

Investigating the scope of textual metrics for learner level discrimination and learner analytics



Nicolas Ballier



Thomas Gaillat

Problem statement

Learning a language

- For individuals > requires regular assessments for both learners and teachers
- For institutions > a growing demand to group learners homogeneously & fast
- Assessment within CEFR framework

Need for automatic assessment tools

Problem statement

Automatic Language assessment tools

- Traditional method: labour intensive, short-context and rule-based exercises
 - focus on specific errors or language forms
 - No reverse engineering for feature explanation
- Supervised learning methods
 - Criterial features: Syntactic, semantic and pragmatic
 - Criterial features: complexity metrics
- Meaningful linguistic feedback for learners
 - Scope: word, sentence and text

Research Questions

- Which scopes do textual metrics have ?
- Which metrics & scopes can be identified as predictors for CEFR levels?

Outline

1. Previous work on language assessment
2. Method
 - 2.1. Corpus
 - 2.2. Annotation and metrics
 - 2.3. Metrics and scopes for feedback
 - 2.4. Processing and data set
 - 2.5. Experimental setup
3. Results
4. Discussion and next steps

Previous work in level assessment

Converging methods in automatic learner language analysis

- Automatic Scoring Systems for learner language (Shermis et al., 2010; Weigle, 2013) > Shared-tasks: Spoken CALL (Baur et al., 2017) & CAP18 (Arnold, et al. 2018)
 - Automatic learner error analysis (Leacock, 2015)
 - Automatic learner language analysis: criterial features (Crossley et al, 2011; Hawkins & Filipović, 2012) , complexity metrics (Lu 2010, 2012; Lissón and Ballier 2018)
- > Our proposal: Criterial features for meaningful feedback to learners

Method

- Corpus
- Annotation and metrics
- Metrics and scopes for feedback
- Processing and data set
- Experimental setup

Corpus

CELVA.Sp (University of Rennes). L2 Corpus of English for Specific Purposes (Pharmacy, Computer Science, Biology, Medicine)

- French L1 component
- 55 000 word learner corpus
- 282 Written essays of different individuals (students L1 to M1)
- Mapped onto the six CEFR levels with DIALANG

Number of writings	A1	A2	B1	B2	C1	C2
French L1	27	63	125	43	19	3

Annotation & metrics

Annotation and metrics tools

- TreeTagger (POS tagging) (Schmid 1994)
- L2SCA (Lu 2010)
- R Quanteda library: texstat (lexdiv & readability) (Benoit 2018)

Metrics

- Syntactic e.g. amount of coordination, subordination
- Lexical diversity e.g. density, sophistication
- Readability (level of difficulty of a text)

Metrics and scopes: a taxonomy for learner feedback

- We classify according to the types of variables used in each formula
- 3 examples
 - $ARI = 0.5ASL + 4.71AWL - 21.34$
 - Word.size.characters (one of the variables focuses on word size in relation to characters)
 - Sentence.size.words (one of the variables focuses on sentence size in relation to words)
 - CN/C = Complex Nominals / Clauses
 - Sentence.component.component (one of the variables focuses on a specific sentence component in relation to another)
 - $W =$ Total number of words in a text
 - Text.size.words (one of the variables focuses on text size in relation to words)

Metrics and scopes: a taxonomy for learner feedback

Word.size.characters: ARI ARI.simple Bormuth Bormuth.GP Coleman.Liau Coleman.Liau.grade Coleman.Liau.short Dickes.Steiwer
DRP, Fucks, nWS nWS.2, Traenkle.Bailer Traenkle.Bailer.2 Wheeler.Smith

Word.size.syllables: Coleman Coleman.C2 meanWordSyllables, Farr.Jenkins.Paterson, Flesch Flesch.PSK Flesch.Kincaid FOG
FOG.PSK FOG.NRI FORCAST FORCASTRGL, Linsear.Write, LIW, nWS nWS.2 nWS.3 nWS.4 Wheeler.Smith

Sentence.size.words: (n words/nsent) MLS MLT MLC ARI family, Bormuth family, Dale.Chall family, Farr.Jenkins.Paterson Fucks WS.3
nWS.4 Flesch Flesch.PSK Flesch.Kincaid FOG FOG.PSK

Sentence.size.characters: Danielson.Bryan family, Dickes.Steiwer,

Sentence.size.syllables : DRP ELF Flesch Flesch.PSK Flesch.Kincaid FOG FOG.PSK FOG.NRI RIX SMOG SMOG.C
SMOG.simple Strain

