

Gesture–speech combinations and early verbal abilities

New longitudinal data during the second year of age

Micaela Capobianco, Elena Antinoro Pizzuto and
Antonella Devescovi

University of Rome “La Sapienza” / Institute of Cognitive Sciences and
Technologies /

This study provides new longitudinal evidence on two major types of gesture–speech combination that play different roles in children’s early language. We analysed the spontaneous production of 10 Italian children observed monthly from 10–12 to 23–25 months of age. We evaluated the extent to which the developmental trends observed in children’s early gesture–word and word–word productions can predict subsequent verbal abilities. The results indicate that “complementary” and “supplementary” gesture–speech combinations predict subsequent language development in a different manner: While the onset of “supplementary” combinations predicts the onset of two-word combinations, the use of “complementary” combinations at 12 and 18 months predicts the vocabulary and the ability to produce more words utterances at 2 years of age. Moreover, the results suggest that both “complementary” and “supplementary” crossmodal combinations are good predictive indexes of early verbal skills during the second year of age.

Keywords: pointing, deictic gestures, supplementary and complementary combinations, early verbal skills, prevention

1. Introduction

This study is framed in relation to recent and older research on the multimodal characteristics of early communication and language development, with reference to the role of gesture–speech in the development of complex verbal utterances. We focus on the transition period from the onset of a one-word utterance, around 12 months of age, to the appearance of two and multiword combinations, at around 1–24 months (Butcher & Goldin-Meadow, 2000; Camaioni, Aureli,

Bellagamba & Presaghi, 2004; Capirci, Caselli, Iverson, Pizzuto & Volterra, 2002; Capirci, Contaldo & Volterra, 2003; Capirci, Contaldo, Caselli & Volterra, 2005; Capobianco, 2015; Colletta & Guidetti, 2012; Goldin-Meadow & Butcher, 2003; Guidetti, 2002; Iverson, Capirci, Volterra & Goldin-Meadow, 2008; McEachern & Haynes, 2004; Pizzuto, Capobianco & Devescovi, 2003; Pizzuto, Capobianco & Devescovi, 2005; Pizzuto, Capobianco & Devescovi, 2009; Stefanini, Bello, Caselli, Iverson & Volterra, 2009; Volterra, Caselli, Capirci & Pizzuto, 2005). Drawing on a large corpus of longitudinal data – with numerous observations over time – the primary objective of the present study is to contribute to a clearer understanding of the specific roles that different types of gesture–speech combinations play in children’s early language.

As highlighted in the literature, so-called *crossmodal or transmodal* combinations between gesture and speech represent a remarkable portion of children’s utterances in the first 2 years of life; these comprise words accompanied by deictic gestures (DGs), particularly POINTING; the other deictic gestures are “REQUEST” and “SHOW” (Capirci et al., 1996; Pizzuto & Capobianco, 2009). DGs all fulfil the basic function of “drawing the interlocutor’s attention towards something in the environment,” as has been extensively described in the literature (Capirci et al., 1996). The three types of DGs can be defined as follows: POINT: Extending the index finger – or in some cases, the full hand – toward an object, location, or person; SHOW: Holding up an object in the adult’s line of sight; and REQUEST: Extending the arm toward an object, location, or person, sometimes with a repeated opening and closing of the hand.

Although authors agree that children produce a considerable quantity of these combinations before they use two words together, there still are discrepancies with respect to the role that different authors ascribe to combinations with *different information* (see below). Such discrepancies may often be explained based on theoretical and methodological differences in the classification and interpretation of children’s early gestural–vocal and vocal-only productions.

Goldin-Meadow & Morford (1990) were the first to propose what we view as an important distinction between “complementary” and “supplementary” gesture–word combinations essentially involving POINTING gestures and words. Specifically, Goldin-Meadow & Morford (1990) defined gesture–word utterances where the two modalities refer to the “same semantic element,” and in these authors’ view, this redundantly enhances the meaning expressed by the word as “complementary” (for example, POINTING at a glass and saying “glass”). In subsequent work, using a different terminology and partially different datasets, Butcher and Goldin-Meadow (2000) and Goldin-Meadow & Butcher (2003) delineated that in “supplementary” utterances, DGs and speech instead refer to two distinct semantic elements, as in POINTING at a glass and saying “outside”; here,

POINTING is equivalent to the noun “glass” at both the semantic and functional levels, while the word “outside” represents a predicative act. Alternatively, “supplementary” utterances may involve combinations where the gesture singles out a referent that is different from that of the word (e.g., POINTING at a glass and saying “mummy”). The authors underlined that with the second type of combinations (“where the gesture refers to a referent that is different from that of the word”), the child produces two different elements in the same utterance.

In longitudinal research including six American children, Butcher and Goldin-Meadow (2000) and Goldin-Meadow and Butcher (2003) found that the onset of combinations of pointing gestures and speech having the same meaning (“complementary”) marks the beginning of the integration of the two modalities in child development. The combinations where pointing gestures and speech express different information (*two distinct semantic elements*) are significantly related to the onset of two-word combinations. The children who first combine pointing gestures and speech to express different pieces of information also seem to be first to produce early word combinations. The authors observed that combinations of pointing gestures and speech with different information (“supplementary”) seem to play an important role in the development of two-word combinations, and consequently, they have predictive value for the development of word utterances.

Similar findings were reported by McEachern & Haynes (2004) in a longitudinal study on a sample of 10 American children observed from the ages of 16 to 21 months. These researchers identified significant development patterns only in “supplementary” combinations; in contrast, “complementary” combinations tended to remain stable over the period of observation. The authors observed an important role of “supplementary” combinations in the acquisition of the early verbal skills. A different developmental model, proposed by Capirci et al. (1996, 2003, 2005), Pizzuto et al. (2005), and Pizzuto & Capobianco (2005), stresses the importance of using comparable criteria in the analysis of gesture-word and word-word combinations. This model extends the distinction between “complementary” and “supplementary” combinations to two-word productions to study differences and links in the information organization of distinct utterances. The authors also pointed out that the deictic elements (gestural or vocal) in “complementary” combinations are not redundant; rather, they specify or disambiguate the intended referent. For example, the child POINTS at a doll and says “baby” or POINTS at a dish and says “this.” In this respect, “complementary” combinations fulfil a primary “naming” function (Capobianco, 2015; Pizzuto, 2002; Pizzuto et al., 2009; Pizzuto et al., 2005; Volterra et al., 2005).

