Assessing Dutch as a Second Language Learners' Writing Through Syntactic Complexity

Nafal Ossandón, Sarah Bernolet, Orphée De Clercq

Background

Syntactic complexity 🤝 Writing quality

- Syntactic complexity features are informative of writing quality (Crossley & McNamara, 2014)

Dutch as a Second Language (DSL)

- Automatic writing assessment tools have focused so far mostly on English / English as a Second Language
- Early DSL interlanguage: simple NP + VP constructions, repetition of syntactic patterns.

Research Questions

1. Is there longitudinal growth in the syntactic patterns of the DSL participants as measured by syntactic complexity indices?
2. Are these syntactic complexity indices predictive of human ratings of text quality?
3. How much difference do the chosen indices pinpoint between DSL writing and L1 (expert) writing?

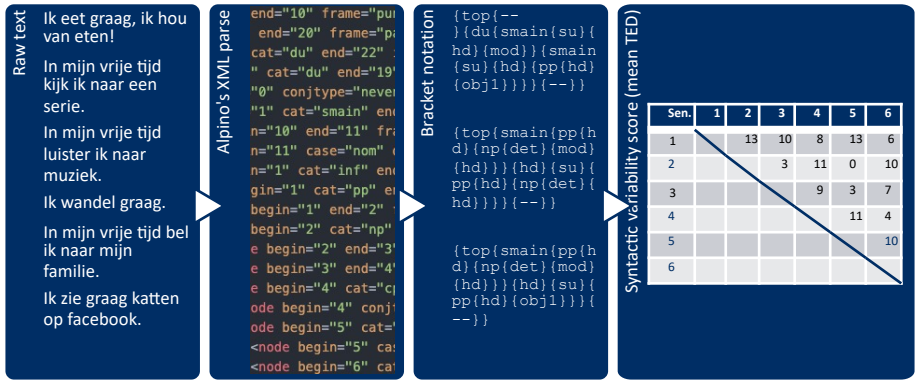
Methods

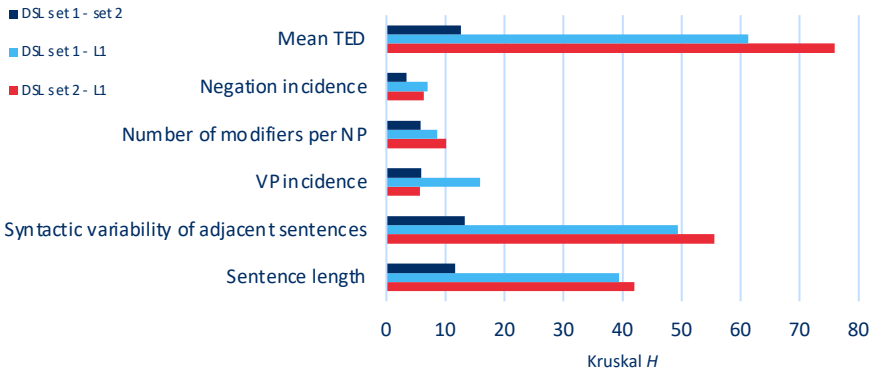
- **Who?** DSL students at Linguapolis' *Dutch as a Foreign Language in an Academic Context* program (N=15)
  - Two measurement moments (Jan '21, Mar '21)
- **Quality?** Text quality evaluations by a Linguapolis instructor + and writings by Dutch native *Multilingual Professional Communication* students (N=16)
- **Features.** Computed by **T-Scan** + engineered with information extracted from **Alpino**'s parse trees
- **Analysis.** Kruskal-Wallis analysis of variance between both DSL + Kendall's Tau correlation between DSL indices and quality scores

Universiteit Antwerpen | Linguapolis

Indices of syntactic complexity	
Word frequency	S-bar incidence
Sentence length	Infinitive clause incidence
Text length in words	VP incidence
Syntactic variability (TED) of adjacent sentences	Number of modifiers per NP
Clause incidence	Number of words before the main verb
PP incidence	Negation incidence
Subject relative clause incidence	Mean Tree Edit Distance (TED)

- ✓ TED between linear representations of XML syntactic trees parsed by Alpino (Van Noord, 2006) measured using the APTED algorithm (Pawlik & Augsten, 2016).
- ✓ Phrase counts and number of words before the main verb extracted from Alpino parsing





Results

- Longitudinal progress: No significant difference
- Quality scores DSL indices: No significant relationship
  - Weak negative association between VP incidence and quality evaluation score (τ=-.58, p=<.01), weak positive association between overall syntactic variability and quality score (τ=.37, p=<.01).
- 6/13 indices show significant difference in complexity between L1 and L2 written texts.

Discussion

- Syntactic complexity indices **don't** seem to be able to measure linguistic **development** in **early** learners of Dutch.
- Syntactic **variability**, measured as the Tree Edit Distance (TED) between the syntactic parse trees of two sentences is indicative of the **quality** of a Dutch text, even for early L2 writers.
- Syntactic variability and indices of nominal complexity can be used as proxies of writing quality assessment.

Limitations

- Exploratory pilot study with a **small** dataset collected within one specific academic-oriented program
- Quality assessment limited to syntactic features.
- Accuracy of the Alpino parser has not been investigated for L2 writing.

Acknowledgements

Thank you Rianne van Lieburg and Edwige Sijyeniyo, Sarah Abattoui and Yasmin Mohamed Moubachir for the datasets, Astrid Elbers for the quality evaluations and Maureen Lambrechts for the insights on DSL evaluation methods.

References

Crossley, Scott A. and Danielle 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.

Pander Maat, H.L.W.; Kraf, R.L.; van den Bosch, Antal; van Gompel, Maarten; Kleijn, S.; Sanders, T.J.M.; van der Sloot, Ko. 2014. T-Scan: a new tool for analyzing Dutch text. *Computational Linguistics in The Netherlands journal*, 4:53–74.

Pawlik, Mateusz and Nikolaus Augsten. 2016. Tree edit distance: Robust and memory-efficient. *Information Systems*, 56:157–173.

Van Noord, Gertjan. 2006. At last parsing is now operational. *TALN06. Verbum Ex Machina. Actes de la 13e conference sur le traitement automatique des langues naturelles*, 20–42.

@nohstns Nafal.OssandonHostens@Student.UAntwerpen.be

GHENT UNIVERSITY | University of Antwerp