# Longitudinal Studies of Variation Sets in Child-directed Speech

**Mats Wirén, Kristina Nilsson Björkenstam, Gintarė Grigonytė and Elisabet Eir Cortes**
Department of Linguistics
Stockholm University
SE-106 91 Stockholm, Sweden
{mats.wiren, kristina.nilsson, gintare, elisabet.cortes}@ling.su.se

## Abstract

One of the characteristics of child-directed speech is its high degree of repetitiousness. Sequences of repetitious utterances with a constant intention, *variation sets*, have been shown to be correlated with children's language acquisition. To obtain a baseline for the occurrences of variation sets in Swedish, we annotate 18 parent–child dyads using a generalised definition according to which the varying form may pertain not just to the wording but also to prosody and/or non-verbal cues. To facilitate further empirical investigation, we introduce a surface algorithm for automatic extraction of variation sets which is easily replicable and language-independent. We evaluate the algorithm on the Swedish gold standard, and use it for extracting variation sets in Croatian, English and Russian. We show that the proportion of variation sets in child-directed speech decreases consistently as a function of children's age across Swedish, Croatian, English and Russian.

## 1 Introduction

### 1.1 Background and motivation

Child-directed speech has many characteristics that set it apart from adult-directed language, such as shorter utterances, lower speech rate, fewer disfluencies, lower syntactic complexity, greater modulation of $F_0$ and high repetitiousness (Broen, 1972). Here is an example of the latter property from our data:[1]

> You can put the animals there.
> You can take the pig and the cat and put
>     them there.
> Can you put them there?
> Good.
> Can you put the pig there too?

Sequences of such (partial) self-repetitions with a constant intention have been called *variation sets*, and have been shown to account for a large proportion of the language that children hear (Küntay and Slobin, 1996; Clark., 2009, p. 37).[2]

Why does this phenomenon occur? To some extent, repetitiousness may serve simply to capture and maintain the child's attention, but our intuitions tell us that it is likely to also facilitate language learning for infants. For example, it may allow for effective segmentation of phonetic material (Bard and Anderson, 1983), and it has been shown to be a predictor of syntax growth (Hoff-Ginsberg, 1986; Hoff-Ginsberg, 1990; Waterfall, 2006). In a similar vein, investigating social and attentional cues in word learning, Frank et al. (2012), point out that the temporal proximity and continuity of repetitious language create supportive contexts where partial understanding of individual utterances can lead to fuller understanding.[3]

But variation sets have also been shown to benefit artificial language learning. In an experiment on this, Onnis et al. (2008) showed that adults exposed to input with varation sets performed better in phrase segmentation and phrase-boundary judgement tasks than a control group who heard

---

[1]Translation of utterances from the MINGLE-3 corpus (Björkenstam and Wirén, 2014) with parental speech to a child aged 1;3 (compare Section 3).

[2]"Variation set" is actually a misnomer, since the idea is that the order of individual utterances is important.

[3]It is also interesting to note that child-directed signing shares many characteristics of child-directed speech, such as prosodic exaggeration, lexical and syntactic simplification, and repetition (Masataka, 2000), and these shared characteristics include variation sets (Hoiting and Slobin, 2002).

the same input in scrambled order without variation sets. They note that "[f]rom a computational standpoint, the key characteristic of variation sets is that local mechanisms of alignment and comparison allow even memory-limited learners to discover structure that they would otherwise miss" (Onnis et al., 2008, p. 424).

## 1.2 Related work

Early studies of child-directed language dealing with partial and exact repetition include Broen (1972), Snow (1972), Kaye (1980) and Hoff-Ginsberg (1986; 1990). For example, Broen (*ibid.*, p. 29, 43) tracked "clusters of sequential sentences" where "the meaning remains constant". Snow (*ibid.*) found more partial and exact repetitions to 2-year olds than to 10-year olds.

Küntay and Slobin (1996) introduced the term "variation set", by which they meant a contiguous sequence of repetitions with varying form but constant intention. They pointed out that the core of a variation set (and the main vehicle for expressing the intention) is almost always a verb, with optionally expressed arguments. (In the above example from the MINGLE-3 corpus, this verb would be "put".) The possible variations were taken to be "(1) lexical substitution and rephrasing, (2) addition and deletion of specific referential terms, and (3) reordering of constituents" (Küntay and Slobin, 2002, p. 6). Their definition did not include exact repetitions, however. Furthermore, it appears that in order for a new utterance to be considered a member of a existing variation set, the new utterance has to satisfy the above conditions for *all* of the previous utterances taken to be in the set.