Sentence.components: Verb Phrase (VP) Clauses (C) T-Units (T) Dependent Clauses (DC) Coordinate Phrases (CP) Complex
Nominals (CN)

Sentence.components.components: C/S (Sentences) VP/T C/T DC/C DC/T T/S CT CT/T CP/T CP/C CN/T CN/C
Traenkle.Bailer family (prepositions & conjunctions)

Text.size.words: W

Text.size.sentences: S Coleman.Liau family (n sentences/n words) Linsear.Write

Text.variation.words: TTR C (Log TTR) R (root TTR) CTTR U S Maas lgV0 lgeV0, Dickes.Steiwer

Text.repetitions.types: Yule's K Simpson's D Herdan's Vm

Text.sophistication.wordsDaleChallList: Bormuth Bormuth.GP Bormuth, Dale.Chall family, DRP, Scrabble

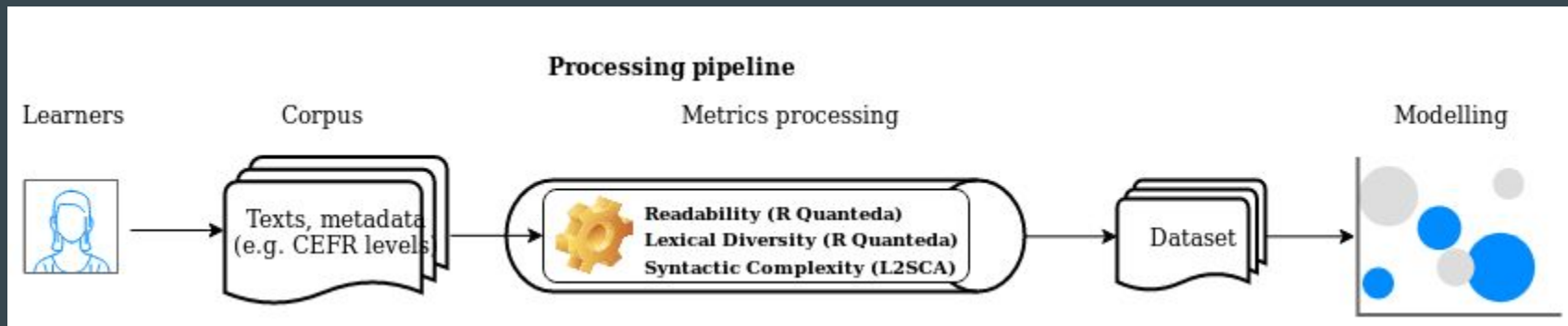
Text.sophistication.wordsSpacheList: Spache Spache.old

Processing pipeline and Data set

Features


- 83 metrics for each text
- Outcome variable: CEFR levels

Pipeline



Experimental setup

Supervised learning approach

- Dataset: Train (80%) and test set (20%) - random selection
- Method: R RandomForest (ntree = 500, mtry = 6) 

Stage 1: Classification in CEFR levels

Stage 2: Feature explanation per CEFR level

Results: Stage 1

6 CEFR class classification

- Poor

3 CEFR classes A, B, C

	precision	Recall	F1-score
A	0.5	0.53	0.51
B	0.84	0.78	0.81
C	0.33	1	1

Results: Stage 1 (2)

