Computerized Corpus Based Investigation of the Use of Multi Word Combinations and the Developmental Stages by Tertiary Level EFL Learners

Ali Şükrü Özbay[1] and M. Naci Kayaoğlu[2]
Karadeniz Technical University, Trabzon, TURKEY[1,2]
ozbay@ktu.edu.tr[1], naci@ktu.edu.tr[2]

Abstract

It is an old consensus by now that the students of English as a Foreign Language all around the globe are facing a variety of problems that are likely to make their ability to write expository and academic essays fail to meet the required criteria in terms of lexical complexity, phraseological richness and syntactic variation. Part of the reason for the conduct of the present study, which is mainly a computerized contrastive learner corpus analysis of BAWE and KTUCLE corpora, is to make an analysis of the overall frequency, variation and development of multi-word combinations (lexical combinations) as well as investigating the common multi word combinations in terms of underuses, overuses and misuses. For this purpose, the expository and academic essays written by tertiary level EFL students of English in 1st and 2nd years were taken as the samples for the study. Lexical and phraseological investigation of KTUCLE corpus, which was solely compiled to serve the objectives of this study, entailed the presence of another similar sized academic corpus and, as a result, British Academic Written English (BAWE) was selected as the reference academic learner corpus for the study. In the selection of the samples for the study purposive sampling methodology was used and the analysis was done by comparing the language use of the learners who were categorized into two distinct levels with respect to their present language proficiency levels through Oxford Online Placement Test and the levels of the subjects (n: 120) were determined based on the scores they got from the online test. The data for the study were both qualitative and quantitative in nature. Quantitative analysis of the learner corpora findings was followed by a qualitative analysis of the learners' responses to the open ended interviews. The result of this computerized and contrastive learner corpus analysis revealed the following findings: First of all, it was observed that KTUCLE learner corpus turns out to be much less complex in terms of lexical diversity, phraseological variation and density than the reference corpus BAWE. The

second finding was that the multi word combination samples from KTUCLE were observed to possess overuses and underuses of some of the most frequent combinations. Yet another finding was a high level of diversity in terms of the incorrect use of multi word combinations as a result of the learners' limited word stock rather than from their ignorance of these multi-word combinations.

1. Introduction

It is an old consensus by now that multi word combinations are very frequent, "crucially important" and common in English language particularly in terms of written texts (Athayde, 2001, p. 10). According to (Renouf, 1988) they are the type of combinations which are vitally important for learners and constitute a significant part in the world of English Language Teaching (ELT, thereafter). Surprisingly enough, their significance for the ELT world has not been noted very often in literature and few linguists like Sinclair and (Renouf, 1988) emphasized the intensive penetration of these constructions in English language based on various corpus analyses. While (Altenberg, 2001) states that these types of constructions are also present in numerous other languages, many multi word combinations (MWCs, hereafter) are not applied "cross-linguistically" (p.196). It is also the case that majority of existing studies on MWCs investigated to the structural aspects of these constructions rather than learners` tendency towards them and their teaching in ELT classrooms. Nevertheless, thanks to the developments in corpus tools and computers and with the start of computerized learner corpus studies in recent decades, the actual significance of MWCs has been understood better in linguistics circles (Altenberg B. &., 2001).

(Polenz, 1963) asserts that MWCs are the prime examples of nominalization on which almost every language highly depends, and therefore he claims that these structures have very common usages in English language. Furthermore, they are pre-determined linguistic issues which are constructed in accordance with general and unchangeable or semi-changeable rules of language as encountered in Sinclair's Idiom Principle (Nesselhauf, 2005, p. 111). MWCs are typical examples of Sinclair's idiom principle. According to this principle, words are regarded as structures following a certain sequence. These (semi)-pre-constructed phrases are constituted in accordance with certain language rules, and require single or restricted choices. Therefore, language users generally do not have any free choice to change or transform them (Sinclair, 1991, p. 110). In their comprehensible study, (Erman, 2000) found that in more than half of all researched authentic texts, that equals a percentage of 55, the idiom principle was used. (Vincze, 2008) asserts that MWCs are common in language use, and frequently present 'unique behaviour', therefore, they usually cause a problem to natural language process. She also claims "In natural language processing, one of the most challenging task is the proper treatment of collocations, which term comprises multi verb constructions as well". This shows that MWCs are notably frequent and mostly pre-constructed structures in English. For this reason, these constructions have been one of the most complicated issues for even advanced learners of English (Nesselhauf, 2005, p. 111). The "scarcity" of the MWC based studies and the lack of "quantification" related to these studies for EFL learners especially are among the reasons this study has been done for (Nesselhauf, 2005, p. 112). Yet, another factor behind this study was the fact that several existing related studies have largely ignored the affective factors behind MWC development of EFL learners. Therefore, part of the reason this study has been conducted is to find out the affective factors behind tertiary level EFL learners' tendency towards MWCs through an open ended interview. Because, we believe that investigation of these tendencies are likely to shed light on the learners' true feelings and thoughts regarding the MWCs in English and give some clues about why and how they learn or fail to learn these structures. All and all, we believe that like many other lexical and phraseological units, MWCs are also important elements of almost