Küntay and Slobin's study was based on transcripts of everyday interaction between a Turkish-speaking mother and her child over a seven-month period, during which the child was between 1;8 and 2;3 years. The finding was that 21% of the utterances occurred within variation sets, and that these sets were positively associated with children's acquisition of specific verbs. A follow-up study of transcripts of another Turkish-speaking mother and a child (at age 1;3 and 2;0 years) showed how the communicative functions of the variation sets changed as a function of age (Küntay and Slobin, 2002).

Waterfall (2006) provided the first longitudinal study of variation sets in English, based on 12 mother–child dyads with children between 1;2 and 2;6 years. Waterfall's (2006, p. 21) definition of variation set is somewhat different from Küntay and Slobin's, though it is not clear what effect that has in practice. Basically, she defines a variation set as a sequence of utterances that belongs to the same conversational turn, that relates to the same event or situation, that "have similar or related meanings", and shares at least one noun or verb. Again, it appears that these conditions should hold between all utterances within the set, and like Küntay and Slobin, she did not include exact repetitions. Also, she allowed up to four non-related intervening utterances in a variation set. Waterfall found that children's production of nominal and verbal structures was correlated with peaks in the parents' use of that structure in variation sets. She also found a moderate decrease in the proportion of utterances that are part of variation sets as a function of age, from 17% at 1;2 years to 12% at 2;6 years.

Attempts at automatic extraction of variation sets naturally focus on form rather than function. Brodsky et al. (2007) suggest a simple definition of a variation set as a sequence of utterances where each successive pair of utterances has a lexical overlap of at least one element, excluding words on a stoplist (which includes high-frequency words). Variation sets are thus extracted by comparing pairs of successive utterances for repeated words, resulting in sets with at least one non-stoplisted word in common. Using an automated procedure of this kind, Brodsky et al. obtain a proportion of 21.5% of the *words* in Waterfall's (2006) corpus occurring in variation sets, and 18.3% of the words in the English CHILDES collection (MacWhinney, 2000). Similar studies have been performed by Onnis et al. (2008) and Waterfall et al. (2010). For example, when Onnis et al. used an automated procedure based on Waterfall's (2006) criteria on the Lara corpus from CHILDES (involving one child between 1;9 and 3;3 years), they obtained a proportion of 27,9% of the utterances being inside variation sets.

## 1.3 The problem

For the purpose of this work, we assume that variation sets play a role in language learning for infants, but we are agnostic as to the precise nature of this role. Rather, the aim is to investigate the longitudinal behaviour of variation sets using a

definition which subsumes earlier work but where the repetitiousness may also be, on the one hand, semantic (with no or very little surface repetition) and, on the other hand, prosodic or non-verbal (while displaying exact repetition). To obtain a baseline for the behaviour of this phenomenon in Swedish, we develop a gold standard for variation sets. To facilitate further empirical investigation, we introduce a surface algorithm which we evaluate on the gold standard and apply to Croatian, English and Russian.

## 2 Criteria for variation sets

### 2.1 Basic criteria

A starting-point for our work is Küntay and Slobin's (1996; 2002) definition, which takes variation sets to be sequences of utterances with the same communicative intention but with small differences in form. Basically, our definition subsumes Küntay and Slobin's, but we extend it in certain ways. First, along with Brodsky et al. (2007), we extract variation sets (whether manually in the gold standard or automatically using the algorithm) by comparing successive pairs of utterances: first–second, second–third, etc. Also, up to two intervening utterances (such as interjections) by the parent are allowed any time in a sequence (similarly to Snow (1972, p. 251) and Brodsky et al. (2007)). Furthermore, we allow for verbal input from the child within variation sets. The rationale for this is that our data covers the ages 0;7–2;9 years (see Section 3), and especially in the early dyads the children are still learning to take turns. As for constant intention, we make one exception from this, following Küntay and Slobin (2002): we include question–answer sequences where the parent provides both the question and the answer in variation sets.

