

TOWARDS A FRAMEWORK FOR BUILDING A TOOL TO ASSIST L2 WRITING BASED ON SEARCH ENGINES CAPABILITIES

The Case of English Phrases and Collocations

Grami M. A. Grami

Department of European Languages, King Abdulaziz University, Jeddah, Saudi Arabia

Basem Y. Alkazemi

Department of Computer Science, Umm Al-Qura University, Makkah, Saudi Arabia

Keywords: ESL Writing, Collocation, Lexical Phrases, Search Engines.

Abstract: Writing is one of the most difficult skills to learn and it gets more complicated when students learn to write in another language. In fact, results of language proficiency tests such as IELTS shows a systematic tendency of Arab students scoring less in writing than any other skill. Obviously there are various reasons that complicate the task of ESL writing but we focus here on the incorrect combination of words, more specifically collocations and lexical phrases, and its relation to L1 interference. We propose an alternative approach in teaching ESL writing which utilizes common search engines in finding out not only correct usage of words but systematic types of errors so they can be avoided. Moreover, students can use such an approach to validate their writing style in their coursework.

1 INTRODUCTION

Writing is probably the most difficult language skill for many ESL/EFL students which becomes evident when we examine results of proficiency tests such as IELTS. In 2009 for instance, the average score of IELTS academic test takers in writing according to Cambridge ESOL: Research Notes (2010) was 5.51 (the maximum score is 9) which was the lowest band scored ever and well below the overall average of 5.88.

When we closely inspect the writing result of Arab test takers (See table 1 below) we discover that they scored the lowest mean (4.89) of any linguistic background which begs the questions 'why?' and 'how can their writing be improved?'

There are many reasons that make Arab ESL writers struggle but we are trying to focus on the area of combining words in this project. L2 writers in general encounter difficulties when attempting to produce accurate English sentences using the right combination of words that also fit into the correct

Table 1: Mean IELTS score of some first languages.

Academic	Listening	Reading	Writing	Speaking	Overall
Amharic	4.78	5.64	5.62	6.11	5.60
Arabic	5.14	4.96	4.89	5.65	5.23
Bengali	5.85	5.44	5.54	5.87	5.74
Chinese	5.72	5.85	5.19	5.28	5.57
Dutch	7.95	7.79	6.79	7.60	7.60

contexts of usage. A key element for such difficulty usually corresponds to the interference of the mother language when constructing sentences, followed by instant interpretations into English which we believe is a contributor to ESL writing difficulty among Arab learners.

One resource learners can use to check their sentences and the context in which they occur is the Internet search engine. In fact, the literature shows that search engines could function as free online resources readily available to many ESL learners and for various purposes. In our case, the use of

search engines could yield further information about how phrases and collocations are used, by who and to what extent. Apparently mainstream engines are general tools by default and as a result are very likely to produce different results some of which are confusing or even misleading. To filter the general outcome, certain measures are required including observing the number of results and where they come from. In other words, the more pages that use exactly the same phrase/expression, the more likely the phrase in question is correct. Similarly, the more results that originate from carefully edited sources such as recognised institutions, media organisations, government bodies and global corporations, the more likely these results are to be trusted. The hypothesis of this research therefore is that "search engines can improve learning English writing at the lexical phrases/collocations level."

This project would be divided into three stages, the first which is the subject of this study involves testing our theory against results generated by Google and setting criteria for judging incorrect phrases and collocations used by Arab ESL learners and this stage is the main focus of this paper. The second stage would be to conduct an empirical study in a university-level ESL writing course where students would be using Google to check phrases or collocations they are unsure about. The final part of the study would build on the findings and recommendations of the two previous stages and design a tool which can be incorporated into readily available document processing programmes (e.g. MS-Word™ and Open Word) and which can compare certain phrases in texts against the actual results of recognised search engines like Bing™ and Google™. We intend to identify a new application of an already existing technology. In theory, the tool should assist L2 writers by significantly reducing the time required to verify every phrase needs checking and by suggesting various validating techniques (filters) which can be selected individually.

