Infants' speech and gesture production in Mozambique and the Netherlands

Chiara de Jong (c.dejong_3@tilburguniversity.edu)
Paul Vogt (p.a.vogt@uvt.nl)
Tilburg centre for Cognition and Communication, Tilburg University
P.O. Box 90153, 5000 LE, Tilburg, The Netherlands

Abstract

In this paper, we explore the cultural differences in the production of speech and speech+gesture combinations by infants at the age of 17-18 months in Mozambique and the Netherlands. We found that Dutch infants produce more speech and gestures compared to Mozambican infants. Infants in both communities make most use of content words. The results further show that Dutch infants make more use of proximal pointing than Mozambicans, whereas Mozambicans make more use of the offering gesture. Finally, the amount of semantically coherent speech+gesture combinations of the Mozambican infants is higher than of the Dutch infants.

Keywords: Child language acquisition; culture; infant speech; infant gesture; semantic coherence; speech+gesture combinations.

Introduction

For infants in the early stages of language, gestures can allow them to express certain thoughts at a time they may be unable to express those thoughts fully in the spoken modality (Butcher & Goldin-Meadow, 2000; Goldin-Meadow & Butcher, 2003). Additionally, early gesture use in infants has been found to be a predictor of their later language development (Rowe, Özçalışkan & Goldin-Meadow, 2006, 2008). Despite the importance of gestures for language development, little knowledge is available about the use of (co-speech) gestures by infants in a cross-cultural perspective, especially in non-Western societies. In this paper we investigate the amount of speech and co-speech gestures produced by infants in two cultures. Additionally, we investigate the relation between different kinds of gestures and accompanying speech.

It is well established that there are considerable individual and cross-cultural differences in the language-learning environment of children (Hoff, 2006). For instance, concerning the amount of infant-directed speech (IDS; Lieven, 1994), the communicative intentions of IDS (LeVine et al., 1994; Vogt, Mastin & Schots, 2015) and style of co-speech gestures addressed to infants (Gogate, Maganti, & Bahrick, 2015; Tamis-LeMonda, Song, Leavell, Kahana-Kalman, & Yoshikawa, 2012). Similar results have been found concerning the influence of socio-economic status on the language environment of children (Hart & Risley, 1995; Rowe & Goldin-Meadow, 2009). Moreover, the various differences concerning the input to children have been found to affect children's development of language and other communicative skills (Hart & Risley, 1995; Rowe & Goldin-Meadow, 2009; Tamis-LeMonda et al., 2012). However, while such cultural differences concerning the learning environment of children are well known, they mostly concern the verbal environment or more Westernized cultures. Still little is known about (the role of) gestural communication in children ('s learning environments) from non-Western societies, such as those from Africa.

Studies have shown that the amount of gestural input, just like vocal input, varies across sociocultural settings and is not just dependent of SES or parents educational level (Callaghan et al., 2011; Iverson, Capirci, Volterra, & Goldin-Meadow, 2008; Salomo & Liszkowski, 2013; Vogt & Mastin, 2013). A recent observational study from Mozambique has obtained some novel insights. For instance, it was found that infants in urban Mozambique were stimulated much more frequently compared to infants from rural Mozambique across their whole gesture repertoire and were exposed to three times as much speech and co-speech gestures, which correlated to differences found in their vocabulary development (Vogt & Mastin, 2013). Also the communicative intentions of IDS differ considerably between the Netherlands, and urban and rural Mozambique (Vogt, et al., 2015). In particular, IDS in the Netherlands contained more utterances with a cognitive intention, whereas IDS in urban Mozambique contained more utterances with a socio-emotional intention, and IDS in rural Mozambique contained most of the imperatives. Additionally, the amount of IDS in the Netherlands was almost twice as much as the input in urban Mozambique, and about ten times as much as in rural Mozambique (Vogt et al., 2015).

The observed differences correspond to differences in caregiving practices between the communities, which can be explained based on lifestyles of the different socio-cultural communities (Keller, 2012). In a Western context, parents focus more on the cognitive development of the infant, such as language development. In a non-Western context the focus is more on situation-oriented socialization, with more attention for the development of motoric skills.

