

PLIN Linguistic Day 2021 - May 7, 2021



Exploring linguistic complexity in learner English applied to business

Sophie Belan

Univ Nantes - EA 1162 CRINI

Thomas Gaillat

Univ Rennes - EA 3874 LIDILE
CRINI

Julie McAllister

Univ Nantes - EA 1162



UNIVERSITÉ DE NANTES



Introduction: background

- Large-scale task-based blended language learning programme for first year business English undergraduate students at University of Nantes (Northwest France).
 - Pre- and post-tests in the form of a timed written production used to evaluate the impact of the programme on student learning outcomes (n=312).
 - Pre-test showed heterogeneous student proficiency levels (25% = B2/C1; 49% = B1).
 - A third of students progressed between pre-test and post-test (n=105).
- Analysis of productions (B1→B2 progression) showed statistically significant positive results for Accuracy and Fluency variables but not for Complexity (McAllister & Belan, 2014).
- We want to compare these learners with native speakers in terms of complexity.

Research Question

Is there a difference in linguistic complexity between native and learner English in the domain of business?

What does it mean for learners and course material designers ?

Method

Two corpora

BNC (Burnard, 2007) subset

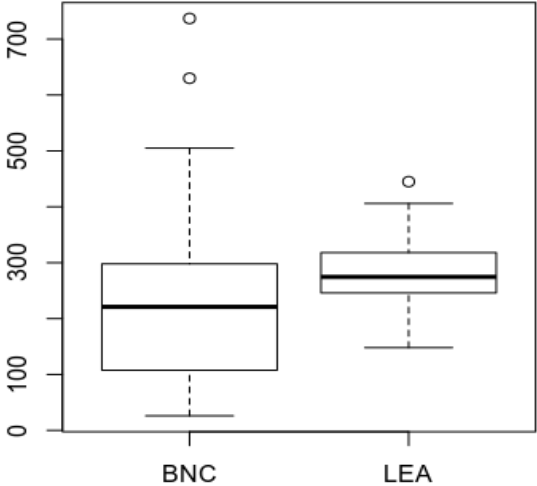
- Extraction: We exploit the classcode feature of Lee's taxonomy (Lee, 2001).
- This category includes two commerce-related types of newspaper articles : "W newsp brdsht nat: commerce" and "W newsp other: report"
- 15,692 documents
- Random sample of 50 texts.

LEA (Langues Étrangères Appliquées / Applied Foreign Languages) subset

- Random sample of 50 B2 level post-tests of First year Business English students participating in blended learning programme (L1) at Nantes University
- Extracted from corpus of 89 B2 post-tests / 312 total number of post-tests
- Task: writing a short newspaper article on the problem of counterfeiting / Time: 60 minutes/ 200-250 words
- 50 handwritten productions were transcribed manually for computer analysis

Data set sizes

Number of words



Metrics

Metrics computation

- VizLing a pipeline (Gaillat et al., 2021).
 - 86 textual metrics related to lexical diversity, syntactic complexity and readability.
 - POS and syntactic annotation with CoreNLP (Manning et al., 2014), L2SCA (Lu, 2010), a syntactic complexity analyser, and Quanteda (Benoit et al., 2018), a R package `textstat_lexdiv()` and `textstat_readability()`.
 - See metric description at <https://www.iris-database.org/iris/app/home/detail?id=york:939020>
- LancsLex: online vocabulary analysis tool - Lancaster University
 - Lexical sophistication of texts
 - Comparison with New General Service List (2,500 most frequent English words based on four language corpora) (Brezina, V. & Gablasova, D., 2015 & 2017) -
 - 3 metrics new-GSL 500, new-GSL 1000, new-GSL 2500
 - Description at <http://corpora.lancs.ac.uk/vocab/>

Data set

100 texts are processed.

2 outliers in BNC sample

Matrix of 98 observations x 89 variables (text metrics)

Evaluation

Pairwise comparisons of metrics across the two datasets.

1. Testing normality with Shapiro test
2. Testing difference with Welch t-test and Wilcox test depending on normality assumption.

Results

56 significant metrics ($p < 0.05$)

Significant metrics according to Welch test for normally distributed variables

- Readability: ARI.NRI Coleman.C2 LIW Dale.Chall Dale.Chall.old Dale.Chall.PSK Scrabble
- Lexical diversity: TTR
- Syntactic complexity C S

Significant metrics according to Wilcox test for non-normally distributed variables

- Readability: Danielson.Bryan Danielson.Bryan.2 NDWC ARI ARI.simple Bormuth.GP Bormuth.MC Dickes.Steiwer DRP ELF Farr.Jenkins.Paterson Flesch.Kincaid FOG.PSK Fucks RIX Spache Spache.old Strain Traenkle.Bailer Wheeler.Smith
- Lexical diversity: W lgeV0 lgV0 Maas U Yule's I (repetition)
- Lexical sophistication: GSL_500 GSL 500 & 1000, Total GSL
- Syntactic complexity: CP.T C.T CP CT DC T.S C.1 CN.C CN.T meanSentenceLength MLC MLS MLT S.1 T VP

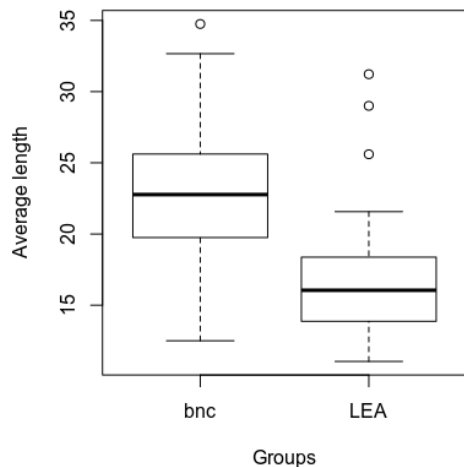
Full results available at <http://bit.ly/PLIN2021-results>