every language all over the world as well as have unique characteristics in the ways they follow certain linguistic patterns. The fact that they are largely used by native speakers point to the case that if an EFL learner   has a native-like proficiency norm in his/her mind or wants to reach particularly an advanced level or near-nativeness s/he has to pay a special attention to these structures as well as improve his-her knowledge of these pre-constructed combinations and frequently apply them..

The use of corpus for lexical investigation is not a recent phenomenon but it`s full significance and value has, in the last decade, been realized especially after the introduction of computerized corpus tools by a much larger group of linguists all around globe. Sixty years ago when it was introduced into the world linguistic research there were very few "enthusiasts" and it was the "obsession of a small group which received little or no recognition from either linguistics or computer science" (Leech, 1991, p. 25). Things seemingly took quite a different turn for corpus since then! (Sinclair J. , 2005), one of the pioneers of modern corpus linguistics, defines the term *corpus* as " a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research" (p.16). As one of the types of corpora, learner corpus "which is a computerized textual database of the language produced by foreign language learners" began to be used largely (Leech, Preface, 1998, p. 10). These computerized corpus tools have had the capability to store and process and analyse learner language in a way that was beyond the imagination previously. The main purpose in collecting a learner corpus is to gather objective data that can help to describe learner language, which is regarded necessary for valid theory and research (Granger, 2002, p. 25). Learner corpora are important because they provide a deviation from the standard norm and the extent of deviation is measured through the investigation of authentic natural language data with several different focus and motivation such as theoretical, pedagogical issues as well as needs of the learners. Part of the reason a learner corpus based methodology was used in the study is that the use of MWCs indicates the level and extent of the EFL learners` communicative competence which requires making appropriate choices of register and style. It does not only refer to a learner's ability to apply and use grammatical rules, but also to form correct utterances, and know how to use these utterances appropriately. Therefore, corpus is a very useful tool for the language learners to have communicative competence. According to (Krieger, 2003) the important point of corpus linguistics is to reveal the patterns of authentic language use through analysis of actual usage (p.36). The aim of a corpus based analysis is not to generate theories of what is possible in the language. Corpus linguistics concern with the usage patterns of the empirical data and implications about the language behaviour. Corpus linguistics provides a more objective view of language than that of introspection, intuition and anecdotes in this respect.

In the scope of the study, the affective factors for the MWC development were also investigated. Second language acquisition is the study of what is learned of a second language and what is not learned. It defines why most second language learners do not achieve the same degree of knowledge and proficiency in a second language as they do in their mother tongue (Gass, 2008). It also deals with developmental processes of the learners who generate a language system known as an interlanguage, a term coined by Selinker. Learners` strategic competence and communication strategies are frequently used in examining interlanguage. The writings of foreign language learners' can be evidence to their limited knowledge of L2. The best way to understand the learners' interlanguage continuums is to control their writing products by using learner corpora. In the scope of this study, a learner corpus designed by the researchers with a strict design criteria was used for this purpose.

2. Research Questions

1. What is the overall frequency of acceptable MWCs usages in the academic essays of the tertiary level EFL learners?
2. What are the affective factors behind the multi word combination development of the tertiary level EFL learners?

3. Methodology

This corpus-based descriptive study, in which a principled collection of natural texts was analysed for a purpose by using computer-based quantitative analytical techniques and interpreted qualitatively investigated the MWC use and the developmental factors behind them by tertiary level EFL learners.