It is well established that the input in speech and gesture tends to correlate with children's development of speech and gesture (Hart & Risley, 1995; Iverson et al., 2008; Rowe & Goldin-Meadow, 2009; Salomo & Liszkowski, 2013). We therefore expect that the observed differences in the input to infants from the Netherlands and Mozambique will reflect differences in the *production* of speech and gesture between infants from the Netherlands and Mozambique. However, little is known about how infants in non-Western communities produce gesture and speech in everyday communication.

Salomo and Liszkowski (2013) have shown that the gesture productions of infants from the Netherlands, China

and a Mayan culture varied considerably with Chinese infants producing most gestures, followed by Dutch and Mayan infants. Assessing another three cultures (Canada, Peru and India) in a pointing task, Callaghan et al. (2011) found that significantly more Canadian infants produced declarative pointing gestures compared to Indian infants of the same age, while the number of Peruvian infants who pointed was in between.

The amount of gestures infants produce predicts the early stages of their spoken language development. Rowe et al. (2006; 2008) have shown that parents' use of gestures did not directly influence children's vocabulary size at 42 months. However, parental gesture use did influence the amount of gestures produced by infants at 14 months, which in turn was a significant predictor of vocabulary development at 42 months. Moreover, Rowe and Goldin-Meadow (2009a) found that children from low SES families produced fewer gestures than children from high SES families, which related to their parents' gestural input.

Children's gesture production is of importance since it does not only predict later vocabulary size, but is also involved in other developmental milestones, like the onset of the two-word stage. Children who are unable to express two words within one utterance can make use of a speech+gesture combination to express a two-word idea (Butcher & Goldin-Meadow, 2000; Goldin-Meadow & Butcher, 2003; Rowe & Goldin-Meadow, 2009b). The age at which children start to produce these speech+gesture combinations predicts the onset of two-word utterances (Iverson, et al., 2008; Iverson & Goldin-Meadow, 2005; Rowe & Goldin-Meadow, 2009b). Speech+gesture combinations conveying different information were found to emerge prior to the onset of two-word combinations for all cultures studied so far (Goldin-Meadow, 2009; Iverson et al., 2008; Iverson & Goldin-Meadow, 2005), but these do not include non-Western cultures.

The purpose of this paper is to explore the cultural differences in the speech and gestures infants produce at the age of 17-18 months. In particular, we compare the speech and gesture production of children from middle-class families in the Netherlands with children from low SES families in urban Mozambique. We assess infants' amounts and semantic categories of speech in combination with gesture production, as well as the mean length of utterances and speech+gesture utterances they produce. While investigating the relation of these differences with other developmental milestones, such as vocabulary development or the onset of two-word utterances, would be interesting, this is left for future research.

Methods

Participants and field sites

This study was based on data from earlier research of Mastin and Vogt (2016) and Vogt et al. (2015). The data consists of video recordings of infants in the Netherlands and urban Mozambique at three different ages, when infants were 1;1, 1;6 and 2;1 years old. For the purpose of the present study, only data recorded at 1;6 was used. Since the infants in rural Mozambique did not express many gestures and speech at the age of 1;6, that field site was excluded.

Table 1: Demographic information. *Note:* [a]Maternal education data from one Mozambican participant is missing.

Participant information	Mozambique (N = 11)	Netherlands (N = 11)
Female / male	5/6	5/6
Avg age (SD)	1;5.11 (29)	1;5.10 (20)
Education level mothers		
No education	1[a]	0
Primary education	8	0
Secondary education	1	0
Higher education	0	11

The field site in urban Mozambique was made up of two adjacent residential suburbs in Maputo. The community is relatively poor, with low parental education, and is has a market-based lifestyle. The Dutch sample consists of highly educated, native Dutch families living in the Tilburg region. The sample reflects a typical Western, middle-class, post-modern urban community that has a knowledge-based lifestyle.

