Onomatopoeia, gestures, actions and words:
How do caregivers use multimodal cues in their communication to children?

Gabriella Vigliocco (g.vigliocco@ucl.ac.uk)
Yasamin Motamedi, Margherita Murgiano, Elizabeth Wonnacott, Chloe Marshall, Iris Milan Maillo,
Pamela Perniss

Abstract

Most research on how children learn the mapping between words and world has assumed that language is arbitrary, and has investigated language learning in contexts in which objects referred to are present in the environment. Here, we report analyses of a semi-naturalistic corpus of caregivers talking to their 2-3 year-old. We focus on caregivers' use of non-arbitrary cues across different expressive channels: both iconic (onomatopoeia and representational gestures) and indexical (points and actions with objects). We ask if these cues are used differently when talking about objects known or unknown to the child, and when the referred objects are present or absent. We hypothesize that caregivers would use these cues more often with objects novel to the child. Moreover, they would use the iconic cues especially when objects are absent because iconic cues bring to the mind's eye properties of referents. We find that cue distribution differs: all cues except points are more common for unknown objects indicating their potential role in learning; onomatopoeia and representational gestures are more common for displaced contexts whereas indexical cues are more common when objects are present. Thus, caregivers provide multimodal non-arbitrary cues to support children's vocabulary learning and iconicity – specifically – can support linking mental representations for objects and labels.

Keywords: language development; word learning; iconicity; onomatopoeia; co-speech gestures; child directed speech; naturalistic observation.

Introduction

Understanding how children acquire language, its onset and the developmental path thereafter - is one of the great challenges for the social sciences, with critical implications for education and for intervention in atypically developing children. Vocabulary learning is a central part of language development and is characterized as a hard problem: How do children know that the sounds people produce are 'words' for objects, actions and properties? At the core of most existing proposals is the long-held assumption that language is purely arbitrary: there is no recognizable link between a label and the corresponding referent in the world (e.g. between the English word *dog* and the furry, four-legged animal; de Saussure, 1916). Arbitrariness makes the task of learning words especially hard: how can children learn the correct referent in a visually cluttered world (where multiple objects, actions and properties are all possible candidates for a given label), or even worse, when the objects, actions and properties talked about are absent from the immediate environment?

However, in addition to being arbitrary, language presents also other types of form-meaning mapping characterized by a more transparent and motivated link (Dingemanse et al. 2015). For example, iconicity, across languages, can be found in the phonology of words, e.g. in onomatopoeia such as *meow* or *drip*. This expressive richness is particularly prominent once we look at the multimodal communicative context in which language is learnt: prosodic modulations (e.g. prolonging a vowel to indicate prolonged extension, *loooong*), iconic, representational gestures (e.g., tracing an up and down movement with the index finger while talking about a bouncing object), points and hand actions with objects (e.g., showing a toy hammer to a child or showing how to use the toy hammer) also contribute to the meaning of the message. In vocabulary learning, these iconic and indexical communicative cues may scaffold the mapping between words and world (Perniss et al., 2010; Perniss & Vigliocco, 2014).

Such cues have been previously documented. Onomatopoeia are over-represented early on in children's language development, both in children's vocabularies (Laing, 2014) and in the input they receive (Perry et al., 2017), though this prevalence declines as children age. Points have been reported as the most common gestures used by caregivers especially with very young children (under the age of 2, Iverson et al. 1999; Özçalıskan & Goldin-Meadow 2005), helping to isolate the referent from a complex scene and to link it to the provided label. Though points are common, iconic gestures are also present in parental input from early on in child development (Rowe et al., 2008), and present in the gestural repertoires of children (Acredolo & Goodwyn, 1988). Furthermore, Rowe et al. (2008) showed that parents' gesture use (including points and iconic gestures) predicts children's gesture use, which in turn predicts later vocabulary development, suggesting the importance of such cues for overall language development. Lastly, research has shown a link between direct manipulation of objects (i.e., hand actions) in caregiver-child interaction and children's learning (see Rohlfing 2011 for a review). However, most previous studies focus on a single cue (e.g. gestures or hand actions), rather than considering how the different cues are used together (and together with speech). Cartmill et al. (2013) find that the quality of parental communication, operationalised as how predictable certain words are given the surrounding context (e.g. speech, gesture, surrounding

