The Earliest Utterances in Dialogue: Towards a Formal Theory of Parent/Child Talk in Interaction*

Jonathan Ginzburg and Sara Moradlou
CLILLAC-ARP (EA 3967) & Laboratoire Linguistique Formelle (LLF) (UMR 7110)
& Laboratoire d'Excellence (LabEx)–Empirical Foundations of Linguistics (EFL)
Université Paris-Diderot, Sorbonne Paris Cité, Paris, France
yonatan.ginzburg@univ-paris-diderot.fr, sara.moradlou@gmail.com

Abstract

Early, initial utterances by children have received relatively little attention from researchers on language acquisition and almost no attempts to describe them using a formal grammar. In this paper we develop a taxonomy for such utterances, inspired by a study of the Providence corpus from CHILDES and driven by the need to describe how the contents of early child utterances arise from an interaction of form and dialogical context. The results of our corpus study demonstrate that even at this early stage quite intricate semantic mechanisms are in play, including non-referential meaning, akin to non–specific readings of quantifiers. We sketch a formal framework for describing the dialogue context and grammar that underlies such utterances. We consider very briefly and informally how some such utterances emerge from parent/child interaction.

1 Introduction

The early stages of a process are crucial in understanding its subsequent development. Although there has been some work in this area, which we summarize below, it seems true to say that the early, initial utterances by children have received relatively little attention from researchers on language acquisition and almost no attempts to describe them using a formal grammar.[1] Given that parents and carers can make sense of much of what young children say, we assume the mechanisms of this understanding process deserve formal analysis and, unless compelling reasons to the contrary be given, incorporation within some notion of grammar. It is clear that such a notion will rely, even more than is the case for adult spoken interaction, on a detailed theory of context.

In this paper we develop a taxonomy for early child utterances. In contrast to previous work, summarized in section 2 which was strongly based on speech act theory and paid little attention to the fine structure of semantic combinatory mechanisms, our own taxonomy, developed in section 3, based on the Providence corpus from CHILDES (Demuth et al., 2006), is driven by the need to describe how the contents of early child utterances arise from an interaction of form and dialogical context. The results of our corpus study, described in section 4, demonstrate that even at this early stage quite intricate semantic mechanisms are in play, including non-referential meaning, akin to non–specific readings of quantifiers. In section 5 we sketch a formal framework for describing the dialogue context and grammar that underlies such utterances, showing that even at this initial stage, the child grammar is in a sense continuous with adult grammar. In section 6, we consider very briefly and informally how some such utterances emerge from parent/child interaction.

2 Literature Review

Previous work on categorizing children's utterances is mainly based on Speech Act Theory (Searle, 1969; Austin, 1975). The work on speech act analysis of child language attempts to characterize the nature of parent-child interactions and its links to language learning. These approaches to language acquisition view verbal forms as means

*We acknowledge support from Lab(oratory of)Ex(cellence)–EFL (ANR/CGI), in particular for providing a graduate fellowship to the alphabetically second author. We would like to thank Eve Clark, Judit Gervain, and three anonymous reviewers for DialDam for very useful comments on an earlier draft.

[1] For brief discussion in the context of a proposal concerning the evolution of grammar, see (Jackendoff and Wittenberg, 2014).

of expressing communicative intents and emphasize the role of function (e.g., (Bates, 1976; Bloom, 1967; Bruner, 1975; Dore, 1975; Dore, 1974; Ninio, 1992; Ninio and Bruner, 1978; Ninio and Snow, 1988)) as opposed to viewing learning as a mapping from form to meaning. Thus, there is no attempt to describe the fine structure of descriptive contents, which as we will see, already involves quite intricate combinatorial mechanisms. Nor is there explication of how these arise drawing on contextual information.

Ninio (1992) shows a strong correlation between single word utterances of children and those of their mothers, indicating a high degree of form-function specificity (see also, (Bruner, 1975; Shatz, 1979; Snow, 1972)). This co-variation between parent speech and child's utterances is the basis of most acquisition theories that consider conversation as an important factor. Bruner (1981) notes that mothers used highlighting or fore-fronting of the objects extensively, when introducing them to children, and also routinized individualized characteristic ways of preparing for presentation when the child was not attending to them (e.g., calling by name). Based on his observations that the interactions between mother and child follow highly regular patterns and that these patterns evolve as the child becomes more and more competent in language use, Bruner (1981) argues that this *Language Assistance System* plays an important role in children's language development.

We briefly describe here two classification schemata based on the speech acts approach we mentioned above: Dore (1974) sets out to explain the development of adult speech acts repertoire; using the data collected in a longitudinal study of two subjects in their single-word stage of language production, he categorizes *Primitive Speech Acts* of these infants into 9 types each of which differs with the others at least in one feature, either in form or function. INCA[2] (Ninio and Wheeler, 1986) and its abridged version (INCA-A) (Ninio et al., 1994) are annotation schemas that code communicative intents in two parts: level of verbal interchange, which is defined as a series of utterances that serve a unitary interactive function, (e.g., negotiating immediate activity, discussing joint focus of attention, etc) and utterance level speech acts (e.g., requesting, proposing an action, etc.). These systems are meant to code communication attempts for both adults and children in different stages of acquisition; this makes the number of types to chose from quite big for the coder and the annotation work rather difficult.

