Revisiting cohesion in academic writing: A corpus-based analysis of EFL learners' use of conjunctions

Abdulaziz B Sanosi

Department of English Language and Literature, College of Sciences and Humanities, Prince Sattam Bin Abdulaziz University, 16273, Al-Kharj, Saudi Arabia

ABSTRACT
Grammatical and semantic devices writers use to maintain text cohesion have received significant attention for over forty years. Among these devices, conjunctions are particularly emphasized for their role in enhancing communication and the quality of academic writing. Despite extensive research on how EFL learners use conjunctions in their academic writings, the existing body of work remains inadequate. Research findings have consistently shown that EFL learners misuse conjunctions. Consequently, conjunction instruction techniques have been integrated into learning materials and academic writing programs. Nevertheless, the persisting issue necessitates further investigation using current research methods and writing conventions. To address this gap, this study aimed to evaluate EFL learners' usage of conjunctions by employing a contemporary research method. It employed the corpus linguistic method to analyze the use of conjunctions by EFL learners in a 125,000-word corpus and compared it with a native speaker corpus, using Halliday and Hassan's categorization as the analytical framework. It was hypothesized that EFL learners misuse conjunctions in both quantity and quality. The study revealed varied uses of conjunctions in the two corpora. However, these variations were not uniform across all levels, as some aspects of conjunction usage showed similarities between the corpora. The findings suggest the need for updating conjunction instruction and learning materials. Further research, adopting corpora and longitudinal approaches, is recommended to validate these findings.

Keywords: Academic writing; cohesive devices; conjunction misuse; logical connectors

INTRODUCTION

The property of text being a *unified whole* in terms of its semantic connotation has remained a topic of interest for a considerably long time. Researchers have studied the formal structures and semantic features that make a text an integrated, understandable, and meaningful unit. Halliday and Hassan (1974) distinguished between written (or spoken structures) as texts or independent groups of sentences, implying that a text has specific features such as being coherent and intact. Then it could be inferred that a text is not just a sequence of sentences strung together but a sequence of units connected in some contextually appropriate ways.

This "connectedness" is called cohesion, and it is a key feature that makes a text understandable. It reflects the writer/speaker's ability to express themselves clearly. A range of linguistic tools is used to maintain the cohesion of the text. Broadly, they are termed cohesive devices, and one of the best instances of them is conjunctions. Conjunctions are found to have a significant role in enhancing communication both verbally and in writing. Conjunctions play a significant role in enhancing communication, both in speech and writing. Their importance is even greater in writing, where there are fewer contextual clues to help understand the text.

*Corresponding Author
Email: a.assanosi@psau.edu.sa

According to Halliday and Hassan (1976), cohesion is attainable to native speakers, who are aware of the linguistic devices that maintain the clarity and lucidity of their texts, though they are not consciously aware of using these devices. When it comes to EFL learners' use of conjunctions, however, it is found that they misuse conjunctions (Jama, 2024; Saeed, 2023) and that there is a positive relationship between using conjunctions and writing quality (He, 2020). Accordingly, helping EFL students use conjunctions correctly is expected to improve their communication skills and writing quality. Nevertheless, the available research on this issue does not match the topic's importance.

Regarding the available studies, there is a paucity of research that analyzes the state of conjunction use by EFL learners (Unubi, 2016), and the development of cohesion among EFL learners (Yamaai, 2023). On the other hand, the available studies employed research designs that either analyzed a limited set of conjunctions or were based on relatively small samples that may not provide accurate results (Bhoomanee & Pothisuwan, 2019, Hosseinpur & Pour, 2022). Therefore, the research gap is represented in the lack of comprehensive, representative, and generalizable studies that thoroughly analyze the usage of conjunctions by EFL learners. Attempting to bridge such research gaps, the present study aimed at revisiting this issue in applied linguistic literature by adopting a more contemporary method, corpus linguistics. The research will utilize two corpora: a learner corpus and a reference one to generate its findings. The study aimed to test two hypotheses which are:

H_1: EFL learners use conjunctions differently from native speakers in quantity and type.

H_2: EFL learners use conjunctions differently from native speakers regarding quality and syntactic placement.

For H_1, it is hypothesized that EFL learners excessively use conjunctions. This hypothesis is based on previous studies on EFL learners' use of conjunctions (Granger & Tyson, 1996; Hosseinpur & Pour, 2022; Saeed, 2023).

As far as the H_2 is concerned, it is hypothesized that EFL learners use conjunctions in initial positions more than native speakers do which can be considered a mark of misuse of conjunction. The claimed hypothesis is based on Granger and Tyson (1996), Lydia (2020), and Tocalo et al. (2022) studies. Additionally, the hypothesis is supported by Chang and Swales's (1999) and Hyland and Jiang's (2017) lists of informality features in academic writing which included using conjunctions in the initial position.

By testing the two hypotheses, the researcher aims to generate further implications for teaching conjunctions to EFL learners based on a modern research method and hence support or refute the previous claims about conjunction use by EFL learners. Further, comparing such use against native speakers' use is envisaged as an added value to research besides analyzing conjunction use according to a well-known and widely adopted classification of conjunctions, that is Halliday and Hassan's (1976) categorization, which is based on their functional grammar concept that has been discussed in several Halliday's other works. Functional grammar provides a set of techniques to analyze the text. For Halliday, a text is a procedure by which specific options are selected to represent language. Writers and speakers utilize grammatical and lexical elements to maintain text unity and meaningfulness, which is known as cohesion and can be exemplified by conjunctions.

Cohesion

Cohesion is defined by Halliday and Hassan (1976) as the meaning relations present in the text and hence mark it as a text. Alawerdy and Alalwi (2022) explained this relation by stating that it is the link between discourse components where the sense of each component can be understood by referring to other components. While Shah et al. (2021) described cohesion as "the meaningful unity among larger units of texts" (p. 433), cohesion is maintained by applying and employing specific linguistic features termed cohesive devices, resources, and components, among other terms. Halliday and Hassan (1976) specified four cohesive components, which are (1) Reference, (2) Substitution and ellipsis, (3) Lexical cohesion, and (4) Conjunctions.