### 2.2 Surface and semantic repetitiousness

A difference compared to previous work that we are aware of is that we aim at capturing a continuous scale of surface and semantic repetitiousness, where, at one extreme, the repetitiousness may be purely semantic without any surface similarity at all. Here is an example of this from our data, with approximate translations:

Titta här då!
  (But look here!)
Har du sett vilka tjusiga byxor?

(Have you seen the fancy pants?)
Kolla!
  (Look!)

The intention in each of these utterances is to make the child look at the pants, but there is no overlap whatsoever in form between the utterances.
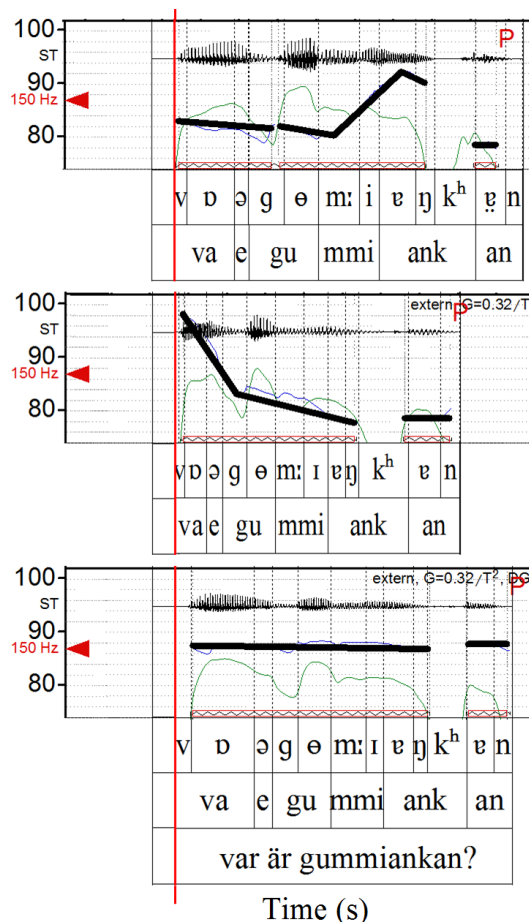


Figure 1: Phonetic and prosodic analysis of a repetition of the Swedish phrase "Var är gummi-ankan?" ("Where is the rubber duck?"), uttered by a male speaker. The utterances are ordered with the first utterance on top. The $y$-axis represents frequency (semitones) and the $x$-axis represents time (seconds). Black thick horizontal lines show stylised intonation based on tonal perception.

### 2.3 Multimodal variation

Contrary to many previous studies of variation sets (Küntay and Slobin, 1996; Küntay and Slobin, 2002; Waterfall, 2006), we include exact (verbatim) repetitions in our definition of variation sets. This is motivated by the result of a study that we made of three dyads in the multimodally annotated MINGLE-3 corpus (see Section 3). When-

ever word-for-word repetition occurred in the written transcript of the three dyads, we found consistent patterns of prosodic variation in the parents' speech, involving pitch, timing and/or stress, and typically also variation of their non-verbal cues, involving eye gaze direction, deictic gestures or object manipulation.

Figure 1 shows a phonetic and prosodic analysis of a variation set from our data with three exact repetitions of the Swedish utterance "Var är gummiankan?" ("Where is the rubber duck?").[4] The vertical line indicates time-synchronized starts of the repetitions. In the analysis window for each repetition, a black thick horizontal line shows stylized intonation based on tonal perception (perceived pitch). A downwards tilted line means falling intonation (from brighter to deeper voice), upwards tilted means rising intonation. In the background, the waveform and intensity (thin line) can be seen. The annotation rows beneath each repetition contains phonetic transcription in IPA (top row) and syllable segmentation (second row). The third row at the bottom contains an orthographic annotation.[5]

Here, the first utterance (shown at the top of the figure) initially displays relatively flat intonation, and then rising intonation with a peak on the first syllable in the noun "ankan" ("duck"), with a fall on the last syllable. In contrast, the second utterance has shorter duration and falling intonation throughout. Finally, the third utterance has completely flat intonation, with duration similar to the first utterance but with a prolongation of the first syllable, corresponding to the adverb "var" ("where").

