CORPUS-BASED ANALYSIS OF TRANSITION WORDS

Satyawati Surya

English Literature Department, Faculty of Cultural Sciences,
Mulawarman University
Email Correspondence: suryasatyawati@yahoo.com

Abstrak: *This study is regarded as a corpus-based investigation which is an empirical study based on corpus linguistics approaches. By using the device of corpora, the investigation over one aspect of writing, that is achieving coherence and cohesion by using transitional signals will be actualized. Corpus approaches are utilized to find the data needed to anchor the usage of transitional signals in producing text or discourse. For this purpose, the researcher would like to find information about the frequency used among the transition words only. The researcher uses a sketch machine of BROWN Corpus to find the data. The aims of the research are: to search for the most order of transitional signal frequency in the English and Indonesian versions; to find out the most order of transitional signal frequency based on the function in English and Indonesian versions. This research belongs to qualitative and quantitative methods. The data found are collected and displayed by using a list of words, table, and description in order to answer the research questions. that the most order of transitional signal frequency in English are: and, but, also, while in the Indonesian version are: dan, juga. The writers of English and Indonesia have the same preference, style, and diction on the function of contrast (but and tetapi), additional information (and, dan), and concluding ideas (finally and pada akhirnya).*

Keywords*: Corpus Linguistic Approach, Brown Corpus, transition words*

Abstrak: Studi ini termasuk penelitian corpus yang merupakan penelitian empiris berdasarkan pendekatan corpus linguistik. Dengan menggunakan alat corpora, penelitian dari aspek menulis, yaitu bagaimana membuat tulisan yang kohesif dan koheren dengan menggunakan penghubung bisa teraktualisasi. Pendekatan corpus digunakan untuk mencari data yang dibutuhkan untuk menguatkan penggunakan penghubung dalam menghasilkan teks atau wacana. Untuk tujuan inilah peneliti mencari informasi mengenai frekuensi dari penghubung tersebut. Peneliti menggunakan sketch machine Brown Corpus untuk menemukan data tersebut. Tujuan dari penelitian ini adalah: untuk mencari frekuensi penghubung yang paling banyak digunakan di Bahasa Inggris dan Bahasa Indonesia, dan untuk mencari penghubung yang paling banyak digunakan berdasarkan fungsinya di Bahasa Indonesia dan Bahasa Inggris. Studi ini termasuk penelitian kualitatif dan kuantitatif. Data yang diperoleh ditampilkan menggunakan kata-kata, table, dan deskripsi untuk menjawab pertanyaan penelitian. Penghubung yang paling banyak frekuensinya dalam Bahasa Inggris adalah: and, but, also, Sementara itu untuk Bahasa Indonesia adalah: dan, juga. Penulis Indonesia dan Inggris mempunyai pilihan, gaya, dan diksi yang sama pada penghubung yang berfungsi menyatakan sesuatu yang berlawanan (but dan tetapi), yang berfungsi menambah informasi (and, dan), dan yang berfungsi menyimpulkan (finally and pada akhirnya).

Kata Kunci: Pendekatan Corpus Linguistik, Brown Corpus, Penghubung

CaLLs

Sesanti

Seminar Nasional Bahasa, Sastra, dan Seni
"Membumikan Kesadaran Multikultural Masyarakat Melalui Ibu Kota Negara Nusantara (IKN)"

A. INTRODUCTION

This study is regarded as a corpus-based investigation which is an empirical study based on corpus linguistics approaches. The corpus linguistics approach is the study of language in use through corpora (Bennet, 2010). By using the device of corpora, the investigation over one aspect of writing, that is achieving coherence and cohesion by using transitional signals will be actualized. Writing is an important skill that needs to be established as a tool of communication.

Actually, writing is a productive skill that is necessary to master English learners. There are some aspects contributing to potential results for good composition. Byrne (1993) asserts that to be able to communicate successfully, we have to do some effort to ensure the text produced can be interpreted by readers by managing our sentence structure, so they are linked together and sequenced. In addition, he states that organizing ideas in such a way is important to learn in order to be understood.

