



About

- We combine empirically **broad modeling of complexity** with **high-quality error annotations** to track early German writing acquisition
- Results:
 - Overall, **writing becomes more accurate and complex**
 - In elementary school, accuracy discriminates grade-levels best
 - In secondary school, complexity discriminates grade-levels best
 - These findings well **generalize across writing topics**
- The complexity code is becoming part of CTAP (Chen & Meurers, 2016): <http://purl.org/ctap>
- Our script to extract accuracy measures from KCT can be found at <https://github.com/zweiss/KCTErrorExtractor>

Data

- Karlsruhe Children's Text corpus (Lavalley, Berkling & Stüker, 2015)
- 1,701 German texts of students in elementary and secondary school
- We study 1,633 texts containing ≥ 10 words by students aged 7 to 15

	1/2	3/4	5/6	7/8	all		1/2	3/4	5/6	7/8	all
Elementary	203	524	0	0	727	Wolf	133	353	0	0	466
Realschule	0	0	297	236	533	Park	90	171	0	0	261
Hauptschule	0	0	165	208	373	Future	0	0	332	333	665
						Idol	0	0	130	111	241
all	203	524	462	444	1,633	all	203	524	462	444	1,663

Assessment of Writing Performance

- (Second) language performance is analyzed along the dimensions of complexity, accuracy, and fluency (CAF triad)
- CAF is used for L1/L2 proficiency and readability assessment (Weiss & Meurers, 2018; Yoon & Polio, 2016; Bulté & Housen, 2014)
- Complexity refers to the **elaborateness and variability of language** while accuracy is defined as a **native speaker-like error rate**
- 308 complexity features** of theoretical linguistic domains and psycho-linguistic language processing and use measures (Weiss, 2017)
- 41 accuracy features** based on KCT annotations and inferred from comparing the target hypothesis with students' original writings

Studies 1 and 2: Grade-Level Predictions

- We classify grade-levels across- (study 1) and within schools (study 2)
- No invariable features (most common value $\geq 90\%$) \rightarrow 262 features
- SMO classifier with linear kernel using 10 iterations of 10-fold CV
- Meta information includes writing topic and school track

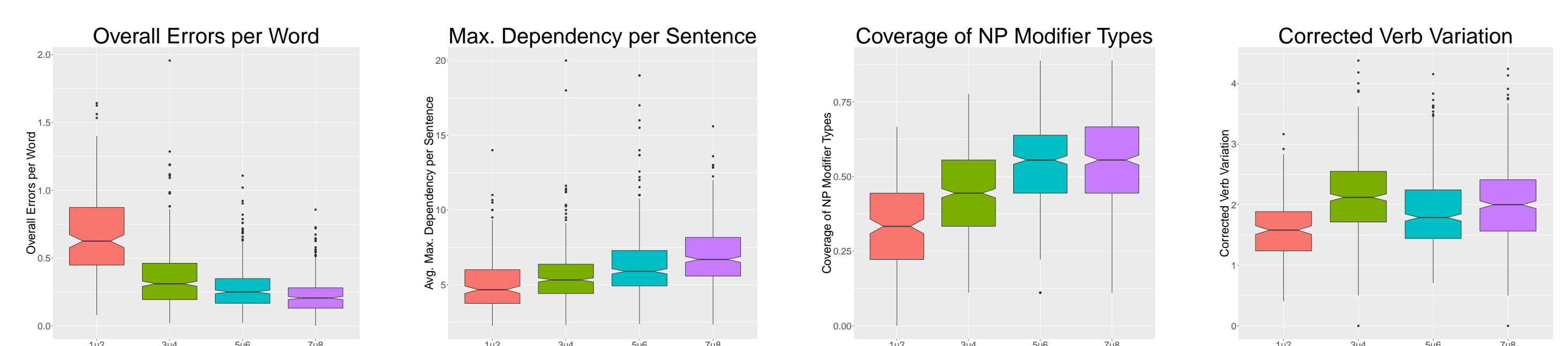
	Grades 1/2 to 7/8			Grades 1/2 and 3/4			Grades 5/6 and 7/8		
	Size	μ -Acc.	SD-Acc.	Size	μ -Acc.	SD-Acc.	Size	μ -Acc.	SD-Acc.
Majority baseline	1	32.08	0.14	1	71.72	0.35	1	51.15	0.27
All Features + Meta	264	72.68	1.94	256	82.81	2.11	258	65.66	2.13
All Features	262	70.96	2.01	255	82.60	1.97	255	63.71	1.82
Complexity	225	68.35	2.25	218	77.93	2.42	220	64.16	1.63
Error Rate	37	54.47	2.11	37	81.56	1.27	35	54.34	2.48
Lexical	31	60.10	1.69	31	77.32	1.92	30	62.74	1.58
Discourse	48	60.10	1.66	46	75.18	1.71	45	57.13	1.75
Phrasal	41	61.29	1.73	39	76.77	2.18	41	57.64	2.10
Clausal	26	52.95	1.56	26	72.44	0.49	25	58.70	2.37
Morphological	27	56.45	1.47	27	71.72	0.35	27	54.31	2.39
Language Use	30	45.45	1.28	30	71.72	0.35	30	55.73	2.34
Human Processing	20	42.18	1.55	19	71.72	0.35	18	52.67	1.90