Next, we describe our taxonomy for classifying children's early utterances, with annotation effort, and dialogue dynamics in mind.

3 Corpus Study

We annotated the odd number files from 11 to 15 months for Naima and Lily of Providence corpus (Demuth et al., 2006) in CHILDES (MacWhinney, 2000) with utterance types based on Fernández and Ginzburg (2002)'s taxonomy of adult non-sentential utterances (NSU).[3] However, these adult NSU types do not cover all of the NSUs observed in child language; furthermore, some adult NSU categories do not occur at all in the early stages of acquisition. We developed our taxonomy for the early stage of child language using Naima's utterances in her one-word stage: we manually categorized the utterances into one of the types we will discuss shortly, based on their form and the conversation function they served, trying to maximize the number of phenomena covered by our taxonomy. We only retained the types that occurred in more than 4 percent of the utterances in at least one of the observation sessions. The motivation for this was to exclude utterances that occurred very rarely; we applied the threshold frequency on sessions instead of the complete development set to capture development of types over time.

Below we describe each class of utterances with examples; these classes are organized in three broad categories: labeling types are the utterance types that refer to the visual scene. The second class of types are those that follow up on parent's utterances. Attention directing types are initiated by the child and play a role in managing joint attention.

3.1 Labeling types

Visual Object Pointing

We classified word or word-like utterances that referred to entities in the visual field in order to label them as **VisObjP**. This type was usually accompa-

[2]Inventory of Communicative Acts

[3]We used the Providence corpus since it is multimodal and children are recorded there from very early stages of speaking.

nied by pointing (example 1) or reaching-out gestures (example 3), and gaze towards the object that was being named.

(1) CHI: yyy .
pho: ˈkʌkə
sit: CHI is pointing to the microphone
MOT: yeah , that's my microphone .
MOT: you have one too . (Naima at 0;11.28)

(2) CHI: yogurt .
MOT: yogurt .
MOT: is it good ? (Naima at 1;0.28)

(3) CHI: bike yyy .
pho: ˈbæk ˈbæbæ
sit: CHI picks up toy bike
MOT: oh your bike . (Naima at 1;1.25)

Visual Predication

When a property of an entity or an event in the visual situation was predicated by an utterance, it was annotated with **VisPred**. This type is similar to VisObjP in the co-occurring gestural-proxemic cues, and in its dependence on the visual situation.

(4) MOT: hey , you dropped an odio . CHI: down . sit: CHI is picking up the cereal. (Naima at 0;11.28)

(5) sit: book reading activity. CHI: baby . MOT: and there's a baby . CHI: sleeping . pho: ʃLiːpi MOT: yes the baby is sleeping . MOT: on the bed . (Naima at 1;2.23)

(6) CHI: big . pho: ˈbiːgə sit: CHI is trying to put a toy chicken into a cup. MOT: oh the chick is too big for that small cup . MOT: it doesn't fit in . (Naima at 1;3.12)

Visual Onomatopoeic Utterances

VisOno class covers onomatopoeic utterances (animal, machine, and though less often, human sounds like imitations of laughing and crying) when it is triggered by an entity or event in the visual attention of the child:

(7) CHI: baa baa baa baa baa baa baa . sit: CHI is pointing at a sheep in a picture book. MOT: that's right , that's the sheep going baː baː . (Naima at 1;0.28)

3.2 Types following up on parent's utterances

Short Answer

Utterances in response to parent's Wh-questions or implicit Wh-questions (usually using pausing, intonations, and gestural cues) were categorized as **ShortAns**. This was irrespective of the correctness of child's answer, as long as the utterance was word-like and it could be interpreted as an answer taking into account child's non-verbal behavior and parent's interpretation of the utterance:

(8) MOT: who's that coming in the door? CHI: Daddy . MOT: yes that's right . (Naima at 0;11.28)

(9) MOT: what else is here ? CHI: duckling . pho: ˈgʌkəlɪ MOT: duckaling , that's a duck . MOT: quack quack , (.) and + . . . (Naima at 1;0.28)

Repetition Acknowledgment

This class contains utterances that follow up on parent's previous utterance(s) by repeating (part of) it. **RepAckWord** was distinguished from pure imitations (**Imit**) based on child's participation in the conversation and annotators' judgment of the nature of the repetition: utterances that functioned purely as practice for pronunciation were annotated as **Imit**. In addition, if child repeated an utterance that was not directed to her or she did not seem to be paying attention, that utterance was not considered a repetition acknowledgment but as an imitation (compare 12 and 10). We also considered onomatopoeic utterances that were related to parents previous utterance (for example, imitating an animal that has been mentioned by the parent) as repetition acknowledgment and tagged them with **RepAckOno**.

(10) MOT: that baby has a bottle , did you notice that ? CHI: bottle . MOT: yeah , baby has a bottle . (Naima at 1;0.28)

(11) MOT: it's a shovel . CHI: shovel . MOT: just like your shovel . (Naima at 1;3.12)

Imitation

(12) MOT: I went to +//. FAT: xxx . MOT: pain d'avignon +//. CHI: yyy . pho: ˈpliː MOT: pain d'avignon yesterday . CHI: yyy . pho: ˈpliː MOT: play . FAT: wash xxx first . MOT: brioche bread and some +//. CHI: brioche yyy yyy . (Naima at 1;2.23)

3.3 Attention directing types

Call

This category contains instances of "Daddy" and variations of "Mommy" used as a means of directing or establishing shared attention:

(13) CHI: Mama ? MOT: yeah ? MOT: that's my tea , you can't drink my tea , babies don't drink tea . (Naima at 1;0.28)

Request

The utterances that functioned as requests for entities using surface forms analogous to those in VisObjP, as in (14), were annotated with **ReqObj**, whereas the requests with forms referring to events or properties of entities were tagged as **ReqPred**, like in (16) and (15).

