# Insights from a Learner Corpus as Opposed to a Native Corpus about Cohesive Devices in an Academic Writing Context

**Ceylan Yangın Ersanli**

Department of Foreign Languages, Faculty of Education, Ondokuz Mayıs University, Turkey

**Abstract**   This study reports on the insights from an EFL learner corpora (a total of 151 essays and 49,690 words) generated from essays collected over the years in a Turkish state university from freshmen students enrolling in the Advanced Writing course. The comparison of cohesive devices in the non-native corpus (NNC) with those in a native corpus (NC) reveals the overuse and misuse of some cohesive devices by Turkish EFL learners. The study specifically aimed to show the use of cohesive devices in learner essays. The frequency counts of cohesive devices in both the NNC and NC were compared to draw conclusions about the macrostructure of the collected essays. Finally, this study makes some suggestions for improvement in the organisation of essays by non-native EFL learners.

**Keywords**   Cohesive Devices, Native Corpus, Non-native Learner Corpus, Writing Development

## 1. Introduction

New technologies, including computers, have changed every aspect of our lives and education is not an exception. Today, a wide variety of educational software provides opportunities for teachers to design their lessons so that they can meet their students' needs (Lambic, [1]). Besides, the use of computer software, (in particular, software that creates a concordance, known as a concordancer) to process and analyse a large databank of natural texts (corpus) allows us to discover patterns of authentic language use. The immense language data compiled from written texts or the transcription of speech provides empirical data about language behaviour, rather than relying on a subjective view gathered through introspection and intuition.

Over the past few decades, the corpus-informed approach to language teaching has gained prominence, and the pedagogical value of this corpus is acknowledged in syllabus design, materials development and classroom activities (Barlow, [2]).The utilisation of corpora in language classrooms has a lot to offer in terms of vocabulary, grammar, language use and discourse patterns of given text types (Gledhill, [3]; Hyland, [4]; Tribble, [5]). When used effectively, corpora might lead to student-centred learning through discovery. In other words, Data-Driven Learning (DDL) pushes the students to assess authentic language use by using authentic materials, exploratory tasks, and activities rather than those composed for pedagogical purposes or traditional teacher-led activities and materials (Johns, [6]). The underlying rationale for Data-Driven Learning is the principle that "what learners can find out themselves is better remembered than what they are simply told" (Ellis, [7, p.163]). The students, thus, draw conclusions about language use and develop an awareness and eventually attain learner autonomy. Although the use of corpora is welcomed in the field of education, there is still a call for more research to provide empirical evidence as to its usefulness (Varley, [8]).

Choosing the corpora that will serve best for the aims of instruction is significant. A general language corpus might be compiled from fiction, academic discourse, newspaper articles, and casual conversation and may involve several registers. However, a specific corpus, such as a spoken corpus or scientific essays, consists of one of the sub-registers. As language use differs according to register, such as formal or informal registers for instance, choosing the right type of corpus for reference is essential. Using specialised corpora is emphasised in EAP as it can cater to the needs of a specific group of learners.

A review of related literature about corpus linguistics and pedagogy shows numerous studies on the use of collocations, which is found to be one of the most problematic aspects to master (Altenberg & Granger, [9]; Chen, [10]; Liu, [11]; Nesselhauf, [12]; Shei & Pain, [13]). These studies show that corpora-referenced instruction and learning is more effective in collocation learning and retention than traditional learning (Cobb & Horst, [14]; Çelik, [15]; Daskalovska, [16]; Tseng, [17]). Corpus-based research also informed L2 material writers to make principled decisions to emphasise and

prioritise this in textbooks (Biber & Reppen, [18]; Gabrielatos, [19]; Frazier, [20]; Romer, [21]). Frequency information provides insight into words and structures that are central to language use (Romer, [22]; Kennedy, [23]; Conrad, [24]) and thus helps teachers and material designers on what aspects to emphasise and introduce first. Fewer studies, however, focus on the effect of corpora use on students' attitudes and performance in writing in the target language. Corpus analysis can be a useful source in writing instruction since it reveals patterns of actual language use. Yoon and Hirvela [25] find that exploring a corpus helped students to learn the common usage patterns of words, which eventually led to an increased confidence in writing. Gaskell and Cobb [26] guided learners to use online corpora to edit their writing drafts and correct their own grammatical errors.