Results & visualisations of specific metrics

Recommendations for learners:

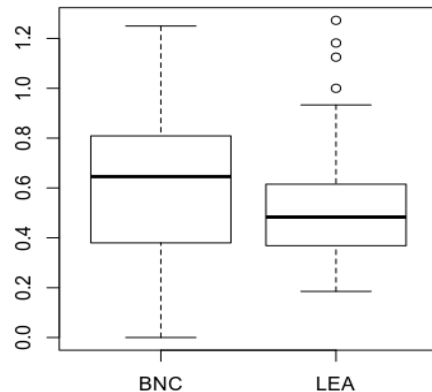
Write longer sentences

Boxplots of average length of sentences



More subordination
(watch for composition though)

Ratio Finite clauses per T-Unit

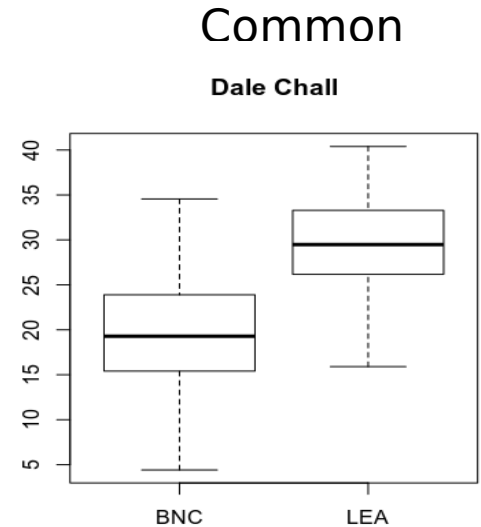
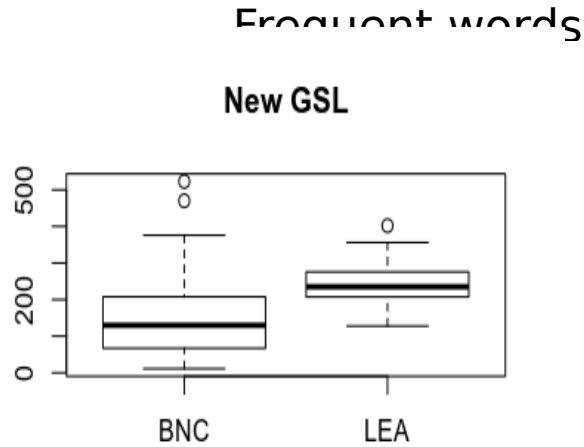
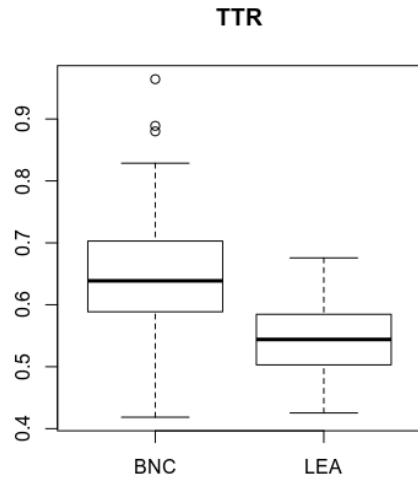


Results & visualisations of specific metrics

Recommendations for learners

Introduce more variety in vocabulary words

Use more uncommon



Discussion

What it means for Business English students

- Business English is not just about terminology
- Impact on course design (Granger, 2015)
- Developing automated diagnostic tools for teachers
- BNC sample: issues with “Finance commerce” classcode XML category (Lee, 2001).

Thank you for your attention

Julie, Sophie and Thomas :-)

References

Brezina, V. & Gablasova, D. (2015). Is there a core general vocabulary? Introducing the New General Service List. *Applied Linguistics* 36 (1): 1-22. <https://doi.org/10.1093/applin/amt018>

Brezina, V. & Gablasova, D. (2017). How to Produce Vocabulary Lists? Issues of Definition, Selection and Pedagogical Aims. A Response to Gabriele Stein. *Applied Linguistics*, 38 (5), 764–767. <https://doi.org/10.1093/applin/amx022>

Burnard, L. (2007). The British National Corpus, version 3 (BNC XML Edition). Oxford University Computing Services. <http://www.natcorp.ox.ac.uk/>

Gaillat, T., Knefati, A., & Lafontaine, A. (2021). Towards a Data Analytics Pipeline for the Visualisation of Complexity Metrics in L2 writings. *Proceedings of the 16th Workshop on Innovative Use of NLP for Building Educational Applications*, 123–129. <https://www.aclweb.org/anthology/2021.bea-1.13>

Granger, S. (2015). The contribution of learner corpora to reference and instructional materials design. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge Handbook of Learner Corpus Research* (Cambridge Handbooks in Language and Linguistics, pp. 485-510). Cambridge: Cambridge University Press.

Lee, D. (2001). Genres, registers, text types, domains and styles: Clarifying the concepts and navigating a path through the BNC jungle. *Faculty of Arts - Papers (Archive)*, 37-72.

[McAllister, J., & Belan, S. \(2014\). L'anglais de spécialité en LEA à la croisée des domaines: Étude de l'acquisition du lexique spécialisé. *ASp. la revue du GERAS*, 66, 41-59.](https://doi.org/10.4000/asp.4564)
<https://doi.org/10.4000/asp.4564>