Acquiring relational meaning from the situational context. What linguists can learn from analyzing videotaped interaction.

Barend Beekhuizen, Rens Bod, and Arie Verhagen[*]

1. The difficulties of acquiring relational meaning

The acquisition of linguistic expressions of relational[1] meaning is typically regarded as a difficult task (Gentner 1978, Gleitman 1990). What hampers this process, as compared to the acquisition of expressions of object reference, is not so much the complexity of relational concepts per se, which children seem to understand well before their first birthdays for many types of relational concepts (Mandler 2006, Casasola, Bhagwat, and Ferguson 2006, Sootman-Buresh, Woordward, and Brune 2006), but the question which subset of the large set of relations perceivable in situational context should be mapped to a linguistic expression. Furthermore, as especially Gleitman (1990) argues, not all relations expressed are directly perceivable, as some pertain to mental states of intentional agents and others are simply not present in the here-and-now of the speech situa- tion. As Gleitman succinctly puts it (p. 5): “[T]here is *not enough* information in the whole world to learn the meaning of even simple verbs, or [. . .] there is *too much* information in the world to learn the meaning of [. . .] verbs.”

In this paper, we evaluate Gleitman’s double problem. Is the context indeed too poor or too rich for the learner to find situational, contextual correlates for the linguistic items he perceives that can help him bootstrap the meaning of those items? We address this question for a broad class of word types, both relational (verbs and prepositions) and non-relational (nouns and adjectives) ones.[2] We do so by using a corpus of child-directed language that contains precise information on the behavior of the participants and several states in the world. This informa- tion is derived from annotations of the behavior of the caregiver and child, based on videotaped interaction in the setting of a simple game.

Note that our focus is mainly on the situation rather than on cognition. Work by Gentner and colleagues on relational categories (see e.g. Gentner and Kurtz 2005) has provided us with much insight in the cognitive mechanisms at work when acquiring relational meanings of words. This paper looks at the other side of the same coin: what information is available in naturalistic situations on which these mechanisms operate. Is there any relation between two objects that could be construed as 'containment' present in the environment in which the language-learning child operates when the word *in* is uttered?

We believe the development and use of methods like these, in combination with computational modeling techniques, is paramount for the development of a more comprehensive usage-based theory. If we assume that the interaction be- tween caregivers and children is the locus of the development of the acquisition of things like construal and symbolic mappings, we must at least have a decent understanding of what happens in those situations. Without analyzing actual be- havioral data, we will not know what the information available is, and any form of theorizing about the mechanisms used in the acquisition of symbolic mappings can effectively not be evaluated, as the interpretive step from a controlled experi- ment tapping into the hypothesized mechanisms and the situation 'in the wild' is unwarranted. It is exactly this understanding that is currently too weakly present in acquisitionist research.

This precise method has, to the best of our knowledge, not been used before in language acquisition research, but similar work has been done on object reference using video data in which the present objects and social cues such as eye gaze and pointing were coded (Frank, Goodman , and Tenenbaum 2009, Frank, Tenenbaum, and Fernald 2013). Because of this, we discuss the considerations in developing the data in some detail first (section 2). In section 3 we then address the question of the usefulness of the context for acquiring relational meaning by looking at a quantitative measure of association between linguistic elements and aspects of the situation in which these elements are produced. We limit ourselves to verbs, and other relational elements that encode conceptualizations of physical relations and actions.[3]

2. Developing a corpus of contextualized language

If we want to understand how informative the situations are for the language-learning child, we need to have material that captures the situational context of every utterance in detail and that is situated in a relatively ecologically valid setting. Regular corpora contain little information on the context, and deriving information about the context from the language in order to find associations between the two, which is often done in computational modeling studies of word-meaning acquisition, leads to circularity.

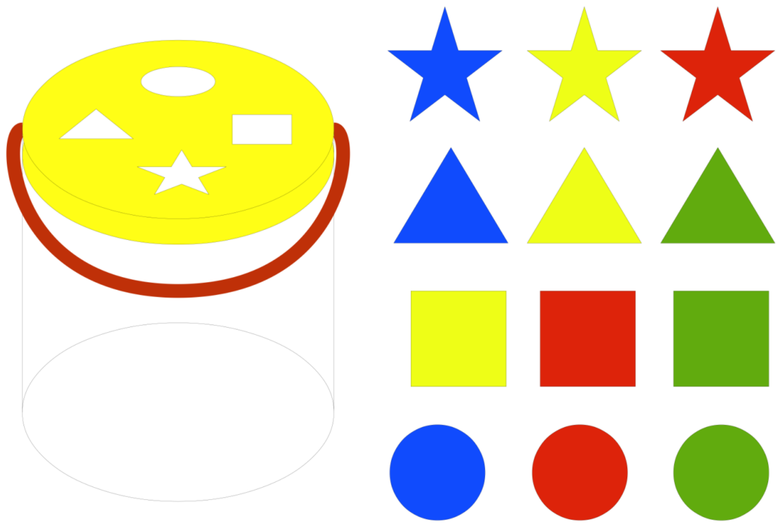
Because of this lack of data, we decided to develop a dataset on the basis of videotaped interactions in Dutch between children and their mothers, which was recorded for other purposes. The child-directed language was transcribed by the first author, and the situation was described according to a formalized coding scheme by two assistants. In the following sections, we discuss the choices and the justification for them, as well as the coding procedure and evaluation in some detail, so as to provide other researchers with a starting point for doing similar research.

2.1. The nature of the underlying data

Our material consists of 32 videotaped fragments, each containing a unique mother-daughter dyad playing a full game of putting blocks in holes of a toy. The mothers are all middle-class, native speakers of Dutch who were living in or close to the Dutch town of Leiden at the time. The fragments are on average 4 minutes and 54 seconds long, ranging from 3 minutes and 12 seconds to 7 minutes and 5 seconds. The game starts with the observer giving the toy to the mother, and ends with the mother returning the toy to the observer. The observer does not take part in the interaction, so the mother and the daughter are jointly focussed on playing the game.

The game-related objects are displayed in figure 1. The game typically takes place on the table or on the floor, and there are few non-game objects that are manipulated during the game. A typical game consists of the mother or daughter opening the bucket and getting all blocks out. Then the daughter tries to fit each of them back in, sometimes with the help of the mother, but always with verbal comments, suggestions and questions by the mother. Several children lose interest after a while and wander off, while others are so excited about their success that the game is played two or three times during the fragment.

