Prosody in Bilingual Caregiver’s Infant-directed Speech: Cues for Infants’ Acquisition of their Languages’ Intonational Structure

Marina Kalashnikova1,2, Cristina Naranjo1

1Basque Center on Cognition, Brain and Language, San Sebastian, Spain
2Ikerbasque, Basque Foundation for Science, Bilbao, Spain

m.kalashnikova@bcbl.eu, cristinanaranjoortega@gmail.com

Abstract

Adults rely on various cues to differentiate among utterance types (e.g., declarative vs. interrogative utterances), including the conversational context, morpho-syntactic markers, and intonation. Infants have limited access to many of these language-specific cues, but they can rely on prosodic information in their speech input to learn the intonational patterns that differentiate utterance types in their language. Unlike monolinguals, bilingual infants must also use this information to differentiate the intonational patterns of each of their native languages.

This study investigated the role that prosodic adjustments in bilingual caregivers’ infant-directed speech (IDS) play in facilitating this task for young bilingual infants (12-30 months). This was assessed in naturally produced IDS and adult-directed speech (ADS) by Spanish-Basque bilinguals. These two languages differ drastically in their lexicon, morpho-syntactic structure, and prosody. We measured pitch and intonational contours of mothers’ infant-directed and adult-directed productions of declarative, wh- and yes/no interrogative utterances. Across utterance types and languages, IDS prosody was exaggerated compared to ADS. Additionally, intonation contours in Spanish and Basque IDS magnified the differences between wh- and yes/no interrogative utterances. These findings suggest that bilingual caregivers’ IDS may contain prosodic cues that facilitate infants’ early acquisition of the intonational patterns of their two languages.

Index Terms: infants, bilingualism, infant-directed speech, intonation, prosody, pitch

1. Introduction

Infants are sensitive to prosodic information in their speech input from their first days of life [1]. This early sensitivity becomes progressively attuned to their native language or languages enabling infants to rely on prosodic cues for segmenting and extracting meaningful linguistic information from continuous speech [2]-[4]. Prosodic cues at the phrase-level, in particular, can provide infants with the necessary information for learning the intonational structure and differentiating among utterance types used in their native language [5], which is essential for extracting the meaning carried by the utterances and later learning the lexical, morphological, and syntactic markers used to construct each utterance type. Bilingual infants, who are acquiring two languages from birth, face the additional task of using these prosodic cues to differentiate two languages and their individual intonational patterns. This study investigated the prosodic cues present in bilingual infants’ early linguistic input, i.e., bilingual caregivers’ infant-directed speech (IDS), which could facilitate this challenging task for these infants.

Adults who are proficient users of their native language or languages can rely on various cues to successfully differentiate among different utterance types, including the context of a given communicative interaction, lexical and morphological markers present in an utterance, its syntactic structure, and its intonation among others. Not all these cues are available for young infants who are still in the process of acquiring their native language. Even though there is evidence that infants are sensitive to some syntactic distinctions between declarative and interrogative utterances (e.g., word order) as early as at 12 months [6], it is likely that in their first year of life, they rely on prosody to differentiate among utterance types. Several prosodic components mark phrase-level intonation (e.g., duration and intensity), but adults [7] and infants have been shown to rely primarily on pitch cues (mean, range, and contour). In fact, at 5 months, infants can differentiate utterance types based solely on pitch information [8]. Frota and colleagues [9] tested 5-6- and 8-9-month-old Portuguese-learning infants’ ability to discriminate rising and falling pitch contours, which correspond to declarative and yes/no-questions in their native language. Infants showed successful discrimination at both ages. Interestingly, this early sensitivity to phrase-level pitch patterns appears to be already modulated by infants’ native language intonational structure since infants whose language input does not contain direct mappings between these intonational patterns and the two utterance types (English-learning infants) did not succeed in this task [10]. Therefore, infants did not simply attend to pitch patterns in these experiments, but their discrimination was guided by the pitch patterns that they have encountered in their specific language environments.