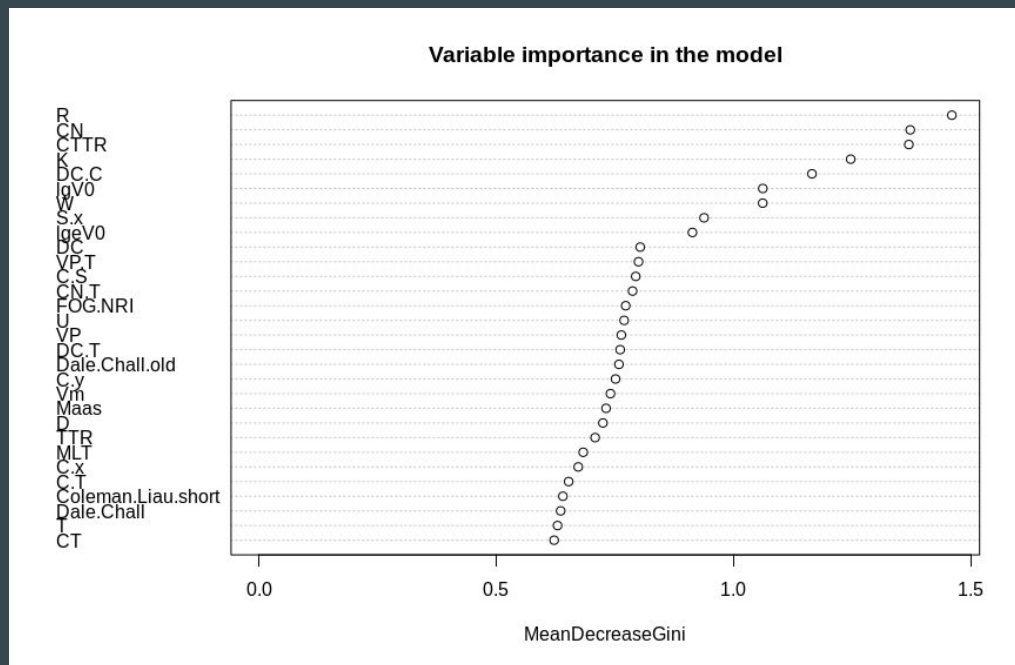
B1 vs B2 class

	Precision	Recall	F1-score
B1	0.91	0.73	0.81
B2	0.20	0.50	0.28

Stage 2: Feature explanation

Significant features for B level:

- Root & Corrected & log TTR
- Complex Nominal (CN)
- Yule's K
- Dependent clauses/clauses
- Number of Words, sentences



Results: Scopes and predictability

B learners are sensitive to:

Metrics	Scope
Root & Corrected & log TTR	Word.density
Complex Nominals (CN)	Sentence.component
Dependent clauses/clauses	Sentence.component.component
Number of Words, sentences	Text.size.words Text.size.sentences
Yule's K	Text.repetitions.types

Discussion

Linguistic features correlate with levels which gives and insight in Interlanguage.
Features linked to scopes can be used to advise learners.

But...

Refine scope taxonomy

Binary classification in 2 levels: beginner (A1,A2,B1) and advanced (B2,C1,C2)

More metrics: Syntagmatic relationships in ratios BUT what about features based on paradigmatic relationships ?

Next steps

- More LIs, features & texts for training and testing
- Linguistic microsystems as metrics
- Building an online system for CEFR prediction (Ulysses PHC project: Universities of Paris Diderot and Insight NUI Galway Ireland)

Thank you

thomas.gaillat@univ-rennes2.fr

nicolas.ballier@univ-paris-diderot.fr

References

- Chen, Xiaobin, and Detmar Meurers. 2016. "Characterizing Text Difficulty with Word Frequencies." *Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications*, 84–94.
- Crossley, S. A., Salsbury, T., McNamara, D. S., & Jarvis, S. (2011). Predicting lexical proficiency in language learner texts using computational indices. *Language Testing*, 28(4), 561–580.
- Crossley, Scott A., Tom Salsbury, Danielle S. McNamara, and Scott Jarvis. 2011. "Predicting Lexical Proficiency in Language Learner Texts Using Computational Indices." *Language Testing* 28 (4): 561–80.
- Díaz-Negrillo, Ana, Nicolas Ballier, and Paul Thompson, eds. 2013. *Automatic treatment and analysis of learner corpus data. Studies in Corpus Linguistics 59*. Amsterdam, Pays-Bas, Etats-Unis: John Benjamins Publishing Co.
- Ellis, Rod. 1994. *The Study of Second Language Acquisition*. Oxford, United Kingdom: Oxford University Press.
- Geertzen, Jeroen, Theodora Alexopoulou, and Anna Korhonen. 2013. "Automatic Linguistic Annotation of Large Scale L2 Databases: The EF-Cambridge Open Language Database (EFCamDat)." In *Proceedings of the 31st Second Language Research Forum*, edited by R. T. Miller, K. I. Martin, C. M. Eddington, A. Henery, N. Miguel, A. Tseng, A. Tuninetti, and D. Walter. Carnegie Mellon: Cascadilla Press.
- Granger, Sylviane, Gaëtanelle Gilquin, and Fanny Meunier, eds. 2015. *The Cambridge Handbook of Learner Corpus Research*. Cambridge: Cambridge University Press.
- Hawkins, John A., and Luna Filipović. 2012. *Criterial Features in L2 English: Specifying the Reference Levels of the Common European Framework*. United Kingdom: Cambridge University Press.
- Khushik, Ghulam Abbas, and Ari Huhta. 2019. "Investigating Syntactic Complexity in EFL Learners' Writing across Common European Framework of Reference Levels A1, A2, and B1." *Applied Linguistics* amy064.
- Kim, Minkyung, and Scott A. Crossley. 2018. "Modeling Second Language Writing Quality: A Structural Equation Investigation of Lexical, Syntactic, and Cohesive Features in Source-Based and Independent Writing." *Assessing Writing* 37: 39–56.
- Kyle, Kristopher, Scott Crossley, and Cynthia Berger. 2018. "The Tool for the Automatic Analysis of Lexical Sophistication (TAALES): Version 2.0." *Behavior Research Methods* 50 (3): 1030–46.
- Lissón, Paula, and Nicolas Ballier. 2018. "Investigating Lexical Progression through Lexical Diversity Metrics in a Corpus of French L3." *Discours. Revue de Linguistique, Psycholinguistique et Informatique. A Journal of Linguistics, Psycholinguistics and Computational Linguistics*, no. 23. <https://doi.org/10.4000/discours.9950>.
- Lu, Xiaofei. 2010. "Automatic Analysis of Syntactic Complexity in Second Language Writing." *International Journal of Corpus Linguistics* 15 (4): 474–496.
- . 2012. "The Relationship of Lexical Richness to the Quality of ESL Learners' Oral Narratives." *The Modern Language Journal* 96 (2): 190–208.
- . 2014. *Computational Methods for Corpus Annotation and Analysis*. Dordrecht: Springer.
- Pilán, Ildikó, and Elena Volodina. 2018. "Investigating the Importance of Linguistic Complexity Features across Different Datasets Related to Language Learning." In *Proceedings of the Workshop on Linguistic Complexity and Natural Language Processing*, 49–58. Santa Fe, New-Mexico: Association for Computational Linguistics.
- Tono, Yukio. 2013. "Automatic Extraction of L2 Criterial Lexicogrammatical Features across Pseudo-Longitudinal Learner Corpora: Using Edit Distance and Variability-Based Neighbour Clustering." In *L2 Vocabulary Acquisition, Knowledge and Use: New Perspectives on Assessment and Corpus Analysis*, edited by Camilla Bardel, Christina Lindqvist, and Batia Laufer, 149–176. *Eurosla Monographs Series 2. The European Second Language Association*.
- Weigle, S. C. (2013). English language learners and automated scoring of essays: Critical considerations. *Assessing Writing*, 18(1), 85–99.