objects), predicts child vocabulary size at 54 months. Though this suggests that the multiplex nature of child-directed communication might scaffold language learning, they do not analyse the information provided by different cues. Moreover, most studies have focused on learning contexts where label and referent co-occur spatially and temporally (e.g., when objects are present in the visual scene, or words are uttered while actions are ongoing). However, displaced contexts (i.e., when objects are absent) can also provide learning opportunities and previous research indicates that children do learn in these contexts (e.g., Tomasello, Stroberg & Akhtar 1996).

Here, we provide a first investigation that comprehensively assesses the distribution of iconic and indexical cues both in learning contexts in which objects are present, and contexts in which they are absent. We expect to find that iconic cues (onomatopoeia and representational gestures) will be especially important in displaced contexts because iconicity can evoke perceptual or auditory features of the object, in this way providing an imagistic link with the referent and help in bringing it to the 'mind's eye. Both iconic and indexical (points and actions with objects) cues can single out referents, when present, in complex and messy visual scenes, and thus provide cues to solve the referential ambiguity problem.

We use a semi-naturalistic method in which we video-recorded caregivers interacting with their child talking about objects (provided by the experiments) which were either known and unknown to the child. We introduced this manipulation as cases in which the child is unfamiliar with the object and its label are more clearly learning episodes. Moreover, we manipulated whether the objects talked about are either present or absent. We focus on children aged 2 to 3 years old as this is a time of remarkable vocabulary growth in which all critical elements of child-directed language are present, communication about displaced referents is present, and finally, at which children are assumed to be able to understand and produce iconic gestures (Özcaliskan & Goldin-Meadow, 2005).

Method

Participants. Thirty-four caregiver-child dyads participated in the study. The language used between the caregiver and the child was British English. All children included in our sample were aged between 24 and 42 months.

Materials. We used toys from four categories: foods, musical instruments, animals and tools. We chose these categories because they are very common for children of this age and because they offer opportunities for vocal and manual iconicity. We created sets of 6 toys from each of the four categories, such that each set contained 3 toys known and 3 toys unknown to the child (based on parental reports). Toys were selected for each child from a larger set of about 20 toy

items per category, each of which were used for a roughly equal number of participants.

Procedure. Caregiver-child interactions took place in the families' homes. Before the session, caregivers were given a list of toy names from our full list and they were asked to indicate whether their child knew those objects and those words. They were also asked to fill in the Oxford Communicative Development Inventory (OCDI)[1]. During the session, two experimenters visited the family, and recorded interactions with two videocameras (one focusing on the caregiver, one focusing on the child and the interaction space). One experimenter checked the correct working of the videocameras while the other carried out the manipulations. The interactions were carried out at a table with the caregiver and the child sitting at 90 degrees from each other. Caregivers were asked to interact with their child in a natural way, as they usually did, but to try to talk about each of the objects provided. Drawings of the set of toys was given to the caregiver to help them remember which toys were in the set. The order of object present vs absent was counterbalanced across participants. When the interaction started with objects present, the experimenter brought to the table 6 toys from one category (e.g., animals) and left the room. The dyad talked about these toys for 3-5mins, then the experimenter re-entered the room, asked the child to help in tidying up the toys and then left the room for the displaced condition asking the caregiver and child to continue to talk (again for 3-5mins) about the toys they just played with. The experimenter then reappeared with a new set of toys until all toy categories had been used. When the toy absent condition came first, the caregiver was asked to begin talking about the toys that were about to come while she was going to get them from another room (caregivers were first familiarised with the toys). After 3-5mins, the experimenter brought in the set of toys, repeating this process for all four categories. The whole recording session lasted approximately 45-60 mins.

