World Journal of Corpus Linguistics

Vol. 1, No. 2 – 2025, Page. 38-47| Available online at
https://pusatpublikasi.com/index.php/corpus/index

THE USE OF CORPUS LINGUISTICS IN PREPARING ENGLISH TEACHING MATERIALS FOR SPECIFIC PURPOSES IN THE FIELD OF METEOROLOGY

Ni Made Dwi Utari Pusparini
Mahasiswa Magister Ilmu Linguistik, Program Pascasarjana, Universitas Warmadewa
dwipusparini63@gmail.com

Abstract
English for Specific Purpose is an English learning approach that focuses specifically on the use of English in certain scientific and professional fields. This research discusses preparing English teaching materials using the features in AntConc software which is one of the linguistic corpus software. The aim research is to know how to use of Corpus Linguistics in Preparing English Teaching Materials for Specific Purposes in the Field of Meteorology. The method used in this study is a combined method by applying quantitative methods to obtain research data and qualitative methods in the analysis process to obtain conclusions. The results showed that there were different words used 1269 in the total of words including repeated 5359 words. These words are classified into classes of nouns, verbs, adjectives, adverbs, abbreviations, etc.

Keyord : Teaching materials, Corpus Linguistics, ESP

Abstrak
English for Specific Purpose adalah pendekatan pembelajaran bahasa Inggris yang memfokuskan secara khusus pada penggunaan bahasa Inggris dalam bidang keilmuan dan profesional tertentu. Penelitian ini membahas tentang penyusunan bahan ajar bahasa Inggris dengan menggunakan fitur-fitur pada software AntConc yang merupakan salah satu software korpus linguistik. Tujuan penelitian adalah untuk mengetahui bagaimana penggunaan Corpus Linguistics dalam Mempersiapkan Bahan Ajar Bahasa Inggris untuk Keperluan Khusus di Bidang Meteorologi. Metode yang digunakan dalam penelitian ini adalah metode gabungan dengan menerapkan metode kuantitatif untuk memperoleh data penelitian dan metode kualitatif dalam proses analisis untuk memperoleh kesimpulan. Hasil penelitian menunjukkan terdapat perbedaan kata yang digunakan sebanyak 1269 kata termasuk kata yang diulang sebanyak 5359 kata. Kata-kata ini diklasifikasikan ke dalam kelas kata benda, kata kerja, kata sifat, kata keterangan, singkatan, dll.

***Kata Kunci** : Bahan ajar, Corpus Linguistics, ESP*

I. INTRODUCTION

In this era of globalization, foreign language skills, especially English, an international language, are highly prioritized. The spread of use of English is now increasing and proliferating. As one of the international languages, English has many vital roles in scientific learning. At the tertiary level, English which is learned by students in tertiary institutions, especially in vocational education, should be knowledge of English needed in specific fields, not just general knowledge. Fauzia (2013) stated that the critical role of English in vocational education itself, as seen from three aspects, namely 1) aspects of graduates, 2) aspects of graduate users and 3) aspects of the nation's and state's educational progress. One of the tertiary institutions in Indonesia has set English as a gen eral course that students in a study program must follow. This also applies to the Meteorology Study Program at the High School of Meteorology, Climatology, and Geophysics. English in higher education is classified based

on its purpose into two, namely General English and English for Specific Purposes or English for specific purposes is an approach to learning English that focuses explicitly on the use of English in the scientific field. And certain professions. In theory, General and Special English do not significantly differ when implemented in the class. However, ESP or English for Specific Purpose introduces and focuses on the field's context Hutchinson & Waters (1987).

Seeing the importance of English in higher education, especially vocational education, English for Specific Purposes is very much needed in learning. Thus the ESP teaching materials at the College of Meteorology, Climatology, and Geophysics of the Meteorology Study Program must, of course, contain vocabulary and terminology commonly used in the field of meteorology and their use by the profession of a meteorologist. However, in Indonesia, ESP meteorology teaching materials that suit the abilities of Indonesian students are minimal. Therefore, a study is needed to determine the linguistic features of meteorology books and articles, including terminology and grammar. Based on this, the formulation of the problem in this study focuses on vocabulary and terminology as well as linguistic features of the words used in the meteorological article book entitled 'Predicting Hurricane Trajectories using a Recurrent Neural Network' written by Sheila Alemany, Jonathan Beltran, Andrian Perez, and Sam Ganzfried.