2 LITERATURE REVIEW

2.1 Collocations

These are closely related concepts and are often considered to be problematic areas among L2 learners even advanced ones especially in the lexical level of ESL writing as reported by various experts in the field (Halliday & Hassan, 1976; Sanders & Pander-Matt, 2006; Xiu-lian, 2007; Yin, 2009; and Stapleton & Radia, 2010)

Halliday & Hassan (1976) mention that collocation is part of lexical cohesion, and it is somehow associated with corpus linguistics as mentioned by Lombard (1997). It is defined by Halliday (1961) as "the syntagmatic association of lexical items, quantifiable, textually, as the probability that there will occur at n removes (a distance of n lexical items) from an item x, the items a, b, c ...". In layman terms, collocation refers to certain words commonly used together which co-occur more often than only by chance. Examples of collocation in English include the use of verbs like 'do' and 'make' and adjectives 'quick' and 'fast' with certain nouns, for instance one can say 'do your homework' and 'make a sandwich' but it is unusual to swap the verbs in these commands even if syntactically correct and the same applies to 'fast train' and 'quick shower'. (Guo & Zhang, 2007) This is one reason why collocation is confusing because 'do' and 'make' are almost synonymous to many ESL students who would assume, probably when applying L1 analogy, that they are interchangeable. Verb + noun collocation is possibly the most common type but there are also verb + adverb (vividly remember), adverb + adjective (fully aware), adjective + noun (excruciating pain), and noun + noun (ceasefire agreement) collocations.

2.2 Lexical Phrases

On the other hand, English phrases - more specifically lexical phrases - are defined by Nattinger and DeCarrico (1992: 1) as "chunks of language of varying length ... that occur more frequently and have more idiomatically determined meaning than language that is put together each time." Other definitions like the one provided by Lindstromberg (2000) also suggest that phrases should be treated like units of language rather than a collection of words which therefore makes phrases largely inflexible. For example, 'by and large' means 'generally' and 'as well' means 'too'. In this sense, collocations and lexical phrases are similar as both refer to fixed association of words. Lindstromberg (ibid.) mentions that collocation is the wider term and it refers to both fixed lexical phrases and relatively loose association of words.

2.3 Teaching Collocation and Phrases in ESL Classroom

Language teaching experts have recognised the need to teach and learn more collocations and phrases in L2 classroom. The common approach is to elicit lists

of words that commonly collocate with others (e.g. verbs such as 'do', 'make', 'take', 'have' and 'break') and explicitly teach them to students. The same technique applies to common English phrases. One good example of such effort is Oxford Collocations Dictionary for Students which is also available online.

To us however, teaching word lists is not always a practical solution. There are many reasons why we take this line of thinking one of them is neatly explained by Altenberg (1991) who mentions that almost 70% of words are part of recurrent combinations and English phrases count to the thousands. Another reason would be the time and effort required to go through each and every word/phrase and the limited success when it comes to actual production later on. We therefore argue that teaching lists of words with no reference to the context in which they are used could significantly reduce students' learning achievement, a point confirmed by Lindstroberg (2000), and Nattinger and DeCarrido (1992).

2.4 L1 Interference

Language transfer or L1 interference has been a central point in second language acquisition (SLA) and language teaching and therefore it has been well-documented and researched (e.g. Odlin, 1989; White et al., 1991; Lightbrown & Spada, 1997; Brown, 2000; Picard, 2002; and Bordag & Pechmann, 2007). In general terms, this phenomenon happens when language learners apply knowledge from their mother tongue to a second language, which in our case would be applying Arabic structures into English. (Ryan & Meara, 1991; and Fender, 2008) We more specifically argue that a major contributor to the incorrect usage of collocations and phrases among Arab ESL learners is the interference from similar structures in L1.