3.1. Setting, Sampling and Participants

The setting of this study was the Western Languages and Literature Department of Karadeniz Technical University, Trabzon, Turkey. The participants were tertiary level Turkish EFL students at a middle size university in Trabzon, the East of Turkey. Participants of the study included the prep, first and the second year students from the Department of English. The student portion of the participant population was diverse in age. Their ages ranged from approximately 17 to 23. Of the120 student participants, 96 are females and 24 were males. The sampling procedure used by the researcher is purposive sampling; the individuals meet the criteria for the study. In an attempt to understand the participants' current language levels, Oxford Online Placement Test was implemented to all the participants. Based on the scores the participants were divided into groups of two or three for further research purposes.

3.2. Instruments

The NNES corpora were built from the argumentative essays written in English by native speakers of Turkish. The learner corpora were built following the criteria of design suggested in (Hunston, 2002). In this study, the NNES corpus is composed of samples of manuscripts in English of Turkish speaking-learners in the Department of English of Karadeniz Technical University. Corpus investigation and an open ended interview were used to obtain data regarding the affective factors behind MWC development of learners. The researchers collected the data after obtaining necessary approval and consent for collecting data. The researchers also delivered an informed consent form reminding the students about the aim of the study, confidentiality, voluntariness, authority and anonymity of the participation. The Oxford Online Placement Test was prepared by the Oxford University and is used widely all around the globe for language placement purposes. The only criterion for the study was that all participants were affiliated with the scores they got from the online placement exam by Oxford University. The database of this study consists of a learner corpus that was compiled by the researcher and the name of the corpus is Karadeniz Technical University Corpus of Learner English (KTUCLE). It contains essays that are argumentative in character and the selected sample for the present study is a total of 500.000 words.

3.3. Open-ended Interview

An open ended interview was conducted with the subjects. Each of these interviews lasted around 15 minutes. The answers were short and this eased the transcription of the data. This part presented the student responses based upon the analysis of the affective factors behind learner tendency towards

MWCs. The responses from the each participant were used in order to investigate the affective factors (positive/negative) behind MWC development of students.

4. Data Analysis

This study aimed to find out the affective factors behind multi word combination development of learners on the basis of the analysis of the most common MWCs in English *have, make, take, give* and *do* in the learner (non-native) corpus called KTUCLE (Karadeniz Technical University Computerized Learner Corpus). It also benefited from a reference/control corpus called BAWE (British American Written English), which is designed for the similar purpose, in order to compare MWC findings in our non-native corpus KTUCLE to the findings obtained by native users in it. In addition, the study investigated the role of the mode and amount of language instruction behind learners' tendency towards MWCs, especially in learning them.

In order to collect data, corpus investigation and open ended interviews were used, and the study was conducted with 120 EFL students currently having education in English Language and Literature Department of Karadeniz Technical University.

This part presents the results of MWCs found in KTUCLE. Based on the learner corpus (KTUCLE) data, all the multi word combinations with *make, have, take, give and do* were extracted and listed separately. First of all, all MWCs in KTUCLE are demonstrated in Table 1. As seen in the Table 1 below, 39 MWCs with "have" are found in students' writings in learner corpus. Verb-noun combinations with have are commonly used by the students. Then, it is possible to say that students know much more MWCs with have than others. It may be because of the fact that it is a very common verb and students frequently encounter constructions with "have". Also, this tendency on the part of high use of MWCs with "have" may stem from the students' extensive exposure to MWCs with "have" in their course books. MWCs with "make" are counted as 24 in students' writings. "Make" is also another common verb so the students may have a tendency to use it. In addition, as the use and meaning of "make" is similar to our native language, Turkish, students make a positive transfer and learn them more easily.

MWCs with "take" "*give*" and "*do*" are no more than a few. A close scrutiny of the table reveals that although the number of MWCs is relatively limited to a few, almost all the existing ones are the ones that are transferable from L1 of the learners. This relatively scarce use of MWCs in the table accounts for the fact that the tertiary EFL learners that fall within the scope of this study preferred to use phrases rather than single word tokens within a limited. This unpopularity seems to stem from such things as the lack of exposure to them, the lack of interest in them, or the lack of information about them. When all these factors combined, it becomes natural the students may not use them frequently and consciously.