Infants’ early language input consists primarily of infant-directed speech (IDS) [11]. Compared to adult-directed speech (ADS), IDS is characterised by exaggerated prosody, manifested mainly in heightened mean pitch, expanded pitch range, and exaggerated pitch contours [12]-[14]. Prosodic exaggeration in IDS serves the function of attracting and maintaining infants’ attention to the speech signal [15], and there is evidence that it may facilitate infants’ perception of phrase-level prosodic cues, such as cues to clausal boundaries [16] and utterance completion [17]. Geffen and Mintz [18] directly investigated the production of declarative and interrogative utterances in IDS by English-speaking caregivers. English uses a combination of intonational, lexical, and syntactic cues to differentiate declarative utterances and two types of interrogative utterances: wh-questions (differentiated from declaratives by lexical and syntactic cues but not intonation, e.g., “where did you go?”) and yes/no-questions
(differentiated from declaratives by intonation and in some cases, but not all, by syntactic cues, e.g., “did you go to the park?” vs. “you went to the park?”). Their results showed that when mothers spoke to 8- to 10-month-old infants who did not have access to lexical and syntactic cues to differentiate these utterances, they produced similar intonational patterns as reported for ADS. That is, IDS contained pitch-based cues to differentiate yes/no-questions from the other two utterance types, but not to differentiate wh-questions from declaratives.

Similarly to monolingual infants, bilingual infants are sensitive to the prosodic patterns of their two native languages from birth [19], and this early-language-specific prosodic knowledge likely guides their ability to extract lexical and morpho-syntactic information from continuous speech in each of their languages [2], [20]. It is fair to assume that like monolinguals, bilinguals extract this language-specific information from their language input, but evidence about the prosodic and linguistic properties of bilingual IDS remains limited. Existing studies have suggested that at the phonetic level, bilingual caregivers produce similar adjustments in IDS of both their languages compared to IDS of monolingual caregivers’ one language [21], [22], but this has not been investigated with respect to phrase-level intonational patterns.

This study addresses this question by assessing the prosodic properties of IDS produced by bilingual caregivers to bilingual infants in Spanish and Basque. This bilingual population and language pair are optimal for this study. First, all caregivers were native speakers of Basque and Spanish. The families were recruited in the Guipuzcoa province of the Basque Country in Spain where the two languages hold an official status [23]. Second, Spanish and Basque have different lexical, syntactic, and intonational structures allowing for a direct comparison of language-specific adjustments in maternal IDS. Spanish is a Romance language that follows the Subject-Verb-Object canonical word order. Wh-questions are constructed by placing an interrogative word in the sentence-initial position, whereas (intonation seeking) yes/no-questions are structurally identical to declaratives [24]. Basque is a linguistic isolate, and it follows the Subject-Object-Verb canonical word order. Wh-questions are constructed by placing an interrogative word in the sentence-initial position and inverting the subject and verb positions, whereas (intonation seeking) yes/no questions are structurally identical to declaratives. (Note that in Basque, an optional interrogative particle can be used in yes/no-questions [25], but it is not described here as this type of utterances was not included in the analyses.)

<table>
<thead>
<tr>
<th></th>
<th>Declarative</th>
<th>Wh-question</th>
<th>Yes/no question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spanish</td>
<td>Falling (L)</td>
<td>Falling (L) or Rising (H)</td>
<td>Rising (H)</td>
</tr>
<tr>
<td>Basque</td>
<td>Falling (L)</td>
<td>Falling (L) or Rising (H)</td>
<td>Falling (L) or Rising-falling (HL)</td>
</tr>
</tbody>
</table>

Table 1 summarises the dominant intonation contours of each sentence type in Spanish and Basque. Declarative sentences in the two languages are predominantly marked by falling intonation, and in the case of wh-questions, both languages allow for the use of falling or rising intonation, with the choice depending on the communicative context and the speaker’s intent [26], [27]. Critically for this study, the two languages differ in the intonational patterns used for yes/no-questions. In Spanish, the predominant pattern is rising intonation, so this type of question is differentiated from both declaratives and wh-questions based on pitch-based cues [26]. On the contrary, in Basque, yes/no questions are realised with two possible contours: falling or the complex rising-falling contour, so prosodically, these utterances are highly similar to declaratives and wh-questions [27].