Research also indicates that learner corpora can be used directly in classroom teaching. Students' written production can be a good indicator of their linguistic competence. Creating a learner corpus composed of students' own writing can provide a source for learning, discovering and correcting errors (Seidlhofer, [27]; Mukherjee & Rohrbach, [28]). Comparing the native and non-native corpora of learners helps identify different uses, such as overuse and misuse of some logical connectors in non-native students' essays (Milton & Tsang, [29]; Granger & Tyson, [30]; Peng, [31]). Such comparison might raise awareness of cohesion and coherence and help avoid mistakes and thus write more authentic texts.

More research into learner corpora will provide information about the proficiency of specific groups of learners and the patterns they make use of and/or common errors that they make. Therefore, this study aimed to report on the use of cohesive devices by Turkish EFL learners at tertiary level in an academic writing course. The frequency and variety of cohesive devices used by Turkish EFL learners were compared with those in native academic essays. The non-native learner corpus consists of 151 essays collected over five years by the researcher. The native reference corpus is the British Academic Written English (BAWE), consisting of 2,761 essays, which has been developed by the University of Warwick, the University of Reading, and the University of Oxford Brookes (Heuboeck, Holmes and Nesi [32]. The present study seeks to find out whether Turkish EFL learners' use of cohesive devices differs from those of native speakers in an academic writing context. If they do, to what extent and in what ways do Turkish EFL learners use different cohesive devices in their academic essays? The participants in the study are majoring in an English-medium program and it is important that they possess better academic writing skills.

## 2. Method

This study was descriptive in nature and adopted a primarily quantitative framework to display Turkish EFL learners' use of cohesive devices in academic essays. The data was derived from the frequency counts of learner and native corpora, and through the evaluation of student essays in terms of cohesive devices. Qualitative analysis, alternatively, looks into the use of cohesive devices in terms of appropriateness and accuracy.

**Participants and the Context of the Study**

A total of 151 students enrolled in the Advanced Academic Reading and Writing course participated in the study with 151 sample essays. The essays were collected over five years in subsequent five academic years in a large state university in Turkey. The participants were in their first year of four-year education program. They were chosen according to the non-probability convenience sampling method suggested by Creswell [33] since all of them were available during the course of the study. The participants were majoring in The English language Teacher Education program (ELTE hereafter). The ELTE program was an English-medium program, which required advanced level language proficiency. The participants either passed a proficiency exam or attended a one-year intensive language program to reach a proficiency level to follow the courses in English.

The Advanced Academic Reading and Writing course aimed to enable the participants to write in different academic genres, and essay types such as addition, summation, apposition, result, contrast and transition to fulfill academic requirements as university students. Among the course objectives were writing paragraphs and essays in accordance with the academic writing rules and standards. Appropriate use of cohesive devices was regarded as a part of cohesion and coherence and significant for producing effective essays.

## 3. Data Collection and Analysis

The data was collected from 151 student essays from junior year graduate students in the Advanced Writing course over five years. All the participants were enrolled in an Advanced Academic Reading and Writing course, in which they were expected to write essays in different essay types and on various topics. A five-year collection of essays were used to create a non-native learner corpus. The non-native learner corpus consisted of 49,690 words and 4140 sentences. Greenbaum [34] states that for an analysis of professional texts, a language corpus of 20,000-30,000 is sufficient. Therefore, it is thought that the corpus was large enough to illustrate patterns in Turkish EFL learners' writing. As the non-native learner corpus is a collection of essays from subsequent academic years rather than a single year, it was believed that it could better represent authentic language production of a specific learner group. The learner corpus was processed via Antconc (Anthony, [35]), which is free and user-friendly software for concordancing and text analysis. Antconc allows users to search for word clusters and sort words by frequency by yielding minimum and

maximum number of appearances of a specific word. Frequency lists of cohesive devices were generated using Antconc and the appropriateness and accuracy of cohesive devices in the context they were used were exemplified by extracts from student essays. The native student essays corpora, on the other hand, came from the British Academic Written English, consisting of 2,761 essays, 6,506,995 words and 269,413 sentences. The corpus is made up of student assignments in three different universities (Oxford Brookes, Reading and Warwick) from 35 disciplines. The British Academic Written English Corpus is available for use and research upon request.

