

Predicting vocabulary development from co-speech gestures: Duration or occurrences, that's the question

Paul Vogt (p.a.vogt@uvt.nl)

Ingrid Masson-Carro (i.masson@uvt.nl)

J. Douglas Mastin (j.d.mastin@uvt.nl)

Tilburg centre for Cognition and Communication (TiCC), School of Humanities,
Tilburg University, The Netherlands.

Abstract

In this paper, we investigate whether or not the duration of exposures to co-speech gestures predicts later vocabulary development better than occurrence frequencies. To this aim we examine the impact of child-directed co-speech gestures on vocabulary development in infants from two cultural groups within Mozambique. We find that duration and occurrence are strongly related and both can predict later vocabulary development almost equally well. In addition, we find considerable cultural differences in the amount and style of co-speech gestures addressed to infants, as well as the way these predict later vocabulary development.

Keywords: vocabulary development, co-speech gestures, methodology, cultural differences.

Introduction

It has been established that co-speech gestures addressed to infants are strong predictors for later vocabulary development (Iverson, Capirci, Longobardi, & Caselli, 1999; Pan, Rowe, Singer, & Snow, 2005). One explanation is that the amount of gestures addressed to children correlates to the amount of gestures produced by children, which in turn predicts later vocabulary size (Rowe, Özçaliskan, & Goldin-Meadow, 2008). Another explanation for this finding is that co-speech gestures may help children to establish joint attention (Bates, Camaioni, & Volterra, 1975), which allows children to acquire the correct word-meaning mappings (Tomasello & Todd, 1983). Although both explanations are likely and not mutual exclusive, we will use the second explanation as the motivational basis for the present study.

Researchers usually measure the occurrence frequencies (occurrence, for short) of gestures (Iverson et al., 1999; Pan et al., 2005; Rowe et al., 2008). Tomasello and Todd (1983) have shown that the amount of time (i.e. *duration*) children engage in joint attention interactions predicts later vocabulary development. Assuming that at least certain child-directed gestures, such as pointing, help infants to establish joint attention, it seems likely that duration of gestures predicts later vocabulary size. Why do researchers tend to measure occurrence rather than duration? It seems plausible that holding a pointing gesture provides the addressee more opportunity to identify the referent than short points. Iverson, Longobardi, Spampinato, and Caselli (2006) have suggested, in a study concerning children with

Down's syndrome, that holding gestures addressed to children for the duration of verbal utterances aids children to maintain attention to a referent, thus providing opportunity for these children to infer the relation between the utterance and the reference. Only few studies have measured the duration of children's, as well as child-directed gestures, but have not done so in relation to vocabulary development (Cochet & Vauclair, 2010). Moreover, no studies were found in the literature that have systematically investigated whether occurrence or duration predict later vocabulary development differently. Therefore, we investigate whether duration of certain co-speech gestures addressed to infants can predict later vocabulary better than occurrence, while for other gestures occurrence is a better predictor.

We investigate these questions using observations of natural interactions between 1;1 years old infants and their social environment obtained in rural and urban Mozambique (Vogt & Mastin, 2013). The reason for using observations, rather than experiments, is that they provide us better insights into how children are stimulated by their natural social surroundings, which constitute the infants' learning environment. By using these data, we additionally gain insights into the differences between the ways children from rural and urban Mozambique are addressed and how these relate to vocabulary development.

Studying the relation between child-directed co-speech gestures and vocabulary development in a non-industrialised country is interesting, because there are differences in the amount of social interactions infants across cultures engage in (Brown, 2011). However, it is unclear whether such differences also relate to the duration of gestures. Moreover, Keller (2012) has argued that caregivers in non-industrial communities tend to value the development of social and motor skills more than object-oriented cognitive skills. It is therefore interesting to explore whether there are clear rural-urban differences regarding the types of gestures addressed to infants and how this relates to vocabulary development.

