Linguistic complexity as a way to probe into genre differences?

Zarah Weiss

based on work done in collaboration with
Nina Hicks, Thomas Studer, and Detmar Meurers

Department of Linguistics & LEAD Research Network
University of Tübingen

Workshop on Complexity and Register (CAR21), Berlin, November 19th, 2021
How do we define complexity?

Figure from Bulté & Housen (2014)
CAF in Second Language Acquisition research

- CAF triad describes dimensions of **language performance** (Housen et al. 2012)
  - Assessed to characterize L2 proficiency and development
  - Also used for readability and writing quality assessment (Vajjala & Meurers 2012; Crossley et al. 2019)

- **Complexity** = elaborateness and variation of language
  - **Accuracy** = native-like and error-free language use
  - **Fluency** = native-like production speed and smoothness

- **Adequacy** suggested as 4th CAF dimension (Pallotti 2009)
  - Or: Native-like elaborateness and variation of language?
Task effects on complexity, accuracy, fluency

- **Task effects** have been shown to influence CAF (Yoon & Polio 2017a; Yang et al. 2015; Yoon 2017)
  - L2 corpora rarely control for comparable task conditions (Tracy-Ventura & Myles 2015; Yoon & Polio 2017a)

- Debated which **task factors** contribute to task effects
  - Functional task factors (e.g., genre, topic)
  - Cognitive task factors (e.g., shared context, code complexity)

- Effect of cognitive task factors on CAF trade-offs debated
  - Limited Attentional Capacity Model (Skehan 1996)
  - Cognition Hypothesis (Robinson 2001)
Task effects on complexity, accuracy, fluency

- **Task effects** have been shown to influence CAF (Yoon & Polio 2017a; Yang et al. 2015; Yoon 2017)
  - L2 corpora rarely control for comparable task conditions (Tracy-Ventura & Myles 2015; Yoon & Polio 2017a)

- Debated which **task factors** contribute to task effects
  - Functional task factors (e.g., genre, topic)
  - Cognitive task factors (e.g., shared context, code complexity)

- **Effect of cognitive task factors on CAF trade-offs debated**
  - Limited Attentional Capacity Model (Skehan 1996)
  - Cognition Hypothesis (Robinson 2001)

⇒ No consensus on interaction between task factors, proficiency and CAF (sub)dimensions across languages
Focus of today’s talk

1. Move from focus on proficiency to **focus on genre**
   ⇒ Can we use linguistic complexity to characterize genre differences in L2 writing?
   ⇒ Does this generalize to writings by (near-)native speakers?

2. Consider **multilingual dimension** of genre differences
   ⇒ Does our generalization apply across languages?

3. Zoom into informative **linguistic differences** between genres
   ⇒ Which linguistic features characterize genre differences?
Automatic assessment of complexity

- Extract 489 features of German linguistic complexity, language use, human processing, and discourse cohesion
- It has been successfully used for the assessment of, e.g.,
  - Text readability and L2 proficiency
    (Weiss et al. 2021; Weiss & Meurers 2018, 2019b, 2021)

1http://ctapweb.com
Automatic assessment of complexity

- Extract 489 features of German linguistic complexity, language use, human processing, and discourse cohesion
- It has been successfully used for the assessment of, e.g.,
  - Text readability and L2 proficiency
    (Weiss et al. 2021; Weiss & Meurers 2018, 2019b, 2021)
  - Academic language development (Weiss & Meurers 2019a)

1http://ctapweb.com
Automatic assessment of complexity

- Extract 489 features of German linguistic complexity, language use, human processing, and discourse cohesion
- It has been successfully used for the assessment of, e.g.,
  - Text readability and L2 proficiency
    (Weiss et al. 2021; Weiss & Meurers 2018, 2019b, 2021)
  - Academic language development (Weiss & Meurers 2019a)
  - Teachers’ grading behavior (Weiss et al. 2019)

1 http://ctpweb.com
Automatic assessment of complexity

- Extract 489 features of German linguistic complexity, language use, human processing, and discourse cohesion
- It has been successfully used for the assessment of, e.g.,
  - Academic language development (Weiss & Meurers 2019a)
  - Teachers’ grading behavior (Weiss et al. 2019)
  - Spoken language in the classroom (Weiss et al. 2022)

1http://ctapweb.com
Automatic assessment of complexity

- Extract 489 features of German linguistic complexity, language use, human processing, and discourse cohesion

- It has been successfully used for the assessment of, e.g.,
  - Academic language development (Weiss & Meurers 2019a)
  - Teachers’ grading behavior (Weiss et al. 2019)
  - Spoken language in the classroom (Weiss et al. 2022)

- Features in multilingual Common Text Analysis Platform¹ (Chen & Meurers 2016; Weiss et al. 2021)
  - Intersection of 200+ features for 🇬🇧 🇩🇪 🇫🇷 🇪🇸
  - Soon: Italian, Arabic, Portuguese, Turkish

¹http://ctapweb.com
Linguistic complexity as a way to probe into genre differences?
Zarah Weiss

Motivation
Analysis system
Multilingual analysis
Complexity research
Cross-lingual readability assessment
Proficiency and task assessment
Probing into genre differences
SWIKO corpus
Predicting text genre
Zooming in on features

Summary & outlook
References

Complex NPs across languages

Deutsch: Hever Castle, eines der schönsten Schlösser der Welt und Wohnsitz von Anne Boleyn, hat einen geheimnisvollen Reiz.