HAVE	MAKE	TAKE	DO & GIVE
Have a claim	Make a remark	Take care of	Do translation
Have a laugh	Make a choice	Take a break	Do research
Have a tendency	Make a mistake	Take action	Do jokes
Have access	Make a deduction	Take into account	Do investigation
Have a score	Make an invention	Take a decision	Do exercise
Have a difference	Make a plan	Take a role	Do harm
Have an order	Make a connection	Take into consideration	
Have an effect	Make a guide	Take a notice of	Give direction
Have trouble	Make an exception		Give reaction
Have experience	Make statement		Give

Have a question	Make an impression		encouragement
Have a conclusion	Make comparison		Give an example
Have an interest	Make benefit from		Give
Have an agreement	Make a difference		information
Have a wish	Make a profit		
Have a risk	Make progress		
Have applications	Make analysis		
Have a desire	Make a research		
Have a rest	Make an assessment		
Have a proud	Make a contribution		
Have a break	Make use of		
Have a balance	Make judgments		
Have determination	Make reference		
Have a solution	Make a speech		
Have a dream			
Have a feel			
Have an influence			
Have an operation			
Have a connection			
Have a doubt			
Have a reaction			
Have a function			
Have confidence			
Have relation			
Have an impact			
Have value			
Have a look at			
Have a dislike			
Have an objection			
39	24	8	6 & 5

Table 1: MWCs with have, make, take, do and give

Table 2 presents MWC results found in KTUCLE and British Academic Written English (BAWE), which was applied as a native reference corpus for this study. Based on both corpora data, most commonly verb–noun combinations with were all listed in the following Table 2.

The Table 2 below shows the results from BAWE within linguistics and KTUCLE in order to make a comparison between MWC usages in a non-native corpus and those in a native one. Some of the most frequent MWCs were given in the table. All these data reveal that MWC findings obtained from the KTUDELL students are close to native-like norms with some exceptions of overused ones. However, from the table it is clear that MWCs used by KTUCLE students are generally similar. For instance, while *take care of* was used 141 times, other MWCs were used in a rather low rate. Moreover, except for *take care of, have an effect* was used the most, that is 28 times. This reveals that there is an excessive imbalance with some MWCs in KTUCLE corpus.

MWCs	BAWE hits	KTUCLE hits	LL Score	Overuse	Underuse
Take into account	35	10	14.71		+
Have impact	11	14	0.36	-	-
Have an influence	16	3	9.77		+

Take into consideration	14	8	1.66	-	-
Make use of	23	9	6.34		+
Have access	6	7	0.08	-	-
Make a decision	14	12	0.15	-	-
Have control (over)	10	3	3.98		+
Take care (of)	7	141	148.79	+	
Make a distinction	8	1	6.20		+
Have an effect	42	28	2.82	-	-
Make an effort	6	10	1.01	-	-
Make reference	9	1	7.36		+
Have a meaning	21	6	8.83		+
Have knowledge	10	10	0	-	-
Have a tendency	7	2	2.94	-	-
Make a mistake	6	12	2.04	-	-
Make a choice	9	4	1.97	-	-
Give an example	11	17	1.30	-	-
Give information	13	11	0.17	-	-
Make judgement	8	1	6.20		+
Make a connection	6	1	3.96		+

Table 2: MWCs' Overuse and Underuse in BAWE hits and KTUCLE hits

As to the affective factors behind MWC development of learners four main categories were found. Affective factors are very important determinants for learners' failure or success in writing. Therefore, they cannot be underestimated because they may be guide for coping with writing problems of learners or making them more encouraged in order to let them to improve their writing skills (Langer, 2004). The common affective factors related to L2 or foreign language learning are emotions, self-esteem, empathy, anxiety, attitude and motivation, but other various general terms are also present such as pedagogical, psychological, gender, personal, environmental and cultural factors (Shumin, 2002, pp. 204-206). As in writing, multi word units in English such as MWCs, which have been very challenging for second or foreign language learners, have such kinds of affective factors on learners' tendency and learning process (Langer, 2004).

As seen in the Table 3 below, pedagogical factors were subcategorized into five sub-factors: background knowledge, lack of background knowledge, amount and mode of instruction, lack of linguistic and/or vocabulary knowledge and lastly, poor writing skills. According to interview responses, 36 students think that background knowledge about MWCs is very important and they had this background information during their high school or university prep. class years. However, 35 students say that they have difficulty in learning MWCs because of their lack of background knowledge. According to 21 students, amount and mode of instruction is very effective in learning these multi-unit structures, and they are aware of them thanks to lessons. 17 and other 14 students think that lack of linguistic or/and vocabulary knowledge and poor skills respectively are main obstacles for them to begin learning MWCs. These reveal that background knowledge is very important because these may facilitate for the students to reconstruct and to analyse these combinations in their minds more easily in accordance with their previous information. The lessons which they have and their knowledge on linguistic, vocabulary and general writing skills may also affect the learners' failure or success in MWC development.

