

P17-0716:1

Reliability of automatic linguistic annotation: native vs non-native texts

Elena Volodina¹, David Alfter¹, Therese Lindström Tiedemann*, Maisa Lauriala*, Daniela Piipponen*
*=University of Helsinki, ¹= University of Gothenburg

“L2 profiles”

Development of Lexical and Grammatical Competences in Immigrant Swedish

Project financed by

Riksbankens jubileumsfond (2018–2020 (2021))

P17-0716:1

<https://rj.se/anslag/2017/development-of-lexical-and-grammatical-competences-in-immigrant-swedish/>

Collaboration:

Swedish language, University of Gothenburg
“Nordica”, Dept of Finnish, Finnougric and
Scandinavian Studies, University of Helsinki

PI: Elena Volodina, Gothenburg

Homepage:

<https://spraakbanken.gu.se/en/projects/l2profiles>



Swedish (L2) Profile

– lexical, grammatical, morphological.

Exercise Generator

Tools

Experimental prototypes

Swedish Profile



<https://spraakbanken.gu.se/larkalabb/svlp>

Language Acquisition Reusing **Korp**

Lexical profile

Grammatical profile

Morphological profile

Swedish (L2) Profile

– lexical, grammatical, morphological.

Exercise Generator

Tools

Experimental prototypes

Swedish Profile



[http](http://larkalabb/svlp)

[/larkalabb/svlp](http://larkalabb/svlp)

Language Acquisition Reusing **Korpus**

Reliable?

Lexical profile

Grammatical profile

Morphological profile

Manual checking

Sparv pipeline (Borin et al. 2016)

- Lemmatization
- Part-of-speech tagging
- Multi-word expression detection
- Word sense disambiguation
- Dependency parsing

Course book texts

Learner essays

Corrected
Learner essays

15 texts
per
dataset;

3 texts per
proficiency
level

Hypotheses

1. Pipelines trained on a standard language (L1) do not perform as well on non-standard language varieties such as learner language (L2)
2. Normalization of non-standard language, e.g. through error correction, improves tool performance

Hypotheses

1. Pipelines trained on a standard language (L1) do not perform as well on non-standard language varieties such as learner language (L2)

2. Normalization of non-standard language, e.g. through error correction, improves tool performance

Corpus	Lemma	PoS	DepRel
L1 Coctail	0.93±0.0	0.98±0.0	74.89
L2 orig	0.90±0.02	0.95±0.0	63.01
L2 norm	0.93±0.02	0.97±0.0	69.02

Table 2: Lemmatization and PoS tagging: accuracy and standard deviation; Dependency: LAS

Hypotheses

1. Pipelines trained on a standard language (L1) do not perform as well on non-standard language varieties such as learner language (L2)

2. Normalization of non-standard language, e.g. through error correction, improves tool performance

Corpus	Lemma	PoS	DepRel
L1 Coctail	0.93±0.0	0.98±0.0	74.89
L2 orig	0.90±0.02	0.95±0.0	63.01
L2 norm	0.93±0.02	0.97±0.0	69.02

Table 2: Lemmatization and PoS tagging: accuracy and standard deviation; Dependency: LAS

	#tokens excl punct	MWE			WSD
		Precision	Recall	F1	Accuracy (corr/tot)
L1 Coctail	1900	0.80	0.71	0.75	0.84±0.03
L2 orig	3635	0.90	0.72	0.80	0.82±0.07
L2 norm	3565	0.85	0.78	0.81	0.83±0.04

Table 3: Number of correctly identified MWEs including precision, recall and F1 score and automatic word sense disambiguation (WSD)

Conclusions

Reliable?

- yes, for lemmatization, POS tagging, WSD & MWE detection
- not really for DepRel

Hypotheses?

1. yes, in general, but depend on linguistic features
2. yes, even if sometimes marginally

Thank you!

Questions? Comments? Suggestions?



[Link to the presentation](#)