Figure 1. The toy and the twelve blocks



All children were around sixteen months at the time of the recording. This means that their ages are within a desirable time window for this type of observational study. Before the age of one, we cannot expect children to understand the symbolic nature of language. When children's early intentional understanding enters the picture, they are able to start learning linguistic symbols (Tomasello 2003, 21-28). However, as we get closer to the child's second birthday, chances increase of the child being able to talk itself, in which case the child is often directing the conversation by means of single-word utterances triggering responses from the caregiver. The age of sixteen months squarely falls within the period in which children understand the symbolic, communicative nature of language, but are just beginning to connect word forms and meanings. Because of this, we can make the assumption that this is actually the kind of situated child-directed language input that children receive.

2.2. Developing a coding scheme

What aspects of the situational context are worth taking into account? The participants perform various actions related to the game, and hence the spatial states of the game objects change continuously. The most straightforward aspects of the scenes we could describe thus are changes in spatial relations between objects and the behavior of participants leading to these changes. We found that with nine predicates, we were able to cover almost all of the participants' behavior in the fragments. The top nine predicates, given in table 1 and described as English verbs, reflect the building blocks of object manipulation: there are descriptions of bringing an object under manual control (**reach** and **grab**), and of letting go of this control (**let_go**). In between, one can do all sorts of things with objects: **move** them from one location to another, position them viz a viz a location, exert **force** upon them and letting them go (as in pushing, throwing) and showing them to another participant. Furthermore, we can **point** to non-held objects. Situations that do not fit into this set of categories are assigned the label **other**, and in case of doubt between two existing codes, the label **unclear** could be used. Finally, when one of the participants' behavior is not visible, because that participant is outside of the camera frame or behind some occluding object, the predicate **out-of-view** is assigned. For a more detailed description of these predicates, we point to the coding manual.[4]

Each of these predicates dictates a number of roles together with which they form a semantic predicate-argument structure. The predicate **grab**, meaning to bring something under one's manual control, has a grabber and a grabbed object, and possibly an instrument of grabbing other than the hands (a spoon, the mouth), **move** has a mover, a moved object and a source and goal location. The second column of table 1 describes the valency of the predicates. As to the fillers of these roles, all roles marked with subscript *O* require objects to fill them, whereas the *S* subscripts describe a role to be filled with a spatial predicate (see below). The objects either come from a closed class of descriptions of game-related objects (table 2) or are assigned freely but consequently by the coder, in the case of objects that are not part of this closed set.

Table 1. Behavioral predicates and their roles

predicate	*arguments*
reach	(theme$_{O/S}$, [instrument$_{O}$])
grab	(theme$_{O}$, [instrument$_{O}$])
show	(theme$_{O}$, recipient$_{O}$, [instrument$_{O}$])
position	(theme$_{O}$, location$_{S}$, [instrument$_{O}$])
move	(theme$_{O}$, source$_{S}$, goal$_{S}$, [instrument$_{S}$])
hit	(theme$_{O}$, [instrument$_{O}$])
let_go	(theme$_{O}$, [instrument$_{O}$])
force	(theme$_{O}$, source$_{S}$, goal$_{O/S}$ [instrument$_{O}$])

point	($theme_O$, $recipient_O$, [$instrument_O$])
other	n.a.
unclear	n.a.
out-of-view	n.a.

Table 2. Object labels for game-related arguments. Note that the abbreviations **ci**, **sq**, **st** *and* **tr** *stand for circular, square, star-shaped and triangular respectively, and* **bl**, **re**, **gr** *and* **ye** *for blue, red, green and yellow respectively.*

predicate	*arguments*
to	toy
bu	bucket
ha	handle
li	lid
ho(ci|sq|st|tr)$_1$	hole in the lid with a certain $shape_1$
b(bl|re|gr|ye)$_1$(ci|sq|st|tr)$_2$	block with a certain $color_1$ and $shape_2$
mo	mother
ch	child
ob	observer
table	table
floor	floor
air	air

Independent spatial predicates are used in the case of some arguments (sources, goals, locations) and in order to describe salient events of the game. The spatial relations are given as English prepositions, but, again, rely upon descriptions that the coders used when coding the states. The four relations that are coded, are **in** for containment, **on** for horizontal support, **at** for all other forms of physical contact between two objects and **near** for all forms of non-contact but a salient proximity between two objects. Two pairs of game-related objects are independently described because of their relation to the success of the participants. The first is the lid being on or not-on (**off**) the bucket, and the second that of each of the blocks being in or not-in (**out**) the bucket. These relations are only coded if there is a change in situation.

When developing a set of predicates and objects like this, two important questions arise. First: does the handcrafted set have the right degree of granularity? Second, and perhaps more important: does the set reflect perceivable relations and not leave out perceivable relations. Regarding the first question, the set of primitives was developed so as to be as specific as possible, while still being codable by an instructed coder. In evaluating the coding procedure, there are some demarcation issues between predicates that can be solved by decomposing them further, but doing so would involve methods beyond coding video data to obtain the information.

As to the question whether the coding scheme is not too specific: or aim was to code primarily the physical behavior, and as little intention as possible. Actions at a larger scale, such as 'getting something out of something' often consist of parts that are simple predicates like **grab** and **move**, but are as a whole intentional, goal-directed actions. Inferring these from the perceived simple action schemas is a future direction we are currently looking into (see also section 2.5).

Turning to the perceptual availability of the predicates, we believe the behavioral predicates reflect very basic, primitive sensory-motor experiences that are available from at least 12 months of age (Meltzoff and Moore 1995). These may then be grouped into culture-specific gestalts, combining properties of the shape of the manipulated object, the presence and nature of instruments etc. (see for an approach to the way languages encode these types of features Narasimhan and Kopecka (2012)), but as such, we expect the basic building blocks of these analytically complex concepts to be perceptually available independent of culture.

The same holds, mutatis mutandis, for the spatial relations and properties and category labels of objects. Especially in the case of the former, the work by Baillargeon and colleagues (Baillargeon and Wang 2002) shows that notions like 'containment', 'nearness' and 'physical contact' have developed well before the child's first birthday. The predicate **on**, denoting horizontal support, is arguably more problematic. Choi (2006) presents evidence that children understand support relations later than containment relations, a fact that she ascribes to the many ways in which there can be a support relation, such that a common conceptual core is harder to abstract. Also, some potential perceptual sources of spatial relations that are encoded in different languages, such as pointwise vs. surface contact, tight fit vs. loose fit, are not coded in the data. Again, decomposing spatial relations into more primitive notions of relations (e.g., contact, surroundedness of figure by ground on both the horizontal and vertical axes, similarity in shape) combined with properties of the figure and the ground (mass-like or discrete, animacy) can provide a promising avenue (see e.g. Feist 2000 for such an approach), but coding observational data with such fine-grained distinctions (as opposed to setting up experimental situations) might prove too difficult for coders. Summarizing: The coded predicates are all available, but might not reflect all perceptually available relations. Nevertheless, we believe that we can still address the question whether the context is not too rich, as we code perhaps not all, but at least many predicates that function as noise for learning a word.

