

Discourse Markers in the Academic Writing of Arab Students of English: A Corpus-based Approach*

Abeer Q. Taweel
Al-Hussein Bin Talal University, Ma'an, Jordan

Abstract—This study aims to shed light on the discourse markers used in the academic writing of Arab students of English as a second language within the framework of corpus linguistics. By so doing, an attempt will be made to examine the use of the discourse marker expressing attitude, sequence, cause and result, addition, and comparing and contrasting. For comparison purposes, similar-sized authentic corpus will be used to examine the learners' use, overuse, and underuse of the target markers. Moreover, the study will provide a detailed account of the possible reasons contributing to the disparity between the two corpora in terms of the use of the target markers. Results show that learners use more discourse markers than native speakers. While this is a general tendency, it still remains feasible to attribute the disparity between the two corpora to learners L1 influence where some of the overused markers spring out naturally and smoothly as they have rhetorical functions in learners' native tongue.

Index Terms—discourse, markers, learners, corpus, misuse

I. INTRODUCTION AND LITERATURE REVIEW

Discourse Analysis is a multi-faceted field in which linguists can delve beyond what is read or heard. This field studies so many aspects and forms of communication. By and large, discourse analysis studies two facets, spoken and written forms. One aspect of discourse analysis is discourse markers (DM, henceforth) which are used in oral or written forms. DMs are inseparable from any communication. Many linguists have studied DMs in different terminologies, but still working under the umbrella of expression DMs. For example, *discourse connectives* was the term used by Blakemore (1987) instead of DMs. Redeker (1990), on contrast, refers to them as discourse operators. Generally speaking, DMs make an integral part of everyday speech and everyday writing practice; they are also important words and phrases that are used in any kind of communication, and generally cannot be avoided. Some examples of DMs are “*anyway, alright, well, so, okay, you know, I mean ... etc.*”. So, what does a DM mean? To have a clear idea, let's have a look at different definitions, which are discussed from dictionaries and viewpoints of a number of linguists. On the one hand, many dictionaries, in general, define DMs as words and phrases used in speaking and writing to 'signpost' discourse (BBC). DMs primarily include a set of words or phrases such as *right, as I say, to begin with, okay, right, anyway*. We use DMs to connect, organize and manage what we say or write or to express attitude “Cambridge Dictionary”. Linguists, on the other hand, have their own point of views in defining DMs. Fraser (1990) defines DMs as linguistic elements that are predominantly used in oral conversations to relate units of discourse to each other. Schiffrin also (1987, p. 31) defines DMs as “sequentially dependent elements which bracket units of talk that signal relationships between immediately adjacent units of talk, and which have thus a coherence building function on a local coherence level.” Swan (2005) defines discourse markers as words and expressions that we use to portray the structure of our discourse. They serve the purpose of connecting what we are saying, what we have said and what will be said.

Having a look at some definitions, it can be noticed that DMs play a crucial role in communication, but they are confined to spoken communication. Therefore, in this study, the focus will be on discourse markers used in academic writing of Arab students of English as a second language within the framework of corpus linguistics. The process of writing is not an easy task for students. It is a demanding process that needs exercised efforts on the part of the writer. Cohesion and coherence are the core of any communication - writing as an aspect of communication. To achieve such an arduous process – achieving cohesion and coherence, learners of writing, especially academic writing, may resort to the use of phrases or words that make their writing look better or well-connected in some way. This connection (well-formedness) can be fulfilled by the use of DMs. One thing to say here is that the importance of these DMs is using them in academic writing for a communicative reason. They are important in clarifying the communicative intentions of the interlocutors (Schiffrin, 1987), to say, the writer and the reader. Besides, following Schiffrin (1987), DMs show the sequence of events and the relation between the elements of communication.

* This study was carried out during a sabbatical leave granted by Al-Hussein Bin Talal University, Jordan

Wei (2013) points that “DMs are advantageous in which they allow speakers or writers to make their presence felt by pausing. DMs also provide guidance to the audience as to how the text is structured so that they can easily decode and infer what the writer has encoded and referred. Appropriately used DMs in writing helps provide arrangements particularly in introductions and conclusions to academic writing”.

The advantages of using DMs are of a great number and multitude. They help the readers understand where to stop and where to continue reading in any piece of writing. They can also help them predict and as a result they can speed up and fully digest the reading at their hands. A plethora of examples of these pieces of writing are available in corpora, which give hints on the use of the DMs in different texts, academic writing of Arab students in particular. DMs also help the students learn how to organise their writing in deciding the beginning and the ending of their work. Perhaps the discussion of the advantages of DMs should incorporate some observation on the functions that such markers might fulfil in discourse.

