discussion

in language labeling. The authors attributed the advantage to potentially greater engagement in language-related conversations in bilingual homes. Moreover, the results from Study 3 indicate that, although language reflections are associated with enhanced language labeling abilities, these conversations are not evenly distributed in bilingual homes. This highlights the diversity of the bilingual experience and the impact it can have on bilingual development. It also suggests that certain experiences can enhance children's language awareness even without exposure to two languages, which could be also valuable for supporting the language acquisition process of monolingual children. However, it's important to keep in mind that the findings are based on correlational data. While the idea that exposure to language reflections stimulates children's language labeling makes theoretical sense, it is also possible that children with higher levels of language awareness, those who can label the languages of people around them, might prompt more language-related conversations from their caregivers.

# Developmental Trajectory of Speaker-Language Associations

The three studies included in the present thesis provide an outline of the developmental trajectory of speaker-language associations throughout the bilingual language acquisition process, from 4 months to 5 years of age.

Before the first year of life, infants as young as 4 months of age demonstrated to be able to track the language used by different speakers, evidenced by the expression of language-based preferences for speakers. This guides their selective attention toward speakers, influencing their social interactions and potentially their language acquisition process. However, infants seemed unable to match each speaker with the language they used. If infants under one year of age cannot retain who speaks which language, sticking to speaking only one language to the child may not be as effective in facilitating bilingual infants' language separation as previously anticipated (Barron-Hauwaert, 2004). Accordingly, strategies like the one parent one language may not be particularly advantageous, at least in the early stages of bilingual development. Bilingual infants may rely on other aspects to support language separation, like paying special attention to the distinctive auditory characteristics of each language rather than to the person who is using the language. Nevertheless, the studies from the present thesis examined the formation of face-language associations with unfamiliar speakers. Infants might be able to learn and retain the language used by speakers from their environment, especially their caregivers. This may occur as a result of cumulative experience with

their face-language combination. Short exposures to new speakers might not been enough for infants to form an association and retain the specific face-language pairing, at least before the first year of life.

After the first year of life, children start exhibiting different behaviors that serve as indicators of the formation of speaker-language associations. Pragmatic differentiation emerges generally around age 2, which indicates that by that age bilingual children know what language speakers predominantly use and have the required linguistic abilities to adapt their language use accordingly. Considering the absence of robust speaker-language associations before the first year of life, it would be possible that these associations develop at some point between age 1 and age 2.

However, there are big individual differences regarding the emergence and the expression of pragmatic differentiation. Despite almost half of the 1-year-old infants showed no pragmatic differentiation due to very limited verbal skills, several of those who were already able to speak showed levels of language use adaptation comparable to those of older children. Moreover, children were better at pragmatic differentiation when they were required to use their dominant language rather than their non-dominant. Taken together, these findings suggest that linguistic proficiency and the lexical availability in each language may be one of the most important factors influencing both the age at which pragmatic differentiation emerges and its level of adequacy. Potentially, children might form speaker-language associations soon after their first birthday but lack the linguistic skills to express them by adapting their language use to their interlocutor.

By age 3, children are able to verbally indicate the languages spoken by different individuals. These explicit reflections are an undeniable sign that children have already developed well-established associations between

## Chapter 5: General Discussion

speakers and the languages they use. Moreover, this skill helps them become more attuned to their sociolinguistic environment, deepening their understanding of how languages are connected to individuals and society.

## Limitations and Future Directions

The studies presented in this thesis represent the first comprehensive examination of the developmental trajectory of speaker-language associations throughout bilingual development, spanning from 4 months to 5 years of age. While the previously discussed findings make a significant contribution to the field, they also have certain limitations that must be addressed and considered in future research.

The first limitation is related to the language-based preferences for speakers observed within the first year of life. Despite the exploration of the developmental trajectory of these preferences was not one of the main goals of the thesis, results from Study 1 and Study 2 revealed an interesting pattern that differed depending on what languages were used by the speakers. While in Study 1 10-month-old infants showed a novelty preference for the speaker of the non-native/non-dominant language, this transition was not observed in Study 2 where 10-month-old infants showed no preference. As previously discussed, the observed differences might be attributed to the fact that the non-native/non-dominant language belonged to a different rhythmic class compared to the infants' native/dominant language or to the infants' lack of prior experience with that language. To determine which factor is responsible for the differences, future research should explore the development of these language-based preferences by testing infants with two distant but familiar languages or two close languages, with one of them being unfamiliar to the infants. This would reveal which aspects are more relevant when assessing the influence of language on speaker perception.

## Chapter 5: General Discussion

Findings from Study 1 and Study 2 also revealed that infants before the first year of life were unable to retain the face-language pairings and match each speaker with their corresponding language. This ability was assessed in the context of learning the face-language combinations of two unfamiliar speakers. Accordingly, these findings might not be generalizable to the formation and retention of face-language associations of familiar speakers, such as the infants' caregivers. Infants in their first year of life may have already associated a language with the speakers they are frequently exposed to. To evaluate this theory, the formation of face-language associations should be assessed with infants' caregivers. Future research aiming to address this gap in the literature should consider two main technical challenges associated with this approach. First, it requires recording stimuli from both caregivers for each participant, significantly increasing the time needed to prepare the experimental task. Second, in many cases, caregivers are a couple consisting of one male and one female. In these situations, it would be unclear whether infants are matching each speaker's face with the corresponding language or if they are relying on the audiovisual gender correspondence between the faces and the voices.

The findings from Study 3 indicate that by age 2, children can form speaker-language associations, as evidenced by their ability to adapt their language use to match the language predominantly spoken by an interlocutor. The results also suggest that these associations might be established even earlier but are not expressed due to children's limited verbal skills before age 2. However, the data collected for this thesis do not provide sufficient evidence to confirm this hypothesis. Subsequent research should explore this possibility by using measures of speaker-language associations that do not rely on children's language proficiency.

## Chapter 5: General Discussion

One approach could involve assessing audiovisual associations through attention-based tasks, replicating Study 1 and Study 2 with older infants.

Understanding how different language strategies influence bilingual development is crucial, as it is a major concern for caregivers raising children who are learning two languages. Study 3 suggests that certain strategies may modulate children's ability to adapt their language use to their interlocutor. Specifically, learning one language at home and the other outside seemed to enhance children's pragmatic differentiation. However, the underlying aspects related to this potential advantage remain unclear. Moreover, the obtained results may be influenced by the way pragmatic differentiation was measured in Study 3, where most questions examined children's language use when interacting with their caregivers, who both used the same language. Future research could address this issue by assessing children's interactions with other familiar speakers who did not use the same language as the caregivers, allowing for the evaluation of pragmatic differentiation across the two of the children's languages.

Lastly, the use of parental report measures may present certain limitations. The accuracy of these reports can vary significantly compared to other experimental measures, depending on the aspects being assessed. Although the findings related to pragmatic differentiation and language labeling obtained from parental reports are consistent with previous research, future studies should directly compare questionnaire results with experimental data from the same children. This approach would help evaluate the validity of the parent-report items developed in this study and confirm whether these abilities can be reliably assessed through parental reports.

