



Gestural and Prosodic Development Act as Sister Systems and Jointly Pave the Way for Children's Sociopragmatic Development

Iris Hübscher^{1*} and Pilar Prieto^{2,3}

¹ URPP Language and Space, University of Zurich, Zurich, Switzerland, ² Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain, ³ Facultat de Traducció i Ciències del Llenguatge, Universitat Pompeu Fabra, Barcelona, Spain

OPEN ACCESS

Edited by:

Judit Gervain,
Centre National de la Recherche
Scientifique (CNRS), France

Reviewed by:

Lucy Tamsin Dipper,
City University of London,
United Kingdom
Francesca Marina Bosco,
University of Turin, Italy

*Correspondence:

Iris Hübscher
iriscarmela.huebscher@uzh.ch

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 30 November 2018

Accepted: 13 May 2019

Published: 12 June 2019

Citation:

Hübscher I and Prieto P (2019)
Gestural and Prosodic Development
Act as Sister Systems and Jointly
Pave the Way for Children's
Sociopragmatic Development.
Front. Psychol. 10:1259.
doi: 10.3389/fpsyg.2019.01259

Children might combine gesture and prosody to express a pragmatic meaning such as a request, information focus, uncertainty or politeness, before they can convey these meanings in speech. However, little is known about the developmental trajectories of gestural and prosodic patterns and how they relate to a child's growing understanding and propositional use of these sociopragmatic meanings. Do gesture and prosody act as sister systems in pragmatic development? Do children acquire these components of language before they are able to express themselves through spoken language, thus acting as forerunners in children's pragmatic development? This review article assesses empirical evidence that demonstrates that gesture and prosody act as intimately related systems and, importantly, pave the way for pragmatic acquisition at different developmental stages. The review goes on to explore how the integration of gesture and prosody with semantics and syntax can impact language acquisition and how multimodal interventions can be used effectively in educational settings. Our review findings support the importance of simultaneously assessing both the prosodic and the gestural components of language in the fields of language development, language learning, and language intervention.

Keywords: gesture acquisition, prosody acquisition, sociopragmatic development, multimodal communication and learning, child language acquisition

INTRODUCTION

Human face-to-face interaction is essentially multimodal and constitutes an organized combination of gesture and speech patterns that jointly convey relevant meanings. For example, speakers typically express focus in discourse through the use of prosodic prominence together with gestural beat movements (e.g., hand, arm, and head movements), which are temporally associated with the focused constituent. Researchers agree that gestures and speech form an integrated system from both a phonological and a semantic standpoint (McNeill, 1992; Goldin-Meadow, 1998; Graziano and Gullberg, 2018). Experimental work largely supports the gesture–speech integration hypothesis in both language production and language comprehension. For example, Kelly et al. (2010) showed how congruent combinations of iconic gestures and speech, expressed in lexical items (for example,

saying “chop” while making a chopping hand gesture), as opposed to incongruent combinations of gesture and speech, are integrated through mutual and necessarily bidirectional interactions during language comprehension.

In the last few decades, prosody has been shown to play a pivotal role within the gesture–speech integration system. First, from a temporal point of view, crosslinguistic investigations have shown that the prominent portions of co-speech gestures (e.g., gesture strokes and gesture apexes) tend to temporally align with prosodically prominent positions in speech, (e.g., pitch-accented syllables and peaks of rising pitch accents) (see De Ruiter, 1998 for Dutch; Jannedy and Mendoza-Denton, 2005; Rochet-Capellan et al., 2008 for Brazilian Portuguese; Esposito et al., 2007 for Italian; Ferré, 2011; Esteve-Gibert and Prieto, 2013 for French; Leonard and Cummins, 2011; Loehr, 2012 for English; Ambrazaitis and House, 2017 for French; Esteve-Gibert et al., 2017 for Catalan). In addition, several studies have highlighted the fact that prominent positions, as well as prosodic boundaries, shape gestural coordination patterns and speech planning processes, revealing that prosodic structure acts as an anchoring structure for gesture movements (Esteve-Gibert and Prieto, 2013; Krivokapiæ, 2014; Esteve-Gibert et al., 2017; Graziano and Gullberg, 2018).

Several studies have demonstrated the independent roles of prosody and gesture as sociopragmatic markers (see Kendon, 2004; Kita, 2009; Prieto, 2015 for the role of prosody in pragmatic marking). When looking at the field of pragmatics and pragmatic development more specifically, it becomes clear that the field has been dominated by the analysis of verbal expression. The standard definition of pragmatics refers to the ability to use propositional language in a given context. However, it has been suggested that prosody and non-verbal expressions, such as gestures, also convey meaning in a given context (Bara, 2010). The Bara definition views this type of cue as being something outside of language. This review is based on the assumption that verbal language is only one aspect of communication; that communication is multimodal and includes both visual and vocal expression (e.g., Perniss, 2018).

Recent work has shown that not only do both prosody and gesture convey pragmatic meaning, but, crucially, that they can also be regarded as two sides of the same coin in the conveyance of sociopragmatic meaning (see Krahmer and Swerts, 2005, 2009 for an overview of the audiovisual prosody framework). Brown and Prieto (unpublished) contend that prosody and gesture work in parallel to jointly encode a set of sociopragmatic meanings related to information structure, speech act information, epistemic stance, or politeness. Speakers tend to mark focused elements through prosodic prominence markers such as pitch accentuation, along with head nods, manual beat gestures and eyebrow movements (Dohen and Loevenbruck, 2009 for French; Kim et al., 2014 in English). The presence of these visual features enhances the perception of contrastive focus (e.g., Krahmer and Swerts, 2007; Prieto et al., 2015). Epistemic stance meanings, such as uncertainty or ignorance, are also expressed in both the gestural and prosodic domains: speakers express uncertainty by using facial cues involving eyebrow raising and furrowing, ‘funny faces’ and

by means of prosodic cues such as fillers, delays, and rising intonational pitch contours (Krahmer and Swerts, 2005; Dijkstra et al., 2006; Roseano et al., 2016). Similarly, speakers tend to signal politeness by using prosodic cues (e.g., by using a slower speech rhythm and a quieter and lower-pitched voice), together with more mitigated gestures and slower gestural movements (Winter and Grawunder, 2012; Brown et al., 2014; Hübscher et al., 2017a; Brown and Winter, 2018). All of the above supports the hypothesis of Bolinger (1986, p. 199), who suggested that “gesture and speech, specifically intonation, are a single form in two guises, one visible and the other audible,” and argued that gesture and intonation were closely related parts of a single gestural complex that stemmed jointly from the same semantic intent.

Despite the importance of gesture and prosody in communication, research on children's pragmatic development, has largely focused on the acquisition of lexical and morphosyntactic structures. Language acquisition research has noted the precursor role of prosody and gesture, but chiefly when considering early stages of language development. While prosodic features have been regarded as bootstrapping features for early speech segmentation, word learning, and syntactic development (see Cavalho et al., 2018, for a review), gestures (and specifically deictic gestures, e.g., pointing) have been shown to be strong predictors of early lexical and syntactic acquisition (Capirci et al., 1996; Iverson and Goldin-Meadow, 2005; Özçalışkan and Goldin-Meadow, 2005). However, there is sufficient evidence to hypothesize that prosodic and gestural patterns might be very important components of sociopragmatic development. First, Infant-Directed Speech (IDS) tightly integrates the prosodic and gestural properties of language, which in turn are profitably used by children as they learn to speak. When using IDS, caregivers often exaggerate both the prosodic features of their speech (higher mean pitch, expanded pitch contours, slower speech rate, and longer pauses) and their gestural behavior (more pronounced head and hand movements, more exaggerated facial expressions) in a manner analogous to the acoustic enhancement that occurs in their speech (e.g., see Smith and Strader, 2014). Second, recent results show that infants are extremely sensitive to the temporal synchrony between gestural movements and prosodic patterns. For example, using materials with congruent and incongruent prominence matching patterns, Esteve-Gibert et al. (2015), showed how 9-month-old infants are able to reliably detect whether a manual deictic gesture is aligned with the corresponding metrical patterns (see also Kitamura et al., 2014 for an experiment showing how 8-month-old infants successfully detect matching visual and prosodic displays of continuous IDS speech). Third, from a production point of view, at the onset of word production, infants temporally coordinate their metrically prominent syllables with pointing gestures in an adult-like way (Butcher and Goldin-Meadow, 2000; Esteve-Gibert and Prieto, 2014). Finally, Igualada et al. (2015) showed that the first uses of temporally synchronous pointing-speech combinations by 12-month-old infants have a high predictive value for language (vocabulary) and syntactic outcomes at 18 months (see also

Murillo et al., 2018). Given that there is already evidence that gesture and prosody jointly constitute precursor systems for lexical acquisition and syntactic/semantic development (i.e., moving from one to two-word stage), this evidence points to the need to assess the possibility that gesture and prosody also jointly constitute a precursor system for pragmatic development.

Though little is known about the developmental path followed by prosodic and gestural patterns in relation to language acquisition, and specifically in relation to pragmatic development, some researchers have highlighted the need to study gestural and prosodic patterns in tandem with development. Recently, Snow (2017) proposed that intonation and gesture are sister systems marking the same pragmatic functions by focusing on the analysis of protodeclarative sentences in 8- to 16-month-old infants. He claimed that the child's use of pitch prominence within an utterance vocally designates or "points to" the information focus of the utterance, similarly to a pointing gesture. Very recently, Esteve-Gibert and Guellai (2018) reviewed a set of developmental studies showing that gestures are tightly linked to prosody at the temporal and the pragmatic levels and across different developmental stages. The authors stressed the importance of clarifying the links between gesture and prosody at different stages in language acquisition. While Snow (2017) looked at the development of focus in initial speech acts and Esteve-Gibert and Guellai (2018) focused on the role of the temporal and functional integration of gesture-prosody, the present review paper will expand on these results and assess in a systematic way recent evidence showing that gesture and prosody, working in parallel, jointly constitute children's early development of socio-pragmatic meanings.

The main aim of this paper is to focus on the relevance of prosody and gesture, and the links between the two, in the context of children's sociopragmatic development. The article first reviews evidence on the role of gesture, Section "Gesture as a Precursor and Predictor of Lexical and Syntactic Development," and prosody, Section "Prosody as a Precursor and Predictor of Lexical and Syntactic Acquisition," as independent precursors and predictors of lexical and syntactic development in early infancy. Section "Developmental Evidence That Gesture and Prosody Are Precursors of Pragmatic Development" analyzes evidence that points to the precursor role of both prosody and gesture patterns across four separate components of pragmatic meaning, namely speech act distinctions, information structure, epistemic stance, and politeness marking. First, precursor speech act marking and information focus strategies are assessed in infants aged between 12 and 18 months. Second, epistemic and politeness marking strategies are assessed in 3- to 5-year-old preschoolers. Further complex pragmatic meanings such as irony and deceit will not be part of this review, but see Bosco et al. (2013) for a full review of these meanings, highlighting both perception and production and the inclusion of different means of expression. The focus of this investigation lies in the joint analysis of multimodal prosodic and gesture patterns by considering different types of gestures (hand, as well as head, face, and shoulders) along

with prosodic features as well as pitch (e.g., intensity, duration, and voice quality).

GESTURE AS A PRECURSOR AND PREDICTOR OF LEXICAL AND SYNTACTIC DEVELOPMENT

Gesture has been found to play a significant role in children's early language and communicative development. According to Tomasello (1995), the start of intentional communication in infants requires two abilities: (a) the ability to distinguish means and goals in their own actions and in their interlocutor's actions; and (b) the ability to engage in joint attention frames. By 9 months, infants are involved in complex triadic interactions between ego (the baby), alter (e.g., a parent), and an object of attention (e.g., a ball) (Striano and Tomasello, 2001). In this context, gesture, in the form of pointing, along with eye-gaze and vocalization patterns, is a key ingredient in the management of joint attention.

