Assessing the Impact of Local Adaptation in Child-Adult Dialogue: A Recurrence-Quantificational Approach

Robert M. Grimm
Utrecht Institute of Linguistics OTS
Utrecht University
r.m.grimm@students.uu.nl

Raquel Fernández
Institute for Logic, Language & Computation
University of Amsterdam
raquel.fernandez@uva.nl

1 Introduction

An important question in the study of dialogue is to what extent interlocutors converge on shared linguistic representations. Building on work by Dale and Spivey (2006) and Fernández and Grimm (2014), we make use of recurrence quantificational analysis (RQA) to investigate such linguistic convergence in child-caregiver dialogue. We use *convergence* as a cover term for possibly different adaptation mechanisms (e.g. priming, repetition), not all of which may be known, and without committing ourselves to the primacy of any one mechanism. However, we do assume that *convergence* is locality-dependent, since presumably the underlying mechanisms are unlikely to act on utterances that are far apart in time.

RQA (Eckmann et al., 1987) involves the construction of recurrence plots—structures which plot two data series against one another, and which allow for the extraction of further quantitative measures. We use recurrence-plot-derived measures in order to independently measure the influence of two possible constraints on the extent to which words and syntactic structures are used in both the child's and the caregiver's speech:

(1) the general use of a linguistic element in the other interlocutor's speech, and
(2) the reuse of a linguistic element in temporally close child-adult turns (i.e., *convergence*).

2 Method: Turn-based Recurrence Plots

Following Fernández and Grimm (2014), we construct turn-based recurrence plots. Given a child-caregiver dialogue, all child and adult turns are extracted; indexed by time, the child's turns are placed on the y-axis, and the adult's turns are placed on the x-axis. Every point in the resultant coordinate system then corresponds to a pair of turns. If we colour points according to the similarity of their turns, with black for maximal and white for minimal similarity, we often see a dark *diagonal line of incidence*—the set of points which compare adjacent turns. Two examples are given in Figure 1.

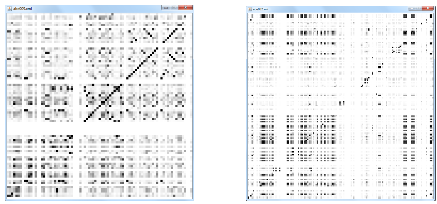

Figure 1: Recurrence plots from dialogues in the Kuczaj corpus (CHILDES database).

3 Procedure

We aim to measure the impact of factors (1) and (2) above on the frequency of linguistic elements in child and adult speech, respectively. We conduct two analyses: one focusing on the child's speech and one focusing on the adult's speech. For both analyses, we concentrate on the usage of three types of linguistic elements: content words, function words, and part-of-speech tag bigrams (POS bigrams). That way, we aim to measure a more meaning-driven usage (content words) and a more syntactically oriented usage (function words) of lexical items. POS bigrams are assumed to correspond roughly to syntactic structures and are the most syntactically oriented element type.

Our data come from three corpora in the CHILDES database (MacWhinney, 2000): Abe from the Kuczaj corpus; and Adam and Sarah from the Brown corpus. Given a linguistic element E and a dialogue, we construct a Boolean recurrence plot, where two turns are given a similarity score of 1 just in case both the child and adult turn contain E, and of 0 otherwise. We next take the sum of scores for points that correspond to turns which are at most two turns apart and which both contain E. This yields E's *raw recurrence*. Division by the frequency of E in the dialogue yields E's *recurrence*—a frequency-independent measure of convergence. By subtracting from E's frequency the number of occurrences of E which fall within the area around the diagonal defined by $d = 2$, we similarly obtain a convergence-independent measure of frequency—E's *occurrence* (in the child's speech and in the adult's speech, respectively). In sum, given some linguistic element E, we calculate the following measures for each dialogue transcript:[1]

- *recurrence*
- *child / adult occurrence*
- *child / adult frequency*

Table 1 shows the sample sizes for each element type. Average measures over all transcripts in each corpus then form the basis of multiple linear regression models: *recurrence* and *child/adult occurrence* serve as predictors; and *adult frequency* and *child frequency* act as response variables.

corpus	cont. words	fun. words	POS big.
Abe	193 (552)	65 (172)	70 (226)
Adam	292 (553)	78 (168)	71 (195)
Sarah	254 (506)	53 (170)	50 (222)

Table 1: Sample sizes. Number of elements whose average *raw recurrence* is significantly different from the randomized condition and the total number of elements in the corpus (in parentheses).