One of the requirements of producing good writing is constructing coherent and cohesive paragraphs. In order to create paragraphs that are smoothly connected is by using coherent and cohesive devices. Such kind of devices in the form of words and phrases connect sentences and paragraphs together, creating a smooth flow of ideas (Zemach & Rumisek, 2005). Furthermore, they propose transition words and phrases in English that are used to connect sentences and ideas together which relate them to one another. The following table shows common transitions suggested used by Zemach and Rumisek (2005) based on the function of each transition.

Chronology	Comparison	Contrast	Additional Information	Example	Cause and Effect	Concluding Ideas
before after then since first,second while when	likewise compared to similarly	however on the other hand but yet in spite of in contrast eventhough	and also in addition in fact furthermore moreover	for example in general generally for instance specifically in particular	therefore so thus as a result because	in conclusion in summary finally to summarise

CaLLs
Sesanti

Seminar Nasional Bahasa, Sastra, dan Seni
"Membumikan Kesadaran Multikultural Masyarakat Melalui Ibu Kota Negara Nusantara (IKN)"

		instead			

Corpus approaches are utilized to find the data needed to anchor the usage of transitional signals in producing text or discourse. For this purpose, the researcher would like to find information about the frequency used among the transition words only. Moreover, it is beneficial if the data found is compared with the transitional signal used by Indonesian writers to compose discourse. By doing this, it will give potential data and reflection on the style and diction used by American writers and Indonesian ones.

The researcher uses a sketch machine of BROWN Corpus to find the data. BROWN Corpus is an electronic collection of text samples of American English, the first major structured corpus of varied genres. This corpus first set the bar for the scientific study of the frequency and distribution of word categories in everyday language use. Compiled by Henry Kučera and W. Nelson Francis at Brown University, in Rhode Island, it is a general language corpus containing 500 samples of English, totaling roughly one million words, compiled from works published in the United States in 1961 (https://en.wikipedia.org/wiki/Brown_Corpus accessed on March 29, 2023).

The sketch machine is an application of BROWN Corpus which can be browsed by using the link:

https://app.sketchengine.eu/#dashboard?corpname=preloaded%2Fententen20_tt31_1.

The sketch machine of BROWN corpus is an application device that is supposed to register first to get access to use the machine. In the dashboard, the users can choose any language, get info, and manage the corpus. There are some rooms that the users can choose based on their interests which encompass: word sketch, thesaurus, parallel concordance, N-grams, trends, one-click dictionary, word sketch difference, concordance, wordlist, keywords, text type analysis, and bilingual terms. The users choose word sketch to search collocation and word combination. Thesaurus is used to search synonyms and similar words. Parallel concordance is for translation search. N-grams are utilized to search multiword expressions. Trends are utilized for the purpose of diachronic analysis and neologisms. The one-click dictionary is utilized for automatic dictionary drafting. Word sketch difference is for searching to compare the collocation of two words. Concordance has a

CaLLs
Sesanti

Seminar Nasional Bahasa, Sastra, dan Seni
"Membumikan Kesadaran Multikultural Masyarakat Melalui Ibu Kota Negara Nusantara (IKN)"

function to find examples of use in context. The wordlist contains a frequency list. Keywords contain terminology extraction. Text type analysis shows statistics of the whole corpus. And bilingual terms contain bilingual terminology extraction.

This research applies the corpus approach methodology to search for the data of frequency to find the most order of transitional signals in text or discourse which are frequently used by the writers to produce their writing, and also to compare between the frequency of transitional signals used for English and Indonesian writing. For these purposes, it is necessary to formulate the research questions to guide the project.

The following statements are the research questions of the research which become the guidance for accomplishing this project:

1. How is the most order of transitional signal frequency in the English and Indonesian versions?
2. How is the most order of transitional signal frequency based on the function in English and Indonesian versions?

B. METHOD

This research belongs to qualitative and quantitative methods. Qualitative researchers seek to understand a phenomenon by focusing on the total picture rather than breaking it down into variables. The goal is a holistic picture and depth of understanding rather than a numeric analysis of data (Ary, Jacobs, Sorenson, Razavieh, 2010). The corpus-based investigation is applied to search the data for finding the most order of the frequency of transitional signals and also classifying the most order based on the function. In order to find the data of frequency, the researcher applied word sketch and wordlist available in the dashboard of the sketch machine. The data found are collected and displayed by using a list of words, table, and description in order to answer the research questions.