The article is one of the materials provided by the STMKG meteorology study program lecturer. On page 7, this article contains a list of how to predict storms with a Recurrent Neural Network system. With such complete information, it can be ensured that the English terminology and vocabulary needed by students of the Meteorology Study Program can be extracted from the article.

English for Specific Purposes, often abbreviated as ESP, is one of the well-known fields of applied linguistics and has been developing for a long time, even starting in the early 1970s. In line with its development, in Indonesia itself, the term ESP is also familiar. However, this term is still limited and only familiar to particular academics who are involved in teaching English only (Kusni, 2007). (Robinson, 1991, p. 5) also defines ESP, namely, "It is generally used to refer to the teaching and learning of a foreign language for a utilitarian purpose of which there is no doubt" (Robinson,

1991). The specific purposes in ESP are generally related to either one's profession or one's academic studies (Dudley-Evans, 2001).

Thus, ESP is the teaching and learning of foreign languages, especially English, whose approach and assumptions are different from General English. Nur (2018) explains that ESP, in this case, is more inclined to learn a language in context rather than problems with grammar and the structure of the language itself. Therefore, authentic contexts and the use of language in unique situations associated with specific fields can evoke enthusiasm for ESP learning.

In tertiary institutions, especially in vocational education Rusmala (2018) explains that the role of English courses (General English), moreover ESP is huge, especially when students are studying or when using various media, tools, or machines that the language of instruction is in English and what is even more important is when students graduate and start entering the world of work.

Dudley-Evans et al., (1998) divide EOP into two parts: English for Professional Purposes, which is divided into English for Medical Purposes, and English for Business Purposes. Then English for Vocational Purposes is divided into Pre-Vocational English and Vocational English. This EOP focuses on work, namely how to use English or the need for English in professional fields such as medicine and business. In English for Vocational Purposes, Pre-Vocational English focuses on preparing yourself to get a job and hone interview skills. Meanwhile, Vocational English is more focused on training for more specific trades or jobs.

Corpus linguistics is a quiet study of language based on the extensive collection of language uses stored in the corpora/corpus. The size of the corpus is related to the number of the corpus. There are three aspects to consider in understanding the concept (Baker, 2995)First, the corpus is primarily a collection of texts that are generated electronically and can be analyzed automatically or semi-automatically. Second, the corpus not only contains a collection of written texts but also includes utterances. Third, the corpus may also include many texts of diverse origin sources, for example, from various authors and speakers and on various topics. In Baker's opinion, there are four criteria for understanding the corpus in a broad concept: shape, size, representative, and open and closed. The database in the linguistic corpus is computerized, which is made for linguistic research.

Hunston (2002) defines 'corpus' as a collection of samples of natural language consisting of several sentences from a series of written texts or recordings collected for linguistic study. The text in spoken and written language is then arranged systematically. The corpus is said to be "natural" because the texts collected are produced and used somewhat and not artificially (as is). According to McEnery & Wilson (1996)the linguistic corpus is a comprehensive system containing methods and principles for applying corpus in language research and teaching or learning. Several corpus software can be used to analyze data quantitatively, one of which is Antconc. Anthony is suitable for this research because Antconc has the features needed to achieve research objectives. The aim research is to know how to use of Corpus Linguistics in Preparing English Teaching Materials for Specific Purposes in the Field of Meteorology.

II. METHOD

This study uses a mixed method (mixed method), namely qualitative and quantitative methods. At the beginning of the research, the method used was quantitative, which functioned to obtain the data to be examined, and qualitative methods were used in the analysis process. The data that becomes the corpus in this study is a text journal article entitled Predicting Hurricane Trajectories using a Recurrent Neural Network. The following are the steps in analyzing the data and features used in this study:

1) In the early stages, quantitative methods were used to obtain quantitative data, which was carried out with the help of the Ant Conc concordance software.
2) The meteorological articles used as data sources will be converted, initially in pdf format, to plain text so Ant Conc can process them.
3) Using the frequency feature. This feature generates a list of all the words used in the article based on their frequency of occurrence, starting from the most frequent to the word that appears only once.
4) The list generated by the software is then sorted in the reduction process to separate words that have whole meanings and other words such as abbreviations, pieces of syllables, or numbers.
5) Classify significant words into their respective word classes according to their function in the sentence.
6) After obtaining the selected data, the words are analyzed qualitatively to find the linguistic features attached to each word class and then draw conclusions.

