

Reading demands & language proficiency in foreign language learning

An exploration of linguistic complexity modeling

Zarah Weiss Detmar Meurers

University of Tübingen
Germany

*Text Comprehension: What We Know About the Dance Between
Reader, Text, and Task in Reading Comprehension*
AERA Symposium, April 12., 2021

Foreign Language Readability and Proficiency

Reading demands
and language
proficiency in
foreign language
learning

Zarah Weiss and Detmar
Meurers

- ▶ Language input fosters foreign language (FL) learning
 - ▶ Challenging input at $i+1$ (Krashen 1981)
 - ▶ Zone of Proximal Development (ZPD; Lantolf et al. 2015)
- ▶ Identification of FL reading materials requires
 1. analysis of FL proficiency of language learner
 2. analysis of FL reading materials
 3. match of proficiency and reading level
- ▶ We use linguistic complexity analysis to link FL writing complexity and FL reading materials:
 - ▶ Is the input appropriately challenging (“+1”)?
Empirical focus here: German as a FL

Motivation

Complexity analysis

Analysis 1: Readability

Analysis 2: Proficiency

Linking readability and proficiency

Syntactic complexity

Lexical complexity

Morphological complexity

Cohesion

Summary and Outlook

References



Automatic analysis of linguistic complexity

- ▶ Complexity, Accuracy, and Fluency used to describe language performance in SLA (Housen et al. 2019)
 - ▶ complex language is **elaborate and varied** (Ellis 2003)
- ▶ Characterization of proficiency and development (Housen et al. 2019; Ortega 2012; Crossley & McNamara 2014)
- ▶ For German, we analyze 489 complexity feature:
 - ▶ covering syntax, lexicon, morphology, cohesion and human language processing and use measures
 - ▶ successfully used for
 - ▶ proficiency assessment: Weiss & Meurers (2019a,b, 2021)
 - ▶ readability assessment: Weiss & Meurers (2018)

Reading demands
and language
proficiency in
foreign language
learning

Zarah Weiss and Detmar
Meurers

Motivation

Complexity analysis

Analysis 1: Readability

Analysis 2: Proficiency

Linking readability
and proficiency

Syntactic complexity

Lexical complexity

Morphological complexity

Cohesion

Summary and
Outlook

References



First analysis: Readability

- ▶ We collected 1,446 graded GFL reading texts
 - ▶ by a leading European publisher of FL reading materials
- ▶ Contains easy, medium, and advanced texts
 - ▶ linked to CEFR levels A2, B1/B2, and C1 by publisher
- ▶ Automatic prediction of reading levels A2, B1/B2, and C1
 - ▶ Trained ordinal random forest with 10-fold cross-validation
 - ▶ Accuracy = 89.0% in 3-way classification

Reading demands
and language
proficiency in
foreign language
learning

Zarah Weiss and Detmar
Meurers

Motivation

Complexity analysis

Analysis 1: Readability

Analysis 2: Proficiency

Linking readability
and proficiency

Syntactic complexiy

Lexical complexiy

Morphological complexity

Cohesion

Summary and
Outlook

References



Second analysis: Proficiency

- ▶ MERLIN: cross-sectional German as a FL data (N=1,033; Wisniewski et al. 2013)
 - ▶ elicited in official standardized language certification tests
 - ▶ Learner writing rated by two experts on Common European Framework of Reference (CEFR) scale
- ⇒ Automatic CEFR classification (A1, A2, B1, B2, C):
- ▶ trained ordinal random forest with 10-fold cross-validation
 - ▶ accuracy for 5-way classification = 70.0%

Zooming in on relevant features across domains

- ▶ 373 features are informative (info gain) for both data sets
- ▶ Complexity feature values for reading texts are higher than those for learner productions for

	A2	B1/B2	C1	#
Syntax	38%	25%	7%	60
Lexicon	82%	67%	56%	95
Morphology	88%	73%	35%	48
Cohesion	97%	86%	72%	36
Language use	91%	71%	37%	109

- ▶ We zoom in on features to explore where an adaptive “+1” relation can be observed in the different domains
 - ▶ syntax, lexicon, morphology, cohesion

