Titles That Announce Argumentative Claims in Biomedical Research Articles

Heather Graves
Roger Graves
Department of English and Film Studies
The University of Alberta
Edmonton, Alberta, Canada
`graves1,hgraves@ualberta.ca`

Robert E. Mercer
Mahzereen Akter
Department of Computer Science
The University of Western Ontario
London, Ontario, Canada
`mercer@csd.uwo.ca`

Abstract

In the experimental sciences authors use the scientific article to express their findings by making an argumentative claim. While past studies have located the claim in the Abstract, the Introduction, and in the Discussion section, in this paper we focus on the article title as a potential source of the claim. Our investigation has suggested that titles which contain a tensed verb almost certainly announce the argument claim while titles which do not contain a tensed verb have varied announcements. Another observation that we have confirmed in our dataset is that the frequency of verbs in titles of experimental research articles has increased over time.

1 Introduction

In this paper we are interested in determining what is being claimed in articles in experimental (not clinical) biomedical literature, in particular. Claims have been studied in the argumentation literature from many different standpoints (White, 2009). Rhetorical structure theory was developed from systemic functional linguistics to map connections among texts (Mann and Thompson, 1987); Argumentative zoning was developed from Swales' CARS model of moves made in research articles (Teufel, 1999; Teufel and Moens, 1999; Teufel and Moens, 2002). Toulmin-based analysis has also been used to map the argumentative structure of articles (Toulmin, 1958 2003; Jenicek, 2006; Reed and Rowe, 2006; Graves et al., 2013; Graves, 2013). With these models of argument in mind, we view the claim of a scientific argument as the conclusion that the authors infer from known information and new information (results from an experiment or other forms of observations). Past studies locate the claim in the Abstract (Kanoksilapatham, 2013), at the end of the Introduction (Swales, 1990; Swales and Najjar, 1987; Kanoksilapatham, 2005; Kanoksilapatham, 2012), and in the Discussion section (Kanoksilapatham, 2005). Our observations of changes in the genre of the research article have led us to perform a preliminary investigation of titles with the outcome being a provisional typology.

2 Method

The Genia Tagger uses the Penn Treebank Tagset. In the following we mention the verb tags from this tagset: VB – base form, VBD – past tense, VBG – gerund, VBN – past participle, VBP – present tense non-3rd person singular, VBZ – present tense 3rd person singular. We applied these tags to the dataset of biomedical article titles and abstracts used in this preliminary study has been taken from MEDLINE, the well-known biomedical bibliographic repository that contains over 19 million citations and abstracts for about 81% of these citations from approximately 5600 journals (NLM, 2013 accessed 3 February 2014). We have curated a small database using *biotextEngine* and some locally developed tools.

3 Analysis

For each title we collect the following:

- cumulative frequency of all verb categories

- whether the title contains a VBP, VBZ, or passive verb

- whether the title contains a nominalization

4 Findings

Our analysis so far has identified three typologies. The articles can be categorized according to genre, purpose and structure. For titles with verbs the claim of the title is repeated several times: in the
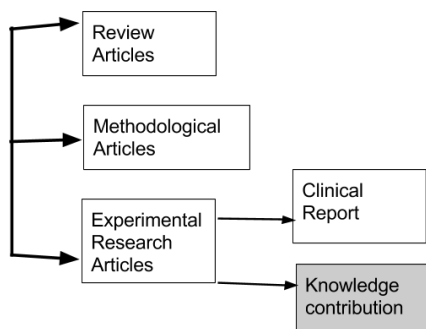
Figure 1: Genre typology

Abstract, Introduction, and Discussion sections. For articles without verbs, the claim does not appear in the title or introduction (it does appear in the abstract and discussion sections). A third finding: the frequency of verbs in titles of experimental research articles has increased over time.

5 Discussion

We believe that our methods for identifying titles could lead to better literature search techniques. If researchers are able to identify the claim of an article from a search of titles alone, they will be able to evaluate the relevance of each article more efficiently. We suspect that the increase in titles with verbs and claims in them is an emerging trend, possibly the result of explicit editorial policy. One side effect of including claims in titles may be higher quality writing by the authors. Another result from using verbs in titles could be the automation of claim extraction. Finally, having research scientists use clear language to state their claim can have the added benefit of making knowledge translation more effective by lessening the difficulty of reading scientific texts. This, in turn, might afford greater access to the research outcomes by clinical practitioners (one of the main readerships of biomedical research).

References

Heather Graves, Shahin Moghaddasi, and Azirah Hashim. 2013. Mathematics is the method: Exploring the macro-organizational structure of research articles in mathematics. *Discourse Studies*, 15:421–438.

Heather Graves. 2013. The trouble with Toulmin for teaching argument in science. In *11th Annual Technology for Second Language Learning Conference: Technology and Teaching Writing for Academic Disciplines*. ms.

Milos Jenicek. 2006. How to read, understand, and write discussion sections in medical articles: An exercise in critical thinking. *Med. Sci. Monitor*, 12.

Budsaba Kanoksilapatham. 2005. Rhetorical structure of biochemistry research articles. *English for Specific Purposes*, 24:269–292.

Budsaba Kanoksilapatham. 2012. Structure of research article introductions in three engineering sub-disciplines. *IEEE Transactions on Professional Communication*, 55:294–309.

Budsaba Kanoksilapatham. 2013. Generic characterisation of civil engineering research article abstracts. *3L: The Southeast Asian Journal of English Language Studies*, 19:1–10.

William C. Mann and Sandra A. Thompson. 1987. Rhetorical structure theory: Description and construction of text structures. In G. Kempen, editor, *Natural language generation: New results in artificial intelligence, psychology and linguistics*, pages 85–95. Dordrecht: Martinus Nijhoff.

U.S. National Library of Medicine NLM. 2013 (accessed 3 February 2014). "ncbi: Pubmed overview. http://www.ncbi.nlm.nih.gov/entrez/query/static/overview.html.

Chris Reed and Glenn Rowe. 2006. Translating Toulmin diagrams: Theory neutrality in argument representation. In David Hitchcock and Bart Verheij, editors, *Arguing on the Toulmin model: New essays in argument analysis and evaluation*, pages 341–358. Dordrecht: Springer.

John Swales and Hazem Najjar. 1987. The writing of research article introductions. *Written Communication*, 4:175–192.

John Swales. 1990. *Genre Analysis: English in Academic and Research Settings*. Cambridge Applied Linguistics. Cambridge University Press.

Simone Teufel and Mark Moens. 1999. Argumentative classification of extracted sentences as a first step towards flexible abstracting. In Inderjeet Mani and Mark Maybury, editors, *Advances in automatic text summarization*, pages 155–171. MIT Press.

Simone Teufel and Marc Moens. 2002. Summarizing scientific articles: Experiments with relevance and rhetorical status. *Computational Linguistics*, 28(4):409–445.

Simone Teufel. 1999. *Argumentative Zoning : Information Extraction from Scientific Text*. Ph.D. thesis, School of Cognitive Science, University of Edinburgh.

Stephen Toulmin. 1958-2003. *The uses of argument*. Cambridge University Press.

Barbara White. 2009. *Annotating a corpus of biomedical research texts: Two models of rhetorical analysis*. Ph.D. thesis, The University of Western Ontario, Canada.