Coding of caregiver communication. The caregiver communicative behaviour was coded in the following manner.
(1) Speech. Data was transcribed by *utterance*, which is our unit of analysis (Berman & Slobin 1994). Lexical elements were transcribed further for *onomatopoeias* (including lexical onomatopoeia as well as sound effects) and for explicit mention of the referent (the toys in our sets) label. For each utterance, we coded the *topic*, as the specific toy (or multiple toys) that each utterance referred to, regardless whether labels were produced or not. Utterances were assigned to the known/unknown condition on the basis of their topic. Utterances not about the toy referents were coded as "other" for topic and were not included in any analysis.
(2) Points: gestures (using the index finger or the whole hand) that single out a referent by pointing to it*;*

---

(3) Iconic/representational gestures: gestures that represent referents by e.g. depicting aspects of their shape or manipulation.

(3) Hand actions: We coded hand actions and movements performed while holding or manipulating an object. These were divided into (i) *deictic* (i.e., showing) and (ii) *depicting* (e.g., demonstrating the use of a tool). Hand actions were only coded for the toys we provided, and thus can only occur in the toy present condition.

Thus, we distinguish in our coding between iconic cues (onomatopoeia and representational gestures) and indexical cues such as points. Another category is hand actions. Hand actions are indexical in that they direct attention to the referent. They can however, differ and we coded separately those that depicted some properties of the referent (depicting hand actions) from hand actions that showed the object to the child (deictic hand actions). We consider the distinction between iconic and indexical cues to reflect two different manners in which cues can be non-arbitrary: iconic cues *stand for* the object; indexical cues provide a visual link to the object but they do not stand for it. Fig. 1 shows screenshots of the different categories.

Figure 1. Examples of communicative behaviors coded in the different categories.

Results

Before looking at the distribution of the multimodal cues, we examined the distribution, across our four conditions, of caregiver utterances. Figure 2 illustrates how often parents talk about objects across conditions. Parents talk more when toys are present, and talk more about items unfamiliar to the child than those that are familiar. The larger number of utterances with objects present may indicate that it is easier to maintain the child's attention, or greater ease of production about present objects than about objects that need to be recalled.
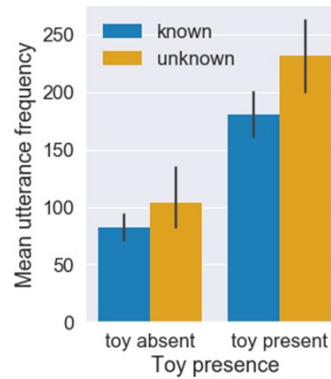

Figure 2. Mean utterance frequency

Iconic and indexical cue use. The primary aim of this study is to understand whether, and how, parents use onomatopoeia, representational gestures, points and hand actions (deictic and depicting) in their interactions with children. As such, we analysed whether age, familiarity (known vs. unknown object) and presence (present vs. absent) affect the use of each cue type.

Analyses use logistic mixed effects models to asses which factors affect the presence of absence of different cues. Age of the child (in months), presence or absence of the object, and familiarity of the label (known/unknown) were included as centered fixed effects, as well as their interaction, and the centered fixed effect of category (category is a control variable). We included a random intercept for participant with random slopes of presence/absence and label familiarity, plus their interaction. Dependent variables are presence/absence of each cue – referent label, onomatopoeia, representational gesture, point, hand action – in an utterance in a given condition). This model structure is used in all models throughout this section, unless otherwise specified. In the interest of space, only the effects of interest are reported here. Full results from the models can be found at https://osf.io/yegxh/.

First, we find that parents make use of all of these cues: approximately 39% of all utterances (11, 755 out of 30,283 utterances) in the dataset are modified by at least one iconic or indexical cue. Figure 3 shows the proportion of each cue across conditions in the study. Second, the proportion of points is low in comparison to the other, especially manual, cues. We attributed this to the affordances of the interaction context: toys were in close proximity to the caregiver and the child, therefore hand actions in this context can take the place of points (indeed, deictic hand actions represent 57% of all hand actions). Deictic and depicting hand actions are by definition only present when objects are present. These are more common for unknown than known objects and their frequency is not modulated by the children's age. *Points* are more common when objects are present, but we find no modulation regarding the familiarity of the label, or based on the age of the child.

