

Attitude in Academic Writing (CAAW): Corpus Compilation and Annotation

Authors : Hortènsia Curell, Ana Fernández-Montraveta

Abstract : This paper presents the creation, development, and analysis of a corpus designed to study the presence of attitude markers and author's stance in research articles in two different areas of linguistics (theoretical linguistics and sociolinguistics). These two disciplines are expected to behave differently in this respect, given the disparity in their discursive conventions. Attitude markers in this work are understood as the linguistic elements (adjectives, nouns and verbs) used to convey the writer's stance towards the content presented in the article, and are crucial in understanding writer-reader interaction and the writer's position. These attitude markers are divided into three broad classes: assessment, significance, and emotion. In addition to them, we also consider first-person singular and plural pronouns and possessives, modal verbs, and passive constructions, which are other linguistic elements expressing the author's stance. The corpus, Corpus of Attitude in Academic Writing (CAAW), comprises a collection of 21 articles, collected from six journals indexed in JCR. These articles were originally written in English by a single native-speaker author from the UK or USA and were published between 2022 and 2023. The total number of words in the corpus is approximately 222,400, with 106,422 from theoretical linguistics (*Lingua*, *Linguistic Inquiry* and *Journal of Linguistics*) and 116,022 from sociolinguistics journals (*International Journal of the Sociology of Language*, *Language in Society* and *Journal of Sociolinguistics*). Together with the corpus, we present the tool created for the creation and storage of the corpus, along with a tool for automatic annotation. The steps followed in the compilation of the corpus are as follows. First, the articles were selected according to the parameters explained above. Second, they were downloaded and converted to txt format. Finally, examples, direct quotes, section titles and references were eliminated, since they do not involve the author's stance. The resulting texts were the input for the annotation of the linguistic features related to stance. As for the annotation, two articles (one from each subdiscipline) were annotated manually by the two researchers. An existing list was used as a baseline, and other attitude markers were identified, together with the other elements mentioned above. Once a consensus was reached, the rest of articles were annotated automatically using the tool created for this purpose. The annotated corpus will serve as a resource for scholars working in discourse analysis (both in linguistics and communication) and related fields, since it offers new insights into the expression of attitude. The tools created for the compilation and annotation of the corpus will be useful to study author's attitude and stance in articles from any academic discipline: new data can be uploaded and the list of markers can be enlarged. Finally, the tool can be expanded to other languages, which will allow cross-linguistic studies of author's stance.

Keywords : academic writing, attitude, corpus, english

Conference Title : ICACL 2025 : International Conference on Applied Corpus Linguistics

Conference Location : Brussels, Belgium

Conference Dates : March 29-30, 2025