## Chapter 5: General Discussion

Chapter 6:  
**CONCLUSIONS**



# Conclusions

This thesis provides an outline of the developmental trajectory of speaker-language associations throughout language acquisition, with a particular focus on bilingual infants and children. These associations were assessed at two developmental stages: before the first year of life, by measuring the formation of audiovisual face-language associations in infants from 4 to 12 months, and afterward, by collecting parental reports on pragmatic differentiation and language labeling in children from 1 to 5 years. Furthermore, potential factors that could modulate the formation and expression of the associations were also assessed.

More specifically, three research questions were addressed: 1) whether the ability to form face-language associations emerges within the first year of life, and if it is influenced by the linguistic background of the infant or their exposure to parental code-switching, 2) the general developmental trajectory of pragmatic differentiation, and if it is influenced by children's language dominance and the family language strategy followed at home, and 3) the general developmental trajectory of language labeling, and if it is influenced by exposure to language reflections at home.

Regarding the first research question, the findings from the present thesis indicate that primary face-language associations can be formed as early as 4 months of age, but that more robust associations may not develop until after the first year of life. Moreover, the formation of these associations is not influenced by infants' linguistic background or their exposure to parental code-switching. Interestingly, the present findings also revealed a developmental pattern regarding language-based preferences for speakers. While 4-month-old infants show a visual

## Chapter 6: Conclusions

preference for speakers that use their native or dominant language, at 10 months of age infants prefer to look at speakers that use a language that is non-native or non-dominant to them. However, this reversed preference only seems to be expressed when that language is present in the environment or shares some similarities with the infants' native language. This developmental pattern may have significant implications for language acquisition, particularly for bilingual infants, as it helps ensure more balanced exposure to both languages being learned throughout their development.

Concerning the second research question, the obtained results confirm a general establishment of pragmatic differentiation around age 2. However, almost half of the children under that age showed an earlier emergence as long as their language production skills allowed it. These findings suggest that the main limitation for this language adaptation is more likely to be related to children's language proficiency rather than to them not knowing what language corresponds to each speaker. From age 3 onwards, no significant age-related improvements in pragmatic differentiation can be observed. Interestingly, children exposed to one language at home and one outside show higher levels of pragmatic differentiation from an earlier age, although the reasons for this advantage require further exploration. Moreover, children are better at adapting their language use to their interlocutor when they are required to use their dominant language.

Lastly, regarding the third research question, the findings revealed that language labeling starts appearing consistently around age 3, although several children exhibited this ability earlier. Children exposed to high levels of language reflections at home are more likely to label the languages used by speakers. However, the impact of language reflections

## Chapter 6: Conclusions

is no longer relevant after age 4, when nearly all children have developed this ability.

In conclusion, the findings indicate that, early on, infants can track the languages spoken by others, influencing their social interactions and potentially their language acquisition. Although robust speaker-language associations are not evident before the first year of life, they evolve gradually and become more complex throughout development. Once children begin speaking, these associations guide their language use and enable them to explicitly identify the languages present in their sociolinguistic environment. Moreover, this thesis highlights the importance of certain linguistic factors, though not all, in the formation and expression of speaker-language associations.

## Chapter 6: Conclusions

Chapter 7:  
**REFERENCES**



# References

- Abercrombie, D. (1967). Elements of General Phonetics. In *Elements of General Phonetics*. Aldine Atherton.  
<https://doi.org/10.1515/9781474463775>
- Akhtar, N., Menjivar, J., Hoicka, E., & Sabbagh, M. A. (2012). Learning foreign labels from a foreign speaker: the role of (limited) exposure to a second language. *Journal of Child Language*, 39(5), 1135–1149.  
<https://doi.org/10.1017/S0305000911000481>
- Atagi, N., & Sandhofer, C. M. (2020). Early Language Environments Predict Aspects of Explicit Language Awareness Development. *Language Learning*, 70(2), 464–505.  
<https://doi.org/10.1111/LANG.12381>
- Baglin, J. (2014). Improving your exploratory factor analysis for ordinal data: A demonstration using FACTOR. *Practical Assessment, Research and Evaluation*, 19(5), 5.  
<https://doi.org/https://doi.org/10.7275/dsep-4220>
- Bahrack, L. E., Hernandez-Reif, M., & Flom, R. (2005). The Development of Infant Learning About Specific Face-Voice Relations. *Developmental Psychology*, 41(3), 541–552.  
<https://doi.org/10.1037/0012-1649.41.3.541>
- Bahrack, L. E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental Psychology*, 36(2), 190–201.  
<https://doi.org/10.1037//0012-1649.36.2.190>
- Bahrack, L. E., & Lickliter, R. (2012). The role of intersensory redundancy in early perceptual, cognitive, and social development. *Multisensory Development*.  
<https://doi.org/10.1093/ACPROF:OSO/9780199586059.003.0008>
- Bahrack, L. E., Netto, D., & Hernandez-Reif, M. (1998). Intermodal Perception of Adult and Child Faces and Voices by Infants. *Child Development*, 69(5), 1263–1275. <https://doi.org/10.1111/j.1467-8624.1998.tb06210.x>
- Barrera, M. E., & Maurer, D. (1981). Discrimination of Strangers by the Three-Month-Old. *Child Development*, 52(2), 558.  
<https://doi.org/10.2307/1129174>

- Barron-Hauwaert, S. (2004). Language Strategies for Bilingual Families: The One-Parent-One-Language Approach. In *Language Strategies for Bilingual Families: The One-Parent-One-Language Approach*. <https://doi.org/10.1017/s0272263105250283>
- Berko, J. (1958). The Child's Learning of English Morphology. *WORD*, 14(2–3), 150–177. <https://doi.org/10.1080/00437956.1958.11659661>
- Birulés, J., Bosch, L., Brieke, R., Pons, F., & Lewkowicz, D. J. (2019). Inside bilingualism: Language background modulates selective attention to a talker's mouth. *Developmental Science*, 22(3), e12755. <https://doi.org/10.1111/DESC.12755>
- Bloom, L., & Lahey, M. (1979). Language Development and Language Disorders. *Language*, 55(4), 945.
- Boisferon, A. H. de, Kubicek, C., Gervain, J., Schwarzer, G., Loevenbruck, H., Vilain, A., Fort, M., Méary, D., & Pascalis, O. (2021). Language familiarity influences own-race face recognition in 9- and 12-month-old infants. *Infancy*, 26(4), 647–659. <https://doi.org/10.1111/INFA.12404>
- Bosch, L. (2024). Acquisition of segmental phonology in child bilingualism. In *The Cambridge Handbook of Bilingual Phonetics and Phonology* (pp. 339–358). M. Amengual (Ed.).
- Bosch, L., & Ramon-Casas, M. (2014). First translation equivalents in bilingual toddlers' expressive vocabulary: Does form similarity matter? *International Journal of Behavioral Development*, 38(4), 317–322. <https://doi.org/10.1177/0165025414532559>
- Bosch, L., & Sebastián-Gallés, N. (1997). Native-language recognition abilities in 4-month-old infants from monolingual and bilingual environments. *Cognition*, 65(1), 33–69. [https://doi.org/10.1016/S0010-0277\(97\)00040-1](https://doi.org/10.1016/S0010-0277(97)00040-1)
- Bosch, L., & Sebastián-Gallés, N. (2001). Evidence of Early Language Discrimination Abilities in Infants From Bilingual Environments. *Infancy*, 2(1), 29–49. [https://doi.org/10.1207/S15327078IN0201\\_3](https://doi.org/10.1207/S15327078IN0201_3)
- Brookes, H., Slater, A., Quinn, P. C., Lewkowicz, D. J., Hayes, R., & Brown, E. (2001). Three-month-old infants learn arbitrary auditory-visual pairings between voices and faces. *Infant and Child Development*, 10(1–2), 75–82. <https://doi.org/10.1002/icd.249>
- Byers-Heinlein, K. (2013). Parental language mixing: Its measurement