Syntactic complexity

Motivation

Complexity analysis

Analysis 1: Readability

Analysis 2: Proficiency

Linking readability and proficiency

Syntactic complexi

Lexical complexi

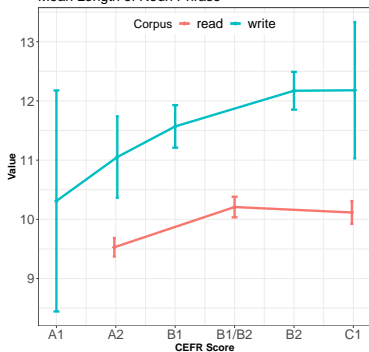
Morphological complexi

Cohesion

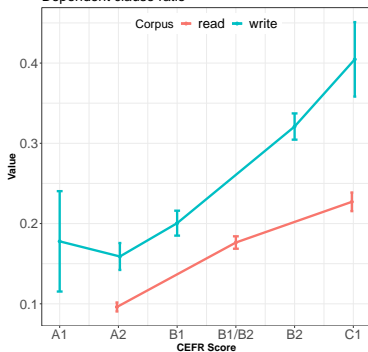
Summary and Outlook

References

Mean Length of Noun Phrase



Dependent clause ratio



- ▶ Systematic increase syntactic elaborateness in both
- ▶ but values for reading texts are systematically lower
⇒ no syntactic challenge in reading texts

Lexical complexity

Motivation

Complexity analysis

Analysis 1: Readability
Analysis 2: Proficiency

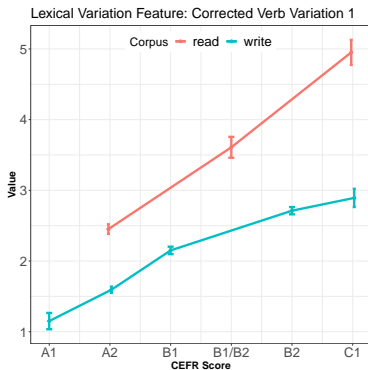
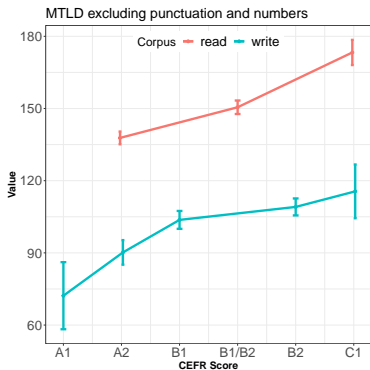
Linking readability and proficiency

Syntactic complexi
Lexical complexi

Morphological complexi
Cohesion

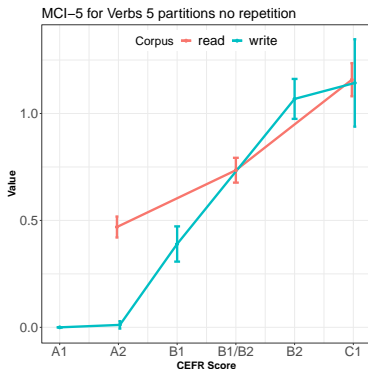
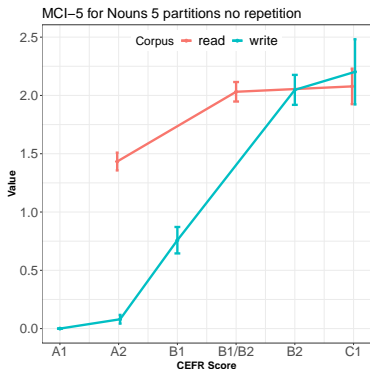
Summary and Outlook

References



- ▶ Systematic increase in lexical variation for both
- ▶ MTLD and corrected verb variation higher for reading texts
⇒ texts offer challenge for learners in lexical domain

Morphological complexity



- ▶ Systematic increase in morphological variation for both
 - ▶ Morphological complexity index measures POS-specific TTR for morphological units (Brezina & Pallotti 2019)
- ▶ Nominal and verbal MCI higher for A and B1 levels
 - ⇒ Texts offer challenge for beginning/intermediate

Cohesion

Motivation

Complexity analysis

Analysis 1: Readability
Analysis 2: Proficiency

Linking readability and proficiency

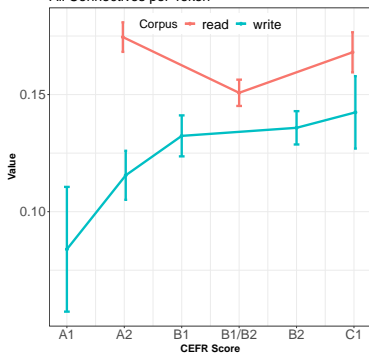
Syntactic complexiy
Lexical complexiy
Morphological complexiy