2.3. The coding procedure

The predicate-argument structures described above were coded using ELAN, a piece of software for video annotation.[5]Regular intervals of three seconds were created, so that the coders had to describe the events that took place within that window of time. Predicates taking place in two subsequent windows were only coded in the window in which they started. The behavior of the mother and the child was coded on different tiers, as were the spatial states of the game-objects.

The coders, a second-year and a fourth-year student at Leiden University, were paid as research assistants and worked in total 56 and 64 hours respectively on the project and were supervised by the first author. They received six hours of training. The first two hours were spent on an in-depth explanation of the predicates, arguments and objects, after which the coders and the first author jointly coded a fragment in four hours. After this, the coders reported to feel confident about the task. Apart from the training, they received a manual with a reference sheet to be used during the coding itself. This document contained detailed descriptions about the predicates, decision trees for anticipated demarcation problems and a general instruction about the workflow using ELAN. The coders and the supervisor had contact at every working day about unclear cases and more general issues.

2.4. Coding the behavioral data: insights and figures

The coders worked for 56 and 64 hours on coding fragments. In this time, they coded 23 and 19 fragments, respectively. Each coder did one fragment twice and there was an overlap of three fragments between the coders. Five fragments were discarded due to a low visibility of the actions (three cases) or children who failed to play the game for the majority of the time (two cases). On average, coding a fragment took the coders two hours and thirty-four minutes, with barely any difference between the coders (2 hours and 31 minutes, and 2 hours and 38 minutes respectively). This resulted in 175 minutes of coded material, out of which 157 minutes were useful, unique fragments. The average rate of coding thus was one minute of video per approximately 37 minutes of coding time.

The coders reported few problems with the procedure. Recurrent difficulties were the difference between **force** (i.e. moving without grabbing) and **hit**, which indeed is a distinction along a continuum, viz. that of the duration of the impact, which is more pointwise in hitting and more durative in forcing. This can be overcome if the dimensions of the manual contact are split out: one has to decide upon a cut-off point for the length of the impact to demarcate **force** from **hit**.

Also, the difference between **position** and **move** was felt to be unclear at times. Again, these predicates seem to form a continuum, with the amount of motion with respect to the ground being low for the former and high for the latter predicate. Truly distinguishing them would require the interpretation of intentions of the agents: one positions something in order to establish a different relation between the undergoing object and the location, whereas one moves something to change the object's location from a source to a goal location.

In order to evaluate the reliability of the coding, we can calculate the inter- and intracoder agreement. Three games were coded by both coders and each coder did one fragment twice. We measure the agreement by checking, for each predicate and for each time frame of three seconds whether at least one instance of that predicate was coded. A good measure for agreement is Cohen's kappa (Cohen 1960), which looks at the extent to which two codings of the same situation deviate (either two codings of one situation by the same coder or two codings of one situation by two different coders). A κ of 1 means perfect agreement, and a κ of 0 complete disagreement. This measure is often used in studies using human coders (e.g., Carletta 1996). Table 3 lists the Cohen's kappa scores for each predicate. A $\kappa \geq 0.80$ is generally taken to be the standard for very reliable coding (though often lower scores, above $\kappa \geq 0.60$ are taken to be acceptable as well).

We can see that several predicates are coded very reliably: **show**, **move**, **let_go**, **point** and **grab**. The figures for **hit** are reasonably good, but the predicate is very infrequent, and all scores of 1.00 for this predicate are due to single predicates. **Reach** was coded unreliable across coders, but reliably within. This was due to the fact that one coder regarded them to be 'implied' by other predicates (**grab**, **hit** and sometimes **force**) and only coded **reach** when these other predicates did not apply. **Position**, finally, was coded variably between coders and within coders. This may have to do with the fact that, as the coders reported, position and move were found difficult to discriminate. How to create a more transparent coding scheme for the motion-positioning continuum remains to be seen. For now, we will retain the predicate, but in future work, this action will need to be demarcated clearer for the coders. The space predicates, finally, were coded reliably overall, except for two low scores, one for **out** and one for **in**. In the first case, one coder coded **out** in the time window after the one in which the other coder coded it, so effectively, they agree upon the predicate, but placed in a different time frame. Part of the explanation of the low score for **in** is the disagreement on positioning: one coder used more **position** labels (as opposed to non-actions) and in many cases the positioning took place in the air.

Table 3. Cohen's kappa per fragment and coder

		Intercoder argeement			Intracoder agreement	
predicate	agent	fragment 1	fragment 2	fragment 3	coder 1	coder 2
reach	mother	0.00	0.29	0.50	0.80	0.82
	daughter	0.57	0.35	0.75	0.66	0.78
show	mother	1.00	1.00	0.91	0.91	1.00
	daughter	0.80	1.00	0.91	0.91	1.00
move	mother	0.77	0.92	0.85	0.81	0.92
	daughter	0.79	0.93	0.86	0.85	0.91
position	mother	0.70	0.00	0.69	0.80	0.00
	daughter	0.59	0.38	0.69	0.67	0.39
hit	mother	0.00	1.00	1.00	1.00	1.00
	daughter	0.66	1.00	1.00	0.75	0.00
let_go	mother	0.96	0.84	0.89	0.90	0.88
	daughter	0.85	0.83	0.79	0.92	0.84
force	mother	1.00	1.00	1.00	1.00	1.00
	daughter	0.00	0.75	1.00	0.27	0.00
point	mother	0.56	0.89	1.00	0.89	0.89
	daughter	0.59	0.89	1.00	0.90	0.89
grab	mother	1.00	0.90	0.87	0.81	0.83
	daughter	0.85	0.82	0.85	0.85	0.87
out		0.00	1.00	1.00	1.00	0.79
off		1.00	0.83	1.00	1.00	1.00
in		0.49	0.92	1.00	0.91	0.87
on		0.83	0.91	0.49	1.00	1.00

2.5. Deriving information

The codings resulting from this process were paired with all utterances that take place within each time window, to give us a corpus of child-directed language enriched with detailed situational information. Table 4 gives us a sample of this corpus.[6]

Table 4. A sample of the coding and transcribed language paired

time	*type*	*action/utterance*
0’00	**situation**	hold(mo,li)
	language	een. nou jij een.
	gloss	one. now you one.
0’03	**situation**	position(mo,to,on-floor) grab(ch,byetr) move(ch,byetr,on- floor,near-hoci)
	language	nee daar.
	gloss	no there.
0’06	**situation**	point(mo,hotr,ch) hold(mo,li) position(ch,byetr,near-hoci)
	language	nee lieverd hier past ie niet.
	gloss	no sweetheart here fits he not.
0’09	**situation**	point(mo,hotr,ch) letgo(mo,lid) move(mo,byetr,near-hoci,near-hotr) grab(ch,bblst) move(ch,bblst,on-floor,in-air) grab(mo,byetr) letgo(ch,byetr)
	language	hier in. kijk e(en)s. een twee.
	gloss	here in. look once. one two.