- and the relation of mixed input to young bilingual children's vocabulary size\*. *Bilingualism: Language and Cognition*, 16(1), 32–48. <https://doi.org/10.1017/S1366728912000120>
- Byers-Heinlein, K. (2015). Methods for studying infant bilingualism. In *The Cambridge Handbook of Bilingual Processing* (pp. 133–154). Cambridge University Press. <https://doi.org/10.1017/cbo9781107447257.005>
- Byers-Heinlein, K., Burns, T. C., & Werker, J. F. (2010). The Roots of Bilingualism in Newborns. *Psychological Science*, 21(3), 343–348. <https://doi.org/10.1177/0956797609360758>
- Byers-Heinlein, K., Morin-Lessard, E., & Lew-Williams, C. (2017). Bilingual infants control their languages as they listen. *Proceedings of the National Academy of Sciences of the United States of America*, 114(34), 9032–9037. <https://doi.org/https://doi.org/10.1073/pnas.1703220114>
- Byers-Heinlein, K., & Werker, J. F. (2009). Monolingual, bilingual, trilingual: Infants' language experience influences the development of a word-learning heuristic. *Developmental Science*, 12(5), 815–823. <https://doi.org/10.1111/J.1467-7687.2009.00902.X>
- Carter, R. (2003). Language awareness. *ELT Journal*, 57(1), 64–65. <https://doi.org/10.1093/ELT/57.1.64>
- Clerc, O., Fort, M., Schwarzer, G., Krasotkina, A., Vilain, A., Méary, D., Lœvenbruck, H., & Pascalis, O. (2022). Can language modulate perceptual narrowing for faces? Other-race face recognition in infants is modulated by language experience. *International Journal of Behavioral Development*, 46(2), 83–90. <https://doi.org/10.1177/01650254211053054>
- Colomer, M., Hwang, H. G., Burke, N., & Woodward, A. (2023). Development of infants' preferential looking toward native language speakers across distinct social contexts. *Developmental Psychology*. <https://doi.org/10.1037/DEV0001669>
- Curtin, S., Byers-Heinlein, K., & Werker, J. F. (2011). Bilingual beginnings as a lens for theory development: PRIMIR in focus. *Journal of Phonetics*, 39(4), 492–504. <https://doi.org/10.1016/J.WOCN.2010.12.002>
- David, A., & Wei, L. (2008). Individual Differences in the Lexical Development of French–English Bilingual Children. *International Journal of Bilingual Education and Bilingualism*, 11(5), 598–618.

<https://doi.org/10.1080/13670050802149200>

- De Houwer, A. (2007). Parental language input patterns and children's bilingual use. *Applied Psycholinguistics*, 28(3), 411–424.  
<https://doi.org/10.1017/S0142716407070221>
- DeAnda, S., Bosch, L., Poulin-Dubois, D., Zesiger, P., & Friend, M. (2016). The Language Exposure Assessment Tool: Quantifying Language Exposure in Infants and Children. *Journal of Speech, Language, and Hearing Research : JSLHR*, 59(6), 1346.  
[https://doi.org/10.1044/2016\\_JSLHR-L-15-0234](https://doi.org/10.1044/2016_JSLHR-L-15-0234)
- Decasper, A. J., & Fifer, W. P. (1980). Of Human Bonding: Newborns Prefer their Mothers' Voices. *New Series*, 208(4448), 1174–1176.
- Decasper, A. J., & Prescott, P. A. (1984). Human newborns' perception of male voices: Preference, discrimination, and reinforcing value. *Developmental Psychobiology*, 17(5), 481–491.  
<https://doi.org/10.1002/DEV.420170506>
- Dodd, B. (1979). Lip reading in infants: Attention to speech presented in- and out-of-synchrony. *Cognitive Psychology*, 11(4), 478–484.  
[https://doi.org/10.1016/0010-0285\(79\)90021-5](https://doi.org/10.1016/0010-0285(79)90021-5)
- Fassbender, I., Teubert, M., & Lohaus, A. (2016). The development of preferences for own-race versus other-race faces in 3-, 6- and 9-month-old Caucasian infants. *European Journal of Developmental Psychology*, 13(1), 152–165.  
<https://doi.org/10.1080/17405629.2015.1073585>
- Fecher, N., & Johnson, E. K. (2018). The native-language benefit for talker identification is robust in 7.5-month-old infants. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 44(12), 1911–1920. <https://doi.org/10.1037/XLM0000555>
- Fecher, N., & Johnson, E. K. (2019a). Bilingual infants excel at foreign-language talker recognition. *Developmental Science*, 22(4).  
<https://doi.org/10.1111/DESC.12778>
- Fecher, N., & Johnson, E. K. (2019b). By 4.5 Months, Linguistic Experience Already Affects Infants' Talker Processing Abilities. *Child Development*, 90(5), 1535–1543.  
<https://doi.org/10.1111/CDEV.13280>
- Fecher, N., & Johnson, E. K. (2022). Revisiting the talker recognition advantage in bilingual infants. *Journal of Experimental Child Psychology*, 214, 105276.

<https://doi.org/10.1016/J.JECP.2021.105276>

- Fecher, N., Paquette-Smith, M., & Johnson, E. K. (2019). Resolving the (Apparent) Talker Recognition Paradox in Developmental Speech Perception. *Infancy, 24*(4), 570–588.  
<https://doi.org/10.1111/infa.12290>
- Field, T. M., Cohen, D., Garcia, R., & Greenberg, R. (1984). Mother-stranger face discrimination by the newborn. *Infant Behavior and Development, 7*(1), 19–25. [https://doi.org/10.1016/S0163-6383\(84\)80019-3](https://doi.org/10.1016/S0163-6383(84)80019-3)
- Fifer, W. P., & Moon, C. (1989). Psychobiology of newborn auditory preferences. *Seminars in Perinatology, 13*(5), 430–433.
- Frank, M. C., Vul, E., & Saxe, R. (2012). Measuring the Development of Social Attention Using Free-Viewing. *Infancy, 17*(4), 355–375.  
<https://doi.org/10.1111/j.1532-7078.2011.00086.x>
- Genesee, F. (1989). Early bilingual development: One language or two? *Journal of Child Language, 16*(1), 161–179.  
<https://doi.org/10.1017/S0305000900013490>
- Genesee, F., Boivin, I., & Nicoladis, E. (1996). Talking with strangers: A study of bilingual children's communicative competence. *Applied Psycholinguistics, 17*(4), 427–442.  
<https://doi.org/10.1017/s0142716400008183>
- Genesee, F., Nicoladis, E., & Paradis, J. (1995). Language differentiation in early bilingual development. *Journal of Child Language, 19*(3), 633–658. <https://doi.org/10.1017/S0305000900011600>
- Goodz, N. S. (1989). Parental Language Mixing in Bilingual Families. *Infant Mental Health Journal, 10*(1), 25–44.  
<https://psycnet.apa.org/record/1989-32403-001>
- Grammont, M. (1902). Observations sur le langage des enfants. In *Mélanges linguistiques* (pp. 61–82).  
<https://www.mpi.nl/publications/item2326062/observations-sur-le-langage-des-enfants>
- Graven, S. N., & Browne, J. V. (2008). Auditory Development in the Fetus and Infant. *Newborn and Infant Nursing Reviews, 8*(4), 187–193.  
<https://doi.org/10.1053/J.NAINR.2008.10.010>
- Heron-Delaney, M., Damon, F., Quinn, P. C., Méary, D., Xiao, N. G., Lee, K., & Pascalis, O. (2017). An adult face bias in infants that is modulated by face race. *International Journal of Behavioral*