In fact, we believe there is ample evidence from the literature and our own investigation to support this theory. For example, one unusual combination of words, supposedly to form an awkward expression, is what organisers of the 2010 Saudi students' conference in the UK used for a slogan which reads 'from different soils into one soil'. As far as we are aware, no such expression exists in English and to make sure we consulted Cambridge Dictionary of Idioms in addition to more general search engines to look for similar combination of words but came up with nothing, a simple search for the exact phrase using quotation marks in Google returned no results. We believe the aforementioned

slogan can only be traced to a relatively common expression used in Arabic journalism and simply has just been literally translated into English. Another example we encountered was learners writing "I want to register my voice in MP3 format" indicating that they actually want to record or tape their voice into a digital recorder.

2.5 Search Engines and ESL Writing

Many experts recognise the important role played by technology and online resources in modern ESL learning. Stapleton & Radia (2009) for instance believe technology contribution to the field of L2 writing has been known as early as when word processing programmes became widely available. Lincoln (2003) more to the point of this study recommends ESL teachers to explicitly teach their students how to use search engines as part of their learning.

However, although the literature of educational technology acknowledges the existence of such a technique among ESL student writers as using search engines to check phrases and collocations (Stapleton & Radia, 2009; and Guo & Zhang, 2007), it vaguely describes how these students actually use these resources. The available literature in fact hardly answers basic questions like how widespread is this practice?, from where have students learned this technique?, what measures do they use to filter search results?, what renders a phrase/collocation acceptable?, how often do students use this technique?, and are students qualified to use general search engines in demanding situations like assessed ESL writing?

Another issue that may affect available text processing software such as MS Word is that they cannot identify certain incorrect collocations nor can they show how popular a phrase/collocation is. For instance, a phrase we considered in this paper was 'from different soils into one soil', which if was searched in Google returns no results, i.e. it does not exist, unlike MS Word which shows no style errors at all.

3 METHODOLOGY

For the first stage of this study, we gathered samples of possibly incorrect phrases and collocations from original texts written by Arab ESL students (n = 37). We then checked these combinations of words against Google by using some preset criteria to filter the returned search results which are the number of

returned results of the possibly incorrect combination and the alternative combination, the format (.doc(x) and .pdf against other formats of less academic association) and their source (institutional, academic, governmental against other sources of less restricted nature).

As for the raw number of results, we were looking for figures to indicate the popularity or otherwise of an expression used by ESL learners. Sources of documents were checked to determine whether results come from trusted websites and/or official documents (academic, organisational, institutional and/or governmental).

Additionally, results of incorrect phrases and collocations usage can be used to determine the scale of the problem, and by inspecting the geographical information of these results, it can also be determined if these errors are more prone to be committed by ESL learners of specific linguistic backgrounds e.g. Arab learners. In other words, we intend to establish whether certain errors are originated from geographic domains more than it would be possible only by chance.

We identified three possible categories of incorrect usage and set measures to deal with these different possibilities accordingly; if a collocation/phrase yields no similar results then we judge it 'isolated and incorrect' then try to guess what the writer intends to say usually by referring to corresponding ideas in his/her L1. If however its return results are found commonly in Arab domains

but not other sources then we examine the possibility of L1 interference and how widespread it is. Finally, if a collocation/phrase does not exist in English but is very common among ESL learners of different backgrounds then we categorise it under 'common errors' regardless of the writer's mother tongue. In every case we proposed an alternative option which we think is more accurate and we check our alternatives against Google as well.

4 RESULTS AND DISCUSSION

The results show a definite answer when it comes to choosing between two possible collocations/phrases; in every case investigated, the alternative option significantly outnumbered the original phrase/collocation. We therefore recommend using search engines results to indicate which string of words is more likely to be correct if students are in doubt choosing between more than one possible combination. (See table 2 below)

As for the different filters used, the location filter (Arabic domains) can tell us - to some extent - if an error is common among learners from this particular background. We for instance have identified two incorrect usages of the preposition 'from' which were found in texts written by Arabs and we can relate this phenomenon to L1 interference. In other words, a language teacher can now address this

Table 2: Google Results of Original Phrases.

