Research Visit to Louvain-la-Neuve

Elizaveta Smirnova

Plan

- Centre and its team
- Methods of studying corpora
- Prospects for the laboratory

When?
Where?
Why?

- November 10-24, 2019
- The Centre for English Corpus Linguistics, at the Université catholique de Louvain
- Research Visits to foreign universities funded by the HSE Scientific Fund:

- To familiarize with the academic environment of European universities

- To establish contacts with foreign scholars

- To carry out some research using a foreign university's facilities

CECL

The Centre for English Corpus Linguistics (CECL) specializes in the collection and use of corpora for linguistic and pedagogical purposes. Its main areas of focus are learner and multilingual corpora. At the end of the 1980s, the CECL pioneered the study of learner corpora with the *International Corpus of Learner English* (ICLE). A truly international enterprise, ICLE contributed to the integration of corpora in language acquisition studies and to their use in various applications ranging from language teaching to natural language processing (https://uclouvain.be/en/research-institutes/ilc/cecl/about.html).

CECL Team

- Fanny Meunier
- Magali Paquot
- Gaetanelle Gilquin
- Sylviane Granger

Feedback on our project

Rationale behind choosing the markers

Assessment – how can it help learners to improve their texts? Tips?

State-of-the-art Research

- Complexity in SLR:
  - Influence of task complexity
  - Quantitative ways of describing complexity
  - Automated tools for capturing complexity
  - Complexity of different proficiency levels
  - Complexity as a predictor of L2 competence
- LEAD

Measuring phraseological complexity

Phraseological units under investigation: grammatical dependencies

- (verb + noun) in a direct object relation
- (adjective/noun + noun) in an adjectival modifier head relation
- (adjective/adverb + adjective/adverb/verb) in an adverbial modifier head relation

Diversity	Sophistication
Root type-token ratios of grammatical dependencies [T/√N]	Pointwise mutual information (MI) means for three grammatical dependencies
	Proportion of grammatical dependencies in collocational bands (MI-based)
	Ratio of academic to total dependencies

Paquot (2018, 2019)

Measuring phraseological complexity

Phraseological units under investigation: grammatical dependencies

- (verb + noun) in a direct object relation
- (adjective/noun + noun) in an adjectival modifier head relation
- (adjective/adverb + adjective/adverb/verb) in an adverbial modifier head relation

MI scores of adverbial modifier dependencies in L2 English

0 < MI < 1:	*clearly negative; clearly described; important enough; measure + typically; represent + directly; very theoretical*
6 < MI:	*mutually exclusive; fiercely debated; scarcely tenable; evenly distributed; firmly rooted*

Collocational bands	
Below threshold	Dependency appears fewer than 5 times in the reference corpus
Non-collocational	MI < 3
Collocational: low	3 ≤ MI < 5
Collocational: medium	5 ≤ MI < 7
Collocational: high	MI ≥ 7

Research Questions

- RQ 1: To what extent do phraseological complexity measures contribute to the prediction of L2 Dutch assessment across the B1 and B2 CEFR levels?
- RQ 2: How do measures of phraseological complexity compare to traditional measures of syntactic and lexical complexity in the assessment of L2 Dutch proficiency?
- RQ 3: How does the role of phraseological complexity in L2 Dutch assessment compare to the results observed for L2 English in Paquot (2018)?

Regression model

Model fit

Model $\chi^2(27) = 378.88$ p < 0.01 | Nagelkerke's R^2: .38

Predictors	B	S.E.	β	O.R.	z	p-value
Intercept**	-15.98	5.46	1.45	4.24	-2.93	< 0.01
CEFR B2	0.34	6.33	-0.38	0.69	0.05	0.96
Morphemes per word	-1.56	2.9	-0.1	0.91	-0.54	0.59
Proportion freq5000 words	-0.23	2.65	-0.02	0.98	-0.09	0.93
MTLD lemma	-0.03	0.02	-0.37	0.69	-1.57	0.12
PMI mean dobj**	0.34	0.08	0.09	1.45	4.13	< 0.01
PMI mean amod*	0.15	0.08	0.19	1.21	2	0.05
PMI mean advmod**	0.71	0.16	0.38	1.46	4.35	< 0.01
RTTR dobj**	0.56	0.13	0.44	1.55	4.26	< 0.01
RTTR advmod**	1.44	0.36	1.24	3.45	4	< 0.01
High PMI dobj*	1.29	0.52	0.19	1.21	2.48	<0.05
High PMI amod	-0.55	0.34	-0.53	0.59	-1.61	0.11
Words per sentence*	1.85	0.72	0.60	1.83	2.56	< 0.05
Coordinated sub clause*	1.06	0.42	0.23	1.26	2.53	< 0.05
D-level above 4	-0.57	0.31	-0.43	0.65	-1.85	0.06
Density nominalizations	0.13	0.16	0.27	1.31	0.82	0.41
Density passives**	0.22	0.06	0.3	1.35	3.76	< 0.01
DD subject verb**	1	0.28	0.32	1.38	3.63	< 0.01
DD article noun	0.83	0.51	0.15	1.17	1.64	0.1
DD verb complement	-0.42	0.22	-0.14	0.87	-1.88	0.06
B2:Morphemes per word	5.75	3.47	0.35	1.42	1.66	0.1
B2:freq5000	6	3.13	0.47	1.59	1.91	0.06
B2:MTLD lemma**	0.05	0.02	0.68	1.98	2.69	< 0.01
B2:RTTR advmod*	-0.95	0.37	-0.81	0.44	-2.53	< 0.05
B2:High PMI amod	0.62	0.35	0.6	1.82	1.77	0.08
B2:Words per sentence*	-1.98	0.79	-0.65	0.52	-2.51	< 0.05
B2:D-level above 4	0.53	0.34	0.4	1.49	1.55	0.12
B2:Density nominalizations*	-0.37	0.18	-0.76	0.47	-2.1	< 0.05

VUB VRIJE UNIVERSITEIT BRUSSEL UCLouvain

Prospects: Markers

- Test the markers we have (general English vs academic discourse)
- Extend the list (LEAD?)
- Classify according to functions?

Prospects: Learner Spoken Corpus

- Check the accuracy of the texts
- Collect more texts - record presentations at iTell 2020
- Transfer audio to text
- Transcribe them
- Transcription guidelines

Partnership with University of Vigo

- Интернет вещей - Хуан де ла Пенья (Juan de la Peña)
- Экономика и бизнес http://ecobas.webs.uvigo.es/index_en.php
- Центр исследований телекоммуникационных технологий http://atlanttic.uvigo.es/

Thank you for your attention!