Conformity and implementation of the CARS model in science writing: a data-driven approach

Charles Lam

C.Lam@leeds.ac.uk

Language Centre, University of Leeds

November 16, 2023

Goal of the study

How much do published writers follow the Create-A-Research-Space model? (Swales, 1990, 2004)

$\rightarrow$ How can instructors better guide students in their writing?

2

Related Works

Charles Lam
C.Lam@leeds.ac.uk

1. "Create-A-Research-Space" (CARS) model (Swales, 1990, 2004) has been the *de facto* gold standard of the organisation for introduction in science writing.

Related Works

1. "Create-A-Research-Space" (CARS) model (Swales, 1990, 2004) has been the *de facto* gold standard of the organisation for introduction in science writing.
2. Data-driven Learning (DDL) (Boulton, 2010; Mizumoto et al., 2017; Lu et al., 2021)
   - "Soft DDL": instructors to be informed by DDL, students do not necessarily need to learn about the data-driven language studies (Gabrielatos, 2005)

Charles Lam
C.Lam@leeds.ac.uk

Related Works

Conformity and implementation of the CARS model in science writing: a data-driven approach

Charles Lam
C.Lam@leeds.ac.uk

1. "Create-A-Research-Space" (CARS) model (Swales, 1990, 2004) has been the *de facto* gold standard of the organisation for introduction in science writing.

2. Data-driven Learning (DDL) (Boulton, 2010; Mizumoto et al., 2017; Lu et al., 2021)
   - ▶ "Soft DDL": instructors to be informed by DDL, students do not necessarily need to learn about the data-driven language studies (Gabrielatos, 2005)

3. Discipline-specific studies (just a few examples)
   - ▶ Computer science (Orr, 1999; Maher and Milligan, 2019)
   - ▶ Mathematics (McGrath and Kuteeva, 2012; Kuteeva and McGrath, 2015)
   - ▶ Biology (Samraj (2005) on conservation sciences; Humphrey and Hao (2013) on UG students)

Related Works

1. "Create-A-Research-Space" (CARS) model (Swales, 1990, 2004) has been the *de facto* gold standard of the organisation for introduction in science writing.

2. Data-driven Learning (DDL) (Boulton, 2010; Mizumoto et al., 2017; Lu et al., 2021)
   - "Soft DDL": instructors to be informed by DDL, students do not necessarily need to learn about the data-driven language studies (Gabrielatos, 2005)

3. Discipline-specific studies (just a few examples)
   - Computer science (Orr, 1999; Maher and Milligan, 2019)
   - Mathematics (McGrath and Kuteeva, 2012; Kuteeva and McGrath, 2015)
   - Biology (Samraj (2005) on conservation sciences; Humphrey and Hao (2013) on UG students

This study:

A non-prescriptive account for the organisation of the Introduction section in biological sciences.

Ongoing Work: Data & Method

Conformity and
implementation of
the CARS model in
science writing: a
data-driven
approach

Charles Lam
C.Lam@leeds.ac.
uk

Obtain papers from PubMed / bioArxiv (n=100)

$\downarrow$

Extract Introduction Sections

Figure 1: Procedures of data collection and annotation

Ongoing Work: Data & Method

Conformity and
implementation of
the CARS model in
science writing: a
data-driven
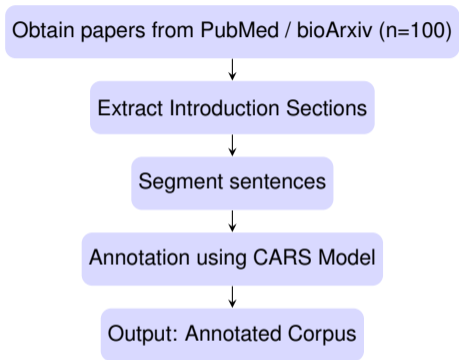approach