(14) CHI: Mommy . MOT: yes Naima . CHI: water . MOT: you want some more water . (Naima at 1;1.25)

(15) CHI: more . pho: ˈmːon MOT: oh more ? MOT: okay , here's a big piece of wheat . MOT: put in that one . MOT: make sure you chew that , okay ? (Naima at 1;0.28)

(16) CHI: Daddy . FAT: yes baby . FAT: you look so serious and earnest . CHI: up . CHI: Daddy . MOT: up . FAT: up oh . FAT: okay baby . FAT: you said up . FAT: pick me up ? (Naima at 1;2.23)

Types		Ages				
		0;11.28	1;0.28	1;1.25	1;2.23	1;3.12
Vis	ObjP	43.3	30.5	32	32.3	25.6
	Pred	7.8	0.9	2.7	2.4	8.1
	Ono	0	1.7	1.3	5.7	1.9
RepAck	Word	3.3	22.9	14.7	13.7	13.4
	Ono	1.1	5.1	0	2.4	3.1
Imit		10	4.2	6.7	4	4.7
ShortAns		10	12.7	12	4.8	11.6
Call		11.1	6.8	18.7	17.7	7.2
Req	Obj	0	0.9	4	9.7	2.5
	Pred	0	1.7	0	1.6	4.7
Multi-word		0	0	0	1.6	7.5
% covered		86.7	87.3	92	96	90.3
# annotated		90	118	75	124	320
# unintelligible		162	104	66	142	193
# total		299	245	171	356	636

Distribution of utterance types over time, in percentage.
Ages in years;months.days

Table 1: Distribution of types for Naima

Types		Ages				
		1;1.02	1;2.02	1;2.30	1;3.27	1;4.25
Vis	ObjP	10.3	18.2	0	0	24.1
RepAck	Word	24.1	9.1	0	0	6.9
Imit		13.8	4.5	10	0	6.9
ShortAns		48.3	59.1	90	71.4	55.2
Call		0	4.5	0	28.6	6.9
% Covered		96.6	95.5	100	100	100
# annotated		29	22	10	7	29
# unintelligible		107	215	468	723	345
# total		148	234	478	732	378

Distribution of utterance types over time, in percentage.
Ages in years;months.days

Table 2: Distribution of types for Lily

4 Results

Table 1 summarizes the type frequencies in Naima's speech for the types retained in the taxonomy after applying the above mentioned threshold to remove very rare types. In this table, percent covered is the ratio of utterances that our taxonomy could account for, over total number of annotated utterances. It is worth noting that we only annotated the first instance of an utterance when it was repeated multiply in a sequential manner. We also applied this metric to a new set of transcripts, odd files of Lily from the Providence corpus: our taxonomy achieved high coverage for both Naima (86.7–96%) and Lily (95.5–100%). Distribution of types for Lily is provided in Table 2.

The most frequent utterances for Naima labeled entities in the visual scene with VisObjP, and short answers accounted for most of Lily's utterances. Naima was a precocious talker whereas Lily, though good at answering Wh-questions, did not often initiate conversations verbally. This points to a possible shortcoming of our taxonomy: annotations were mainly driven by word or word-like utterances. Extra-linguistic cues were only used to guide category assignment to a somewhat intelligible verbal act and did not merit annotation on their own. This is a good first approximation. Nonetheless, the gestural actions initiating conversations also play an important role in language acquisition (Kelly, 2011). In the example below Lily uses pointing at pictures in a book and flipping pages as requests for labels:

(17) CHI: yyy . pho: ˈɪ sit: CHI turns the page back and points at it MOT: fish ! CHI: yyy . pho: ˈɛ MOT: turtle: . MOT: fish ! MOT: turtle . sit: CHI flipping page back and forth (Lily at 1;2.02)

Another reason we think this might be fruitful for the study of the early stages of child language is the high proportion of 'failed' (viz incomprehensible to the adult) utterances in the files we annotated; 39% of the utterances were unintelligible in Naima's files and this number goes even further up to 94% for Lily.

Repetition acknowledgments usually happened when a new label was provided by the caregivers. This is in line with the results of (Clark, 2007):

(18) sit: CHI crawling toward stuffed animal dog MOT: yeah there's puppy honey ! MOT: do you see puppy ? MOT: puppy's [: puppy is] over there . CHI: puppy . pho: ˈhʌbeɪ (Lily at 1;2.02)

For Naima, VisObjP Category becomes less frequent as she acquires new ways of referring to objects and moves to the two word stage, as suggested by the emergence of two word predication (e.g., "sleepy daddy") and other multi-word utterance denoting relations that are more sophisticated than simple labeling, at age 1;2.23 and 1;3.12, and also the increase in proportion of VisPred utterances. The 7.8% for VisPred at age 0;11.28 goes against this trend; but after taking a closer look at these utterances we discovered that this session was where Naima learned to say "down". The analysis of VisPred forms (Table 3) shows that the form diversity of VisPred utterances goes up with age. Similar analyses for other types in our taxonomy might prove useful for gaining more insight into children's developmental paths.