III. RESULT AND DOSCUSSION

Antconc Analysis

At this stage, the researcher initially converted articles related to meteorology studies in PDF form into TXT to be processed in the Ant Concordance software. Through the wordlist feature, the document is processed by the software. It produces a list of words in the radiology book based on the frequency (number of occurrences) according to corpus linguistics.

Table 1. List of 100 words resulting from the Ant Conc process based on their frequency (Freq.)

Words types : 1269 **Word Tokens : 5359**

No	Frek	Leksikal	No	Frek	Leksikal
1	392	the	51	15	based
2	204	of	52	15	dynamical
3	162	and	53	15	more
4	131	to	54	15	or
5	101	a	55	14	complexity

6	89	hurricane	56	14	errors
7	81	in	57	14	it
8	67	is	58	14	over
9	56	grid	59	14	statistical
10	51	model	60	14	values
11	48	for	61	13	accuracy
12	47	rnn	62	13	an
13	46	as	63	13	behavior
14	45	are	64	13	set
15	43	from	65	13	their
16	39	with	66	13	using
17	38	data	67	12	al
18	37	by	68	12	be
19	37	hurricanes	69	12	could
20	35	that	70	12	et
21	32	this	71	12	forecasting
22	31	neural	72	12	learning
23	27	our	73	12	locations
24	26	models	74	12	networks
25	24	latitude	75	12	nonlinear
26	24	network	76	12	points
27	24	we	77	12	prediction
28	23	trajectories	78	12	rnns
29	22	at	79	12	scale
30	22	used	80	12	sparse
31	21	recurrent	81	12	than
32	20	hidden	82	11	forecast
33	20	longitude	83	11	output
34	20	nhc	84	11	predict
35	20	training	85	11	value
36	20	was	86	11	weather
37	19	employed	87	11	weight

38	18	trajectory	88	11	wind
39	18	were	89	10	center
40	17	figure	90	10	collected
41	17	not	91	10	layers
42	17	on	92	10	results
43	17	time	93	10	system
44	16	can	94	10	track
45	16	due	95	10	when
46	16	each	96	9	about
47	16	error	97	9	atlantic
48	16	features	98	9	between
49	16	methods	99	9	complex
50	16	these	100	9	dynamic

Word types indicate the types of individual words used in the article. The number of words used is 1269, which will then be reduced, classified, and analyzed. Meanwhile, word tokens are the total number of words in the book, including repeated words; token words are the total number of words resulting from the sum of the frequency of each word. Words have been taken in the article as much as 5359 times.

Data Reduction

The analysis phase continued with data reduction. At this stage, the researcher focuses on the field data that has been collected. The data used is frequency data which has been reduced by separating significant words and abbreviations from other words. This stage determines the degree of relevance to the research objectives.

Table 2. Basic word list

No	Frequency	Basic word
1	89	hurricane
2	56	grid
3	31	neural
4	24	latitude

5	24	network
6	20	longitude
7	18	trajectory
8	12	scale
9	12	sparse
10	11	forecast
11	11	predict
12	11	weather
13	11	wind
14	10	track
15	9	atlantic

Table 3. List of abbreviations

No	Frequency	Basic word
1	47	rnn
2	20	nhc
3	12	al
4	12	et
5	12	rnns

Data Classification

The researcher classifies the selected data at this data classification stage, namely the meaningful essential words in table 2. The classification is selected by selecting and separating the list of words based on their respective word classes. According Gelderen (2010) who explains that lexical categories (word classes) in English are syntactically divided into five categories, namely Noun/Noun (N), Verb/Verb (V), Adjective/Adjective (Adj), Adverb/Adverb (Adv) and prepositions (P).