In this model, pointing gestures are not considered to replace nouns or verbs; instead, they are like vocal deictics (e.g., “this,” “that,” “here”). For this reason, the combination is not interpreted as “redundant”; rather, it is aimed at

“disambiguating/singling out” the referent. In “supplementary” combinations, the two elements may refer to the same or to two different referents. In both cases, each element of the combination adds information to the other. Typical examples of “supplementary” combinations are two-word utterances, such as “much+balls” or “open+book,” but crossmodal combinations can also emerge, such as POINTING (at a book) and saying the word “nice.” As noted in Pizzuto et al. (2005), some of the gesture–word utterances that Butcher & Goldin-Meadow (2000) and Goldin-Meadow & Butcher (2003) coded as “supplementary” can be classified as “complementary” using more conservative criteria. For example, Butcher & Goldin-Meadow (2000) and Goldin-Meadow and Butcher (2003) classified utterances like POINTING at a cow and saying “moo” as “supplementary,” thus assigning a “predicative” meaning to onomatopoeic words. In the encoding system adopted by Pizzuto et al. (2005) and Pizzuto & Capobianco (2008), the same utterance would be classified as “complementary”: The onomatopoeic word “moo” is assigned a “nominal” meaning (simplest form of “cow”) rather than the action meaning of “moo = to bellow.” If the onomatopoeic word is regarded as a “verb” rather than a “nome,” the crossmodal combination could be encoded as “supplementary.”

Drawing on longitudinal evidence on Italian children, Pizzuto et al. (2005) and Pizzuto & Capobianco (2005) observed six children from 10 and 24 months of age using an encoding system adapted from Capirci et al. (1996). The authors showed that pointing–word crossmodal combinations are primarily of the “complementary” type, whereas word–word combinations are mostly “supplementary.” Thus, children tend to “name” in crossmodal combinations – for example, when they POINT and say “cup” – and to “predicate” in vocal productions where they use two names in the word–word combination (Pizzuto et al., 2009). This pattern differs from that found in adults and older children, who use a variety of gestures in combination with speech to add information. Adults do point at things and name them, and they do sometimes point at something and make a comment about it, but they also do other things with gestures that these extremely young children do not do (Kendon, 2016; McNeill, 2000; Mayberry & Nicoladis, 2000).

In a more recent study, Fasolo & D’Odorico (2012) found that complementary gesture-plus-word combination use at 18 months were related with lexical skill and mean length of utterance (MLU) – but not with utterance complexity – at 24 months; supplementary gesture-plus-word combination use at 18 months was related with utterance complexity but not with lexical skill or MLU. Ultimately, the limited number of children examined and/or of the observations over time per child carried out have not allowed the application of appropriate statistical analyses to evaluate the significance of different developmental patterns and their predictive value in the development of different early verbal skills in a more detailed way.

The present study aims to explore and assess the specific contribution of “complementary” and “supplementary” combinations of gestures and speech in predicting the onset of early two-word utterances and the development of more general verbal skills assessed at the end of the second year of life. To accomplish this, we use a relatively large longitudinal corpus stemming from observations of 10 Italian children observed monthly during the second year of age.

2. Method

2.1 Participants

The sample included 10 typically developing Italian children (six boys and four girls) who participated in a larger study on early language development in full-term compared to preterm children (Capobianco, 2006; Capobianco, Pizzuto & Devescovi, 2007).

All children showed typical development, evaluated through spontaneous observations and language standardized tools (Capobianco, 2006). The socioeconomic levels of the children’s families were calculated using the Hollingshead Index (Hollingshead, 1975). Table 1 provides basic information on each child in terms of gender, weeks and weight at birth, family socioeconomic level, the time period during which the child was observed, and the number of observations made.

Table 1. Characteristics and videotaping observations of each child

Participants	Gender	Gestational Age (weeks)	Weight (grams)	Hollingshead Index*	Period of observations (months)	Number of observations
1	M	38.3	3040	3	12–25	11
2	M	41	3550	5	12–24	9
3	M	40	3300	5	10–23	13
4	M	39.5	3200	3	10–25	15
5	M	40.3	3500	4	12–24	11
6	M	39.6	3700	2	12–23	10
7	F	39.1	3400	2	12–25	9
8	F	40	3260	4	10–25	14
9	F	39.6	3530	2	12–25	14
10	F	40	3720	4	12–25	14
Mean		40	3420	3.6		12

* Score: 1–2 = low; 3 = middle; 4–5 = high

As can be seen in Table 1, five children came from families with a high socioeconomic level, two from middle-class families, and three from lower-class families. Thus, the children in this study had extremely different socioeconomic circumstances. However, the assessment of the socioeconomic levels of the participants in this study was performed for descriptive purposes. It was not possible to explore any link between the socioeconomic variable and raw linguistic skills given the need for a larger sample for this type of analysis.

2.2 Data collection procedures

Following a procedure that was successfully employed in previous studies (e.g., Capirci et al., 1996; Pizzuto et al., 2005), the children were observed at home during videotaped sessions lasting 45 minutes on average while they spontaneously interacted with their mothers in three different contexts (each lasting 15 minutes), as follows: playing with new examples of unfamiliar objects (a set of toys provided by the experimenter), playing with familiar objects, and having a meal or snack. Observation sessions were scheduled monthly, but as often happens in longitudinal studies of this sort, this schedule could not always be followed, and some sessions were skipped. The data collected included 10–13 monthly records of each child in the developmental period between 10–12 and 23–25 months [Table 1].

2.3 Coding and language assessment

All children's spontaneous gestural and vocal productions were transcribed and analyzed using Child Language Data Exchange System (CHILDES), standard procedures (Mac Whinney, 1997). CHILDES is a system used to share and analyze conversational interactions. Three major tools for child language research are incorporated into the system, as follows: – the CHILDES database of transcripts, – the CHAT system for transcribing and coding data, and – the CLAN programs for analyzing children's verbal and nonverbal productions. The communicative statuses of the children's utterances made up of gestural and vocal elements were assessed using the criterion proposed by Thal & Tobias (1992) and Capirci et al. (1996), where “gestures and words are considered communicative if they were accompanied by eye contact with the child's interlocutor.” Both two-element utterances and gesture–word combinations were classified into the “complementary” and “supplementary” type, as shown in Table 2.

For the present study, to allow appropriate comparisons with studies involving English-speaking children, the analysis of combinations was limited to those including a DG and those that represented the most remarkable part of children's crossmodal productions. Two verbal measures assessed at the 2-year threshold were

considered, as follows: (1) size of vocabulary (types of words; number of distinct word types) and (2) sentence complexity assessed via frequency (tokens) of utterances of two or more words. For the present work, gestural and vocal elements and utterances observed in the spontaneous productions were identified, interpreted, and encoded according to the coding scheme used by Pizzuto et al. (2005) and Pizzuto & Capobianco (2005), which was adapted from Capirci et al. (1996). Table 2 summarizes our coding scheme with examples drawn from the children's productions.