English: Hever Castle, one of the most beautiful castles in the world and the residence of Anne Boleyn, has a mysterious charm.

French: Le château de Hever, l’un des plus beaux châteaux du monde et la résidence d’Anne Boleyn, a une allure mystérieuse.
Complex NPs across languages

Hever Castle, eines der schönsten Schlösser der Welt und Wohnsitz von Anne Boleyn, hat einen geheimnisvollen Reiz.

Hever Castle, one of the most beautiful castles in the world and the residence of Anne Boleyn, has a mysterious charm.

Le château de Hever, l’un des plus beaux châteaux du monde et la résidence d’Anne Boleyn, a une allure mystérieuse.
Hever Castle, one of the most beautiful castles in the world and the residence of Anne Boleyn, has a mysterious charm.

Le château de Hever, l’un des plus beaux châteaux du monde et la résidence d’Anne Boleyn, a une allure mystérieuse.
Complex NPs across languages

Hever Castle, eines der schönsten Schlösser der Welt und Wohnsitz von Anne Boleyn, hat einen geheimnisvollen Reiz.

Hever Castle, one of the most beautiful castles in the world and the residence of Anne Boleyn, has a mysterious charm.

Le château de Hever, l’un des plus beaux châteaux du monde et la résidence d’Anne Boleyn, a une allure mystérieuse.
Complex NPs across languages

Hever Castle, eines der schönsten Schlösser der Welt und Wohnsitz von Anne Boleyn, hat einen geheimnisvollen Reiz.

Hever Castle, one of the most beautiful castles in the world and the residence of Anne Boleyn, has a mysterious charm.

Le château de Hever, l’un des plus beaux châteaux du monde et la résidence d’Anne Boleyn, a une allure mystérieuse.
Broad multilingual complexity modeling

**Syntax (N=60)**
- phrasal elaboration, clausal elaboration, syntactic variation
  (e.g., Chen, 2017; Kyle, 2016; Wolfe-Quintero et al., 1998)

**Human processing (N=21)**
- cognitive costs, dependency lengths
  (e.g., Gibson, 2002; Shain et al., 2016; Weiss 2017)

**Lexicon (N=86)**
- richness, variation, density
  (e.g., Lu, 2012; McCarthy & Jarvis, 2010; Wolfe-Quintero et al., 1998)

**Language use (N=28)**
- frequencies, age of active use
  (e.g., Chen & Meurers, 2018; Wolfe-Quintero et al., 1998)

**Morphology (N=41)**
- derivation, inflection, compounding, morpheme variation
  (e.g., Brezina & Pallotti, 2019; Weiss & Meurers, 2019)

**Cohesion (N=26)**
- connectives, pronouns, grammatical transitions, sentence overlap
  (e.g., Barzilay & Lapata, 2008; Crossley et al., 2016; Graesser et al., 2004; Todirascu, 2013)

**Syntactic complexity as a way to probe into genre differences?**

**Zarah Weiss**

**Motivation**

**Analysis system**

**Multilingual analysis**

**Complexity research**
- Cross-lingual readability assessment
- Proficiency and task assessment

**Probing into genre differences**
- SWIKO corpus
- Predicting text genre
- Zooming in on features

**Summary & outlook**

**References**

\[ \sum \cap = 262 \]
Broad multilingual complexity modeling

Linguistic complexity as a way to probe into genre differences?
Zarah Weiss

Motivation

Analysis system
Multilingual analysis

Complexity research
Cross-lingual readability assessment
Proficiency and task assessment

Probing into genre differences
SWIKO corpus
Predicting text genre
Zooming in on features

Summary & outlook

References

— Syntax (N=60)
  — phrasal elaboration, clausal elaboration, syntactic variation
    (e.g., Chen, 2017; Kyle, 2016; Wolfe-Quintero et al., 1998)

— Lexicon (N=86)
  — richness, variation, density
    (e.g., Lu, 2012; McCarthy & Jarvis, 2010; Wolfe-Quintero et al., 1998)

— Language use (N=28)
  — frequencies, age of active use
    (e.g., Chen & Meurers, 2018; Wolfe-Quintero et al., 1998)

— Cohesion (N=26)
  — connectives, pronouns, grammatical transitions, sentence overlap
    (e.g., Barzilay & Lapata, 2008; Crossley et al., 2016; Graesser et al., 2004; Todirsca, 2013)