Charles Lam
C.Lam@leeds.ac.
uk

Obtain papers from PubMed / bioArxiv (n=100)

Extract Introduction Sections

Segment sentences

Annotation using CARS Model

Output: Annotated Corpus

Figure 1: Procedures of data collection and annotation

4

Expected Outcome & Implications

- ▶ Output: Annotated Corpus (n=100)
- ▶ Potential for training & fine-tuning existing large language models with few-shot learning (Bansal et al., 2019; Wang et al., 2020; Brown et al., 2020)
- ▶ Practical use: GUI for writing instructors; self-assessment in writing

  Cf. AntMover (Anthony, 2003, 2016) which comes only with a small training set from computer science; AWSUM (Mizumoto et al., 2017)
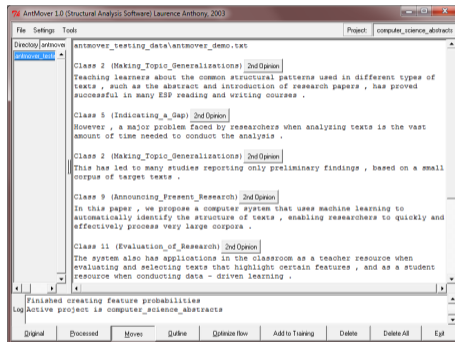

Figure 2: Screenshot of AntMover 1.1.0

Conformity and
implementation of
the CARS model in
science writing: a
data-driven
approach

Charles Lam
C.Lam@leeds.ac.
uk

Thank you!

References

Anthony, L. (2003). *AntMover (Version 1.1.0) [Computer Software]*.

Anthony, L. (2016). Introducing corpora and corpus tools into the technical writing classroom through data-driven learning (DDL). In *Discipline specific writing*, pages 162–180. Routledge.

Bansal, T., Jha, R., and McCallum, A. (2019). Learning to few-shot learn across diverse natural language classification tasks. *arXiv preprint arXiv:1911.03863*.

Boulton, A. (2010). Data-driven learning: on paper, in practice.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.

Gabrielatos, C. (2005). Corpora and language teaching: Just a fling or wedding bells? *TESL-EJ*, 8(4):1–37.

Humphrey, S. and Hao, J. (2013). Deconstructing written genres in undergraduate biology. *Linguistics & the Human Sciences*, 7.

Kuteeva, M. and McGrath, L. (2015). The theoretical research article as a reflection of disciplinary practices: The case of pure mathematics. *Applied Linguistics*, 36(2):215–235.

Lu, X., Yoon, J., Kisselev, O., Casal, J. E., Liu, Y., Deng, J., and Nie, R. (2021). Rhetorical and phraseological features of research article introductions: Variation among five social science disciplines. *System*, 100:102543.

Maher, P. and Milligan, S. (2019). Teaching master thesis writing to engineers: Insights from corpus and genre analysis of introductions. *English for specific purposes*, 55:40–55.

McGrath, L. and Kuteeva, M. (2012). Stance and engagement in pure mathematics research articles: Linking discourse features to disciplinary practices. *English for Specific Purposes*, 31(3):161–173.

Mizumoto, A., Hamatani, S., and Imao, Y. (2017). Applying the bundle–move connection approach to the development of an online writing support tool for research articles. *Language Learning*, 67(4):885–921.

Orr, T. (1999). Genre in the field of computer science and computer engineering. *IEEE Transactions on Professional Communication*, 42(1):32–37.

Samraj, B. (2005). An exploration of a genre set: Research article abstracts and introductions in two disciplines. *English for specific purposes*, 24(2):141–156.

Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge University press.

Swales, J. M. (2004). *Research genres: Explorations and applications*. Cambridge University Press.

Wang, Y., Yao, Q., Kwok, J. T., and Ni, L. M. (2020). Generalizing from a few examples: A survey on few-shot learning. *ACM computing surveys (CSUR)*, 53(3):1–34.