The MINGLE annotation scheme: Multimodal annotation of parent-child interaction in a free play setting (version 1.0)*

Kristina Nilsson Björkenstam†
Department of Linguistics, Computational Linguistics
Stockholm University
9 januari 2013

1 Introduction

“In order to [reduce] the complexity of the grammar induction problem it is necessary to construct multimodal data bases in which nonlinguistic aspects of interaction ... are encoded in enriched representations of the PLD [primary linguistic data]” [5, p. 207]

Recent data and theory posits that language acquisition is grounded in innate general cognitive abilities such as intention-reading and sensitivity to joint attention, (visual and auditory) pattern-finding, and the ability to imitate [24, 14]. Research also shows the impact of social interaction on language acquisition in infants [14, 15]. The parent’s ability to establish and maintain joint attention, read the child’s signals (e.g., eye gaze), and to align his/her speech with the child’s focus of attention are key elements in the interactive learning process [8, 22].

Thus, a cognitive model of language learning must be dialogue-driven and multimodal to reflect how parent and child interact, using words, eye gaze, and object manipulation. We present a scheme for multimodal annotation of parent-child interaction. The purpose is to add verbal and non-verbal annotation to a corpus of longitudinal video and sound recordings of parent-child dyads. In this guideline, we describe the transcription of adult and child speech and vocalizations, and the annotation of both empty-hand gestures and object-related actions by both adults and children.

*This research is part of the project “Modelling the emergence of linguistic structures in early childhood”, funded by the Swedish Research Council as 2014-0555-50015-31.
†Contact: kristina.nilsson@ling.su.se
Adults modify their speech when talking to infants, for example by using fewer words per utterance, more repetitions, and decreased syntactic complexity. Prosodically, infant-directed speech (ID-speech, or “motherese”) is characterized by high affective salience and specific adjustments of acoustic characteristics such as higher F0, wider F0 excursions, distinctive F0 contours, slower tempo, and longer pauses.¹ Adults use ID-speech to regulate infants’ arousal levels, convey affective intent, and capture their infants’ attention; and studies show that infants prefer ID-speech to adult-directed speech [9, 12].

This modification is not modality-specific: signing caregivers typically sign slowly to infants, with high levels of repetition and exaggerated movements [16], and both deaf infants and hearing infants without previous exposure to sign language show greater attentional and affective responsiveness to infant-directed signing than adult-directed signing [17, 18].

Adults also modify their gestures when talking to young children, with frequent use of concrete pointing, display of objects, demonstration and enactment, and emblems (culturally defined gestures). Gesture types that are common in adult interaction such as representative, metaphoric, and emphatic gestures are less frequent in ID-speech [19]; [11] found that mothers used gestures less frequently in interaction with their infants as compared to interaction with adults, and that the ID-gestures reinforce and disambiguate spoken utterances, rather than add new information.

This type of “gesture” is mirrored in what Brand and colleagues [4] call “mitionese”: in a study of voluntary bodily movement with and without objects in communication with infants or adults, they focused on eight features: range of motion, rate, repetitiveness, proximity to partner, enthusiasm, interactiveness, punctuation,² and simplification, and found that mothers significantly modify their actions to infants, as compared to adults. In a related study, [13] found that infants respond to motionese; Variations in amplitude and repetition affect infant attention, while level of repetition affect the way the infant explored the objects.

In this project, we want to investigate both unimodal (e.g., repetition of utterances) and multimodal synchrony in an interactive setting, based on annotated data.

This annotation guideline is structured as follows. Section 2 describes the data: the participants and the objects, the setting, and the longitudinal data. An overview of the multimodal annotation structure is presented in section 3. The annotation process and the different annotation tasks are described in section 4: the verbal annotation in section 4.1, the first step of discourse segmenting in section 4.2, the non-verbal annotation in section 4.3, and the second step of discourse annotation in section 4.4. Finally, in section 5, different methods for analysis of the annotated data are outlined.