— Human processing (N=21)
  — cognitive costs, dependency lengths
    (e.g., Gibson, 2002; Shain et al., 2016; Weiss 2017)

— Morphology (N=41)
  — derivation, inflection, compounding, morpheme variation
    (e.g., Brezina & Pallotti, 2019; Weiss & Meurers, 2019)

— Syntax (N=60) ∩ Lexicon (N=86) ∩ Language use (N=28) ∩ Cohesion (N=26) = 262

▶ Clausal elaboration (dependent clauses per sentence)
▶ Phrasal elaboration (modifiers per noun phrase)
▶ Syntactic variation (parse tree edit distances)
Broad multilingual complexity modeling

Syntax ($N=60$)
- phrasal elaboration,
- clausal elaboration,
- syntactic variation
(e.g., Chen, 2017; Kyle, 2016; Wolfe-Quintero et al., 1998)

Human processing ($N=21$)
- cognitive costs,
- dependency lengths
(e.g., Gibson, 2002; Shain et al., 2016; Weiss 2017)

Cohesion ($N=26$)
- connectives, pronouns, grammatical transitions,
- sentence overlap
(e.g., Barzilay & Lapata, 2008; Crossley et al., 2016; Graesser et al., 2004; Todirascu, 2013)

Language use ($N=28$)
- frequencies,
- age of active use
(e.g., Chen & Meurers, 2018; Wolfe-Quintero et al., 1998)

Morphology ($N=41$)
- derivation,
- inflection,
- compounding,
- morpheme variation
(e.g., Brezina & Pallotti, 2019; Weiss & Meurers, 2019)

Lexicon ($N=86$)
- richness,
- variation,
- density
(e.g., Lu, 2012; McCarthy & Jarvis, 2010; Wolfe-Quintero et al., 1998)

Syntax ($N=60$) ∩ Human processing ($N=21$) ∩ Cohesion ($N=26$) ∩ Language use ($N=28$) ∩ Morphology ($N=41$) ∩ Lexicon ($N=86$) = 262

- Richness (type-token ratios)
- Variation (POS-specific type-token ratios)
- Density (noun density = nouns per lexical word)
Broad multilingual complexity modeling

Syntax (N=60)
- phrasal elaboration,
- clausal elaboration,
- syntactic variation
  (e.g., Chen, 2017; Kyle, 2016; Wolfe-Quintero et al., 1998)

Human processing (N=21)
- cognitive costs,
- dependency lengths
  (e.g., Gibson, 2002; Shain et al., 2016; Weiss 2017)

Cohesion (N=26)
- connectives, pronouns, grammatical transitions,
  sentence overlap
  (e.g., Barzilay & Lapata, 2008; Crossley et al., 2016; Graesser et al., 2004; Todirascu, 2013)

Language use (N=28)
- frequencies,
- age of active use
  (e.g., Chen & Meurers, 2018; Wolfe-Quintero et al., 1998)

Lexicon (N=86)
- richness, variation,
- density
  (e.g., Lu, 2012; McCarthy & Jarvis, 2010; Wolfe-Quintero et al., 1998)

Morphology (N=41)
- derivation, inflection,
- compounding, morpheme variation
  (e.g., Brezina & Pallotti, 2019; Weiss & Meurers, 2019)

\[ \Sigma \cap \cap \cap = 262 \]

- Mean and SD word frequency from frequency data bases
- Here: frequencies per million words for TV captions
- Separate for different POS or word categories
Broad multilingual complexity modeling

**Syntax (N=60)**
- phrasal elaboration, clausal elaboration, syntactic variation
  (e.g., Chen, 2017; Kyle, 2016; Wolfe-Quintero et al., 1998)

**Human processing (N=21)**
- cognitive costs, dependency lengths
  (e.g., Gibson, 2002; Shain et al., 2016; Weiss 2017)

**Cohesion (N=26)**
- connectives, pronouns, grammatical transitions, sentence overlap
  (e.g., Barzilay & Lapata, 2008; Crossley et al., 2016; Graesser et al., 2004; Todirascu, 2013)

**Lexicon (N=86)**
- richness, variation, density
  (e.g., Lu, 2012; McCarthy & Jarvis, 2010; Wolfe-Quintero et al., 1998)

**Language use (N=28)**
- frequencies, age of active use
  (e.g., Chen & Meurers, 2018; Wolfe-Quintero et al., 1998)

**Morphology (N=41)**
- derivation, inflection, compounding, morpheme variation
  (e.g., Brezina & Pallotti, 2019; Weiss & Meurers, 2019)

**Human processing (N=21)**
- cognitive costs, dependency lengths
  (e.g., Gibson, 2002; Shain et al., 2016; Weiss 2017)

**Cohesion (N=26)**
- connectives, pronouns, grammatical transitions, sentence overlap
  (e.g., Barzilay & Lapata, 2008; Crossley et al., 2016; Graesser et al., 2004; Todirascu, 2013)