We calculated inter-annotator agreement scores using annotations done for a portion of file seven of Naima (approximately 25 minutes of conversation), by three other coders external to the project. The analysis of mismatches showed that RepAck

	ages				
VisPred	0;11.28	1;0.28	1;1.25	1;2.23	1;3.12
# different forms	2	1	2	3	10
'down'	5	1	1	1	4
'good'	1		1		
'hot'				1	
'sleeping'				1	
'big'					7
'tall'					5
'heavy'					2
'stuck'					2
'off'					2
'dirty'					1
'on'					1
'kiss'					1
'clap'					1
total #	7	1	2	3	26

Frequency of VisPred forms over time
Ages in years;months.days

Table 3: Frequency of VisPred for Naima

	r2	r3	r4
r1	86.66%, .83	89.19%, .85	80%, .60
r2		94.03%, .92	82.22%, .66
r3			82.05%, .78

Percent agreement %, Cohen's kappa

Table 4: Inter-annotator agreement evaluation

and Imit categories are difficult to distinguish. Most mismatches however, were the result of the different choices by different coders for utterances to annotate, in the case of repetitive utterances. Removing these instances (i.e. including only utterances that both coders regarded as non repetitive), improved the scores significantly. These percent agreement, and Cohen's kappa coefficient numbers are presented in Table 4.

5 Modelling parent/child interaction

5.1 Initial Assumptions

One of our main hypotheses is that parent (P)/early child (C) interaction involves mechanisms that, by and large, exist in regular adult/adult interaction. This is true for the principles underlying the coherence of such interaction and the types of utterances made by both parties. There is of course a basic asymmetry in that C is expected to adapt to P's linguistic system and not vice versa (though there are parents who attempt the latter to a certain extent, as we will see.).

The main apparent differences derive from the fact that that much of the time C does not respond or responds in a fashion that is not comprehensible to P. Nor does P limit him/herself to uttering 'comprehensible language', in contrast to adult interaction where a basic presupposition exists of using 'shared language' (violating this is viewed, minimally, as arrogance.). However, given the flexibility of turn taking and the existence of dialogical repair mechanisms, this means that in the short term the conversation does not break down, whereas in the long term much positive evidence exploitable for learning gets produced.

5.2 Dialogue GameBoards

We use the dialogue framework KoS (see e.g. (Ginzburg and Fernández, 2010; Ginzburg, 2012) for details) as the framework for describing P/C interaction. On the approach developed in KoS, there is actually no single context—instead of a single context, analysis is formulated at a level of information states, one per conversational participant. This assumption is particularly useful for modelling an asymmetric type of interaction as here. The dialogue gameboard represents information that arises from publicized interactions. Its structure is given in (19)—the *spkr,addr* fields allow one to track turn ownership, *Facts* represents conversationally shared assumptions, *Pending* and *Moves* represent respectively moves that are in the process of/have been grounded, *QUD* tracks the questions currently under discussion, though not simply questions qua semantic objects, but pairs of entities which we call *InfoStrucs*: a question and an antecedent sub-utterance (the *focus establishing constituent* (FEC) that partially specifies a subsequent focal utterance.[4]

(19) DGBType $=_{def}$

$$\begin{bmatrix} \text{spkr: Ind} \\ \text{addr: Ind} \\ \text{utt-time : Time} \\ \text{c-utt : addressing(spkr,addr,utt-time)} \\ \text{Facts : Set(Proposition)} \\ \text{Pending : list(locutionary Proposition)} \\ \text{Moves : list(locutionary Proposition)} \\ \text{QUD : poset(Infostruc)} \end{bmatrix}$$

DGBs are useful means of conceptualizing an adult's public context in dialogue interaction. To

[4] On the whole, for current purposes one could restrict attention to QUDs consisting solely of questions. However, FECs potentially play a significant role in learning, as hinted in section 6.

what extent is this plausible for young children? It is plausible to assume that child representations of context are in several ways less detailed than adult ones, though some of the burden can be associated with incomplete mastery of the conversational rules we discuss shortly, rather than the representations. At the same time, given the evidence for common ground (Tomasello, 1988; Clark, 2013), for the ability to participate in simple games (amply demonstrated in Providence), and awareness of 'topic continuity' across utterances justifies the existence of (some notions corresponding to) FACTS, MOVES, and QUD respectively. We return to the issue of Pending below when we discuss metacommunicative interaction.

