



# Chapter 1

## Building an IEEE Journal Corpus



## 1.1 Lead-in

Corpus can be used as a computer-based tool for description of language, and enables analysis of a certain scope of language (Bieber, 1993). For example, a subject-specific corpus would help analyze the styles of language use in a particular subject.

For this textbook, a corpus of about 550,000 words gathered from 87 articles published during 2016—2021 in 11 IEEE (Institute of Electrical and Electronic Engineers) journals is built for the analysis of rhetorical moves and steps within each section of a research article as well as the specific language styles used accordingly. These articles were selected from journals on different topics. It should be admitted that the average selection of 8 articles per journal may not be able to fully present the variations of rhetorical moves and language styles. However, the corpus is well managed in terms of representativeness through the following criteria:

(1) The journals are with impact factors above 5, i.e. *Internet of Things*, *Transactions on Automatic Control*, *Transactions on Automation Science and Engineering*, *Transactions on Communications*, *Transactions on Geoscience and Remote Sensing*, *Transactions on Circuits and Systems for Video Technology*, *Transactions on Mobile Computing*, *Transactions on Multimedia*, *Transactions on Pattern Analysis and Machine Intelligence*, *Transactions on Vehicular Technology*, *Computational Intelligence Magazine*, and they cover a wide spectrum of topics within IEEE field;

(2) The articles are either highly cited or newly published (within three years);

(3) The articles contain four basic sections: Introduction, Methods, Results and Discussion;

(4) The articles are written by native English speakers;

(5) The articles are selected based on academic rigor, high-quality writing and clear presentation.

It is notable that the corpus is open to new articles from different journals.

## 1.2 Coding

After data collection, we can convert the articles into plain text and then proceed to data coding. Coding consists of two phases: One is manual annotation of moves and steps, and the other is automatic semantic tagging of words with a parser.

### 1.2.1 Move Coding

Based on the functional segments of a text, different Moves and Steps in the four main

sections are identified, as shown in the following table:

**Table 1.1** Moves and Steps in a research article

Sections	Moves	Steps	
Introduction	Establishing a research territory	Claiming the centrality	
		Making topic generalization(s)	
		Reviewing items of previous research	
	Establishing the niche	Establishing the niche	Counter-claiming
			Indicating a research gap
			Question-raising
			Continuing a tradition
	Occupying the niche	Occupying the niche	The purpose of the current research
			The main features of the current research
Stating the outline of the article			
Methods	Contextualizing the study methods	Stating the purpose of the methods	
		Referencing previous works	
		Identifying the methodological approaches	
	Describing the study	Describing the study	Rationalizing the methods
			Describing the data
			Describing experimental/study procedures
Establishing credibility	Establishing credibility		
Results	Contextualizing results	Providing background information	
		Listing procedures or methodological techniques	
		Stating research questions or hypothesis	
		Describing aims and purposes	
		Referring to previous research	
		Summarizing main results	
	Reporting results	Reporting results	Showing results by figures or tables
	Summarizing results	Summarizing results	Instantiating results
Invalidating results			