**Lexicon (N=86)**
- richness, variation, density
  (e.g., Lu, 2012; McCarthy & Jarvis, 2010; Wolfe-Quintero et al., 1998)

**Language use (N=28)**
- frequencies, age of active use
  (e.g., Chen & Meurers, 2018; Wolfe-Quintero et al., 1998)

**Morphology (N=41)**
- derivation, inflection, compounding, morpheme variation
  (e.g., Brezina & Pallotti, 2019; Weiss & Meurers, 2019)

**Morphological complexity index (Brezina & Pallotti 2019)**
- to measure variability of morphemes for specific POS

\[ \Sigma \land \land \land = 262 \]
Broad multilingual complexity modeling

Syntax (*N*=60)
- phrasal elaboration, clausal elaboration, syntactic variation
  (e.g., Chen, 2017; Kyle, 2016; Wolfe-Quintero et al., 1998)

Human processing (*N*=21)
- cognitive costs, dependency lengths
  (e.g., Gibson, 2002; Shain et al., 2016; Weiss 2017)

Lexicon (*N*=86)
- richness, variation, density
  (e.g., Lu, 2012; McCarthy & Jarvis, 2010; Wolfe-Quintero et al., 1998)

Language use (*N*=28)
- frequencies, age of active use
  (e.g., Chen & Meurers, 2018; Wolfe-Quintero et al., 1998)

Morphology (*N*=41)
- derivation, inflection, compounding, morpheme variation
  (e.g., Brezina & Pallotti, 2019; Weiss & Meurers, 2019)

Cohesion (*N*=26)
- connectives, pronouns, grammatical transitions, sentence overlap
  (e.g., Barzilay & Lapata, 2008; Crossley et al., 2016; Graesser et al., 2004; Todirascu, 2013)

- Overlap of POS/arguments between
  1. Adjacent sentences = local overlap
  2. Any two sentences = global overlap

Σ = 262
Broad multilingual complexity modeling

Linguistic complexity as a way to probe into genre differences?
Zarah Weiss

Motivation

Analysis system

Multilingual analysis

Complexity research

Cross-lingual readability assessment
Proficiency and task assessment
Probing into genre differences
SWIKO corpus
Predicting text genre
Zooming in on features

Summary & outlook

References

Average and maximal integration costs based on Dependency Locality Theory by Gibson (2000)
Multilingual complexity assessment

1. Shared NLP pipeline to obtain sentences, words, POS tags, constituency/dependency parses, lemmas, stems (see for technical details Weiss et al. 2021)

2. Identify linguistic constructions with extraction rules
   - Rules for language-specific features
   - Frequency features from Subtlex-US/DE and Lexique (Brysbaert et al. 2011a,b; New et al. 2001)
   - Language-specific rules for syllables, POS, constituents
   - Language-independent rules for all other

3. Fully language-independent calculation of feature ratios
Cross-lingual readability assessment

- Predict reading levels on multi-lingual L2 reading corpus (Weiss et al. 2021)
  1. Can we predict L2 reading levels for German and English?
  2. Does this generalize across languages?
  3. How do reading levels linguistically differ from each other

2www.spotlight-online.de, www.deutsch-perfekt.de
Cross-lingual readability assessment

- Predict reading levels on multi-lingual L2 reading corpus (Weiss et al. 2021)
  1. Can we predict L2 reading levels for German and English?
  2. Does this generalize across languages?
  3. How do reading levels linguistically differ from each other

- Experiment on Spotlight corpus from L2 magazine articles
  - English (N=3,285), German (N=1,446)
  - Beginning (A2), medium (B1/B2), advanced (C1)

---

2www.spotlight-online.de, www.deutsch-perfekt.de
Readability classification results on Spotlight

- Train/test ordinal random forest using 301 features
  - Model 1: distinguish reading levels on English data
  - Model 2: distinguish reading levels on German data

- Evaluation with accuracy in with 10-fold cross-validation
  - Model 1: 74.5% for English; baseline: 46.5%
  - Model 2: 88.0% for German; baseline: 52.8%

- Language-specific models generalize to some degree to the other language (despite considerable performance drop)

- Investigation of informative features show comparable reading level differences
  - Especially wrt. lexical domain and frequency features
Illustrate feature differences for English/German

- More frequent vocabulary in English than German texts
  - Likely due to compounding and inflection in German
- Word frequency decreases with increasing reading levels

▶ More frequent vocabulary in English than German texts
  ▶ Likely due to compounding and inflection in German
  ▶ Word frequency decreases with increasing reading levels
Proficiency and task effects

- Predict German L2 proficiency of learners from their written essays (Weiss & Meurers 2019b)
  1. Can we predict proficiency from learner writing?
  2. Which task influences do we observe?
  3. How do proficiency, task, and complexity interact?
Proficiency and task effects

- Predict German L2 proficiency of learners from their written essays (Weiss & Meurers 2019b)
  1. Can we predict proficiency from learner writing?
  2. Which task influences do we observe?
  3. How do proficiency, task, and complexity interact?

