

Disfluency in child-directed speech

Kristina Nilsson Björkenstam¹, Mats Wirén¹ & Robert Eklund²

¹Department of Linguistics, Stockholm University, Stockholm, Sweden

²Department of Culture and Communication, Linköping University, Linköping, Sweden

Abstract

We report results from a longitudinal study of the rate and location of disfluencies in child-directed speech, using data for children between 0;6 and 2;9 years. We compare these results to adult-directed speech by the same speakers.

Introduction

From a language acquisition perspective, disfluency (for example, “uh” and “um”) is interesting because it could arguably make learning harder. Put differently, it looks like yet another manifestation of the poverty of the stimulus. Seen from this perspective, it is natural that child-directed utterances are not only short and slow, but also highly fluent compared to adult-directed speech (ADS). Even though the adult disfluency rate increases with the age of the child, child-directed speech (CDS) is consistently less disfluent than ADS (Broen, 1972). However, it has recently been shown that disfluencies contain information that helps the child to interpret the input from a certain age: disfluencies tend to occur before words that are unfamiliar, infrequent or new in the discourse, and thereby provide a cue about a speaker's intended referent or communicative intention (Kidd, White & Aslin, 2011). To corroborate this finding, we must begin by investigating the disfluencies that children hear at different ages. To this end, we report results from a longitudinal study of the rate and location in utterances of disfluencies in child-directed speech, using data for children between 0;6 and 2;9 years.

Fluency and disfluency in child-directed speech

Spontaneous speech in adult–adult conversations typically includes disfluencies such as filled pauses, segment prolongations, hesitations, repetitions, and truncated words at a rate of about 6% of all words uttered (Eklund & Wirén, 2010; Fox Tree, 1995).

When talking to young children, adults modify their speech, e.g. by using fewer words per utterance, slower speech rate, more

repetitions, and decreased syntactic complexity compared to ADS (Broen, 1972). Typically, CDS is described as *fluent* speech (Clark, 2009:36). Over time, as caretakers use longer, more complicated utterances at a faster speech rate, the disfluency rate increases accordingly; Kidd, White and Aslin (2011) report that filled pauses occur at a rate of 1/1000 words in speech directed at 2-year olds in the CHILDES database, and that this rate increases with the age of the child. This can be compared to a reported filled pause incidence of 1.9% to 4.4% in scientific works covering the period 1959 to 2007 (Eklund, 2010:25).

The most prevalent type of disfluency is the filled pause (FP), e.g. *um*, *öh*. Eklund and Wirén (2010) list five hypotheses regarding the function(s) of FPs in speech:

- 1) Floor-holding hypothesis
- 2) Help-me-out hypothesis
- 3) Self-monitoring/error-detection hypothesis
- 4) Many-options hypothesis
- 5) Attention-getting signal

Eklund and Wirén (2010) point out that these hypotheses are not mutually exclusive and that FPs may serve more than one function, but that there is strong support for the many-options hypothesis. In a CDS scenario, the first two hypotheses are less likely than the latter three since the adult is typically very attentive to vocalizations by the child.

Corpus data

The data consist of audio and video recordings of free play sessions in a recording studio at the Phonetics laboratory at Stockholm University. The free play sessions are in most cases followed by a session when the parent and the experiment leader chat informally while working through The Swedish Early Communicative Development Inventory (SECDI, a version of the MacArthur Communicative Development Inventory) with the child in the room.

The data consist of 31 recordings of four children (age 6–33 months), three girls and one boy, interacting with their Swedish-speaking mothers or fathers (mean recordings/child 7, range 11–5).

All utterances by both parent and child in these audio and video recordings have been transcribed using ELAN. The utterances by the parents have been orthographically transcribed, with additional labels for features like laughter, onomatopoeia, and disfluency according to the MINGLE annotation guidelines (Nilsson Björkenstam, 2012). Utterances interpreted as exclamations, appeals, or orders are marked with an exclamation mark, and questions with a question mark. Utterances interpreted as adult-directed are labeled as such, while the default is child-directed speech. A subset of this data, named MINGLE-2, has also been annotated with eye gaze, hand gestures, and object-related actions (Nilsson Björkenstam & Wirén, 2012).

MINGLE-4 consists of a total of about 59600 words, with about 24100 words ADS, and 35500 words CDS. Due to the set-up of the experiment these recordings originate from, there is little (or in some cases no) ADS in sessions recorded with older children (>16 months). The CDS word average per session is 1145 (range 565–2305), while the ADS average is 778 (range 0–4203).