Although this is just a small study, the fact that variation is here being systematically manifested through prosody and/or non-verbal cues when the wording is constant fits well with our general impression of exact repetitions. It is because of this multimodal variation that we include verbatim repetitions in variation sets.

## 3 Data

Our data consists of transcripts of Swedish child-directed speech from the MINGLE-3 corpus (Björkenstam and Wirén, 2014), consisting of 18 longitudinal dyads with three children (two girls, one boy) recorded between the ages of 7 and 33 months with six dyads per child, all of which is multimodally annotated. The complete duration of the 18 dyads is 7:29 hours (mean duration 24:58 minutes). The video and audio recordings were made from naturalistic parent–child interaction in a recording studio at the Phonetics Laboratory at Stockholm University (Lacerda, 2009). The children were interacting alternately with their mothers (10 dyads) and fathers (8 dyads). The scenario was free play.[6] The ELAN annotation tool (Wittenburg et al., 2006) was used for transcription of parent and child utterances, as well as annotation of eye gaze, deictic gestures and object manipulation (Björkenstam and Wirén, 2014). The transcripts have been automatically annotated with part-of-speech and morphosyntactic tags using Stagger (Östling, 2013), followed by manual correction.

## 4 Creating a gold standard

The manual annotation of variation sets started with analysis of four dyads, based on a guideline according to the criteria in Section 2. The same criteria were applied throughout all age groups. The annotations were made in ELAN, using timelines to code the extensions of variations sets across utterances, and taking into account both verbal and non-verbal input from parent and child from transcriptions, audio and video.

Each of the four dyads was annotated by two coders independently. The resulting annotations were merged, and a third annotator marked cases of disagreement. This resulted in an interannotator agreement (measured as set overlap between annotators) of 78%. The remaining 14 dyads were annotated by one annotator. During this phase, a classification of communicative intention based on the Inventory of Communicative Acts-Abridged (Ninio et al., 1994) was added. This classification was evaluated by comparing four representative dyads annotated by three independent annotators, resulting in a Fleiss's kappa of 0.63.

---

[4]The examples in this section are from the MINGLE-3 corpus (Björkenstam and Wirén, 2014), with approximate English translations.

[5]The intonation analyses were done in Prosogram (Mertens, 2004), and the figures were compiled from Prosogram and Praat (Boersma, 2001).

[6]Some of the data (transcripts and audio) is available through the Swedish section of CHILDES as the Lacerda files (MacWhinney, 2000).

Table 1: Results of the longitudinal study of Swedish variation sets (also used as gold standard in Table 3). The third row shows the proportions of child-directed utterances that are in variation sets. Each figure is obtained by first calculating the proportion per dyad and then averaging the proportions over all dyads in the respective age group. Boldface indicates statistically significant difference to boldfaced neighbour ($z$-test of sample proportions; respectively, $z = 8$, $p < 0.0001$, $z = 2.3$, $p < 0.02$, $z = 8.2$, $p < 0.0001$, two-tailed). The fourth row shows the proportions of exact repetitions within variation sets. Each figure is obtained by first calculating the proportion per variation set and averaging over the dyad, then averaging over all dyads in the respective age group.

| Longitudinal study of Swedish variation sets | Group 1 0;7–0;9 | Group 2 1;0–1;2 | Group 3 1;4–1;7 | Group 4 2;3–2;9 |
|---|---|---|---|---|
| Number of dyads | 5 | 5 | 5 | 3 |
| Number of child-directed utterances | 1032 | 1421 | 1492 | 724 |
| Proportion of utterances that are in variation sets | **50%** | **34%** | **30%** | **14%** |
| Proportion of exact repetitions in variation sets | 24% | 16% | 13% | 10% |

# 5 Results: Gold standard variation sets

In order to obtain a baseline for how the proportion of utterances that are in variation sets varied as a function of age of the children, we grouped the dyads according to child age in the following four data sets:

Age group 1: 0;7–0;9 (7–9 months)
Age group 2: 1;0–1;2 (12–14 months)
Age group 3: 1;4–1;7 (16–19 months)
Age group 4: 2;3–2;9 (27–33 months)

As shown in Table 1, our gold standard displayed a consistent decrease in the proportion of utterances in variation sets over time, from 50% for age group 1 to 14% for group 4. The proportion of verbatim repetitions in variation sets also decreased, from 24% for age group 1 to 10% for group 4.