In this study, each speaker contributed four speech samples: IDS in Spanish and Basque and ADS in Spanish and Basque. We calculated three measures for each sample to capture the degree of prosodic exaggeration and intonational adjustments in IDS compared to ADS: mean pitch height, pitch range, and utterance boundary tones (i.e., final intonation). In line with previous research on IDS, we predicted that pitch height and range would be exaggerated in this register compared to ADS in both Spanish and Basque. Furthermore, we predicted that despite this overall prosodic exaggeration, mothers would preserve the language-specific intonational patterns for each utterance type in IDS in both languages.

2. Method

2.1. Participants

Twenty mother-infant (11 female infants) dyads participated in this study. In 6 dyads, infants were 13 months old (M = 13 months, SD = 0.6) and in 14 dyads 30 months old (M = 29 months, SD = 0.7). Mothers (M age = 34.8 years, SD = 4.1) were native speakers of Basque and Spanish, having acquired both languages at birth, or one language at birth and the other since pre-school (note that preliminary analyses yielded no effects of maternal non-dominant language age of acquisition). They reported that, on a daily basis, they used their dominant language (Spanish for 13 mothers, Basque for 7 mothers) on average for 58% of time and their non-dominant language for 32% of time (note that the percentages do not add up to 100% because most mothers also used a third language, English in most cases, but never at home or with their infants). Infants were growing up in a bilingual environment with exposure to Spanish and Basque from birth. On average, infants’ dominant language exposure was 59% (Spanish for 11 infants, Basque for 9 infants), and 41% for their non-dominant language. Dyads were only included in the study if the mother reported feeling comfortable using both languages to speak with her child.

2.2. Materials

Two versions of a story book, suitable for young children, were created for this study. The two versions were direct Spanish and Basque translations, so their content was identical. The stories consisted of a total of 123 utterances, of which 48 were designated as the target utterances for analyses: 16 declaratives, 16 wh-questions, and 16 yes/no questions.

2.3. Procedure

The IDS and ADS recordings were conducted online using a video-conferencing platform. Mothers were informed in advance that the study consisted of two parts; one in which their infant would need to be outside of the room (and cared for by another family member), and another in which they would interact with their infant. Mothers were asked to use a computer or a tablet and to use an external microphone during the sessions to ensure the best possible audio quality. They also received additional instructions about reducing external noise during the
sessions (e.g., ensuring that pets and other family members were not in the room, switching off electric appliances, etc.).

At the start of the experiment, the mother and an experimenter (female native bilingual speaker of Spanish and Basque) started the video call, and the experimenter provided detailed instructions about the experiment and collected informed consent. First, the ADS session was completed. In this session, the mother was asked to read the story book to the experimenter, once in Spanish and once in Basque. After the ADS recordings, the mother was asked to bring the infant to the room and read the story book with them, once in Spanish and once in Basque. The order of the languages in the ADS and IDS sessions was counterbalanced across participants. In the two sessions, mothers were instructed to read the stories naturally but to avoid modifying their voice to sound like the characters. In addition, they were asked to try to read the story verbatim, but they were told not to worry if they were interrupted by their infant or wanted to say additional things to them.

2.4. Analyses

Acoustic analyses of the recordings were conducted in Praat [28]. First, the recordings were segmented into utterances, and target utterances suitable for analyses were selected. An utterance was suitable if (1) it was read verbatim from the storybook, and (2) it did not have any background noise. A total of 1996 utterances were suitable. Next, boundary syllables were identified for each utterance for boundary-tone coding. The boundary syllables comprise all syllables occurring after the last stressed syllable in an utterance.

The boundary syllable was used to identify the shape of the boundary tone, i.e., the tone that occurs immediately after the last pitch accent in an utterance. We used a simplified version of the ToBi system for prosodic annotation [29], which is based on the Autosegmental Metrical model [30]. A simplified version was used that consisted solely in the identification of utterance boundary tones and classification into four tone categories: two single tone categories, H and L, and two complex bi-tonal categories, LH and HL. H describes high and rising intonation contours, and L describes low or falling intonation contours. In the bi-tonal contours, the tone first undergoes rising or falling followed by a sharp fall or rise, for the HL and LH respectively. This simplified coding allowed for a direct cross-linguistic comparison of the dominant boundary intonation contours in our analyses. Finally, custom Praat scripts were used to extract the F0 mean, F0 min, and F0 max values in Hz for each utterance. Given the logarithmic nature of speech perception, these values were transformed to a MEL scale (using the mel function of the EtmuR package [31] in R), and Pitch Range was calculated as F0 max – F0 min.

