



HAL
open science

Investigating learners' progression in French as a Foreign Language: vocabulary growth and lexical diversity

Paula Lissón, Nicolas Ballier, English Linguistics

► To cite this version:

Paula Lissón, Nicolas Ballier, English Linguistics. Investigating learners' progression in French as a Foreign Language: vocabulary growth and lexical diversity. CUNY Student Research Day, Mar 2018, New York United States. hal-01705828

HAL Id: hal-01705828

<https://hal.science/hal-01705828>

Submitted on 26 Feb 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Investigating learners' progression in French as a Foreign Language: vocabulary growth and lexical diversity

Paula Lissón, Nicolas Ballier

Université Paris Diderot (USPC) – CLILLAC-ARP (EA 3967) – Department of English Linguistics

ABSTRACT

This poster presents the use of lexical diversity metrics (cf. Lu, 2012; Jarvis, 2013; Bentz & Buttery, 2014) and vocabulary growth curves (cf. Evert & Baroni, 2006; Ballier & Gaillat, 2016) for the assessment of lexical progression in learners of French as a Foreign Language.

In order to investigate how these two techniques monitor lexical progression, an experiment was carried out with a semi-longitudinal corpus of learners' written productions of French L3.

Results show that in our corpus, lexical diversity metrics only capture an increase in vocabulary from the first to the third production of each student, whereas vocabulary growth curves capture lexical progression across the three productions.

These techniques are presented as a complementary and automatic way to assess lexical progression in learners' written assignments.

CONTACT

Paula Lissón (MA student)
 English Linguistics,
 Nicolas Ballier (Prof. of
 English Linguistics)

 Université Paris Diderot,
 CLILLAC-ARP (EA 3967)
paula.lisson@etu.univ-paris-diderot.fr
nicolas.ballier@univ-paris-diderot.fr

Introduction

- ❖ corpus-based evaluation of 14 lexical diversity metrics as measures of longitudinal progression in written productions
- ❖ random forests (Breiman, 2001) are used to see whether lexical diversity metrics' scores capture enough vocabulary diversity progression to predict the production wave.
- ❖ vocabulary growth curves and vocabulary growth ratios are also calculated for each production wave

Description of the corpus

- Spanish University students enrolled in English studies (major) and in French studies (minor).
- 24 students, A1 level of the CEFRL, beginners
- Three waves of written productions were collected (October – November – December). Between 100 and 150 words.
- Topics:
 - a) describe your house,
 - b) talk about a famous person,
 - c) explain a film

Metrics

Table 1. Lexical diversity metrics used in the study

Metric	Formula
TTR	V/N
MSTTR	V/N (fragments of n tokens)
MTLD	$V/\text{factors}$ (segments with the stabilization point of TTR)
MATTR	Mean of moving TTR (window technique)
MTLD-MA	Factors and window technique combined
Herdan's C	$\log V / \log N$
Guiraud's RTTR	V/\sqrt{N}
Carrol's CTTR	$V/2\sqrt{N}$
Uber Index (U)	$(\log N)^2 / \log N - \log V$
Summer's Index (S)	$\log(\log V) / \log(\log N)$
Yule's K	$K = 10^4 \frac{[\sum_{m=1}^N V(m, N)(m/N)^2] - N}{N^2}$
Maas a	$a^2 = (\log N - \log V) / \log N^2$
Maas log	$\log V_0 = \log V / \sqrt{1 - \frac{\log V^2}{\log N}}$
HDD-D	For each type, the probability of finding any of its tokens in a random sample of 42 words taken from the same text