- Across schools, combining accuracy and complexity performs best
- Across schools, lexicon, discourse, and phrasal complexity develop
- In secondary school, clausal complexity is particularly important

Zooming in on Students' Writing Development

- Information gain ranking on elementary and secondary school data to inspect the most informative feature of each feature set
- Two-tailed t-tests to test if grade-level differences are significant

Feature Name	Feature Set	Elementary school				Secondary school			
		1/2	3/4	t	p	5/6	7/8	t	p
Overall errors / W	Error Rate	0.68	0.37	11.53	.000	0.28	0.22	5.60	.000
Corrected verb variation	Lexical	1.62	2.13	-11.55	.000	1.88	2.01	-3.03	.003
P(Subject \rightarrow Nothing)	Discourse	0.15	0.10	3.40	.001	0.05	0.06	-1.35	.177
Avg. NP modifier types	Phrasal	0.31	0.42	-8.93	.000	0.52	0.52	-0.21	.831
Conjunction clauses / S	Clausal	0.11	0.13	-0.96	.339	0.13	0.18	-3.47	.001
Finite verbs / verb	Morph.	0.82	0.81	1.63	.105	0.71	0.70	0.88	.381
Pct. LW in Subtlex	Language Use	0.04	0.05	-1.71	.089	.085	.077	1.82	.069
DLT-IC (M) / finite verb	Human Processing	1.09	1.11	-1.96	.051	1.22	1.25	-1.65	.099
Capitalization errors / W	Error Rate	0.15	0.07	9.87	.000	0.05	0.04	5.61	.000
Yule's K	Lexical	217.	153.	7.21	.000	152.	128.	5.60	.000
Temp. connectives / S	Discourse	0.73	0.63	1.85	.066	0.47	0.62	-4.10	.000
Verb modifiers / VP	Phrasal	0.29	0.49	-4.85	.000	0.55	0.65	-2.86	.004
Verbs / t-unit	Clausal	1.67	1.57	-0.97	.333	1.69	1.81	-3.18	.002
Derived nouns / noun	Morph.	0.02	0.02	-0.38	.708	0.04	0.05	-2.66	.008
Pct. LW in dlexDB	Language Use	0.62	0.60	1.60	.111	0.60	0.63	-3.27	.001
(Σ max. dep.) / S	Human Processing	5.12	5.60	-2.64	.009	6.30	6.97	-4.59	.000



- We observe for many features a significant incremental development (e.g., error rate and average maximal dependency length)
- However, some features only develop in certain school types (e.g., NP modifier ratio) or show unclear patterns (e.g., verb variation)

Study 3: Cross-Topic Testing

- We test the cross-topic performance of the classifiers from study 2
- We perform cross-topic training and testing of classifiers using all features (+ meta), all complexity features, and all error rate features

	Elementary School			Secondary School		
	Train	Test	Acc.	Train	Test	Acc.
Majority baseline	<i>n.a.</i>	Park	65.52	<i>n.a.</i>	Idol	50.01
All Features + Meta	-	-	-	Future	Idol	62.66
All Features	Wolf	Park	76.63	Future	Idol	59.33
Complexity	Wolf	Park	68.58	Future	Idol	59.34
Error Rate	Wolf	Park	81.61	Future	Idol	55.19

- The elementary school accuracy model fully generalizes
- The secondary school complexity model shows only a small drop

Current Limitations and Future Work

- Our current studies are restricted by using quasi-longitudinal data that approximates development through grade-level bins
- Our follow-up study will investigate this development in a genuinely longitudinal corpus spanning several weeks (Berkling, 2018)

References

- Berkling, K. (2018). A 2nd Longitudinal Corpus for Children's Writing with Enhanced Output for Specific Spelling Patterns. In *LREC-2018*. pp. 2262–2268.
- Bulté, B. & A. Housen (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing* 26(0), 42 – 65.
- Chen, X. & D. Meurers (2016). CTAP: A Web-Based Tool Supporting Automatic Complexity Analysis. In *Proceedings of the Workshop on CLALC @ COLING*. Osaka, Japan, pp. 113–119.
- Lavalley, R., K. Berkling & S. Stüker (2015). Preparing Children's Writing Database for Automated Processing. In *LTLT@ SLATE*. pp. 9–15.
- Weiss, Z. (2017). Using Measures of Linguistic Complexity to Assess German L2 Proficiency in Learner Corpora under Consideration of Task-Effects. Master's thesis, University of Tübingen, Germany.
- Weiss, Z. & D. Meurers (2018). Modeling the Readability of German Targeting Adults and Children: An Empirically Broad Analysis and its Cross-Corpus Validation. In *Proceedings of the 27th COLING*.
- Weiss, Z. & D. Meurers (2019). Analyzing Linguistic Complexity and Accuracy in Academic Language Development of German across Elementary and Secondary School. In *Proceedings of the 14th BEA*.
- Yoon, H.-J. & C. Polio (2016). The Linguistic Development of Students of English as a Second Language in Two Written Genres. *TESOL Quarterly* pp. 275–301.