Method

Participants

Data were collected longitudinally from two cultural groups in Mozambique: one urban and one rural. In each community we recruited 14 families with an infant of approximately 1;1-year old (rural: $M=1;1.8$, $SD=0;0.26$;

urban, $M=1;1.6$, $SD=0;0.28$) at the start of the study. Participants in the rural community were monolingual Changana speakers (a Southern Bantu language spoken in the South of Mozambique). In the urban community, infants were raised bilingually in Portuguese (the official language) and Ronga (a local language that is mutually intelligible with Changana). In both communities, there was an appropriate representation of both genders, and there were no significant differences in family size. Mothers' education level was generally very low, but slightly higher in the urban community.

Procedure

The present study followed a longitudinal design, with three data-collection points when infants were 1;1, 1;5, and 2;1 years old. Each participant family was visited twice during each time period. In these visits, one experimenter and one local research assistant were present. During the first visit, the infant and other family members were videotaped so that they could accommodate to the presence of the research team and filming procedures. Prior to each recording session, the families were instructed to continue their daily activities as if the researchers were not present. During the second visit, the infants were recorded with their caregivers during 45 to 75 minutes of free behaviour. Subsequently, the caregivers were interviewed to fill in the MacArthur-Bates Communicative Development Inventory (MBCDI) short infant version (Fenson et al., 2000), adapted for the local languages, to assess the infants' vocabulary development.

Coding

All videos collected in the second visits were coded for approximately 30 minutes ($M=27:57$; $SD=01:52$) in which the infants were on-camera, awake, not interacting with the researchers, and not too distressed. In this report, we focus on data recorded at infants' age of 1;1.

Gesture coding A range of speech-accompanying non-verbal behaviours produced during episodes of *joint engagement* (Mastin & Vogt, 2013) were coded according to the following eleven categories (Cohen's kappa is 0.67 'substantial agreement' on the aggregate gesture coding):

Proximal pointing The speaker points at an object in the near vicinity using his index finger or hand, in order to draw the infant's attention to the target object.

Distal pointing The speaker points at an object placed far away using his index finger or hand, in order to draw the infant's attention to the target object.

Non-manual pointing The speaker points at an object using any part of his body except the hands, in order to draw the infant's attention to the target object (e.g. head-pointing).

Showing The speaker holds an object, drawing the infant's attention to it.

Demonstrating The speaker manipulates an object to show the infant how that object is used, or the type of actions that can be performed upon it.

Reaching The speaker 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 speaker offers (or gives) an object to the infant.

Taking The speaker takes an object from the infant's possession.

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.

Ritualised play This category accounts for all ritualised interactions or displays that occur between infants and communication partners. For instance, dancing, clapping hands or turn-taking games, such as patty-cake.

Embody The speaker performs a physical action onto the infant's body, such as redirecting or correcting its position, or taking its hand to lead the way (cf., Zukow-Goldring, 1996).

Request for attention This category comprises any gesture that seeks for the attention of the infant.

Data analyses

The duration frequencies were calculated by dividing the sum of the durations of a specific co-speech gesture by the total duration of each video. This normalization was carried out, because the total duration of each video did not sum up to exactly 30. Occurrence frequencies were obtained by counting the occurrence of each gesture. The Mann-Whitney U test was used to assess whether frequencies differed significantly across communities. We used Spearman's rank correlation to correlate these frequencies with expressive vocabulary at 1;5 and 2;1 as measured by the MBCDI.

Results

Figure 1 shows the mean values of co-speech gesture duration (top) and occurrence (bottom) in urban and rural areas. The graphs show considerable differences between the amounts of child-directed gestures in both communities, with children in urban areas receiving more gestural input than children in rural areas. If we look at duration frequencies, *demonstration* ($U=45$; $p<.05$), *reaching* ($U=46$; $p<.05$), *offering* ($U=47$; $p<.05$), *ritualised play* ($U=52$; $p<.05$), *embody* ($U=35$; $p<.01$) and *request attention* ($U=36$; $p<.01$) last significantly longer in the urban community. Occurrence frequencies of most of these gestures are also significantly higher in the urban area: *Showing* ($U=38$; $p<.01$), *demonstrating* ($U=45$; $p<.05$), *reaching* ($U=42$; $p=.01$), *offering* ($U=28$; $p<.01$), *embody* ($U=31$; $p<.01$) and *request for attention* ($U=33$; $p<.01$).