¹But see [10] for a discussion of cultural aspects on pitch.
²[4] use the word “punctuation” to mean segmentation of discourse.
2 Data

2.1 Participants and objects

- Participants: Parent (mother or father) and Child
- Target objects: Siflu (a brown toy monkey) and Kucka (a yellow toy rabbit)
- Other objects: for example a small bag, a toy bucket, a hat, dresses for the toys, a toy bottle, a jigsaw puzzle, a farm house with animals, a book.

2.2 Setting

Data consists of video recordings (using two cameras) of free play sessions in a recording studio at the Phonetics laboratory at Stockholm University, and audio recordings (via wire-less microphones). The studio is equipped with a set of toys, including the target objects, and the parents are asked to play with these toys. The parent is told the names of the target objects at the beginning of the session.

2.3 Video and audio data

The data consist of longitudinal dyads at 7-8, 12-13, 17-19, and 27-31 months. The length and “information density” of the recordings vary greatly depending on the mood of both the child and the parent. We also see differences in the kind of play due to development.

2.4 Annotation tool

ELAN. URL: http://tla.mpi.nl/tools/tla-tools/elan/

3 Multimodal annotation of parent-child interaction

Overview of the current structure:

- Discourse annotation
  - Segments (ELAN tier: Segment)
  - Repetition between parent and child (ELAN tier: Repetition)
  - Focus and Initiative (ELAN tier: Focus)

- Verbal annotation
  - Transcription parent (ELAN tier: Parent)
  - Transcription child (ELAN tier: Child)
• Object-mention parent (ELAN tier: P-Speech)
• Object-mention child (ELAN tier: C-Speech)

• Non-verbal annotation
  – Eye gaze parent (ELAN tier: P-Gaze)
  – Eye gaze child (ELAN tier: C-Gaze)
  – Object-related actions parent (ELAN tier: P-ObjAct)
  – Object-related actions child (ELAN tier: C-ObjAct)
  – Gesture function parent (ELAN tier: P-GestFunc)
  – Gesture function child (ELAN tier: C-GestFunc)

• Analysis
  – Analysis of synchrony across modalities – parent based on P-Obj, P-Gaze, P-ObjAct, P-GestFunc. (ELAN tier: P-Analysis)
  – Analysis of synchrony across modalities – child, co-temporal with P-Obj; analysis based on C-Obj, C-Gaze, C-ObjAct, C-GestFunc. (ELAN tier: C-Analysis)

4 The annotation process

The order of the annotation layers presented here differs from the order in the previous section. This section describes the practical annotation work, and therefore the different layers are presented in the order of the annotation process.

Please note: All tiers must have the name listed here! Compare with previous annotations, use the ELAN search function, discuss difficult cases with your colleagues. Use a comments tier!

4.1 Verbal annotation

All utterances of parents and children are transcribed, using two separate ELAN tiers. We also mark each verbal mention of the target objects (object mentions) by either parent or child in two separate ELAN tiers.

4.1.1 Transcription parent

ELAN tier: Parent

An utterance is a vocal contribution from the parent. The scope of an utterance is determined from the speech signal, where the goal is to make the annotation scope as precise as possible without losing any acoustic information. The scope of an utterance is ultimately determined by the transcriber, who must take into consideration both the context of each utterance, the situation (free play, interaction between parent and child), and difference between speakers, etc.
The transcription is orthographical, with additional labels for features like laughter, onomatopoeia, and disfluency. The labels are listed in 1 on page 15. A set of shorthand labels for common features are listed in Table 2 on page 16. Additional labels may be added. If so, discuss with the team members and add the new label to this list.

Utterances that the transcriber interprets as exclamations and/or appeals/orders are marked with an exclamation mark, and utterances interpreted as questions are marked with a question mark. No other punctuation is used. All words are written in small caps, except for proper names, e.g.:

- kommer du ihåg vem Sifru var då?
- ja! vem var det?