According to the findings obtained from the interviews, personal factors were subcategorized into four sub-factors: using internet, personal interest or experience in writing, using dictionary and watching English movies, reading English magazines, listening English music etc. 83 students say that they learned these structures while using internet. 11 students say that they are familiar with these

constructions due to their interest or experience in writing; and therefore, they use MWCs in their essays. 9 students think that they generally use dictionary in order to learn new lexical items, and by this way, they also came across and learned these structures. 7 students think that their extensive watching, listening and reading help them to learn these MWCs. In other words, learners' personal attitude towards writing and of course towards these combinations may have crucial effect on their MWC development.

AFFECTIVE FACTORS			
Pedagogical Factors	**Personal Factors**	**Psychological Factors**	**Motivational Factors**
Background Knowledge **(36)**	Using Internet **(83)**	Perception towards Writing (positive) **(38)**	Effort for having a better Writing Skill **(31)**
Lack of Background Knowledge **(35)**	Personal Interest in Writing [Previous Writing Experience] (poem, prose, story etc.) **(11)**	Writing Apprehension/ Anxiety **(8)**	Intrinsic and/ or extrinsic Mot. **(28)**
Amount and Mode of Instruction **(21)**	Using Dictionary **(9)**		Lack of Motivation **(13)**
Lack of Linguistic and/or Vocabulary Knowledge **(17)**	Watching English movies, series etc./ Reading English magazines, books etc. / Listening music etc. **(7)**		
Poor Writing Skills **(14)**			

Table 3: Affective Factors in the MWC development

As seen in the Table 3 above, psychological factors were subcategorized into two sub-factors: perception towards writing positively and writing apprehension. According to the diaries, while 38 students think that they learned MWCs because of their general perception towards writing, 8 students say that their writing apprehension hinders them to learn these multi-word units. It is clear that learners' psychological aspect towards writing itself may be encouraging or discouraging in general.

The above table also shows the motivational factors with its sub-categories: effort for having a better writing skill, intrinsic and/or extrinsic motivation and lack of motivation. 31 students say that they try to learn these structures in order to have better writing skills. Among 28 students, some have individual and spontaneous motivation towards writing and the others have motivation due to external parameters; and this tempts them to learn MWCs in order to have better writing skills. 13 students think that they have lack of motivation, and so even do not try to look for these combinations. These factors clarify that motivation may be significant component for MWC development of learners.

Consequently, as understood from these four main factors and fourteen sub-factors, it may be UNdeniable that there are many situations (whether positive or not) which have great impact on learners' tendency towards MWCs. Therefore, both teachers and students should not ignore these crucial factors; instead, should increase their awareness and pay much more attention to them.

5. Results and Discussion

In this study, both qualitative and quantitative research methods were used. With the help of a corpus software program AntConc 3.0.1, MWC findings which involve the most common MWCs *make, take, have, give* and *do* were found in KTUCLE (Karadeniz Technical University Computerized Learner English), and these were reassessed in accordance with the findings in a native corpus called BAWE (British Academic Written English). However, the quantitative data obtained through corpora did not allow for in-depth explanation of the students' real thoughts and feelings about their intrinsic and extrinsic application of MWCs. For this reason, qualitative data, that is, interviews were also incorporated in the design of the study. The qualitative data collected complemented the findings of the quantitative data. The research questions in this study were:

1. What are the affective factors behind the multi verb construction (MWCs) development of the tertiary level EFL learners?
2. What is the overall frequency of acceptable MWCs usages in the academic essays of the tertiary level EFL learners?

The first question asked is about which affective factors come into play for the tertiary level EFL learners' tendency towards MWCs. Overall results from the students' interviews revealed that four main affective factors had a great role in MWC development of the learners. These were pedagogical, personal, psychological and motivational factors. Interviews also brought the light the sub-factors constituting the main ones, and according to the students' thoughts and feelings in their diaries, these sub-categories were integrated into the main factors. All these factors showed that psychology, motivation, pedagogy and personality are very important determinants and specifically *using internet* at the rate of % 35 was found having the most significant role in learning and application of MWCs. This sub-category also indicates that the personal factors have crucial effects on MWC development of learners. Besides, retrospective protocols with students provide a deep insight to real base of MWC application among them. According to the students' answers during protocols, personal, pedagogical and motivational factors were obtained to a large scale. Almost half of them said that they developed their MWC knowledge in terms of their extensive activities such as listening to music, watching series and movies, reading books, magazines, newspapers and so on.

