

# Corpus Linguistics and Data-Driven Learning for ESP: The analysis of “Methods” sub-genre



Sébastien Louvigné<sup>(\*)</sup>

Jie Shi<sup>(\*\*)</sup>

<sup>(\*)</sup> Graduate School of Information Systems

<sup>(\*\*)</sup> Graduate School of Informatics and Engineering

The National University of Electro-Communications, Tokyo.

## Abstract

RA is a genre comprised of the typical IMRAD structure, i.e. Introduction, Methods, Results and Discussion. Hence, genre-based analysis of the RA is generally considered to be challenging for EFL users due to its complexity in nature, though such analysis is essential for teaching, learning and applications. Without sufficient genre and linguistic knowledge of the RA, EST students and researchers tend to face obstacles at various levels when employing the correct structure or organization and appropriate linguistic expressions for clear communication of their research. As part of a longitudinal study on the RA genre, this study focuses on the following three aspects: to determine the linguistic features of the common moves in the “Methods” section of RA using corpus analysis, to compare and analyze the linguistic features of the “Methods” with those of the “Abstract” section, the “Introduction” section and of the whole RA from the same dataset.

## Research Article (RA) genre

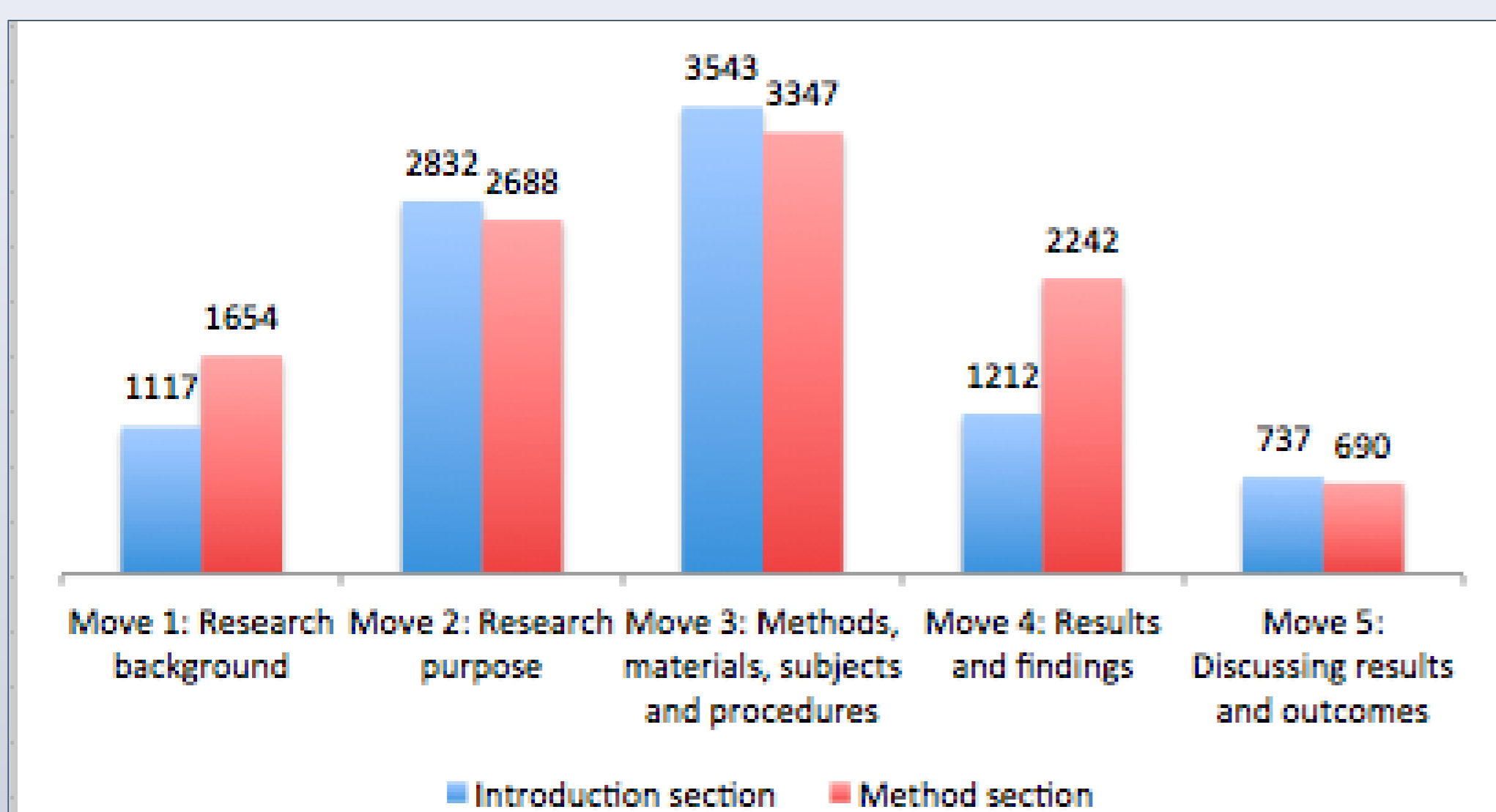
- ◆ Important Academic Genre
  - ◆ Final outcome of research
  - ◆ Demonstrates needs, contributions, justification
- ◆ Challenging genre for communicating ideas  
(Swales, 1990; Kanoksilapatham, 2007)
  - ◆ Different sub-genres (e.g. Intro, Methods, Results)
  - ◆ Different functional units: words, linguistic expressions for communicative purposes
- Lack of knowledge of RA → Difficulty to clearly communicate research ideas
  - ◆ English Foreign Language (EFL)
  - ◆ Correctly employ & organize linguistic expressions

