



Bottom-up vs top-down: Refining the methodology for Compiling an Academic Phrasebank

Anni Jürine^{1,2} & Djuddah A.J. Leijen²

Estonian Military Academy
University of Tartu



THE EIGHTH INTERNATIONAL CONFERENCE ON WRITING ANALYTICS
Winterthur Switzerland,
5-6 September 2019,
<https://writinganalytics.zhaw.ch>

An Academic Phrasebank

- **WHAT?**

- Collection of phrases typically found in academic texts
- Organised by rhetorical moves (Morley n.d.)

- **WHY?**

- Appropriate vocabulary (Chitez et al. 2015)
- Awareness of the functions
- Helps students starting to write (Rapp & Kauff 2018)

- **FOR WHOM?**

- Students and researchers
- L1 and L2 writers (Davies & Morley, 2018)

Academic Phrasebank for Writing in Estonian

- 750 academic expressions in Estonian
- Organised by functions
- Organised by sections of papers
 - Introduction
 - Main Body
 - Conclusion
 - General functions
- Not discipline-specific
- An online-resource: www.teadustekst.ut.ee

Aim of the study

- Validate the methodology for compiling an academic phrasebank
- By comparing the functions (and expressions)
 - Bottom-up approach
 - Top-down approach
 - *Introduction* section

1. Bottom-up: Data

Bottom-up

- Defended Ma/MSc and BA/BSc theses
- Repository of University of Tartu
- Selected semi-randomly from each of the four faculties:
 - Arts and Humanities
 - Social Sciences
 - Medical Sciences
 - Science and Technology
- 200 theses ~ 2.2 mil words

Top-down

1. Bottom-up: Procedure

Bottom-up

- **Starting from expressions**
- Finding the key phrases
 - Most common relevant N-grams
 - 3- and 4-grams (n=382)
- Searches in the corpus
 - 10,417 segments
- Sifting the expressions (Morley n.d.)
- Associating expressions with functions

Top-down

1. Bottom-up: Results

Bottom-up

- Naming the topic
- Underlining the importance of the topic
- Establishing the niche
- Stating the aim
- Posing research questions
- Setting up a hypothesis
- Indicating the structure of a thesis

Top-down

2. Top-down: Data

Bottom-up

- Defended Ma/MSc and BA/BSc theses
- Repository of University of Tartu
- Selected semi-randomly from each of the four faculties:
 - Arts and Humanities
 - Social Sciences
 - Medical Sciences
 - Science and Technology
- **200 theses ~ 2.2 mil words**

Top-down

- Defended Ma/MSc and BA/BSc theses
- Repository of University of Tartu
- Selected semi-randomly from each of the four faculties:
 - Arts and Humanities
 - Social Sciences
 - Medical Sciences
 - Science and Technology
- **28 theses ~ 17,500 words**

2. Top-down: Procedure

Bottom-up

- **Starting from expressions**
- Finding the key phrases
 - Most common relevant N-grams
 - 3- and 4-grams (n=382)
- Searches in the corpus
 - 10,417 segments
- Sifting the expressions (Morley n.d.)
- Associating expressions with functions

Top-down

- **Starting from functions**
- Segmentation into T-units (n = 1,036)
- Coding for functions (Geisler, 2018)
- Inter-rater reliability substantial (15% data, Kappa = 0.61, p < .001), 95% CI (0.525, 0.696)
- Refining the coding scheme
- Sifting expressions

2. Top-down: Results

Bottom-up & Top-down

- **Giving background information**
- Naming the topic
- Underlining the importance of the topic
- Establishing the niche
- Stating the aim
- Posing research questions
- Setting up a hypothesis
- Indicating the structure of a thesis
- **Main theories/concepts/Terms**
- **Outlining Methods & Data**
- **Presenting limitations**
- **Stating the value of the research**

Discussion

- The two approaches share a bulk of the functions
- The top-down approach identified some additional functions
 - Do not associate with particular linguistic patterns
 - *Giving Background Information*
 - *Introducing Main Theories/Concepts/Terms*
 - Are infrequent
 - *Presenting Limitations*
 - *Stating the Value of the Research*
 - Are primarily associated with another section
 - *Outlining Methods & Data*

Conclusion

- Bottom-up approach
 - Detecting expressions
 - Large datasets
- Top-down approach
 - Determining the rhetorical structure
 - Smaller datasets
- **Top-down method serves as valuable addition**

Future research

- Comparison with Phrasebanks in other languages
 - English
 - Dutch
 - Danish
 - Romanian
 - ...
- Technical applications
 - Usability?
 - Technical tool?

Selected references

Chitez, Madlina, Christian Rapp, Otto Kruse 2015. Corpus-supported academic writing: how can technology help? – Critical CALL – Proceedings of the 2015 EUROCALL Conference, Padova, Italy. Dublin: Researchpublishing.net, 125–132.

Jürine, A. (2019a). Akadeemiliste Väljendite Varamu [Academic Phrasebank for Writing in Estonian]; www.teadustekst.ut.ee.

Jürine, A. (2019b). Akadeemiliste väljendite varamu – keeleressurss, mis on abiks kirjutama õppimisel ja õpetamisel [Academic Phrase Bank for Writing in Estonian: a Language Resource to aid in the Learning and Teaching of Writing]. Emakeele Seltsi Aastaraamat, Volume 64, pp 279–289.

Morley, J. (n.d). Phrasebank: a University-wide Online Writing Resource; <https://tinyurl.com/lgbsxd6>.

Rapp, Christian, Peter Kauf 2018. Scaling academic writing instruction: evaluation of a scaffolding tool (Thesis Writer). — International Journal of Artificial Intelligence in Education 28 (4), 590—615.



Academic Phrasebank for Writing in Estonian

Anni Jürine, Djuddah A.J. Leijen



Thank you!

Contact:

anni.jurine@mil.ee