# 6 Automatic extraction of variation sets

The method that we use for extracting variation sets is deliberately surface-based to allow us to determine how far this can bring us relative to our gold standard, which is based on both surface and semantic criteria. As mentioned above, the algorithm performs a stepwise comparison of pairs of successive utterances. The criterion for including two successive utterances in a variation set is that the difference between them (regarded as strings) does not fall below a certain similarity threshold. Additionally, following Brodsky et al. (2007) and others, we allow for sequences of maximally two intervening dissimilar utterances that do not obey this condition.

For string comparison, we used Ratcliff–Obershelp pattern recognition (Black, 2004) as implemented in the Python module `difflib`.[7] We refer to the variation-set extraction algorithm using this as "difflib ratio", DLR.[8] When comparing two strings, the matcher returns a value between 0 and 1. A value of 1 corresponds to an exact repetition, and 0 corresponds to two utterances without any overlap of words. By using this value as a parameter, we can obtain a threshold for the desired degree of similarity. The threshold can either be selected arbitrarily, or learned from evaluation against the gold standard variation sets. When evaluated against the gold standard, the optimal similarity threshold was 0.55 (see Figure 2).

We experimented with including information from the part-of-speech tagging of the transcripts (see Section 3) in such a way that the pair of strings compared consisted of both the words and their part-of-speech tags. Our intuition was that this might give us a more refined analysis, for example, by distinguishing cases of homonymy. This version of the algorithm turned out not to improve performance, however (see Figure 2), and was therefore dropped.

---

[7] https://docs.python.org/2/library/difflib.html#module-difflib.

[8] We also experimented with another standard technique for calculating string similarity, namely, edit distance, also known as Levenshtein distance (Levenshtein, 1966). In the end we found DLR to perform slightly better relative to the gold standard, however.
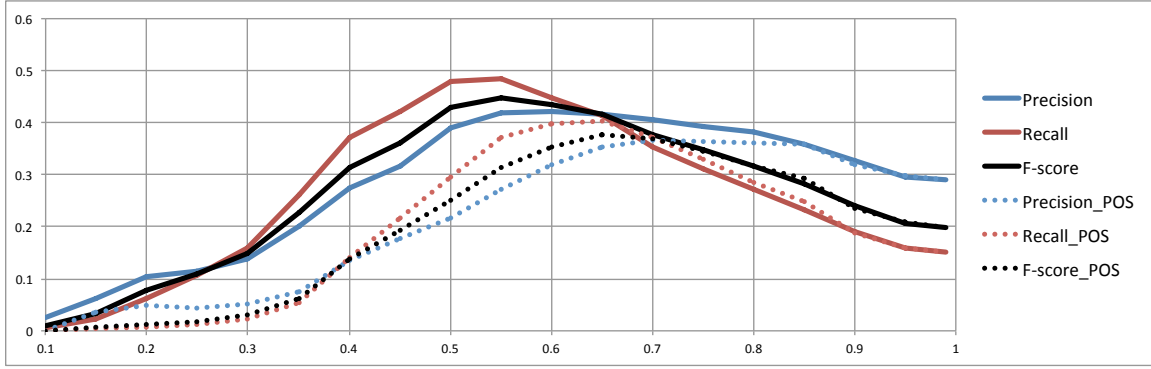
Figure 2: Results of strict matching with the DLR similarity measure on raw (solid lines) and part-of-speech tagged data (dotted lines). Similarity level thresholds on $x$-axis; precision, recall and $F$-score on $y$-axis.

Table 2: Example variation set from the gold standard (utterance 3–4) and utterances exrqcted by the algorithm (utterance 1–4).

| Example utterances | Member of gold set | Extracted by algorithm |
|---|---|---|
| 1. Ska vi lägga ner nånting i i väskan då? (Are we going to put something in in the bag then?) | – | Yes |
| 2. Va? (Huh?) | – | Yes |
| 3. Ska du lägga ner kossan i väskan kanske? (Are you going to put down the cow in the bag maybe?) | Yes | Yes |
| 4. Ska vi lägga ner kossan? (Are we going to put down the cow?) | Yes | Yes |