- Development*, 41(5), 581–587.  
<https://doi.org/10.1177/0165025416651735>
- Hillairiet de Boisferon, A., Dupierrix, E., Quinn, P. C., Løevenbruck, H., Lewkowicz, D. J., Lee, K., & Pascalis, O. (2015). Perception of Multisensory Gender Coherence in 6- and 9-Month-Old Infants. *Infancy*, 20(6), 661–674. <https://doi.org/10.1111/inf.12088>
- Hulstijn, J. H. (2005). Theoretical and empirical issues in the study of implicit and explicit second-language learning: Introduction. *Studies in Second Language Acquisition*, 27(2), 129–140. <https://doi.org/10.1017/S0272263105050084>
- Johnson, E. K., Westrek, E., Nazzi, T., & Cutler, A. (2011). Infant ability to tell voices apart rests on language experience. *Developmental Science*, 14(5), 1002–1011. <https://doi.org/10.1111/j.1467-7687.2011.01052.x>
- Kandhadai, P., Danielson, D. K., & Werker, J. F. (2014). Culture as a binder for bilingual acquisition. *Trends in Neuroscience and Education*, 3(1), 24–27. <https://doi.org/10.1016/J.TINE.2014.02.001>
- Kandhadai, P., Hall, D. G., & Werker, J. F. (2017). Second label learning in bilingual and monolingual infants. *Developmental Science*, 20(1), 1–14. <https://doi.org/10.1111/DESC.12429>
- Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L., & Pascalis, O. (2007). The other-race effect develops during infancy: Evidence of perceptual narrowing. *Psychological Science*, 18(12), 1084–1089. <https://doi.org/10.1111/j.1467-9280.2007.02029.x>
- Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences of the United States of America*, 104(30), 12577–12580. <https://doi.org/10.1073/PNAS.0705345104>
- Kinzler, K. D., Shutts, K., Dejesus, J., & Spelke, E. S. (2009). Accent trumps race in guiding children’s social preferences. *Social Cognition*, 27(4), 623–634. <https://doi.org/10.1521/soco.2009.27.4.623>
- Kinzler, K. D., Shutts, K., & Spelke, E. S. (2012). *Language-based social preferences among children in South Africa*. <https://doi.org/10.1080/15475441.2011.583611>
- Krasotkina, A., Götz, A., Höhle, B., & Schwarzer, G. (2018). Perceptual narrowing in speech and face recognition: Evidence for intra-

- individual cross-domain relations. *Frontiers in Psychology*, 9(SEP).  
<https://doi.org/10.3389/FPSYG.2018.01711/PDF>
- Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science (New York, N.Y.)*, 218(4577), 1138–1141.  
<https://doi.org/10.1126/SCIENCE.7146899>
- Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, 9(2). <https://doi.org/10.1111/J.1467-7687.2006.00468.X>
- Lansing, C. R., & McConkie, G. W. (2003). Word identification and eye fixation locations in visual and visual-plus-auditory presentations of spoken sentences. *Perception and Psychophysics*, 65(4), 536–552.  
<https://doi.org/10.3758/BF03194581>
- Lanza, E. (1992). Can bilingual two-year-olds code-switch? *Journal of Child Language*, 19(3), 633–658.  
<https://doi.org/10.1017/S0305000900011600>
- Lecanuet, J.-P., Granier-Deferre, C., Jacquet, A.-Y., Capponi, I., & Ledru, L. (1993). Prenatal discrimination of a male and a female voice uttering the same sentence. *Early Development and Parenting*, 2(4), 217–228. <https://doi.org/10.1002/EDP.2430020405>
- Legacy, J., Reider, J., Crivello, C., Kuzyk, O., Friend, M., Zesiger, P., & Poulin-Dubois, D. (2017). Dog or Chien? Translation Equivalents in the Receptive and Expressive Vocabularies of Young French-English Bilinguals. *Journal of Child Language*, 44(4), 881.  
<https://doi.org/10.1017/S0305000916000295>
- Levi, S. V., & Pisoni, D. B. (2007). Indexical and Linguistic Channels in Speech Perception. In *Psycholinguistic Phenomena in Marketing Communications* (pp. 203–220).  
<https://doi.org/10.4324/9781003064411-11>
- Lewkowicz, D. J., & Hansen-Tift, A. M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences of the United States of America*, 109(5), 1431–1436.  
<https://doi.org/10.1073/PNAS.1114783109>
- Liu, S., Xiao, N. G., Quinn, P. C., Zhu, D., Ge, L., Pascalis, O., & Lee, K. (2015). Asian infants show preference for own-race but not other-race female faces: The role of infant caregiving arrangements. *Frontiers in Psychology*, 6(MAY), 593.

<https://doi.org/10.3389/fpsyg.2015.00593>

- Liu, S., Xiao, W. S., Xiao, N. G., Quinn, P. C., Zhang, Y., Chen, H., Ge, L., Pascalis, O., & Lee, K. (2015). Development of visual preference for own- versus other-race faces in infancy. *Developmental Psychology*, *51*(4), 500–511. <https://doi.org/10.1037/A0038835>
- Macchi Cassia, V., Bulf, H., Quadrelli, E., & Proietti, V. (2014). Age-related face processing bias in infancy: evidence of perceptual narrowing for adult faces. *Developmental Psychobiology*, *56*(2), 238–248. <https://doi.org/10.1002/DEV.21191>
- May, L., Baron, A. S., & Werker, J. F. (2019). Who can speak that language? Eleven-month-old infants have language-dependent expectations regarding speaker ethnicity. *Developmental Psychobiology*, *61*(6), 859–873. <https://doi.org/10.1002/DEV.21851>
- 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. [https://doi.org/10.1016/0010-0277\(88\)90035-2](https://doi.org/10.1016/0010-0277(88)90035-2)
- Mercure, E., Kushnerenko, E., Goldberg, L., Bowden-Howl, H., Coulson, K., Johnson, M. H., & MacSweeney, M. (2019). Language experience influences audiovisual speech integration in unimodal and bimodal bilingual infants. *Developmental Science*, *22*(1), e12701. <https://doi.org/10.1111/DESC.12701>
- Miller, C. L. (1983). Developmental changes in male/female voice classification by infants. *Infant Behavior and Development*, *6*(2–3), 313–330. [https://doi.org/10.1016/S0163-6383\(83\)80040-X](https://doi.org/10.1016/S0163-6383(83)80040-X)
- Miller, C. L., Younger, B. A., & Morse, P. A. (1982). The categorization of male and female voices in infancy. *Infant Behavior and Development*, *5*(2–4), 143–159. [https://doi.org/10.1016/S0163-6383\(82\)80024-6](https://doi.org/10.1016/S0163-6383(82)80024-6)
- Mitchell, L., Tsui, R. K. Y., & Byers-Heinlein, K. (2022). Cognates are advantaged in early bilingual expressive vocabulary development. *PsyArXiv*. <https://doi.org/10.31234/osf.io/daktp>
- Molnar, M., Gervain, J., & Carreiras, M. (2014). Within-rhythm Class Native Language Discrimination Abilities of Basque-Spanish Monolingual and Bilingual Infants at 3.5 Months of Age. *Infancy*, *19*(3), 326–337. <https://doi.org/10.1111/infa.12041>
- Montanari, S. (2008). Pragmatic differentiation in early trilingual