We make one modification to the DGB, one which is in any case also required for modelling adult/adult interaction. We introduce an extra field to FACTS which we dub *Vis(ual)Inf* of type *RecType* (cf. MSOA in (Grosz, 1977; Poesio and Rieser, 2011)). This represents the dialogue participant's (view of) the visual situation and attended entities. The basic structure of this type is given in (20a). A concrete example is given in (20b): a visual situation involving a doll with spot on her head, where the spot is the attentional focus:

(20) a. $\text{VisInf} = \begin{bmatrix} \text{VisSit : RecType} \\ \text{InAttention : Ind} \\ \text{c1 : member(InAttention,VisSit)} \end{bmatrix}$

b. $\begin{bmatrix} \text{VisSit} : \begin{bmatrix} \text{x: Ind} \\ \text{c1 : doll(x)} \\ \text{y : Ind} \\ \text{c2 : head(y,x)} \\ \text{z : Ind} \\ \text{c3 : spot(z)} \wedge \text{On(z,y)} \end{bmatrix} \\ \text{InAttention = VisSit.z : Ind} \\ \text{c1 : member(InAttention,VisSit)} \end{bmatrix}$

5.3 Conversational Rules

The basic units of change are mappings between dialogue gameboards that specify how one gameboard configuration can be modified into another on the basis of dialogue moves. We call a mapping between DGB types a *conversational rule*. The types specifying its domain and its range we dub, respectively, the *preconditions* and the *effects*, both of which are supertypes of DGBType.

An example of such a rule, taken from (Ginzburg, 2012), needed to analyze querying and assertion interaction is given in (21). QSPEC is what characterizes the contextual background of reactive queries and assertions. (21) says that if q is QUD–maximal, then subsequent to this either conversational participant—hence, the turn underspecification characterized by the type $TurnUnderspec$—may make a move constrained to be q–specific (i.e. either About or Influencing q).

(21) QSPEC

$$\begin{bmatrix} \text{pre: } \begin{bmatrix} \text{qud} = \langle \text{i, I} \rangle \text{: poset(InfoStruc)} \end{bmatrix} \\ \text{effects : TurnUnderspec} \\ \wedge_{merge} \begin{bmatrix} \text{r : AbSemObj} \\ \text{R: IllocRel} \\ \text{LatestMove =} \\ \text{R(spkr,addr,r) : IllocProp} \\ \text{c1 : Qspecific(r,i.q)} \end{bmatrix} \end{bmatrix}$$

QSPEC highlights a feature of KoS's dialogue semantics crucial for P/C interaction: the fact that a speaker can straightforwardly answer their own question. Such cases get handled because turn taking is abstracted away from querying: this allows either conversationalist to take the turn given the QUD-maximality of q.

Given how one sided interaction can be, also crucial are rules controlling downdating of questions without receiving responses (see section 8.3 in (Ginzburg, 2012)).

5.4 Metacommunicative Interaction

Metacommunicative interaction is handled in KoS by assuming that in the aftermath of an utterance u it is initially represented in the DGB by means of a *locutionary proposition* individuated by u and a grammatical type T_u associated with u. If T_u fully classifies u, u gets grounded, otherwise clarification interaction ensues regulated by a question inferrable from u and T_u. If this interaction is successful, this leads to a new, more detailed (or corrected) representation of either u or T_u. This is also the basis for an account of interactive word learning (Macura, 2007; Larsson and Cooper, 2009).

In early child utterances, much of the time the adult does not react to incomprehensible utterances, but such reaction is certainly not rare.[5] At this stage the child does not initiate clarification interaction, but she clearly is sensitive to feedback about her utterances, both in terms of form

[5] *MOT: chew and swallow . *CHI: yyy . xpho: ba: *MOT: ba: . *MOT: what is ba ? *MOT: what d'you mean ba ? (from: naima2)

and meaning (Gallagher, 1977). This justifies the need for modifiable utterance representations akin to locutionary propositions—as we suggest below presumably incorporating probabilistic notions—and to some notion like PENDING. Given that the child does not initiate such interaction, how does it arise? We believe this is an instance of domain–specific reasoning about the dialogue, as we now explain.

5.5 Visually accessible chat

So far we have mentioned entirely domain general principles of interaction. In adult/adult interaction the activity type influences the interaction e.g. with respect to issues that arise without explicitly being introduced (cf. differences between conversations in a bakery, a train station, or among friends in a café.). Similar considerations apply here. As far as the parent is concerned, s/he confronts the following challenge—what to discuss with an interlocuter who, much of the time, does not respond in a comprehensible fashion and whose knowledge of language is very incomplete. The parent can talk about that which is visible and susceptible to linguistic description. Using the theory of conversational genres developed in (Larsson, 2002; Ginzburg, 2012), one may characterize this genre as *visually accessible chat* in which at any given point an inferrable issue is: *what word can one use to refer to the visually most prominent entity?*

This is the basis for our account of how the utterances (22(a),(b)) get the italicized readings.

(22) a. *MOT: should we comb her hair ? *MOT: with a comb ? *MOT(a): comb. $\mapsto$ *This entity can be referred to as a **comb*** (from: naima2)

b. sit: CHI reaching for MOT's microphone *MOT: that's a microphone . *CHI: microphone [?] . sit: CHI looking and pointing at MOT's microphone *MOT: mi:crophone . *MOT (b): a microphone . $\mapsto$ *This entity can be referred to as a **microphone*** (from: naima 4)

5.6 Initial child grammar

In this section we show how to formally characterize the utterance types which make up the taxonomy in section 3. For this purpose we use HPSG$_{TTR}$ (Ginzburg, 2012), a variant of the grammatical formalism Head–driven Phrase Structure Grammar (Sag et al., 2003). In specifying the child's grammar—a set (or type) of utterance types, we need to distinguish the comprehension grammar and production grammar. What we specify here are the production types (for interactions where the child is the speaker), but this is clearly distinct from the (presumably more extensive set/type of) comprehension types (for interactions where the child is the addressee).[6]