- Merlin corpus of 1,033 German L2 texts (Wisniewski et al. 2013)
  - Elicited in standardized testing at CEFR levels A1 to C1
  - Three task prompts per level ⇒ 15 different task prompts
  - Texts rated by two experts on CEFR scale (A1 to C2)

- Focus on task rather than individual task factors
  ⇒ Nearly all factors fully confounded with test level
Proficiency and task classification results

- Train/test ordinal random forests
  - 489 features extracted from form-based target hypothesis
  - Model 1: predict L2 proficiency of learners
  - Model 2: predict task prompt for text

- Evaluation with accuracy in 10-fold cross-validation
  - Model 1: 70.0%; 5-way classification; baseline: 32.0%
  - Model 2: 89.8%; 15-way classification; baseline: 8.0%
Proficiency and task classification results

- Train/test ordinal random forests
  - 489 features extracted from form-based target hypothesis
  - Model 1: predict L2 proficiency of learners
  - Model 2: predict task prompt for text

- Evaluation with accuracy in 10-fold cross-validation
  - Model 1: 70.0%; 5-way classification; baseline: 32.0%
  - Model 2: 89.8%; 15-way classification; baseline: 8.0%

⇒ Complexity features are more indicative for task differences than for proficiency differences
Zooming in on the Merlin corpus

- Subset of the Merlin corpus used for classification studies
  - Focus on B1/B2 learners taking B1/B2 tests (N=337)
- Test level confounds with genre, formality, & structured-ness
Summary of findings

- Task differences play central role in learners’ L2 complexity and partially outweighs seeming proficiency differences.
Summary of findings

- Task differences play central role in learners’ L2 complexity and partially outweighs seeming proficiency differences
  - Morphological complexity, expl. cohesion, language use
Task differences play central role in learners’ L2 complexity and partially outweighs seeming proficiency differences

- Morphological complexity, expl. cohesion, language use
- Lexical and phrasal complexity characterize proficiency differences despite task sensitivity
Summary of findings

- Task differences play central role in learners’ L2 complexity and partially outweighs seeming proficiency differences
  - Morphological complexity, expl. cohesion, language use
  - Lexical and phrasal complexity characterize proficiency differences despite task sensitivity
  - Human processing and clausal measures stable
Summary of findings

- Task differences play central role in learners’ L2 complexity and partially outweighs seeming proficiency differences
  - Morphological complexity, expl. cohesion, language use
  - Lexical and phrasal complexity characterize proficiency differences despite task sensitivity
  - Human processing and clausal measures stable

- Repeated observation of “test level” effect on complexity
  - Confounds with genre, formality, & structured-ness
Summary of findings

- Task differences play central role in learners’ L2 complexity and partially outweighs seeming proficiency differences
  - Morphological complexity, expl. cohesion, language use
  - Lexical and phrasal complexity characterize proficiency differences despite task sensitivity
  - Human processing and clausal measures stable

- Repeated observation of “test level” effect on complexity
  - Confounds with genre, formality, & structured-ness

- Similar findings on Falko GT corpus (Siemen et al. 2006)
  - Corpus of German L2 writing elicited in curricular writing courses at the Georgetown University
  - No deconstruction in task characteristics possible
Complexity and genre

- Previous research demonstrates interaction between tasks, proficiency and complexity domains
- Confounded task variables in corpora do not allow to distinguish the source of these task effect
  - Are the observations really due to overall task difference or can we identify specific contribution of certain task factors?
  - Do findings generalize to L1 writing and across languages?
Complexity and genre

- Previous research demonstrates interaction between tasks, proficiency and complexity domains
- Confounded task variables in corpora do not allow to distinguish the source of these task effect
  - Are the observations really due to overall task difference or can we identify specific contribution of certain task factors?
  - Do findings generalize to L1 writing and across languages?
- Probe into genre differences in L2 writing across languages
SWIKO corpus

- Multilingual task-based Swiss learner corpus of 410 oral & 1,803 written texts by lower secondary school students
  - Currently augmented and further analyzed in two projects (WETLAND, DiCoi; 2021–2024)

---

3 https://centre-plurilinguisme.ch/en/research/swiss-learner-corpus-swiko
Linguistic complexity as a way to probe into genre differences?