The results reported here are from 11 infants from the Dutch community and 11 from urban Mozambique. The average age at the time of recording was 1;5.10 years (Table 1). In Mozambique, local research assistants explained the general purpose and recording procedures of the study to the families in their native language (Portuguese or Changana). In the Netherlands this was done by one of the researchers. Signed informed consent forms were obtained from the infants' mothers. Table 1 shows some basic demographics of our sample. For more details consult Vogt et al. (2015).

Data collection procedures

The families were visited at their home. During those visits naturalistic interactions were recorded of infants' daily, social environment by instructing those people who were present to carry on with what they were doing prior to our visit and to ignore our presence as well as possible. In Mozambique each visit was preceded by an accommodation session a week ahead in order to familiarize infants and their families with the presence of a foreign person who observes them with a video camera. In the Netherlands such an accommodation session was considered unnecessary since the infants have been exposed to video equipment before and the researcher was not foreign to them.

The infants were filmed for 45 to 75 minutes to obtain at least 30 minutes of clear and naturalistic footage for analysis. Video recording was done using a tripod whenever possible, placed approximately 5 meters from where most of the interactions occurred. In smaller confined areas, recording was done by hand-held camera.

Coding procedures

The videos were coded for 30 minutes during which the infant was displaying natural behaviour for a prolonged duration (i.e.: not asleep, off camera and interacting with or disturbed by the experimenters). Prolonged periods of breastfeeding (more than 2 minutes) were excluded too, since this could have introduced a bias toward dyadic interactions. The videos have been annotated for infant engagements (Mastin & Vogt, 2016), infant-directed speech and infants' speech (Vogt et al., 2015), and (co-speech) gestures directed to and produced by infants (Vogt & Mastin, 2013). The coding schemes were developed based on the literature and has been piloted and adapted with videos from an earlier pilot study from Mozambique. In this article, we only present the results of infants' speech and gestures. Interrater agreements are assessed and reported based on 35% cross-coding of data, except for the gestures of which only 15% was cross-coded.

Infant speech All speech produced by the infants during those 30 minutes was transcribed by research assistants who were native speakers of the language. The Mozambican data was subsequently translated to English. For the purpose of the present study, a distinction was made between *meaningful* and *meaningless* utterances. An utterance was considered meaningful when the speech was intelligible, or a clear meaning was conveyed. Only meaningful utterances were analysed in this study.

To investigate what type of words infants produced, we coded the semantic category of each utterance. In case the utterance contained only one word, this is the semantic category of that word; in case it was a multiword utterance, we coded the semantic category of the most prominent word in the utterance. We coded the infants' speech into the following three semantic categories (Cohen's kappa=0.90):

Content words. A word was considered to be a content word if it bears meaning on itself (e.g. 'daddy').
Demonstratives/deictics. A word was considered a demonstrative or deictic if it refers to something else, which can be another time, space, person or object or if it was meant to refer someone's attention to something (e.g. 'look'). These are words that cannot be understood without any additional contextual information (e.g. 'there').
Other. The category that was coded as 'other' consisted of words that did not belong to the previous categories, such as affirmatives or interjections (e.g. 'yes', 'so').

Gestures The gestures produced by the infants were coded during episodes of joint engagement. A gesture is broadly defined as any physical activity with the hand or body that has a clear communicative intent (Zukow-Goldring, 1996). For this research the coding from Vogt and Mastin (2013) and Vogt, Masson-Carro and Mastin (2015) was adopted. The following categories were distinguished (Cohen's kappa=0.58):

Proximal pointing. The infant points at an object in the near vicinity using his index finger or hand, in order to draw the attention of another individual to the target object.
Distal pointing. The infant points at an object placed far away using his index finger or hand, in order to draw the attention of another individual to the target object.
Showing. The infant holds an object, drawing another individual's attention to it, without any direct imperative.
Reaching. The infant moves his hand towards a target object with the intention to obtain it, but does not obtain said object. Also requests for objects by extending the hand were included in this category.
Offering. The infant offers (or gives) an object to another individual with the intention of transferring the object.
Taking. The infant takes an object from the other individual's possession while it has not been offered by the other.
Conventional gestures. This category comprises gestures that are symbolic of nature, such as emblematic gestures, but also gestures that bear an iconic relationship with their referent. For example, waving bye-bye, or indicating the size of the target object with the hands.
Ritualized play. This category accounts for all ritualized interactions or displays that occur between infants and communication partners. For instance, dancing, clapping hands or turn-taking games, such as patty-cake.
Request for attention. This category comprises any gesture that seeks for the attention of the addressee.