- development. *Journal of Child Language*, 36(3), 597–627.  
<https://doi.org/10.1017/S0305000908009112>
- Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development*, 16(4), 495–500.  
[https://doi.org/10.1016/0163-6383\(93\)80007-U](https://doi.org/10.1016/0163-6383(93)80007-U)
- 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. <https://doi.org/10.1037/0096-1523.24.3.756>
- Nazzi, T., Jusczyk, P. W., & Johnson, E. K. (2000). Language Discrimination by English-Learning 5-Month-Olds: Effects of Rhythm and Familiarity. *Journal of Memory and Language*, 43(1), 1–19. <https://doi.org/10.1006/jmla.2000.2698>
- Nicoladis, E. (1998). First clues to the existence of two input languages: Pragmatic and lexical differentiation in a bilingual child. *Bilingualism: Language and Cognition*, 1(2), 105–116.  
<https://doi.org/10.1017/s1366728998000236>
- Nicoladis, E., & Genesee, F. (1996). A Longitudinal Study of Pragmatic Differentiation in Young Bilingual Children. *Language Learning*, 46(3), 439–464. <https://doi.org/10.1111/J.1467-1770.1996.TB01243.X>
- Nicoladis, E., & Secco, G. (2000). The role of a child’s productive vocabulary in the language choice of a bilingual family. *First Language*, 20(58), 003–028.  
<https://doi.org/10.1177/014272370002005801>
- Nygaard, L. C., & Tzeng, C. Y. (2021). Perceptual integration of linguistic and non-linguistic properties of speech. In *The Handbook of Speech Perception* (pp. 398–427). John Wiley & Sons, Ltd.  
<https://doi.org/10.1002/9781119184096.ch15>
- Orena, A. J., Mader, A. S., & Werker, J. F. (2022). Learning to Recognize Unfamiliar Voices: An Online Study With 12- and 24-Month-Olds. *Frontiers in Psychology*, 13, 1695.  
<https://doi.org/10.3389/FPSYG.2022.874411/BIBTEX>
- Patterson, M. L., & Werker, J. F. (2003). Two-month-old infants match phonetic information in lips and voice. *Developmental Science*, 6(2), 191–196. <https://doi.org/10.1111/1467-7687.00271>
- Pons, F., Bosch, L., & Lewkowicz, D. J. (2015). Bilingualism Modulates

- Infants' Selective Attention to the Mouth of a Talking Face. *Psychological Science*, 26(4), 490–498.  
<https://doi.org/10.1177/0956797614568320>
- Poulin-Dubois, D., Serbin, L. A., & Derbyshire, A. (1998). Toddlers' Intermodal and Verbal Knowledge About Gender. *Merrill-Palmer Quarterly*, 44(3), 338–338.  
<https://go.gale.com/ps/i.do?p=AONE&sw=w&iissn=0272930X&v=2.1&it=r&id=GALE%7CA55416357&sid=googleScholar&linkaccess=fulltext>
- Poulin-Dubois, D., Serbin, L. A., Kenyon, B., & Derbyshire, A. (1994). Infants' Intermodal Knowledge About Gender. *Developmental Psychology*, 30(3), 436–442. <https://doi.org/10.1037/0012-1649.30.3.436>
- Quay, S. (2008). Dinner conversations with a trilingual two-year-old: Language socialization in a multilingual context. *First Language*, 28(1), 5–33. <https://doi.org/10.1177/0142723707083557>
- Quinn, P. C., Yahr, J., Kuhn, A., Slater, A. M., & Pascalis, O. (2002). Representation of the gender of human faces by infants: a preference for female. *Perception*, 31(9), 1109–1121.  
<https://doi.org/10.1068/P3331>
- Ramus, F., Dupoux, E., & Mehler, J. (2003). The psychological reality of rhythm classes: Perceptual studies. *15th ICPhs Barcelona*, 1–6.  
<http://cogprints.org/3079/>
- Ramus, F., Hauser, M. D., Miller, C., Morris, D., & Mehler, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science*, 288(5464), 349–351.  
[https://doi.org/10.1126/SCIENCE.288.5464.349/SUPPL\\_FILE/JASYN\\_TREV.AIF](https://doi.org/10.1126/SCIENCE.288.5464.349/SUPPL_FILE/JASYN_TREV.AIF)
- Redlinger, W. E., & Park, T. Z. (1980). Language mixing in young bilinguals. *Journal of Child Language*, 7(2), 337–352.  
<https://doi.org/10.1017/S030500090000266X>
- Rennels, J. L., & Davis, R. E. (2008). Facial experience during the first year. *Infant Behavior and Development*, 31(4), 665–678.  
<https://doi.org/10.1016/J.INFBEH.2008.04.009>
- Richoiz, A. R., Quinn, P. C., De Boisferon, A. H., Berger, C., Loevenbruck, H., Lewkowicz, D. J., Lee, K., Dole, M., Caldara, R., & Pascalis, O. (2017). Audio-visual perception of gender by infants emerges earlier for adult-directed speech. *PLoS ONE*, 12(1), e0169325.