Zarah Weiss

Motivation
Analysis system
Multilingual analysis
Complexity research
Cross-lingual readability assessment
Proficiency and task assessment
Probing into genre differences
SWIKO corpus
Predicting text genre
Zooming in on features
Summary & outlook
References

SWIKO corpus

- Multilingual task-based Swiss learner corpus of 410 oral & 1,803 written texts by lower secondary school students³
  - Currently augmented and further analyzed in two projects (WETLAND, DiCo; 2021–2024)
- Analysis of two sub-corpora of Swiss student texts
  - 1,002 L2 texts: 176🇩🇪+329🇫🇷+497🇬🇧
  - 585 in language of schooling (LoS): 332🇩🇪+152🇫🇷+101🇬🇧

³https://centre-plurilinguisme.ch/en/research/swiss-learner-corpus-swiko
SWIKO corpus

- Multilingual task-based Swiss learner corpus of 410 oral & 1,803 written texts by lower secondary school students⁴
  - Currently augmented and further analyzed in two projects (WETLAND, DiCoI; 2021–2024)
- Analysis of two sub-corpora of Swiss student texts
  - 1,002 L2 texts: 176🇩🇪+329🇫🇷+497🇬🇧
  - 585 in language of schooling (LoS): 332🇩🇪+152🇫🇷+101🇬🇧

⁴https://centre-plurilinguisme.ch/en/research/swiss-learner-corpus-swiko
SWIKO corpus

- Multilingual task-based Swiss learner corpus of 410 oral & 1,803 written texts by lower secondary school students
  - Currently augmented and further analyzed in two projects (WETLAND, DiCoI; 2021–2024)
- Analysis of two sub-corpora of Swiss student texts
  - 1,002 L2 texts: 176 🇩🇪 + 329 🇫🇷 + 497 🇬🇧
  - 585 in language of schooling (LoS): 332 🇩🇪 + 152 🇫🇷 + 101 🇬🇧

---

3 https://centre-plurilinguisme.ch/en/research/swiss-learner-corpus-swiko
SWIKO corpus

- Multilingual task-based Swiss learner corpus of 410 oral & 1,803 written texts by lower secondary school students³
  - Currently augmented and further analyzed in two projects (WETLAND, DiCoi; 2021–2024)
- Analysis of two sub-corpora of Swiss student texts
  - 1,002 L2 texts: 176\(\text{DE}\)+329\(\text{FR}\)+497\(\text{EN}\)
  - 585 in language of schooling (LoS): 332\(\text{DE}\)+152\(\text{FR}\)+101\(\text{EN}\)
- Corpus characteristics
  - Text length across languages & tasks: \(\mu = 65(\pm 34)\) words
  - Annotated with form-based, local normalization
  - Systematically varies genre, formality, structured-ness

³https://centre-plurilinguisme.ch/en/research/swiss-learner-corpus-swiko
### SWIKO tasks used to elicit data across languages

<table>
<thead>
<tr>
<th></th>
<th>01</th>
<th>02</th>
<th>03</th>
<th>04</th>
<th>05</th>
<th>06</th>
<th>07</th>
<th>08</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Genre</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>arg.</td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>des.</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td><strong>Formality</strong></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>form.</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>inform.</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Structured</strong></td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Linguistic complexity as a way to probe into genre differences?

Zarah Weiss

Motivation

Analysis system
Multilingual analysis

Complexity research
Cross-lingual readability assessment
Proficiency and task assessment

Probing into genre differences

SWIKO corpus
Predicting text genre
Zooming in on features

Summary & outlook

References
### SWIKO tasks used to elicit data across languages

<table>
<thead>
<tr>
<th>Genre</th>
<th>Arg.</th>
<th>Des.</th>
<th>01</th>
<th>02</th>
<th>03</th>
<th>04</th>
<th>05</th>
<th>06</th>
<th>07</th>
<th>08</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formality</td>
<td>Form.</td>
<td>Inform.</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Structured</td>
<td>Yes</td>
<td>No</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

01 answer 5 out of 12 personal questions (e.g., what is your favorite animal/book/food?)
## SWIKO tasks used to elicit data across languages

<table>
<thead>
<tr>
<th></th>
<th>01</th>
<th>02</th>
<th>03</th>
<th>04</th>
<th>05</th>
<th>06</th>
<th>07</th>
<th>08</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Genre</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>arg.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>des.</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Formality</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>form.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✗</td>
</tr>
<tr>
<td>inform.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✗</td>
</tr>
<tr>
<td><strong>Structured</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✗</td>
</tr>
<tr>
<td>no</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✗</td>
</tr>
</tbody>
</table>

02 describe four graphics about animals in Switzerland
SWIKO tasks used to elicit data across languages

<table>
<thead>
<tr>
<th></th>
<th>01</th>
<th>02</th>
<th>03</th>
<th>04</th>
<th>05</th>
<th>06</th>
<th>07</th>
<th>08</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genre</td>
<td>arg.</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>des.</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Formality</td>
<td>form.</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>inform.</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Structured</td>
<td>yes</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>no</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

03 discuss popularity of vacation destinations from a list
SWIKO tasks used to elicit data across languages

<table>
<thead>
<tr>
<th></th>
<th>01</th>
<th>02</th>
<th>03</th>
<th>04</th>
<th>05</th>
<th>06</th>
<th>07</th>
<th>08</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Genre</strong></td>
<td></td>
<td></td>
<td></td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Formality</strong></td>
<td></td>
<td></td>
<td></td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Structured</strong></td>
<td></td>
<td></td>
<td></td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

04 discuss list of top 1–5, 95–100 most important inventions
### SWIKO tasks used to elicit data across languages