4 Results

The regression models are summarized in Tables 2 and 3.[2] Results for the child's and adult's speech are very similar, suggesting that the two interlocutors adapt to the other's speech via the same underlying mechanisms. Regression coefficients also do not differ much across corpora, indicating that different child-caregiver dyads adapt to one another in similar ways.

Comparison of the predictor values sheds light on the impact of (1) general use in the other interlocutor's speech and (2) convergence of both interlocutors' speech in determining the frequency of a linguistic element in the child's/adult's speech. *Occurrence* takes the larger value for most element types; general use in the other's speech thus appears to have a stronger impact on the usage of content words (almost exclusively affected by occurrence) and of function words (affected by both predictors, though more strongly by occurrence). The usage of POS bigrams, lastly, is more strongly affected by *recurrence*. More syntactic elements may thus be prone to a stronger influence of convergence. We also found that the most frequent items within each element type are much more strongly affected by convergence than by general use (space constraints prevent us from elaborating on this result). Since the most frequent elements account for a disproportionately large part of the language produced, this suggest that the majority of both interlocutors' dialogue contributions may in fact be shaped through convergence.

corpus	*element type*	*adult occ.*	*recurrence*	R^2
Abe	con. words	0.79 * * *	0.08	0.61
	fun. words	0.43 * * *	0.48 * * *	0.68
	POS big.	0.20 *	0.72 * * *	0.78
Adam	con. words	0.83 * * *	0.09 **	0.66
	fun. words	0.67 * * *	0.34 * * *	0.79
	POS big.	0.20 *	0.71 * * *	0.76
Sarah	con. words	0.85 * * *	0.18 * * *	0.66
	fun. words	0.57 * * *	0.50 * * *	0.77
	POS big.	0.49 * * *	0.46 * * *	0.85

Table 2: Multiple linear regression models for the predictors *adult occurrence* and *recurrence*, with *child frequency* as response variable.

corpus	*element type*	*child occ.*	*recurrence*	R^2
Abe	con. words	0.80 * * *	0.01	0.63
	fun. words	0.43 * * *	0.54 * * *	0.73
	POS big.	0.30 * * *	0.65 * * *	0.83
Adam	con. words	0.86 * * *	0.03	0.74
	fun. words	0.73 * * *	0.24 * * *	0.80
	POS big.	0.23 **	0.71 * * *	0.79
Sarah	con. words	0.82 * * *	-0.11 * * *	0.72
	fun. words	0.82 * * *	-0.01	0.65
	POS big.	0.24 **	0.71 * * *	0.85

Table 3: Multiple linear regression models for the predictors *child occurrence* and *recurrence*, with *adult frequency* as response variable.

5 Future Work

In future work, we aim to utilize a longitudinal design in order to track developmental changes in how specific element types are influenced by the two factors we have studied here.

[1] Importantly, we only consider E for analysis if its average *raw recurrence* differs significantly from a baseline condition where the child's turns are randomly shuffled (one-sided t-test, $p \leq 0.05$).

[2] We use the following convention to indicate significance: *** : $p \leq 0.001$, ** : $p \leq 0.01$, * : $p \leq 0.05$

References

Rick Dale and Micheal J. Spivey. 2006. Unraveling the dyad: Using recurrence analysis to explore patterns of syntactic coordination between children and caregivers in conversation. *Language Learning*, 56(3):391–430.

Jean-Pierre Eckmann, Sylvie Oliffson Kamphorst, and David Ruelle. 1987. Recurrence plots of dynamical systems. *Europhysics Letters*, 4(9):973–977.

Raquel Fernández and Robert Grimm. 2014. Quantifying categorical and conceptual convergence in child-adult dialogue. In *Proceedings of the 36th Annual Conference of the Cognitive Science Society*, Quebec City, Canada.

Brian MacWhinney. 2000. *The CHILDES Project: Tools for Analyzing Talk: Volume I: Transcription format and programs, Volume II: The database*. Lawrence Erlbaum Associates.