In terms of syntax, we follow the approach of (Ginzburg and Sag, 2000) to non-sentential utterances treating these as constructions with a single daughter, which constitutes the head.[7] Semantically, the context, represented within the field *dgb-params*, plays a crucial role via QUD, VisSit or Pending, providing the main predicate and/or the conversational move type.[8]

RepAck As with their use by adults, RepAck utterances are not straightforward to analyze: they can be viewed as bare acknowledgements ('an utterance containing this word was just uttered'.) or they can be viewed as singling out a word because the child is testing their pronunciation or understanding of the word. We propose that the conventional meaning of such utterances is essentially:[9] *child acknowledges that an utterance including the word word$_i$ happened*. (23) captures this by imposing segmental parallelism between a sub-utterance u1 of the prior (maximally pending) utterance and the AckRep utterance:[10]

[6] Indeed such a distinction probably needs to be drawn for the adult as well, e.g., to capture the difference between a carer of a given child and a random adult who interacts with child. But that is a somewhat more controversial case.

[7] In (Ginzburg and Sag, 2000) the category of these constructions was verbal, as by assumption this was the category of root utterances. We do not adopt this assumption here, which in any case is not ultimately tenable even for adult grammars, for a variety of interjections. We utilize a type *root*, whose explication we leave for another occasion. We are grateful to Joan Bresnan in conversation for alerting us to this issue.

[8] We present the types here in isolation. In a more detailed presentation one would extract some more general types and infer the 'leaves' of the type hierarchy using inheritance.

[9] (Clark, 2007) suggests that such utterances invariably involve recently acquired words. One could, in principle, include such a restriction in the construction description; it is unclear, however, whether carers are sensitive to this.

[10] Segmental parallelism is captured by imposing identity at the *type* level between u1 and the AckRep utterance. The ability to capture such parallelism distinguishes TTR from standard typed feature structure approaches to grammar.

(23)
$$\begin{bmatrix}\text{cat = root : syncat}\\ \text{dgb-params :}\\ \begin{bmatrix}\text{MaxPending : LocProp}\\ \text{u1 :LocProp}\\ \text{c1: member(u1, MaxPending.sit.constits)}\\ \text{A : Ind}\\ \text{B : Ind}\\ \text{c1 : address(A,B)}\end{bmatrix}\\ \text{hd-dtr : }\begin{bmatrix}\text{tune =u1.sit-type.phon : Type}\\ \text{phon : } \textit{tune}\end{bmatrix}\\ \wedge_{merge}\ \text{sign}\\ \text{cont = Acknowledge(A,u1) : IllocProp}\end{bmatrix}$$

VisObjP and VisPred In both cases the visually prominent entity plays a key role. For VisObjP it is simply that *entity InAttention has uttered word's descriptive property*, as in (24a). For VisPred the *property associated with uttered word is predicated of entity InAttention*, as in (24b).[11]

(24) a.
$$\begin{bmatrix}\text{cat = root : syncat}\\ \text{hd-dtr.cont : }\begin{bmatrix}\text{x : Ind}\\ \text{c1 : P(x)}\end{bmatrix}\\ \text{dgb-params :}\\ \begin{bmatrix}\text{v : }\begin{bmatrix}\text{VisSit : RecType}\\ \text{InAttention = hd-dtr.cont : Ind}\\ \text{member(InAttention,VisSit)}\end{bmatrix}\end{bmatrix}\\ \text{cont = P(hd-dtr.cont.x) : Prop}\end{bmatrix}$$

b.
$$\begin{bmatrix}\text{cat = root : syncat}\\ \text{hd-dtr.cont : }(\begin{bmatrix}\text{x : Ind}\end{bmatrix})\text{RecType}\\ \text{dgb-params :}\\ \begin{bmatrix}\text{v : }\begin{bmatrix}\text{VisSit : RecType}\\ \text{InAttention : Ind}\\ \text{member(InAttention,VisSit)}\end{bmatrix}\end{bmatrix}\\ \text{cont = hd-dtr.cont}(\begin{bmatrix}\text{x=InAttention}\end{bmatrix}) : \text{Prop}\end{bmatrix}$$

Short Answer In the adult grammar this is a much discussed construction (Morgan, 1973; Ginzburg and Sag, 2000; Merchant, 2004; Ginzburg, 2012), with controversy whether such constructions are underlyingly sentential or whether the resolution is effected semantically. In this case there is a clear preference for a semantically-based approach since it presupposes less syntactic knowledge for the child: semantically it involves predication—the question predicating of the fragment; it leaves open the issue whether the child is aware of parallelism between the interrogative whP and the fragment, as required in the adult case.[12]

(25)
$$\begin{bmatrix}\text{cat = root : syncat}\\ \text{hd-dtr.cont : }\begin{bmatrix}\text{x : IND}\end{bmatrix}\\ \text{dgb-params :}\\ \begin{bmatrix}\text{max-qud : UnaryWhQuestion}\end{bmatrix}\\ \text{cont = max-qud(hd-dtr.cont.x) : Prop}\end{bmatrix}$$

ReqObj This is a class that is particularly interesting from a semantic point of view as this involves, arguably, the child expressing *non-referential* contents—a request for water does not involve asking for a specific portion, ditto when asking for more (wheat biscuit)—a well known puzzle in semantics first pointed out in (Quine, 1956) and satisfactorily solved in (Montague, 1974).[13] The TTR implementation of the latter analysis is based on (Cooper, 2005), in a non-higher order version proposed in (Ginzburg, 2012).[14,15] The type associated with ReqObj, given in (26a) uses the content of the word uttered by the child as the argument for the (illocutionary) Request predicate. This is exemplified for the utterance 'biscuit' in (26b)—the record type $\begin{bmatrix}\text{x : Ind}\\ \text{c1 : biscuit(x)}\end{bmatrix}$ represents the desire whose fulfillment the child requests—intuitively any witness for that type—an entity that is a biscuit—will do.