<table>
<thead>
<tr>
<th>Task Description</th>
<th>01</th>
<th>02</th>
<th>03</th>
<th>04</th>
<th>05</th>
<th>06</th>
<th>07</th>
<th>08</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Genre</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>arg.</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>des.</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td><strong>Formality</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>form.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>inform.</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td><strong>Structured</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yes</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>no</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

05 write free self-description for a school exchange
**SWIKO tasks used to elicit data across languages**

<table>
<thead>
<tr>
<th></th>
<th>01</th>
<th>02</th>
<th>03</th>
<th>04</th>
<th>05</th>
<th>06</th>
<th>07</th>
<th>08</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Genre</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>arg.</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>des.</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Formality</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>form.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>inform.</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Structured</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yes</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>no</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

06 briefly present one of the presented topics (8 options)
<table>
<thead>
<tr>
<th></th>
<th>01</th>
<th>02</th>
<th>03</th>
<th>04</th>
<th>05</th>
<th>06</th>
<th>07</th>
<th>08</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Genre</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>arg.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>des.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td><strong>Formality</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>form.</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>inform.</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Structured</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yes</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>no</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

07 discuss if school should start and end later (10am–5:30pm)
SWIKO tasks used to elicit data across languages

<table>
<thead>
<tr>
<th>Genre</th>
<th>01</th>
<th>02</th>
<th>03</th>
<th>04</th>
<th>05</th>
<th>06</th>
<th>07</th>
<th>08</th>
</tr>
</thead>
<tbody>
<tr>
<td>arg.</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>des.</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>form.</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>inform.</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Structured</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yes</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>no</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

08 discuss exchanging foreign language classes in school against obligatory language-exchange for six months
Classification experiment

- Predict text genre (arg./des.) from L2 writing in 🇩🇪 🇫🇷 🇬🇧
  - Can we train a model that generalizes across all languages?
  - Does the model generalize form L2 to LoS?
Classification experiment

- Predict text genre (arg./des.) from L2 writing in 🇩🇪 🇫🇷 🇬🇧
  - Can we train a model that generalizes across all languages?
  - Does the model generalize form L2 to LoS?

- Set-up for random forest classifier (best of 5 algorithms)
  1. Use variable features available for 🇩🇪 🇫🇷 🇬🇧: 200/262
  2. Train/test on L2 data with 10-fold cross-validation
  3. Apply model trained on L2 on LoS data

- Classifier performance in accuracy
  - Across folds: 80.4%[79.6; 81.3] (baseline: 51.1%)
  - On LoS data: 84.3% (baseline: 50.1%)
  ⇒ Clear improvement from L2 to LoS texts
Discussion

1. Can we train a model that generalizes across all languages?
   ▶ Yes, our model successfully learns linguistic genre characteristics that generalize across languages

2. Does the model generalize form L2 to language of schooling?
   ▶ Yes, a genre model trained on L2 data generalizes to language of schooling data
   ▶ Higher performance could indicate that linguistic genre differences are more pronounced in LoS data
Which linguistic text differences inform the model?

- Investigate variable importance of features in model
  - Compares avg. accuracy difference with and w/o each feature
  - Normalized with standard error

- Compare top & bottom 35 features grouped by domain & features commonly associated with argumentative texts
Most informative features

- Lexical measures of noun and modifier density and variation
- Clausal elaboration, syntactic variation and word frequency
Noun and modifier density

▶ Descriptive texts show higher noun & lower modifier density than argumentative texts
  ▶ Also representative for noun and modifier variation features
Descriptive texts contain less frequent vocabulary than argumentative texts (across all languages)
Clausal elaboration

- Argumentative texts contain longer, more elaborate clauses
  - Matches contrast argumentative vs. narrative texts in English L2 writing (Yoon & Polio 2017a,b; Biber & Conrad 2019)
  - Effect only for French and English, inverse for German
Least informative features

- Implicit cohesion and human processing measures
  - Partially in line with findings on Merlin/Falko
- Other: relative clauses & PP/VP complexity
Argumentative texts often found to contain more complex NPs than narrative texts in previous research (Yoon & Polio 2017b)

- English: argumentative texts > complex NP than descriptive texts
  - Effect absent in French and German argumentative texts
Descriptive texts contain more pre- and postnominal modification across languages (effect weaker for German)

Languages differ with regard to the amount of modification
Discussion

- Model informed by frequency, syntactic & lexical complexity
  - Descriptive: more variable noun use, rare words
  - Argumentative: more variable modifiers, clausal elaboration
Discussion

- Model informed by frequency, syntactic & lexical complexity
  - Descriptive: more variable noun use, rare words
  - Argumentative: more variable modifiers, clausal elaboration
- Uninformative: human processing, cohesion, & verb MCI
  - Matches Merlin/Falko findings for human processing
  - Cohesion: proficiency level or other task factors?
Discussion

- Model informed by frequency, syntactic & lexical complexity
  - Descriptive: more variable noun use, rare words
  - Argumentative: more variable modifiers, clausal elaboration

- Uninformative: human processing, cohesion, & verb MCI
  - Matches Merlin/Falko findings for human processing
  - Cohesion: proficiency level or other task factors?