Cohesion

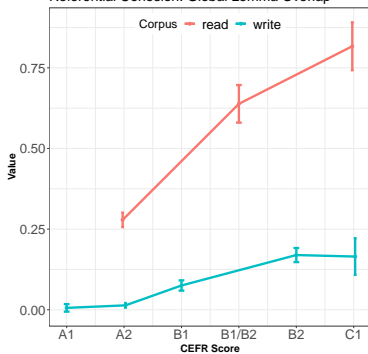
Summary and Outlook

References

All Connectives per Token



Referential Cohesion: Global Lemma Overlap



- ▶ Reading texts systematically use more connectives and higher implicit cohesion
 - ⇒ Texts offer challenge with respect to text cohesion

Summary and Outlook

- ▶ Leveled reading material generally more complex than learner production in most domains but syntax.
 - ▶ most prominent for beginning and intermediate levels
 - ▶ to be clarified: How much of this is due to gap between active and passive language competence?
- ⇒ Which level of challenge in which domain best fosters foreign language learning for which kind of learners?
 - ▶ More research needed on the nature of the adaptivity best fostering foreign language learning.
 - ▶ Linguistic complexity analysis provides fine-grained instrument to carry out such research.

References

- Brezina, V. & G. Pallotti (2019). Morphological complexity in written L2 texts. *Second language research* 35(1), 99–119.
- Crossley, S. A. & D. S. McNamara (2014). Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners. *Journal of Second Language Writing* 26, 66–79.
- Ellis, R. (2003). *Task-based Language Learning and Teaching*. Oxford, UK: Oxford University Press.
- Housen, A., B. De Clercq, F. Kuiken & I. Vedder (2019). Multiple approaches to complexity in second language research. *Second Language Research. Special Issue on Linguistic Complexity* 35(1), 2–31.
- Krashen, S. D. (1981). The Fundamental Pedagogical Principle in Second Language Teaching. *Studia Linguistica* 35(1–2), 50–70. URL <http://www3.interscience.wiley.com/journal/121500538/abstract?CRETRY=1&SRETRY=0>.
- Lantolf, J. P., S. L. Thorne & M. E. Poehner (2015). Sociocultural theory and second language development. In *Theories in second language acquisition: An introduction*, Routledge New York, NY.
- Ortega, L. (2012). Interlanguage complexity: A construct in search of theoretical renewal. In B. Szmrecsanyi & B. Kortmann (eds.), *Linguistic complexity: Second language acquisition, indigenization, contact*, Berlin: de Gruyter, pp. 127–155.
- 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 International Conference on*

Computational Linguistics (COLING). Santa Fe, New Mexico, USA.
<https://www.aclweb.org/anthology/C18-1026>.

- Weiss, Z. & D. Meurers (2019a). Analyzing Linguistic Complexity and Accuracy in Academic Language Development of German across Elementary and Secondary School. In *Proceedings of the 14th Workshop on Innovative Use of NLP for Building Educational Applications (BEA)*. Florence, Italy: Association for Computational Linguistics.
- Weiss, Z. & D. Meurers (2019b). Broad Linguistic Modeling is Beneficial for German L2 Proficiency Assessment. In A. Abel, A. Glaznieks, V. Lyding & L. Nicolas (eds.), *Widening the Scope of Learner Corpus Research. Selected Papers from the Fourth Learner Corpus Research Conference*. Louvain-La-Neuve: Presses Universitaires de Louvain.
- Weiss, Z. & D. Meurers (2021). Analyzing the linguistic complexity of German learner language in a reading comprehension task: Using proficiency classification to investigate short answer data, cross-data generalizability, and the impact of linguistic analysis quality. *International Journal of Learner Corpus Research* 7(1), 84–131.
- Wisniewski, K., K. Schöne, L. Nicolas, C. Vettori, A. Boyd, D. Meurers, A. Abel & J. Hana (2013). MERLIN: An online trilingual learner corpus empirically grounding the european reference levels in authentic learner data. In *ICT for Language Learning 2013, Conference Proceedings, Florence, Italy*. Libreriauniversitaria. it Edizioni.

Reading demands
and language
proficiency in
foreign language
learning

Zarah Weiss and Detmar
Meurers

Motivation

Complexity analysis

Analysis 1: Readability

Analysis 2: Proficiency

Linking readability
and proficiency

Syntactic complexiy

Lexical complexiy

Morphological complexity

Cohesion

Summary and
Outlook

References