Speech gesture combinations In order to determine whether the gestural and vocal modalities are unified into one communication system, gesture and speech combinations were further coded. Unification is characterised by both semantic coherence and temporal synchrony (Butcher & Goldin-Meadow, 2000; Goldin-Meadow & Butcher, 2003). In case an utterance was not accompanied by a gesture, the utterance was coded as *speech only*. When an utterance was accompanied by a gesture, this was coded as a *speech+gesture combination*. A combination was categorized as *semantically coherent* when it was combining gesture with meaningful and related speech (Butcher & Goldin-Meadow, 2000). Following Goldin-Meadow and Butcher (2003), semantic coherence was further divided into two categories (Cohen's kappa=0.66):

Semantically congruent. A speech+gesture combination is semantically congruent when both speech and gesture bear the same meaning. An example of such a combination is a child pointing at a dog and uttering the word 'dog'.
Semantically related. A speech+gesture combination is semantically related when the meaning of the speech is different from the meaning of the gesture, but combined they form a semantically coherent utterance. For instance, a child shows a pair of glasses while uttering 'mommy', meaning that the pair of glasses is their mother's.

Table 2. Average number of utterances per infant, specified by semantic category. *Notes*: +p<.10; *p<.05; **p<.01; ***p<.001.

	Mozambique		Netherlands	
	Mean (SD)	Mean % (SD)	Mean (SD)	Mean % (SD)
All utterances	39.64 (27.22)		73.36 (41.76)+	
Content word	28.18 (22.26)	66.18 (28.34)	36.45 (25.91)	51.32 (23.80)
Deictic/ demonstratives	9.18 (11.27)	28.05 (30.06)	14.82 (14.30)	16.71 (11.74)
Other	2.18 (1.72)	5.78 (6.05)	22.00 (12.47)	31.98 (21.19)***
Speech+gesture	10.00 (6.96)		22.09 (14.90)*	
Content word	5.18 (3.95)	52.23 (27.20)	12.09 (10.26)	56.40 (29.68)
Deictic/ demonstratives	4.55 (4.87)	46.04 (27.87) +	6.55 (7.03)	26.49 (19.77)
Other	0.27 (0.47)	1.72 (3.17)	3.36 (3.72)	17.11 (24.88)*
Speech only	29.73 (21.64)		51.27 (30.14)	
Content word	23.09 (18.67)	81.77 (20.64)**	24.36 (17.61)	48.91 (23.30)
Deictic/ demonstratives	4.64 (7.70)	11.54 (19.81)	8.27 (7.63)	13.26 (9.97)
Other	1.91 (1.51)	6.69 (6.83)	18.64 (10.74)	37.84 (20.10)***

Results

Infant speech

Table 2 shows the average number of utterances and speech+gesture combinations expressed by the infants. In Mozambique the average number of utterances (including both speech only and speech-gesture combinations) per child was M=39.64 (SD=27.22), against M=73.36 (SD=41.76) in the Netherlands. A Mann-Whitney U test showed that this difference tended towards significance (U=31.00, p=.056). Of the total number of utterances, 30.11% was combined with a gesture in the Netherlands (n = 22.09 average per child) and 25.23% in Mozambique (n = 10.00 average per child), which resulted in a significantly larger amount of speech+gesture combinations in the Netherlands compared to Mozambique (U=26.00, p=.023). To analyse the semantic categories of speech between Mozambican and Dutch infants, we compared the mean percentages of each category. In both communities gestures were most often accompanied by content words, 52.23% in Mozambique and 56.40% in the Netherlands. However, in Mozambique, a gesture was more frequently accompanied by a demonstrative or deictic (46.04%) than in the Netherlands (26.49%). This difference is marginally significant (U=34.00, p=.088). In Mozambique, gestures were significantly less frequently accompanied by *other* words (1.72%) than in the Netherlands (17.11%; U=26.00, p=.023).