The first three cohesive ties usually operate backwards. In other words, every sentence in a text exhibits some form of cohesion with another one in the text (Alawerdy & Alalwi, 2022). There are exceptions, such as the case with the cataphoric references. However, in the case of conjunctions, the fourth grammatical component of cohesion, the situation is different because they serve as transitional elements that bridge ideas, thoughts, and statements, enabling the reader or listener to follow the logical progression and relationship between different parts of the text.
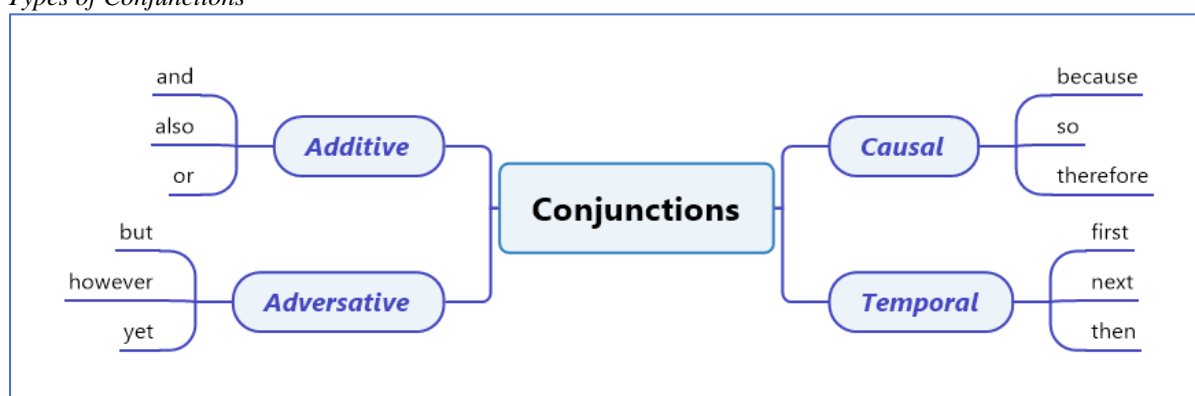
Conjunctions

Conjunctions are linguistic tools that are used to indirectly express some cohesive relationship through certain meanings (Halliday and Hasan, 1976). Other components of the discourse are assumed by these meanings. As a result, conjunctive elements are considered cohesive ties, as they represent a form of cohesion where sentences in a text are connected using a set of ties. These ties help to place the message in an interpretative frame (Hosseinpur & Pour, 2022). More generally, other scholars conceive conjunctions as "linkers or connectors that join two words, phrases, clauses or sentences together, either in speech or writing"

(Unubi, 2016, p. 203). The present study adopts this definition for conjunctions. According to this definition, conjunctions act differently than the other cohesive ties explained in Haliday and Hassan's (1976) classification since "they express certain meanings which presuppose the presence of other components in the discourse." (p. 226). These features enable them to be marked with their semantic role and their importance in communication and storytelling (Sayah & Fatemi, 2013). However, "[e]ven at a reasonably advanced level, connectors are difficult to master" (Granger & Tyson, 1996, p. 24) which can be traced back to "the similarity of conjunction forms and functions and the interference from L1" (Özçelik, 2023, p. 476), and the students' limited repertoire of cohesive devices which causes repetition of conjunction (Rokhaniyah et al., 2022). The issue becomes thornier for EFL students who find conjunction use problematic (Hassoon, 2023; Kurniati, 2019).

A well-known and widely used taxonomy of conjunctions, which is adopted in the current study, is that provided by Halliday and Hassan (1976). In their suggested categorization, they divided conjunctions into four categories. Their scheme, with a few examples for each category, is presented in Figure 1 below.

Figure 1
Types of Conjunctions

Note. Adapted from Halliday, M. and Hassan, R. (1976). *Cohesion in English*. Longman Group Limited. pp. 238-261.

Additive conjunctions "structurally coordinate or link by adding to the presupposed items" (Sayah & Fatemi, 2013, p. 135). This cohesive relation can be simply used to express addition or alternative. Sometimes, they may also be used to negate such a cohesive relation, as in the case of *neither* or *nor*. Adversative conjunctions are employed when there is a relation "contrary to expectation" (Halliday & Hassan, 1976, p. 250) due to the content or the communication setting. The adversative relation can occur at the sentence level between clauses or across sentences. In the case of the causal relation, the reason for what is being stated is expressed, formulating a relation with its result or consequence. Temporal conjunctions are used to show relationships or order between ideas, such as next, third, and then. (Jama, 2024).

Research on conjunctions was not limited to their types and functions. Subsequent studies have also focused on their syntactic positioning in discourse. Granger and Tyson (1996) stated that using conjunctions in the initial positions of a sentence is not a language-specific feature but a common practice of EFL learners. Using conjunctions in such a position is not always an error, as some conjunctions can occur both internally and externally (Halliday & Hassan, 1976).

Nevertheless, it was later marked as a feature of informality in academic writing and hence a sign of poor writing quality of EFL learners that decreases over time as "a dearth of evidence regarding how informality changes over the years is evident." (Tocalo et al., 2022, p. 101). Chang and Swales (1999) listed ten features of informal writing, including what they termed "beginning a sentence with a conjunction or a conjunctive adverb." (p. 148). Moreover, the overuse of conjunctions, especially additives, was found as a significant feature of EFL writers of different L1 backgrounds, for instance, Turkish (Özçelik, 2023), Indonesian (Kurniati, 2019), Malaysian (Kashiha, 2022), Chinese (Liu, 2021), and Arabic (Alawerdy & Alalwi, 2022; Saeed, 2023). All of these studies confirmed the overuse hypothesis with findings generated from EFL and ESL students of various L1 backgrounds.

Previous studies have also identified other problems related to the use of conjunctions by EFL/ESL learners. For example, a recent study by Jama (2024) that investigated conjunction use among Saudi undergraduate EFL learners found several issues, such as the creation of sentence fragments, meaningless sentences, and poor punctuation when using conjunctions. Similar

studies, e.g. Hassoon (2023) and Saeed (2023), in this strand, revealed significant errors in EFL learners' conjunction use such as conjunction placement and semantic misuse. Another study i.e. (Sayah & Fatemi, 2013) found a positive transfer of the mother tongue (Persian) on the use of conjunctions and that this use is positively correlated with an advanced level. Nevertheless, they spotted many flaws in EFL learners' use of conjunctions that needed treatment such as inappropriate syntactic and semantic use of conjunctions. Such studies imply a significant effect of the Hallidayan work on cohesion in language teaching.

Although several works investigated cohesion in EFL, ESL, and EAP settings, these investigations have been limited in scope or have faced other limitations (Xuanhong, 2019). Furthermore, available research on the EFL students' use of the conjunction is not proportionate either with the importance of the issue or with the concerns it receives from teachers and curriculum designers. Subsequently, the available studies still acknowledge the poor use of conjunctions both on quantitative and qualitative bases and their tendency to use specific types of conjunctions (Zewitra et al, 2023). This indicates that further work is needed to generate findings that could help inform more effective teaching methods and appropriate use of conjunctions.