<https://doi.org/10.1371/journal.pone.0169325>

- Righi, G., Westerlund, A., Congdon, E. L., Troller-Renfree, S., & Nelson, C. A. (2014). Infants' experience-dependent processing of male and female faces: Insights from eye tracking and event-related potentials. *Developmental Cognitive Neuroscience*, 8, 144–152. <https://doi.org/10.1016/J.DCN.2013.09.005>
- Ronjat, J. (1913). *Le développement du langage observé chez un enfant bilingue*. H. Champion.
- Sai, F. Z. (2005). The role of the mother's voice in developing mother's face preference: Evidence for intermodal perception at birth. *Infant and Child Development*, 14(1), 29–50. <https://doi.org/10.1002/ICD.376>
- Schott, E., Tamayo, M. P., & Byers-Heinlein, K. (2023). Keeping track of language: Can monolingual and bilingual infants associate a speaker with the language they speak? *Infant and Child Development*, 32(3), e2403. <https://doi.org/10.1002/icd.2403>
- Sebastián-Gallés, N., Albareda-Castellot, B., Weikum, W. M., & Werker, J. F. (2012). A Bilingual Advantage in Visual Language Discrimination in Infancy. *Psychological Science*, 23(9), 994–999. <https://doi.org/10.1177/0956797612436817>
- 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. <https://doi.org/10.2307/1165783>
- Spelke, E. S., & Owsley, C. J. (1979). Intermodal exploration and knowledge in infancy. *Infant Behavior and Development*, 2(1), 13–27. [https://doi.org/10.1016/S0163-6383\(79\)80004-1](https://doi.org/10.1016/S0163-6383(79)80004-1)
- Tare, M., & Gelman, S. A. (2010). Can you say it another way? Cognitive factors in bilingual children's pragmatic language skills. *Journal of Cognition and Development : Official Journal of the Cognitive Development Society*, 11(2), 137. <https://doi.org/10.1080/15248371003699951>
- Tsui, R. K. Y., Gonzalez-Barrero, A. M., Schott, E., & Byers-Heinlein, K. (2022). Are translation equivalents special? Evidence from simulations and empirical data from bilingual infants. *Cognition*, 225, 105084. <https://doi.org/10.1016/J.COGNITION.2022.105084>
- Uttley, L., de Boisferon, A. H., Dupierrix, E., Lee, K., Quinn, P. C., Slater,

- A. M., & Pascalis, O. (2013). Six-month-old infants match other-race faces with a non-native language. *International Journal of Behavioral Development, 37*(2), 84–89.  
<https://doi.org/10.1177/0165025412467583>
- Volterra, V., & Taeschner, T. (1978). The acquisition and development of language by bilingual children. *Journal of Child Language, 5*(2), 311–326. <https://doi.org/10.1017/S0305000900007492>
- Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: evidence for a bias for speech in neonates. *Developmental Science, 10*(2), 159–164. <https://doi.org/10.1111/J.1467-7687.2007.00549.X>
- Walker-Andrews, A. S., Bahrack, L. E., Raglioni, S. S., & Diaz, I. (1991). Infants' Bimodal Perception of Gender. *Ecological Psychology, 3*(2), 55–75. [https://doi.org/10.1207/s15326969eco0302\\_1](https://doi.org/10.1207/s15326969eco0302_1)
- Walton, G. E., Bower, N. J. A., & Bower, T. G. R. (1992). Recognition of familiar faces by newborns. *Infant Behavior and Development, 15*(2), 265–269. [https://doi.org/10.1016/0163-6383\(92\)80027-R](https://doi.org/10.1016/0163-6383(92)80027-R)
- Watkins, M. W. (2018). Exploratory Factor Analysis: A Guide to Best Practice. *Journal of Black Psychology, 44*(3), 219–246.  
[https://doi.org/10.1177/0095798418771807/ASSET/IMAGES/LARGE/10.1177\\_0095798418771807-FIG1.JPEG](https://doi.org/10.1177/0095798418771807/ASSET/IMAGES/LARGE/10.1177_0095798418771807-FIG1.JPEG)
- Weikum, W. M., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastián-Gallés, N., & Werker, J. F. (2007). Visual language discrimination in infancy. *Science (New York, N.Y.), 316*(5828), 1159. <https://doi.org/10.1126/science.1137686>
- Werker, J., & Curtin, S. (2005). PRIMIR: A Developmental Framework of Infant Speech Processing. *Language Learning and Development, 1*(2), 197–234. [https://doi.org/10.1207/S15473341LLD0102\\_4](https://doi.org/10.1207/S15473341LLD0102_4)
- Werker, J. F., & Byers-Heinlein, K. (2008). Bilingualism in infancy: first steps in perception and comprehension. *Trends in Cognitive Sciences, 12*(4), 144–151.  
<https://doi.org/10.1016/J.TICS.2008.01.008>
- Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*(1), 49–63.  
[https://doi.org/10.1016/S0163-6383\(84\)80022-3](https://doi.org/10.1016/S0163-6383(84)80022-3)
- Yu, M. E., Fecher, N., & Johnson, E. K. (2024). Learning to identify talkers: Do 4.5-month-old infants distinguish between unfamiliar males?

## Chapter 7: References

*JASA Express Letters*, 4(1), 15203.

<https://doi.org/10.1121/10.0024271/2932825>

## Chapter 7: References

Chapter 8:  
**APPENDICES**



# Appendix 1: Language Separation and Awareness Questionnaire

## Before you start:

If you have more than one child, please answer these questions with regard to your youngest child over the age of 1. Only one caregiver per child needs to complete the questionnaire, however feel free to discuss the questions with other family members.

As we are looking at children of different ages, it is possible that some questions will ask about things that your child does not yet do. In those cases, you will find options to indicate this. Please answer as accurately as possible.

## **General Questions:**

### About yourself

1. What is your relationship to the child?
  - a. Mother
  - b. Father
  - c. Other: [Open answer]
2. Where in Quebec do you currently live? Please state the name of the town or city.[Open answer] + *I do not live in Quebec*
3. What is the highest level of education you have completed?
  - a. Some high school or less
  - b. High School
  - c. CEGEP or DEP (or equivalent)
  - d. Bachelor's degree (or equivalent)

- e. Master's degree (or equivalent)
  - f. PhD or MD
  - g. Other: [Open answer]
4. What language(s) do you usually use when speaking to your child? [Check all that apply]
- a. French
  - b. English
  - c. Other: [Open answer]
5. What percentage of your interactions with your child are in each language? The percentages must total 100.
- a. French: %
  - b. English: %
  - c. Other: %
6. How would you rate your proficiency in **French**? 1 indicates very low proficiency, and 10 indicates very high proficiency.
- a. Understanding: [1 = not at all proficient ... 10 = highly proficient]
  - b. Speaking: [1 = not at all proficient ... 10 = highly proficient]
7. How would you rate your proficiency in **English**? 1 indicates very low proficiency, and 10 indicates very high proficiency.
- a. Understanding: [1 = not at all proficient ... 10 = highly proficient]
  - b. Speaking: [1 = not at all proficient... 10 = highly proficient]
8. Which of the following options best describes your family configuration?
- a. I am raising my child alone
  - b. I am raising my child with their other parent who lives in the same household

- c. I am raising my child with their other parent who lives in a different household
- d. Other: [Open answer]

### About the other parent

9. What is the highest level of education the other parent has completed?
- a. Some high school or less
  - b. High School
  - c. CEGEP or DEP (or equivalent)
  - d. Bachelor's degree (or equivalent)
  - e. Master's degree (or equivalent)
  - f. PHD or MD
  - g. Other: [Open answer]
10. What language(s) does the other parent usually use when speaking to your child?[Check all that apply]
- a. French
  - b. English
  - c. Other: [Open answer]
11. What percentage of the other parent's interactions with your child are in each language? The percentages must total 100.
- a. French: %
  - b. English: %
  - c. Other: %
12. How would you rate the other parent's proficiency in **French**?  
1 indicates very low proficiency, and 10 indicates very high proficiency.
- a. Understanding: [1 = not at all proficient ... 10 = highly proficient]
  - b. Speaking: [1 = not at all proficient ... 10 = highly

proficient]

13. How would you rate the other parent's proficiency in **English**?

1 indicates very low proficiency, and 10 indicates very high proficiency.