2.4 Coding of specific gestures and word types

Following suggestions from Capirci et al. (1996) and Iverson, Capirci & Caselli (1994), three types of DGs were transcribed and analyzed. These were produced by the children throughout the period under examination and their function was to direct adults' attention towards objects, people, places, or events in the interaction context. These were as follows: **POINTING**, with the index finger or the whole open hand directed to objects, places, or events in the surrounding environment and looking alternatively at the adult and at the object; **REQUEST**, involving opening and closing the hand, most frequently shifting the eyes from the adult to the requested object; and **SHOWING**, where an object is held in the adult's line of sight as a way of drawing attention to the object. In some cases, a gesture produced without holding any object in the hand was also classified as **SHOWING**. An example of this is a child showing his/her hand, for instance, because it is "dirty," with the clear intention of showing it to his/her mother. In the studies by Pizzuto et al. (2005) and Pizzuto & Capobianco's (2005), real words were distinguished from other vocal elements, such as vocalizations, which have different characteristics and functions although they have a communicative function in children's production. For this study, only representative (rw) and deictic (dw) words were taken into account, and vocalizations were excluded. Representative words (rw) include all content words that, in adult language, are assigned to the class of proper and common nouns ("mummy," "food"), verbs ("broken," "look"), adjectives ("good," "nice," "beautiful"), positive and negative expression ("yes," "no"), adverbs ("because," "then"), interjections ("good"), greetings ("bye-bye"), and prepositions ("above," "below"). The rw category also includes onomatopoeic forms (ow; e.g., "bau-bau" for dog, "miao" for cat, "brum-brum" for car, "amme," for eating; see Examples in a.2 in Table 2). The dw include all demonstrative ("this," "that") and locative ("here," "there") forms, personal pronouns ("I," "you," "we"), and possessive pronouns ("mine," "yours").

Table 2. Coding scheme and notational conventions with illustrative examples of the children's gestural and vocal production and verbal measures valued to explore their predictive role*

A. Types of gestures and words in the children utterances	
a.1 gesture	
• Deictic (DG):	POINT, SHOW, REQUEST
a.2 words	
• Representational (rw): (conventional and onomatopeic)	mamma, pappa, piccolo, guarda, sì, no <mommy, food, little, look, yes, no> cra + cra = rana, bau = cane, miao = gatto
• Deictic (dw):	qua, là, questo, quello, io, tu, mio, tuo <here, that, this, there, I, you, mine, yours>
B. Two element utterance: Modality, Components, information conveyed	
b.1 Crossmodal (DG-w):	
Complementary (&)	
b.1.1	
DG & rw	POINT (cat) & "miao" <cat: gatto> REQUEST (doll) & "doll" <doll: bambola>
b.1.2	
DG & dw	POINT (ball) & "ball" <ball: palla>
Supplementary (+)	
b.1.3	
DG + rw	POINT (chocolates) + "much" <chocolates: cioccolatini; much: tanti>
b.1.4	
GD + dw	SHOW (ball) + "I" <ball: palla; I: io>
b.2 Unimodal vocal (w-w): Complementary (&) and Supplementary (+)	
b.2.1	
dw & rw	this & food; here & baby <questa & pappa, qua & bimbo >
b.2.2	
rw + rw	much + bolls, book + open <tante + bolle, apri + libro>
b.2.3	
dw + rw	I + mamy, that + beautiful <io + mamma, questo + bello>
b.2.4	
dw + dw	I + that, that + mine <io + quello, mio + questo>
C. Verbal measure: (1) onset of two words, (2) verbal complexity, (3) repertoire of words around two years oldc.1	
onset of two words:	onset age of two words utterance in each child's development
c.2	
verbal complexity:	token of two and more words productions in the last three months in each child
c.3	
repertoire of words:	type of words produced in the last three months in each child

* Adapted from Capirci et al. (1996) and Piavto et al. (2005), in Capobianco (2006). The difference categories of gestures, words and utterance are described in the text. Abbreviations used throughout the text for each major category are given in parentheses (i.e. DEICTIC GESTURES = DG). Gestures are represented by English labels, in CAPITALS. Words are given in lower case letters in English and Italian translations.

2.5 The information relationship in crossmodal and two-words utterances

For this work, only *crossmodal* combinations made up of one DG and one word (rw or dw) and two-word combinations were considered, distinguishing between “complementary” (&) and “supplementary” combinations. “Complementary” combinations, signaled by an “&” character, included all gesture–word or word–word combinations in which both elements referred to the same referent in the context and those in which one of the elements in the combination disambiguated/singled out the object, person, event, or place referred to. In “complementary” combinations, a DG may be produced in combination with a rw (e.g., the child POINTS at the ball and says “ball”: GD&rw), an onomatopoeic word (the child POINTS at the cat and says “miao”: DG&ow), or a deictic word (the child POINTS at the plate and says “this”: DG&dw). The same function is also expressed in two-word combinations made up of one dw and one rw, for example, “this & ball” or “here & plate” (dw&rw). In contrast to the views expressed by Butcher & Goldin-Meadow (2000), Goldin-Meadow & Butcher (2003), and McEachern & Haynes (2004), who considered complementary combinations “redundant,” but in line with remarks made by Pizzuto et al. (2005), we do not view the DG, which is the pointing, to be assimilated to the word to which it refers (“ball” in the example above). Rather, in this work, we suggest that it retains its meaning of “deixis”/“reference” (“this” is the verbal form) relative to the context to which it refers. We think that in these combinations, the DG is not assimilated to the meaning of the word it is combined with. Children use the “complementary” crossmodal combination to “name” or “specify” an element of the interaction context.

“Supplementary” utterances, indicated with a “+” character, referred to either a single referent (most frequently) or to two referents. In all cases, each combined element added information to the other one (see Examples b.1.3 and b.1.4 in Table 2). Examples of “supplementary” utterances include combinations of one DG and one word (rw or dw), for example, POINTING at a chocolate and saying “a lot” (DG+rw) or two words (representative or deictic), as in the combination “baby+nice” (rw+rw).

Another important methodological aspect is the interpretation of gesture and speech combinations produced by children as “complementary” versus “supplementary.” For example, Butcher & Goldin-Meadow (2000) and Goldin-Meadow & Butcher (2003) classified utterances like POINTING at a cow and saying “moo” as “supplementary,” thereby assigning a “predicative” meaning to the onomatopoeic word. In the encoding system adopted by Pizzuto et al. (2005) and Pizzuto & Capobianco (2008), in contrast, such utterances are classified as “complementary”: A nominal meaning (“cow”) rather than a verbal one is assigned to the onomatopoeic word “moo.”