The lack of sufficient research on the relationship between conjunction use and writing quality has led to uncertainty in the findings. This issue has been a subject of debate for a long time (He, 2020). It is still not conclusively determined whether a high use of conjunctions necessarily indicates high-quality writing. Some studies have found no significant correlation between the use of cohesive devices and overall coherence (Granger & Tyson, 1996), suggesting that the presence of conjunctions may not be a definitive condition for producing well-written essays. Moreover, some research has even reported a negative correlation between conjunction use and writing quality (Al Shamalat & Abdul Ghani, 2020). However, other studies have found that the use of syntactic features, including conjunctions, does correlate with students' language proficiency (Xuanhong, 2019; Zewitra, 2023). Additionally, the appropriate use of conjunctions has been identified as a crucial factor in maintaining good narrative structure (Sayah & Fatemi, 2013) and can contribute to the overall quality of EFL learners' written texts (Jama, 2024).

As the relationship between conjunction use and writing quality remains a complex and unresolved issue, with mixed findings from various studies. The present study aimed to revisit the issue of conjunction use, aiming at enhancing research in the area as the problem of poor conjunction use is still persistent in the EFL setting (Alawerdy &

Alalwi, 2022; Hassoon, 2023; Hosseinpur & Pour, 2022; Liu, 2021). It adopts the corpus linguistic method as it is believed to provide trustful evidence based both on quantitative measures of the syntactic units and qualitative basis of the semantic criteria. Accordingly, it is envisaged that the findings from this study will offer valuable insights to inform the teaching and use of conjunctions in various local and global contexts.

METHOD
To generate its findings, the researcher followed a series of procedures, including selecting the research design, compiling the learner corpus, and extracting the reference corpus. Subsequently, the researcher processed the data using corpus analysis software and applied statistical analysis techniques to test the research hypotheses and produce the research findings. These procedures are discussed below.

Design
The present study embraces the corpus linguistic method, which has witnessed a marked rise in recent decades by using corpora as a sole data source or triangulating the corpus linguistic method with methods and data from other areas in linguistics (Egbert & Baker, 2020). It was also claimed that Systematic Functional Grammar (SFG), of which conjunction is an important aspect, has always emphasised authentic linguistic data obtained from society which makes it share a similar theoretical stance with corpus linguistics (Yang & Li, 2020). Corpora are also believed to provide more trustful evidence on language use as it is based on authentic data (Meyer, 2023) either written or spoken, not for the research, and hence reflect actual language use. Moreover, the ease of accessing corpora through words or word forms has contributed to the focus of many corpus studies on various aspects of grammatical structure (Stefanowitsch, 2020).

Corpora
The researcher utilized two corpora to inform the research findings. The first corpus was compiled for the purpose of the study, and it was named Arab Learners Academic Written English Corpus (ALAWEC). For benchmarking, an extract from the British Academic Written English (BAWE) was used.

ALAWEC corpus
The learner corpus (ALAWEC) incorporates 125314 words. The steps and techniques of building learner corpora suggested by pioneer corpus linguists (Gries, 2022, Meyer, 2023; Stefanowitsch, 2020; Yang & Li, 2020) were followed. These steps include collecting the text, annotating, storing, and conducting statistical processes. ALAWEC incorporates 100 files written by EFL college

students at Prince Sattam bin Abdulaziz University, Saudi Arabia, on different topics related to their studies. The student levels range from upper intermediate to advanced. The texts were collected from regular assignments of four courses which are Advanced Writing, Applied Linguistics, CALL, and Graduation Project. The assignments were in the form of essays and research articles. Writing Essay course assignments include essays of different types i.e., narrative, argumentative, and persuasive. They contain topics related to academic, personal and other various topics that enhance the diversity of the corpus. Assignments from the courses are in the form of essays and research reports on different applied linguistics topics. All files were anonymized and converted to .txt format using Anthony's (2020) software AntFileConverter. For cleaning purposes, the researchers used the RegEx feature of EmEditor Professional (2019) software. All noise data were deleted, including figures, special characters, non-English characters, empty lines, and formatting symbols. The output files were renamed according to the file type, writer's gender, and number. The files were then assigned to the learner corpus folder.

The reference corpus
A comparable corpus was extracted from the British Academic Written English (BAWE) corpus for reference purposes. Since the suggested corpus incorporates academic essays written by native speakers, it is believed to provide a reliable benchmark for comparing Arab EFL writing. The selected files were extracted according to specific criteria. Referring to the corpus documentation file, 100 texts with the following criteria were selected: (1) L1: English. (2) Discipline: Arts and Humanities (AH). (3) Genre family: essay; Macro type: simple assignment (4) Course: BA English and Linguistics. The criteria for selecting the reference corpus were set to make it as comparable as possible to the learner corpus. For instance, only essays written by undergraduate students were chosen, and the genre selected was the simple assignment type of essay. The nominated files were then extracted from the corpus using a Python code which extracts them based on the file name and assigns them to the reference corpus folder. Ultimately, the two corpora used for the research have the features reported in Table 1 below:

Table 1
Corpus Statistical Information

	ALAWEC	BAWE
Files	100	100
Types	7851	13006
Tokens	125314	125347
TTR*	0.06	0.10
Average	1253.1	1253.5

* Type/Token Ratio

The two corpora are comparable in terms of word count, the number of files, and average file length. However, BAWE is more lexically diverse than ALAWEC. This is represented in the large number of word types used in the BAWE and resulted in the larger TTR therein. Although the number of words in the two corpora was relatively small compared to typical corpus linguistics studies, it is considered sufficient based on the research questions. This is because the data can provide a snapshot of the use of conjunctions, which can be taken as a realistic reflection of the situation. These findings can be used as a foundation for further research.

Data Analysis
Before executing the data analysis, the researcher specified twelve conjunctions as representative of the four categories of conjunctions provided by Halliday and Hassan (1976, p. 242), see Figure 1 above. The process of selection was guided by the wordlist feature of the corpus analysis software. Wordlists provide valuable insights by identifying the most frequent words in a corpus, along with their frequency and dispersion across the corpus

files. Frequency and dispersion are considered the standard linguistic analysis techniques (Gries, 2022). This data-driven approach, grounded in the corpus itself, was believed to ensure that the research is well-informed and reflective of the actual language usage patterns present in the data and hence reflects the proficiency level of the learners in using conjunctions. For each category of conjunctions, all the examples provided by Halliday and Hassan (1976) were searched for, and the three most frequent words were chosen.

The collected data were analyzed in two phases. First, the researchers used the corpus analysis software #LancsBox (Berzina et al., 2020) to determine all instances of the selected conjunctions in the corpora. The Key Word in Context (KWIC) feature was used to compute frequency, the relative frequency (Rel. Freq.) (number of conjunctions per ten thousand tokens), and dispersion measures (the range) of each conjunction in the corpus. Relative frequency was calculated to compare the density of conjunction in the two corpora, while dispersion was used to investigate how evenly the conjunctions were distributed throughout each corpus. For this stage,

the researcher used the simple search feature of the software where the keyword was searched for by simply typing the targeted word in the software interface to attain comparable results (see Appendix A for an example of KWIC search for conjunctions). Results from this stage were meant to test the first hypothesis of the research (the overuse hypothesis).