## Research Purpose

- ◆ Corpus Linguistics for:
  1. Linguistic features of common moves in “Methods” section of RAs (Cheng, 2011),
  2. Compare & Analyze linguistic features with “Introduction” section and whole RAs from the same dataset.
- ◆ 3rd part of longitudinal study  
(Louvigné et al., 2013; Louvigné et al., 2014)
- ◆ Dataset: 543 research articles  
(*Journal of Optics, 1997-2001*)
- ◆ Concordancer: AntConc Corpus Linguistics Tool  
(Anthony, 2011)

## Rhetorical Moves

- ◆ Similarities with usage in Introduction section
- ◆ Stronger usage of “Results” rhetorical move in “Method” section



## Lexicon Frequency & N-gram

	Methods		Introduction		Research article	
	Freq.	Word	Freq.	Word	Freq.	Word
1	227	shown in fig	230	in this paper	785	due to the
2	208	due to the	169	this paper we	709	a function of
3	189	a function of	126	the use of	644	as a function
4	166	as a function	99	due to the	636	in order to
5	157	in order to	94	in order to	493	the number of
6	153	the number of	83	as well as	469	the case of
7	141	the case of	78	based on the	430	in this paper
8	132	function of the	69	a number of	401	with respect to
9	128	is given by	69	one of the	385	as well as
10	110	with respect to	66	properties of the	785	due to the
11	108	part of the	62	can be used	379	function of the
12	96	in fig a	57	of the laser	350	is shown in
13	95	in the case	56	the case of	349	the use of
14	94	in this case	53	the number of	346	in this case
15	94	is shown in	52	of the optical	345	in fig the
16	93	in fig the	48	because of the	328	of the laser
17	90	as shown in	46	in section we	326	is given by
18	87	as well as	46	the presence of	324	in the case
19	86	in terms of	46	with respect to	323	in fig a
20	84	note that the	45	be used to	300	as shown in

Methods		Research article	
Freq.	Word	Freq.	Word
66	as a function of the	237	as a function of the
23	as a function of time	61	this work was supported by
20	intracavity field distribution in an	60	in the case of the
20	three-dimensional intracavity field distribution in	57	as a function of time
19	the Fourier transform of the	49	it is well known that

## Discussion & Conclusion

- 1) This study of the academic Method sub-genre analyzed linguistic features and rhetorical moves over 543 RAs.
- 2) “Results” Rhetorical move largely used in “Method” section.
- 3) Most frequently used N-grams: Similarity between Method section and whole RAs.
- 4) Frequent use of 3-gram describing methods (e.g. figures, equations).
- 5) Future works: extend to other sections, research areas

## References

- Anthony, L. (2011). AntConc 3.2.4 (Computer software). Waseda University, Tokyo, Japan: <http://www.antlab.sci.waseda.ac.jp/>
- Cheng, W. (2011). *Exploring Corpus Linguistics, Language in action*. Routledge.
- Kanoksilapatham, B. (2007). Rhetorical moves in biochemistry research articles. In *Discourse on the Move: Using Corpus analysis to describe discourse structure* (pp. 73-120). Amsterdam: John Benjamins.
- Louvigné, S., Shi, J., & Sharmin, S. (2014). A Corpus-based Analysis of the Scientific RA Genre and RA Introduction. In ICAMechS International Conference on Advanced Mechatronic Systems Proceedings (pp. 123-127), Kumamoto, Japan, August 10-12, 2014.
- Swales, J. M. (1990). *Genre analysis: English in academic research settings*. Cambridge University Press.