- Results partially in line with research on genre differences in English between narrative and argumentative texts
Discussion

- Model informed by frequency, syntactic & lexical complexity
  - Descriptive: more variable noun use, rare words
  - Argumentative: more variable modifiers, clausal elaboration

- Uninformative: human processing, cohesion, & verb MCI
  - Matches Merlin/Falko findings for human processing
  - Cohesion: proficiency level or other task factors?

- Results partially in line with research on genre differences in English between narrative and argumentative texts

- Most features behave similar across languages despite presumed proficiency and language differences
Summary & outlook

1. Broad linguistic complexity modeling helps distinguish text genres for English, German, & French texts
Summary & outlook

1. Broad linguistic complexity modeling helps distinguish text genres for English, German, & French texts

2. Genre generalizes across two instances of languages: L2 & language of schooling
Summary & outlook

1. Broad linguistic complexity modeling helps distinguish text genres for English, German, & French texts

2. Genre generalizes across two instances of languages: L2 & language of schooling

3. Linguistic genre differences in lexicon, syntax and frequency, not in human processing and cohesion
   ▶ Not fully in line with previous findings

⇒ Complex interplay of task factors, proficiency & complexity needs to be investigated in more detail
Linguistic complexity as a way to probe into genre differences?

Zarah Weiss

Motivation

Analysis system
Multilingual analysis

Complexity research
Cross-lingual readability assessment
Proficiency and task assessment

Probing into genre differences
SWIKO corpus
Predicting text genre
Zooming in on features

Summary & outlook

References


References


Linguistic complexity as a way to probe into genre differences?

Zarah Weiss

Motivation

Analysis system

Multilingual analysis

Complexity research

Cross-lingual readability assessment

Proficiency and task assessment

Probing into genre differences

SWIKO corpus

Predicting text genre

Zooming in on features

Summary & outlook

References


B1/B2 tasks in the Merlin corpus

<table>
<thead>
<tr>
<th></th>
<th>01</th>
<th>02</th>
<th>03</th>
<th>04</th>
<th>05</th>
<th>06</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test level</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B1</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B2</td>
<td></td>
<td></td>
<td></td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Genre</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>arg.</td>
<td></td>
<td></td>
<td></td>
<td>×</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>des.</td>
<td></td>
<td></td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Formality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>form.</td>
<td></td>
<td></td>
<td></td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>inform.</td>
<td></td>
<td></td>
<td>×</td>
<td>×</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>Structured/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>code complexity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yes</td>
<td></td>
<td></td>
<td>×</td>
<td>×</td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>no</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>×</td>
<td>×</td>
</tr>
</tbody>
</table>

- B1: announce a visit, happy birthday, happy new year
- B2: Apply for job, work complaint, request job information
Integration costs are higher for texts with B2 scores

Within proficiency levels, they are stable across tasks
B2 test takers write more implicitly cohesive texts
  ▶ Limited within task difference between B1 and B2 texts

Overall differences between B1 and B2 texts can be explained by the elicitation bias of CEFR testing
  ▶ B2 tests were more often taken by B2 than B1 learners
B2 test takers write more implicitly cohesive texts
  ▶ Limited within task difference between B1 and B2 texts

Overall differences between B1 and B2 texts can be explained by the elicitation bias of CEFR testing
  ▶ B2 tests were more often taken by B2 than B1 learners
Discourse Complexity

⇒ B2 test takers write more implicitly cohesive texts
  ▶ Limited within task difference between B1 and B2 texts
⇒ Overall differences between B1 and B2 texts can be explained by the elicitation bias of CEFR testing
  ▶ B2 tests were more often taken by B2 than B1 learners
Language Use

Tasks for B2 testing elicit more rare vocabulary

- Within tasks, texts by B1 and B2 learners do not differ
- This pattern holds for all frequency features we measured

Overall differences between B1 and B2 texts can be explained by the elicitation bias of CEFR testing
Tasks for B2 testing elicit more rare vocabulary
▶ Within tasks, texts by B1 and B2 learners do not differ
▶ This pattern holds for all frequency features we measured

Overall differences between B1 and B2 texts can be explained by the elicitation bias of CEFR testing
Tasks for B2 testing elicit more rare vocabulary
- Within tasks, texts by B1 and B2 learners do not differ
- This pattern holds for all frequency features we measured

Overall differences between B1 and B2 texts can be explained by the elicitation bias of CEFR testing
Clausal Complexity

⇒ B2 texts use more elaborate subordination than B1 texts
  ▶ This holds across measures of clausal complexity

⇒ We see some task-specific variation at B1 and B2
  ▶ Specific tasks elicit more or less clausal elaboration
  ▶ This holds across measures of clausal complexity