When infants produced speech without a gesture, Mozambican infants significantly more often produced a content word (81.77%) compared to the Netherlands (48.91%; U=18.00, p=.004). For speech only, the percentages of deictics and demonstratives were similar for both cultures. However, Dutch infants produced significantly more *other* words (37.84%) than in Mozambique (6.69%; U=4.00, p<.001).

Infant gesture

Table 3 shows the mean frequencies with which infants produced the various gestures in speech+gesture combinations. The gesture that was used the most in Mozambique was offering (35.50%), whereas in the Netherlands this was proximal pointing (45.27%). Comparison of gesture use in Mozambique with the Netherlands using the Mann-Whitney U test showed that gesture use only differed significantly for these two gestures: proximal pointing (U=2.00, p<.001) and offering (U=20.00, p=.007). For all the other gestures, no significant difference across the two cultures was observed.

Table 4 shows to what extent the speech+gesture combinations were coherent. While the total number of speech+gesture combinations was higher in the Netherlands, the mean percentage of coherent combinations in Mozambique (M=92.86; SD=14.99) was significantly higher than in the Netherlands (M=80.06; SD=15.23; U=20.00, p=.007). Interestingly, in Mozambique infants produced relatively more semantically congruent combinations (M=39.07; SD=40.84) than in the Netherlands (M=9.70; SD=9.07), while Dutch infants produced relatively more semantically related combinations (M=70.36; SD=15.19) compared to Mozambicans (M=53.79; SD=39.58). A Chi square showed that there was a significant association between country and semantic congruency (χ^2(1, 353)=14.33, p<.001).

Table 3: Average gesture use in numbers and frequencies. *Note:* **p<.01, ***p<.001.

	Mozambique		Netherlands	
	M (SD)	% (SD)	M (SD)	% (SD)
Proximal pointing	0.73 (1.42)	4.53 (8.22)	10 (7.92)	46.64 (24.00)***
Distal pointing	1.18 (1.47)	17.07 (19.63)	2.64 (3.47)	10.05 (9.52)
Showing	0.73 (1.01)	5.66 (8.78)	2.09 (2.81)	10.21 (8.83)

Reaching	1.64 (2.25)	14.75 (19.91)	1.91 (2.74)	7.91 (10.47)
Offering	3.55 (4.74)	32.63 (23.30)**	1.82 (4.24)	6.40 (12.27)
Taking	0.27 (0.47)	2.20 (4.49)	0.82 (2.59)	4.45 (7.84)
Conventional	0.36 (0.67)	2.26 (4.25)	1.91 (2.59)	10.77 (19.52)
Ritualized play	0.91 (1.58)	15.06 (27.52)	0.64 (1.12)	2.67 (5.30)
Attention request	0.64 (1.80)	5.84 (15.26)	0.27 (0.65)	0.91 (2.24)

Table 4: Percentage of semantically coherent speech+gesture combinations conveying congruent or related information. *Note:* **p<.01.

	Mozambique		Netherlands	
	M (SD)	% (SD)	M (SD)	% (SD)
Coherent	9.64 (7.00)	92.86 (14.99)**	18.27 (12.48)	80.06 (15.23)
Congruent	5.18 (4.40)	39.07 (40.84)	16.00 (10.70)	9.70 (9.07)
Related	4.45 (6.33)	53.79 (39.58)	2.27 (2.80)	70.36 (15.19)

Discussion

The objective of this paper is to explore the cultural differences in the speech and gestures production of infants at the age of 17-18 months from the Netherlands and Mozambique. Since the sample size was small and the nature of the research exploratory, generalizability is difficult. Moreover, the two communities do not only differ in culture but also in socioeconomic status. Many of the differences that we find may therefore not only be explained due to cultural differences but also to SES differences. Many of the differences that we found are in line with those found in other studies of the same nature (Hart & Risley, 1995; Rowe & Goldin-Meadow, 2009a).