The p-value of differences between the frequency of conjunctions in the two corpora was computed to confirm or refute this hypothesis. An independent samples t-test was used to compare the mean frequency of conjunction use between the two groups.

The t-test was selected as it is suitable for analyzing the two corpora, as it provides a p-value that allows for the acceptance or rejection of the null hypothesis. In this study, as well as in most corpus linguistics research, the p-value threshold was set to 0.05 (Gries, 2022). This means that if the p-value is equal to or greater than 5%, there is not enough evidence in the corpus to reject the null hypothesis. In other words, the difference in conjunction use between the two groups would not be considered statistically significant.

In the second phase, a qualitative analysis was conducted to test the second hypothesis of the present study. According to Chang and Swales (1999), Granger and Tyson (1996), and Lydia (2020), using conjunctions in the initial position of the sentence is considered an instance of the improper use of conjunction in academic writing. However, there are specific conjunctions which can be used in initial positions without being considered a misuse of conjunctions, for example, *however* and *therefore*. Furthermore, some writers tend to use introductory clauses to set the stage for the

independent clause. In such a situation, a conjunction can safely be used in the initial position, for instance *because*. In the case of the two temporal conjunctions: *first* and *finally*, they are more likely to occur in initial positions. There is no clear specification of conjunctions that can occur in sentence-initial positions; however, after reviewing previous works on conjunction positioning (i.e. Chang & Swales, 1999; Halliday & Hassan, 1976; Tocalo et al., 2022), the researcher specified five conjunctions to investigate their occurrence in initial positions in the two corpora. The selected conjunctions are: *and, or, but, so,* and *then* which were chosen because they were found to be the most likely to be located in the internal position connecting clauses according to the studies reviewed.

This principle formulated the second hypothesis of the present research, which addresses conjunction misuse. Accordingly, the researcher used the smart search feature of the #LancsBox software (Berzina et al. 2020) to determine all instances of initial positioning of the selected conjunctions in the corpus. Subsequently, the instances of initial conjunctions will be calculated, and the statistical significance between the frequencies will be computed to test this hypothesis. Appendix B shows an example of the output of this phase of analysis. The findings generated from the two stages of the analysis are presented and discussed below.

FINDINGS
Overall Use of Conjunctions
Findings pertaining to the global use of conjunctions by EFL learners are shown in Table 2 below.

Table 2
The Overall Use of Conjunctions by EFL Learners

Category	Conjunction	Frequency	Rel. Freq.*	Range**	Total
Additives	and	4959	359.7	100%	
	also	405	32.3	95%	**5837**
	or	473	37.7	88%	
Causal	because	302	24.1	84%	
	so	257	20.5	65%	**600**
	therefore	41	3.2	32%	
Adversatives	but	332	26.4	76%	
	yet	9	0.7	7%	**395**
	however	54	4.3	33%	
Temporal	first	178	14.2	76%	
	then	122	9.74	56%	**353**
	finally	53	4.2	39%	
Overall					**185**

* *Note*. Relative frequency per 10k
** The dispersion of the conjunction among the corpus files.

The findings revealed that the most used category of conjunctions is the additive, followed by causal, adversatives, and finally, temporal conjunctions. The learners show excessive use of

the conjunction *and* which amounts to more than half of the conjunctions used. Very low use of the conjunction *yet* is also witnessed in the data. The range represents the proportion of corpus files that

include at least one instance of the linguistic element in question (Gries, 2022). In the present study, the range of use of the conjunction is proportionate with the frequency, as the high range of use is associated with the high use of conjunctions. For example, the conjunction *(and) is* are detected in all the corpus files with a range of 100 % while the range of using *yet* is only 7 %. This implies that the data is distributed evenly and that there are few cases of excessive use of conjunction by a single writer, which may distort the data.

The results can then be compared with the native speaker's use of conjunctions which are presented in Table 3 below.

Table 3
Overall native speakers' use of conjunctions

Category	Conjunction	Frequency	Rel. Freq.	Range	Total
	and	3318	264.7	100%	
Additives	also	300	23.9	86%	**4029**
	Or	411	32.7	86%	
	but	281	22.4	86%	
Adversatives	yet	30	2.3	20%	**548**
	however,	237	18.9	76%	
	because	137	10.9	58%	
Causal	So	168	13.4	63%	**447**
	therefore	142	11.3	56%	
	first	115	9.1	53%	
Temporal	then	197	15.7	73%	**338**
	finally	26	2	19%	
Total					**5362**

Although there is a clear difference in the frequency of conjunction use between the two corpora, the details of conjunction use are comparable. Again, the conjunction *and* represents more than half of the conjunctions used and *yet* is the least used one. Furthermore, additives are the most used conjunctions, and temporal is the least used one. Adversative conjunctions, however, are not in the same order as they were used more than causal conjunctions, which was not the case in the ALAWEC corpus. The range of conjunction use is again balanced and even across the corpus files. To provide a visual illustration of the comparison, Figure 2 displays the overall usage of conjunctions in the two corpora. The figure demonstrates how the two groups employed different conjunctions, the order in which they are used, and the differences between the overall usage and the specific types of conjunctions.

Figure 2
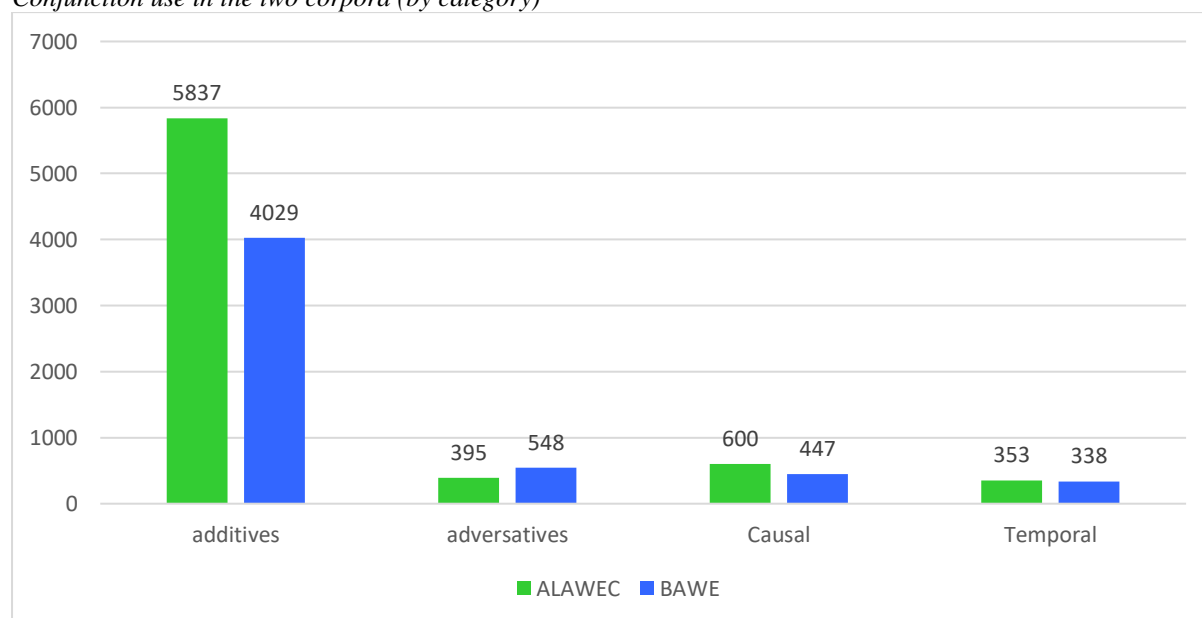*Conjunction use in the two corpora (by category)*