Disfluency annotation

In MINGLE-4, the following disfluency categories are annotated: truncated words and phrases, prolongations, hesitations, and filled pauses. Below, examples from both CDS and ADS are presented.

Truncated words (marked by &word):

- 1) CDS: *ska du göm& gömma Kucka i väskan?* (“are you going to &hi hide Kucka in the bag?”)
- 2) ADS: *ja hon brukar det i alla fall när jag ger henne &bo tandborsten* (“yes she does at least when I give her the &br toothbrush”)

Truncated phrases (marked as &(phrase)):

- 3) CDS: *&(här kommer nä) här kommer nämligen Kucka* (lit. “&(here comes ac) here comes actually Kucka”)
- 4) ADS: *&(titta kan) titta förstår hon* (lit. “&(look knows) look understands she”)

Prolongations (marked with :):

- 5) CDS: *kan det vara en ee ha:j?* (“can that be a ee sha:rk?”)

Hesitations (marked with _):

- 6) ADS: *ee hon förstå_r kom hit* (“ee she understa_nds come here”)

Filled pauses (e.g. ee, eh, uu, uh, öö, öh)

- 7) CDS: *kani& ee Kucka måste ha den där* (“the rabbi& ee Kucka needs that”)

Note that the primary annotation task was orthographic transcription, not disfluency annotation, and thus our results may underestimate the true disfluency rate.

Categorization of filled pauses

We distinguish between filled pauses in initial, internal, and final position within an utterance, clause, and/or phrase.

Initial: the FP occurs in the beginning of an utterance, e.g.:

- 8) a. ADS: *ee jaha ee det gör hon ju rätt ofta faktiskt* (“ee yeah ee she does that quite often actually”)
- b. CDS: *ee är du hungrig?* (“ee are you hungry?”)

Internal: the FP is located within a clause, or within a phrase, e.g. a verb phrase (“sees my keys” in 9a), a proper name (“Kucka”, “Ulla” in 10a, b), or in the beginning of (“roosters” in 11a) or within a noun phrase (“her different nicknames” in 11b):

- 9) a. ADS: *i hissen då får hon ee se mina nycklar och så* (“In the lift then she ee sees my keys”)
- b. CDS: *&(ska vi) ska vi ee hitta namn till allihopa?* (“&(shall we) shall we ee make up names for all of them?”)
- 10) a. ADS: *men kollar ni alltså på det hon gör nu när ee Ulla och jag pratar* (“but do you look at what she is doing now when ee Ulla and I are talking”)
- b. CDS: *här kan du få ee Kucka* (“here you can have ee Kucka”)
- 11) a. ADS: *vi hade ee tuppar också* (“we hade ee roosters as well”)
- b. ADS: *ja hon förstår ju sitt eget namn och hon förstår sina olika ee smeknamn* (“yes she understands her own name and she understands her different ee nicknames”)

Final: the FP marks the end of an utterance:

- 13) ADS: *igår så tittade hon och hennes pappa på en tavla ee* (“yesterday she and her father looked at a painting ee”)

Data extraction

For this study, we divide all utterances into two categories, adult-directed (AD) or child-directed (CD). We further categorize utterances based on the age of the child, and the gender of the caretaker.

Based on the disfluency annotation described above, disfluent utterances were extracted using the ELAN search tool. The categorization of filled pauses into initial, internal, or final position was performed manually.

Results

Disfluency in child-directed speech

Table 1 shows the disfluency frequency and rate per 100 words in ADS and CDS utterances in MINGLE-4. As shown, there is a difference between ADS (2.60 disfluencies/100 words) and CDS (0.88 disfluencies/100 words). This difference is statistically significant given a Log-Likelihood test (Log-Likelihood value 262.03, $p < 0.0001$).

Table 1. Disfluency frequency, word frequency, and disfluency rate per 100 words in Adult-Directed and Child-Directed speech in MINGLE-4.

	Disfl	Words	Disfl/100 w
AD	627	24109	2.60
CD	314	35485	0.88

In Table 2, the ADS and CDS utterances are divided in two categories based on the age of the child: infants (7–12 months) and one-year olds (13–24 months). Table 2 shows that the disfluency rate per 100 words for ADS is the same regardless of the age of the child present during recording (Log-Likelihood value 0.04), but that there is a significant increase of disfluency in CDS as the children develop (Log-Likelihood value 18.10, $p < 0.0001$).