Given the fine-grained descriptions of the behavior of the participants, it is possible to derive more information about the situation without having to code it manually. An important aspect of understanding a situation is understanding the intentions and goals of the participants. Although intentions are not directly perceivable, they can be inferred from the actions the participants partake in (cf. Fleischman and Roy 2005). It is likely that a person grabbing a block has some goals involving that block. Suppose we assume that the participants understand the global outline of the game, we could say that one participant’s grabbing of a block can make the other participant infer that the first participant has (or: should have) the goal of putting that block in the bucket through the hole. We are cur- rently exploring this direction further.

On a lower, less intentional level, it is interesting to know whether there is a match in shape or not between a block and a hole when a participant moves and positions a block next to a hole. Furthermore, the blocks can be split out in such a way that **grab**(**mother**,**blue-triangular-block**) is given as **grab**(e_i,e_j), **blue**(e_j), **triangular**(e_j), **block**(e_j), **mother**(e_i).

2.6. The language

The child-directed language used in the 157 minutes of useable material consists of 7842 word tokens in 2492 utterances, on average 3.14 words per utterance. Many utterances consist of merely *ja.* 'yes' or *goed zo!* 'well done!'. The word tokens contain 480 types and 355 word lemmas (i.e., lemmas without inflectional and diminutive morphological marking).[7] Table 5 gives an overview of lemma statistics split over some categories. Nouns and adjectives are lumped, because many adjectives are used nominally. We can see many game-related lemmas in the top of the frequency distributions of the verbs, nouns and adjectives, and prepositions and spatial particles.

Table 5. Lemma frequencies split out over part of speech type and ranked. The gloss PRT is given when the word primarily function as a discourse particle for which no translation equivalent exists in English. (red.) means that this is a conventional, reduced form that is morphosyntactically distinct from the full form. The bottom row gives the number of lemmas and tokens of that part of speech type.

verbs		*nouns/adjectives*		*prepositions/ spatial particles*		*other*	
lemma	*n*	lemma	*n*	lemma	*n*	lemma	*n*
doen `do'	260	goed `good'	197	in `in'	287	ja `yes'	361
gaan `go'	205	mooi `beautiful'	58	op `on'	54	die `that'	306
zijn `be'	185	ander `other'	49	uit `out'	43	zo `like that'	288
kijken `look'	178	moeilijk `difficult'	46	met `with'	25	maar `but'	259
moeten `should'	127	mama `mom'	36	af `off'	15	d'r `there (red.)'	191
proberen `try'	70	keer `time'	26	door `through'	15	hier `here'	168
komen `come'	60	deksel `lid'	25	van `of'	10	eens `PRT'	162
kunnen `can'	51	blok `block'	24	voor `for'	10	nee `no'	158
pakken `grab'	40	rond `round'	23	aan `on,at'	9	een `a'	147
passen `fit'	38	leuk `nice'	19	naar `towards'	8	je `you (red.)'	143

zullen `shall'	35	ster `star'	17	bij `at'	6	hij `he'	138
zitten `sit'	33	makkelijk `easy'	15	om `in order to'	4	dat `that'	132
halen `get'	32	puzzel `puzzle'	13	na `after'	3	he `huh'	115
hebben `have'	30	gat `hole'	11	neer `down'	5	nog `still'	112
maken `make'	28	spel `game'	11	heen `-ither'	5	niet `not'	106
draaien `turn'	26	stuk `piece'	10	naast `next to'	2	we `we (red.)'	105
stoppen `stick'	18	vierkant `square'	9	achter `behind'	1	nou `PRT'	98
zien `see'	13	emmer `bucket'	8	boven `above'	1	\'e\'en `one'	96
weten `know'	12	meid `girl'	8	buiten `outside'	1	daar `there'	86
willen `want'	12	groen `green'	7	dichtbij `close'	1	deze `this'	85
74 lemmas 1640 tokens		103 lemmas 765 tokens		20 lemmas 505 tokens		158 lemmas 4932 tokens	

3. A first exploration

3.1. Data

To see if the situational context can be informative for the child trying to assign a meaning to a word, we investigate the association between coded aspects of the situation and certain lemmas. If we can find a strong association between certain lemmas and aspects of the situation related to their meaning, we can establish that the situational context is rich enough and not too rich to learn the meaning, or at least point the child in the right direction.

The starting point is a corpus of pairs of utterances (*U*) and situations (*S*), based on the 157 minutes of coded and transcribed interactional data discussed in the previous section. An utterance consists of the string of all lemmas $w_1 \ldots w_n$, and the situations are reresented as a set of features $f_1 \ldots f_n$ that are present within the three-second time window in which the utterance takes place. So, every datapoint *d* is a pair *U*,*S*, where $U = w_1 \dots w_n$ and $S = \{f_1 \dots f_n\}$.

For this exploration, we simplified the representation of the situation. Whereas we coded the information as predicate-argument structures, we discarded all arguments and used the predicates as atomic features. The predicate-argument structure **grab**(**mother**,**bucket**) and **on**(**lid**,**bucket**) thus become grab and on respectively. Apart from the behavioral predicates and spatial predicates that take place at a certain point in time, the following features are also included:

- properties of blocks and holes present among the arguments of the predicates coded at that moment in time (e.g., **red**, **star**),
- category labels for all objects that are arguments of those predicates (e.g. **mother**, **block**, **table**),
- two derived predicates: **match** and **mismatch**, that apply if there is a coded spatial relation between a block and a hole in the time window and there is either a match in the shape of both or a lack thereof, respectively.

In total, this procedure gives us 89 feature types, with an average of 12.2 features per situation.

To give a fuller example, consider the utterance in (1), which is paired with the situation in (2):

(1) *allemaal* *even* *d'r* *uit* *halen*
all PARTICLE there out remove.INFINITIVE
'Let's get them all out.'