2.6 Data analysis

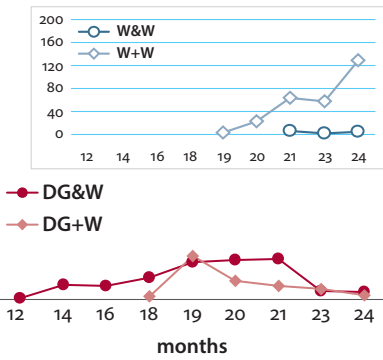
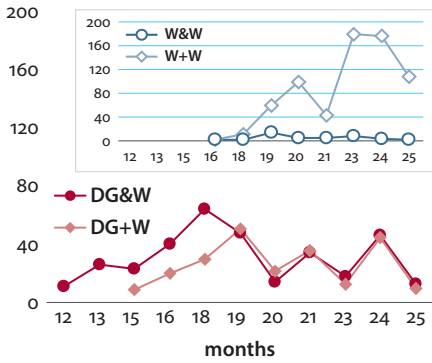
The frequency of combinations of gestures and words and two or more words produced by each child during development was examined via an individual profile description to identify the common development trends across children. A *polynomial regression* analysis, appropriate in studies with measures that are repeated over time, was also used to analyze the mean trend of each combination type, R square percentage, and significance of individual development patterns observed over time that were calculated for the 10 children. To assess the predictive value of the use of “complementary” and “supplementary” combinations of gestures and words with respect to subsequent verbal development, the *Pearson* product–moment correlation test was used ($p < .05$ and $< .01$).

The frequencies of the two combination types observed at 12, 15, and 18 months of age were correlated with the onset of two-word utterances and two measures of verbal development assessed over each child’s last three observation sessions (at 22–25 months). These two measures were as follows: (1) verbal complexity: mean frequency (tokens) of combinations of two or more words and (2) vocabulary size: repertoire (types) of words. Using the same correlation test, the study observed the predictive value of the onset and use of “complementary” and “supplementary” crossmodal combinations with the onset of two-word combinations and verbal abilities during the second year of age. Finally, if necessary, the specific weight (standard β coefficient) of the combinations of gestures and speech correlating with the same verbal measure was calculated through a *multiple regression* analysis (standard model) to single out the best predictor for that verbal skill assessed just before 2 years of age.

3. Results and discussion

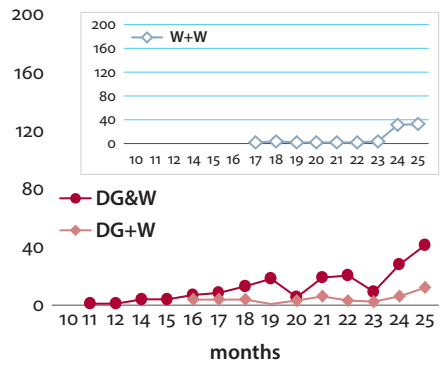
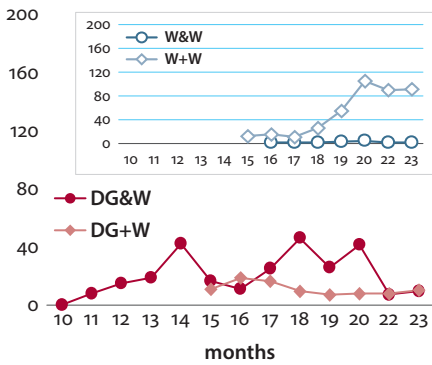
3.1 Complementary and supplementary crossmodal and two-word utterances: Individual profiles and estimated development across children

For the production of each child, Figures 1–10 show the frequency of crossmodal combinations of gestures and speech with “complementary” (DG&w) and “supplementary” (DG+w) relationships in relation to the corresponding vocal combinations with “complementary” (w&w) and “supplementary” (w+w) relationships.



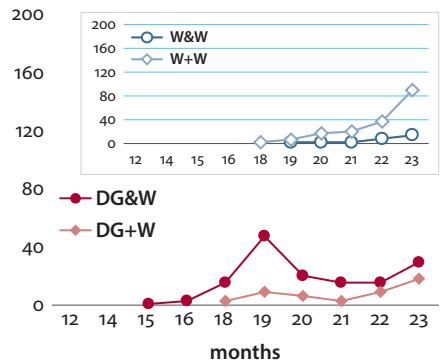
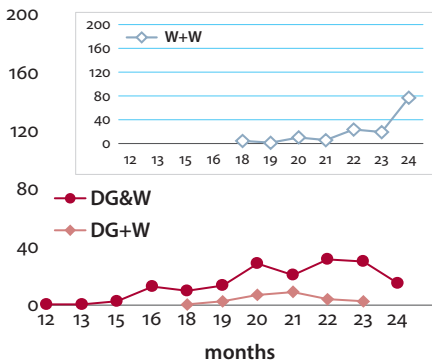
1.

2.



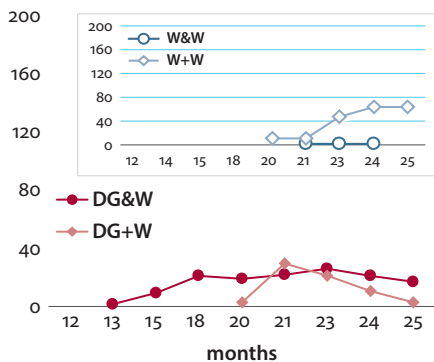
3.

4.

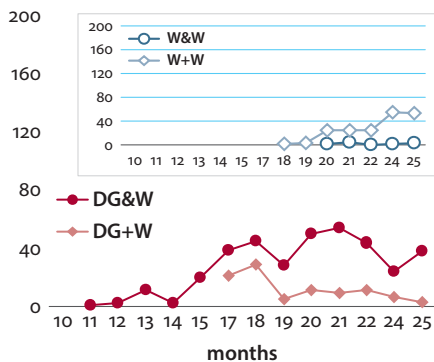


5.

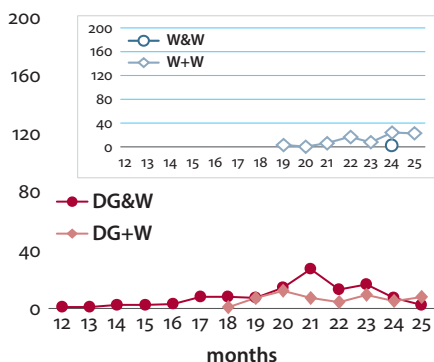
6.



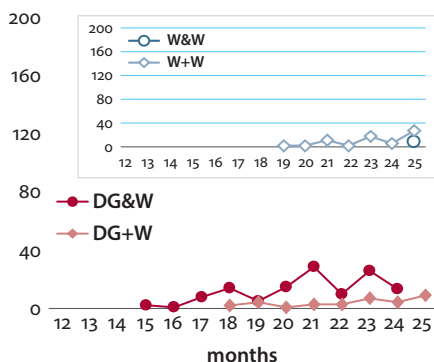
7.