Original Text	Raw Search Results	Filtered Results (by location)	Filtered Results (by format)	Filtered Results (by source)	Alternative Choice	Search Results
"register my voice"	28,500	2 Pages	8 .pdf 1 .doc	none	"record my voice"	146,000
"different soils into one soil"	none	none	None	none	--	--
"Speak in English"	797,000	33,900 Pages	33,700 .pdf 3,820 .doc	3,060 .gov 7,860 .edu 1,120 .ac.uk	"Speak English"	8,220,000
"Get my advantages"	3	none	None	none	"Get my rewards"	34,100
"near from my family"	5	1 Page	None	none	"Near my family"	985,000
"better from " (comparison)	*	1,330 Pages	None	none	"better than "	91,900,000
"to talk English"	299,000	7,630 Pages	14,100 .pdf 1,020 .doc 8 .ppt	4,240 .edu 321 .gov 489 .ac.uk	"to speak English"	9,040,000
"His days are finished"	2,310	none	3 .pdf 1 .doc	none	"His days are over"	89,900

problem by asking students not to constantly translate meanings from their mother tongue.

Another example of L1 interference is the incapacity to differentiate between verbs such as 'record' and 'register' and how to use them in their proper contexts. Again, this mistake has occurred in texts written by Arabs which further supports the hypothesis that L1 interference is widespread indeed.

However, in extreme cases when a whole phrase originated only in Arabic is wholly translated into English with no regard to L2 conventions we found that no similar results were found in any other website. The only case we came across was using a relatively common expression in Arabic by a group of Arab students in the UK to promote their conference but we also accept that it is not uncommon to see more of the same. Google yields no results and we could not come with an expression that conveys the same meaning.

In few occasions, the choice of words might be a personal preference or follows conventions of formality and having a fewer number of results does not always mean that our alternative choice is correct and the original is not. For example 'talk English' shows much less results than our preference 'speak English' but as the former collocation abundantly appears in edited academic and official websites as well as revised documents, one cannot reject it and accept the latter simply because it shows significantly more results.

The filters are interesting methods to determine various characteristics of phrases. We already have discussed that 'location' helps us understand if an error is common among certain groups of learners. We also found that results filtered by source can be stricter than results by format.

5 RECOMMENDATIONS FOR ESL CLASSROOM

Our recommendation for language teachers therefore is to avoid translating whole expressions from the mother tongue to the target language and focus instead on teaching expressions and phrases commonly used in English. We also suggest that it is recommended that when students write in a foreign language that they follow its conventions without constantly referring back to their mother tongue. In fact, almost all the errors we identified can be traced back to Arabic in some cases with no regard to L2 conventions at all as in the case with 'from different soils into one soil'.

As for using search engines, we suggest that if a result appears in great numbers in educational and official websites then it should be treated as a correct combination of words. Finally, although the main purpose of the study is to aid ESL writers the tool we aim to develop can be used innovatively to serve other purposes as well. For instance, search engines can show results from specific regions and in websites of certain languages which means one can check how widespread an error is among learners from a specific background and compare that to others. The identification of these errors can further help the research into L1 interference and the role of context in ESL learning.

6 LIMITATIONS AND FUTURE RESEARCH

Our investigation of the writing samples indicates a widespread problem of incorrect usage of phrases and collocations among Arab ESL writers chiefly due to the interference of their L1. However, it is difficult to find systematic patterns of errors within writing of students of a certain background without an empirical study that involves asking students to write about topics very likely to generate such patterns and - in our case - students can also check combination of words they are not sure about against search engines and from observing them doing so we shall be more informed about the techniques used and if these can be integrated in our tool model.

Our attention therefore should move to actual ESL Arab writers who use Google and other search engines to help them determine whether a phrase/collocation they use is acceptable. This proposed empirical study would be the second stage of our project in which we aim to gather as much information about search techniques and incorrect usage of words as possible. The results should help us better understand how search engines work and how can we further develop various methods to have as accurate results as possible. All the data gathered would then be considered when we finally design an open-source support tool which can be prompted to search certain phrases and classify results according to the filters we suggest.