Student essays of the given group were analyzed in terms of frequency and variety of cohesive devices. The frequency counts gathered were used to explore and diagnose students' use of cohesive devices to construct the macrostructure of their essays. Sample extracts were excerpted to exemplify non-native use of cohesive devices. Also, examples from the native corpus were provided for comparison and deeper insight.

The quantitative analysis yielded a comparison of the raw frequencies of cohesive devices in both corpora. The cohesive devices were further classified according to their types such as addition, summation, apposition, result, contrast and transition. After the frequencies were obtained, the results were rendered comparable by using frequencies by the ten thousand, which is referred to as normalising the frequencies. A common test of significance used in evaluating corpus studies is *log-likelihood*. This is a statistical test used to compare the fit between two models. The numerical data needed to do the *log-likelihood* test was frequency in corpus A, frequency in corpus B, the total number of words in corpus A, and the total number of words in corpus B. The likelihood was calculated using a web-based wizard.

## 4. Findings and Discussion

The findings in the tables below show the raw frequencies of cohesive devices (CDs) in both corpora, and the normalised frequencies by ten and one thousand.

**Table 1.** Overall frequencies of cohesive device usage per 10,000 words

|  | Non-Native Corpus | Native Corpus |
| --- | --- | --- |
| Word Count | 49,690 | 6,506,995 |
| Raw frequency of CDs | 1947 | 85,910 |
| CDs/10,000 words | 392 | 132 |

**Table 2.** Overall frequencies of cohesive device usage per 1,000 sentences

|  | Non-Native Corpus | Native Corpus |
| --- | --- | --- |
| Sentence Count | 4,140 | 269,413 |
| Raw frequency of CDs | 1947 | 85,910 |
| CDs/1,000 sentences | 470 | 319 |

As Table 1 and Table 2 show, the quantitative analysis revealed that Turkish writers overuse cohesive devices both at word and sentence level. The word counts indicate that Turkish students at tertiary level tend to use three times as many cohesive devices in their academic writing.

Table 3 presents the observed frequencies of cohesive devices by category, their relative frequencies in texts, and log-likelihood values.

For more insight into use of cohesive devices, below are two excerpts from the corpora:

Despite a few similarities they have, women's outlook on life is very different from that of men. Whereas women usually think with their emotions, men use their logic….First of all, women think love is indispensable for marriage. However, men do not …. Secondly, men are ambitious when it comes to career… Nonetheless, they have some similarities too. In short, men and women are generally very different from each other. *(Non-native student corpus)*

Much more reproductive choice is now available to women… this, combined with shifting social and economic opportunities for women, has led to an increase in the number of childless women. However the anticipated number of children per woman in Europe and the USA is still near or above two… showing that many are still having children. In this essay I will explore why women have children, even though there is now more opportunity for them not to, and why those who do not have children do not do so. *(text 0001d, BAWE corpus)*

**Table 3.** The log-likelihood values of cohesive devices by category

| Types of cohesive devices | O1 NNS Corpus | %1 | O2 NS Corpus | %2 | Log-likelihood |
| --- | --- | --- | --- | --- | --- |
| Enumeration & addition | 812 | 1.63 | 23,114 | 0.36 | + 1190.28 |
| Summation | 115 | 0.23 | 740 | 0.01 | + 459.01 |
| Apposition | 104 | 0.21 | 6,575 | 0.10 | + 43.45 |
| Result & inference | 485 | 0.98 | 26,876 | 0.41 | + 271.78 |
| Contrast & concession | 351 | 0.71 | 25,538 | 0.39 | + 99.65 |
| Transition | 80 | 0.16 | 3,067 | 0.05 | + 82.36 |

As can be seen, the excerpt from the non-native corpus is loaded with several cohesive devices and is characterized by much shorter sentences. The sample from the BAWE, however, includes fewer cohesive devices and the sentences are much longer. The ideas are bind together with craftsmanship displaying effective variation in sentence patterns and length. The non-native learners, on the other hand, appear to be using shorter sentences with a variety of cohesive devices to link ideas.

The table below shows comparison of two randomly selected essays from both corpora.