8.



9.



10.

Figures 1–10. Token of “complementary” *versus* “supplementary” cross-modal and vocal combinations in each child’s production at different age

Despite important individual differences observed in the frequency of use of the two combinations types, all children showed similar developmental patterns and stages. The combinations of gestures and speech with a “complementary” relationship (DG&w) appeared first, around 12.5 months of age on average, with a range of 10–15 months; these prevailed over “supplementary” combinations (DG+w) in almost all children’s productions. “Supplementary” combinations appeared later in development, at around 17.5 months of age on average, with a wider range of 15–20 months; such combinations generally showed lower frequencies compared to “complementary” combinations in all observed children (the two types of combinations only showed overlapping values in subject 1, from 19 months of age onward).

In contrast to what happens for crossmodal combinations, in two-word combinations, “supplementary” relationships prevailed, in accordance with the findings of Capirci et al. (1996), Pizzuto et al. (2005), and Pizzuto & Capobianco (2005). Two-word combinations with a “supplementary” relationship (w+w)

appeared first, around 18 months of age on average (with the exception of subject 1, for whom they appeared at the same age), and prevailed remarkably over two-word combinations with a “complementary” relationship (w&w) for the entire observed period. Figure 2 shows the analysis of the estimated trends carried out in the 10-child sample.

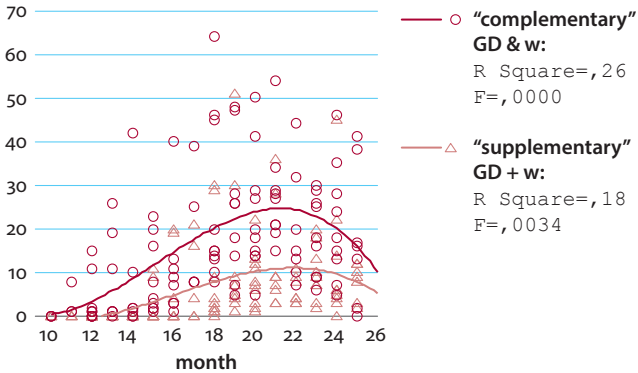


Figure 2. Trend development of “complementary” versus “supplementary” cross-modal combinations estimated across ten children at the different age

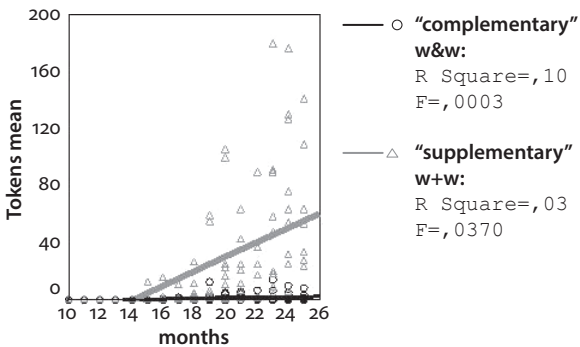


Figure 3. Trend development of “complementary” versus “supplementary” vocal combinations estimated across ten children at the different age

The development of the two types of crossmodal combinations presented similar nonlinear developmental trends. “Complementary” and “supplementary” cross-modal combinations were characterized by a curve that presented a maximum production period and a subsequent decrease, with frequencies around similar values to those recorded in the early development period. “Complementary” crossmodal combinations presented the highest frequencies at 18–22 months of age ($F < .01$), while “supplementary” combinations did so at 20–23 months of age ($F < .01$). The percentage of variance in “complementary” combinations explained

by the development trend (27%) was higher than the corresponding “supplementary” percentage (19%). This suggests less variability among the 10 children for the “complementary” crossmodal combination trend (POINTING&w), as also shown by the data in individual profiles (Figure 1–10).

The trend of two-word combinations, in contrast, was characterized by an incremental straight line, with a much steeper rise for “supplementary” combinations (w+w). Indeed, “complementary” two-word combinations (w&w) showed a less significant increase compared to “supplementary” combinations ($F < .05$ vs. $F < .01$), as well as a much lower variance percentage (3% for “complementary” vs. 10% for “supplementary”), suggesting the presence of greater individual difference in the children’s use of “supplementary” vocal productions. This variance was linked to the variability of the frequency of each child’s use of the combinations in the observed period.

The results of this study suggested that children’s crossmodal productions were somewhat biased toward conveying complementary information, while the productions were biased towards conveying supplementary information: There was an asymmetry between complementary and supplementary combinations during the transition from one to two words: Complementarity (&) prevailed in crossmodal combinations, while supplementarity (+) was more prevalent in two-word combinations (Pizzuto & Capobianco, 2005; Pizzuto et al., 2005; Pizzuto et al., 2009).

During early verbal acquisition, children mainly use one gesture (POINTING) and one word to disambiguate/single out the referent in the context (for example, POINTING at the doll and saying “doll”); in the corresponding vocal combination, they tend to add information (for example, “baby+nice”). Furthermore, the developmental pattern of “complementary” combinations estimated across 10 children in this study turned out to be significant for age, in contrast to McEachern & Haynes’s (2004) findings concerning 10 American children observed over a shorter period of time at 16–21 months of age. Similarly, as a confirmation of the results attained by Goldin-Meadow & Butcher (2003) and McEachern & Haynes (2004), all children in this study started combining two words only once they were also able to produce “supplementary” crossmodal combinations (DG+w).

3.2 The predictive value of crossmodal (complementary and supplementary) combinations taken and observed in development

In light of the development patterns resulting from individual profiles and the assessment of “complementary” and “supplementary” crossmodal combinations trends in the 10 children, the predictive values of the two types of crossmodal combinations observed at 12, 15, and 18 months of age were calculated for the different verbal skills assessed in the same children just before 2 years of age.

Table 3 shows the results for the following correlations between the tokens of “complementary” and “supplementary” combinations observed at 12, 15, and 18 months of age and the verbal measures examined around 2 years of age: a) the frequency (token) of combinations of two or more words (as verbal complexity measure) and b) repertoire (type) of words (as vocabulary size measure), calculated over the last 3 months’ observation sessions for each child and onset age of two-word utterances.