The pointing gesture has been studied the most as it presents the clearest case of declarative reference (i.e., directing an interlocutor's attention) within a joint attention framework. Pointing gestures are thought to be used in all human societies (Kita, 2003). In development, pointing is first used by 9- to 12-month-old infants well before they are able to produce their first words (e.g., Bates, 1976; Bates et al., 1979; Acredolo and Goodwyn, 1985; Iverson et al., 1994; Özçalışkan and Goldin-Meadow, 2005). While pointing is deictic (i.e., directive) in nature, infants also make use of other arm movements that serve an iconic function. They convey information about the attributes of an object or mimic actions, such as flapping arms to illustrate flying (Bates et al., 1979). Therefore, gesture comes before speech in language development and gives infants a channel to express declarative reference before they are able to do so through lexical means.

While some studies have demonstrated that gesture precedes verbalization, others have shown that the use of pointing gestures can *predict* lexical development (e.g., Capirci et al., 1996; Iverson and Goldin-Meadow, 2005; Özçalışkan and Goldin-Meadow, 2005; Rowe and Goldin-Meadow, 2009; Igualada et al., 2015). For example, Iverson and Goldin-Meadow (2005) showed that the use of gesture predicts the later use of verbal cues, thereby laying the foundation for the appearance of lexical meanings in speech. By observing the gestural and verbal development of three children between the ages of 10 and 24 months at monthly intervals, these authors found that referents that featured initially in children's gestural repertoire appeared some months later in their spoken vocabulary. For example, if children pointed to a ball at 14 months, the word ball was highly likely to appear in their vocabulary at 18 months.

Similarly, just as deictic gestures act as precursors to children's initial acquisition of single words, children combining deictic gestures with speech precedes the production of two-word utterances (Capirci et al., 2005; Iverson and Goldin-Meadow, 2005; Özçalışkan and Goldin-Meadow, 2005; Pizzuto and Capobianco, 2005; Goldin-Meadow, 2009). Studies have found

that those children who were the first to produce gesture-plus-word combinations conveying two elements in a proposition (e.g., pointing at a chair while saying “sit”) were also those who first produced two-word combinations (“sit chair”). Özçalışkan and Goldin-Meadow (2005) investigated whether the production of supplementary gesture–speech combinations, (e.g., *eat* + point at cookie) at 14, 18, and 22 months, foreshadowed oncoming changes in children’s speech, thus serving as a forerunner of linguistic advances. They found that the number of supplementary gesture–speech combinations produced by the children significantly increased between the ages of 14 and 22 months. Children first produced various constructions in gesture–speech combinations before the same constructions appeared entirely as speech.

Thus, deictic gestures play clear precursor and predictor roles in lexical and syntactic acquisition, particularly in the early stages of language development. However, less is known about the potential role of pointing gestures in the development of pragmatic skills, given that pointing is used to express basic pragmatic functions. It has been shown that, between the ages of 11 and 12 months, infants begin to produce communicative pointing gestures that have a declarative or requestive pragmatic function (Cochet and Vaclair, 2010). Similarly, infants at 12 months of age have been shown to be able to use the shape of the pointing gesture for specific pragmatic intentions, namely to request an object, inform about its presence, or express interest in it (see Kovács et al., 2014; Liszkowski, 2014 for a review; Tomasello et al., 2007). However, little is known about the relationship between the use of these gestures and the acquisition of pragmatics, or about how different shapes of pointing gestures combine with distinct prosodic patterns to express speech act distinctions. This is partly due to the fact that linguistic pragmatics has been regarded as a primarily verbal phenomenon. Moreover, gestural communication does not disappear with the emergence of word production, and even very young children can encode epistemic stance or politeness information through gesture and prosody. In Section “The Development of Speech Act Marking in Infancy,” we present evidence that, acting as sister systems, prosody and gesture function as precursors in the development of children’s comprehension and deployment of speech acts.

PROSODY AS A PRECURSOR AND PREDICTOR OF LEXICAL AND SYNTACTIC ACQUISITION

Like gestures, prosodic skills have been shown to arise very early in infants’ communicative development and play an important role in early language acquisition. Special attention has been paid to how prosody facilitates the segmentation of early speech into phonemes, syllables, and syntactic structures (see Cavalho et al., 2018 for a review). In their key study, Gleitman and Wanner (1982) proposed that acoustic cues in speech may help infants to detect syntactic boundaries before lexical knowledge is available (see also Morgan and Newport, 1981; Pinker, 1987; Gleitman et al., 1988). This proposal has become known as the

prosodic bootstrapping hypothesis. Three important elements are highlighted in this hypothesis, namely that (1) syntax is reliably correlated with prosodic properties; (2) infants are sensitive to the acoustic properties of speech; and (3) infants benefit from these cues when they are processing speech.

Importantly, prosody scaffolds not only children’s early segmentation and syntactic abilities but also their lexical acquisition (see Thorson, 2018 for an overview). As we have noted above, IDS is characterized by slower speech rates and exaggerated pitch excursions, which highlight vocalic and consonantal contrasts and thus help children in their construction of phoneme inventories (Kuhl et al., 1991; Werker et al., 2007; Cristià, 2011). It has been found that the use by a caregiver of such slower speaking rates and more emphatic vowel properties help 21-month-olds to better learn and remember new words (Ma et al., 2011). Thorson and Morgan (2014) showed that word learning is enhanced both when words appear at prosodic phrase boundaries and when new words carry a more prominent pitch accent. Thus, the exaggerated prosodic properties of IDS serve to make the relationship between prosodic form and lexical identity uniquely salient to children (Saint-Georges et al., 2013; Thorson, 2018).

Prosody, like gesture, has been found to provide children with acoustic cues that serve to bootstrap their grammatical and lexical development. However, research also suggests that young infants are able to use prosody for early intentional communication and the marking of speech act information well before the one-word stage (Papaeliou and Trevarthen, 2006; Sakkalou and Gattis, 2012; Esteve-Gibert and Prieto, 2013). Papaeliou and Trevarthen (2006) investigated various acoustic and prosodic parameters of 10-month-old English-acquiring children, comparing them playing by themselves with when they were with their mother. The results of their experiment showed that intentional vocalizations directed at their mother were shorter and displayed higher pitch values than the non-intentional vocalizations they produced when alone. In a later study, Esteve-Gibert and Prieto (2013) analyzed a total of 2,701 vocalizations from a longitudinal corpus of four Catalan-babbling infants aged 0;7 to 0;11 months. They found that infants used different prosodic patterns not only to distinguish communicative vocalizations from investigative ones and to express intentionality, but also to mark speech act information. Specifically, they found that requests and expressions of discontent displayed a wider pitch range and longer duration patterns than responses or statements. Sakkalou and Gattis (2012) tested if 14- and 18-month-old British infants would imitate more intentional actions than accidental ones, when the difference between the two actions was based on the prosody of the word accompanying the actions. During the test phase, the experimenter modeled two different actions on the toy. The intentional action was accompanied by a ‘there’ in a purposeful and satisfied tone of voice, and the accidental action was accompanied by ‘woops’ with a dissatisfied and surprised tone of voice. They found that both 14- and 18-month-old infants were able to understand the intention of the experimenter by responding differently to accidental actions than to intentional actions. Then, in a follow-up experiment, they replaced the

English words 'there' and 'whoops' with novel Greek words that were not known to the infants. Here they found that, purely on the basis of prosody, 18-month-old infants, and to a lesser extent, 14-month-old infants, more often imitated intentional actions than accidental ones. This demonstrates that infants are able to understand the pragmatic value of prosody since they are able to link prosodic properties to intentionality. To summarize, these results show that prelexical infants successfully use a set of prosodic patterns to signal intentional speech and speech act information.

Following on from the above review of research into the separate roles of gesture and prosody in early language development, the next section will discuss recent empirical evidence showing how gesture and prosody work together as an integrated system, and specifically act as precursors in children's pragmatic development.

DEVELOPMENTAL EVIDENCE THAT GESTURE AND PROSODY ARE PRECURSORS OF PRAGMATIC DEVELOPMENT

This section reviews empirical evidence on the role of gesture and prosody in early pragmatic development in terms of four pragmatic components, namely speech act marking, information focus, epistemic stance, and politeness. This review is based on extensive literature searches carried out during the past few years, by adopting an interdisciplinary approach on gesture-, prosody-, and pragmatic-development more generally, including studies stemming from the field of linguistics on the one hand and developmental psychological on the other hand or from the intersections of both. **Table 1** summarizes the main features of the studies that have taken into account prosody and non-verbal expressive means, i.e., gestures in the same developmental sample. Converging evidence will be shown from the development of the four above-mentioned socio-pragmatic components that prosody and gestures develop in a parallel and complementary fashion, both from a perception and a production point of view, and that they act as precursors of pragmatic development.

The Development of Speech Act Marking in Infancy

Speech acts provide a useful starting point for the analysis of communicative intent, as many researchers regard them as a bridge between the so-called pre-linguistic and early linguistic stages of language development (see Cameron-Faulkner, 2014 for an overview). Speech acts are a key facet of pragmatic development and also serve as the best example for demonstrating the key role of gesture and prosody in children's pragmatic development. The most detailed analysis of speech act development was carried out by Bates et al. (1975). They focused on the development of two speech acts, (proto)imperatives and (proto)declaratives, and found that both of these speech acts appeared around the age of 10 months

in three Italian-speaking children. These speech acts were first expressed gesturally, and then developed into locutionary acts. Building on this study, Tomasello et al. (2007) proposed that pointing was a communicative act, intended to direct an interlocutor's attention toward an object for one of three reasons: (1) to help the interlocutor with some information that might be of interest to them (*declarative informative pointing*), (2) to share attention with the interlocutor about an object or event (*declarative expressive pointing*), or (3) to request an object from the interlocutor (*imperative pointing*). Focusing on perception, research has shown that, at 12 months, infants understand whether an adult has produced an attention-directing act such as pointing in order to request an object, inform about its presence, or express interest in it (see Behne et al., 2012; Kovács et al., 2014; Liszkowski, 2014 for a review; Tomasello et al., 2007). Similarly, it has been demonstrated that infants aged from 0;7 to 0;11 can also use prosody to signal the specific pragmatic meaning of intentional vocalizations, with requests and expressions of discontent displaying wider pitch excursions and longer durations, and statements and responses displaying narrower pitch ranges and shorter durations (Esteve-Gibert and Prieto, 2012).

As mentioned above, much of the literature addressing early infant communication abilities comprises separate assessments of either gesture (mainly pointing gestures) or speech modalities, and only a few studies investigate this issue in an integrated fashion (Grünloh et al., 2015; Murillo and Capilla, 2016; Aureli et al., 2017; Esteve-Gibert et al., 2017). From a production perspective, these studies have shown that, across languages, the early production of pointing gestures with declarative or requestive functions are accompanied by specific intonational patterns. For instance, the pattern seen by Grünloh and Liszkowski (2015), where 14-month-old Dutch infants produced flat intonations with requestive pointing gestures and rising contours with declarative pointing gestures, seems to be found in children from many different language communities. On the other hand, language-group specific effects have also been identified. For example, Spanish-acquiring infants (15 months) produced flat intonations with declarative pointing and rising contours with requestive pointing (Murillo and Capilla, 2016), similar to Italian-learning infants (also at 15 months) (Aureli et al., 2017). Esteve-Gibert et al. (2016) showed that Catalan child-directed speech contains specific prosody-gesture multimodal combinations expressing distinct speech act information. In their study, nine caregiver-infant dyads played three games designed to elicit pointing acts with either an expressive, imperative, or informative pragmatic meaning. Results showed that in the three pragmatic situations, caregivers used different combinations of pointing gesture features (e.g., hand shape, gesture duration, and the gesture's lexical affiliate) and prosodic features (intonation contour, pitch range, and mean syllable duration) to express the three different social intentions.