The two graphs of Figure 1 further show that within each community duration and occurrence frequencies reveal similar patterns. To assess how well both measures are related, we calculated the Spearman correlations between

the duration and occurrence frequencies for all gestures. For the urban community, most gestures have a correlation coefficient $r > 0.86$ ($p < 10^{-5}$), except for *reaching* ($r = 0.71$; $p = .005$), *offering* ($r = 0.59$; $p = .028$) and *embody* ($r = 0.74$; $p = .003$). For the rural community, all gestures have a correlation $r > 0.92$ ($p < 10^{-6}$), except for *showing* ($r = 0.83$; $p = .0005$) and *taking* ($r = 0.71$; $p = .006$). So, duration and occurrence frequencies of each *single* co-speech gesture are highly similar.

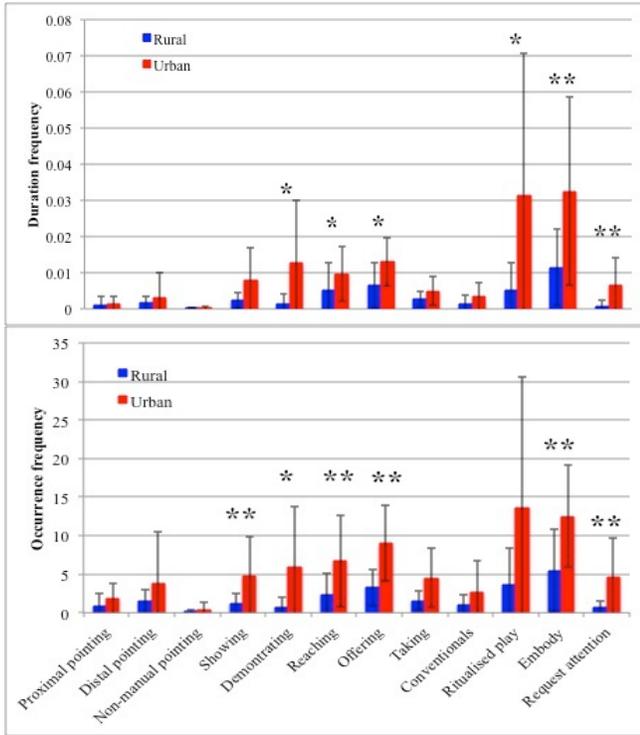


Figure 1: Means and Standard Deviations of duration and occurrence frequencies of gestures in urban and rural areas. Intercultural differences: * $p < .05$; ** $p < .01$.

Table 1 shows the correlations of the duration and occurrence frequencies with vocabulary development at 1;1 and 2;1. At a first glance we can see that very few gestures correlate significantly to vocabulary size, but there are some cultural differences. For example, only positive correlations occur in the urban area, whereas in the rural area there is only one significant, but negative, correlation. In particular, we can see that in the urban community, *ritualised play* yields significant correlations between duration and vocabulary size at 1;5 ($r = 0.59$; $p < .05$), and for occurrence correlations are found at both ages ($r_{1;5} = 0.54$ and $r_{2;1} = 0.56$; $p < .05$). In addition, duration and occurrence of *request attention* correlates significantly with vocabulary size at 2;1 ($r_{dur} = 0.54$ and $r_{occ} = 0.65$; $p < .05$). In the rural area, the duration frequency of *distal pointing* correlates negatively with vocabulary development at 2;1 ($r = -0.57$; $p < .05$). However, no significant correlations are found for the occurrence frequencies of co-speech gestures in the rural community.

