Phraseological Complexity in Oral and Written L2 French: Work in progress report

Nate Vandeweerd | Magali Paquot | Alex Housen
LCR, Warsaw, Sept 13 2019
L2 French Phraseology

Mostly top-down, manual identification of formulaic language:

- **quantity** of formulaic language increases from A2<B2<C2 (Forsberg & Bartning, 2010)
- Advanced long-term residency L2 French speakers have less **diversity** of multi-word units compared to NS (Erman et al. 2015)
- Increase in **productive collocation knowledge** of VERB+DOBJ collocations from B2 < C1 levels (Forsberg Lundell et al., 2018)

The Gap

Phraseological **complexity** (diversity and sophistication) in L2 French.
Phraseological complexity (Paquot, 2019)

**Data:** Vespa Corpus
- L2 English
- 98 linguistics essays
- holistic CEFR grades

**Method**

- Phraseological complexity measures
  - diversity: RTTR (root type token ratio)
  - sophistication: Mean MI (mutual information) score, Academic Collocation List
- Lexical measures → Lexical Complexity Analyzer (Lu, 2012)
- Syntactic measures → L2 Syntactic Complexity Analyzer (Lu, 2010)

Regression model to predict CEFR Level
Phraseological complexity (Paquot, 2019)

Results

- lexical and syntactic complexity measures: non-significant increase
- phraseological complexity measures:
  - diversity: no significant increase in RTTR
  - sophistication:
    - ADV+VERB: significant increase in mean MI (B2 → C1/C2)
    - ADJ+NOUN: significant increase in mean MI (B2 → C2)
    - VERB+DOBJ: significant increase in mean MI (B2/C1 → C2)
- Variance (regression model):
  - Mean MI ADJ+NOUN (11%)
  - Mean MI ADV+VERB (14%)
  - Mean MI VERB+DOBJ (15%)
Research Questions

RQ1: How does phraseological complexity compare in L2 French at different proficiency levels?

RQ2: To what extent does phraseological complexity relate to lexical complexity in L2 French?
The Current Study: Partial replication of Paquot (2019)

Data: Leerdercorpus Frans
- L2 French (L1 Dutch)
- Argumentative essays
- Texts assessed by trained CEFR Raters → B2 (n=29), C1(n=109), C2(n=36)

Method

- Phraseological complexity measures (Hubert Naets, CENTAL)
  - diversity: RTTR (root type token ratio)
  - sophistication: Mean MI (based on FRCOW16 corpus)
- Lexical complexity measures (Thomas François, CENTAL)
  - diversity (TTR/100w)
  - sophistication (absent words from Gougenheim 2000)
Preliminary Results:

RQ1. Proficiency ~ Phraseological measures
Phraseological diversity by level

*p<0.05

(r = 0.03)
Dependencies per level:

nathan.vandeweerd@uclouvain.be

Phraseological sophistication by level

*p < 0.05

(r = 0.20)

(r = 0.30)
Preliminary Results:

RQ2. Phraseological vs. lexical measures

nathan.vandeweerd@uclouvain.be
Random Forest Model (1500 permuted trees)

- party::cforest
- Phraseological complexity measures
  - diversity: RTTR (root type token ratio)
  - sophistication: Mean MI (based on FRCOW16 corpus)
- Lexical complexity measures
  - diversity (TTR/100w)
  - sophistication (absent words from Gougenheim 2000)
Model Predictions

- Baseline: 66%
- Model accuracy: 77%
- Accuracy > Baseline (p < 0.001)

<table>
<thead>
<tr>
<th>Predictions</th>
<th>B2</th>
<th>C1</th>
<th>C2</th>
</tr>
</thead>
<tbody>
<tr>
<td>B2</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C1</td>
<td>20</td>
<td>111</td>
<td>18</td>
</tr>
<tr>
<td>C2</td>
<td>0</td>
<td>0</td>
<td>18</td>
</tr>
</tbody>
</table>
Preliminary results

**RQ1:** How does phraseological complexity compare in **L2 French** at different proficiency levels?

- Significant (but small) increase in phraseological diversity (ADJ+NOUN) from C1→ C2 → in contrast to Paquot (2019)
- Slight increase in sophistication for proficiency levels (but only significant for ADV+VERB from B2/C1 → C2) → same as Paquot (2019) (but NS for VERB+DOBJ, ADJ+NOUN)

**RQ2:** To what extent does phraseological complexity relate to lexical complexity in **L2 French**?

- Distinguishing between C1 and C2 level texts seems to be at least partially dependent on diversity and sophistication of ADJ+NOUN and diversity of VERB+DOBJ
- but lexical sophistication also plays a role → in contrast to Paquot (2019)
Discussion
Reasons for divergent results from Paquot (2019)?
• measures too coarse-grained?
• topic effect? (Paquot, Naets Gries, forthcoming)
• stylistic preferences of French (Vinay & Darbelnet, 1995, p. 101)

Verb phrases:
to review → passer en revue

Adjectival phrases:
a hopeless undertaking → une entreprise sans espoir
medical students → des étudiants en médecine

Adverbial phrases:
answered angrily → répondu avec colère
Next Steps

- integrate syntactic measures (possibly more important in distinguishing B/C)
- integrate more (better) measures of diversity and sophistication
  - at the lexical level (cf. Jarvis, 2013)
  - at the phraseological level
    - more aspects of diversity
    - collocational bands (cf. Paquot, 2018)
    - association as a multi-dimensional phenomenon (Gries, 2019)
    - or diversity within collocation bands?
- qualitative look at the phraseological units being used in the learner texts
  → measurement validity
  - looking beyond 2-word combinations?
    - prepositional modifiers (passer *en revue*, une entreprise *sans espoir*)
  → lexical bundles?
Eh?
References