Studying comprehension, Esteve-Gibert et al. (2017) showed that infants are able to infer meaning from multimodal cues, (e.g., pointing gestures alongside prosodic patterns), before they are able to use spoken language forms. In two different experiments, they tested whether 12-month-old infants would

TABLE 1 | Overview of the studies that have studied the development of different pragmatic functions by including gestural and/or prosodic cues.

Pragmatic function	References	Design	Ages	Gesture cues	Prosodic cues	Production/ Perception	Language acquiring
Speech act	Bates et al., 1975	Observational	0;2, 0;6, 1;0	Pointing gestures	–	Production	Italian
Speech act	Cochet and Vauclair, 2010	Observational	0;11 – 3;01	Pointing gestures	–	Production	French
Speech act	Behne et al., 2012	Experimental	1;0	Pointing gestures	–	Production/ Perception	German
Speech act	Prieto et al., 2012	Observational	0;11 – 1;8	–	Intonation contours	Production	Catalan and Spanish
Speech act	Sakkalou and Gattis, 2012	Experimental	1;2 – 1;6	–	Prosodic contours	Perception	English
Speech act	Esteve-Gibert and Prieto, 2012	Observational	0;7 – 0;11	–	Pitch and duration	Production	Catalan
Speech act	Kovács et al., 2014	Experimental	1;2	Pointing gestures	–	Perception	Hungarian
Speech act	Grünloh and Liszkowski, 2015	Experimental	1;2	Pointing gestures	Intonation contour	Production	Dutch
Speech act	Esteve-Gibert et al., 2016	Experimental	0;11 – 1;2	Pointing gestures	intonation, contour, pitch range and mean syllable duration	Production	Catalan
Speech act	Murillo and Capilla, 2016	Observational/ Semi-structured play	0;9 – 1;3	Pointing gestures	Several acoustic parameters	Production	Spanish
Speech act	Frota et al., 2016	Observational	1;0 – 2;4	–	Intonation and duration	Production	Portuguese
Speech act	Aureli et al., 2017	Experimental	1;0, 1;3, 1;6	Pointing gestures	Duration, mean F0, key, final contour	Production	Italian
Speech act	Esteve-Gibert et al., 2017	Experimental	1;0	Pointing gestures	Pitch range and syllable duration	Perception	Dutch
Information focus	Nicoladis et al., 1999	Observational	2;0 – 3;6	Gestures (iconic, deictic, beat)	–	Production	French-English bilingual
Information focus	Ito et al., 2012	Experimental	6;0 – 8;11	–	Pitch accent	Perception	English
Information focus	Mathew et al., 2014	Experimental	5;0 – 7;11	Beat gestures	Rhythmic marking of speech	Production	English
Information focus	Igualada et al., 2015	Experimental	1;0	Pointing gestures	Pitch accent	Production	Catalan
Information focus	Kurumada and Clark, 2017	Experimental	4;0-4;11	–	Pitch and duration	Perception	English
Information focus	Snow, 2017	Observational/semi-structured play	0;8 – 1;16	Pointing gestures	Intonation	Production	English
Information focus	Murillo et al., 2018	Observational/ Semi-structured play	0;9, 1;0, 1;3, 1;6	Pointing gestures	Vocalizations	Production	Spanish
Information focus	Esteve-Gibert et al., 2019	Experimental	5;0 – 5;11	Head nods	Syllable duration marking	Perception	French
Epistemic stance	Moore et al., 1993	Experimental	3;0 – 6;11	–	Intonation	Perception	English
Epistemic stance	Liszkowski et al., 2008	Experimental	1;0	Pointing gestures	–	Production	Dutch
Epistemic stance	Kim et al., 2016	Experimental	3;0 – 4;11	Head tilt, head shake, shoulder shrug, looking away	–	Production	German

(Continued)

TABLE 1 | Continued

Pragmatic function	References	Design	Ages	Gesture cues	Prosodic cues	Production/ Perception	Language acquiring
Epistemic stance	Hübscher et al., 2017a	Experimental	3;0 – 5;11	Head tilt, raised eyebrows	Intonation	Perception	Catalan
Epistemic stance	Bartz, 2017	Observational	1;2 – 3;6	Hand flip (PUOH gesture)	–	Production	English
Epistemic stance	Hübscher, 2018	Experimental	3;0 – 5;11	Head tilt and raised eyebrows	Intonation, fillers and vowel lengthening	Production	Catalan
Politeness	Bates, 1976	Experimental	3;0 – 7;11	–	Intonation	Perception and Production	Italian
Politeness	Shochi et al., 2009	Experimental	9;0 – 10;11	Facial cues	Intonation	Perception	Japanese
Politeness	Hübscher et al., 2018	Experimental	3;0-3;11	Facial cues	Intonation	Perception	Catalan
Politeness	Hübscher et al., 2019	Experimental	3;0– 5;11	Body and facial cues	Pitch, duration, intensity, voice quality	Production	Catalan

be able to link specific prosodic patterns and hand gestures to a speaker's underlying social intention (declarative, expressive or requestive). While in the first experiment they tested the role of prosody together with lexical and gestural cues, in the second they tested only the role of prosody and gesture. The results of the first experiment showed that the infants were able to use prosody and gesture to understand the intentions behind an attention-directing act, while the results of the second experiment illustrated that even when there was no lexical information, when the cue was controlled across experimental conditions, infants were able to infer a speaker's social intentions purely on the basis of prosody and manual gestures. This shows that infants are able to understand different speech acts at 12 months, and suggests that prosody and gestural patterns work in parallel to play a role in infants' early pragmatic development.

Children's socio-cognitive and lexical skills increase in concert with their use of prosodic features as they grow older. For example, Portuguese, Catalan, and Spanish children between 1 and 2 years of age produce some basic, pragmatically appropriate, intonation contours, which reflect a variety of speech acts such as vocatives, statements, or requests, and express various intentions (Prieto et al., 2012; Frota et al., 2016). Taken together, these findings demonstrate that both gestural and prosodic cues to speech act information develop in an integrated fashion in the early stages of communicative development and well before children can make the same speech act distinctions by verbal means (see Cameron-Faulkner, 2014).

The Development of Information Focus

Across languages, the use of prosodic prominence has been highlighted as one of the most reliable cues to various features of information structure, in particular the marking of focus. Prosodic prominence is by no means the only such cue, however. Cross-linguistically, focus can also be expressed by means of specific focus morphemes, syntactic movement of constituents, gestural prominence, or combinations of these elements.

Two recent review chapters (Ito, 2018 on focus perception and Chen, 2018 on focus production) on the subject of the acquisition of prosodic focus highlight the fact that research has produced mixed results about the age at which children begin to develop the ability to use prosodic prominence to mark information structure, with values that vary from 3 to 6 years. This variation may be contingent on the sorts of experimental tasks and measures used (Ito, 2018) or the type of focus marking used in a particular language (Chen, 2018). Specifically, Chen (2018) proposes a cross-linguistic theory for the acquisition of prosodic focus according to which, the specific means of marking focus in a particular language will have a predictable impact on the age at which children are able to produce it, and the order in which different types of focus will be acquired.

Ito (2018) points out that the assumption behind many studies about the development of prosodic focus is that the marking of the information structure of words (new, given, contrastive) must await the development of higher-level discourse abilities. However, while it is true that the acquisition of contrastive focus marking comes later in development, and has been separately studied (e.g., Speer and Ito, 2009; Ito et al., 2012;

Kurumada and Clark, 2017), we have some evidence suggesting that information focus might be acquired much earlier. Ito (2018, p. 251) states that, “in theory, infants who are learning to detect word boundaries in running speech may be simultaneously developing the ability to compute gross informational weight for parts of speech that bear prosodic prominence.” In fact, recent studies have shown that children benefit from the prosodic and gestural characteristics of IDS for lexical learning (see Thorson, 2018 for a review). As noted above, a series of studies have shown that the use of prosodic prominence expressing focus, in English, facilitates word learning in toddlers under the age of 2 (e.g., Fernald and Mazzie, 1991; Ma et al., 2011; Thorson and Morgan, 2014).

While the abovementioned studies showed that, from a production perspective, infants are sensitive to prosodic prominence cues signaling the focus referent for lexical learning, studies on early intonational development have shown that children younger than 2 years are able to use prosody to mark novel events and emphasis (Esteve-Gibert and Prieto, 2018; Frota and Butler, 2018). Perhaps the earliest evidence of a more emphatic prosodic realization of information focus is that described by Murillo and Capilla (2016). In their study, the authors explored the acoustic properties of vocalizations in the transition to first words, when they were accompanied with pointing gestures. The duration and pitch properties of the vocalizations were found to be expanded and more similar to those of adult speech when produced with pointing gestures with a declarative function.

It is our view that the combination of gesture (pointing) and prosody (most prominent prosody) is initially used by infants to mark information focus. Following Snow (2017), we suggest that intonation and gesture act as sister systems in the signaling of information focus in protodeclarative sentences used by infants between 8 and 16 months. In the regression patterns he detected in language development, Snow (2017, p. 184) found that pitch prominence, like pointing gestures, were used by children to designate or “point to the information focus of the utterance.” He contends that “the comparison between intonation and gesture shows that the two systems differ markedly in form but they have the same pragmatic function.” For example, protodeclaratives and falling intonation both convey the speaker's intention to make a statement or share an experience about an object of joint attention.

Further evidence comes from the behavior of 12-month-old children in the declarative pointing task used by Igualada et al. (2015). In the task, children produced a pointing gesture, which was temporally synchronous with the pitch accent in the associated word, as a strategy to successfully initiate and maintain joint attention. We interpret the ability of these 12-month-olds to selectively use this multimodal communicative strategy, to engage an adult in joint attention by drawing his or her attention toward an unseen event or object, as an early manifestation of information focus. The feature that marks novelty in speech, and also predicts later lexical and grammatical development, is the close temporal alignment of both prosody and gesture. Igualada et al. (2015) found that the presence of simultaneous pointing-speech combinations, as opposed to

speech-only and pointing-only productions, at 12 months old positively correlated with lexical and grammatical development at 18 months. Later in development, the same predictive pattern arises. A recent study investigated the changes in the temporal synchrony between gesture and speech of multimodal communicative behaviors in the crucial transition from babbling to two-word productions (Murillo et al., 2018). It found that, at 15 months, the proportion of overlap between gestural and vocal elements, and the proportion of gesture strokes overlapping with vocalizations, were both related to lexical development 3 months later. Both studies thus provide evidence about the importance of temporal coordination between gesture and speech for early communication and language learning.

A step forward in the early patterns of pointing gesture-speech integration appears later on, at around 4–5 years of age, when children start developing discourse-based strategies. Manual beat gestures (e.g., hand, arm, or head movements typically associated with prosodic prominence) are used by children to focus specific information in discourse (Nicoladis et al., 1999; Capone and McGregor, 2004; Mathew et al., 2014; Esteve-Gibert et al., 2019). Beat gestures accompanying prosodic prominences constitute a clear example of a more refined multimodal gesture-speech unit used as a discourse-integrated focus-marking strategy. Interestingly, several studies have also shown that beat gestures have a beneficial effect on 4-year-old children's recall and comprehension of narratives (Llanes-Coromina et al., 2018).