The results revealed that infants from the Netherlands produced twice as many utterances than infants from Mozambique (Table 2). This is also the case for speech+gesture combinations and for speech only productions, but the latter difference was not significant. These differences reflect the difference observed in the amount of IDS, which was also twice as much in the Netherlands compared to urban Mozambique (Vogt et al., 2015).

Looking at the type of utterances, infants from both countries most often produced content words. For utterances not accompanied by gestures, we found that the percentage of content words is higher in Mozambique compared to the Netherlands. This may partly be caused by the low proportion of *other* words (like affirmatives or interjections) observed in Mozambique compared to the proportion of other words produced in the Netherlands. As a result, the utterances are redistributed over content words and deictics, which may cause a skewed distribution.

The higher proportion of content words in the absence of a gesture may also be related to a relatively frequent use of gestures when producing a deictic or a demonstrative. When combining speech with a gesture, Mozambican infants tend to produce more deictics and demonstratives compared to Dutch infants, though this difference is only marginally significant. This tendency may be explained by the high frequency of the Mozambican deictic '*im*', which means 'look at my gesture'. 'Im' is a high frequency word in Changana and Mozambican Portuguese, especially by and for children, and is practically always accompanied by a gesture, thus boosting the frequency of deictics and demonstratives in combination with a gesture.

The relatively low frequency of the *other* semantic category in Mozambique may reflect the lack of diversity of the IDS, because the more diverse the IDS, the more diverse infants' vocabularies tend to grow (Hart & Risley, 1995). Hart and Risley demonstrated that the diversity of IDS strongly relates to socio-economic status or maternal education. Since the maternal education in Mozambique is relatively low, the diversity of IDS may be low, thus explaining low diversity in the infants' production. Future analyses on the type-token ratio should inspect whether this is indeed the case here.

Looking at the different gestures infants produce (Table 4), we see that while Dutch infants produced more than twice as many gestures, the relative frequencies are highly similar for both cultures except for the amount of proximal pointing and the number of offering gestures observed. Dutch infants produced substantially more proximal pointing gestures than Mozambican infants. Almost half of the gestures Dutch children produce are pointing gestures to an object in close proximity. This is consistent with results from other cross-cultural studies (Callaghan et al., 2011; Salomo & Liszkowski, 2013), and may be explained by the fact that pointing facilitates joint attention and word learning, which are fostered more in Western middle class societies than non-Western middle class societies (Keller, 2012).

The higher frequency of the offering (including giving) gesture in Mozambique might be due to the importance of sharing in non-Western cultures as part of the social responsibility (Keller, 2012). Another explanation can be the high amount of imperatives directed to children in Mozambique (Vogt et al., 2015), resulting in children offering their caregivers the things they asked for.

Finally, we observe that in both countries infants' speech+gesture combinations attained a high level of semantic coherence, and thus in both cultures children have a unified communication system in which gesture and speech are integrated to convey one single message (Butcher & Goldin-Meadow, 2000). Despite the larger number of speech+gesture combinations observed in the Netherlands, Mozambican infants produced proportionally more semantically coherent combinations. Moreover, these combinations often convey the same information in Mozambique and are thus semantically congruent, whereas

in the Netherlands the combinations often convey different information, and are thus semantically related. Given the suggested associations between semantic congruency and temporal synchrony, and between semantic relatedness and the onset of two-word utterances (Goldin-Meadow & Butcher, 2003), these findings could predict related cultural differences in the development of temporal synchrony and two-word utterances.

Acknowledgments

We express our gratitude to Doug Mastin, Ingrid Masson-Carro and various other research assistants for their contributions in collecting and annotating the data. Additionally we offer a special thanks to our participant families for their kind hospitality and collaboration.