Table 4. List of words and their classes

No	Word Class	Word Example	Number of Words
1	Kata Benda (noun)	hurricane, grid, model, data, latitude, trajectory, figure, time, error, features, methods, complexity, values, accuracy, behavior, set, locations, points, prediction, scale, forecast, output, weather, weight, wind, center, layers, results, system, track Pronoun : our,	
2	Kata Kerja (verb)	sparse, employed, predict, collected, using, forecasting, learning	
3	Kata Sifat (adjective)	neural, based, recurrent, hidden, atlantic, complex, dynamic	
4	Kata Depan (preposition)	of, to, in, as, by, at	
6	Singkatan	rnn, nhc, al, et, rnns	
7	Lain-lain	Konjungsi: and, or, from, about artikel (an, a, the)	

Data Analysis

The final stage is data analysis using qualitative methods. Based on the theory put forward by Gelderen (2010) and Quirk. At al (1980) then found the features of each word class, which can be concluded as follows:

1) *The ending –s as a singular and plural noun marker*
   - *Countable noun : trajectories, features, grid,*
   - *Uncountable noun : data, behavior, scale*
   - *Abstract noun : system, prediction, complexity*

   There are also noun pronouns such as we, our, their, where, which, whose
2) *The collected verbs are:*
   - *Main verbs : is, are, were, was, be, used, sparse, employed, predict, collected, using, forecasting, learning, destroyed, forced, presented*
   - *Auxiliary verbs : is, are, was, were, has, have*
   - *Modal auxiliary verbs : can, could*
3) There are types of active sentences and passive sentences and tenses in the form of Simple present, Simple past, Present continuous, Simple present perfect, also found the use of gerunds.
4) There is the addition of the ending –er for smaller, more for more accurately, most for most destructive
5) There are adverbs that appear in the article
   - *Adverb of manner : specifically, accurately,*
   - *Adverb of frequency : currently*
6) There is a conjunction
   - *Coordinate conjunctions: and, or, but*
   - *Subordinate conjunction: although, because, however*

IV. CONCLUSION

ESP or English for Specific Purposes, introduces and focuses on the field's context. English for Specific Purposes is needed in a lesson. The controlled vocabulary will then be used in preparing the English for Specific Purposes module of teaching materials for the field of meteorology. These words will be arranged based on the grammar found in the analysis and then compiled in the form of text in modules that include language competencies, namely listening and reading, and will be used in speaking and writing. By obtaining a list of meteorological vocabulary and terminology as well as the grammar and linguistic features of these words needed in the preparation of English for Specific Purposes teaching materials, it can be concluded that the use of Corpus Linguistics is beneficial and quite effective in preparing ESP teaching materials for the Meteorology Study Program.

REFERENCES

Baker, C. L. (1995). *English Syntax*. the MTI Press.

Dudley-Evans, T. (2001). *English for Specific Purposes. In R. Carter, & D. Nunan (Eds.), The Cambridge Guide to Teaching English to Speakers of Other Languages*. CCambridge University Press.

Dudley-Evans, T., St John, M. J., & Saint John, M. J. (1998). *Developments in English for Specific Surposes: A Multidisciplinary Approach*. Cambridge university press.

Gelderen, E. V. (2010). *An introduction to the grammar of English (Revised edition)*. John Benjamins Publishing Company.

Hunston, S. (2002). *Corpora in Appllied Linguistics*. Cambridge University Press.

Hutchinson, T., & Waters, A. (1987). *English for Specific Purposes*. Cambridge University Press.

Kusni. (2007). Reformulasi Perancangan Program ESP di Perguruan Tinggi. *Linguistik Indonesia*, *35*(1), 63–72.

McEnery, T., & Wilson, A. (1996). *(1996). Corpus Linguistics*. Edinburgh University Press.

Nur, M. (2018). Penerapan Esp di Perguruan Tinggi Umum (Non English Majors) melalui Pendekatan Content-based Instruction-cbt. *Mabasan*, *12*(1), 86–103.

Rusmala, M. (2018). Peran Mata Kuliah Bahasa Inggris dalam Pendidikan Vokasi di STKOM Sapta Computer Kalsel. *Seminar Nasional Pendidikan: Peningkatan Kualitas Pendidikan Tinggi, Dasar Dan Menengah*.