Table 3. Correlations between token of crossmodal combinations (complementary and supplementary) at 12, 15, 18 months and different verbal abilities assessed in last three months in the sample of ten children

Predictor	Verbal measures								
	two and more words token (last three observations mean)			words type (last three observations mean)			two words onset		
	12 m.	15 m.	18 m.	12 m.	15 m.	18 m.	12 m.	15 m.	18 m.
Crossmodal									
Token of combinations									
Token of “complementary”	.660*	ns	.832**	ns	ns	.845**	ns	ns	ns
Token of “supplementary”	–	.773*	.774*	–	ns	ns	–	ns	ns

Positive correlations were observed between use of DG-word “complementary” combinations at the ages of 12 and 18 months and verbal complexity around 2 years of age ($r = .660$; $r = .832$; $p < .05$). The token of “complementary” combinations observed at 18 months was also correlated with the type of words around 2 years of age. The token of DG-word “supplementary” combinations (+) observed at the age of 15 and 18 months was only positively correlated only with verbal complexity at 2 years of age ($r = .773$; $r = .774$; $p < .05$). No token of crossmodal “supplementary” combinations exhibited correlation with types of words around 2 years of age. Unlike the earlier development period, both tokens of “complementary” and “supplementary” combinations at 18 months correlated positively with verbal complexity at 2 year of age ($r = .832$, $p < .01$; $r = .774$, $p < .05$, respectively), while only the token of “complementary” combinations at 18 months correlated to types of words ($r = .845$, $p < .01$). Although both combinations had a predictive value for later verbal development at 18 months of age, when the specific contribution of the two types of combinations (“complementary” and “supplementary”) was considered for prediction purposes, a significantly higher relative weight was found for “complementary” combinations ($\beta = .09$, $p < .01$) than “supplementary” ones ($\beta = -.13$, $p = ns$). Thus, at 18 months of age, the token of “complementary” combinations in children’s productions seemed to better predict verbal complexity around 2 years of age. Finally, no correlation was observed between the token of

crossmodal combinations (“complementary” and “supplementary”), as observed during development and the age of onset of two-word productions. The token of different types of gesture and word combinations before 2 years of age does not seem to predict the onset of the early word combinations.

The correlation data observed at 18 months of age confirm the data resulting from individual profiles and from the estimated trends (Figures 1–10, 2, and 3), where the maximum production of “complementary” and “supplementary” crossmodal combinations was observed in all children in the study along with the onset of early two-word productions. However, further statistical analysis indicated a more significant predictive value for “complementary” combinations; at 18 months of age, their use is correlated not only with verbal complexity but also with the type of words produced over last 3 months of observation sessions for each child. The same results on the predictive role of the use of “complementary” combinations at 18 months were also observed in Fasolo and D’Odorico (2012). However, we also observed an early predictive value of “complementary” combinations used at 12 months for verbal complexity around 2 years.

3.3 Predictive value of the onset of crossmodal (complementary and supplementary) combinations

The patterns resulting from individual profiles and from the trends estimated across the children were confirmed by a positive and highly significant correlation ($r = .904$, $p < .01$) between the age of onset of “supplementary” crossmodal combinations (DG+w) and the age of onset of the early two-word combinations (2w).

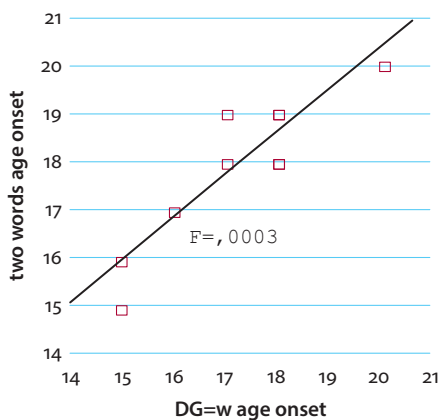


Figure 4. Correlation between age onset of supplementary crossmodal and two words utterances

Indeed, the age of onset related to the early word combinations was neither predicted by the frequency of “complementary” and “supplementary” combinations at any time of development (as already observed in Table 3) nor by the age of onset of “complementary” crossmodal combinations. This means that the children in the sample who first started using gesture and word combinations expressing a predicative relationship during development were also first to combine two words (2w). For the 10 children in this study, the age onset of “supplementary” crossmodal combinations (+) significantly predicted the children’s ability to produce the early 2w combinations (as Butcher & Goldin-Meadow 2000 and Goldin-Meadow & Butcher 2003).

The correlation data suggest that children started to produce two words together only if they were also able to express a “supplementary” relationship at the crossmodal level. This highlights the specific role played by the onset of “supplementary” combinations in predicting the onset of early two-word combinations; however, it is not sufficient to ascribe a unique and general role to this type of combination in predicting verbal skills in the transition from 1 to 2 years of age. During development, the two combination types play specific roles in predicting verbal skills at the end of the second year of age. Based the data from 10 children in this study, when considered in terms of their frequency of use (token) over time, complementary combinations (&) are good predictors of the frequency (token) of word combinations and word repertoire (type) over the last 3 months’ observation sessions for each child. Instead, the onset or presence of “supplementary” crossmodal combinations in children’s productions a good indicator of the age of onset of early two-word utterances.

A more accurate analysis of the predictive value of both combinations (“complementary” and “supplementary”) against their token and/or age of onset in predicting various verbal measures that are examined around 2 years of age enabled us to precisely assess both combinations’ specific roles and to stress the important contribution of “complementary” combinations at different developmental stages (against the token at 12 months, and especially, at 18 months of age). This allowed us to predict the children’s later verbal skills for both verbal complexity and word size measured at 2 years of age. These results help to explain the predictive role of crossmodal combinations of a DG and a word in the explosion of vocabulary (as for repertory words) and grammar skills at 2 years, in relation to the age of onset and frequency of use observed before 2 years of age, at 12, 15, and 18 months (Capobianco, 2015; Murillo & Belincho, 2012).

4. Conclusions

The results of this study, which included a larger number of children and was based on monthly observation during the second year of age, confirm the findings of previous studies (Butcher & Goldin-Meadow 2000, Colletta & Guidetti, 2012; Fasolo & D'Odorico, 2012; Goldin-Meadow & Butcher, 2003; Mc Eachern & Haynes, 2004; Murillo & Belincho, 2012; Pizzuto et al., 2009); moreover, the results provide new data on the development of crossmodal combinations (DGs and words) providing different types information (complementary and supplementary) in the transition to early word development and two-word combinations. The new elements of this work concern the survey results and the methodology used. In fact, the children's data are linked with the specific methodology used to elicit spontaneous production and engage in collection, and analysis in this study, observed monthly via interaction with their mothers. The comparison between individual profiles and developmental trends allowed us to study aspects of individual variability and evolutionary sequences common to all children (Capobianco, 2006; Capobianco, 2015; Tambelli, Cimino, Cerniglia & Ballarotto, 2015).