Figure 2 compares the frequencies of conjunctions in the ALAWEC and BAWE corpora. The ALAWEC corpus shows higher frequencies of additives and causal conjunctions, while the BAWE corpus has a higher frequency of adversatives. Both datasets exhibit similar frequencies in the temporal category. These results suggest variations in the usage of conjunctions between the two corpora.

Results for the t-test which account for the significance of the difference in frequency of conjunctions in the two corpora are presented in Table 4 below.

Table 4
Significance of difference in conjunction use in the two corpora

Conjunction	t	p	Significance
and	-8.08	0.001	Yes
also	2.9	0.004	Yes
or	1.22	0.22	No
but	0.69	0.488	No
yet	-2.2	0.029	Yes
however	-7.56	0.001	Yes
because	4.52	0.001	Yes
so	1.95	0.052	No
therefore	5.04	0.001	Yes
first	2.22	0.027	Yes
then	3.06	0.003	Yes
finally	2.4	0.017	Yes

Quality of Conjunction Use

Using conjunction in the initial sentence position was considered a criterion for improper use of conjunctions that was adopted as a marker of EFL learners' use (Granger & Tyson, 1996; Lydia, 2020; Tocalo et al., 2022). In the current study, the result pertaining to this phenomenon is presented in Table 5 below.

Table 5
Using conjunctions in the initial position in the two corpora

Conjunction	ALAWEC		BAWE		p-value	significance
	Freq	%*	Freq	%*		
And	73	1.47%	8	0.24%	0.001	Yes
Or	3	0.63%	6	1.46%	0.202	No
But	38	11.45%	39	13.88%	0.982	No
So	62	24.12%	20	11.90%	0.022	No
Then	20	16.30%	26	13.20%	0.659	No
Total	**176**	**2.45%**	**73**	**1.36%**	**0.296**	**No**

* Note. The percentage is counted out of the total hits of the same conjunction

The results show that using conjunctions in the initial position is roughly similar in the two corpora. Apart from the case of the conjunction *and*, there is no statistically significant difference in using conjunctions by the two groups. According to this finding, the second hypothesis of the research is rejected. Consequently, it can be stated that Arab EFL learners use most conjunctions in initial positions in the same way native speakers do. Implications of these results are discussed in the following subsection.

DISCUSSION

The present study investigated EFL learners' use of conjunctions in their academic writing. It utilized a corpus of 125000 words to check the density of conjunctions used in the corpus, then compared it to a reference native corpus. Further, the two corpora were analyzed to investigate initial-position conjunctions. The researcher adopted Halliday and Hassan's (1976) categorization of conjunctions to specify the targeted research items. The yielded data were used to test the two hypotheses that EFL learners misuse conjunctions in terms of both quantity and quality.

The study's findings demonstrate an overuse of conjunctions by EFL learners when compared to native speakers' use. The overall use of conjunctions in the learner corpus was around 7000 hits compared to around 5000 hits in the native corpus. Nevertheless, there are some conjunctions which are used by native speakers more than by EFL learners which are, *however, then, therefore, and yet* respectively. The results also demonstrate an excessive use of the additive conjunction *and* by the two groups representing half of the detected conjunctions in each corpus. The distribution of using conjunctions in the two corpora is normal and consistent with the frequency of the conjunctions used. For example, the most used conjunction *and* is used in all the corpora files, and the least used one

yet is used in limited files, which implies little probability of uneven distribution of conjunctions across the corpora files. This overuse of conjunction sometimes presents an improper use, especially in the case of EFL Learners. For example, the corpus incorporates statements such as:

> "*I started by watching American movies **and** try to focus on what the characters are saying **and** how their pronunciation **and** the way they're talking…*" (ess_m_16).

Although the sentence is grammatically correct, it is clear that the excessive use of ***and*** affects the overall quality of the text.

These results support previous findings relating to EFL learners' use of conjunctions which indicated overuse of this linguistic feature (Granger & Tyson, 1996; Kashiha, 2022). Also, the excessive use of the conjunction *and* and other additives was discussed and found to be a mark of both native and non-native learners (Saeed, 2023). This phenomenon is justified by the fact that *and* can act as a connector between words, clauses and sentences and hence used recurrently. Another reason for the heavy use of *and* was proposed by Sayah & Fatemi (2013), who suggested that this conjunction is "*simple*" and one of the early acquired instances of connectors which are replaced with more "*sophisticated*" (p. 140) conjunctions by increasing age or levels. Consequently, it is more likely to be observed in EFL settings.

Although there are no clear-cut specifications of *simple* and *sophisticated* conjunctions in the previous literature, the researcher believes that there is an agreement that *but* is simpler than *yet* and *however* (Sayah & Fatemi, 2013), though both indicate adversative functions. Also, using *because* is easier than using *therefore*, though both indicate causal functions. These insights are implied by comparing EFL learners' to natives' use of these conjunctions. EFL learners employed *simple* conjunctions more than native speakers and vice versa. Likely, the differences in the use of conjunctions between EFL learners and native speakers can be attributed to their respective proficiency levels. EFL learners tend to use common, high-frequency conjunctions like *and* and *but* because they are easier to learn and use. On the other hand, native speakers, with their stronger command of the language, tend to use more sophisticated conjunctions. This disparity in conjunction usage may be a result of the differences in language input and exposure. Native speakers have access to a wider range of linguistic resources, such as various media, educational materials, and everyday interactions, which allows them to acquire a more extensive repertoire of conjunctions. In contrast, EFL learners have more limited exposure to authentic, natural English, which restricts the range of conjunctions they can learn and incorporate into their language use. Further, the instructional

focus on specific conjunctions might be another possible reason for this difference. These results entail further research to establish taxonomies on types and instances of conjunction use by EFL learners along their developmental learning stages. Such taxonomies should be based on solid research findings such as corpus-based research and longitudinal studies. They will subsequently serve as invaluable resources for conjunction instruction and academic writing training programs.