(2) reach(mother, bucket) grab(mother, bucket) move(mother, bucket, on-floor, in-air) position(mother, bucket, in-air) out(blocks, bucket)

From the predicate-argument structures, we extract for the purpose of this study the features listed above. This gives us the situation *S* given in example (4). The lemmas of the words in the utterance in (1) are given in (3).

(3) *U* = (allemaal, even, d'r, uit, halen, gewoon, zo)

(4) *S* = {**reach, grab, move, position, air, block, bucket, floor, mother, blue, green yellow, red, in, on, out, round, square, star, triangle**}

3.2. Associating language and situation

Suppose we are interested in seeing whether the lemma *ster* 'star' is associated with the (non-relational) feature **star**. We check in our dataset of utterance-situation pairs for every sentence whether the lemma ster is used in the utterance (*ster* ∈ *U*) and whether the feature star is present in the situation (**star** ∈ *S*). This gives us the 2 × 2 table given in table 6.

Table 6. 2 × 2 table for the lemma *ster* and the feature **star**

	star ∈ *S*	**star** ∉ *S*	row sum
ster ∈ *U*	16	1	17
ster ∉ *U*	748	1727	2475
column sum	764	1728	2492

What we can see here is that the lemma *ster* is used almost exclusively in situations in which the feature **star** is present, that is: in situations in which a star-shaped object is an argument to an action or space predicate taking place at that time. However, in an overwhelming majority of cases in which the feature **star** is present, the lemma *ster* is not used. Despite this fact, we can see that the lemma and the feature are clearly associated: there is a higher chance of finding the feature if the lemma is present than if the lemma is not present (94.1% vs. 30.2%) and a slightly higher chance of finding the lemma if the feature is present than if it is not present (2.1% vs. 0.1%).

This association is reflected in a highly significant association using the Fisher exact test (p = 6.6399e−8). Note that the association between lemmas and features resembles that between syntactic environments and words in certain slots in collostructional analysis (Stefanowitsch and Gries 2003). Where Stefanowitsch and Gries establish behavioral profiles for constructions, we use the same statistics to show what the situational profiles are for lemmas. Following Stefanowitsch and Gries (2005), we take the negative natural logarithm of the Fisher exact p-value as a more readily interpretable association measure. This effectively means that the lower the probability of this and more extreme distributions, the higher the association value. For the feature **star** and the lemma *ster*, the association value is −ln 6.6399e−8 = 16.528. Note that, as we use the natural logarithm, association values greater than 2.995 and 4.605 reflect associations significant at $\alpha = 0.05$, resp. $\alpha = 0.01$.

Importantly, the Fisher exact test and derived association value will inform us about both poverty and excessive richness of the context. If the intended meaning is not grounded in some perceivable feature (i.e., the context is too poor) the lemma will be associated with other features that make little sense from an adult point of view. On the other hand, if there are too many features consequently present in the situational context of a lemma, the lemma will be associated about equally with all of them. A high association between a lemma and a feature thus means that that feature is found more often with that lemma than with other lemmas and that that lemma is found more often with that feature than with other features.

Now, we can calculate the association values between the lemma at hand (i.c., *ster*) and all possible features. Doing so and sorting the features on their association values with *ster*, from high to low, we can determine the situational profile for the lemma (table 7).

Table 7. Features with the highests association values with the lemma star.

feature	*association value*
star	16.528
air	7,354
in	6.035
move	3.589
ball	3391

What we see is that the lemma is most strongly associated with what we conceive to be its meaning, namely the feature star. Apart from that, we find noise, with the second strongest association being between the lemma and the feature air. We can repeat this procedure for a range of lemmas. Table 8 gives the top three strongest associated features for the four words referring to the colors of the blocks, and table 9 does so for the words referring to the shape of the blocks.

What we can see from these tables, is that in five out of eight cases, the arguably correct feature is the one with the strongest association with the lemma. The cases in which the correct feature is not associated with the lemma are either very low in frequency (*blauw* (r = 3), *driehoek* (r = 4)), or have an 'unfortunate' split in time windows (three out of nine cases of groen). This means that the child grabbed a green block, and a second later, just within the next time window the mother comments on the block, using the lemma *groen*. Because the child was not *doing* anything with the block in the time window in which the utterance was produced, there is no predicate in which it is involved and hence **green** is not included in *S*.

Table 8. The lemmas for color and the features most strongly associated with them

rood 'red'		*groen* 'green'		*blauw* 'blue'		*geel* 'yellow'	
feature	assoc.	feature	assoc.	feature	assoc.	feature	assoc.

red	3.985	**move**	3.111	**circle**	3.430	**yellow**	4.494
let_go	2.061	**lid**	2.787	**out**	2.530	**let_go**	1.654
in	1.912	**green**	2.671	**mother**	1.907	**mother**	1.447

Table 9. The lemmas for shape and the features most strongly associated with them

vierkant 'square'		*driehoek* 'triangle'		*rond* 'round		*ster* 'star'	
feature	assoc.	feature	assoc.	feature	assoc.	feature	assoc.
square	5.167	**yellow**	3.226	**round**	19.351	**star**	16.528
force	4.423	**point**	2.394	**floor**	6.253	**air**	7.354
match	3.439	**mother**	1.645	**mismatch**	4.259	**in**	6.035

3.3. Relational terms and their situational contexts

In this section, we investigate to what extent the situation can guide a learner towards the meaning of relational terms. We consider eight lemmas that relate to aspects of the game. Linguistic items that do relate to the limited world of the game are action verbs, spatial prepositions and terms referring to properties of the blocks and holes (colors and shapes, as we have discussed in the previous section). Let us turn to the former two categories now.

The utterances contain many verbs pertaining to game-related actions. We find verbs of placement, like *stoppen* 'put in', *halen* 'remove, get'. Then there are verbs of manipulation, such as *duwen* 'push', *pakken* 'grab' and *draaien* 'turn, rotate'. Finally, of interest are verbs denoting relations between blocks and holes, such as *horen* 'belong' and *passen* 'fit'. For all of these relations, a range of modals and light verbs is also used, especially in combination with certain argument-structure constructions and spatial particles. We will not go into these in this exploration. Here, we discuss four verbs, namely *halen*, *pakken*, *passen* and *proberen* 'try'. The former three are members of the categories of placement verbs, manipulation verbs and verbs pertaining to stative relations. The latter is a frequently used verb in the game situation but has no clear correlate in the coded predicates. We discuss it to see what sort of situational profile it displays.

The prepositions and verbal particles also form an interesting category. The caregivers often talk about the spatial relations of the game objects, especially in combination with the three types of verbs discussed above (someone pushing something in something, someone getting something off something, something fitting in something). Four prepositions we will discuss are *in* ‘in’, *op* ‘on’, *uit* ‘out’ and *af* ‘off’.