Crucially, iconic cues that can be used across all our four conditions (onomatopoeia and representational gestures) show a clear effect of both toy presence and familiarity. In particular, for *onomatopoeia* we find that caregivers use them more often when toys are absent. We also find an interaction between familiarity and toy presence. When toys are present, caregivers use onomatopoeia more with known items. However, when toys are absent, we see the reverse, such that onomatopoeia occur more for unknown items. Interestingly, onomatopoeia decrease as the child's age increases; parents use fewer onomatopoeia with older children. For *representational gestures*, we see that these are overwhelmingly used when toys are absent and for unknown objects. No effect of age is observed. Table 1 summarises the model results.
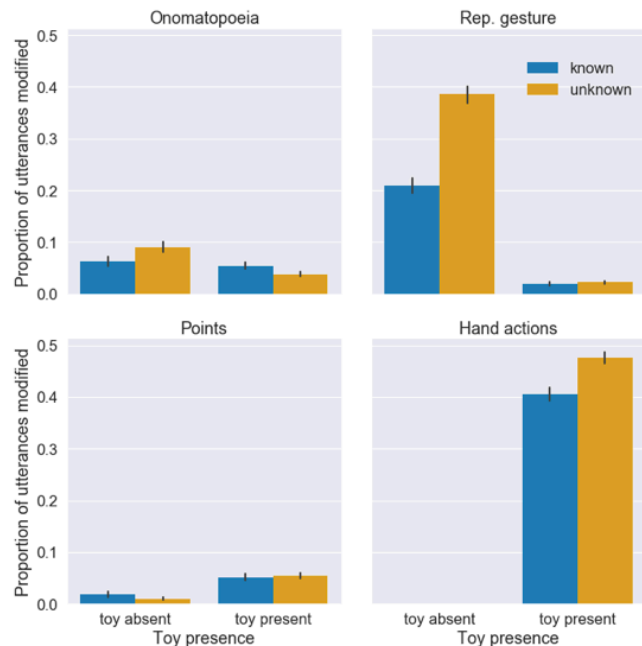
Figure 3. Proportion of onomatopoeia, representational gestures, and points across conditions (toy presence – situated vs displaced, and status – known vs. unknown). Hand actions can only be found in the situated condition. Error bars represent 95% confidence intervals.

Within utterances, cues can co-occur with other cues within and across modalities (e.g., in an utterance we may have a hand action and a representational gesture; or we may find an
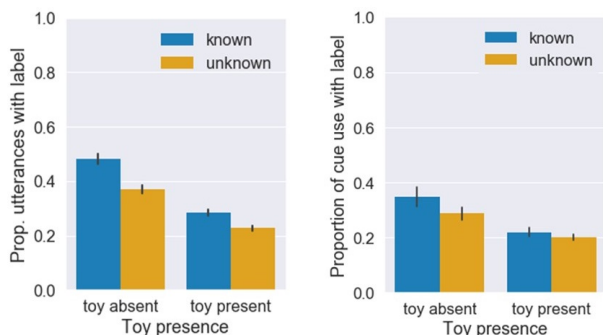
Figure 4. a) Mean proportion of label use, and **b)** Mean proportion of multimodal cues that co-occur with a label.

onomatopoeia and a hand action). Co-occurrences between cues (e.g., between onomatopoeia and hand actions) were remarkably rare in the dataset, occurring in only approximately 3% of parent utterances. Although we have a high proportion of cue modifications, we do not see a high proportion of cases where multiple cues co-occur.

Label use. Finally, we looked how often parents use explicit labels for the objects (e.g. saying the word 'cat'). We find that parents use explicit referent labels more when the objects are not present, and tend to use the label more for familiar objects than unfamiliar objects. Analysis of a model predicting referent use confirms this: a decrease in referent use for unknown labels, compared with known ones (β=-0.43, SE=0.06, z=-7.60, p<0.001), and a decrease in referent use in the toy present condition, compared to the toy absent (β=-0.90, SE=0.08, z=-11.22, p<0.001). We also address the question of whether and when any of our multimodal cues co-occur with explicit naming of referents. If cues specifically help to link label and referent, then we might expect that use of multimodal cues occur in close proximity to the referent.