Across languages, contrastive focus marking consistently makes a later appearance than information focus marking. Speer and Ito (2009) and Ito et al. (2012) showed that children could only interpret intonation contour signaling contrasts after the age of 6. Kurumada and Clark (2017) showed that children may be sensitive to contrastive focus intonation as early as age 4. Looking at the production of contrastive focus by considering prosody and gesture alongside each other, Esteve-Gibert et al. (2019) found that 5-year-old French children already use head nods to mark contrastive focus and tend to use this strategy well before they apply prosodic strategies (specifically, syllable-duration-marking strategies). Interestingly, studies on a variety of languages, such as German, English, and Mandarin Chinese, using focus particles like *only*, *also*, or *even*, have shown that children acquire these particles by ages 4 or 5 (Crain et al., 1994; Paterson et al., 2003; Notley et al., 2009; Höhle et al., 2016).

The results reviewed in this section suggest that prelexical children already have considerable knowledge of information focus deriving from a combination of prosodic and gestural features. From the point of view of perception, infants benefit from the use of prosodic and gestural prominence features marking focus in IDS for lexical learning. From the point of view of production, at the prelexical stage infants use integrated productions of prosodic and gestural prominence to initiate joint attention scenarios, and further, to signal simple realizations of information focus. These integrated multimodal units can be regarded as precursors of information focus marking because they represent information status at a rudimentary level. At a later stage, multimodal integration of gesture and prosodic prominence at the discourse level is used by older children in

their early use of beat gestures to mark discourse novelty and discourse contrast.

The Development of Epistemic Stance

Prosody and gesture can be strong markers of epistemic stance across languages. Speakers use different prosodic cues such as slower speech rate and higher pitch levels (Krahmer and Swerts, 2005; Roseano et al., 2016), and gestures such as eyebrow furrowing, head tilt, or shoulder shrugging, to indicate ignorance or uncertainty states (Krahmer and Swerts, 2005; Dijkstra et al., 2006).

Twelve-month-old infants already demonstrate sensitivity to other people's epistemic states, using contextual cues to distinguish between knowledgeable and ignorant partners. In two experiments, Liszkowski et al. (2008) found that 12-month-olds pointed more often to an object whose location was unknown to their adult interlocutor than to an object whose location was known to the adult. However, it is not until the preschool years that children start to fully understand another person's beliefs or attitudes and can verbalize such things themselves. Research on the development of epistemic stance, and more specifically, knowledge state (certainty and uncertainty), has focused on the acquisition of lexical and verbal epistemic markers (see Matsui, 2014 for an overview over linguistic expressions of certainty and evidentiality). To situate the age frame for acquisition of lexical and grammatical epistemic markers in relation to perception, it has been found that English-speaking children are able to understand the contrast between *I know* and *I think* by age 4 (Moore et al., 1989). At a similar age they also understand the contrast between *must* and *might* (Moore et al., 1990). Similar results have been found for Korean (Choi, 1995; Papafragou et al., 2007), Cantonese (Lee and Law, 2001; Tardif et al., 2004), Turkish and Puerto Rican Spanish (Shatz et al., 2003), Japanese (Matsui et al., 2006), and Japanese and German (Matsui et al., 2009). Even earlier comprehension of speaker uncertainty has been found in Japanese, where at age 3 children were already able to detect a speaker's uncertainty encoded through grammatical particles (*yo* = direct evidence vs. *tte* = heresy), compared to a later acquisition of verbal forms that served the same function (Matsui et al., 2006).

Drawing accurate conclusions about another speaker's knowledge is quite a difficult task that requires both conceptual and linguistic maturity. One early study by Moore et al. (1993) compared children's (3–6 years old) comprehension of mental state lexicon to their comprehension of mental state prosody. Children had to listen to contrasting pairs of statements by two puppets and guess the location of a hidden object. Each statement pair either differed with respect to the mental state verbs, e.g., *know* vs. *think* or *think* vs. *guess*, or with respect to terminal pitch contour (e.g., falling or rising). While 3-year-olds were not able to use either lexicon or prosody to detect where the object was, 4-year-olds started to do so significantly better in the *know* vs. *think* and falling vs. rising pitch contrast conditions. Furthermore, the *think* vs. *guess* condition was much harder, even for the 5-year-old children, than the *know* vs. *think* condition. Testing whether gestural and prosodic cues might facilitate this task for children, a recent experimental

study investigated 3- to 5-year-old Catalan-speaking children's understanding of a speaker's uncertainty on the basis of lexical cues (*potser* 'maybe'), intonational cues (rising intonation), or gestural cues (head tilt and raised eyebrows) (Hübscher et al., 2017b). In a between-subjects design, the children were either exposed to the lexical condition (where they received lexical and gestural cues to uncertainty) or the intonation condition (where they were exposed to intonational and gestural cues to uncertainty). Within each condition, three different presentation formats were used (audio-only, visual-only and audiovisual) as within-subject variables. The results showed that the 3-year-olds performed significantly above chance level when detecting a speaker's uncertainty on the basis of facial gestural cues, and also performed significantly better when dependent on intonation compared to lexical cues. Similarly, Armstrong et al. (2018) tested 3- to 5-year-old children's understanding of another type of belief state, a speaker's incredulity, encoded solely in prosody and gesture, and found that the children performed best when they had access to both cues together, providing similar evidence for a parallel development of prosody and gesture in children's understanding of another speaker's belief state.

Several studies have provided evidence that children in their preschool years already display some awareness of their own knowledge state (Pillow and Anderson, 2006; Balcomb and Gerken, 2008; Nilsen et al., 2008; Lyons and Ghetti, 2011, 2013; Paulus et al., 2013). Others have found that children do not acquire this feeling of knowledge before they are 7 (e.g., Beck and Robinson, 2001; Flavell et al., 2001; Pillow and Anderson, 2006). Looking at the production of one's own knowledge state, it has been detected that English-speaking children at age 3 are able to verbally report on their knowledge state when they are in either total ignorance or total knowledge states, but they struggle to signal partial ignorance until they reach school age (Rohwer et al., 2012). Yet evidence from the analysis of multimodal signaling of ignorance and uncertainty shows that children can do this at an even earlier age. First, with regard to signaling ignorance, a longitudinal study with 64 US-English children who were recorded every 4 months from 14 to 42 months found that at around age 2 the children begin to signal their ignorance through gestural cues such as flipping their palms upward and outward or to the side (Bartz, 2017). At 22 months, one-fifth of the sample had been observed producing such a hand flip to signal ignorance, and by 42 months almost half had done so. Verbal statements of ignorance, such as *I don't know*, emerged later between 22 and 26 months and then increased rapidly. Furthermore, observing children's gestural and other non-verbal cues, Kim et al. (2016) showed that when they had only partial access to objects hidden in a box, 4-year-olds displayed their uncertainty by employing a set of uncertainty/ignorance gestures like tilting their head to one side, shaking their head, shrugging their shoulders, or looking away. While they were not yet able to verbally report on their knowledge state, they used gestural cues to mark their uncertain knowledge state. Another study focused on Catalan children's uncertain stance signaling (Hübscher, 2018). A total of 40 3- to 5-year-old Catalan children participated in a guessing game involving five easy and previously touched and seen objects, and five difficult and not previously seen objects. The results showed

that children could already signal uncertainty at age 3 using a set of gestural cues (such as tilted head and raised eyebrows) and prosodic cues (rising intonation, fillers and vowel lengthening), and only later began to employ lexical devices like *maybe* or *I think it is*.

In sum, research has provided converging evidence that preschool-aged children first comprehend and produce knowledge state information, in particular, uncertainty, through gestural and prosodic cues before they do so through appropriate lexical marking. Thus we can view these epistemic multimodal markers as playing a precursory and foundational role during sociopragmatic development.

The Development of Politeness

There is no consensus in the research about when children's awareness of politeness emerges (Shatz and Gelman, 1973; Andersen, 1977; Hollos, 1977; James, 1978; Axia and Baroni, 1985; Baroni and Axia, 1989). Based on studies investigating children's understanding of lexical cues in relation to a speaker's politeness, Andersen (1977) suggested that English-speaking children start to understand how different lexical politeness markers should be used, depending on the status asymmetry between the speaker and addressee, at around age 4. It has also been found that children's ability to appropriately use the different lexical and morphosyntactic forms and conventions conveying politeness takes quite some time to develop in a number of languages, including English (e.g., James, 1978; Ervin-Tripp and Gordon, 1986; Sealey, 1999), French (Marcos and Bernicot, 1994; Ryckebusch and Marcos, 2004), Swedish (Aronsson and Thorell, 1999), Norwegian and Hungarian (Hollos and Beeman, 1978) and Greek (Georgalidou, 2008). In Japanese, in which a wide array of politeness attitudes must be acquired, it has been suggested that children's awareness of the use of the complex politeness forms in gratitude and apology situations begins as early as age 6, but does not approximate that of adults until much later, between ages 13 and 16 (Long, 2010). One of the only and most comprehensive studies with a focus on preschool children's developing perception (and production) of politeness by including intonation as a cue is Bates's (1976) study. Bates (1976) experimentally tested whether 60 Italian children aged 3–6 could perceive politeness as encoded in lexical cues or through prosody only. The results showed that the children had acquired the understanding of *per favore*, please, as a politeness marker by age 3, but the use of gentle intonation as a strategy only became significant after 4 years of age.

Some studies have found that facial expressions and affective prosody can act as facilitatory cues in the early detection of politeness by children. One study investigated Japanese 9- and 10-year-olds' ability to infer a speaker's politeness vs. impoliteness (Shochi et al., 2009). It found that the presence of facial cues was clearly beneficial for the processing of politeness and impoliteness meanings in this age range. In order to investigate the early comprehension of politeness, and focusing on intonation as a cue, a recent study addressed the question of whether 3-year-old children would be better able to infer a speaker's polite stance in a request on the basis of intonational cues or facial cues, or through a combination of the two cues (Hübscher et al., 2018).

The results showed that children were able to infer significantly above chance, which one of two speakers produced more friendly requests for all three conditions. This demonstrated that children could already detect politeness by exclusively using prosodic or gestural cues at an early age when they were still acquiring the various lexical and morphosyntactic politeness markers.

A recent study investigated children's production of politeness-related cues in request interactions with varying sociopragmatic factors of cost and social distance (Hübscher, 2018; Hübscher et al., 2019). Surprisingly, already at 3 years of age and increasingly more so between 4 and 5 years, children prosodically adapted their speech in a way similar to adults, when the cost of the request was increased. They also modified their gestural behavior when talking to an adult they did not know, by showing, for example, more raised eyebrows, head tilts, smiles, or slumped shoulders. In general, preschoolers seem able to use both prosody and gesture as mitigating strategies when making requests, either when they imply a high cost or are addressed to a stranger. These results are in line with recent work on the adult expression of politeness indicating a cross-linguistic tendency to use prosodic and gestural mitigating strategies when the interlocutor is of higher status. More specifically, in relation to prosody it has been found that speakers' polite speech has a slower speech rate, a lower intensity, lower pitch, and also the use of a breathy voice quality (Grawunder and Winter, 2010; Winter and Grawunder, 2012; Hübscher et al., 2017a). Gesture goes in a similar direction. While in English it has been shown that adults display an array of gestural mitigation cues such as raised eyebrows, direct body orientation, and a tense, closed position with small gestures, accompanied by a soft voice (Tree and Manusov, 1998), in Korean speakers use fewer gestures and less physical contact in the polite register, while at the same time using more head nods and bows (Brown et al., 2015).

In short, there is increasing evidence that gesture and prosody not only play a precursor role in children's pragmatic development during infancy, but also, acting in tandem, continue to play a vital role when children are already able to talk and are acquiring more complex socio-pragmatic skills. This provides clear proof of the integrated nature of gesture and prosody and their shared role as foundational stones in sociopragmatic development.

