Linguistic Alignment in Collaborative **Coding Tasks**

A Comparison between Groups of Middle School and Undergraduate Students

Amanda Griffith, Zane Price, Xiaoyi Tian

METHODS

Data: Transcribed spoken-dialogue corpora (Table 1)

Middle school (MS) corpora

- 7812 utterances
- 33 pairs (66 individuals) of students
- Science classrooms (topics e.g., revolution)
- Block-based programming environment Snap!

Undergraduate (UG) corpora

- 7297 utterances
- 15 pairs (30 individuals) of students
- CS1 Java lab sessions (topics e.g., calculator)
- Text-based programming environment *Eclipse*

Linguistic Alignment Measurements

Syntactic Local Alignment

- target: one utterance or turn
- prime: previous 10 utterances from the other speaker
- **SILLA:** For each (prime, target) pair, the number of overlapped syntax rules normalized by the product of prime length and target length (Wang et al, 2014).
- **nSILLA:** Normalized SILLA. The number of overlapped syntax rules normalized by the average SILLA from all pairs that have the same product of length.

Lexical Global Alignment

- Calculated for All Words (entr-All) and for the Top 25 words (entr-25)
- **entr-All**: negated absolute value of the difference between the number of times speaker 1 used a particular word divided by the total number of words uttered. Ranges between 0 and - ∞ , 0 being perfect match.
- i amandagriffith, zaneprice5, tianx@ufl.edu
- Github repo: sylvia935/ling-alignment





Middle schoolers: more syntactic alignment, less lexical alignment **Undergraduates:** more lexical alignment,

less syntactic alignment



Middle school

BACKGROUND

Pair programming

- Working side-by-side at one computer
- Driver (operator) and Navigator (observer)
- Two programmers switch roles frequently

Why alignment?

Linguistic alignment is considered an important communicative feature (Xu & Reitter, 2015). Syntactic and lexical alignment has been found to correlate with task success (Reitter & Moore, 2014).

RESULTS

- student groups.

RQ1: Do students align with their partners syntactically and lexically?

• Yes. Both middle school and college students align with their dialogue partners both syntactically and lexically.

• Compared the alignment scores with the baseline scores (Table 2 & Figure 3)

RQ2: Are there differences in linguistic alignments between pairs of middle school and undergraduate students?

• Yes. Local syntactic alignment (SILLA) of middle school students' utterances are higher than college students utterances.

• Global lexical alignment (entr2-all) of middle school groups are lower than college

RQ3: How do syntactic alignment associated with lexical alignment in two groups? • Linear least square regression on group-level alignments (Table 3 & 4) No significant correlation between syntactic and lexical alignment

• Syntax rule counts per sentence is negatively correlated with lexical alignment



ogramming system *Snap!* used by the middle schoo



Figure 2. Undergrad (left) and Middle school (right) programming pairs solving problems together

L. Descriptive Statistics for Middle school (MS) and undergraduate (UG) corpora

Corpora	stats	# total turn	# groups	# speakers	duration	# turns	# words	# syntax rules
					(min)	per group	per turn	per turn
MS	mean	7812	36	72	31.22	236.73	8.84	9.4
	stdev	-		-	4.36	122.64	10.65	11.1
UG	mean	7297	15	30	53.2	486.47	9.14	8.4
	stdev	-	-	-	18.88	190.18	11.62	11

Table 2. Syntactic and lexical alignment and baseline scores for both corpora

Local Syntactic Alignment				Global Lexical Alignment (entr2)										
	Middle school		Undergraduate		Middle school			Undergraduate						
		SILLA			SILLA		All	All words	Top25	Top25	All	All words	Top25	Top25
	SILLA	baseline	nLLA	SILLA	baseline	nLLA	words	baseline	words	baseline	words	baseline	words	baseline
Mean	0.024	0.022	1	0.019	0.018	1	-0.58	-0.72	-0.42	-0.59	-0.48	-0.62	-0.41	-0.47
Std	0.019	0.016	0.693	0.024	0.016	0.913	0.1	0.07	0.13	0.11	-0.1	0.08	0.15	0.11
Median	0.021	0.02	1	0.016	0.025	0.95	-0.59	-0.72	-0.41	-0.56	-0.47	-0.6	-0.37	-0.45



Figure 3. Actual and baseline syntactic alignment scores (smoothed) over turns for UG corpora. Actual alignment score is almost always higher than the baseline.



Figure 4. Syntactic alignment scores (SILLA and nLLA) over turns for both corpora

 Table 3. Linear Regression: Middle School Lexical Alignment (entr-all)

n=33; R-squared: 0.355

	Estimate	Standard Error	t Ratio	Prob> t
(Intercept)	-0.427	0.233	-1.84	0.0767
Number of Syntax Rules	-0.03	0.008	-3.60	0.0012
SILLA	0.377	0.235	1.6	0.1202
nLLA	-0.328	0.222	-1.48	0.1504

 Table 4. Linear Regression: Undergraduate Lexical Alignment (entr-all)

n=15; R-squared: 0.708

	Estimate	Standard Error	t Ratio	Prob> t
(Intercept)	-0.387	0.266	-1.45	0.1745
Number of Syntax Rules	-0.425	0.011	-3.73	0.0033
SILLA	0.422	0.262	1.61	0.1353
nLLA	-0.164	0.24	-0.68	0.5077