7 CONCLUSIONS

Having reviewed the literature and assessed the scale and widespread of the problem, we would argue that there is a feasible chance ESL students' writing can

be improved using Google and similar search engines in tackling the problematic areas of using collocation and lexical phrases. The incorrect usage of these items happens in large part due to L1 interference as we attempted to establish in this study. It is however not a very practical solution to expect students to read and remember every lexical phrase and collocation list available. We therefore propose an alternative approach in using search engines which is based on two concepts; simplicity and learning by doing.

REFERENCES

- Altenberg, B. (1991) 'Amplifier collocations in spoken English', In S. Johansson & A. Stenstrom (eds.) *English computer corpora*, pp 127 – 147, Berlin: Mouton de Gruyter.
- Brown, H. D. (2000) *Principles of Language Learning and Teaching*, 4th ed., London: Longman.
- Cambridge ESOL: Research Notes, Issue 40/ May 2010, Cambridge University Press.
- Fender, M. (2008) 'Spelling Knowledge and Reading Development: Insights from Arab ESL Learners', *Reading in a Foreign Language*, 22, pp. 19-42.
- Guo, S. and Zhang, G. (2007) 'Building a customised Google-based collocation collector to enhance language learning', In *British Journal of Educational Technology*, Vol. 38 No. 4, pp 747 – 750.
- Halliday, M. A. K. (1961) 'Categories of the Theory of Grammar', In *Word*, Vol. 17, pp 241 – 492.
- Halliday, M. A. K. & Hassan, R. (1976) *Cohesion in English*. London: Longman.
- Lightbown, P. M. and Spada, N. (1997) 'L1 constraints in interlanguage judgments of grammaticality', Paper presented at the *17th Annual Second Language Research Forum*, Michigan State University, East Lansing.
- Lincoln, K. (2003) 'Teaching Search Engines to ESL Students: Avoiding the Avalanche', In *The Internet TESL Journal*, Vol. 9 No. 6, available on <iteslj.org>, accessed on 11 August 2010.
- Lindstromberg, S. (2003) 'My good-bye to the lexical approach', In *Humanising Language Teaching*, Vol. 5 No.2.
- Lombard, R. J. (1997) *Non-native speaker collocations a Corpus-driven characterization from the writing of native speakers of Mandarin*, PhD Thesis, MI: UMI.
- Nattinger, J. R. & DeCarrico, J. S. (1992) *Lexical Phrases and Language Teaching*, Oxford: Oxford University Press.
- Odlin, T. (1989) *Language Transfer: Cross-Linguistic Influence in Language Learning*, Cambridge: Cambridge University Press.
- Ryan, A. and Meara, P. (1991) 'The Case of the Invisible Vowels: Arabic Speakers Reading English Words', In *Reading in a Foreign Language*, Vol. 7 No.2, pp 531 – 540.
- Sanders, T. & Pander Maat, H. (2006) 'Cohesion and Coherence: Linguistic approaches', In Brown. K. et al. (eds.) *Encyclopedia of Language and Linguistics*, 2nd edition, volume 2. London: Elsevier.
- Spada, N. and Lightbown, P. M. (1993) 'Instruction and the development of questions in L2 classrooms', In *Second Language Acquisition*, Vol. 15, pp 205 – 224.
- Stapleton, P. and Radia, P. (2009) 'Tech-era L2 writing: towards a new kind of process', In *ELT Journal*, Vol. 64 No. 2, pp 175 – 183.
- White, L., Spada, N., Lightbown, P. M., and Ranta, L. (1991) 'Input enhancement and L2 question formation', In *Applied Linguistics*, Vol. 12, pp 416 – 432.
- Xiu-lian, Y. (2007) 'Lexical approach to college English teaching', In *Sino-US English Teaching*, Vol. 4 No. 10, pp 22 – 24.