What are the situational profiles for the eight items we selected? Tables 10 and 11 present the top five most strongly associated features per lemma.

Table 10. Four verbs and the features most strongly associated with them

passen ‘fit’		*pakken* ‘grab’		*halen* ‘get’		*proberen* ‘try’	
feature	assoc.	feature	assoc.	feature	assoc.	feature	assoc.
mismatch	35.693	**hit**	4.866	**bucket**	24.822	**square**	10.946
hole	23.372	**unknown1**	4.132	**out**	19.181	**block**	9.023
near	19.563	**unknown2**	3.447	**on**	7.016	**mismatch**	7.109
block	11.921	**cheek_child**	2.769	**off**	6.074	**child**	5.525
circle	11.038	**observer**	2.233	**floor**	4.954	**show**	5.286

Table 11. Four prepositions/particle and the features most strongly associated with them

in ‘in		*op* ‘on’		*af* ‘off’		*uit* ‘out’	
feature	assoc.	feature	assoc.	feature	assoc.	feature	assoc.
square	5.167	**yellow**	3.226	**round**	19.351	**star**	16.528
force	4.423	**point**	2.394	**floor**	6.253	**air**	7.354
match	3.439	**mother**	1.645	**mismatch**	4.259	**in**	6.035

Turning to the verbs first, we find two verbs that have readily interpretable situational profiles. The four most strongly associated features for *passen* ‘fit’ are all elements of the meaning we would like to assign to the verb. Used mostly with negation, *passen* is about the match or mismatch (‘doesn’t fit’) between a figure (i.c., the block) and a ground (i.c., the hole) that is salient because the two are in some relevant relation to each other (i.c., nearness, as the child is trying to fit the block in the hole). The fact that the precise figure and ground are associated is obviously due to the nature of the game: the only things that fit or don’t fit are blocks and holes.

The lemma *halen* ‘remove, get’ is in our corpus always used with either *af* ‘off’ or *uit* ‘out’. The things that are typically removed are the lid from the bucket and the blocks from the bucket. Now, the verb *halen* is indeed associated strongly with features related to these two situations: the bucket is a pivotal object out of which or **off** of which objects are removed. These objects then enter into a new spatial relationship with another object (typically a support relation with the floor, hence **on** and **floor**). As the mode of removal varies (grabbing and moving by hand, positioning the bucket in such a way that the blocks fall out), no specific feature pertaining to the behavior (e.g., **move**, **position**) is associated with halen.

It may be argued that if a child acts in a certain way when a certain word is used, the child *understands* that word. Suppose the mother says *Go take the lid off*, and the child subsequently takes the lid off. It does not follow logically that the child does so by having an understanding of the particle verb *take off*, the auxiliary *go* and the noun *lid*. In fact, in situations like the ones occurring in this corpus, it is clear what is expected from the participants. An openable and closeable container with blocks that fit in certain holes and not others needs hardly any instruction to be played with. When the toy is given to the mother and the mother shows the child the toy, and says *Go take the lid off*, the child will just as ready start taking the lid off as when the mother makes a comment about the toy (e.g., *That's a nice puzzle!*) or produces another adhortative utterances (e.g., *Let's do this puzzle.*).

For the other two verbs, the situational context is not so clear. This is to be expected from a verb denoting an intention, such as *proberen*, for which encoding the intentional states of the agents might provide a way out, but not so much for a highly concrete verb like pakken. We expect *pakken* to become associated with the feature **grab**, but in only in 57.5% of utterances containing *pakken*, that feature is present, whereas it is also present in 52.2% of all utterances not containing *pakken*.

What then, is the situational action *pakken* relates to? Out of the forty cases of utterances in which the lemma occurs, thirty-one have a grabbing event that can be identified with the utterance within a time frame of two three-second windows before and after the window in which the utterance was produced. An example is the utterance in example (5), produced within the fifteen-second fragment of situational context in table 12. The mother comments on the choice of the daughter to grab the red, round block and does so a few seconds after the action referred to takes place.

(5)	*pak*	*je*	*weer*	*de*	*zelfde*
	grab	you	again	the	same
	'You picked the same one again!'				

Table 12. The situational context of the utterance Pak je weer dezelfde

time	*predicates*
-2	reach(child,bgrsq)
-1	grab(child,breci) move(child,breci,on-floor,in-air)
0	move(child,breci,in-air,at-hoci)
1	letgo(child,breci) in(breci,bucket)

Widening the window thus seems to make the situation available in which the referred action takes place. Nevertheless, if a learner used a window this wide, the feature **grab** would be present in 93.9% utterance-situation pairs. This presents a great amount of noise for the learner, and so if a child is to learn the meaning of *pakken*, other cues must be used. Now, it falls outside of the scope of this exploration to go much deeper into this matter, but as other cues we can imagine bootstrapping the meaning from pre-established object reference: if we know that the pronoun *die* 'that' refers to a certain block, it should be more likely for the lemma *pakken* to be associated with predicates in which that specific block is an argument. The use of these advanced cues can only start when the child has a basic understanding of other referential expressions in the utterance, and hence cannot be the initial drive behind verb learning. This position resembles that of the Emergentist Coalition Model (Hollich, Hirsh-Pasek , and Golinkoff 2000, Poulin-Dubois and Forbes 2006), in which the validity of different cues for acquiring word meanings changes over ontogenetic time. Another route to pursue would be to look at the argument structure construction and see if those can be associated with certain clusters of meanings and then use that information to bootstrap the meaning of the verb. This is essentially the idea of syntactic bootstrapping (Gleitman 1990, Naigles 1990), but then from an emergentist perspective: both argument structure constructions and verbs are acquired through use and mutually reinforce each others acquisition.

For the prepositions, we see a similar pattern: some have relatively strong associations with the (correct) spatial features and the typical figure and ground objects in those relations (*af* and *uit*). For *op*, the highest four features are object categories (three of which are typical objects found in a support relation), and the fifth is the correct meaning of op. Although the feature **on** only comes fifth, it ranks higher than all other spatial, or in general relational features, and we can hence say that the situational context does provide some evidence for the acquisition of the preposition *op*.