## 7 Results: Automated extraction

### 7.1 Evaluation on Swedish gold standard

We evaluated the algorithm against the gold standard variation sets using two kinds of metrics, which we refer to as *strict* and *fuzzy* matching. Strict matching requires exact matching on the utterance level of the extracted variation set and the corresponding gold standard set, whereas fuzzy matching allows for partial overlaps of the extracted variation set and the gold standard set. In the example in Table 2, only utterance 3 and 4 are members of the gold standard variation set, whereas the algorithm extracts utterances 1–4. Hence, the strict matching metric treats this extracted set as a false positive, whereas the fuzzy matching metric treats it as a true positive. As for fuzzy matching, we need a way of calculating precision for different degrees of overlaps with the gold set. The measure we have adopted for this purpose is *mean average precision* (MAP), see

Croft et al. (2009, p. 313).

Table 3 summarizes the results of extraction of variation sets relative to the gold standard according to the strict and fuzzy metric. Strict $F$-score reaches 0.56 and fuzzy $F$-score reaches 0.82 for age group 1, but $F$-scores gradually decrease with increasing age. Apparently, the variation displayed in the parents' speech becomes less amenable to surface methods as the children grow older. An indirect sign of this increased complexity in variation sets is that the proportion of exact repetitions decreases as the children grow older, as shown in Table 1.

### 7.2 Extraction of variation sets in Croatian, English and Russian

To investigate the behaviour of variation sets in other languages, we ran the algorithm with lon-

Table 3: Evaluation of the algorithm for automatic variation-set extraction against the Swedish gold standard per age group.

| String matching relative to gold standard | Group 1 0;7–0;9 | Group 2 1;0–1;2 | Group 3 1;4–1;7 | Group 4 2;3–2;9 |
|---|---|---|---|---|
| Strict precision | 0.539 | 0.406 | 0.351 | 0.217 |
| Strict recall | 0.581 | 0.406 | 0.446 | 0.333 |
| **Strict *F*-score** | **0.559** | **0.406** | **0.392** | **0.262** |
| Fuzzy precision | 0.774 | 0.627 | 0.505 | 0.324 |
| Fuzzy recall | 0.877 | 0.763 | 0.743 | 0.615 |
| **Fuzzy *F*-score** | **0.822** | **0.689** | **0.601** | **0.425** |

Table 4: Results of the algorithm for automatic variation-set extraction applied to Croatian, English and Russian child-directed utterances from CHILDES. The rows show the number of utterances in each age group, the average proportion of utterances that are in variation sets, and the average proportion of exact repetitions in the variation sets, with figures having being calculated in the same way as in Table 1.

| Language | Features of the data set | Group 1 0;7–0;9 | Group 2 1;0–1;2 | Group 3 1;4–1;7 | Group 4 2;3–2;9 |
|---|---|---|---|---|---|
| **Croatian** Kovacevic | Total number of utterances | 39 | 217 | 408 | (no data) |
| | Utterances in variation sets | 85% | 54% | 50% | – |
| | Exact repetitions in variation sets | 0.0% | 8.5% | 4.9% | – |
| **English (UK)** Lara | Total number of utterances | (no data) | (no data) | 926 | 391 |
| | Utterances in variation sets | – | – | 54% | 44% |
| | Exact repetitions in variation sets | – | – | 7.8% | 6.9% |
| **Russian** Protassova | Total number of utterances | (no data) | (no data) | 1088 | 545 |
| | Utterances in variation sets | – | – | 35% | 24% |
| | Exact repetitions in variation sets | – | – | 6.3% | 4.6% |

gitudinal corpora in Croatian,[9] English[10] and Russian from CHILDES (MacWhinney, 2000).[11] Although it was not possible to find a perfect correspondance with the age groups for Swedish, Table 4 shows how the selection of languages and transcripts from CHILDES partly matches the Swedish data. As shown in Table 4, both the proportion of variation sets and the proportion of exact repetitions as far, as can be seen, decrease consistently for Croatian, English and Russian.