Another justification for the research sample's use of conjunction may be traced to the type of conjunction used. Researchers in EFL have maintained that native speakers tend to use subordination conjunctions while native speakers use coordinating conjunctions more (Shah et al., 2021). For other researchers, such as Halliday, developmental progression for grammatical features indicates that language development starts with coordination and continues with subordination (Saricaoglu, et al., 2021). Though the present study did not analyze the conjunction use in terms of their type, the results indicate that EFL learners used the coordinating conjunctions *and, but,* and *or* more than native speakers, while native speakers used *however* and *therefore* more than nonnative learners.

Considering the categories of conjunctions used by both groups, it is interesting that both groups follow a similar order of conjunction use. Additives are the most employed, whereas temporals are the least adopted. Temporal conjunctions were found to be the least used ones by EFL Learners in some previous studies, for example, both Jama (2024) and Bhoomanee and Pothisuwan (2019) found that nonnative learners of English faced problems in using temporal conjunctions compared to native. A state that can be traced back to the nature of this type of conjunctions which entails mastery of subordination which is known to be acquired at advanced stages of learning.

The rank of adversative conjunctions is, nevertheless, different from the current study's findings; it was the second for native speakers but the third for EFL learners. This difference was also found by Hosseinpur and Pour (2022), who pointed out that EFL learners underused adversative conjunctions. However, the revealed differences between using causal and adversative conjunctions between the two corpora are trivial when considering the overall use of conjunctions by each group, as presented in Figure 2 above. These results hold strong implications for designing learning materials of conjunctions based on the order of acquisition of each category of conjunctions.

As far as the quality of conjunctions used in the two corpora is concerned, the results revealed that there are no significant differences in using conjunctions in the initial positions of sentences between the two studied corpora. Overall, EFL

learners placed 2.45 % of the used conjunctions in the initial positions, while native speakers used 1.36 % of the conjunctions in such a position. The only significant difference is pertaining to positioning the conjunction (*and),* which was placed by EFL learners 73 times in the corpus corresponding to 8 cases by the native speakers. Although it is acceptable to use a range of conjunctions in sentence-initial positions, using (*and)* to start a sentence has been considered an improper use of formal academic writing techniques. However, the ALAWEC corpus contains several instances of this use such as "most speakers of the Arabic language fall into the wrong pronunciation, such as (v and f). **And** there are letters that are pronounced and not written" (gp_m_11) and "computer-assisted language learning is becoming more and more popular. **But** expanding its engagement in education is a protracted" (ess_m_1). This finding implies that EFL learners lack sufficient knowledge of the academic writing conventions regarding the syntactic requirements of using conjunctions.

Results pertaining to the initial positioning of conjunctions contradict previous findings that suggest this phenomenon to be a marker of EFL writing (Granger & Tyson, 1996; Xia, 2020). The reason for such contradiction is probably that such views have changed over time as informality in academic writing has recently become a trend (Hyland & Jiang, 2017; Tocalo et al., 2022). A probable reason for the shift to informality may be the elevated use of social media by EFL learners. Chatting in cyberspace is marked with conceptual orality that makes chatters tend to use more informal language in their chat as noted by Lydia (2020). As language is dynamic and affected by the surrounding socio-cultural changes, these facts should be taken into consideration. Subsequently, this result entails updating views that evaluate EFL learners' use of conjunctions in initial positions. After all, starting a sentence with a conjunction is no longer a serious error, as it has been depicted by old theories on academic techniques that view academic writing as conservative and impersonal (Matsuda & Nouri, 2020). Accordingly, instructional materials on conjunction use should also focus on the syntactic placement of conjunctions to foster using conjunctions correctly and conventionally, of these, explicit instruction techniques on conjunction can improve producing more cohesive texts as proved by Yamaai (2023) and Alawerdy and Alalwi (2022). On the other hand, further research is required to explore the initial positioning of conjunctions by native and foreign English language learners. The generated findings of such research can be used to develop EFL academic writing programs and update the existing ones.

The generalizability of the present study's findings is limited by the fact that it is based on the output of EFL learners with a single L1 background,

that is Arabic. Considering the effect of L1 transfer on using conjunctions which is recognized by several research studies (e.g. Alawerdy & Alalwi, 2022; Rokhaniyah et al., 2022; Sayah & Fatemi, 2013), it cannot be generalized that these findings are globally applicable. To account for such a limitation, further research on EFL learners with different L1 backgrounds is required. Another limitation to acknowledge is that it limited the qualitative analysis of conjunction use to the initial positioning of selected conjunctions. All the necessary measures to select the conjunctions were taken; however, it was extremely difficult to decisively determine which conjunction is proper or improper to use in such a position as it is usually context-dependent. Further, other criteria to decide on the quality of conjunction use should be adopted to better reflect the reality of EFL learners' use of conjunctions. Finally, strict criteria were followed to make the learner and reference corpora as representative (in terms of size) and comparable as possible. However, as is often the case with corpus studies, a perfect match between the two corpora could not be guaranteed. This potential mismatch should be considered as a factor that may affect the study's results.

While the current analysis attempts to provide insights into the patterns of conjunction usage between EFL and native learners, it is important to acknowledge the challenge of conducting a context-dependent evaluation of conjunction quality. The appropriateness and effectiveness of conjunctions are often highly dependent on the specific communicative context in which they are employed. Factors such as coherence, rhetorical function, and overall communicative impact can play a crucial role in determining the quality of conjunction use. Future research may benefit from the development of more robust, context-sensitive frameworks for analyzing conjunction usage.

CONCLUSION

Using conjunctions fosters academic writing and enhances communication. Due to the observable misuse committed of this feature, particularly by EFL learners, researchers are adopting different methods to explore the issue and generate implications for treating it. Previous findings suggest that EFL learners typically overuse conjunctions and misuse them in other ways, such as placing conjunctions at the beginning of sentences where it is often deemed inappropriate. Despite long-standing discussions on these issues, recent studies still acknowledge the poor use of conjunctions by EFL learners. Moreover, the availability of relevant academic writing programs and learning materials suggests that the problem is solved, which is contrary to reality. This discrepancy may be due to the inherent difficulty in

applying theoretical knowledge to practical language use. Effectively applying conjunction usage requires additional supporting factors, such as greater exposure to the language and consideration of developmental acquisition patterns. Accordingly, the present study revisited the issue by adopting the corpus linguistic method to compare the use of EFL learners of conjunctions to native ones. It adopted Halliday and Hassan's (1976) categorization of conjunctions as a benchmark to test its hypotheses.