Table 1: Correlations between gestures and vocabulary

Gesture	Vocab. at 1;5		Vocab. at 2;1	
	Dur.	Occ.	Dur.	Occ.
<i>Urban area</i>				
Proximal pointing	0.17	0.25	0.02	0.04
Distal pointing	0.19	0.18	0.37	0.44
Non-manual point.	0.03	0.06	0.44	0.45
Showing	0.07	0.09	0.02	0.05
Demonstrating	0.27	0.24	0.18	0.31
Reaching	0.28	0.26	0.10	0.35
Offering	0.17	0.42	0.22	0.45
Taking	0.37	0.44	0.22	0.24
Conventional	0.26	0.23	0.24	0.32
Ritualised play	0.59*	0.54*	0.52	0.56*
Embody	0.17	0.29	0.32	0.47
Request attention	0.35	0.31	0.54*	0.65*
<i>Rural area</i>				
Proximal pointing	-0.25	-0.27	0.11	0.14
Distal pointing	0.07	0.23	-0.57*	-0.48
Non-manual point.	-0.38	-0.38	0.21	0.21
Showing	-0.44	-0.45	-0.32	-0.25
Demonstrating	0.07	0.01	0.23	0.09
Reaching	-0.00	0.04	-0.10	0.11
Offering	-0.26	-0.23	-0.23	-0.10
Taking	-0.03	-0.39	-0.27	-0.38
Conventional	-0.20	-0.13	0.17	0.12
Ritualised play	-0.01	-0.11	0.49	0.48
Embody	0.07	0.07	0.35	0.40
Request attention	0.29	0.31	0.28	0.28

Note: * $p < .05$.

Discussion

The main objective of this paper is to investigate whether the duration of co-speech gestures is a better predictor for later vocabulary development than occurrence frequencies. The results show that gestures which correlate significantly with vocabulary (*ritualised play*, *request attention* and *distal pointing*) do so both for duration and occurrence, except for the duration of *ritualised play* and the occurrence of *distal pointing*. A close comparison of the correlation coefficients for both types of frequencies from Table 1 indicates that overall they are quite similar. So, the results suggest that duration and occurrence predict later vocabulary almost equally well. This may be explained by the strong correlations that we found between duration and occurrence.

A question that arises is how strong this correlation should be in order to conclude that duration essentially measures the same as occurrence. A correlation coefficient of $r = 0.59$ between duration and occurrence for *offering* in the urban community, although significant, is relatively low to explain a one-to-one correspondence between the two. This appears in accordance with the substantial difference in correlations with vocabulary between duration and occurrence of urban *offering* (Table 1). Similar observations can be made for urban *reaching* and *embody*, and rural *taking*; for rural *showing* the correlations with vocabulary size do not differ substantially, but here the correlation between duration and occurrence is higher ($r = 0.83$). On the

other hand, urban *demonstration* has a high duration-occurrence correlation ($r > .86$), yet they yield substantial different correlations with vocabulary.

So, there are strong correlations between duration and occurrence for most gestures, and both reveal comparable correlations with vocabulary size. However, it would be premature to conclude that both are measuring the same or that one of them predicts later vocabulary development better than the other. Experimental studies in which duration and occurrence are carefully manipulated could provide additional insights.

The second objective of this paper is to investigate urban and rural differences concerning the way co-speech gestures relate to vocabulary development. Figure 1 clearly shows that urban children are stimulated much more frequently than rural children across all gestures. This is in line with an earlier analysis of the same data where the occurrence of all gestures taken together was compared cross-culturally (Vogt & Mastin, 2013). That study revealed that urban children are addressed about 3.2 times more frequently than their rural counterparts. Moreover, Vogt and Mastin found that the overall occurrence of co-speech gestures addressed to urban infants correlates strongly to later vocabulary size, but no such correlation was found in the rural area. A possible explanation for this lack of correlation in the rural area is that infants there are faced with a changing caregiving system between 1;1 and 1;5, which makes predicting vocabulary size at 1;5 or later difficult from observations at 1;1 (Vogt & Mastin, 2013).