Table 2. Disfluency frequency, word frequency, and disfluency rate per 100 words in Adult-Directed and Child-Directed speech categorized by child age.

	Infants (6–12 mnts)			Toddlers (13–24 mnts)		
	Disfl	Words	Disfl/100 w	Disfl	Words	Disfl/100 w
AD	232	8991	2.58	395	15042	2.63
CD	64	11057	0.58	219	21271	1.03

Table 3 shows the disfluency frequency and rate in ADS and CDS utterances to one-year olds categorized by the gender of the caretaker. As Table 3 shows, there is a difference between male and female speakers in disfluency frequency in ADS (Log-Likelihood value 4.90, $p < 0.05$) but, interestingly, there is no significant difference in CDS (Log-Likelihood value 2.2) between male and female speakers.

Filled pauses in Child-Directed Speech

We find that in our data, the majority of FPs in ADS (70%) occurs in initial position, whereas in CDS, FPs are evenly distributed between initial and utterance-internal position and there are no FPs in final position.

Table 3. Disfluency frequency, word frequency, and disfluency rate per 100 words in Adult-Directed and Child-Directed speech to children age 13–24 months, categorized by the gender of the caretaker.

	Men			Women		
	Disfl	Words	Disfl/100 w	Disfl	Words	Disfl/100 w
AD-1	211	7130	2.96	221	9244	2.39
CD-1	134	11880	1.13	109	11696	0.93

There are only 19 occurrences of FPs in our CDS data, but among these we find patterns of usage for FPs in both initial and internal position, as shown in Table 4.

Table 4. Frequency for Filled Pauses in Adult-Directed and Child-Directed speech in MINGLE-4, categorized by the position of the FP.

	Initial (%)	Internal (%)	Final (%)	TOTAL (%)
AD	174 (70%)	36 (14%)	39 (16%)	249 (100%)
CD	11 (58%)	8 (42%)	0	19 (100%)

Out of 11 initial FPs, 6 occur as attention-getting signals, and 5 precede utterance fragments. The initial FPs as attention-getting signals are followed by the child’s name (e.g. 14), a question (e.g. 15), or an imperative (e.g. 16).

There are only 19 occurrences of FPs in our CDS data, but among these we find patterns of usage for FPs in both initial and internal position. Out of 11 initial FPs, 6 occur as attention-getting signals, and 5 precede utterance fragments. The initial FPs as attention-getting signals are followed by the child’s name (e.g. 14), a question (e.g. 15), or an imperative (e.g. 16).

14) *ee hörru Cornelia* (“ee hey you Cornelia”)

15) *oj! ee ska du dricka upp all min mjölk?* “oi! ee are you going to drink all my milk?”)

16) *ee öppna munnen!* (“ee open your mouth!”)

In the utterance fragments following initial FPs, objects (e.g. 17) or actions (e.g. 18) are named:

17) *ee Kucka*

18) *ee blåsa* (“ee blow”)

The internal FPs in our CDS data (8 occurrences) precede unfamiliar or discourse-new objects referred to by names (e.g. 19) or noun phrases (e.g. “a little tanktop” in 20):

- 19) *kan du mata ee Kucka* (“can you feed ee Kucka”)
20) *kan vara ett ee ett litet linne?* (“could be a ee a little tanktop?”)

Discussion

There is a significant difference in disfluency rate between ADS and CDS in our data, and further, we find a significant increase in the rate of disfluency coupled with increasing child age when comparing CDS directed at infants (age 6 to 12 months) to CDS directed at one-year olds. These results for Swedish CDS are consistent with previous research on English CDS (Broen, 1972; Kidd, White & Aslin, 2011).

Previous research suggests that FPs commonly precede infrequent or discourse-new words, and may be a result of delay in lexical retrieval (Clark & Fox Tree, 2002). Although there are few occurrences of FPs in CDS, we find clear patterns of usage where FPs in initial position tends to function as attention-getting signals or to precede utterance fragments, while the internal FPs precede discourse-new information. However, since disfluencies such as FPs are infrequent in CDS, further data collection and analysis are needed.