C. RESULTS AND DISCUSSION

The following findings show the research results based on thorough observation and classification.

Here is the frequency of transitional signals in English based on the most order:

1. and: 1,098,435,820
2. but: 122,651,396
3. also: 70,097,413
4. so: 60,914,767
5. Then: 39,717,942
6. however: 18,736,061
7. First: 7,250,406
8. instead: 7,231,959
9. thus: 6,099,815
10. therefore: 5,495,270
11. finally: 5,486,759
12. in particular/particularly: 4,686,760
13. generally: 4,099,846
14. Before: 3,050,467
15. specifically: 2,542,361
16. while: 2,077,967
17. compared to: 1,618,849
18. moreover: 1,210,644
19. yet: 1,193,016
20. similarly: 1,144,435
21. furthermore: 1,143,247
22. Since: 1,125,779
23. as a result: 1,017,048
24. likewise: 662,057
25. in fact: 638,611
26. second: 596,045
27. because: 396,956
28. After: 363,171
29. in contrast:122,206
30. in spite of: 102,557

CaLLs
Seminar Nasional Bahasa, Sastra, dan Seni
"Membumikan Kesadaran Multikultural Masyarakat Melalui Ibu Kota Negara Nusantara (IKN)"

Sesanti

31. for instance: 96,024

32. for example: 86,766

33. eventhough: 46,625

34. in addition: 38,921

35. when: 28,561

36. in conclusion: 19.108

37. on the other hand: 18,370

38. in summary:17,382

39. to summarise: 10,744

40. in general: 7,952

Based on the previous list of words containing transitional signals and frequency of words in the English version, the big three of the most order is: and, but, also. The preference of the writers to apply transitional signals in their writing is on using a conjunction (and, but, also) as transitional signals to connect ideas. However, the big three of the least order is: in general, to summarise, in summary. The writers seem seldom to end the text by giving transitional signals to show the final part of the composition. They prefer to end it directly without giving any signals.

The following is the frequency of transitional signals in Indonesia based on the most order:

1. dan: 2,407,235

2. juga: 342,193

3. karena: 286,632

4. tetapi: 148,625

5. namun: 120,813

6. ketika: 119,190

7. Setelah: 108,426

8. kemudian: 99,759

9. jadi: 83,742

10. pertama: 77,232

11. kedua: 67,311

12. sejak: 55,780

![CaLLs Sesanti]

Seminar Nasional Bahasa, Sastra, dan Seni
"Membumikan Kesadaran Multikultural Masyarakat Melalui Ibu Kota Negara Nusantara (IKN)"

13. Sebelum: 54,017

14. misalnya: 40,458

15. meskipun, walaupun: 32,135

16. oleh karena itu: 29,710

17. meskipun, walaupun: 25,790

18. pada saat: 25,523

19. selain itu: 20,037

20. selanjutnya: 19,977

21. khususnya/terutama: 19,669

22. umumnya: 19,021

23. dengan demikian: 14.377

24. tetapi juga: 9.345

25. dibandingkan dengan: 8,250

26. pada umumnya: 8,022

27. pada akhirnya: 6,662

28. demikian pula: 4,588

29. di lain pihak: 3,798

30. secara khusus: 3,118

31. sebagai akibat: 2,251

32. pada kenyataannya: 1,544

33. sebagai tambahan: 1,152

34. sebaliknya: 644

35. ringkasnya: 569

36. secara ringkas: 537

37. malahan: 134

38. kesimpulannya: 83

Based on the previous list of words containing transitional signals and frequency of words in the Indonesian version, the big three of the most ordered is dan, juga, karena. Similar to the English version, Indonesian writers also prefer to use conjunction (dan, juga) to connect ideas. However, the big three of the least order is: kesimpulannya, malahan, secara ringkas.