## 8 Discussion

In our study of the Swedish gold standard, we obtained statistically significant decreases in the proportion of utterances within variation sets as a function of age between all age groups, from 50% for age group 1 to 14% for age group 4 (see Table 1). These differences were also more consistent than in Waterfall (2006), who obtained an overall decrease from 17% for 1;2 years to 12% for 2;6 years (*ibid.*, p. 125). Waterfall's age span was shorter than ours,[12] but its decrease was still less pronounced within the comparable age interval. It is also interesting to see that we obtained the largest proportion of variation sets for the youngest age group (0;7–0;9 years), which was not covered by Waterfall.

The fact that we see larger age-related differences in our data does not seem to be attributable to the inclusion of exact repetitions in our variation sets, judging from the proportiones of these in Table 1. In any case, and as argued in Section 2, the reason for extending the definition of variation sets in this way is motivated by an in-

---

[9] Kovacevic: Vjeran, files 20 (0;10 years) 23 (1;2 years), 33 (1;7 years).

[10] Lara, files 1-09-13 (1;9 years), 2-06-00 (2;6 years).

[11] Protassova: Varv, files 01 (1;6 years), 04 (1;10 years), 06 (2;4 years).

[12] Waterfall's age group 1;2 roughly matches our group 2, 1:6 and 1:8 roughly match our group 3, and 2:2 and 2:6 roughly match our group 4. Our results for age group 1 are new compared to previous studies, however.

depth analysis of a subset of these utterances in our multimodally annotated corpus. We conjecture that when an utterance is repeated verbatim, there is instead multimodal variation that increases the information and helps the child learn from the utterance. As far as we know, our longitudinal figures on proportions of exact repetitions are also the first that have been reported.

Our automatic algorithm for variation set extraction is deliberately surface-based in order to test how far this kind of method can bring us. An independent advantage is that it is easily replicable since it is based on a standard library for string comparison. The algorithm reaches a fuzzy *F*-score of up to 0.82 (strict: up to 0.56) relative to the Swedish gold standard in spite of only using criteria related to form. The *F*-score drops as a function of age, however (see Table 3); we conjecture that this is due to the relation between form and intention becoming less transparent with increased age. That is, as the child develops and learns more language, the parents' variation gets more complex. One way of handling this complexity would be by generalizing the algorithm to recognize intention.

Since the algorithm only uses form-based criteria, it is in principle also language-independent. We obtain consistent decreases of the proportions of utterances in variation sets also when we apply the algorithm to Croatian, English and Russian corpora of child-directed language (see Table 4). Although in this case we have no evaluations, it is interesting to see that the behaviour corresponds to what we expected.

## 9 Conclusion

We have investigated the longitudinal behaviour of variation sets in child-directed speech according to a generalised definition. Variation sets appear to function as a device for effective communication and learning with young children: the speaker repeats the same content while varying the wording, prosody and/or non-verbal cues in order to maximise the chance of comprehension. With increasing age and language comprehension, there is less need for such repetitiousness.

Our study of Swedish covered a larger age span and displayed a more consistent decrease than Waterfall's (2006) study of American English. Our automatic algorithm seems to usefully approximate manual extraction of variation sets at least for lower age groups, and an advantage is that the

algorithm is easily replicable. Applications of the algorithm to Croatian, English and Russian displayed similar decreases in the proportions of utterances in variation sets as a function of ages. We also found that the proportions of exact repetitions are similarly decreasing as a function of age for all languages, and we have demonstrated how multimodal cues seem to provide other dimensions of variation in these utterances.

## References

Ellen Gurman Bard and Anne H. Anderson. 1983. The unintelligibility of speech to children. *Journal of Child Language*, 10(2):265–292.

Kristina Nilsson Björkenstam and Mats Wirén. 2014. Multimodal annotation of synchrony in parent-child interaction. In *Proceedings of The 10th Workshop on Multimodal Corpora: Combining applied and basic research targets, 2014*, Reykjavik, Iceland, May. ELRA.

Paul E Black. 2004. Ratcliff/obershelp pattern recognition. *Dictionary of Algorithms and Data Structures*, 17.

Paul Boersma. 2001. Praat, a system for doing phonetics by computer. *Glot International*, 5(9/10):341–345.

Peter Brodsky, Heidi R. Waterfall, and Shimon Edelman. 2007. Characterizing motherese: On the computational structure of child-directed language. In *Proc. 29th Cognitive Science Society Conference*, Nashville, TN.