DISCUSSION AND CONCLUSION

The present review paper has shown converging evidence from recent research about the pivotal role that prosodic and gesture patterns play in children's sociopragmatic development across languages by focusing on children's communicative behaviors between the ages of 1 and 5. In general, the empirical evidence shows that typically developing children achieve pragmatic milestones in gesture and prosody before attaining the same milestones in their use of spoken language forms, and this pattern is broadly consistent across children, languages, and stages of development. While research has traditionally focused on the role of prosody and gesture in the early acquisition of lexical and morphosyntactic elements, fewer studies have concentrated

on the ways in which prosody and gesture may be involved in how children learn to express pragmatic meanings. It is those few studies that have been the object of our attention in this review.

First, our review has highlighted the fact that early prosodic and gestural patterns in children seem to overlap in terms of the pragmatic functions for which they are used, which range from speech act and information structure marking to stance-related meanings. There is empirical evidence that the two components are closely linked to each other in crucial phases of development, such that gesture and speech are integrated from a temporal and pragmatic point of view into composite multimodal productions. Children learn to perceive and perform speech acts and 12-month-old infants are able to successfully infer whether an adult's pointing gestures, together with specific prosodic patterns, convey a declarative, requestive, or expressive social intention (Esteve-Gibert et al., 2017). In many languages, children are also able to produce multimodal combinations of pointing gestures with specific prosodic patterns to express requestive or declarative acts (Grünloh and Liszkowski, 2015; Murillo and Capilla, 2016; Aureli et al., 2017).

Second, regarding the marking of information focus, in accordance with Snow (2017), we have argued that the multimodal integration of pointing gestures and prosodic prominence apparent at 12 months can be considered an expressive precursor of information focus. Moreover, prosodic focus cues present in IDS cues are widely seen to afford children early access to lexical meanings (see Thorson, 2018 for a review) and probably help them to gain access to the form-meaning mapping of lexical pragmatic items acquired later, like epistemic stance and politeness. Two cross-sectional studies showed that 3-year-olds were able to express knowledge state and politeness meanings by modulating prosodic and gestural cues well before they started using verbal strategies (Hübscher, 2018; Hübscher et al., 2019). When expressing politeness, these 3-year-olds used gestural and prosodic cues of mitigation very similar to those employed by adults to express politeness in contexts involving higher social status interlocutors and higher pragmatic costs. Overall, the evidence supports the claim that children at different stages of development are able to identify and use multimodal units that integrate gestural and prosodic features as markers of sociopragmatic meaning.

The evidence reviewed here has also shown that gesture and pragmatic prosody are not only acquired in tandem by children, but also seem to both precede and predict changes in children's more complex sociopragmatic marking involving lexical forms. Though a few studies have shown that children can begin to express pragmatic meaning by way of gesture alone (e.g., Beaupoil-Hourdel et al., 2015 study illustrating how children undergo a change from expressing negation through embodiment then later move on to symbolic negation, where gesture and speech are completely integrated into composite multimodal productions), recent studies have shown that when gestural and prosodic integration are assessed together, children use this combination of channels to understand and produce a variety of pragmatic meanings before they are able to use spoken language forms involving lexical and morphosyntactic strategies (see sections "The Development of Speech Act Marking

in Infancy," "The Development of Information Focus," "The Development of Epistemic Stance," and "The Development of Politeness").

The precursor role of gesture and prosody should not be surprising. Under normal circumstances children acquire language and develop form-meaning mappings in multimodal settings. It is precisely these multimodal cues that boost children's early access to meanings and help them gain access to the form-meaning mapping of lexical pragmatic items acquired later (e.g., the usage-based approach to language, Tomasello, 2003). The existing research also confirms the important role of IDS, with its exaggerated prosodic and gestural features, in shaping children's language learning trajectory.

Even though most of the evidence tends to point to a synchronous development of prosody and gesture, it is also important to try to disentangle the specific trajectories of prosodic and gestural features across developmental stages and languages. In particular, future studies will need to determine whether prosodic features or gestural features, or both, act as independent or integrated precursors of language learning and pragmatic development in relation to the exact path of acquisition at the onset of a particular meaning. Initial results from some recent work on the precursor role of focus prosody and the use of head nods (Esteve-Gibert et al., 2019), seem to indicate that they are independent of each other. Specifically, the results of this study showed that preschoolers (5-year-olds) do not seem to use typical acoustic-prosodic marking in order to distinguish between old, new, and corrective focus information in discourse, at a time when they already use head nod patterns to mark corrective focus. Future research will need to better disentangle the temporal pathway of the acquisition of prosodic and gestural cues encoding different pragmatic meanings.

In general, our review of previous work has revealed the need for a more integrated and detailed multimodal approach to the study of prosody and gesture within the context of language acquisition. First, very few studies have assessed prosody and non-verbal expressive means, e.g., gestures, on the same developmental sample. Second, while developmental studies on gesture and prosody have focused on the strict analysis of manual gestures and intonational pitch patterns (see also Snow, 2017), recent findings have shown the relevance of analyzing non-manual gestures such as head, face, or shoulder movements and prosodic features other than pitch, like intensity, duration, and voice quality. Future work will need to assess, in a more systematic fashion, the overlapping versus the distinctive functions of gesture and prosody with respect to pragmatic encoding; that is, the degree of systematicity of the prosodic and gestural cues as signals of different pragmatic functions, as well as their potential degree of overlap from a functional point of view. Previous studies have found that often different articulators are jointly involved in the marking of one specific pragmatic function, as it is the case, for example, with a speaker's ignorance: a speaker can use a range of different articulator movements such as head shake, raised eyebrows, lips down, open hand palm up gestures and raised shoulders) to signal unknowingness (see also Brown and Prieto, unpublished). Some of these articulators can also be involved in the marking of other

pragmatic meanings. For example, when marking focus, speakers typically raise their eyebrows too. Kendon (2002) analyzed a large number of 'head shakes' in naturally occurring interactions in Italian and reported eight different kinds of uses for the head shake, with different pragmatic uses. Even though in the context of language acquisition, most of the studies on the development of different pragmatic functions have investigated individual cues, future research will need to more carefully tease out the development of several and partially overlapping cues used to signal a pragmatic function. Ideally this will be done by taking both an observational and an experimental approach, in such a way that they inform each other's findings.

More research will also be needed to confirm the value of simultaneous gesture–speech combinations in specific socio-communicative contexts, and their potential predictive value for later language development. There is evidence that early gesture and prosodic integration may be paving the way for future developments in a child's language skills and that an early command of these multimodal features can have a predictive value in the emergence of later linguistic and grammatical strategies. For example, two studies have shown that the degree to which 12-month-old infants use integrated and simultaneous pointing–speech combinations predicts their later language development. The ability to selectively use this multimodal communicative strategy to successfully initiate and maintain joint attention at 12 months is related to language development at 18 months (Igalada et al., 2015; for similar findings see Murillo et al., 2018).

It would be of interest to gather further evidence from later stages of acquisition such as narrative and discourse development. Although recent evidence points toward the predictor role of iconic gestures (specifically character-viewpoint gestures) used by 5-year-olds for narrative structure at later ages (Demir et al., 2014), little is known about the joint role of both gesture and prosody in the process of narrative development. However, recent work strongly focuses on gestures, as seen in the recent book by McNeill on the analysis of cohesion and clause-linking devices in spontaneous narratives produced by children aged 2 and 3, which traces the changing relationships between speech and gesture in this type of discourse (McNeill, 2016). In our view, in order to gain a complete understanding of the steps by which children acquire language, researchers need to take a multimodal perspective by assessing how the relationship between the four components of gesture, prosody, lexicon, and syntax changes over time. Finally, there is clearly also a need to include more typologically different languages in the analysis in order to assess if and to what degree the relationship between prosody, gesture and other verbal elements of language is shaped by different linguistic systems.

Research has provided evidence that pragmatic development is a multimodal venture in which semantically integrated gesture-prosody units act as foundational elements. Given this, it is not surprising that multimodal language training is being used successfully in language intervention settings. However, these interventions have been developed by separately addressing the gestural and prosodic aspects of language. On the one hand, a set of studies has shown that interventions that teach gesture and

verbal imitation skills to young children with Autism Spectrum Disorder (ASD) lead to greater gains in the rate of language use (Ingersoll and Schreibman, 2006; Ingersoll et al., 2007; Ingersoll, 2008). On the other hand, speech therapists have employed interventions that mainly use prosodic features for treatment. One example is Melodic-Based Communication Therapy, which has been found to improve expressive vocabulary, verbal imitative abilities, and pragmatic skills in children with ASD by combining standard melodies and positive reinforcement (Sandiford et al., 2013). Another example, in this case intended to treat patients with non-fluent aphasia, is Melodic Intonation Therapy (MIT), a technique that uses melody and rhythm to improve the patient's expressive language (Norton et al., 2009). Future research in applied developmental treatments would benefit from more holistic approaches that integrate approaches to language learning by including key components of pragmatic development, specifically prosody and gesture. Interestingly, a recent study assessing gesture- and prosody-based narrative training through beat gestures revealed that training 5-year-old children with the observation of beat gestures boosted their narrative performances (Vilà-Giménez et al., 2019). By the same token, an integrated assessment of these components should be included in present standard assessments of children's pragmatic skills (Phelps-Terasaki and Phelps-Gunn, 2007; Wiig et al., 2013; Carrow-Woolfolk, 2017).

In summary, the evidence presented in this article has shown that children's development of pragmatic skills is inherently multimodal, and that the multimodal features of language pave the way for children's sociopragmatic development later in life. We thus argue for the need to incorporate a multimodal view of language into the field of language acquisition. Prosody and gesture are crucial pieces of the foundational puzzle that children have to solve when developing (socio)pragmatic skills, and it is therefore indispensable that they are included in any enquiry in this field. Although we are still far from having a comprehensive picture of the childhood development of prosody and gesture in communication, a more holistic approach, where the different functions and interactions of prosody and gesture are studied in concert with one another, is sure to enrich the field.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

We would like to thank the reviewers and the editor for valuable comments. This article is partly based on, and was inspired by, the results of research funded by the Spanish Ministry of Science and Innovation (Grant No. FFI2015-66533-P "Intonational and gestural meaning in language" and Grant No. PGC2018-097007-B-I00 "Multimodal Language Learning") and a grant awarded by the Generalitat de Catalunya (Grant No. 2017SGR-925) to the Prosodic Studies Group.