- a. Understanding: [1 = not at all proficient ... 10 = highly proficient]
- b. Speaking: [1 = not at all proficient ... 10 = highly proficient]

### About your child

14. What is your child's birthdate? [Select answer from calendar]

15. What is your child's sex?

- a. Male
- b. Female
- c. Other: [Open answer]

16. Does your child have any developmental or health issues that might affect their language development?

- a. Yes
- b. No

If you answered yes, please specify: [Open answer]

17. How old was your child when they started **regularly hearing** (10 hours per week or more) each of the following languages? Please exclude sources such as TV, radio, videos, etc.

- a. French: [Years][Months]
- b. English: [Years][Months]

18. Is your child **regularly exposed** (10 hours per week or more) to any language(s) other than French and English? Please exclude sources such as TV, radio, videos, etc. [Yes, No]

- a. If yes, which language(s)? [Open answers]
- b. From what age? [Years][Months]

- c. In what context (e.g., grandmother, daycare, preschool, etc)? [Open answer]
19. **Currently**, in a normal week, what percentage of the time does your child hear each language directly spoken to them? Please exclude sources such as TV, radio, videos, etc.
- a. French: %
  - b. English: %
  - c. Other: %
20. Throughout their **whole life**, what percentage of the time has your child heard each language directly spoken to them? Please exclude sources such as TV, radio, videos, etc.
- a. French: %
  - b. English: %
  - c. Other: %
21. How would you rate your child's proficiency in **French** compared to other **children their age**? 1 indicates much lower proficiency than other children, and 10 indicates the same proficiency as other children their age.
- a. Understanding: [1 = Much lower, 10 = Same as other children] + *My child does not yet understand any French*
  - b. Speaking: [1 = Much lower, 10 = Same as other children] + *My child does not yet speak any French*
22. How would you rate your child's proficiency in **English** compared to other **children their age**? 1 indicates much lower proficiency than other children, and 10 indicates the same proficiency as other children their age.
- a. Understanding: [1 = Much lower, 10 = Same as other children] + *My child does not yet understand any English*
  - b. Speaking: [1 = Much lower, 10 = Same as other children]

+ *My child does not yet speak any English*

About your family's approach to languages

23. Do you talk to your child about their bilingualism (e.g., mention that they speak two languages)? [1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Often, 5 = Very often]
- a. (if 4-5) What kinds of things do you say to your child about their bilingualism? Feel free to give examples.  
[Open answer]
24. Does your family follow a specific strategy to raise your child bilingually? (e.g., each caregiver uses a different language, both caregivers use both languages, one language at home and the other outside the home, etc.) [Yes, No]
- a. (If yes) How would you describe your family's language strategy? [Open answer]
- b. (If yes) Do you talk to your child about your family's language strategy? [1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Often, 5 = Very often]
25. Do you talk to your child about the languages that the people around them speak? (e.g., "Grandma speaks French") [1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Often, 5 = Very often]
26. Do you talk to your child about the language they should use with different people? (e.g., "You should speak French with grandma") [1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Often, 5 = Very often]

About your opinion on your child language awareness

27. Does your child seem to tell the difference between **French** and **English**? [1 = Strongly seems they don't, 2 = Somewhat seems

they don't, 3 = Neither seems as if they do or they don't, 4= Somewhat seems they do, 5 = Strongly seems they do]

- a. (If answered 4 or 5): At what age do you think they started telling the difference? [Years] [Months]
- b. What did your child do that made you think they could tell the difference? Feel free to give examples. [Open answer]

28. Does your child seem to notice that different people use different languages? [1= Strongly seems they don't, 2 = Somewhat seems they don't, 3 = Neither seems as if they do or they don't, 4= Somewhat seems they do, 5 = Strongly seems they do]

- a. (If answered 4 or 5): At what age do you think they started noticing? [Years][Months]
- b. What did your child do that made you think they noticed? Feel free to give examples. [Open answer]

29. Does your child seem to notice when someone familiar speaks to them in a different language than usual? (i.e., you usually speak in French with your child, but at a specific moment you speak in English to them). [1= Strongly seems they don't, 2 = Somewhat seems they don't, 3 = Neither seems as if they do or they don't, 4= Somewhat seems they do, 5 = Strongly seems they do]

- a. (If answered 4 or 5): At what age do you think they started noticing? [Years][Months]
- b. What did your child do that made you think they noticed? Feel free to give examples. [Open answer]

## Language Separation and Awareness Questions:

### Noticing two languages

30. Your child seems to notice when somebody speaks one of their languages rather than an unfamiliar language. [1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Often, 5 = Always]
31. Your child seems to notice when somebody speaks a language that is unfamiliar to them. [1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Often, 5 = Always]
32. Your child seems to notice when somebody is speaking one of their languages and switches to their other language. (e.g., Someone says a sentence in English then switches to French) [1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Often, 5 = Always]

### Using words in two languages

Bilingual children learn pairs of words, one for each of their languages, to refer to the same thing.

33. Sometimes, the word pairs may sound very different, like [dog/*chien*] or [duck/*canard*]. When your child knows how to say a word in both of their languages, how often do they use them appropriately depending on the person they are speaking with? (i.e., Chien with French speakers and dog with English speakers) [1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Often, 5 = Always] + *My child does not yet know any word pairs like this*
34. Other times, the word pairs may sound similar but present different pronunciations, like [elephant/*éléphant*] or [banana/*banane*]. When your child knows how to pronounce a word in both of their languages, how often do they pronounce them appropriately depending on the person they are speaking

with? (i.e., *Éléphant* with French speakers and *elephant* with English speakers) [1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Often, 5 = Always] + *My child does not yet know any word pairs like this*

35. When your child does not know a word in one of their languages, they borrow a word from the other language. [1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Often, 5 = Always] + *My child is too young to do this*

36. When your child does not know a word in one of their languages, they ask for the translation in that specific language. (i.e., “What do you call this **in English**?” or “How do you say *dog* **in French**?”) [1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Often, 5 = Always] + *I think my child is too young to do this*

37. When your child knows a word in both languages, they can translate from one language to the other when asked to (i.e., “How do you say *apple* in French?”, “*Pomme*”). [1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Often, 5 = Always] + *I think my child is too young to do this*

38. Your child changes their accent according to which of their languages they are speaking [1 = Not at all, 2 = Slightly, 3 = Moderately, 4 = Quite a bit, 5 = Completely] + *My child is too young to tell*

### Speaking to different people in different languages

39. How frequently does your child speak in **French** to you? (Consider only when your child initiates the interaction) [1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Often, 5 = Always] + *My child does not yet speak in French*

40. How frequently does your child speak in **English** to you? (Consider only when your child starts the interaction) [1 = Never,

2 = Rarely, 3 = Sometimes, 4 = Often, 5 = Always] + *My child does not yet speak in English*

41. When you speak to your child in **French**, how frequently do they respond in **French**? [1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Often, 5 = Always] + *I never speak to my child in French* + *My child does not yet speak in French*

a. (If the answer is 1 or 2) Why do you think this is? [Open answer]

42. When you speak to your child in **English**, how frequently do they respond in **English**? [1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Often, 5 = Always] + *I never speak to my child in English* + *My child does not yet speak in English*

a. (If the answer is 1 or 2) Why do you think this is? [Open answer]