The results for the "complementary" and "supplementary" crossmodal combinations in the first 2 years were linked to a "conservative" methodological choice. The DG retained its meaning of "reference" ("this" in verbal form), and onomatopoeic words were encoded as content words and not as actions ("meow" = "cat" and not the action of meowing). Thus, combinations of a DG and a word – in which the gesture and word refer to the same referent (for example, POINTING to the cat and saying "cat") – does not express a "redundant" meaning; instead, it has the function of "naming" and/or specifying the referent (the cat) in the context of interaction. In addition, those combinations consist of a DG and a "predicate" (verb and adjective) or pronoun ("I", "my") referring to the same element – for example, pointing to the cat and say "beautiful" or "mine," were coded as "supplementary." The methodology we used was based on a corpus of spontaneous productions with monthly observations, which has not been used in other works; this allowed us to study the statistical value of the developmental trends of both types of crossmodal combinations and their predictive value in more stages over time (at 12, 15, and 18 months) for the verbal threshold of 2 years (22–24 months). The predictive value of crossmodal combinations for verbal abilities around 2 years of age were studied longitudinally, considering two measures – frequency of use (token) between 12 and 18 months and age of onset (Goldin-Meadow and Butcher, 2003).

We found that the frequency of use of complementary combinations at 12 months predicted verbal complexity around 2 years of age, and this is of particular interest with respect to prevention and early screening for language disorders. Complementary combinations could be used as an early index of development at

an age where supplementary combinations are absent (Capobianco et al., 2007). Similarly, the data showed that all children begin to produce early two-word utterances only after the emergence of “supplementary” crossmodal combinations. The onset and the use of “supplementary” combination denote a greater lexical and social-cognitive stage of development for the child involving the capacity to use predicates (verbs and adjectives) and pronouns, add the information to the DG and share more complex mental states (theory of mind).

The absence of “supplementary” crossmodal combinations around 24 months may indicate not only language immaturity compared to early verbal development but also social and cognitive immaturity. When the child proves able to share a state of mind with another (for example, by POINTING a ball and saying “beautiful!”), this may allow prediction of linguistic development (Franco, Perucchini & March, 2009; Goldin-Meadow, Brentari, Coppola, Horton & Sengales, 2015; Novack, Goldin-Meadow & Woodard, 2016). The data from this study confirm the link between language and cognition from the first stages of development, where early language skills are expressions of the onset the development social and cognitive abilities.

The importance of observing early development of “complementary” and “supplementary” crossmodal combination to identify risk profiles is linked to the possibility of early intervention to stimulate and increase the use of “multimodal” combinations (“naming” and “predicative” utterances) by adults who interact with the child during the first 2 years of age (Bonifacio & Stefani, 2011; Capobianco, 2015, Levickis, Reilly, Girolametto, Ukoumunne & Wake, 2014). Studies including a greater number of children will be necessary to collect reliable reference data to further investigate the spontaneous production of different gesture–word combinations as predictive indexes of early verbal skills for language difficulty prevention and early diagnosis purposes (Capobianco et al., 2007; Capobianco, 2015; Goldin-Meadow, 2014).

Acknowledgements

This paper is in memory and honour of our co-author Elena Pizzuto, who died in 2011.

References

- Bonifacio, S., & Stefani, L. (2011). *L'intervento precoce nel ritardo di linguaggio. L'intervento precoce nel ritardo del linguaggio. Il modello INTERCT per il bambino parlatore tardivo*. Milano: Franco Angeli.

- Butcher, C., & Goldin-Meadow, S. (2000). Gesture and the transition from one- to two-word speech: when hand and mouth come together. In D. McNeill (Ed.), *Language and gesture* (pp. 235–257). Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511620850.015
- Camaioni, L., Aureli, T., Bellagamba, F., & Presaghi, F. (2004). Traiettorie evolutive nella transizione alla combinazione simbolica. *Giornale Italiano di Psicologia*, 31(2), 307–334.
- Capirci, O., Iverson, J. M., Pizzuto, E., & Volterra, V. (1996). Gestures and words during the transition to two-word speech. *Journal of Child language*, 23, 645–673. doi: 10.1017/S0305000900008989
- Capirci, O., Caselli, M. C., Iverson, J. M., Pizzuto, E., & Volterra, V. (2002). Gesture and the nature of language in infancy: The role of gesture as transitional device en route to two word speech. In D. Armstrong, M. Karchmer, & J. Van Cleeve (Eds.), *Essays in honor of William C. Stokoe. The study of sign languages*: (pp. 213–246).
- Capirci, O., Contaldo, A., & Volterra, V. (2003). Il legame gesto-parola nelle prime fasi di sviluppo linguistico: una prospettiva longitudinale. Poster presented at the XVII National Congress AIP, *Developmental Psychology section*, Bari, Italy, September 22–24, 2003.
- Capirci, O., Contaldo, A., Caselli M. C., & Volterra, V. (2005). From action language through gesture. A longitudinal perspective. *Gestures*, 5(1).
- Capobianco, M. (2015). *Il linguaggio nei primi tre anni di vita. Metodi e tecniche per la valutazione, la prevenzione e l'intervento*. Milano: Franco Angeli.
- Capobianco, M. (2006). *Gesti, parole e prime combinazioni nello sviluppo tipico e primo confronto con bambini nati pretermine*. Doctoral Dissertation, University of Rome 1 “La Sapienza”.
- Capobianco, M., Pizzuto, E., & Devescovi, A. (2007) The role of gesture-speech cross-modal combinations in early language development: New longitudinal evidence. Dissertation presented at the *Third Conference of the International Gesture Society, Integrating Gestures*, Chicago, ILL, Northwestern University in Evanston (abstract in International Society for Gesture Studies “Integrating Gestures” Conference Program, pp. 52–53).
- Coletta, J., & Guidetti, M. (Eds.). (2012). *Gesture and multimodal development*. Amsterdam: John Benjamin. doi: 10.1075/bct.39
- Fasolo, M., & D’Odorico, L. (2012). Gesture-plus-word combinations, transitional forms, and language development. *Gesture*, 12 (1), 1–15. doi: 10.1075/gest.12.1.01fas
- Franco, F., Perucchini, P., & March, B. (2009). Is infant initiation of joint attention by pointing affected by type of interaction? *Social Development*, 55–76.
- Goldin-Meadow, S., & Morford, M. (1990). Gesture in early child language. In V. Volterra & C. J. Erting (Eds.), *From gesture to language in hearing and deaf children*. New York: Springer-Verlag, pp. 249–262 (199–2nd Edition Washington, D.C.: Gallaudet University Press). doi: 10.1007/978-3-642-74859-2_20
- Goldin-Meadow, S., & Butcher, C. (2003). Pointing toward two-word speech. In S. Kita (Eds.) *Pointing: where language, culture and cognition meet* (pp. 85–107). Mahwah, NJ: Lawrence Erlbaum.
- Goldin-Meadow, S., Brentari, D., Coppola, M., Horton, L., & Sengales, A. (2015). Watching language grow in the manual modality: nominals, predicates, and handschapes. *Cognition*, 136, 381–395. doi: 10.1016/j.cognition.2014.11.029
- Goldin-Meadow, S. (2014). How gesture works to change our minds. *Trends in Neuroscience and Education*, Vol. 3(1), 4–6. Elsevier. doi: 10.1016/j.tine.2014.01.002
- Guidetti, M. (2002). The emergence of pragmatics: forms and functions of conventional gestures in young French children. *First Language*, 22, 265–285. doi: 10.1177/014272370202206603