REFERENCES

- Acredolo, L. P., and Goodwyn, S. W. (1985). Symbolic gesturing in language development. *Hum. Dev.* 28, 40–49. doi: 10.1159/000272934
- Ambraziatis, G., and House, D. (2017). Multimodal prominences: Exploring the patterning and usage of focal pitch accents, head beats and eyebrow beats in Swedish television news readings. *Speech Commun.* 95, 100–113. doi: 10.1016/j.specom.2017.08.008
- Andersen, E. (1977). Young children's knowledge of role-related speech difference: a mommy is not a daddy is not a baby. *Papers Presented at the 9th Annual on Child Language Development, Stanford, CA*. doi: 10.1016/j.specom.2017.08.008
- Armstrong, M. E., Esteve-Gibert, N., Hübscher, I., Igualada, A., and Prieto, P. (2018). Developmental and cognitive aspects of children's disbelief comprehension through intonation and facial gesture. *First Lang.* 38:014272371878927.
- Aronsson, K., and Thorell, M. (1999). Family politics in children's play directives. *J. Pragmat.* 31, 25–47.
- Aureli, T., Spinelli, M., Fasolo, M., Consetta Garito, M., Perruchini, P., and D'Odorico, L. (2017). The pointing–vocal coupling progression in the first half of the second year of life. *Infancy* 22, 801–818. doi: 10.1111/inf.12181
- Axia, G., and Baroni, M. R. (1985). Linguistic politeness at different age levels. *Child Dev.* 56, 918–927. doi: 10.1111/j.1467-8624.1985.tb00165.x
- Balcomb, F. K., and Gerken, L. (2008). Three-year-old children can access their own memory to guide responses on a visual matching task. *Dev. Sci.* 11, 750–760. doi: 10.1111/j.1467-7687.2008.00725.x
- Bara, B. G. (2010). *Cognitive Pragmatics*. Cambridge, MA: MIT Press.
- Baroni, M. R., and Axia, G. (1989). Children's meta-pragmatic abilities and the identification of polite and impolite requests. *First Lang.* 9, 285–297. doi: 10.1177/014272378900902703
- Bartz, D. T. (2017). *Young Children's Meta-Ignorance*. Ph.D. dissertation, Harvard University, Cambridge, MA.
- Bates, E. (1976). *Language and Context: the Acquisition of Pragmatics*. New York, NY: Academic Press.
- Bates, E., Benigni, L., Bretherton, I., Camaioni, L., and Volterra, V. (1979). *The Emergence of Symbols: Cognition and Communication in Infancy*. New York, NY: Academic Press.
- Bates, E., Camaioni, L., and Volterra, V. (1975). The acquisition of performatives prior to speech. *Merrill Palmer Q.* 21, 205–226.
- Beaupeil-Hourdel, P., Morgenstern, A., and Boutet, D. (2015). “A child's multimodal negotiations from 1 to 4: the interplay between modalities,” in *Negation and Polarity: Experimental Perspectives*, Vol. 1, eds P. Larrivé, and C. Lee (Cham: Springer), 95–123.
- Beck, S. R., and Robinson, E. J. (2001). Children's ability to make tentative interpretations of ambiguous messages. *J. Exp. Child Psychol.* 79, 95–114. doi: 10.1006/jecp.2000.2583
- Behne, T., Liszkowski, U., Carpenter, M., and Tomasello, M. (2012). Twelve-month-olds' comprehension and production of pointing. *Br. J. Dev. Psychol.* 30, 359–375. doi: 10.1111/j.2044-835X.2011.02043.x
- Bolinger, D. (1986). *Intonation and Its Parts: Melody in Spoken English*. Stanford, CA: Stanford University Press.
- Bosco, F. M., Angeleri, R., Colle, L., Sacco, K., and Bara, B. G. (2013). Communicative abilities in children: an assessment through different phenomena and expressive means. *J. Child Lang.* 40, 741–778. doi: 10.1017/s0305000913000081
- Brown, L., and Winter, B. (2018). Multimodal indexicality in Korean: “Doing Deference” and “Performing Intimacy” through Nonverbal Behavior. *J. Politeness Res. Lang. Behav. Cult.* 15, 25–54. doi: 10.1515/pr-2016-0042
- Brown, L., Winter, B., Idemaru, K., and Grawunder, S. (2014). Phonetics and politeness: perceiving Korean honorific and non-honorific speech through phonetic cues. *J. Pragmat.* 66, 45–60. doi: 10.1016/j.pragma.2014.02.011
- Brown, L., Winter, B., Idemaru, K., and Grawunder, S. (2015). The sound of honorific language: How speech indexes social meaning in Korean, Japanese, German and Russian. *Paper Presented at the IPRA, Antwerp*.
- Butcher, C., and Goldin-Meadow, S. (2000). “Gesture and the transition from one- to two-word speech: When hand and mouth come together,” in *Language and Gesture*, ed. D. McNeill (New York, NY: Cambridge University Press), 235–257.
- Cameron-Faulkner, T. (2014). “The development of speech acts,” in *Pragmatic Development in First Language Acquisition*, ed. D. Matthews (Amsterdam: John Benjamins Publishing Company), 37–52.
- Capirci, O., Iverson, J. M., Pizzuto, E., and Volterra, V. (1996). Gestures and words during the transition to two-word speech. *J. Child Lang.* 23, 645–673. doi: 10.1017/s0305000900008989
- Capirci, P., Contaldo, A., Caselli, M. C., and Volterra, V. (2005). From action to language through gesture: a longitudinal perspective. *Gesture* 5, 155–177. doi: 10.1075/gest.5.1-2.12cap
- Capone, N. C., and McGregor, K. K. (2004). Gesture development: a review for clinical and research practices. *J. Speech Lang. Hear. Res.* 47, 173–186. doi: 10.1044/1092-4388(2004)015
- Carrow-Woolfolk, E. (2017). *Comprehensive Assessment of Spoken Language, Second Edition (CASL-2)*. Torrance, CA: Western Psychological Services.
- Cavalho, A., Dautriche, I., Millotte, S., and Christophe, A. (2018). “Early perception of phrasal prosody and its role in syntactic and lexical acquisition,” in *The Development of Prosody in First Language Acquisition*, eds P. Prieto and N. Esteve-Gibert (Amsterdam: John Benjamins), 17–35. doi: 10.1075/tilar.23.02car
- Chen, A. (2018). “Get the focus right across languages: acquisition of prosodic focus-marking in production,” in *The Development of Prosody in First Language Acquisition*, eds P. Prieto, and N. Esteve-Gibert (Amsterdam: John Benjamins Publishing Company), 295–316.
- Choi, S. (1995). “The development of epistemic sentence-ending modal forms and functions in Korean children,” in *Modality in Grammar and Discourse*, eds J. Bybee, and S. Fleischman (Amsterdam: John Benjamins), 165–204.
- Cochet, H., and Vaclair, J. (2010). Features of spontaneous pointing gestures in toddlers. *Gesture* 10, 86–107. doi: 10.1075/gest.10.1.05coc
- Crain, S., Ni, W., and Conway, L. (1994). “Learning, parsing and modularity,” in *Perspectives on Sentence Processing*, eds C. Clifton, L. Frazier, and K. Rayner (Hillsdale, NJ: Lawrence Erlbaum), 443–467.
- Cristià, A. (2011). Fine-grained variation in caregivers' speech predicts their infants' discrimination. *J. Acous. Soc. Am.* 129, 3271–3280. doi: 10.1121/1.3562562
- De Ruiter, J.-P. (1998). *Gesture and Speech Production*. Ph.D. thesis, University of Nijmegen, Nijmegen.
- Demir, ÖE., Levine, S. C., and Goldin-Meadow, S. (2014). A tale of two hands: children's early gesture use in narrative production predicts later narrative structure in speech. *J. Child Lang.* 42, 662–681. doi: 10.1017/s0305000914000415
- Dijkstra, C., Krahmer, E., and Swerts, M. (2006). Manipulating uncertainty: the contribution of different audiovisual prosodic cues to the perception of confidence. *Paper Presented at the Speech Prosody, Dresden*.
- Dohen, M., and Loevenbruck, H. (2009). Interaction of audition and vision for the perception of prosodic contrastive focus. *Lang. Speech* 52(Pt 2–3), 177–206. doi: 10.1177/0023830909103166
- Ervin-Tripp, S., and Gordon, D. (1986). “The development of children's requests,” in *Communicative Competence: Assessment and intervention*, ed. R. Schiefelbusch (San Diego, CA: College Hill Press), 61–96.
- Esposito, A., Esposito, D., Refice, M., Savino, M., and Shattuck-Hufnagel, S. (2007). “A preliminary investigation of the relationship between gestures and prosody in Italian,” in *Fundamentals of Verbal and Nonverbal Communication and the Biometric Issue*, eds A. Esposito, M. Bratanic, E. Keller, and M. Marinaro (Amsterdam: IOS Press), 45–57.
- Esteve-Gibert, N., Borràs-Comes, J., Asor, E., Swerts, M., and Prieto, P. (2017). The timing of head movements: the role of prosodic heads and edges. *J. Acous. Soc. Am.* 141, 4727–4739. doi: 10.1121/1.4986649
- Esteve-Gibert, N., and Guellai, B. (2018). The links between prosody and gestures: a developmental perspective. *Front. Psychol.* 9:338. doi: 10.1121/1.4986649
- Esteve-Gibert, N., Liszkowski, U., and Prieto, P. (2016). “Prosodic and gesture features distinguish the pragmatic meaning of pointing gestures in child-directed communication,” in *Intonational Grammar in Ibero-Romance: Approaches Across Linguistic Subfields*, eds M. E. Armstrong, N. Henriksen, and M. M. Vanrell (Amsterdam: John Benjamins), 251–275.
- Esteve-Gibert, N., Loevenbruck, H., Dohen, M., and D'Imperio, M. (2019). Preschoolers use head gestures rather than duration or pitch range to signal narrow focus in French. *Paper Presented at the 19th International Congress of Phonetic Sciences, Melbourne*.