References

Butcher, C., & Goldin-Meadow, S. (2000). Gesture and the transition from one- to two-word speech: when hand and mouth come together. In D. McNeil (Ed.), *Language and gesture*. New York: Cambridge University Press.

Callaghan, T., Moll, H., Rakoczy, H., Warneken, F., Liszkowski, U., Behne, T., Tomasello, M., & Collins, W. A. (2011). *Early social cognition in three cultural contexts*. Wiley-Blackwell.

Gogate, L., Maganti, M., & Bahrick, L. E. (2015). Cross-cultural evidence for multimodal motherese: Asian Indian mothers' adaptive use of synchronous words and gestures. *Journal of experimental child psychology*, *129*, 110-126.

Goldin-Meadow, S. (2009). How gesture promotes learning throughout childhood. *Child development perspectives*, *3*, 106-111.

Goldin-Meadow, S., & Butcher, C. (2003). Pointing toward two-word speech in young children. In S. Kita (Ed.), *Pointing: Where language, culture, and cognition meet*. Mahwah, NJ: Erlbaum Associates.

Hart, B. & Risley, T. R. (1995). *Meaningful differences in the everyday experiences of young American children*. Baltimore, MD: Paul H. Brooks.

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, *26*, 55-88.

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*, *28*, 164-181.

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological science*, *16*, 367-371.

Keller, H. (2012). Autonomy and relatedness revisited: Cultural manifestations of universal human needs. *Child Development Perspectives*.

LeVine, R. A., Dixon, S., LeVine, S., Richman, A., Keefer, C. H., Leiderman, P. H., & Brazelton, T. B. (1994). *Child care and culture: Lessons from Africa.* Cambridge, UK: Cambridge University Press

Lieven, E. V. (1994). Crosslinguistic and crosscultural aspects of language addressed to children. In C. Gallaway & B. Richards (Eds.), (1994). *Input and interaction in language acquisition*. New York, US: Cambridge University Press.

Mastin, J. D., & Vogt, P. (2016). Infant engagement and early vocabulary development: A naturalistic observation study of Mozambican infants from 1;1 to 2;1. *Journal of child language*, *43*, 235-264.

Rowe, M. L., & Goldin-Meadow, S. (2009a). Differences in early gesture explain SES disparities in child vocabulary size at school entry. *Science*, *323*, 951-953.

Rowe, M. L., & Goldin-Meadow, S. (2009b). Early gesture selectively predicts later language learning. *Developmental science*, *12*, 182-187.

Rowe, M., Özçalıskan, S., & Goldin-Meadow, S. (2006). The added value of gesture in predicting vocabulary growth. *Proceedings of the 30th annual Boston University conference on language development* (pp. 501-512).

Rowe, M. L., Özçalışkan, Ş., & Goldin-Meadow, S. (2008). Learning words by hand: Gesture's role in predicting vocabulary development. *First language*, *28*,182-199.

Salomo, D., & Liszkowski, U. (2013). Sociocultural settings influence the emergence of prelinguistic deictic gestures. *Child development*, *84*, 1296-1307.

Tamis‐LeMonda, C. S., Song, L., Leavell, A. S., Kahana‐Kalman, R., & Yoshikawa, H. (2012). Ethnic differences in mother–infant language and gestural communications are associated with specific skills in infants. *Developmental Science*, *15*, 384-397.

Vogt, P., & Mastin, J. D. (2013). Rural and urban differences in language socialization and early vocabulary development in Mozambique. In *Proceedings of the 35th Annual Meeting of the Cognitive Science Society* (pp. 3787-3792). Austin, TX: The Cognitive Science Society.

Vogt, P., Mastin, J. D., & Schots, D. M. (2015). Communicative intentions of child-directed speech in three different learning environments: Observations from the Netherlands, and rural and urban Mozambique. *First Language, 35*, 341-358.

Zukow-Goldring, P. (1996). Sensitive caregiving fosters the comprehension of speech: When gestures speak louder than words. *Early development and parenting, 5*, 195-211.