Preliminary statistical analyses included children’s age (13 vs. 30 months) as a factor. It yielded no significant effects, so it was removed from the analyses reported below.

3. Results

3.1. Mean Pitch and Pitch Range

To assess the effects of speech register on maternal pitch production in Spanish and Basque, two repeated-measures ANOVAs were conducted with mean pitch and pitch range as the dependent variables respectively, and Language (Spanish, Basque) and Register (ADS, IDS) as the independent variables (Figure 1). Results showed that mothers produced significantly higher mean pitch in IDS than in ADS, F(1, 17) = 24.467, p < .001, η² = .503. There was no effect of Language, p = .938, or Language by Register interaction, p = .737. Similarly, pitch range was significantly higher in IDS than in ADS, F(1, 13) = 56.276, p < .001, η² = .672, with no effect of Language, p = .431, and no Language by Register interaction, p = .290.

![Figure 1: Mean pitch and pitch range in Basque and Spanish ADS and IDS.](image)

3.2. Intonation Contours

Figure 2 displays the frequency with which each intonation contour occurred in Basque and Spanish ADS and IDS across utterance types. Proportions were calculated out of the total number of utterances within a type produced by an individual mother. In order to compare productions across registers, we selected the dominant contour or contours for each utterance type (based on our predictions in Table 1 and confirmed by our ADS data), and conducted paired-samples t-test analyses with Register (IDS and ADS) as the independent variable.

![Figure 2: Frequency of the four intonation contours in ADS and IDS in Basque and Spanish.](image)

3.2.1. Basque

In Basque ADS, declarative utterances were produced predominantly with the L contour (86%). Question production was more varied: wh-questions occurred with H (42%) or L (37%) contours, and yes/no/questions occurred predominantly with H (50%) contours, but L (25%), and HL (20%) contours were also present.

In Basque IDS, several adjustments to these distributions were observed. Declarative utterances in IDS occurred predominantly with an L (73%) contour. This pattern matches ADS productions, but the frequency of L contours was lower in
IDS compared to ADS, \(t(18) = 2.976, p = .008\), so IDS production was more variable with some occurrences of complex HL and LH contours. Wh-questions in IDS were less likely to occur with an H contour (19%) compared to ADS, \(t(18) = 5.407, p < .001\), and instead, the L contour was predominant (61%), \(t(18) = -5.623, p < .001\). In yes/no-questions, two predominant contours were observed: H (33%), which occurred less frequently than in ADS, \(t(18) = 2.314, p = .033\), and L (37%), which occurred with comparable frequency to ADS, \(p = .053\). As in ADS, some instances of HL (18%) were also observed, and their occurrence frequency did not differ between registers, \(p = .965\).

3.2.2. Spanish

In Spanish ADS, declarative utterances were produced predominantly with the L contour (90%). The interrogative utterances had two dominant contours each: wh-questions occurred with H (50%) and L (39%) contours, and similarly, yes/no-questions occurred with H (54%) and L (35%) contours.

In Spanish IDS, declarative sentences also occurred predominantly with an L (84%) contour, with similar frequency to ADS, \(p = .693\). This was also the case for IDS yes/no-questions, which occurred with H (44%), \(p = .257\), and L (40%) contours, \(p = .483\). However, compared to ADS, wh-questions in Spanish IDS occurred more frequently with the L (66%) contour, \(t(17) = -4.246, p < .001\), and less frequently with the H (23%) contour, \(t(17) = 3.757, p = .002\).

4. Discussion

This study assessed the pitch-based prosodic properties of IDS by bilingual mothers in Spanish and Basque during natural interactions with their bilingual children. As expected, bilingual mothers produced IDS with higher pitch and greater pitch range than ADS in the two languages. Furthermore, mothers produced language-specific adjustments to the intonation contours that differentiated declarative and interrogative utterances in Spanish and Basque.