The analysis of the overall frequencies in the learner corpus suggests that in general EFL learners do not show a numerical difference from native academic corpus writers in MWC use. However, the range and the frequency of MWCs are limited to only a few common MWC structures. This finding is consistent with the findings in (Waibel, 2007) and (Gilquin, 2011) in that the language learners whose native languages (L1) are rich in collocations and word combinations are more likely to frequently use word combinations. The BAWE corpus included more word combinations than the learner corpus but within the scope of the analysis, only those MWC samples found in both corpora were taken for analysis. Naturally, the number of these MWC was limited. L2 proficiency was surely an important factor in this process. In other words, if non-native speakers have good language proficiency and mastery it will be possible that they may produce more MWCs in their writing.

References

Altenberg, B. &. (2001). Recent Trends in Cross-Linguistics Lexical Studies. In *Lexis in Contrast* (pp. 3-48). Amsterdam and Philadelphia: John Benjamins.

Altenberg, B. (2001). *"Contrasting Delexical English Make and Swedish Göra"*. Gothenburg: The University of Gothenburg.

Athayde, G. (2001). *Building a Mean-Downside Risk Portfolio Frontier: Developments in Forecast Combination and Portfolio Choice*. John Wiley and Sons.

Erman, B. &. (2000). The Idiom Principle and the Open Choice Principle. *Text*, 20, 29-62.

Gass, S. &. (2008). *Second Language Acquisition: An Introductory Course.* UK: Routledge.

Gilquin, G. (2011). Corpus linguistics to bridge the gap between World Englishes and Learner Englishes. *Communicación en el siglo XXI*(2), 638-642.

Granger, S. (2002). A Bird's eye view of Learner Corpus Research. In S. L. Computer Learner Corpora, *Granger, S. & Hung, J. & Petch-Tyson, S.* (pp. 3-33). Amsterdam: John Benjamins.

Hunston, S. (2002). *Corpora in Applied Linguistics.* CUP: Cambridge.

Krieger, D. (2003). Teaching Debate to ESL Students: A Six-Class Unit. *The Internet TESL Journal*.

Langer, S. (2004). A Linguistic Test Battery for Support Verb Constructions. *Lingvisticae Investigationes, 27*(2), 171-184.

Leech, G. (1991). The State of the Art in Corpus Linguistics. In A. K. B., *English Corpus Linguistics: Studies in Honour of Jan Svartvik* (pp. 8-29). London: Longman.

Leech, G. (1998). Preface. In S. Granger, *Learner English on Computer* (p. 10). London and NewYork: Addison Wesley Longman.

Nesselhauf, N. (2005). Collocations in a Learner Corpus. How Learner Corpus Analysis can Contribute to Language Teaching. In *A Study of Support Verb Conjunctions* (p. 111). Heidelberg: University of Heidelberg.

Polenz, P. v. (1963). *Function Verbs in Current English.* Düsseldorf: Padagogischer Verl. Schwann.

Renouf, A. (1988). Coding Metalanguage: Issued raised in the Creation and Processing of Specialized Corpora. *Kyote et al. (eds)*, 197-206.

Shumin, K. (2002). Factors to Consider: Developing Adult EFL Students' Speaking Abilities. In *Methodology in Language Teaching: An Anthology of Current Practice* (pp. 204-206). Cambridge: Cambridge University Press.

Sinclair, J. (2005). Corpus and Text - basic Principles. In M. Wyenne, *Developing Linguistic Corpora: A Guide to Good Practice* (pp. 1-16). Oxford: Oxbow Books. Retrieved from http://ahds.ac.uk/linguistic-corpora/

Sinclair, J. (2005). Corpus and Text - basic Principles. In M. Wyenne, *Developing Linguistic Corpora: A Guide to Good Practice* (pp. 1-16). Oxford: Oxbow Books. Retrieved from http://ahds.ac.uk/linguistic-corpora/

Sinclair, J. M. (1991). *Corpus, Concordance, Collocation.* Oxford: Oxford University Press.

Vincze, V. (2008). On the status of Bare Common noun + verb Constuctions. In *In Sinkovics, Bal'azs, LingDok 7. Nyelv'esz-doktoranduszok dolgozatai* (pp. 265-283). Szeged, Hungary: University of Szeged.

Waibel, B. (2007). Phrasal Verbs in Learner English: A Corpus-based Study of German and Italian Students. Freiburg, Albert-Ludwigs-Universität Freiburg.