DMs perform some functions in language depending on their position in text and on the type of the text. Some DMs, for example, show attitude (attitude of the writer) such as e.g. *of course*, *surprisingly*. In addition, DMs show indirectness and vagueness of the writer's views (Hinkel, 1997), this means that academic writers may resort to DMs as a cover of his/ her imprecision in writing. One more function of DMs is addition, e.g. *moreover*, *in addition*, *additionally*, *further*, *further to this*, *also*, *besides*, *what is more etc...*, such DMs are used to add information and give extra knowledge. Comparing and contrasting are also deemed as a part of the functions that DMs perform when using expression like *however*, *on the other hand*, *in contrast*, *yet*. Sequence again can be represented by DMs such as using *oh well*, *but then*, etc (Koops, C. and Lohmann, A., 2013). Schiffrin (1987) defined the sequence function as “*sequentially dependent elements which bracket units of talk.*” (p. 31). Bracketing units of talk means giving sequence of the events depending on the event occurrence and progression in the process of writing.

To take the functions from another angle, let us take a look at some works and classifications of some linguists. DMs perform many functions according to discourse in general, and in academic writing in particular (Btoosh and Taweel, 2011). Muller (2005, p. 9) identifies the functions of DMs as follows.

- a. Establish discourse.
- b. Indicate a boundary in discourse (shift/partial shift in topic).
- c. Introduce a response or a reaction.
- d. Function as fillers or delaying tactics.
- e. Help the speaker in holding the floor.
- f. Influence interaction or sharing between speaker-hearer.
- g. Bracket the discourse either cataphorically or anaphorically.
- h. Mark either foregrounded or backgrounded information

Croucher (2004, p. 40), DMs fulfil formal and informal functions as follows. The formal functions of DMs are:

- a. To indicate a turn in conversation (you know, well)
- b. To identify a digression from the topic under discussion (oh, by the way)
- c. To share a speaker's attitude/sentiment (like)
- d. To frame general conversation

In addition to the formal functions, DMs perform several informal functions, including the following.

- a. To fill pauses in conversation
- b. To act as nervous glitches in speech
- c. To act as a part of our collective lexicon

Cohesion constitutes another significant role that DMs play in discourse. Iseni et al. (2013) assume that there is a strong connection between DMs and coherence and cohesion. Halliday and Hasan (1976) also assume that coherence is a means that makes the sentences semantically well-formed which means that DMs play a vital role in the structure of the text.

Halliday and Hassan (1976) identified four grammatical strategies (cohesive devices) applied to achieve relatedness in communication:

- **Conjunctions:** connect sentences meaningfully “*however, but, though, although... etc.*”. This means that such discourse markers contribute to the consistency of the written text.
- **Substitution:** used to reduce the use of some expressions or structures and reduce redundancy. Such as “*one as the second one, another one ... etc*” (Iseni et al, 2013).
- **Ellipses:** omitting some words from the text to reduce the use of redundant words or expressions.
- **Reference:** is a cohesive device that refers to an item in a sentence and this item may be mentioned early.

According to Iseni et al. (2013), reference in discourse can be made by means of pronouns.

Pronouns are the main resource people have for referring, reference can be subcategorized as exophoric which refers to something outside the text and endophoric which is also subcategorized as follow: anaphoric which points backwards, and cataphoric which refers to forwards in a text.

In spite of the significant role played by these categories, special attention is to be paid to the first strategy, namely, conjunctions, since they can be found in corpus in great numbers. Considering the use of conjunctions in academic writing, it is worth mentioning, here, that the use of conjunctions is meaningful to show the different functions of DMs.

In addition, the use of DMs helps give the text more consistency, but this does not always happen when it comes to learners. A close look at the learners corpus used here shows that DMs are sometimes used erroneously. This may happen for some reason such as the weakness points in grammar or the lack of competence which makes the cohesion of their writing, to some extent, a defect. As a result, the student may resort to the use of DMs to avoid discovering their errors and inability of connecting sentences grammatically to give meaningful sentences or the inability to use the appropriate lexical words which may change the whole meaning. The discussion of DMs in the academic writing of Arab students of English brings a new perspective on the notion of genre patterning.

The notion of genre is very decisive when it comes to the field of discourse analysis. Innumerable studies have engaged in the definition of genre. Since its impact on studies and results, it is very important to define and study the genre in which the DMs may occur. Hyland (2013) defines genre as typified acts of communication.