Shriberg (1996) finds that the FP rate in the Switchboard corpus correlates with gender in that men produce significantly higher rates of FPs than women. We find the same pattern in our data, where there is a significant difference in disfluency rate (including FPs) by male and female speakers when talking to the (female) experiment leader, but interestingly there is no difference in disfluency rate between the male and female speakers when talking to children. Previous studies have reported no gender differences in Swedish as regards filled pauses production (e.g. Bell, Eklund & Gustafson, 2000).

Acknowledgements

This research of the first and the second author is part of the project “Modelling the emergence of linguistic structures in early childhood”, funded by the Swedish Research Council as 2011-675-86010-31. Thanks to the section for Phonetics, SU, for making this data available for us. Thanks also to our team of transcribers: Anna Ericsson, Joel Ivre, and Johan Sjons.

References

- Bell, L., R. Eklund & J. Gustafson. 2000. A Comparison of Disfluency Distribution in a Unimodal and a Multimodal Human–Machine Interface. *Proceedings of ICSLP '00*, Beijing, 16–20 October 2000, 3:626–629.
- Broen, P.A. 1972. *The Verbal Environment of the Language-Learning Child*. ASHA Monographs, No. 17. American Speech and Hearing Association: Washington, DC.
- Clark, E. 2009. *First Language Acquisition*. 2nd Edition. Cambridge University Press: Cambridge.
- Clark, H. & J. E. Fox Tree. 2002. Using *uh* and *um* in spontaneous speaking. *Cognition*, 84(1):73–111.
- Fox Tree, J. E. 1995. The Effects of False Starts and Repetitions on the Processing of Subsequent Words in Spontaneous Speech. *Journal of Memory and Language* 34:709–738.
- Eklund, R. 2010. The Effect of Directed and Open Disambiguation Prompts in Authentic Call Center Data on the Frequency and Distribution of Filled Pauses and Possible Implications for Filled Pause Hypotheses and Data Collection Methodology. *Proceedings of DiSS-LPSS Joint Workshop 2010, The 5th Workshop on Disfluency in Spontaneous Speech and The 2nd International Symposium on Linguistic Patterns in Spontaneous Speech*. University of Tokyo, 25–26 September 2010, Tokyo, Japan, 23–26.
- Eklund, R. & M. Wirén. 2010. Effects of open and directed prompts on filled pauses and utterance production. In: *Proceedings of Fonetik 2010*, Lund University, 23–28.
- Kidd, C., K. S. White & R. N. Aslin. 2011. Learning the Meaning of “Um”: Toddlers’ developing use of speech disfluencies as cues to speakers’ referential intentions. In I. Arnon & E. V. Clark (eds.), *Experience, Variation, and Generalization: Learning a First Language (Trends in Language Acquisition Research)*. Amsterdam: John Benjamins, 91–106.
- Nilsson Björkenstam, K. 2012. The MINGLE annotation scheme: Multimodal annotation of parent-child interaction in a free play setting (v. 1.0). *Papers from the Institute of Linguistics, University of Stockholm (PILUS)*, ISSN 0348-3223:63.
- Nilsson Björkenstam K. & M. Wirén. 2012. Reference to Objects in Longitudinal Parent–Child Interaction. In: *Proceedings of Workshop on Language, Action and Perception (APL)*. Lund, Sweden. [No page numbers.]
- Shriberg, E. E. 1996. Disfluencies in SWITCHBOARD. In: *Proceedings of International Conference on Spoken Language Processing*, Addendum, Philadelphia, PA, 11–14.

Proceedings of Fonetik 2013

The XXVIth Annual Phonetics Meeting
12–13 June 2013, Linköping University
Linköping, Sweden

Studies in Language and Culture
no. 21

Robert Eklund, editor



Linköping University

Conference website: www.liu.se/ikk/fonetik2013

Proceedings also available at: <http://roberteklund.info/conferences/fonetik2013>

Cover design and photographs by Robert Eklund

Photo of Claes-Christian Elert taken by Eva Strangert on the occasion of his 80th birthday

Proceedings of Fonetik 2013, the XXVIth Swedish Phonetics Conference

held at Linköping University, 12–13 June 2013

Studies in Language and Culture, no. 21

Editor: Robert Eklund

Department of Culture and Communication

Linköping University

SE-581 83 Linköping, Sweden

ISBN 978-91-7519-582-7

eISBN 978-91-7519-579-7

ISSN 1403-2570

© The Authors and the Department of Culture and Communication, Linköping University, Sweden

Printed by LiU-Tryck, Linköping, Sweden, 2013