Onomatopoeia	β	SE	z	p
Age	-0.05	0.02	-2.39	0.02
Label familiarity	-0.04	0.13	-0.33	0.74
Presence	-0.53	0.14	-3.77	<0.001
Familiarity*Pres.	-0.85	0.21	-4.10	<0.001
Points				
Age	-0.004	0.02	-0.19	0.85
Label familiarity	-0.14	015	-0.95	0.34
Presence	1.62	0.25	6.46	<0.001
Familiarity*Pres.	0.54	0.34	1.60	0.11
Gesture				
Age	0.006	0.03	0.18	0.86
Label familiarity	0.14	0.20	0.70	0.48
Presence	-3.42	0.19	-18.18	<0.001
Familiarity*Pres.	-0.86	0.33	-2.60	0.009
Hand actions				
Age	-0.02	0.02	-1.10	0.27
Label familiarity	0.25	0.06	4.14	<0.001
Hand action type (deictic-depicting)				
Age	0.03	0.03	0.86	0.39
Label familiarity	0.07	0.14	0.45	0.65

Table 1. Summary of model results from logistic mixed effects models. Output variable given in bold. Note that the model for hand action does not include toy presence, as hand actions are not possible in cases where toys are absent.

All of the cues we coded for can co-occur with explicit labelling of the referent (e.g., naming the referent while producing a hand action or representational gesture; naming the referent in the same utterance in which an onomatopoeia is produced). We found that, overall, the multimodal cues occur with explicit naming of a referent approximately 35% of the time. Figure 4 illustrates referent-cue co-occurrence across conditions. We subsetted rows in the dataset where

any of our four cues were produced, to analyse how co-occurrence between cues and labels differed across conditions. The model revealed both an effect of familiarity (β=-0.32, SE=0.08, z=-4.30, p<0.001), and of toy presence (β=-0.58, SE=0.12, z=-4.98, p<0.001). Label-cue co-occurrence occurs more when toys are absent, and when the label is known to the child.

Discussion

The work reported here aimed to characterize the distribution of iconic and indexical cues in the input to 2-3 year-old children. We see that approximately 40% of the clauses produced by caregivers contains at least one of these multimodal cues which often co-occur with explicit labelling of objects especially when toys were present and unknown to the child.

Iconicity as a bridge between words and world
One main goal was to establish if multimodal cues are differentially distributed across contexts: (i) whether the child knows the object and its label, and (ii) whether the objects being talked about are present in the communicative context vs. absent. The latter manipulation has been introduced in order to assess the extent to which the multimodal communicative strategy of caregivers is responsive to the presence vs. absence of object, and whether they modify their language based on the physical setting in which the communication takes place.

The hypotheses we have test is one in which non-arbitrary cues in learning provide a stepping stone to the child to bridge between words and world (Perniss & Vigliocco, 2014). Both iconic and indexical cues can single our referents when these are present in the environment. Moreover, iconic cues can be used when the objects are absent to bring to the mind's eye properties of referents. Thus, indexical and iconic cues may play an important role in learning. We see that this is the case in our data. With the exception of points, which are equally likely for familiar and unfamiliar objects (and labels), all other cues are more commonly used for unknown objects (learning contexts). Crucially, the iconic cues (onomatopoeia and representational gestures) are also used more often when the objects are not present in the physical environment. The results for iconic cues are in line with previous work using a similar paradigm, where it was found that deaf caregivers modify iconic signs in British Sign Language (BSL) to highlight iconic properties of signs (e.g., enlarging the up-and-down movement path of the arm in the sign HAMMER) far more often when objects where absent than present (Perniss et al., 2017), suggesting that this tendency holds across language modalities. In contexts where the label is known to the child, we see that while representational gestures are still overwhelmingly most common for displaced contexts, this is not the case for onomatopoeia. Another interesting difference between onomatopoeia and representational gestures (as well as all other cues) is that

onomatopoeia show a decrease with age, in the age range we considered (24-42 months). This finding is in line with previous work showing that use of onomatopoeia in caregivers and children's speech decreases from 0.8 to 2 years (Kauschke & Hofmeister, 2002; Kauschke & Klann-Delius, 2007; Laing, 2014). It has been suggested that for spoken languages, iconicity embedded in wordforms as onomatopoeia may act as a bootstrapping mechanism, a sort of protolanguage, guiding infants' attention to the fact that what comes out of the mouth is linked to what happens in the world (see Imai & Kita, 2014; Laing, 2014).