For the sake of disambiguation (and subsequent POS-tagging), some common words with spelling variants are transcribed as follows (based on the standard resource for Swedish, Stockholm-Umeå Corpus):

- sedan/sen → sen
- de/dom; dem/dom → de; dem
- något/nåt; någonting/nånting → något; någonting
- sade/sa → sa
- va/var/vara → vara, var
- varann/varandra → varandra
- sån/sådan; sånt/sådant; såna/sådant → sån, sånt, såna

4.1.2 Transcription child
ELAN tier: Child

An utterance is a vocal contribution from the child. Vocalizations are phonetically transcribed, until later recordings when the child starts to utter words that can be orthographically transcribed. In later recordings, the child’s vocalizations are transcribed as a combination of orthographic words and non-word vocalizations.

4.1.3 Repetition between parent and child
ELAN tier: Repetition

This tier marks repetitions of (part of) utterances between parent and child, e.g., between parent and child (annotation value PC):
Parent: Kucka
Child: guga

Such repetitions can be e.g., imitations, repetitions where the first utterance is expanded when repeated (see the example below), or repair sequences. Repetitions can also occur between child and parent (annotation value CP):

C: där!
P: ja där är Sifu

This tier is added during verbal and vocal transcription. The function of each occurrence will be categorized at a later date – at this time we annotate the scope and the order of participants (e.g., parent-child). The annotation value describes initiative and order (in a CP sequence, the parent repeats the child’s utterance, and vice versa), and length (two repetitions: CPC, three repetitions: PCPC, etc.).

4.1.4 Object-mention parent

ELAN tier: P-Speech
We annotate each verbal mention of either of the target objects by the parent. Example mentions are the names Kucka and Sifu, definite descriptions such as kaninen (“the rabbit”), apan (“the monkey”), or pronouns den (“it”), han (“he”), hon (“she”). The whole NP is added as the annotation value.

Annotation value: S(iflu), K(ucka) + in cases of non-NEs the NP (e.g., kaninen, den, hon), e.g., S han.

4.1.5 Object-mention child

ELAN tier: C-Speech
We also annotate each verbal mention of either of the target objects by the child.

Annotation value: S, K + NP in case of non-NE

4.2 Discourse annotation I

4.2.1 Segments

ELAN tier: Segment

A discourse segment is an interval of a dyad in which any or both of the target objects are in focus. At present, “in focus” means that the target object has been orally referred to by the parent (using names, definite descriptions, or pronouns) at least once during the segment.

The segment starts when the target object is brought into focus by either participant by speech, eye gaze, or hand movement, and ends when focus is
shifted (more permanently) to other objects, by e.g., one of the participants removing the object from view (by throwing, putting away, etc) and introducing a new object. We want to cover the whole interaction so make sure the scope of the segment includes all verbal and non-verbal actions. At present, all non-verbal annotation (described below) is restricted to these segments.

Annotation value: the target object(s) in focus, i.e., Kucka, Siffu, or Kucka+Siffu (alt. Kucka/Siffu).

4.3 Non-verbal annotation

The non-verbal annotation includes eye-gaze, object-related actions, and gestures. Important: turn off the sound and hide the transcription tiers before starting with the non-verbal annotation! Each annotation task should be performed independently of the other tasks, with the results of previous annotation hidden from the annotator.

Traditionally in maternal gesture research, only communicative gestures are selected and coded (see e.g., [20], [11]). Communicative gestures are “accompanied by eye contact with an interactive partner, vocalization, or other clear evidence of an effort to direct the attention of another person in the room” [11, p 63-63]. While gestures involving objects, such as holding up an object toward the partner (SHOW) or touching/tapping an object (INDICATING) can be considered gestures as they communicate information to the listener, they are typically excluded from analysis (see e.g., [11, p. 69] or [19, p 327]). For a different view, see e.g., [21] and [2].

In this annotation scheme, we do not differentiate between communicative and non-communicative gestures during annotation; such classification must be done from a complete annotation. We do, however, differentiate between empty-hand gestures and object-related actions (see below). From a “traditional” gestural research perspective some of the object-related actions can be categorized as manipulative forms of gesture deixis [19, p 327] but we do not classify them during annotation; again, such classification must be done from a complete annotation.