(26) a.
$$\begin{bmatrix}\text{cat = root : syncat}\\ \text{hd-dtr.cont = R : RecType}\\ \text{dgb-params : }\begin{bmatrix}\text{A : Ind}\\ \text{B : Ind}\\ \text{c1 : address(A,B)}\end{bmatrix}\\ \text{cont = Request(A,B, R) : IllocProp}\end{bmatrix}$$

b. Child: biscuit $\mapsto$ Request(A,B,$\begin{bmatrix}\text{x : Ind}\\ \text{c1 : biscuit(x)}\end{bmatrix}$)

[11] To avoid notational clutter, we omit the assertoric illocutionary force associated with these utterances and with short answers.

[12] In (Ginzburg and Sag, 2000; Ginzburg, 2012) this parallelism is captured by constraining the category of the head daughter to be identical to the focus establishing constituent, in this case the sub-utterance corresponding to the *wh*-phrase.

[13] For reasons of space we do not discuss the REQ event type here. This would involve a futurate propositional entity such as an *outcome* (Ginzburg and Sag, 2000), presumably including the child as agent.

[14] The witnessing conditions of the record type that fills the object argument role seem to describe well the fulfillment conditions of a desire.

[15] Of course, as Dimitra Kolliakou (p.c.) has pointed out to us, one could argue that the child does not utilize non-referential contents at this stage, exploiting an image of a recent token or some such.

6 The emergence of the initial grammar

Our ultimate aim is not merely to describe the mechanisms of a single period of child utterances in dialogue, but to develop a theory that can accurately describe the transitions between distinct phases. Such a theory should, to the extent possible, explicate this on the basis of interaction between parent and child, given plausible priors and general learning principles. Developing such a theory has been a long term aim since at least (Bruner, 1981).

Here we sketch quite informally for two of the utterance types discussed in section 5.6 some components of an interaction–oriented theory accounting for their emergence.

RepAck once the child understands that she is expected to participate using words, that turns are assigned to her, and can chunk an utterance, those responses of hers that resemble recently uttered words get differentially positive feedback.[16] RepAcks serve as a probe for the child's ability to imitate correctly and with appropriately fast timing. Feedback from the adult causes the child to adjust her hypotheses about a new word. For such hypothesis adjustment using probabilistic representations in TTR see (Cooper et al., 2013).

Short answer The fundamental problem here is to learn the answerhood relation holding between interrogative utterances and (certain classses of) subsequent utterances. The child gets significant data on this from the parent who responds to the parent's own queries—the child receives evidence for several possible answers to a question and several forms, sentential and non-sentential. But this, in turn, presupposes that the child has some means of classifying utterances as wh–interrogative. Morphosyntactically, we can assume this as a prior. But the issue that remains is distinguishing the meaning of different wh–words, at this stage *where, what, who*, as well as combining these with predicates. For the former, we hypothesize this can be done on the basis of utterance bigrams linking *where*–utterances with deictic gestures/locative utterances to entities in the visual field, in contrast to *what*–utterances which are differentially linked to utterances supplying attributes.

[16] *MOT: that baby has a bottle , did you notice that ? *CHI: bottle . *MOT: yeah , baby has a bottle. (from: Naima03)

7 Conclusions and Further Work

In this paper we develop a taxonomy of early child utterances that allows the fine structure of the semantic content of such utterances to be represented, thereby remedying problems for existing classifications. We offer a formal analysis of such utterances in the frameworks of KoS and HPSG$_{TTR}$. This requires spelling out the dialogue context and interaction since such utterances are strongly context dependent. We also provide a brief sketch of how two classes of such utterances could be acquired interactively.

In ongoing work, we are refining the taxonomy to incorporate gesture and to scale up to later, more complex utterances. We also intend to implement a learning algorithm which will allow us to experimentally test the interactive account of acquisition of certain early utterance types, above all short answers.

References

John Langshaw Austin. 1975. *How to Do Things with Words*. Oxford University Press.

Elizabeth Bates. 1976. *Language and context: The acquisition of pragmatics*, volume 13. Academic Press New York.

Lois Bloom. 1967. Language Development: Form and Function in Emerging Grammars. November.

Jerome Bruner. 1975. The ontogenesis of speech acts. *Journal of Child Language*, 2(01):1–19, September.

Jerome Bruner. 1981. The social context of language acquisition. *Language & Communication*, 1(2):155–178.

Eve V. Clark. 2007. Young children's uptake of new words in conversation. *Language in Society*, 36:157–182.

Eve Clark. 2013. Common ground. In Brian MacWhinney and William O'Grady, editors, *The Handbook of Language Emergence*. Wiley-Blackwell, London.