Iconic vs Indexical cues
We have distinguished points from representational gestures: points don't stand for an object - like representational gestures do - but they direct attention to the object via direct deixis. Just like representational gestures, they are non-arbitrary. Although, in principle iconic cues could be found both when objects are present as well when they are absent, they are far more common in displaced contexts (note however that we observe depicting hand actions in situated contexts); points also, in principle could be found in both contexts, but they are overwhelmingly more common in situated contexts (and deictic hand actions can only be present in situated contexts). This is in line with what was observed in a previous study in BSL where pointing was also much more common in situated than displaced contexts, though the use of indexing to abstract locations in space is common in signed language (Perniss et al., 2017). In the introduction, we mentioned two ways in which non-arbitrary cues can support language development. First, they can help singling out referents when these are present. Both iconic and indexical cues can do this, however, indexical cues may be better placed, as they can be used from earlier age and they provide an unambiguous visual link to the referent. Second, they can help evoking - via imagery - properties of referents that are not present. Iconic cues are best suited to support this type of learning scenario. Note that our study might have called for the use of iconic cues also linked to learning about properties of novel objects. When children were presented with unfamiliar objects, caregivers also used iconicity (especially depicting hand actions) to show the child how the object is used, or how it moves.

How are cues orchestrated?
Previous work suggests that some cues co-occur. For example, Laing et al. (2017) showed that onomatopoeia are usually prosodically marked and Kita (1997) reports that representational gestures tend to co-occur with onomatopoeic and other sound-symbolic words in Japanese. However, we did not observe any tendency for cues to co-occur (although we did not code for prosody in our dataset at this point and therefore we acknowledge that things might be different when considering prosody). For the cues we have considered, it is clear that caregivers choose one cue, presumably on the basis of affordances of the objects (e.g., onomatopoeia for toy animals, representational gesture for tools) to associate to

each utterance. We found that explicit label productions are more likely to co-occur with a multimodal cue when objects are absent and when objects are known to the child. Precisely why this may be the case is unclear; given that labels themselves appear more frequently when toys are absent, it may be that parents use the label in conjunction with cues when toys are absent to make reference to a given object more salient. When the toys are present, it is possible to interact with or point to the toys, making direct reference less necessary.

Conclusions

This study provides a first snapshot of the distribution of multimodal cues in child-directed language. We found a clear indication that iconic as well as indexical cues are well represented in caregivers' input and crucially, they are especially used in those contexts where they may be most useful to children: namely in learning contexts, where the objects and labels talked about are unfamiliar to the child and when the learning occurs in displaced contexts where the objects are not available. It is important to note that the work reported here only provides a partial picture, however. First, the interactions in this study are focussed on contexts of play, which may not be representative of other interactional contexts. Secondly, missing from the current picture is prosodic modulation, which is a key feature of child-directed speech (e.g., Fernald & Simon, 1984; Fernald, 1989; Fernand et al., 1989) and which has been shown to be associated to onomatopoeia (Laing, 2017). Finally, and most important, is the fact that the present work focuses on the communication by the caregiver only, without considering the child's communication, thus giving the impression that the child is a passive receiver of input from caregivers. There is clear evidence this is not the case (e.g. Pereira, Smith & Yu 2014), however, while we plan to code the children's productions, we nonetheless believe that considering the distribution of multimodal cues in caregivers' communication can already provide insight into important questions that has received little attention so far such as which and how cues are used in displaced contexts.

References

Acredolo, L., & Goodwyn, S. (1988). Symbolic Gesturing in Normal Infants. *Child Development*, *59*(2), 450–466.