“Genre analysts set out to offer descriptions of "typified acts of communication" based on the form and purposes of texts. Basically, genres are kinds of broad rhetorical templates that writers draw on to respond to repeated situations; users see certain language choices as representing effective ways of getting things done in familiar contexts.”

These definitions show that genre as a rhetorical template that the writer must choose when the process of writing starts depending on the topic and the idea which is in question. The learner usually has many fields of writing out of which he/ she chooses and works on depending on the characteristics of the field - in other words, the writer can deeply delve throughout the different genres depending on the idea and the topic (subject) such as legalese, journalese, medicine culture, politics, religion, academia ... etc. The learner should also write essays, laboratory reports, case-studies, book reviews, reflective diaries, posters, research proposals. The importance of genre stems from figuring out the characteristics of the type of the text. Therefore, this study will shed light on academic writing as a genre that most, if not all, ESL Arab students must engage themselves in as a major topic that is studied in universities and departments of English basically. In conclusion, genre is a basic notion that the learner of any language must know before starting writing. As a result, the learner should know what DMs suit the type of writing he/ she is doing e.g. formal or informal DMs.

The notion of genre again can be the same as the notion of patterning. This belief comes from the idea that the words genre and pattern mean, in some way, “type/ template”. As noticed in the above definitions by Hyland (2013) and Miller (1984), both of the scholars agreed that genre is a type/ template. Accordingly, academic writing is the genre and the pattern that this study will shed light on. To be more specific, it focuses on the use of DMs in academic writing by ESL Arab students.

The novelty of the current study stems from the analysis of the samples taken from Arab students of natural academic writing work without any intervention from others; the number of occurrences (quantitative); and the objectivity (qualitative) of using DMs by Arab students in academic writing.

II. METHODOLOGY

A. Data Collection

The data in this study have been gained by an access to the Interlanguage Corpus of Arab students of English language. The Interlanguage Corpus consists of roughly 100.000 words of argumentative, evaluative and narrative essays written by university students majoring in English language and literature at five private and public Jordanian universities. Both timed and untimed methods of data collection were used during the course of corpus compilation.

The data representing the native speakers’ writings, on the other hand, have been accessed to from the Louvain Corpus of Native English Essays (LOCNESS). This corpus consists of approximately 300.000 words; the essays in LOCNESS were collected mainly from the writings of American and British native speakers in different universities. Yet, the samples used for the purpose of this study, are a bit equal in terms of the number of tokens (100.000 each). The sample of LOCNESS was given a priority because it was more representative than that in the Interlanguage. This can be said because of the similarity, at first place, and the diversity, in second place, in the topics between the native speakers and the learners’ corpora. It is highly imperative and important to echo that the two corpora are practically comparable in size despite of the differences in the overall number of essays analysed. The data have been prepared for the lexical and statistical analyses benefiting from the tools:

- Platform: Windows 13
- Concordancer and Wordlist: *WordSmith* suite of Tools

B. Data Analysis

The two corpora have been examined with regard to the following aspects.

- (i) The use of the target features in terms of frequency count and semantic functions.
- (ii) The possible reasons for the disparity between the two corpora in terms of the frequency count of the markers under investigation.

III. RESULTS

This section reports on the findings of the study. The findings were presented and analyzed in accordance with the overall frequency of the discourse categories and the individual markers.

A. Frequency of DMs Used

B. Overall Frequency

The figures given in Table (1) relate to the number of instances of occurrence of each category. These figures provide general tendencies and differences between learners and native speakers in terms of the use of the target categories.

TABLE 1.
FREQUENCY COUNT OF DISCOURSE CATEGORIES IN LEARNER AND REFERENCE CORPORA

Type	LC	RC
Additives: and, furthermore, moreover, in addition, additionally, similarly, that is, in other words, for example, also, too, on top of that	3894	3170
Adversatives but, yet, though, although, however, while, on the other hand, on the contrary, nevertheless, instead, nonetheless	628	880
Causatives So, because (of), as, since, hence, therefore, as a result, consequently, otherwise, in view of, that's why, thus	1334	1577
Sequencers and (then), after, later, as long as, until, after that, at the same time, meanwhile, first, second, firstly, secondly, first of all, to begin with, next, finally, to sum up	731	359
Attitude Markers obviously, preferably, (not) surprisingly, clearly, undoubtedly, miraculously, (un)fortunately, of course, predictably, regrettably	30	70
Total	6617	6056

Though the overall frequency of the discourse categories in the learner corpus is higher than that in the reference corpus, the gap between the two is still statistically insignificant. However, accounting for the actual gap between the two corpora requires looking onto the frequency counts of the individual markers. Tables (2), (3), (4), (5) and (6) present the frequency counts of the individual discourse markers in both corpora.