4.3.1 Eye gaze parent

ELAN tier: P-Gaze

We annotate eye gaze by marking whether the parent is looking at the child (annotation value C), the target objects Siffu and Kucka (annotation values S, K), or on any other object (annotation value o). In some cases, these other objects are named, e.g., if the parent is looking at a bag containing the target object Siffu we annotate gaze as “bag_w_S”.

In cases where the parent is looking at the child and the child is holding one of the objects, we annotate gaze as directed at the “larger frame”, i.e., the child. We can extract the fact that the object is also in the parent’s line of sight from
the annotation of the child’s object related actions (that is, that the child is holding the target object).

Annotation values: C, S, K, o

4.3.2 Eye gaze child
ELAN tier: C-Gaze

We annotate the child’s eye gaze as directed at the parent (annotation value P), the target objects Siffu and Kucka (annotation values S, K), or on any other object (annotation value o) in the room. Other objects may be named (see 4.3.1).

Annotation values: P, S, K, o

4.3.3 Object-related actions parent
ELAN tier: P-ObjAct

Object-related actions are hand movements involving objects, including the child (annotation value C), the target objects Siffu and Kucka (annotation value S, K), and any other object (annotation value o). The movements are labeled as Verb_Object or Verb_Object1_Object2. Below, the movement labels are grouped for readability:

- Taking an object
  - Reach_S: the parent reaches for the monkey Siffu (n.b., the hand is moving towards Siffu, but the hand does not yet touch the toy; object-related actions does not equal handling objects!). Reaching is typically followed by grabbing, picking-up and moving the object, but this is not always the case. Thus, we annotate reaching as a separate action from e.g., moving a toy.
  - Grab_S: the parent grabs Siffu with one (or two?) hand(s?), but does not pick/lift the object up (we annotate that separately).
  - Pick-up_S: the parent picks up Siffu.
  - Move_S: the parent moves Siffu from one place to another (i.e., this is about change of location).
  - Hold_K: the parent holds the toy in his/her hands, but does not play with it and is (often) looking at something else.

- Initiating play
  - Show_S: the parent shows Siffu to the child by holding it up in front of the child.
- Shake _S_: the parent holds S up before the child, and shakes/looms the toy.
- Enact _S_: the parent makes S jump, walk, climb, interact with C, etc.
- Tickle _C_ _K_: the parent tickles the child with a tickler

• Play

  - Explore _K_: the parent is exploring the toy, often turning it over or focusing on some feature of the toy (a button, a hat).
  - Manipulate _S_: the parent is manipulating the toy in some way not covered by other movement labels.
  - Touch _S_: the parent touches the object (often with a kind of deictic function, e.g., taps).
  - Peek-hide _S_: Peek-a-boo with Siffu. Step 1: hiding Siffu.
  - Peek-show _S_: Peek-a-boo with Siffu. Step 2: showing Siffu.

• Handing over

  - Put-down _S_: the parent puts the toy “on display”, typically next to or in front of the child on the floor or on a chair, and the toy is still in play/focus.
  - Offer _S_: the parent offers Siffu to the child by holding it close to the child(‘s hands), making it easy for the child to grab the toy.
  - Hand _C_ _S_: the parent hands over Siffu to the child.

• Ending play

  - Put-away _S_: the parent puts the toy on the floor (typically out of reach and/or sight) of the child, and the toy is clearly no longer in focus.

At present, the annotator decides what to call an action, but after analysis of the annotated data we will decide on a set of defined actions to choose from (primarily).

Since the parent has two hands, he/she can perform different actions with both hands at the same time. In such cases, we annotated both actions combined by the plus sign (“+”), e.g.

  - Reach _S_ + Hold-hand _C_: The parent is holding the child’s hand while reaching for Siffu

In such cases, the primary action (above: reaching) is listed first, and the secondary action is listed second. Typically, the primary action is the most recent and the most active. In unclear cases, add “??” and a comment and discuss with the team.
4.3.4 Object-related actions child

ELAN tier: C-ObjAct

Object-related actions are hand movements involving objects, including the parent (annotation value P), the target objects Siffin and Kucka (annotation value S, K), and any other object (annotation value o). The movements are labeled as Verb/Object, or Verb/Object1/Object2.