- Esteve-Gibert, N., and Prieto, P. (2012). Prosody signals the emergence of intentional communication in the first year of life: evidence from Catalan-babbling infants. *J. Child Lang.* 40, 919–944.
- Esteve-Gibert, N., and Prieto, P. (2013). Prosody signals the emergence of intentional communication in the first year of life: evidence from Catalan-babbling infants. *J. Child Lang.* 40, 919–944. doi: 10.1017/S0305000912000359
- Esteve-Gibert, N., and Prieto, P. (2014). Infants temporally coordinate gesture-speech combinations before they produce their first words. *Speech Commun.* 57, 301–316. doi: 10.1016/j.specom.2013.06.006
- Esteve-Gibert, N., and Prieto, P. (2018). “Early development of prosody-meaning interface,” in *The Development of Prosody in First Language Acquisition*, eds N. Esteve-Gibert and P. Prieto (Amsterdam: John Benjamins), 228–246.
- Esteve-Gibert, N., Prieto, P., and Pons, F. (2015). Nine-month-old infants are sensitive to the temporal alignment of prosodic and gesture prominences. *Infant Behav. Dev.* 38, 126–129. doi: 10.1016/j.infbeh.2014.12.016
- Fernald, A., and Mazzei, C. (1991). Prosody and focus in speech to infants and adults. *Dev. Psychol.* 27, 209–221. doi: 10.1037/0012-1649.27.2.209
- Ferré, G. (2011). Functions of three open-palm hand gestures. *Multimodal Commun.* 1, 5–20.
- Flavell, J. H., Flavell, E. R., and Green, F. L. (2001). Development of children's understanding of connections between thinking and feeling. *Psychol. Sci.* 12, 430–432. doi: 10.1111/1467-9280.00379
- Frota, S., and Butler, J. (2018). “Early development of intonation: perception and production,” in *The Development of Prosody in First Language*, eds P. Prieto and N. Esteve-Gibert (Amsterdam: John Benjamins), 145–165.
- Frota, S., Matos, N., Cruz, M., and Vigário, M. (2016). “Early prosodic development: Emerging intonation and phrasing in European Portuguese,” in *Issues in Hispanic and Lusophone Linguistics: Interdisciplinary Approaches to Intonation Grammar in Ibero-Romance*, eds M. Armstrong, M. M. Vanrell, and N. C. Henriksen (Philadelphia, PA: Benjamins), 295–324. doi: 10.1075/ihll.6.14fro
- Georgalidou, M. (2008). The contextual parameters of linguistic choice: greek children's preferences for the formation of directive speech acts. *J. Pragmat.* 40, 72–94. doi: 10.1016/j.pragma.2007.08.009
- Gleitman, L. R., Gleitman, H., Landau, B., and Wanner, E. (1988). “Where learning begins: Initial representations for language learning,” in *Language: Psychological and Biological Aspects*, ed. F. J. Newmeyer (Cambridge: Cambridge University Press), 150–193.
- Gleitman, L. R., and Wanner, E. (1982). “Language acquisition: the state of the art,” in *Language Acquisition: the State of the Art*, eds E. Wanner, and L. R. Gleitman (Cambridge: Cambridge University Press).
- Goldin-Meadow, S. (1998). *The Development of Gesture and Speech as an Integrated System*. San Francisco, CA: Jossey-Bass.
- Goldin-Meadow, S. (2009). How gesture promotes learning throughout childhood. *Child Dev. Perspect.* 3, 106–111. doi: 10.1111/j.1750-8606.2009.00088.x
- Grawunder, S., and Winter, B. (2010). Acoustic correlates of politeness: prosodic and voice quality measures in polite and informal speech of Korean and German speakers. *Speech Prosody* 2010, 14–17.
- Graziano, M., and Gullberg, M. (2018). When speech stops, gesture stops: evidence from developmental and crosslinguistic comparisons. *Front. Psychol.* 9:879. doi: 10.3389/fpsyg.2018.00879
- Grünloh, T., Lieven, E., and Tomasello, M. (2015). Young Children's intonational marking of new, given and contrastive referents. *Lang. Learn. Dev.* 11, 95–127. doi: 10.1080/15475441.2014.889530
- Grünloh, T., and Liszkowski, U. (2015). Prelinguistic vocalizations distinguish pointing acts. *J. Child Lang.* 42, 1312–1336. doi: 10.1017/S0305000914000816
- Höhle, B., Fritzsche, T., and Müller, A. (2016). Children's comprehension of sentences with focus particles and the role of cognitive control: an eye tracking study with German-learning 4-year-olds. *PLoS One* 11:e0149870. doi: 10.1371/journal.pone.0149870
- Hollos, M. (1977). “Comprehension and use of social rules in pronoun selection by Hungarian children,” in *Child Discourse*, eds S. Ervin-Tripp and C. Mitchell-Kernan (New York, NY: Academic Press), 211–223. doi: 10.1016/b978-0-12-241950-8.50017-4
- Hollos, M., and Beeman, W. (1978). The development of directives among Norwegian and Hungarian children: an example of communicative style in culture. *Lang. Soc.* 7, 345–355. doi: 10.1017/S0047404500005765
- Hübscher, I. (2018). *Preschoolers' Pragmatic Development: How Prosody and Gesture Lend a Helping Hand*. Ph.D. thesis, Pompeu Fabra University, Barcelona.
- Hübscher, I., Borràs-Comes, J., and Prieto, P. (2017a). Prosodic mitigation characterizes Catalan formal speech: the frequency code reassessed. *J. Phon.* 65, 145–159. doi: 10.1016/j.wocn.2017.07.001
- Hübscher, I., Esteve-Gibert, N., Igualada, A., and Prieto, P. (2017b). Intonation and gesture as bootstrapping devices in speaker uncertainty. *First Lang.* 37, 24–41. doi: 10.1177/0142723716673953
- Hübscher, I., Garufi, M., and Prieto, P. (2019). The development of polite stance in preschoolers: how prosody, gesture and body cues pave the way. *J. Child Lang.*
- Hübscher, I., Wagner, L., and Prieto, P. (2018). Three-year-olds infer polite stance from intonation and facial cues. *J. Politeness Res.*
- Igualada, A., Bosch, L., and Prieto, P. (2015). Language development at 18 months is related to multimodal communicative strategies at 12 months. *Infant Behav. Dev.* 39, 42–52. doi: 10.1016/j.infbeh.2015.02.004
- Ingersoll, B. (2008). The social role of imitation in autism: implications for the treatment of imitation deficits. *Infants Young Child.* 21, 107–119. doi: 10.1016/j.neuro.2012.10.008
- Ingersoll, B., Lewis, E., and Kroman, E. (2007). Teaching the imitation and spontaneous use of descriptive gestures in young children with autism using a naturalistic behavioral intervention. *J. Autism Dev. Disord.* 37, 1446–1456. doi: 10.1007/s10803-006-0221-z
- Ingersoll, B., and Schreibman, L. (2006). Teaching reciprocal imitation skills to young children with autism using a naturalistic behavioral approach: Effects on language, pretend play, and joint attention. *J. Autism Dev. Disord.* 36, 487–505. doi: 10.1007/s10803-006-0089-y
- Ito, K. (2018). “Gradual development of focus and affect prosody comprehension: a proposal for a holistic approach,” in *Prosodic Development in First Language Acquisition*, eds P. Prieto and N. Esteve-Gibert (Amsterdam: John Benjamins).
- Ito, K., Jincho, N., Minai, U., Yamane, N., and Mazuka, R. (2012). Intonation facilitates contrast resolution: evidence from Japanese adults and 6-year olds. *J. Memory Lang.* 66, 265–284. doi: 10.1016/j.jml.2011.09.002
- Iverson, J. M., Capirci, O., and Caselli, M. C. (1994). From communication to language in two modalities. *Cogn. Dev.* 9, 23–43. doi: 10.1016/0885-2014(94)90018-3
- Iverson, J. M., and Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychol. Sci.* 16, 367–371. doi: 10.1111/j.0956-7976.2005.01542.x
- James, S. L. (1978). Effect of listener age and situation on the politeness of children's directives. *J. Psycholinguist. Res.* 7, 307–317. doi: 10.1007/BF01068112
- Jannedy, S., and Mendoza-Denton, N. (2005). “Structuring information through gesture and intonation,” in *Interdisciplinary Studies on Information Structure 3: Approaches and Findings in Oral, Written and Gestic Language*, ed. S. Dipper (Potsdam: Universitätsverlag Potsdam), 199–244.
- Kelly, S. D., Özyürek, A., and Maris, E. (2010). Two sides of the same coin: speech and gesture mutually interact to enhance comprehension. *Psychol. Sci.* 21, 260–267. doi: 10.1177/0956797609357327
- Kendon, A. (2002). Some uses of the head shake. *Gesture* 2, 147–182. doi: 10.1075/gest.2.2.03ken
- Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.
- Kim, J., Cvejic, E., and Davis, C. (2014). Tracking eyebrows and head gestures associated with spoken prosody. *Speech Commun.* 57, 317–330.
- Kim, S., Paulus, M., Sodian, B., and Proust, J. (2016). Young children's sensitivity to their own ignorance in informing others. *PLoS One* 11:e0152595. doi: 10.1371/journal.pone.0152595
- Kita, S. (ed.) (2003). *Pointing: Where Language, Culture, and Cognition Meet*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Kita, S. (2009). Cross-cultural variation of speech-accompanying gesture: a review. *Lang. Cogn. Process.* 24, 145–167. doi: 10.1080/01690960802586188
- Kitamura, C., Guellai, B., and Kim, J. (2014). Motherese by eye and ear: infants perceive visual prosody in point-line displays of talking heads. *PLoS One* 9:e111467. doi: 10.1371/journal.pone.0111467