Vocabulary Growth Curves (VGC) and vocabulary growth ratios

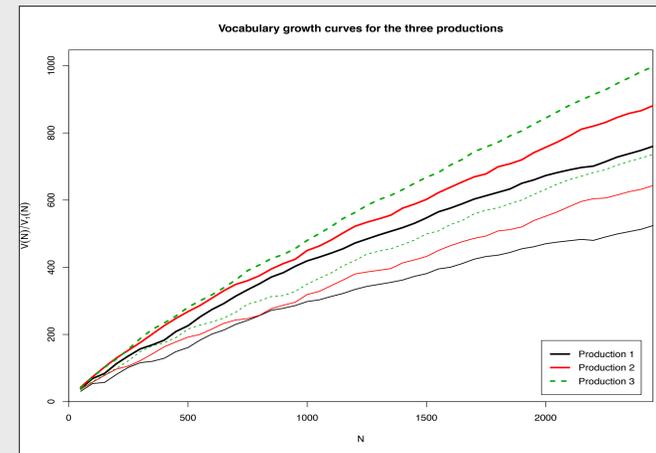


Figure 2: VGC of the three waves of productions and their hapax legomena

Table 3. Vocabulary growth rates for the three productions

	Tokens	Types	f	Growth rate	LNRE model	X2	df	p
Production1	2,464	762	526	0.21347***	fzm	3.26	5	0.7
Production2	2,464	883	644	0.26136***	gigp	5.56	6	0.5
Production1	2,464	762	526	0.21347***	fzm	3.27	5	0.6
Production3	2,464	1,000	736	0.29870***	fzm	8.01	5	0.2
Production2	2,464	883	644	0.26136***	gigp	7.54	6	0.3
Production3	2,464	1,000	736	0.29870***	fzm	8.01	5	0.2

Results (random forests)

Table 2. Confusion matrix of results

n=64	P1	P2	P3
P1	17	1	8
P2	4	15	0
P3	2	5	12

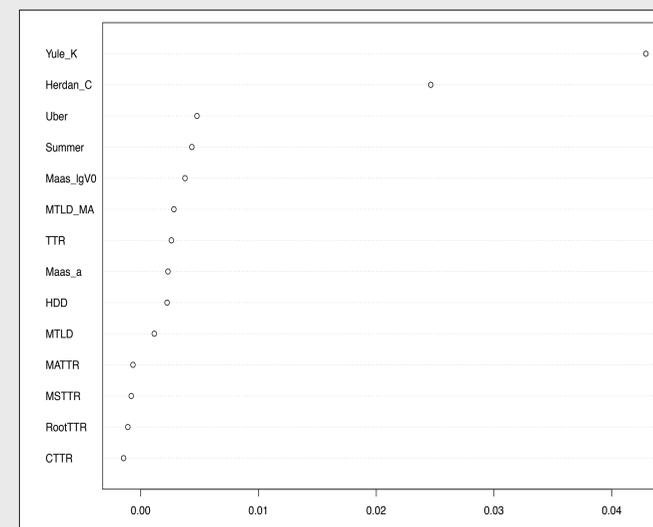


Figure 2. Variable importance in the random forest model

Conclusions & Discussion

1. VGCs and growth rates can be used to assess learners' progression. However, these methods present one major caveat: they do not detect learners' errors such as spelling mistakes, invented or adapted words (mostly related to L1 and L2 influence), which are computed as new types.
2. In our study, learners show a positive progression. However, metrics only reflect this progression from the 1st to the 3rd production, whereas VGC clearly show progression across the 3 productions.
3. VGC and growth ratios appear to be more precise for the detection of semi-longitudinal lexical progression.
4. But, in VGC all learners are considered as a subset of the corpus (all texts cumulated), results might be different assessing VGCs of individual learners.

References

- Baayen, R. H. (2007). Package languageR: data sets and functions for 'Analyzing Linguistic Data' (Version 1.0).
- Ballier, N., & Gaillat, T. (2016). Classification d'apprenants francophones de l'anglais sur la base des métriques de complexité lexicale et syntaxique (Vol. 9, pp. 1–14). Presented at the JEP-TALN-RECITAL 2016.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Evert, S., & Baroni, M. (2006). The zipfR library: Words and other rare events in R. Presented in *all'useR*.
- McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of Michalke, M. (2017). Package koRpus: An R Package for Text Analysis (Version 0.06-5). Retrieved from <http://reaktan.de/?c=hacking&s=koRpus>
- Osborne, J. (2015). Transfer and learner corpus research. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge Handbook of Learner Corpus Research* (pp. 333–356). Cambridge University Press.