Many thanks to:

Xiaofei Lu

Detmar Meurers

Schmid (Treetagger)

Syntactic complexity metrics

meanSentenceLength meanWordSyllables W S VP C T DC CT CP
CN MLS MLT MLC C/S VP/T C/T DC/C DC/T T/S CT/T CP/T
CP/C CN/T CN/C

Readability metrics

ARI ARI.simple Bormuth Bormuth.GP Coleman Coleman.C2 Coleman.Liau
Coleman.Liau.grade Coleman.Liau.short Dale.Chall Dale.Chall.old
Dale.Chall.PSK Danielson.Bryan Danielson.Bryan.2 Dickes.Steiwer DRP ELF
Farr.Jenkins.Paterson Flesch Flesch.PSK Flesch.Kincaid FOG FOG.PSK
FOG.NRI FORCAST FORCAST.RGL Fucks Linsear.Write LIW nWS
nWS.2 nWS.3 nWS.4 RIX Scrabble SMOG SMOG.C SMOG.simple
SMOG.de Spache Spache.old Strain Traenkle.Bailer Traenkle.Bailer.2
Wheeler.Smith

Lexical diversity metrics

TTR C_x R CTTR U S_x K D V_m Maas (a , $\log V_0$ $\log eV_0$)

<https://www.rdocumentation.org/packages/quanteda/versions/0.9.7-17/topics/lexdiv>

https://quanteda.io/reference/textstat_lexdiv.html

L2SCA component definitions

- **Sentence:** a group of words (including sentence fragments) punctuated with a sentence-final punctuation mark, such as a period, question mark, or exclamation mark.
- **Clause:** a structure with a subject and a finite verb, such as an independent, adjective, adverbial, or nominal clause (see, e.g., Hunt 1965; Polio 1997). Non-finite verb phrases are not counted as clauses.
- **Dependent clause:** a finite adjective, adverbial, or nominal clause (e.g., Cooper 1976; Hunt 1965; Kameen 1979).
- **T-unit:** “a main clause plus any subordinate clause or non-clausal structure that is attached to or embedded in it” (Hunt 1970, p. 4).
- **Complex T-unit:** a T-unit with one or more dependent clauses (see, e.g., Casanave 1994).
- **Coordinate phrase:** a coordinate adjective, noun, or verb phrase.
- **Complex nominal:** a noun plus an adjective, possessive, prepositional phrase, adjective clause, participle, or appositive; a nominal clause; or a gerund or infinitive in subject position (see, e.g., Cooper 1976).
- **Verb phrase:** a finite or nonfinite verb phrase.