Bates, D., Bolker, B., Machler, M., & Walker, S. C. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Cartmill, E. a, Armstrong, B. F., Gleitman, L. R., Goldin Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(28), 11278–83.

de Saussure, (1916) *Course in general linguistics*. New York, NY: McGraw-Hill

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity and systematicity in language. *Trends in Cognitive Sciences*, *19*(10), 603–615.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language* 16, 477–501.

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? Child Development, 60, 1497-510.

Fernald, A. & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. Developmental Psychology, 20, 104-113.

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Phil. Trans. R. Soc. B* 369, 20130298.

Iverson, J.M., Capirci, O., Longobardi, E., & Caselli, M. (1999). Gesturing in mother–child interactions. *Cogn. Dev.* 14, 57–75.

Kauschke, C., & Hofmeister, C. (2002). Early lexical development in German: A study on vocabulary growth and vocabulary composition during the second and third year of life. *Journal of Child Language* 29, 735–757.

Kauschke, C. & Klann-Delius, G. (2007). Characteristics of maternal input in relation to vocabulary development in children learning German. In Gülzow, Insa and Gagarina, Natalia (eds.), *Frequency Effects in Language Acquisition. Defining the Limits of Frequency as an Explanatory Concept*, 181-204. Berlin, New York: De Gruyter Mouton.

Kita, S. (1997). Two-dimensional semantic analysis of Japanese mimetics. Linguistics, 35, 379-415.

Laing, C.E. (2014). A phonological analysis of onomatopoeia in early word production. *First Language* 34(5), 387-405.

Laing, C.E., Viham, M. & Portnoy, T.K. (2017). How salient are onomatopoeia in the early input? A prosodic analysis of infant-directed speech. *J. Child Lang.*, *44*, 1117-1139.

Özçalışkan, Ş. & Goldin-Meadow, S. (2005). Do parents lead their children by the hand? *Journal of Child Language* 32(3), 481-505.

Özçalışkan, Ş, Levine, S.C. & Goldin-Meadow S (2013). Gesturing with and injured brain: how gesture helps children with early brain injury learn linguistic constructions. *Journal of Child Language, 40*: 69-105.

Pereira, A.F., Smith, L.B. & Yu. C. (2014). A bottom-up view of toddler word learning. *Psychon Bull Rev, 21*, 178-185.

Perniss, P., Lu, J.C., Morgan, G. & Vigliocco, G. (2017). Mapping language to the world: The role of iconicity in the sign language input. *Developmental Science* DOI: 10.1111/desc.12551.

Perniss P, Thompson T, Vigliocco G. (2010). Iconicity as a

general property of language: evidence from spoken and signed languages. *Front. Psychol*. 1, 1–15.

Perniss, P., & Vigliocco, G. (2014). The bridge of iconicity: from a world of experience to the experience of language. *Phil. Trans. R. Soc. B* 369, 20130300.

Perry, L. K., Perlman, M., Winter, B., Massaro, D. W., & Lupyan, G. (2017). Iconicity in the speech of children and adults. *Developmental Science, 21(3)*, e12572.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rohlfing, K. J. (2011). Meaning in the objects. *Experimental Pragmatics/Semantics*, 151–176, Amsterdam: John Benjamins.

Rowe, M. L., Özçalışkan, Ş., & Goldin-Meadow, S. (2008). Learning words by hand: Gesture's role in predicting vocabulary development. *First Language, 28*(2), 182–199.

Tomasello, M. & Kruger, A. (1992). Acquiring verbs in ostensive and non-ostensive contexts. *Journal of Child Language* 19, 311—33.

Tomasello, M. & Barton, M. (1994). Learning words in non-ostensive contexts. *Developmental Psychology* 30, 639-50

Tomasello, M., Strosberg, R., & Akhtar, N. (1996). Eighteen-month-old children learn words in non-ostensive contexts. *Journal of Child Language* 23, 157-176.

Yu, C. & Smith, L. (2012). Embodied attention and word learning by toddlers. Cognition 125, 244-262.