Interestingly, *ritualised play*, which is the most occurring gesture in the urban area, correlates strongly to vocabulary development in that community. It also shows a tendency towards a significant correlation to vocabulary at 2;1 in the rural area. This is - to some extent - in line with Keller's (2012) observation that caregivers in non-industrialised communities are more concerned with raising their children to develop social and motor skills than cognitive (object related) skills. It is likely that social and motoric stimulation are better facilitated through *ritualised play* gestures.

Deictic gestures, such as *pointing*, *giving*, *reaching*, etc., appear to have little effect on vocabulary development, except for *showing*, which in the rural community reveals a negative correlation. Again, this follows Keller's suggestion that caregivers in non-industrialised communities tend to focus less on object-oriented cognitive development of their children.

To conclude, we did not find evidence that duration of gestures addressed to infants is a better predictor for later vocabulary size than occurrence. It seems likely that a balanced use of both provide an optimal learning environment. However, more research is needed to draw firm conclusions.

Acknowledgments

This research was supported with a VIDI grant from the Netherlands Organization for Scientific Research (NWO, grant number 276-70-018). We express our gratitude to

those Mozambicans who have assisted or participated in this research. Special thanks go to Van Le and Mohammad Hafiz for their assistance in coding the videos.

References

- Bates, E., Camaioni, L., & Volterra, V. (1975). The acquisition of performatives prior to speech. *Merrill-Palmer Quarterly: Journal of Developmental Psychology*.
- Brown, P. (2011). The Cultural Organization of Attention. In A. Duranti, E. Ochs & B. Schieffelin (Eds.), *The Handbook of Language Socialization*: Blackwell Publishing Ltd.
- Cochet, H., & Vauclair, J. (2010). Pointing gestures produced by toddlers from 15 to 30 months: Different functions, hand shapes and laterality patterns. *Infant Behavior and Development*, 33(4), 431-441.
- Fenson, L., Pethick, S., Renda, C., Cox, J. L., Dale, P. S., & Reznick, J. S. (2000). Short-form versions of the MacArthur communicative development inventories. *Applied Psycholinguistics*, 21(1), 95-116.
- Iverson, J. M., Capirci, O., Longobardi, E., & Caselli, C. M. (1999). Gesturing in mother-child interactions. *Cognitive Development*, 14(1), 57-75.
- Iverson, J. M., Longobardi, E., Spampinato, K., & Caselli, C. M. (2006). Gesture and speech in maternal input to children with Down's syndrome. *International Journal of Language & Communication Disorders*, 41(3), 235-251.
- Keller, H. (2012). Autonomy and relatedness revisited: Cultural manifestations of universal human needs. *Child Development Perspectives*.
- Mastin, J. D., & Vogt, P. (2013). Correlations between infant engagement levels and vocabulary development: An observational study of Mozambican infants from 1;1 to 2;1. *Under review*.
- Pan, B. A., Rowe, M. L., Singer, J. D., & Snow, C. E. (2005). Maternal correlates of growth in toddler vocabulary production in low income families. *Child Development*, 76(4), 763-782.
- Rowe, M. L., Özçaliskan, S., & Goldin-Meadow, S. (2008). Learning words by hand: Gesture's role in predicting vocabulary development. *First language*, 28(2), 182-199.
- Tomasello, M., & Todd, J. (1983). Joint attention and lexical acquisition style. *First Language*, 4, 197-212.
- Vogt, P., & Mastin, J. D. (2013). *Language socialization and infants' vocabulary development in rural and urban Mozambique*. Paper presented at the CogSci 2013.
- Zukow-Goldring, P. (1996). Sensitive caregiving fosters the comprehension of speech: When gestures speak louder than words. *Early development and parenting*, 5(4), 195-211.