- Hollingshead, A. B. (1975). *Four Factor Index of Social Status*. Yale University, Department of Sociology, New Haven.
- Kendon, A. (2016). Reflections on the “Gesture First” Hypothesis of Language Origins. *Psychonomic Bulletin and Review*. doi: 10.3758/s13423-016-1117-3
- Iverson, J. M., Capirci, O., Volterra, V., & Goldin-Meadow, S. (2008). Learning to talk in a gesture-rich world: Early communication in Italian vs. American children. *First Language*, 8(2), 164–181. Special Issue: Gestures and communicative development. doi: 10.1177/0142723707087736
- Iverson, J. M., Capirci, O., & Caselli, M. C. (1994). From communication to language in two modalities. *Cognitive Development*, 9, 23–43. doi: 10.1016/0885-2014(94)90018-3
- Levickis, P., Reilly, S., Girolametto, L., Ukoumunne, O. C., & Wake, M. (2014). Maternal behaviors promoting language acquisition in slow-to-talk toddlers: prospective community-based Study. *Journal of Developmental and Behavioral Pediatrics*, 35(4), 274–281. doi: 10.1097/DBP.000000000000056
- Mayberry, R. I., & Nicoladis, E. (2000). Gesture reflects language development: evidence from bilingual children. *Current Directions in Psychological Science*, 9(6), 192–196. doi: 10.1111/1467-8721.00092
- McEachern, D., & Haynes, W. O. (2004). Gesture-speech combinations as a transition to multiword utterances. *American Journal of Speech-language Pathology*, 13, 227–235. doi: 10.1044/1058-0360(2004)024
- McNeill, D. (Eds.) (2000). *Language and gesture*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511620850
- Mac Whinney B. J. (1997). *Il progetto CHILDES. Strumenti per l'analisi del linguaggio parlato*. Edizioni del Cerro.
- Murillo, E., & Mercedes, B. (2012) Changes and predictive value on early lexical development. *Gesture*, 12(1), 16–39. doi: 10.1075/gest.12.1.02mur
- Novack, M., Wakefield, E., & Goldin-Meadow, S. (2016). What makes a movement a gesture? *Cognition*, 146, 339–348. doi: 10.1016/j.cognition.2015.10.014
- Pizzuto, E. (2002). Communicative gestures and linguistic signs in the first two years of life. Paper presented at the *EURESCO Conferences, Brain Development and Cognition in Human Infants*, Maratea, Italy.
- Pizzuto, E., & Capobianco, M. (2008). Is pointing “just” pointing? unraveling the complexity of indexes in spoken and signed discourse. *Gesture*, 8(1), 82–103. doi: 10.1075/gest.8.1.07piz
- Pizzuto, E., & Capobianco, M. (2009). L'indicare e la costruzione del discorso nell'acquisizione delle lingue vocali e della lingua dei segni. *Età evolutiva*, 11, 79–87.
- Pizzuto, E., & Capobianco, M. (2005). The link and differences between deixis and symbols in children's early gestural-vocal system. *Gesture*, 5(1/2), 179–199. doi: 10.1075/gest.5.1-2.13piz
- Pizzuto, E., Capobianco, M., & Devescovi, A. (2009). Gestural-vocal deixis and representational skills in early language development. In C. Abry, A. Vilain, & J. Schwartz (Eds.), *Vocalize to localize* (pp. 179–206), Amsterdam: John Benjamins. doi: 10.1075/bct.13.11piz
- Pizzuto, E., Capobianco, M., & Devescovi, A. (2005). Gestural-vocal deixis and representational skills in early language development. *Interaction Studies*, 6(2), 223–252. doi: 10.1075/is.6.2.05piz
- Pizzuto, E., Capobianco, M., & Devescovi, A. (2003). Gestural-vocaldeixis and representational skills in early language development. Paper presented at *International Congress on “Vocalize to localize: A missing piece in the puzzling route towards language?” CNRS-Institut de la Communication Parlée et Université Stendhal, Grenoble*.

- Stefanini, S., Bello, A., Caselli, M. C., Iverson J. M., & Volterra, V. (2009). Co-speech gestures in a naming task: Developmental data. *Language and Cognitive Processes*, 24, 168–189. doi: 10.1080/01690960802187755
- Tambelli, R., Cimino, S., Cerniglia, L., & Ballarotto, G. (2015). Early maternal relational traumatic experiences and psychopathological symptoms: a longitudinal study on mother-infant and father-infant. *Scientific Report*, 5, 13984. doi: 10.1038/srep13984
- Thal, D., & Tobias, S. (1992). Communicative gestures in children with delayed onset of oral expressive vocabulary. *Journal of Speech and Hearing Research*, 41, 374–384.
- Volterra, V., Caselli, M. C., Capirci, O., & Pizzuto, E. (2005). Gesture and the emergence and development of language. In M. Tomasello, & D. I. Slobin (Eds.), *Beyond nature-nurture - Essays in honour of Elizabeth Bates* (pp. 3–40). Mahwah, N.J.: Lawrence Erlbaum.

Authors' addresses

Micaela Capobianco	Elena Antinoro Pizzuto
Department of Psychology of Developmental Processes and Socialization	National Research Council
via dei Marsi, Rome	Via Nomentana 56
Italy	00161 Rome
	Italy

Antonella Devescovi
 Department of Psychology of Developmental Processes and Socialization
 via dei Marsi, Rome
 Italy

Biographical notes

Micaela Capobianco is a post-doctoral fellow at the University of Roma “La Sapienza”. From 2005 she teaches “Developmental Psychology” at the University “Lumsa” of Rome and, from 2013, is Tutor at the International Telematic University Uninettuno. Her research focuses on the role of gestures and combinations in early language learning in typically developing children, as well as in atypical conditions (e.g. pre-term children) and in the use of different language assessment methodologies in clinical practice.

Elena Pizzuto was a researcher at the “Institute of Cognitive Sciences and Technologies” of the National Research Council, where she coordinated the Sign Language Laboratory (www.istc.cnr.it/sll). Her research was on the linguistic investigation of Italian Sign Language (LIS) in a crosslinguistic, crosscultural perspective, and on language development in hearing and deaf children. She died in January of 2011.

Antonella Devescovi is professor of “Developmental Psychology” at the University of Rome “La Sapienza”, Department of Psychology of Developmental Processes and Socialization. Her research interests are primarily in comparative crosslinguistic investigations of spoken language processing, in children and adults, in typical and atypical conditions.