
Preface: How Should Scientific Phrases Be Learnt?

Kenichi Iwatsuki

International Kimwipe Table Tennis Association

ABSTRACT

KTTA released a tool for assisting scientists in writing a paper.

Introduction

KTTA has always been on the side of scientists. It has been our priority to support scientists by any means so that they can save time for enjoying scientific sports.

One of the most time-consuming but important activities for scientists is writing a paper. Publication brings joy to us because our scientific findings will be delivered to certain people and make the world slightly better or because our list of achievements will become as long as is necessary to find another job. Composition, however, makes us exhausted not only from writing as such but also from imagining very fluent sentences being almost automatically flowed out of native English-speaking scientists, which is usually not real because academic writing is a genre with which they are unfamiliar before they dive into academia.

To mitigate the difficulty, many contributions have been made. Some people ask others to write a paper instead of themselves, which is problematic because it requires reading the whole paper and modifying it if necessary. Translation is also problematic for the same reason; the resulting text must be reviewed carefully by authors. Moreover, a manuscript must be written in some language in the first place, whose labour is not as little as can be ignored.

Therefore, learning academic writing is the easiest way. It consists of logical (or rhetorical) structure and English for academic purposes. A famous structure is IMRaD, standing for Introduction, Methods, Results, and Discussion, to which many papers of experimental sciences conform except journals terribly emphasising impact of published papers and place results in their first sections. Academic English is different from other genres, and wording is very peculiar to it. We do not say "in this paper" in our daily life, and if we do, it must be a parody of scholarly papers (some might say it is true of this preface).

In this preface, we present Scientific Phrases, a web-based application for assisting retrieval of phrasal expressions for academic writing. It searches for phrases that are different from an input but have the same rhetorical function. For instance, if the input is "has not been investigated", one of the output will be "little attention has been paid to", both of which play a role of "showing the lack of past work".

Those phrases are discipline-specific; thus, we collected phrases from four different disciplines: biomedicine, chemistry, computer science, and psychology. The number of disciplines addressed will be increased in the future.

Methods

Data

We used a dataset (Iwatsuki & Aizawa, 2021) that contained tens of thousands of phrases that are classified based on their rhetorical function. It was also divided into four parts based on disciplines.

Although it contains example sentences, they were extracted only from one journal for a discipline. Thus, we collected some other sentences from several open-access journals.

Retrieval

In the first place, a rhetorical function of an input phrase should be recognised. It is not realistic to apply neural models to the recognition because of its computational cost, the most similar phrase is searched for with the input as a query, and its rhetorical function is regarded as the one of the input.

Similar phrases to the input are not helpful because users easily come up with them. Thus, the presented system suggests dissimilar phrases from the input. The dissimilarity is calculated with the Jaccard index (Jaccard, 1912). The thresholds of the dissimilarity are to be defined respectively because diversity of phrases is different among rhetorical functions.

Results

The results are shown in Figure 1. The URI is <https://phrases.ktta.net/>.

KTTA Scientific Phrases

Biomedicine ▼ input a phrase Search

input a phrase

Example sentences © Authors of each paper or publishers.
©2022 KTTA. [Terms of Use & Privacy Policy](#) (利用規約・個人情報)

Figure 1: The results

Discussion

To promote something data-driven for writing or reading scientific papers, access to well-written, high-quality full texts must be assured. The full texts should be provided in some computer-friendly formats such as XML. Mathematical formulae should also be annotated with

reasonable meta data. For example, it is not clear to computers that ab means that a is multiplied by b without it being tagged with '&invisibletimes;' or similar markups.

Reference

1. Iwatsuki, K., & Aizawa, A. (2021). Communicative-Function-Based Sentence Classification for Construction of an Academic Formulaic Expression Database. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 3476–3497). <https://doi.org/10.18653/v1/2021.eacl-main.304>
2. Jaccard, P. (1912). The Distribution of the Flora in the Alpine Zone. *The New Phytologist*, 11(2), 37–50. <https://doi.org/10.1111/j.1469-8137.1912.tb05611.x>

Iwatsuki, K. (2022). Preface: how should Scientific Phrases be Learnt?. *Scientific Sports*, 7(1), 1–3.

© 2022 Kenichi Iwatsuki. Licensed CC BY 4.0. Published by the International Kimwipe Table Tennis Association. <https://journal.ikta.org/articles/7/1/7-1-1-abstract.html>