C. Frequency of Individual DMs

This section sheds light on the individual discourse markers as categorised above. Accordingly, four categories have been set for analysis below; these include additives, adversatives, causatives and sequencers.

1. Additives

TABLE 2.
FREQUENCY COUNT OF ADDITIVES IN LEARNER AND REFERENCE CORPORA

DM	LC	RC
and	3519	2723
furthermore	1	6
moreover	17	5
in addition	37	7
additionally	0	0
similarly	0	5
that is	45	49
in other words	1	3
for example	53	55
also	189	246
too	32	71
TOTAL	3894	3170

The huge gap between the two corpora in terms of the frequency count (3894 vs. 3170) which comes in favour of the learner corpus is a bit not surprising. The figures above show that the coordinating conjunction 'and' is responsible most for the disparity between the two corpora. The overuse of 'and' in the learner corpus is not surprising for two reasons. First, previous research findings show that learners (irrespective of their L1 use more 'ands' than native speakers. Second, native speakers of Arabic are likely to use more 'ands' due to their L1 influence. 'And' in Arabic is as a sentence opener in addition to its primary use as a coordinating conjunction. Also, Arabic uses 'and' where English uses comas to separate series of similar items.

2. Adversatives

TABLE 3.
FREQUENCY COUNT OF ADVERSATIVES IN LEARNER AND REFERENCE CORPORA

DM	LC	RC
but	464	399
yet	4	48
though	11	48
although	37	85
however	22	188
while	52	60
on the other hand	18	18
on the contrary	5	3
nevertheless	1	4
nonetheless	1	0
instead	13	27
TOTAL	628	880

3. Causatives

TABLE 4.
FREQUENCY COUNT OF CAUSATIVES IN LEARNER AND REFERENCE CORPORA

DM	LC	RC
so	369	208
because (of)	466	220
as	412	885
since	36	66
hence	1	8
therefore	19	97
as a result	17	16
consequently	3	5
otherwise	2	7
that's why	3	2
thus	6	63
TOTAL	1334	1577

The gap between the two corpora in terms of the adversatives and causatives count should not be surprising. Unlike additives, adversatives and causatives reflect and require a profound understanding of the text rhetoric. One possible explanation for using less adversatives and causatives by learners is their lack of awareness of academic writing conventions. A careful look at the frequency count above and the learners' actual use of conjunctives show that there are numerous complex contextual relations that are not visible for learners. Another possible explanation for the learners' limited use of conjunctives compared to native speakers can be attributed to learners' L1 itself. Arabic discourse does not implement adversative and causative conjunctions in the same way English does. Such contrast relations in the text are expressed more often via long phrases and sentences rather than one word conjunction.

4. Sequencers

TABLE 5.
FREQUENCY COUNT OF SEQUENCERS IN LEARNER AND REFERENCE CORPORA

DM	LC	RC
and (then)	14	16
after	135	87
later	10	18
as long as	0	7
until	36	31
after that	87	2
at the same time	8	9
meanwhile	1	1
first, firstly, first of all	158	109
second, secondly	94	27
third, thirdly	42	15
next	44	16
finally	102	21
TOTAL	731	359

The explanation of the above figures is simply a matter of L1 influence. The highly frequent use of numeration and listing in Arabic discourse is one possible reason. Like additives, sequencers show clearer relations in the context.

5. Attitude Markers

TABLE 6.
FREQUENCY COUNT OF ATTITUDE MARKERS IN LEARNER AND REFERENCE CORPORA

DM	LC	RC
obviously	1	13
preferably	0	0
(not) surprisingly	3	2
clearly	6	20
undoubtedly	0	5
fortunately	3	2
of course	17	28
regrettably	0	0
TOTAL	30	70

Discourse markers that express attitude simply show how the writers present themselves in the text. Like adversatives and contrastive devices, attitude conjunctions require a deep understanding of the discourse cohesion, rather than just grammatical structure. Xu (2001) found that Chinese undergraduates use more additives in their writings as their competence in English improves. However, they use less attitude markers though. Xu attributes this to their awareness of English writing rules and how they differ from their L1 norms. The more non-native speakers comprehend the complexity of L2 discourse rules, the less they tend to assert themselves in their written English texts. Native speakers of Arabic avoid asserting themselves as writers or showing their opinions and attitudes.