The most frequent (r = 286) preposition, in, however, shows hardly any interesting associations. The reason why the lemma is not associated with the feature in is the same as for *pakken*: the feature is highly frequent in general, and hence there are many occurrences of the feature in which the lemma is not used. Furthermore, in only 51.7% of all instances of *in*, a change of state resulting in a containment relation took place. This might be due to the fact that mothers often encourage their child by saying *doe 'm* *daar maar in* ‘go put it in there’, when the actual putting-in only takes place half a minute later. This can be seen from the fact that the accompanying pointing gesture is highly associated with the lemma *in*. What cues there are that might guide the child towards the intended meaning of *in*, remains to be seen. In any event, the learner would have to be able to surpress a lot of noise in the data and perhaps ignore serveral uninformative instances of the lemma. Perhaps the a priori salience of the containment relation (Choi 2006) helps finding the intended meaning as well.

4. Conclusion and directions

It seems that some relational terms are used in a way that an ideal learner can associate them with the correct features. Others, like *pakken*, *proberen* and *in* do not have any clearly associated features that we would recognize to be valid meaning representations from an adult point of view. Interestingly, *pakken* is a highly concrete verb, denoting a single, atomic predicate in our representation (i.e., **grab**), but there is too much noise for the feature and lemma to be associated. We suggested that other cues might help understand the meaning of *pakken*. In the case of a preposition with a relatively accessible intended meaning like *in*, the situational context provides little help in acquiring that meaning, at least if we see learning as naïve association. How a preposition like *in* is acquired from the situational context, remains to be seen.

It thus seems that Gleitman’s disjunction is only partly right. Some relational meanings do have clear correlates in the perceivable situation and are highly as- sociated with the lemmas denoting these meanings. We discussed eight lemmas, five of which were strongly associated with features related to the meaning they were intended to convey. This should be enough to get a learner (at least) started learning terms with relational meanings. One of the other three, *pakken* ‘grab’, also has grabbing events in its context, only taken a little wider (6 seconds before and 6 seconds after the time frame of the utterance). Surprisingly, *in* ‘in’ has no good association with the feature denoting a containment relation. For the lemmas for which the situational context does not provide reliable grounding, Gleitman’s argument still holds: additional cues are needed to explain their acquisition.

In this study, we presented a method for studying the situational context of child-directed language in more detail. We discussed the decisions and the pro- cess, as this is an often overlooked aspect of the study of child language. The main reason for developing such a dataset is to evaluate claims about the excessive rich- ness (size of the hypothesis space) and the poverty of the situational context. We re-evaluated the statement that the situational context contains both too much in- formation to find the right correlate for a word and too little – with the intended relation or action not being present. On the basis of the developed dataset, we concluded that for several relational meanings, the situational context does con- tain features that are very significantly associated with the word that refers to the relation captured in that feature. Other predicates and prepositions are less strongly associated with features related to their meaning. For these words, it is indeed the case that the hypothesis space is often too big, in the sense that the feature is often present when utterances *not* containing the word are produced, so that the association becomes weaker. Nevertheless, the association can form a starting point for word learning, which is later supplemented with additional cues (other words in the sentence, syntactic frames, social cues, better understanding of the task), a position that has been worked out in more detail in the Emergentist Coalition Model (Hollich et al. 2000).

What this study shows, is that it is possible to *get started* on learning relational semantics on the basis of mere association. The relevance of this analysis for usage-based theorizing is that domain-specific biases for word learning are not needed to start developing a lexicon. However, learning the symbolic mappings remains difficult, and there often is too much information in the situation. One challenge for usage-based approaches is to develop accounts of structured learning (as opposed to merely associative learning) where constructions (for example) play a role in bootstrapping word meanings, and to apply this in a structured way to naturalistic data like the dataset presented in this paper.

Using the atomic-feature representation, we plan to work with actual word learning models (e.g., Fazly, Alishahi, and Stevenson 2010) in order to see if the meaning can indeed be learned by a cognitively realistic incremental model. A next step would be to see if it is possible to have a computational learner induce a simple, (semi-)productive grammar that can be used to interpret unseen utterances.

In our exploration we left aside interesting, but more complex issues, such as the relation between the nature of the child-directed language and the output. For prepositions, it would be interesting to see if there is a correlation between the availability of situational context guiding the learner in the correct direction and the age of acquisition or the overextension patterns of certain lemmas (e.g., the preposition *af* 'off of' and *uit* 'out of' (Bowerman 1995) or the set of *op* 'canonical support', *aan* ‘support by hanging and attachment’ and *in* ‘containment’ and the overextension of *op* to *aan*-situations (Bowerman 1993)).

Modeling known phenomena, whether they were established by experimental method or by observation, gives us a greater insight in the cognitive mechanisms underlying them. However, in order to model the acquisition of meaning, a de- tailed understanding of the context is needed, especially given claims about the nature of that context. With the development of a dataset, we hope to encourage research in language acquisition that takes the situatedness of actual child-directed language into account in a systematic way.

Notes

* Corresponding author: Barend Beekhuizen, Leiden University. Email address: barendbeekhuizen@gmail.com. The work reported here is part of the first author’s PhD-project Constructions Emerging, funded by NWO. The second and third authors have made significant contributions to the development of the described method. The authors would like to thank Marian Bakermans and Rien van IJzendoorn of the department of Child Studies at Leiden University for allowing us to use their data, as well as the two anonymous reviewers for their valuable comments.

1. As a working definition of ‘relational’, we adopt Gentner and Kurtz’s (2005) explanation that “[e]ntity categories can be thought of as first-order partitions of the world (Gentner 1982) and relational categories second-order ways of organizing and linking those first-order partitions”.
2. We realize that it is the case that there are verbs that have non-relational meanings, as well as nouns and adjectives that have relational ones. The particular items we study do not belong to these types.
3. This is not to say that verbs encoding mental states, intentions and dispositions have no ground in the child’s situational understanding independently of language. From a very young age, children have a basic understanding of intentions and emotions (Behrend and Scofeld 2006, Tomasello 2003) that might become associated with certain verbs. Barak, Fazly, and Stevenson (2012) provide an interesting first step towards modelling the acquisition of such meanings.
4. To be found on https://github.com/dnrb/publications/blob/master/annotation_guidelines.p df
5. Developed by The Language Archive at the Max Planck Institute for Psycholinguistics, Nijmegen, and publicly available from http://tla.mpi.nl/tools/tla-tools/elan/. For a description, see Brugman and Russel (2004)

6. The abbreviations used are the same as in Table 2, but are repeated here for the reader's convenience: **to** = toy, **bu** = bucket, **ha** =handle, **li** = lid, **ho** = hole, **blocks** = all blocks collectively, **b** = block, **re** = red, **gr** = green, **bl** = blue, **ye** = yellow, **sq** = square, **st** = star, **ci** = circle, **tr** = triangular, **mo** = mother, **ch** = child, **ob** = observer, **table** = table, **floor** = floor, **air**=air.
7. We take a type to be a unique transcribed representation of the speech, including diminutives and different sorts of inflection. A lemma, on the other hand, is taken to be the core word (whether it is a compound, another derivation or a monomorphemic element). Examples of different types belonging to the same lemma are *ster* 'star', *sterren*, 'star.PL' and *sterretje*, 'star.DIM', that are all types related to the lemma *ster* 'star'. No methodological problems were found in determining the types and lemmas.