Some of the child’s object-related actions are the same as the parent’s actions, so there is an overlap in annotation values between the tiers P-ObjAct and C-ObjAct. But some actions are more frequently (and in some cases exclusively) performed by the children, for example:

• Taking an object
  – Reach/S: the child reaches for the monkey Siffin
  – Grab/S: the child grabs the toy with one or two hands (depending on developmental stage)
  – Pick-up/S: the child picks up Siffin
  – Move/S: the child moves Siffin from point A to B
  – Hold/S: the child holds Siffin, but does not interact with the toy

• Play
  – Touch/K: the child is touching Kucka (finger tips, palm of hand, etc)
  – Taste/K: the child puts the toy to his/her mouth (typically, we cannot see from the video footage if the child really is tasting the toy, but we interpret this action as an intention to taste it)
  – Explore/S: the child is holding the toy while examining it closely, typically turning it around and touching different parts of the toy (eyes, ears, buttons, etc)
  – Hover/K: the child is holding his/her hand over the toy - not quite touching; the hand is “hovering” above the toy
  – Wave/S: The child is holding the object and for example waving/shaking it in the air
  – Mani/S: (manipulate) The child is e.g., banging the toy against something (another toy or the floor)
  – Feed/S:bottle: child is feeding Siffin from a toy bottle

• Ending play
  – Drop/S: the child lets go of the toy, and the toy falls to the floor.
  – Put-away/S: the child puts the toy away/out of reach; the toy is clearly no longer in focus.
In cases where another object is used to perform an action, e.g., brushing someone’s hair with a toy hair brush, and the action verb implicitly names this object it is not annotated:

- Brush_S: Child is brushing SIFFU with a toy hair brush

### 4.3.5 Gesture function parent

ELAN tier: P-GestFunc

We annotated gesture function according to categories commonly used in multimodal annotation of adult interaction [19, 1] and adult-child interaction [20, 23]. All visible actions are coded, that is, we do not classify gestures as communicative and non-communicative at this stage. At present, we do not annotate gesture form (i.e., prep, stroke, hold) [19] but this can be added at a later date.

Because the subject is parent-child interaction (see section 2), we will probably not meet all the gesture types described here.

- Pointing (with finger or hand).

Value: Point-(Index|Hand|0) referent-name (target area)

- Point-Index S: for pointing at SIFFU with the index finger
- Point-Hand C (tummy): for pointing with hand toward the tummy of the child

We annotate all pointing gestures, and differentiate between abstract pointing and concrete pointing (to a referent in the physical context) [19] by adding the name of the referent. In some cases, we also annotated the target area (tummy, head, shoe, etc). Note that in some annotation schemes, the pointing form can have different functions, e.g., deictic, discursive, and representational [7, 6]. At present, we do not annotate pointing function other than abstract/concrete.

- Representational gestures: this category consists of two types of gestures: action gestures/enactment where body parts enact a pattern of action, and size-shape gestures that depict characteristics of an object or event (see e.g., [20, 23]). This category is also called *iconic gestures* by e.g., [19] and [1].

- Conventional, symbolic gestures (emblems): culture-specific gestures where the form-function relation is based on convention, e.g., thumbs-up “OK”, wave “hejda” [1].

- Emphatic: rhythmic beats/batons. Simple, repetitive movements where e.g., the hand moves in time with the speech marking the rhythm [19].

Here, we also mark grooming gestures, e.g., adjusting clothes, hair, etc.
• Grooming

We also annotate clear shifts in posture that indicate where the parent’s attention is:

• Lean fwd: parent is leaning forward (often towards the child, or the “play area”)

• Lean twd K: parent is leaning toward Kucka

In short: annotate all empty-hand gestures, including those that do not fall into any of these categories. Add a comment and “??” for further analysis.