IV. CONCLUSION

This study has examined the use of discourse markers in the academic writing of Arab students of English and English-native speakers. To be more specific, five broad categories of discourse markers express attitude, sequence, cause and result, addition, and comparing and contrasting. The primary purposes of this study were two-fold. The first was to gain insights onto the frequency counts of the target discourse markers in both corpora. The second purpose was to examine the similarities and differences between the two similar-sized corpora in terms frequency count, uses and the possible factors for divergence. Findings have shown that learners use more discourse markers than native speakers. Also, the results indicate that the overuse and underuse of the target discourse markers are mainly attributed to general tendencies and L1 influence. Truly speaking, the artificiality of the ESL students' use of DMs is so obvious. Such overuse could have come merely to implement them which is totally the opposite when a native writer utilizes such DMs.

Note:

The researcher is officially authorized and licensed to use both corpora for academic purpose.

REFERENCES

- [1] Alami, A. (2015). Pragmatic Functions of Discourse Markers: A Review of Related Literature. *International Journal on Studies in English Language and Literature (IJSELL)*, Volume 3, Issue 3, PP 1-10.
- [2] Ali, E. and Mahadin, R. (2016). The Use of Discourse Markers in Written Discourse by Students of English at the University of Jordan. *International Journal of Humanities and Social Science*. Vol. 6, No. 3. PP. 23-35.
- [3] Blakemore, D. (1987). *Semantic constraints on relevance*. Oxford: Blackwell.
- [4] Btoosh, Mousa A. & Taweel, Abeer Q. (2011). Contrastive Rhetoric: Inflation, verbal voices and polyphonic visibility in learners and native speakers' academic writing. *Asian EFL Journal*, 13(3), 205-228
- [5] Btoosh, Mousa A. (2010). *Interlanguage Corpus of Arab Students of English (ICASE)*, USA.
- [6] Croucher, S.M. (2004). Like, you know, what I'm saying: A Study of discourse marker frequency in extemporaneous and impromptu speaking. *National Forensic Journal*, 22(2-3), 38-47.
- [7] Fraser, B. (1990). An Approach to Discourse Markers. *Journal of Pragmatics*, 14, 383- 395.
- [8] Halliday, M. & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- [9] Hinkel, E. (1997). Indirectness in L1 and L2 academic writing. *Journal of Pragmatics*, 27 (1997) PP 361-386
- [10] Hyland, K. (2013). *Genre and Discourse Analysis in Language for Specific Purposes*. Oxford: Wiley-Blackwell.
- [11] Iseni, A. et al (2013). The Role of Discourse Markers and Cohesive Devices in Writing: EFL Students a Case Study. *Anglisticum Journal (IJLLIS)*, Volume: 2. PP 35-48
- [12] Johnstone, B. (2002). *Discourse analysis*. Oxford: Blackstone publishers.
- [13] Koops, C. and Lohmann, A. (2013). *Discourse Marker Sequencing and Grammaticalization*. Berkeley Linguistics Society: The Linguistic Society of America. PP 108-122.
- [14] Miller, C. R. (1984). Genre as social action. *Quarterly Journal of Speech*, 70, 151-162.
- [15] Muller, S. (2005). *Discourse Markers in Native and Non-native English Discourse*. Amsterdam: John Benjamins
- [16] Redeker, G. (1991). Linguistic markers of discourse structure. *Linguistics*, 29, 1139-1172.
- [17] Schiffrin, D. (1987). *Discourse markers*. Cambridge: Cambridge University Press.
- [18] Swan, M. (2005). *Practical English usage*. Oxford University Press, Oxford, UK.
- [19] Wei, S. (2013). The importance of discourse markers in English learning and teaching. *Theory and Practice in Language Studies*, 3(11): 2136-2140.
- [20] Xu, H. (2001). *Metadiscourse: A cross-cultural perspective*. Nanjing: Southeast University Press

Abeer Qasem Al-Taweel was born in Jordan in 1968. She received her PH.D degree in linguistics from Arizona State University, United State of America in 2005.

She is currently an associate professor in the Department of English Language and Literature at Al-Hussein Bin Talal University, Jordan.

She has been teaching in the USA, Saudi Arabia and Jordan for more than 15 years. Her research interests include discourse analysis, semantics, rhetoric and ESL.