References

Baillargeon, Renee and Su-Hua Wang
2002 Event categorization in infancy. *Trends in Cognitive Sciences* 6(2), 85–93.

Barak, Libby, Afsaneh Fazly, and Suzanne Stevenson
2012 Modeling the Acquisition of Mental State Verbs. In *Proceedings of the NAACL-HLT Workshop on Cognitive Modeling and Computational Linguistics (CMCL).*

Behrend, Douglas A. and Jason M. Scofeld
2006 Verbs, Actions, and Intentions. In *Action Meets Word. How Children Learn Verbs*, Kathryn Hirsh-Pasek, and Roberta M. Golinkoff (eds), 286–307. Oxford, UK: Oxford University Press.

Bowerman, Melissa
1993 Typological Perspective on Language Acquisition: Do Crosslinguistic Patterns Predict Development. In Eve V. Clark (ed.) *Proceedings of the Twenty-fifth Annual Child Language Research Forum*, 7–15. Stanford, CA: CSLI Publications.
1995 Learning how to structure space for language: a crosslinguistic perspective. In *Language and Space*, Paul Bloom, Mary. A. Peterson, Lynn Nadel, and Merill F. Garett (eds.), 385–436. Cambridge, MA: MIT Press.

Brugman, Hennie and Albert Russel
2004 Annotating Multimedia/Multi-modal resources with ELAN. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC).*

Carletta, Jean

1996 Assessing agreement on classification tasks: the kappa statistic. *Computational Linguistics* 22(2), 249–254.

Casasola, Marianella, Jui Bhagwat, and Kim T. Ferguson
2006 Precursors to verb learning: Infant's understanding of motion events. In *Action Meets Word. How Children Learn Verbs*, Kathryn Hirsh-Pasek, and Roberta M. Golinkoff (eds), 160–190. Oxford, UK: Oxford University Press.

Soonja Choi
2006 Preverbal Spatial Cognition and Language-specific input: Categories of Containment and Support. In *Action Meets Word. How Children Learn Verbs*, Kathryn Hirsh-Pasek, and Roberta M. Golinkoff (eds), 191–207. Oxford, UK: Oxford University Press.

Cohen, Jacob
1960 A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20(1), 37–46.

Fazly, Afsaneh, Afra Alishahi, and Suzanne Stevenson
2010 A probabilistic computational model of cross-situational word learning. *Cognitive Science: A Multidisciplinary Journal* 34(6), 1017–1063.

Feist, Michele I.
2000 *On* In *and* On. *An investigation into the linguistic encoding of spatial scenes*, PhD thesis, Northwestern University.

Fleischman, Michael and Deb K. Roy
2005 Why Verbs are Harder to Learn than Nouns. Initial Insights from a Computational Model of Intention Recognition in Situated Word Learning, In *Proceedings of the 27th Annual Meeting of the Cognitive Science Society*.

Michael C. Frank, Noah D. Goodman, and Joshua B. Tenenbaum
2009 Using Speakers Referential Intentions to Model Early Cross-Situational Word Learning. *Psychological Science* 20(5), 578–585.

Michael C. Frank, Joshua B. Tenenbaum, and Anne Fernald
2013 Social and discourse contributions to the determination of reference in cross-situational word learning. *Language, Learning, and Development* 9, 1–24.

Gentner, Dedre
1978 On Relational Meaning: The Acquisition of verb meaning. *Child Development* 49, 988–998.

Gentner, Dedre
1982 Why Nouns are Learned Before Verbs: Linguistic Relativity versus Natural Partitioning. In *Language Development. Volume*

2: Language, Thought, and Culture, Stan A. Kuczaj II, (ed.), 301–334. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Gentner, Dedre and Kurtz, Kenneth. J.

2005 Relational Categories. In *Categorization inside and outside the lab*, W. K. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman, and P. W. Wolff (eds),151–175. Washington, DC: APA.

Gleitman, Lila

1990 Sources of Verb Meanings. *Language Acquisition* 1(1), 3– 55.

Hollich, George. J., Kathryn Hirsh-Pasek, and Roberta M. Golinkoff

2000 *Breaking the language barrier: An emergentist coalition model for the origins of word learning*. *Monographs of the society for research in child development* 65(3), 1–135.

Sootman-Buresh, Jennifer, Amanda L. Woordward, and Camille W. Brune

2006 The Roots of Verbs in Prelinguistic Action Knowledge. In *Action Meets Word. How Children Learn Verbs*, K. Hirsh-Pasek, and R. M. Golinkoff (eds), 208–227. Oxford, UK: Oxford University Press.

Jean M. Mandler

2006 Actions organize the infant's world. In *Action Meets Word. How Children Learn Verbs*, Kathryn Hirsh-Pasek, and Roberta M. Golinkoff (eds), 111–133. Oxford, UK: Oxford University Press.

Meltzoff, Andrew N. and M. Keith Moore

1995 Infants' Understanding of People and Things: From Body Imitation to Folk Psychology. In *The Body and the Self*, José L. Bermúdez, Anthony Marcel, and Naomi Eilan (eds), 43–69. Cambridge, MA: MIT Press

Letitia R. Naigles

1990 Children use syntax to learn verb meanings. *Journal of child language* 17(2), 357–74.

Narasimhan, Bhuvana and Anetta Kopecka. (eds.)

2012 *Events of putting and taking: A crosslinguistic perspective.* Amsterdam, The Netherlands: John Benjamins.

Poulin-Dubois, Diane and James N. Forbes

2006 Word, Intention, and Action: A Two-Tiered Model of Action Word Learning. In *Action Meets Word. How Children Learn Verbs*, Kathryn Hirsh-Pasek, and Roberta M. Golinkoff (eds), 262–285. Oxford, UK: Oxford University Press.

Anatol Stefanowitsch and Stefan T. Gries

2003 Collostructions: Investigating the interaction of words and constructions, *International Journal of Corpus Linguistics* 8(2), 209–243.

2005 Covarying collexemes. *Corpus Linguistics and Linguistic Theory* 1(1), 1–43.

Michael Tomasello
2003 *Constructing a language: A Usage-Based Theory of Language Acquisition*, Cambridge, MA: Harvard University Press.