Satyawati Surya
Corpus-Based Analysis of Transition Words

CaLLs

Seminar Nasional Bahasa, Sastra, dan Seni
"Membumikan Kesadaran Multikultural Masyarakat Melalui Ibu Kota Negara Nusantara (IKN)"

Sesanti

It is also a tendency that Indonesian writer seldom uses the signal of conclusion to end their composition.

The following table is the frequency of transitional signals based on the function in the English version.

Chronology	Comparison	Contrast	Additional Information	Example	Cause and Effect	Concluding Ideas
Before (14) After (28) then (5) since (22) first (7) second (26) while (16) when (35)	Likewise (24) compared to (17) similarly (20)	However (6) on the other hand (37) but (2) yet (19) in spite of (30) in contrast (29) even though (33) instead (8)	And (1) Also (3) in addition (34) in fact (25) furthermore (21) moreover (18)	for example (32) in general (40) generally (13) for instance (31) specifically (15) in particular/ particularly (12)	Therefore (10) so (4) thus (9) as a result (23) because (27)	in conclusion (36) in summary (38) finally (11) to summarise (39)

a.   Based on the function of chronology, the occurrence of transitional signals for the most order is: then, first, before, while, since, second, after, when.

b.   Based on the function of comparison, the occurrence of transitional signals for the most order is: compared to, similarly, likewise.

c.   Based on the function of contrast, the occurrence of transitional signals for the most order is:

but, however, instead, yet, in contrast, in spite of, even though, on the other hand.

d.   Based on the function of additional information, the occurrence of transitional signals for the most order is: and, also, moreover, furthermore, in fact, in addition.

e.   Based on the function of example, the occurrence of transitional signals for the most order is:

in particular/particularly, generally, specifically, for instance, for example, in general.

f. Based on the function of cause and effect, the occurrence of transitional signals for the most order is: so, thus, therefore, as a result, because.

g. Based on the function of concluding ideas, the occurrence of transitional signals for the most order is: finally, in conclusion, in summary, to summarise.

The following table is the frequency of transitional signals based on the function in Indonesian version.

Chronology	Comparison	Contrast	Additional Information	Example	Cause and Effect	Concluding Ideas
Sebelum (13) Setelah (7) Kemudian (8) Sejak (12) Pertama (10) Kedua (11) pada saat (18) Ketika (6)	Juga (2) dibandingkan dengan (25) demikian pula (28)	Namun (5) di pihak lain (29) tetapi (4) meskipun (15) sebaliknya (34) walaupun (17) malahan	Dan (1) tetapi juga (24) sebagai tambahan (33) pada kenyataannya (32) selanjutnya (20) selain itu (19)	Misalnya (14) pada umumnya (26) umumnya (22) secara khusus (30) khususnya/terutama (21)	Oleh karena itu (16) Jadi (9) dengan demikian (23) sebagai akibat (31) karena (3)	Kesimpulannya (38) secara ringkas (36) pada akhirnya (27) ringkasnya (35)

a. Based on the function of chronology, the occurrence of transitional signals for the most order is: ketika, setelah, kemudian, pertama, kedua, sejak, sebelum, pada saat.

b. Based on the function of comparison, the occurrence of transitional signals for the most order is: juga, dibandingkan dengan, demilian pula.

c. Based on the function of contrast, the occurrence of transitional signals for the most order is: tetapi, namun, meskipun, walaupun, di pihak lain, sebaliknya.

d. Based on the function of additional information, the occurrence of transitional signals for the most order is: dan, selain itu, selanjutnya, tetapi juga, kenyataannya, sebagai tambahan.

e. Based on the function of example, the occurrence of transitional signals for the most order is:

misalnya, khususnya/terutama, umumnya, pada umumnya, secara khusus.

CaLLs
Sesanti

Seminar Nasional Bahasa, Sastra, dan Seni
"Membumikan Kesadaran Multikultural Masyarakat Melalui Ibu Kota Negara Nusantara (IKN)"

f. Based on the function of cause and effect, the occurrence of transitional signals for the most order is: karena, jadi, oleh karena itu, dengan demikian, sebagai akibat.

g. Based on the function of concluding ideas, the occurrence of transitional signals for the most order is: pada akhirnya, ringkasnya, secara ringkas, kesimpulannya.