The study results revealed that EFL learners overuse conjunctions, especially additive ones. Nevertheless, they follow a roughly similar order of use of types of conjunctions. On the other hand, no statistically significant differences were found between the syntactic placement of selected conjunctions from that of the native speakers. The latter findings could be traced back to the widespread use of informality features in academic writing. Simultaneously, it entails a reconsideration of EFL learners' use of conjunctions based on new findings and conventions in writing. Consequently, an update of the analysis methods of conjunctions use is required..New methods should take into account the native use of conjunctions, as revealed by corpus-based research and the evolving conventions of academic writing, particularly the inclusion of informality features. Further research is required to enhance the present study's findings to proportionate with the importance of the topic. These research studies should utilize larger corpora of EFL learners' writing with diverse L1 backgrounds to avoid potential limitations of the current study. The results of the present study and proposed further research may lead to better implementation of conjunction instruction programs which teach EFL learners academic writing skills at the phrasal, clausal and sentential levels. The proposed program may first determine simple and sophisticated conjunctions and then design learning materials that aim at training students to employ them in a structured approach. Moreover, the focus on the targeted conjunctions should be placed both on semantic and syntactic contexts to enhance learners' competence in acquiring and using conjunctions rather than focusing on their semantic connotations. These programs could also involve exposure to a wider range of language use contexts, explicit instruction on contextual usage, and scaffolding techniques that support the transfer of theoretical knowledge to real-world language production. Accomplishing this, the researcher believes, will lead to more cohesive writing by EFL learners.

REFERENCES

Al Shamalat, R., & Abdul Ghani, C. A. (2020). The effect of using conjunction as cohesive device on the undergraduates' quality of writing in argumentative essays of Jordanian EFL learners. *Arab World English Journal (AWEJ),* (6), 290-301. https://doi.org/10.24093/awej/call6.19

Alawerdy, A., & Alalwi, F. (2022). Enhancing English as a foreign language university students' writing through explicit instruction of conjunctions as cohesive devices: An experimental study. *Front. Psychology, 13*, 1-10. https://doi.org/10.3389/fpsyg.2022.1053310

Anthony, L. (2020). *AntFileConverter* (version 1.2.1) [Computer Software]. Waseda University. http://www.antlab.sci.waseda.ac.jp/

Berzina, V., Weill-Tessier, P., & McEnery, A. (2020). *#LancsBox* (v. 6.0) [Computer software]. http://corpora.lancs.ac.uk/lancsbox

Bhoomanee, C., & Pothisuwan, P. (2019). A corpus-based study: High-frequency use of temporal connectors in written text between native speakers and Thai learners. *The Academic Journal: Faculty of Humanities and Social Sciences Nakhonsawan Rajabhat University, 6*(2), 1-14. https://so05.tci-thaijo.org/index.php/hssnsru/article/view/240716

Chang, Y., & Swales, J. (1999). Informal elements in English academic writing: Threats or opportunities for advanced non-native speakers. In C. Candlin, & K. Hyland, *Writing: Texts, processes and practices* (pp. 145-167). Longman.

Egbert, J., & Baker, P. (2020). Introduction. In P. Baker, & J. Egbert (Eds.), *Using Corpus Methods to triangulate linguistic analysis* (pp. 1-21). Routledge.

Emurasoft_Inc. (2019). *EmEditor Professional* (Version 19.3.2) [Computer Software]. Filepuma. https://www.filepuma.com/download/emeditor_professional_64bit_19.3.2-23779/

Granger, S., & Tyson, S. (1996). Connector usage in the English essay writing of native and non-native EFL speakers of English. *World Englishes, 15*(1), 17-27. https://doi.org/10.1111/j.1467-971X.1996.tb00089.x

Gries, S. (2022). Toward more careful Corpus statistics: Uncertainty estimates for frequencies, dispersions, association measures,

and more. *Research Methods in Applied Linguistics*, 1-19. https://doi.org/10.1016/j.rmal.2021.100002

Halliday, M. K., & Hassan, R. (1976). *Cohesion in English.* Longman Group Limited.

Hassoon, N. H. (2023). Conjunctions in Iraqi media EFL classroom. *Social Science Journal, 13*(1), 2943-2956.

He, Z. (2020). Cohesion in academic writing: A comparison of essays in English written by L1 and L2 University students. *Theory and Practice in Language Studies, 10*(7), 761-770. http://dx.doi.org/10.17507/tpls.1007.06

Hosseinpur, R., & Pour, H. (2022). Adversative connectors use in EFL and native students' writing: A contrastive analysis. *The Electronic Journal for English as a Second Language, 26*(1), 1-18. https://doi.org/10.55593/ej.26101a3

Hyland, K., & Jiang, F. (2017). Is academic writing becoming more informal? *English for Specific Purposes, 45*, 40-51. https://doi.org/10.1016/j.esp.2016.09.001

Jama, I. (2024). Zone of proximal development: Investigating the most usage conjunctions and the common issues written by EFL students at paragraph levels. *World Journal of English Language, 14*(2), 109-117. https://doi.org/10.5430/wjel.v14n2p109

Kashiha, H. (2022). An investigation of the use of cohesive devices in ESL students' essay writing. *Journal of Education and Practice, 13*(18), 11-19. http://doi.org/10.7176/JEP/13-18-02

Kurniati, R. (2019). Conjunctions in Indonesian undergraduate thesis abstracts. *Etnolingual, 3*(1), 27-41. https://doi.org/10.20473/etno.v3i1.12059

Liu, M. (2021). References and conjunctions as cohesive devices in Chinese undergraduate EFL students' argumentative essays. *The Journal of Asia TEFL, 18*(4), 1144-1160. http://doi.org/10.18823/asiatefl.2021.18.4.5.1144

Lydia, I. (2020). *EFL students' deviation from academic to informal writing: The case of third-year students in the department of English at Mouloud Mammeri University of Tizi-ouzou* [Unpublished Master Thesis]. Tizi-Ouzou, Algeria: Mouloud Mammeri University of Tizi-Ouzou.

Matsuda, P. K., & Nouri, M. (2020). Informal writing and language learning. In M. Dressman, & R. Randall, *The handbook of informal language learning* (pp. 75-84). John Wiley & Sons Ltd.