4.3.6 Gesture function child
ELAN tier: C-GestFunc

The gesture function annotation of the child is identical to that of the parent (see 4.3.5): we mark pointing (see e.g., [25]), as well as any gestures that can be classified as representational/Iconic, conventional, and emphatic gestures (above). We also annotate grooming and shifts in posture that indicate focus of attention.

Note that due to the age of the participants, it is unlikely that we will find abstract pointing and emphatic gestures. Concrete pointing is reported from 11 to 12 months [25] and representational gestures from 12 months [19].

Note also that children use their whole bodies for gestures, by for example pointing with a foot [19].

4.4 Discourse annotation II

4.4.1 Focus and Initiative
ELAN tier: Focus

This tier marks shifts in focus of attention.

Object mentions (P-Speech) are categorized as initial-mention (annotation value: 1) and subsequent-mentions (value: 2). We are interested in finding mentions where the parent is introducing or reintroducing an object into focus (initial-mentions), and mentions where the parent is co-refering to an object already in focus (subsequent-mentions).

A shift in focus is detected by jointly examining the utterances and actions of the parent, and the child’s attention marked by speech, gaze, and actions. Is the focus of attention introduced/reintroduced or upheld? For example:

• The first time an object is mentioned is marked as an initial-mention.

• Reintroductions occur when the child’s focus is on an object, but then there is a clear shift of the child’s focus to something else, and the parent
mentions and shows (vocal and non-vocal reintroduction) the first object to the child.

- If the parent has mentioned Kucka, the child is focused on Kucka, and the parent mentions Kucka again, this is a subsequent-mention.

Copy the P-Speech tier, rename it to Focus, and change the values of each mention to either 1 or 2. Annotation values:

- Initial-mention: 1
- Subsequent-mention: 2

Each instance annotated as an initial-mention (1) is further annotated for initiative as either a bring-in (when the parent introduces the object) or a follow-in (when the child introduces the object and the parent follows). Since we are annotating the mentions of the parent, there are no non-vocal initiatives. Annotation values:

- Bring-in: The parent initiates by vocal (and possibly also non-vocal) reference to the object. The child may or may not follow. Additional values:
  - voc: vocal reference to object
  - voc+non-voc: a combination of vocal and non-vocal (object action and/or gesture) references
- Follow-in: The child initiates by vocal and/or non-vocal reference to the object, and the parent answers with a vocal (and possibly non-vocal) reference to the object:
  - C: combinations of voc and/or gaze and/or hand (e.g., voc+gaze+hand)
  - P: voc or voc+non-voc

At present, we do not mark focus for object mentions by the child (C-Speech). This can be added at a later date.

5 Analysis

Our basic hypothesis is that perception of synchrony in multimodal patterns in auditory-visual speech is the device primarily used to reduce complexity in language learning; this analysis is further discussed in [3].

5.1 Analysis of synchrony across modalities

5.1.1 Analysis parent

ELAN tier: P-Analysis
The analysis of invariance combines each utterance of a reference to one of the target objects (ELAN tier P-Speech) with eye gaze (tier P-Gaze), object-related actions (tier P-ObjAct), and gesture functions (tier P-GestFunc) into one pattern. The pattern consists of four digits, either 1 for matching annotation values or 0 for non-matching values. The pattern is composed as follows:

- The first digit in P-Analysis patterns is always 1, since P-speech is always a match.
- If the referent of the utterance in P-speech is Kucka, and eye gaze is directed at Kucka, the annotation values match, and the second digit is a 1. If gaze is not directed at Kucka, the second digit is 0. In cases where eye gaze is directed at the child, and the child is holding an object, this must be considered. In difficult cases, return to the video recording.
- If there is an annotation on tier P-ObjAct, and the object is Kucka (e.g., Hold_K), the third digit is a 1. If there is no object-related action at all (i.e., there is no annotation on tier P-ObjAct), the third digit is 0. If there is an annotation but the object is not Kucka, the third digit is 0.
- If there is a gesture directed at Kucka, the final digit is 1. If the gesture is directed elsewhere, or there is no gesture at all, the final digit is 0.