43. How frequently does your child speak in **French** to the other parent? (Consider only when your child initiates the interaction) [1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Often, 5 = Always] + *My child does not yet speak in French*

44. How frequently does your child speak in **English** to the other parent? (Consider only when your child initiates the interaction) [1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Often, 5 = Always] + *My child does not yet speak in English*

45. When the other parent speaks to your child in **French**, how frequently do they respond in **French**? [1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Often, 5 = Always] + *The other parent never speaks to my child in French* + *My child does not yet speak in French*

a. (If the answer is 1 or 2) Why do you think this is? [Open answer]

46. When the other parent speaks to your child in **English**, how frequently do they respond in **English**? [1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Often, 5 = Always] + *The other parent never speaks to my child in English* + *My child does not yet speak in English*
- a. (If the answer is 1 or 2) Why do you think this is? [Open answer]
47. When your child meets a new person that speaks only **French** to them, how often do they use French with that person? [1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Often, 5 = Always] + *My child does not yet speak in French*
48. When your child meets a new person that speaks only **English** to them, how often do they use English with that person? [1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Often, 5 = Always] + *My child does not yet speak in English*
49. When asked, your child can name the language(s) you normally use when speaking to them. [Yes, No] + *I think my child is too young to do this*
50. When asked, your child can name the language(s) the other parent normally uses when speaking to them. [Yes, No] + *I think my child is too young to do this*
51. When asked, your child can name the language(s) a new person has used when speaking to them. [Yes, No] + *I think my child is too young to do this*
52. When asked what languages they speak, your child:
- a. Cannot name either of their languages
- b. Can name only one of their languages
- c. Can name both of their languages
- + *I think my child is too young to do this*

### Preferring to use different languages

53. Your child prefers to engage with you...

[1 = Much more when you speak French, 2 = Somewhat more when you speak French, 3 = Equally when you speak either of the languages, 4 = Somewhat more when you speak English, 5 = Much more when you speak English]

54. Your child prefers to engage with the other parent ...

[1 = Much more when they speak French, 2 = Somewhat more when they speak French, 3 = Equally when they speak either of the languages, 4 = Somewhat more when they speak English, 5 = Much more when they speak English]

55. Your child prefers to interact with other people ...

[1= Much more when they speak French, 2 = Somewhat more when they speak French, 3 = Equally when they speak either of the languages, 4 = Somewhat more when they speak English, 5 = Much more when they speak English]

56. Your child has explicitly told you that they prefer to speak one of their languages over the other. [Yes, No] + *I think my child is too young to do this*

a. (If yes): At what age?

b. (If yes): Which language?

57. Your child has explicitly told you that they prefer to interact with people that use one of their languages over the other. [Yes, No] + *I think my child is too young to do this*

a. (If yes): At what age?

b. (If yes): Which language?

**Final Open Question:**

58. Would you like to share any other information about your child's languageawareness? [Open answer]

# Appendix 2: Items for the Exploratory Factor Analysis

All items were measured using a 5-point Likert scale [1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Often, 5 = Always] unless otherwise specified below.

## **A. Items related to Language Discrimination:**

1. Notice Familiar Language: The child notices when someone speaks one of their languages (Q30).
2. Notice Unfamiliar Language: The child notices when someone speaks an unfamiliar language (Q31).
3. Notice Language Change: The child notices when someone speaks one of their languages and changes to the other (Q32).

## **B. Items related to Lexical Differentiation:**

4. Use of Translational Equivalents: The child uses translational equivalents appropriately according to the linguistic context (Q33).
5. Word Borrowing: When the child does not know a word in one language, they borrow one from the other language (Q35).
6. Ask for Translations: When the child does not know a word in one language, they ask for the translation in that language (Q36).
7. Translate When Asked: the child can translate words from one language to the other when asked to (Q37).

## **C. Items related to Phonological Differentiation:**

8. Pronunciation of Translational Equivalents: The child pronounces translational equivalents appropriately according to the linguistic context (Q34).

9. Change Accent: The child changes their accent according to the language they are using (Q38).

**D. Items related to Pragmatic Differentiation:**

10. Initiate in Caregiver 1's Most-used Language: The child initiates conversations with caregiver 1 in the language most used by that caregiver (Q39 or Q40, relative to caregiver 1's language use).
11. Initiate in Caregiver 1's Least-used Language: The child initiates conversations with caregiver 1 in the language least used by that caregiver (Q39 or Q40, relative to caregiver 1's language use).
12. Respond in Caregiver 1's Most-used Language: When caregiver 1 speaks to the child in their most-used language they respond in that language (Q41 or Q42, relative to caregiver 1's language use).
13. Respond in Caregiver 1's Least-used Language: When caregiver 1 speaks to the child in their least-used language they respond in that language (Q41 or Q42, relative to caregiver 1's language use).
14. Initiate in Caregiver 2's Most-used Language: The child initiates conversations with caregiver 2 in the language most used by that caregiver (Q43 or Q44, relative to caregiver 2's language use).
15. Initiate in Caregiver 2's Least-used Language: The child initiates conversations with caregiver 2 in the language least used by that caregiver (Q43 or Q44, relative to caregiver 2's language use).
16. Respond in Caregiver 2's Most-used Language: When caregiver 2 speaks to the child in their most-used language they respond in that language (Q45 or Q46, relative to caregiver 2's language use).
17. Respond in Caregiver 2's Least-used Language: When caregiver 2 speaks to the child in their least-used language they respond in that language (Q45 or Q46, relative to caregiver 2's language use).

18. Speak in a New Person's Language (Catalan/French): When someone new only speaks to the child in Catalan (or French) they use that language (Q47)
19. Speak in a New Person's Language (Spanish/English): When someone new only speaks to the child in Spanish (or English) they use that language (Q48)

**E. Items related to language preference and affiliation:**

20. Language Preference with Caregiver 1: The child prefers to engage with caregiver 1 when they use one of the languages (Q53) [Recoded: 0 = No preference, 1 = Slight preference, 2 = Strong preference].
21. Language Preference with Caregiver 2: The child prefers to engage with caregiver 2 when they use one of the languages (Q54) [Recoded: 0 = No preference, 1 = Slight preference, 2 = Strong preference].
22. Language Preference with Other People: The child prefers to interact with other people when they use one of the languages (Q55) [Recoded: 0 = No preference, 1 = Slight preference, 2 = Strong preference].

**F. Items related to explicit language awareness:**

23. Label Language(s) Caregiver 1: The child can name the language(s) caregiver 1 normally use with them (Q49) [Yes/No].
24. Label Language(s) Caregiver 2: The child can name the language(s) caregiver 2 normally use with them (Q50) [Yes/No].
25. Label Language(s) New People: The child can name the language(s) someone new has used with them (Q51) [Yes/No].

26. Label Own Languages: The child can name the languages they speak (Q52) [Recoded: 0 = None of them, 1 = One of them, 2 = Both of them].
27. Explicit Language Preference for Speaking: The child has expressed a preference for speaking one of the languages (Q56) [Yes/No].
28. Explicit Language Preference for Interacting: The child has expressed a preference for interacting with people who speak one of the languages (Q57) [Yes/No].