- Kovács, A. M., Tausz, T., Téglás, E., Gergely, G., and Csibra, G. (2014). Pointing as epistemic request: 12-month-olds point to receive new information. *Infancy* 19, 543–557. doi: 10.1111/infa.12060
- Krahmer, E., and Swerts, M. (2005). How children and adults produce and perceive uncertainty in audiovisual speech. *Lang. Speech* 48, 29–53. doi: 10.1177/00238309050480010201
- Krahmer, E., and Swerts, M. (2007). The effects of visual beats on prosodic prominence: acoustic analyses, auditory perception and visual perception. *J. Mem. Lang.* 57, 396–414. doi: 10.1016/j.jml.2007.06.005
- Krahmer, E., and Swerts, M. (2009). Audiovisual prosody—introduction to the special issue. *Lang. Speech* 52, 129–133. doi: 10.1177/00238309090103164
- Krivokapić, J. (2014). Gestural coordination at prosodic boundaries and its role for prosodic structure and speech planning processes. *Philos. Trans. R. Soc. B Biol. Sci.* 369:20130397. doi: 10.1098/rstb.2013.0397
- Kuhl, P. K., Williams, K. A., and Meltzoff, A. N. (1991). Cross-modal speech perception in adults and infants using nonspeech auditory stimuli. *J. Exp. Psychol. Hum. Percept. Perform.* 17, 829–840. doi: 10.1037/0096-1523.17.3.829
- Kurumada, C., and Clark, E. V. (2017). Pragmatic inferences in context: learning to interpret contrastive prosody. *J. Child Lang.* 44, 850–880. doi: 10.1017/S0305000916000246
- Lee, T. H., and Law, A. (2001). “Epistemic modality and the acquisition of Cantonese final particles,” in *Issues in East Asian Language Acquisition*, ed. M. Nakayama (Tokyo: Kuroshio Publishers.), 67–128.
- Leonard, T., and Cummins, F. (2011). The temporal relation between beat gestures and speech. *Lang. Cogn. Process.* 26, 1457–1471. doi: 10.1080/01690965.2010.500218
- Liszkowski, U. (2014). Two sources of meaning in infant communication: preceding action contexts and act-accompanying characteristics. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369:20130294. doi: 10.1098/rstb.2013.0294
- Liszkowski, U., Carpenter, M., and Tomasello, M. (2008). Twelve-month-olds communicate helpfully and appropriately for knowledgeable and ignorant partners. *Cognition* 108, 732–739. doi: 10.1016/j.cognition.2008.06.013
- Llanes-Coromina, J., Vilà-Giménez, I., Borràs-Comes, J., Kushch, O., and Prieto, P. (2018). Beat gestures help preschoolers recall and comprehend discourse information. *J. Exp. Child Psychol.* 172, 168–188.
- Loehr, D. (2012). Temporal, structural, and pragmatic synchrony between intonation and gesture. *Lab. Phonol.* 3, 71–89.
- Long, C. (2010). Apology in Japanese gratitude situations: the negotiation of interlocutor role-relations. *J. Pragmat.* 42, 1060–1075. doi: 10.1016/j.pragma.2009.09.005
- Lyons, K. E., and Ghetti, S. (2011). The development of uncertainty monitoring in early childhood. *Child Dev.* 82, 1778–1787. doi: 10.1111/j.1467-8624.2011.01649.x
- Lyons, K. E., and Ghetti, S. (2013). I Don't Want to Pick! Introspection on Uncertainty Supports Early Strategic Behavior. *Child Dev.* 84, 726–736. doi: 10.1111/cdev.12004
- Ma, W., Golinkoff, R. M., Houston, D. M., and Hirsh-Pasek, K. (2011). Word learning in infant-and adult-directed speech. *Lang. Learn. Dev.* 7, 185–201. doi: 10.1080/15475441.2011.579839
- Marcos, H., and Bernicot, J. (1994). Addressee co-operation and request reformulation in young children. *J. Child Lang.* 21, 677–692. doi: 10.1017/S0305000900009508
- Mathew, M., Yuen, I., Shattuck-Hufnagel, S., Ren, A., and Demuth, K. (2014). The development of prosodic gestures during parent-child discourse interactions. *Paper Presented at the LabPhon Conference, Munich.*
- Matsui, T. (2014). “Children's understanding of linguistic expressions of certainty and evidentiality,” in *Pragmatic Development in First Language Acquisition*, ed. D. Matthews (Amsterdam: John Benjamins), 295–316.
- Matsui, T., Rakoczy, H., Miura, Y., and Tomasello, M. (2009). Understanding of speaker certainty and false-belief reasoning: a comparison of Japanese and German preschoolers. *Dev. Sci.* 4, 602–613. doi: 10.1111/j.1467-7687.2008.00812.x
- Matsui, T., Yamamoto, T., and McCagg, P. (2006). On the role of language in children's early understanding of others as epistemic beings. *Cogn. Dev.* 21, 158–173. doi: 10.1016/j.cogdev.2005.10.001
- McNeill, D. (1992). *Hand and Mind – What Gestures Reveal about Thought*. Chicago, IL: University of Chicago Press.
- McNeill, D. (2016). *Why We Gesture: The Surprising Role of Hand Movements in Communication*. Cambridge: Cambridge University Press.
- Moore, C., Bryant, D., and Furrow, D. (1989). Mental terms and the development of certainty. *Child Dev.* 60, 167–171. doi: 10.1111/j.1467-8624.1989.tb02706.x
- Moore, C., Harris, L., and Patriquin, M. (1993). Lexical and prosodic cues in the comprehension of relative certainty. *J. Child Lang.* 20, 153–167. doi: 10.1017/S030500090000917x
- Moore, C., Pure, K., and Furrow, D. (1990). Children's understanding of the modal expression of speaker certainty and uncertainty and its relation to the development of a representational Theory of Mind. *Child Dev.* 61, 722–730. doi: 10.1111/j.1467-8624.1990.tb02815.x
- Morgan, J. L., and Newport, E. L. (1981). The role of constituent structure in the induction of an artificial language. *J. Verbal Learning Verbal Behav.* 20, 67–85. doi: 10.1016/S0022-5371(81)90312-1
- Murillo, E., and Capilla, A. (2016). Properties of vocalization- and gesture-combinations in the transition to first words. *J. Child Lang.* 43, 890–913. doi: 10.1017/S0305000915000343
- Murillo, E., Ortega, C., Otones, A., Rujas, I., and Casla, M. (2018). Changes in the synchrony of multimodal communication in early language development. *J. Speech Lang. Hear. Res.* 61, 2235–2245. doi: 10.1044/2018_JSLHR-L-17-0402
- Nicoladis, E., Mayberry, R. I., and Genesee, F. (1999). Gesture and early bilingual development. *Dev. Psychol.* 35, 514–526. doi: 10.1037/0012-1649.35.2.514
- Nilsen, E. S., Graham, S. A., Smith, S., and Chambers, C. G. (2008). Preschoolers' sensitivity to referential ambiguity: evidence for a dissociation between implicit understanding and explicit behavior. *Dev. Sci.* 11, 556–562. doi: 10.1111/j.1467-7687.2008.00701.x
- Norton, A., Zipse, L., Marchina, S., and Schlaug, G. (2009). Melodic intonation therapy: shared insights on how it is done and why it might help. *Ann. N. Y. Acad. Sci.* 1169, 431–436.
- Notley, A., Zhou, P., Crain, S., and Thornton, R. (2009). Children's interpretation of focus expressions in English and Mandarin. *Lang. Acquis.* 16, 240–282. doi: 10.1080/10489220903266669
- Özçalışkan, Ş., and Goldin-Meadow, S. (2005). Gesture is at the cutting edge of early language development. *Cognition* 96, B101–B113. doi: 10.1016/j.cognition.2005.01.001
- Özçalışkan, Ş., and Goldin-Meadow, S. (2009). When gesture-speech combinations do and do not index linguistic change. *Lang. Cogn. Process.* 24:190. doi: 10.1080/016906801956911
- Papaeliou, C. F., and Trevarthen, C. (2006). Prelinguistic pitch patterns expressing ‘communication’ and ‘apprehension’. *J. Child Lang.* 33, 163–178. doi: 10.1017/S0305000905007300
- Papafragou, A., Li, P., Choi, Y., and Han, C. (2007). Evidentiality in language and cognition. *Cognition* 103, 253–299. doi: 10.1016/j.cognition.2006.04.001
- Paterson, K. B., Liversedge, S. P., Rowland, C., and Filik, R. (2003). Children's comprehension of sentences with focus particles. *Cognition* 89, 263–294. doi: 10.1016/S0010-0277(03)00126-4
- Paulus, M., Proust, J., and Sodian, B. (2013). Examining implicit metacognition in 3.5-year-old children: an eye-tracking and pupillometric study. *Front. Psychol.* 4:145. doi: 10.3389/fpsyg.2013.00145
- Perniss, P. (2018). Why we should study multimodal language. *Front. Psychol.* 9:1109. doi: 10.3389/fpsyg.2018.01109
- Phelps-Terasaki, D., and Phelps-Gunn, T. (2007). *TOPL-2 Test of Pragmatic Language: Examiner's Manual*, 2 Edn. Austin, TX: Pro-ed Inc.
- Pillow, B. H., and Anderson, K. L. (2006). Children's awareness of their own certainty and understanding of deduction and guessing. *Br. J. Dev. Psychol.* 24, 823–849. doi: 10.1348/026151005X83360
- Pinker, S. (1987). “The bootstrapping problem in language acquisition,” in *Mechanisms of Language Acquisition*, ed. B. MacWhinney (Hillsdale, NJ: Lawrence Erlbaum Associates), 399–441.
- Pizzuto, E., and Capobianco, M. (2005). The link and differences between deixis and symbols in children's early gestural-vocal system. *Gesture* 5, 179–199. doi: 10.1075/gest.5.1.13piz
- Prieto, P. (2015). Intonational meaning. *Wiley Interdiscip. Rev. Cogn. Sci.* 6, 371–381. doi: 10.1002/wcs.1352
- Prieto, P., Estrella, A., Thorson, J., and Vanrell, M. D. M. (2012). Is prosodic development correlated with grammatical and lexical development? Evidence

- from emerging intonation in Catalan and Spanish. *J. Child Lang.* 39, 221–257. doi: 10.1017/S030500091100002X
- Prieto, P., Pugliesi, C., Borràs-Comes, J., Arroyo, E., and Blat, J. (2015). Exploring the contribution of prosody and gesture to the perception of focus using an animated agent. *J. Phon.* 49, 41–54. doi: 10.1016/j.wocn.2014.10.005
- Rochet-Capellan, A., Laboisière, R., Galván, A., and Schwartz, J. L. (2008). The speech focus position effect on jaw-finger coordination in a pointing task. *J. Speech Lang. Hear. Res.* 51, 1507–1521. doi: 10.1044/1092-4388(2008/07-0173)
- Rohwer, M., Kloof, D., and Perner, J. (2012). Escape from metaignorance: how children develop an understanding of their own lack of knowledge. *Child Dev.* 83, 1869–1883. doi: 10.1111/j.1467-8624.2012.01830.x
- Roseano, P., González, M., Borràs-Comes, J., and Prieto, P. (2016). Communicating epistemic stance: how speech and gesture patterns reflect epistemicity and evidentiality. *Discourse Process.* 53, 135–174. doi: 10.1080/0163853X.2014.969137
- Rowe, M. L., and Goldin-Meadow, S. (2009). Early gesture selectively predicts later language learning. *Dev. Sci.* 12, 182–187. doi: 10.1111/j.1467-7687.2008.00764.x
- Ryckebusch, C., and Marcos, H. (2004). Speech acts, social context and parent-toddler play between the ages of 1;5 and 2;3. *J. Pragmat.* 36, 883–897. doi: 10.1016/j.pragma.2003.10.019
- Saint-Georges, C., Chetouani, M., Cassel, R., Apicella, F., Mahdhaoui, A., Muratori, F., et al. (2013). Motherese in interaction: at the cross-road of emotion and cognition? (a systematic review). *PLoS One* 8:e78103. doi: 10.1371/journal.pone.0078103
- Sakkalou, E., and Gattis, M. (2012). Infants infer intentions from prosody. *Cogn. Dev.* 27, 1–16. doi: 10.1111/bjdp.12135
- Sandiford, G. A., Mainess, K. J., and Daher, N. S. (2013). A pilot study on the efficacy of melodic based communication therapy for eliciting speech in nonverbal children with autism. *J. Autism Dev. Disord.* 43, 1298–1307. doi: 10.1007/s10803-012-1672-z
- Sealey, A. (1999). 'Don't be cheeky': requests, directives and being a child. *J. Sociolinguist.* 3, 24–40. doi: 10.1111/1467-9481.00061
- Shatz, M., Diesendruck, G., Martinez-Beck, I., and Akar, D. (2003). The influence of language and socioeconomic status on children's understanding of false belief. *Dev. Psychol.* 39, 717–729. doi: 10.1037/0012-1649.39.4.717
- Shatz, M., and Gelman, R. (1973). The development of communication skills: modifications in the speech of young children as a function of listener. *Monogr. Soc. Res. Child Dev.* 38, 1–38.
- Shochi, T., Erickson, D., Sekiyama, K., Rilliard, A., and Aubergé, V. (2009). "Japanese children's acquisition of prosodic politeness expressions," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, Stockholm.
- Smith, N. A., and Strader, H. L. (2014). Infant-directed visual prosody: Mothers' head movements and speech acoustics. *Interact. Stud.* 15, 38–54. doi: 10.1075/is.15.1.02smi
- Snow, D. P. (2017). Gesture and intonation are "sister systems" of infant communication: evidence from regression patterns of language development. *Lang. Sci.* 59, 180–191. doi: 10.1016/j.langsci.2016.10.005
- Speer, S. R., and Ito, K. (2009). Prosody in first language acquisition - Acquiring intonation as a tool to organize information in conversation. *Linguist. Lang. Compass* 3, 90–110. doi: 10.1111/j.1749-818X.2008.00103.x
- Striano, T., and Tomasello, M. (2001). "Infant development: physical and Social Cognition," in *International Encyclopedia of the Social and Behavioral Sciences*, eds N. J. Smelser and P. Baltes (Oxford: Pergamon), 7410–7414. doi: 10.1016/b0-08-043076-7/01569-2
- Tardif, T., Wellman, H. M., and Cheung, K. M. (2004). False belief understanding in Cantonese-speaking children. *J. Child Lang.* 31, 779–800. doi: 10.1017/S0305000904006580
- Thorson, J. (2018). "The role of prosody in early word learning: Behavioral evidence," in *The Development of Prosody in First Language Acquisition*, eds P. Prieto and N. Esteve-Gibert (Amsterdam: John Benjamins), 59–78.
- Thorson, J., and Morgan, G. (2014). "Directing Toddler Attention: Intonation and Information Structure," in *Paper Presented at the 38th Annual Boston University Conference on Language Development (BUCLD)*, Boston, MA.
- Tomasello, M. (1995). "Joint Attention as Social Cognition," in *Joint Attention: Its Origins and Role in Development*, eds C. Moore, and P. J. Dunham (Hillsdale, NJ: Lawrence Erlbaum Associates, Inc), 103–130.
- Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.
- Tomasello, M., Carpenter, M., and Liszkowski, U. (2007). A new look at infant pointing. *Child Dev.* 78, 705–722. doi: 10.1111/j.1467-8624.2007.01025.x
- Tree, A. R., and Manusov, V. (1998). Managing face concerns in criticism integrating nonverbal behaviors as a dimension of politeness in female friendship dyads. *Hum. Commun. Res.* 24, 564–583. doi: 10.1111/j.1468-2958.1998.tb00431.x
- Vilà-Giménez, I., Igualada, A., and Prieto, P. (2019). Observing storytellers who use rhythmic beat gestures improves children's narrative discourse performance. *Dev. Psychol.* 55, 250–262. doi: 10.1037/dev0000604
- Werker, J. F., Pons, F., Dietrich, C., Kajikawa, S., Fais, L., and Amano, S. (2007). Infant-directed speech supports phonetic category learning in English and Japanese. *Cognition* 103, 147–162. doi: 10.1016/j.cognition.2006.03.0
- Wiig, E. H., Semel, E., and Secord, W. A. (2013). *Clinical Evaluation of Language Fundamentals-Fifth Edition (CELF-5)*. Bloomington, MN: NCS Pearson.
- Winter, B., and Grawunder, S. (2012). The phonetic profile of Korean formal and informal speech registers. *J. Phon.* 40, 808–815. doi: 10.1016/j.wocn.2012.08.006

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Hübscher and Prieto. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.