For example, the pattern 1 1 0 0 means that the parent says Kucka, and looks at Kucka at the same time. The pattern 1 0 1 0 means for example that the parent says Sifiu and shows Sifiu to the child at the same time.

5.1.2 Analysis child

ELAN tier: C-Analysis

The analysis for the child is similar to the analysis of the parent, with one difference: they child must utter a reference to the object (C-Speech) immediately prior to the (parental) utterance in which P-Speech occurs. There is no strict time limit, but the two utterances must be part of a cohesive stretch of discourse.

The pattern for C-Analysis consists of analysis of the annotation values of the tiers C-Speech, C-Gaze, C-ObjAct, C-GestFunc in combination with the current P-Obj.

Referenser


<table>
<thead>
<tr>
<th>Label</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>??</td>
<td>A note from the annotator to check the transcription.</td>
</tr>
</tbody>
</table>
| ?,(?) | Ambiguity. Can be combined with both labels and words, for example: #VR? ha(?)  
In order to disambiguate from the question mark, add a parenthesis. |
| #     | Descriptions (see the list of feature labels in table 2). The label is placed  
before the utterance in question, for example:  
#IN dår!  
In cases where the description covers part of an utterance, the scope of  
the description is marked by an additional # at the end, for example:  
#VI hej Emil # hej pojken! |
| ()    | Description of non-speech/vocalization combined with word. Example:  
(h)oj! |
| &     | False starts (words), for example:  
och &sko kolla hår då |
| &{(fras)} | False starts (phrases), for example:  
det har vi sagt &förf att han hela tiden eftersom han alltid dreglar  
på fjärrkontrollen” |
| xx    | One unknown word, for example:  
ta den då ta xx bollen  
There can be more than one unknown word in an utterance, for example:  
ta den då ta xx xx” |
| xxx   | A sequence of unknown words, for example: titta bollen xxx |
| *(a-z) | Non-words with a communicative function, e.g., “huh?”, mimicry  
or vocal illustration (“nämnamn” for eating), or sound effects (“hå!”)  
are marked with an asterisk, for example:  
*(nämnamn) |
| :     | Long sounds are marked with a colon, for example: hå: |
| Ex ee | Filled pauses are transcribed as they sound, for example:  
ee mm hm aa öh |
| _     | Hesitation disfluencies when part of a word, e.g., “lijja det tror jag”.  
“neesij” are marked as “i_a det tror jag”, “ne_i” |

Tabell 1: Labels used for transcription.
<table>
<thead>
<tr>
<th>Label</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>#VR</td>
<td>adult-directed speech (typically interaction with the experiment leader); Child-directed speech is default</td>
</tr>
<tr>
<td>#VI</td>
<td>whisper</td>
</tr>
<tr>
<td>#IN</td>
<td>ingestive speech</td>
</tr>
<tr>
<td>#LA</td>
<td>word or phrase in another language. If a phrase, mark scope with a parenthesis, e.g., ‘‘#LA (the thing)’’</td>
</tr>
<tr>
<td>#SK</td>
<td>laughter</td>
</tr>
<tr>
<td>#FL</td>
<td>panting</td>
</tr>
<tr>
<td>#GR</td>
<td>crying, whining</td>
</tr>
<tr>
<td>#SA</td>
<td>singing, humming</td>
</tr>
<tr>
<td>#VL</td>
<td>whistling</td>
</tr>
<tr>
<td>#PU</td>
<td>kissing noise</td>
</tr>
<tr>
<td>#SM</td>
<td>slurping noise, smacking one’s lips (sound, not vocalization)</td>
</tr>
<tr>
<td>#HO</td>
<td>coughing</td>
</tr>
<tr>
<td>#HR</td>
<td>clearing one’s throat</td>
</tr>
<tr>
<td>#LS</td>
<td>noise, disturbance</td>
</tr>
</tbody>
</table>

Tabell 2: Shorthand labels for features.