**Table 4.**   Comparison of two randomly selected essays from both corpora

| Non-native Learner Corpus | | BAWE | |
|---|---|---|---|
| Total words | 582 | Total words | 1480 |
| Total sentences | 41 | Total sentences | 52 |
| Total CD | 27 | Total CD | 43 |
| Cohesive Devices | And x 3 | Cohesive Devices | And x 17 |
| | Because of x 3 | | However x 6 |
| | Because x 3 | | But x 2 |
| | However x 2 | | Not just … but also x 3 |
| | But x 2 | | Therefore x 3 |
| | For instance x 2 | | Although x 3 |
| | Furthermore x 2 | | As x 2 |
| | So x 2 | | During |
| | On the other hand | | Since |
| | In addition | | So |
| | Yet | | Despite the fact that |
| | Whereas | | Firstly |
| | What's more | | Secondly |
| | First of all | | In conclusion |
| | All in all | | |
| | In conclusion | | |

The comparison of the essays yield similar findings as depicted in the excerpts. Non-native learners in the study tend to use a lot more cohesive devices in much shorter essays. They also use a variety of cohesive devices. The sentence count reveals that native students use much longer sentences.

# 5. Conclusions and Suggestions

Based on the discussion of the research findings, it appears that Turkish writers tend to use much shorter sentences and exhibit a striking overuse of cohesive devices in academic written discourse. This may result from a desire to create an elaborative text to get credit. As Flowerdew [36, p. 39] points out, learners may insert too many conjunctions with the expectation of being given credit for them. However, the overuse of linking words makes it more difficult to follow rather than smooth and easy to read. It might be better if they could develop craftsmanship in sentences, combining ideas in a variety of patterns rather than using several cohesive devices to link simple sentences. The Turkish learners may have failed to master cohesion and coherence. To develop Turkish learners' ability to effectively outline and construct essays, they need to be informed about corpora information, and should be trained to choose fewer cohesive devices, but with care. One advantage of corpus-informed studies is the increased awareness of other cohesive devices available and their correct use.

Turkish writers also need help with register requirements and should refer to corpora to discover what cohesive devices are appropriate for academic writing. Corpora studies can increase their awareness of cohesive devices that are common in spoken and written discourse. Concordance lines could also be of help to illustrate the correct usage of cohesive devices.

Further and more detailed studies should be carried out to discover cohesive device use in terms of different essay types. We suggest that learners should be enabled to use corpora as a reference tool when composing essays, and emphasize that this should be a goal in academic writing courses.

# REFERENCES

[1]   Lambic, D. Factors influencing future teachers' adoption of educational software use in classroom. *Croatian Journal of Education*, 2014, 16(3): 815-846.

[2]   Barlow, M. Corpora for theory and practice. *International Journal of Corpus Linguistics*, 1996, 1(1):1-37.

[3]   Gledhill, C. The discourse function of in research article introduction. *English for Specific Purposes,* 2000, 19(2): 115-135.

[4]   Hyland, K. Activity and evaluation: Reporting practices in academic writing. In J. Flowerdew (Ed.), *Academic*

*Discourse.* Harlow, England: Longman, 2002, 115-130.

[5]   Tribble, C. Corpora and corpus analysis: New windows on academic writing. In J. Flowerdew (Ed.), *Academic Discourse*. Harlow, England: Longman, 2002, 131-149.

[6]   Johns, T. From printout to hand-out: Grammar and vocabulary teaching in the context of data-driven learning. *ELT Journal, 1991,* 4: 27-45.

[7]   Ellis, R. *Task-based language teaching and learning*. Oxford: Oxford University Press, 2003.Altenberg, B., & Granger, S. The grammatical and lexical patterning of make in native and non-native student writing. *Applied Linguistics*, 2001, 22(2): 173-194.

[8]   Varley, S. I'll just look that up in the concordance: Integrating corpus consultation into the language learning environment. *Computer Assisted Language Learning,* 2008, 22(2): 133-152.

[9]   Altenberg, B., & Granger, S. The grammatical and lexical patterning of make in native and non-native student writing. *Applied Linguistics*, 2001, 22(2): 173-194.

[10]  Chen, Y. H. The use of corpora in the vocabulary classroom. *The Internet TESL Journal*, 2004, 10(9). Retrieved December 18, 2014, from http://iteslj.org/Techniques/Chen-Corpora.html.

[11]  Liu, C. P. An analysis of collocational errors in EFL writings. In the *Proceedings of the Eighth International Symposium on English Teaching* (p 483-494) Taipei, Taiwan: The Craine Publishing Co, 1999.

[12]  Nesselhauf, N. The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, 2003, 24: 223-242.

[13]  Shei, C., & Pain, H. An ESL writer's collocational aid. *Computer Assisted Language Learning,* 2000, 13/2: 167-182.