According to the documented intonation patterns for Spanish and Basque (Table 1) and to our own ADS analyses, in the two languages declarative sentences were expected to occur predominantly with a falling contour, and wh-questions with rising and falling contours. IDS declarative utterances followed this pattern, but IDS wh-questions occurred more frequently with falling than with rising intonation. Yes/no-questions were expected to occur with predominantly rising intonation in Spanish and falling intonation in Basque, but we observed that both rising and falling contours occurred with similar frequency in ADS and IDS in the two languages. Thus, while maternal ADS did not contain reliable prosodic cues that differentiated wh- and yes/no-questions in the two languages, their IDS magnified the difference between these utterance types by producing wh-questions with more consistent falling intonation than yes/no questions. Therefore, maternal IDS provided infants with prosodic cues that could assist their acquisition of yes/no-utterances, which differ from declaratives solely in intonation but no other lexical or syntactic cues.

The adjustments described above rendered prosodic cues insufficient for the successful differentiation of declarative and wh-interrogative utterances (both occurred predominantly with falling contours in IDS). However, unlike yes/no questions, wh-questions contain additional lexical and syntactic information such as the occurrence of the high-frequency interrogative words [6]. It is also noteworthy that while our study focused on utterance-final intonation patterns, reflecting previous research (e.g., [6]), utterance-initial cues can also be enlisted for utterance type discrimination [26]. In addition, it is highly plausible that IDS accentuated other non-pitch based cues that could assist infants’ differentiation of declarative and interrogative utterances even before they have access to the lexical and syntactic information. These additional cues can include other acoustic adjustments (e.g., increases in intensity, [18]) and visual cues (e.g., raised eyebrows, [32]), which are known to be exaggerated in IDS compared to ADS [33].

Our results concerning the yes/no-question production in Spanish and Basque in ADS were surprising since we did not observe the expected language-specific differences, i.e., rising contours in Spanish and falling contours in Basque. Therefore, we are unable to conclude whether pitch-based cues assessed here could assist bilingual infants’ language differentiation based on intonational patterns. It is noteworthy that previous research has observed convergence between the two intonational patterns due to the close contact of Spanish and Basque in the Basque Country, especially among younger speakers [34]. Furthermore, individual intonation choices in these utterances are subject to effects of individual language background and linguistic attitudes, as well as dialectal variation [25], [27], [35]. Our ADS data provide a baseline for the intonation patterns among Gipuzkoan bilingual female speakers around 35 years of age in a story-reading task, but further research is required to identify the canonical patterns of yes/no-question productions in this population.

Our preliminary analyses failed to show significant infant age effects on maternal IDS, but we acknowledge that this may be due to the small sample size in our younger group. It is plausible that age effects would be observed in a larger sample and wider age range. One possibility is that the adjustments to wh-question intonation (i.e., increased frequency of falling contours) will be absent in IDS to younger infants, and instead rising and complex contours will dominate across utterance types [13]. In their first year of life, infants show a preference for utterances with more variable intonational contours [12], [16] and for prosodically exaggerated pitch in general [15], [36], and caregivers comply by adjusting their IDS accordingly. Similarly, the adjustments reported here would be expected to decrease with age [37], as children develop more advanced receptive and expressive language capacities and become able to access the lexical and syntactic cues that differentiate utterance types in their native languages.

5. Conclusions

Our findings suggest that bilingual caregivers’ IDS may contain prosodic cues that facilitate infants’ early acquisition of the intonational patterns of their two languages. These results add to the growing evidence that IDS may serve the function of facilitating language development during children’s first years of life.

6. Acknowledgements

This research was funded by the Spanish Ministry of Science and Innovation Research Grant, PID2019-105528GA-I00, and a Ramón y Cajal Fellowship, RYC2018-024284-I; by the Basque Government through the BERC 2018-2021 program; and by the Agencia Estatal de Investigación through BCBL Severo Ochoa excellence accreditation SEV-2015-0490.
7. References