Meyer, C. (2023). *English corpus linguistics: An introduction.* Cambridge University Press. http://doi.org/10.1017/9781107298026

Özçelik, N. (2023). A corpus-based study of conjunction use by the Turkish EFL learners. *International Journal of Educational Research Review, 8*(3), 475-483. https://orcid.org/0000-0001-8772-0075

Rokhaniyah, H., Adila, W., & Yulianti, D. (2022). Cohesive devices used in argumentative essays by undergraduate EFL learners. *Langkawi Journal of The Association for Arabic and English*, 32-44. http://doi.org/10.31332/lkw.v0i0.3491

Saeed, L. (2023). Excessive use of additive conjunctions in the writing of EFL students at tertiary level with ref. to the conjunction and. *International Journal of Contemporary Applied Researches*, *10*(2), 43-74

Saricaoglu, A., Bilki, Z., & Plakans, L. (2021). Syntactic complexity in learner-generated research paper introductions: Rhetorical functions and level of move/step realization. *Journal of English for Academic Purposes*, 53, 1-11. https://doi.org/10.1016/j.jeap.2021.101037

Sayah, L., & Fatemi, A. H. (2013). The role of conjunctions in EFL learners' narrative development. *International Journal of Education and Research, 1*(11), 133-146. http://www.ijsac.net/sites/default/files/2021-01/V.5.%20N.12-1-converted.pdf

Shah, M., Gasim, H., & Sibtain, M. (2021). Locus of conjunctions in academic writing: A corpus-driven approach to developing writing among EFL learners. *Pakistan Languages and Humanities Review*, *5*(2)432-443. http://doi.org/10.47205/plhr.2021(5-II)1.35

Stefanowitsch, A. (2020). *Corpus Linguistics: A guide to methodology.* Language Science Press.

Tocalo, A., Racman, S., Guiamelon, K., & Mama, B. (2022). Is Filipino ESL academic writing becoming more informal? *PASAA Journal, 64*, 99-125.

Unubi, A. (2016). Conjunctions in English: Meaning, types and use. *International Journal of Social Science and Humanities Research, 4*(3), 202-213

Xia, G. (2020). A comparable-corpus-based study of informal features in academic writing by English and Chinese scholars across disciplines. *Ibérica, 39*, 119-140. https://doi.org/10.17398/2340-2784.39.119

Xuanhong, G. (2019). *A study of the use of lexical cohesion in Chinese postgraduate writing at a UK university*. [Doctoral dissertation, Durham University]. Durham E-Theses Online. http://etheses.dur.ac.uk/13353/

Yang, B., & Li, W. (2020). Systemic functional linguistics: Interaction, methodology, and applicability. In B. Yang, & W. Li (Eds.), *Corpus-based approaches to grammar, media,*

and health discourse: Systemic functional and other perspectives* (pp. 1-14). Springer. https://doi.org/10.1007/978-981-15-4771-3

Yamaai, J. (2023). *The effects of explicit instruction and corrective feedback on lexis and cohesion with EFL learners* (Publication No. 30812060) [Doctoral dissertation, Temple Univeristy]. ProQuest Dissertations & Theses Global. https://www-proquest-com.sdl.idm.oclc.org/pqdtglobal1/dissertations-theses/effects-explicit-instruction-corrective-feedback/docview/2912214190/sem-2?accountid=142908

Zewitra, Bakhti, K., & Rama, M. (2023). EFL learners' tendency in using conjunctions in writings: Evidence from English Diploma-III. *Professional Journal of English Education, 6*(4), 705-713.

APPENDICES
Appendix A
Search results for the additive conjunctions (and) in the two corpora

Appendix B

Smart search results for the adversative conjunction (but) in initial positions in the two corpora

#LancsBox 6.0

| KWIC | GraphColl | Whelk | Words | Ngrams | Text | Wizard |

Corpora | KWIC: /But/, /But/ ✕

Search

Search /But/ **Occurrences** 38 (3.03) Texts 27/100 ▼ **Corpus** ALAEC ▼ **Context** 7 ▼ **Display Text**

Index	File	Left	Node	Right
1	ess_f (1).txt	village and buys me lots of sweets.	But	my father's working conditions made us move
2	ess_f (5).txt	because I use these skills every day.	But	the skill I am not very good
3	ess_f (8).txt	learning is becoming more and more popular.	But	expanding its engagement in education is a
4	ess_m (1).txt	learning is becoming more and more popular.	But	expanding its engagement in education is a
5	ess_m (10).t	graduate. I don't remember something my childhood,	But	I love to watching my clips. IN
6	ess_m (12).t	lot of attention in the Arab world.	But	many Arabs often mispronounce it when learning
7	ess_m (12).t	the conversion of spoken words into text.	But	this research will help to solve the
8	ess_m (13).t	on the grounds, running and having fun.	But	suddenly a dog appeared from the middle
9	ess_m (13).t	people, me, my brothers and my sister.	But	the dog quickly sped towards me. Suddenly
10	ess_m (13).t	am greatfull for everything in my life.	But	I don't like the crowded places, My
11	ess_m (13).t	alkharj city when I was a kid	,But	in my teens we moved to Riyadh
12	ess_m (14).t	village and buys me lots of sweets.	But	my father's working conditions made us move
13	ess_m (16).t	school, I learned nothing but the alphabet.	But	in my middle school I started to
14	ess_m (16).t	understand most of the words without translation.	But	I still don't know the rules of
15	ess_m (18).t	at school like grammar and simple vocabulary.	But	my method will help you if you
16	ess_m (19).t	I think it is the easiest skill.	But	I am not very good at writing
17	ess_m (20).t	because I use these skills every day.	But	the skill I am not very good

▼▼▼

Search /But/ **Occurrences** 39 (3.11) Texts 22/100 ▼ **Corpus** BAWE.mini ▼ **Context** 7 ▼ **Display Text**

Index	File	Left	Node	Right
1	0139e.txt	of shared ethnical, cultural and religious heritage.	But	these factors cannot alone explain how "a
2	0139i.txt	in Chinese Economic History", Economic History Review.	But	do solutions to these institutional problems exist?
3	0142a.txt	restricted to Mrs. Donoghue's friend (buyer) only.	But,	her friend did not suffer much. So,
4	0152c.txt	at the expense of long-term business development.	But	why, then, were they internationally boycotted for
5	0153d.txt	addressing the flaws present in the dimensions.	But	even then, dimensions and theories will always
6	0153e.txt	matters and may not be worth considered.	But	I realised that this view is biased
7	0153e.txt	as some I think can be trivial.	But	to say that CSR is not important
8	0155c.txt	obtain finance with less Cost of Capital.	But	if discount rates alter over life of
9	0179e.txt	for the "preservation of his own nature".	But	he may also seek power and glory
10	0179e.txt	rationality and logic, the codification of law.	But	most importantly Hobbes gives us the language
11	0179e.txt	pertinent as ever in today's contemporary world.	But	crucially Hobbes fails to create the citizen,
12	0179f.txt	under the guises of ethics and philosophy.	But	it was not until post-revolution Europe that
13	0179f.txt	them, no citizen had yet been created.	But	this brings us back to the endless
14	0179f.txt	vague beginnings of sociology could be associated.	But	as with Aristotle the concept of science
15	0179f.txt	probably translate to current rational choice theory!	But	Comte clearly wanted to add prestige to
16	0179g.txt	of generalised importance in the capitalist societies...	But..	it seems preferable to treat such conflicts
17	0179g.txt	required to gain access to a mortgage.	But	this is representative of structuralised bias against

Filtering complete