Patricia A. Broen. 1972. *The verbal environment of the language-learning child*. ASHA Monographs Number 17, American Speech and Hearing Association. Washington D.C.

Eve V. Clark. 2009. *First Language Acquisition*. Cambridge University Press, Cambridge, UK.

W. Bruce Croft, Donald Metzler, and Trevor Strohman. 2009. *Search Engines: Information Retrieval in Practice*. Addison-Wesley Publishing Company, USA.

Michael C. Frank, Joshua B. Tenenbaum, and Anne Fernald. 2012. Social and discourse contributions to the determination of reference in cross-situational learning. *Language Learning and Development*, 00:1–24.

Erika Hoff-Ginsberg. 1986. Function and structure in maternal speech: Their relation to the child's development of syntax. *Developmental Psychology*, 22(3):155–163.

Erika Hoff-Ginsberg. 1990. Maternal speech and the child's development of syntax: a further look. *Journal of Child Language*, 17:85–99.

Nini Hoiting and Dan I. Slobin. 2002. What a deaf child needs to see: Advantages of a natural sign language over a sign system. In R. Schulmeister and H. Reinitzer, editors, *Progress in sign language research. In honor of Siegmund Prillwitz/Fortschritte in der Gebärdensprachforschung. Festschrift für Siegmund Prillwitz*, pages 268–277. Signum, Hamburg.

Kenneth Kaye. 1980. Why we don't talk 'baby talk' to babies. *Journal of Child Language*, 7:489–507.

Aylin C. Küntay and Dan I. Slobin. 1996. Listening to a turkish mother: Some puzzles for acquisition. In *Social Interaction, Social Context, and Language. Essays in the Honor of Susan Ervin-Tripp*, pages 265–286. Lawrence Erlbaum, Mahwah, NJ.

Aylin C. Küntay and Dan I. Slobin. 2002. Putting interaction back into child language: Examples from Turkish. *Psychology of Language and Communication*, 6:5–14.

Francisco Lacerda. 2009. On the emergence of early linguistic functions: A biological and interactional perspective. In K. Alter, M. Horne, M. Lindgren, M. Roll, and J. von Koss Torkildsen, editors, *Brain Talk: Discourse with and in the brain*, pages 207–230. Media-Tryck, Lund, Sweden.

Vladimir I. Levenshtein. 1966. Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady*, volume 10, pages 707–710.

Brian MacWhinney. 2000. *The CHILDES Project: Tools for analyzing talk*. Lawrence Erlbaum Associates, Mahwah, NJ, 3 edition.

Nobuo Masataka. 2000. The role of modality and input in the earliest stages of language acquisition: Studies of japanese sign language. In C. Chamberlain, J. P. Morford, and R. I. Mayberry, editors, *Language acquisition by eye*, pages 3–24. Erlbaum, Hillsdale, NJ.

Piet Mertens. 2004. The prosogram: Semi-automatic transcription of prosody based on a tonal perception model. In *Proceedings of Speech Prosody 2004*, pages 549–552, Nara, Japan.

Anat Ninio, Catherine E. Snow, Barbara A. Pan, and Pamela R. Rollins. 1994. Classifying communicative acts in children's interactions. *Journal of Communicative Disorders*, 27:157–187.

Luca Onnis, Heidi R. Waterfall, and Shimon Edelman. 2008. Learn locally, act globally: Learning language from variation set cues. *Cognition*, 109(3):423–430.

Robert Östling. 2013. Stagger: an open-source part of speech tagger for swedish. *Northern European Journal of Language Technology*, 3:1–18.

Catherine E. Snow. 1972. Mothers' speech to children learning language. *Child Development*, 43(2):549–565.

Heidi R. Waterfall, Ben Sandbank, Luca Onnis, and Shimon Edelman. 2010. An empirical generative framework for computational modeling of language acquisition. *Journal of Child Language*, 37:671–703.

Heidi R. Waterfall. 2006. *A Little Change is a Good Thing: Feature Theory, Language Acquisition and Variation Sets*. Ph.D. thesis, Department of Linguistics, University of Chicago.

P. Wittenburg, H. Brugman, A. Russel, A. Klassmann, and H. Sloetjes. 2006. Elan: a professional framework for multimodality research. In *Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation*, pages 1556–1559, Genoa, Italy, May. ELRA.