Based on the data stated in the finding searched by using Brown Corpus, it is clear that the style of using transitional signals for achieving coherent and cohesive text for both English and Indonesian language utilized by the writers tends to have similar choices and preferences. The most transition used based on the data of frequency is using conjunction: and, but, also for the writers who use English. Likewise, the writers who use Bahasa Indonesia also tend to utilize conjunction: dan, juga, in order to connect words, phrases, clauses, sentences, and ideas. Conjunctions are very useful to connect words in a series of items and also to connect clauses and sentences in making compound sentences and complex sentences, as well as transitional signals. It might be the reason that the frequency of the conjunction words is numerous.

Moreover, the least transitions used to connect words, phrases, clauses, sentences, and ideas based on the investigation using Brown Corpus are the ones that are used to give a conclusion. Both the writers using English and Bahasa Indonesia seem seldom to use the expressions of to summarise and in summary for English text, and also kesimpulannya and secara ringkas for Indonesian text. In fact, for the concluding paragraph or the last part of the text, the writers sometimes want to remind the readers about important points stated previously or provide signals that they want to end the text. So, they put the transitional signals like in summary, in conclusion at the end paragraph. On the other hand, some writers want the readers to think about the prediction for the following paragraph or give suggestions that make them not give any signals to accomplish the text.

The most order of frequency of transitional signals based on the function in both English and Indonesian writers show the writers' preference for their choice, style, and diction. The most orders based on the function of transitional signals for English and Indonesian text are: for the function of chronology: then (English) and ketika (Indonesia), for the function of comparison: compared to (English) and juga (Indonesia), for the function of

contrast: but (English) and tetapi (Indonesia), for the function of additional information: and (English) and dan (Indonesia), for the function of example: in particular/particularly (English) and misalnya (Indonesia), for the function of cause and effect: so (English) and karena (Indonesia), for the function of concluding idea: finally (English) and pada akhirnya (Indonesia). In conclusion, the writers of English and Indonesia have the same preference, style, and diction on the function of contrast (but and tetapi), additional information (and, dan), and concluding ideas (finally and pada akhirnya).

D.   CONCLUSION

For answering the first research question, that is: How is the most order of transitional signal frequency in the English and Indonesian versions? The answer appoints that the most order of transitional signal frequency in English are: and, but, also, while in Indonesian version are: dan, juga. They are similar in terms of using part of speech "conjunction" as the most order of transitional signal.

Next, the second research question is: How is the most order of transitional signal frequency based on the function in English and Indonesian versions? The answer appoints that the writers of English and Indonesia have the same preference, style, and diction on the function of contrast (but and tetapi), additional information (and, dan), and concluding ideas (finally and pada akhirnya). While for the other function, they show differences in terms of preference of choices in style and diction.

REFERENCES

Ary, D., Jacobs, L.C., Sorensen, C., Razavieh, A. (2010). *Introduction to Research in Education.* 8th edition. Canada: Nelson Education, Ltd.

Bennet, G.R. 2010. *Using Corpora in the Language Learning Classroom: Corpus Linguistics for Teachers.* Michigan ELT. http:/www.press.umich.edu/titleDetailDesc.do?id=371534.

Byrne, D. (1993). *Teaching Writing Skill New Edition.* London & New York: Longman.

Ze,ach, D.E. & Rumisek, L.A. (2005). *Academic Writing From Paragraph to Essay.* London: Macmillan Publishers Limited.

CaLLs
Sesanti

Seminar Nasional Bahasa, Sastra, dan Seni
"Membumikan Kesadaran Multikultural Masyarakat Melalui Ibu Kota Negara Nusantara (IKN)"

Brown Corpus Sketch Mahine
https://app.sketchengine.eu/#dashboard?corpname=preloaded%2Fententen20_tt31_1. Accessed on March 29, 2023.

Wikipedia. https://en.wikipedia.org/wiki/Brown_Corpus   accessed on March 29, 2023.