Robin Cooper, Simon Dobnik, Staffan Larsson, and Shalom Lappin. 2013. Probabilistic type theory and natural language semantics. *Under Review*, xx:xx.

Robin Cooper. 2005. Austinian truth, attitudes and type theory. *Research on Language and Computation*, 3(4):333–362.

Katherine Demuth, Jennifer Culbertson, and Jennifer Alter. 2006. Word-minimality, epenthesis and coda licensing in the early acquisition of english. *Language and Speech*, 49(2):137–173.

John Dore. 1974. A Pragmatic Description of Early Language Development. *Journal of psycholinguistic Research*, 3(4):343–350.

John Dore. 1975. Holophrases, speech acts and language universals. *Journal of Child Language*, 2(01), September.

Raquel Fernández and Jonathan Ginzburg. 2002. Non-sentential utterances: A corpus-based study. *Traitement automatique des languages*, 43(2):13–42.

Tanya M Gallagher. 1977. Revision behaviors in the speech of normal children developing language. *Journal of Speech, Language and Hearing Research*, 20(2):303.

Jonathan Ginzburg and Raquel Fernández. 2010. Computational models of dialogue. In Alex Clark, Chris Fox, and Shalom Lappin, editors, *Handbook of Computational Linguistics and Natural Language*, Oxford. Blackwell.

Jonathan Ginzburg and Ivan A. Sag. 2000. *Interrogative Investigations: the form, meaning and use of English Interrogatives*. Number 123 in CSLI Lecture Notes. CSLI Publications, Stanford: California.

Jonathan Ginzburg. 2012. *The Interactive Stance: Meaning for Conversation*. Oxford University Press, Oxford.

Barbara Grosz. 1977. *The Representation and Use of Focus in Dialogue Understanding*. Ph.D. thesis, Stanford University.

Ray Jackendoff and Eva Wittenberg. 2014. What you can say without syntax: A hierarchy of grammatical complexity. In Fritz Newmeyer and Laurel Preston, editors, *Grammatical Complexity*. Oxford University Press, New York, NY.

Barbara Kelly. 2011. A new look at redundancy in children's gesture and word combinations. In Inbal Arnon and Eve Clark, editors, *Experience, Variation and Generalization: Learning a First Language*. John Benjamins, Amsterdam.

Staffan Larsson and Robin Cooper. 2009. Towards a formal view of corrective feedback. In *Proceedings of the EACL 2009 Workshop on Cognitive Aspects of Computational Language Acquisition*, Athens.

Staffan Larsson. 2002. *Issue based Dialogue Management*. Ph.D. thesis, Gothenburg University.

Zoran Macura. 2007. *Metacommunication and Lexical Acquisition in a Primitive Foraging Environment*. Ph.D. thesis, King's College, London.

Brian MacWhinney. 2000. The childes project: Tools for analyzing talk: Volume i: Transcription format and programs, volume ii: The database. *Computational Linguistics*, 26(4):657–657.

Jason Merchant. 2004. Fragments. *Linguistics and Philosophy*, 27:661–738.

Richard Montague. 1974. The proper treatment of quantification in ordinary english. In Richmond Thomason, editor, *Formal Philosophy*. Yale UP, New Haven.

Jerry Morgan. 1973. Sentence fragments and the notion 'sentence'. In B Kachru, editor, *Issues in Linguistics: papers in honour of Henry and Rene Kahane*. UIP.

Anat Ninio and Jerome Bruner. 1978. The achievement and antecedents of labelling. *Journal of Child Language*, 5(01):1–15, September.

Anat Ninio and Catherine Snow. 1988. Language Acquisition Through Language Use: The Functional Sources of Children's Early Utterances. *Y. Levy, I. Schlesinger, & M.D.S. Braine (Eds.), Categories and processes in language acquisition*, pages 11–30.

Anat Ninio and Polly Wheeler. 1986. A manual for classifying verbal communicative acts in mother-infant interaction. *ERIC*, 3:1–12.

Anat Ninio, Catherine E. Snow, Barbara A. Pan, and Pamela R. Rollins. 1994. Classifying communicative acts in children's interactions. *Journal of Communication Disorders*, 27.

Anat Ninio. 1992. The Relation of Children's Single Word Utterances to Single Word Utterances in the Input*. *Journal of Child Language*.

Massimo Poesio and Hannes Rieser. 2011. An incremental model of anaphora and reference resolution based on resource situations. *Dialogue and Discourse*, 1:235–277.

Willard V Quine. 1956. Quantifiers and propositional attitudes. *the Journal of Philosophy*, 53(5):177–187.

Ivan A. Sag, Thomas Wasow, and Emily Bender. 2003. *Syntactic Theory: A Formal Introduction (second edition)*. CSLI, Stanford.

John R. Searle. 1969. *Speech Acts: An Essay in the Philosophy of Language*, volume 0. Cambridge University Press.

Marilyn Shatz. 1979. How to Do Things by Asking : Form-Function Pairings in Mothers' Questions and Their Relation to Children's Responses. *Child Development*, 50(4):1093–1099.

Catherine E. Snow. 1972. Mothers' Speech to Children Learning Language. *Child development*, 43(2):549–565.

Michael Tomasello. 1988. The role of joint attentional processes in early language development. *Language Sciences*, 10(1):69–88, January.