[14]  Cobb, T., & Horst, M. Reading academic English: Carrying learners across the lexical threshold. In J. Flowerdew & M. Peacock (Eds.), *Research perspectives in English for Academic Purposes*. Cambridge: Cambridge University Press, 2001.

[15]  Çelik, S. Develping collocational competence through web based concordance activities. *Novitas-ROYAL (Research on Youth and Languages),* 2011, 5(2): 273-286.

[16]  Daskalovska, N. Corpus-based versus traditional learning of collocations. Computer *Assisted Language Learning*, 2013, 28(2): 130-144.

[17]  Tseng, F. P. A study of the effects of collocation on the collocational competence of senior high school students in Taiwan. Unpublished master's thesis, National Taiwan Normal University, Taipei, 2002.

[18]  Biber, D., & Reppen, R. What does frequency have to do with grammar teaching? *Studies in Second Language Acquisition*, 2002, 24: 199-208.

[19]  Gabrielatos, C. Corpora and language teaching: Just a fling or wedding bells. *TESL-EJ,* 2005, 8(4):1-37.

[20]  Frazier, S. A corpus analysis of would-clauses without adjacent if-clauses. *TESOL Quarterly,2003,* 37(3): 443-46.

[21]  Romer, U. Using general and specialized corpora in English

language teaching: past, present and future. In M. C. Campoy, B. Beles-Fortuno & M. L. Gea-Valor (Eds.), *Corpus-based Approaches to English Language Teaching*. London: Continuum, 2010, 18-38.

[22]  Romer, U. A corpus-driven approach to modal auxiliaries and their didactics. In J. Sinclair (Ed.), *How to Use Corpora in Language Teaching*. Amsterdam: John Benjamins, 2004, 185-199.

[23]  Kennedy, G. *An introduction to corpus linguistics*. London: Longman Publishing, 1998.

[24]  Conrad, S. M. Will corpus linguistics revolutionize grammar teaching in the 21$^{st}$ century? *TESOL Quarterly,* 2000, 34(3): 548-560.

[25]  Yoon, H., & Hirvela, A. ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing,* 2004, 13(4): 257-283.

[26]  Gaskell, D., & Cobb, T. Can learners use concordance feedback for writing errors? *System*, 2004, 32(3): 301-319.

[27]  Seidlhofer, B. Pedagogy and local learner corpora: working with learning driven data. In S. Granger, J. Hung & S. Petch-Tyson (Eds.), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Philadelphia: John Benjamins, 2002, 213–234.

[28]  Mukherjee, J., & Rohrbach, J. Rethinking applied corpus linguistics from a language-pedagogical perspective: New departures in learner corpus research. In B. Kettemann & G. Marko (Eds.), *Planning, Gluing and Painting Corpora: Inside the Applied Corpus Linguist's Workshop*. Frankfurt: Peter Lang, 2006, 205-232.

[29]  Milton, J., & Tsang, E. S. A corpus-based study of logical connectors in EFL students' writing: Directions for future research. In R. Pemberton & E. S. Tsang (Eds.), *Studies in lexis*. Hong Kong: HKUST Language Centre, 1993, 215-246.

[30]  Granger, S., & Tyson, S. Connector usage in the English essay writing of native and non-native EFL speakers of English. *World Englishes,* 1996, 15(1): 17-27.

[31]  Peng, D. Linking adverbials in the academic writing of Chinese learners: a corpus based comparison. In A. Hardie & R. Love (Eds.), *Corpus linguistics abstract book 2013*. Lancaster: UCREL, 2013.

[32]  Heuboeck, A., Holmes, J. & Nesi, H. *The BAWE Corpus Manual for the project entitled "An Investigation of Genres of Assessed Writing in British higher Education"*. The University of Warwick, The University of Reading, Oxford Brookes University, 2006

[33]  Creswell, J. W. *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*. New Jersey: Prentice Hall Publishing, 2005.

[34]  Greenbaum, S. *The development of the international corpus of English*. London: Longman, 1991

[35]  Anthony, L. AntConc: A learner and classroom friendly, multi-platform corpus analysis toolkit. *Proceedings of IWLeL 2004: An Interactive Workshop on Language e- learning*, pp. 7–13, 2004

[36]  Flowerdew, J. *Discourse in English language education*. Routledge: New York, 2013.