A Corpus-based Study of Adjective-Noun Collocation in Descriptive Essays by Algerian Students of EFL

دراسة مدونية **للمفردات** التراصفية (صفة+ موصوف) في المقالات الوصفية لمتعلمي اللغة الانجليزية كلغة أجنبية

BENCHEIKH Youcef

PhD Researcher at University of Oran 2 Mohamed Ben Ahmed-Algeria

bencheik.y.30@gmail.com

Abstract: The present paper is a corpus-based study of the use adjective-noun collocations in the writing of Algerian students of English as a Foreign Language. The corpus consists of 30 descriptive essays compiled as a classroom assignment. A frequency-based approach was used to identify the most frequent adjective-noun pattern based on Howarth's (1998) model of categorization of lexical collocations. The study aims, *inter alia* to describe the factors influencing the production of these patterns as well as the strategies learners opt for when dealing with collocations. The results demonstrated that 75% of the erroneous collocations are collocator-based error while 13% are collocate-based ones. In addition, 57% of the errors are due to negative transfer from L1.Finally, the study recommends more explicit teaching of collocations in the Algerian university classes of English using corpus- based activities.

Keywords: lexical collocations, collocational errors, negative transfer, learner corpus, EFL learner

الملخص: تهدف هذه الدراسة للكشف عن نوع واحد من المفردات التراصفية بالتحديد, و هو التركيب(صفة + موصوف) في 30 مقالا وصفيا لمتعلمي اللغة الانجليزية كلغة أجنبية وقد تم اعتماد أنموذج Howarth (1998) لتصنيف المفردات التراصفية للكشف عن مثل هذا النوع. تهدف الدراسة من جهة أخرى إلى وصف العوامل المؤثرة في إنتاج هذا النوع التراصفي وكذلك الاستراتيجيات التي يلجأ إليها المتعلمون في تركيب هذا النوع التراصفي. أظهرت النتائج أن النقل السلبي من اللغة الأم هو المتسبب الرئيسي في هذه الأخطاء التراصفية بنسبة 57% بالإضافة إلى الاختيار الخاطئ تارة للصفة 75 % وتارة أخرى للموصوف 13 % في التركيب التراصفي الواحد. توصي الدراسة بتدريس المفردات التراصفية في أقسام اللغة الانجليزية باستخدام تمارين مستنبطة من المدونات النصية ومدونات المتعلمين.

الكلمات المفتاحية: المفردات التراصفية، الأخطاء التراصفية، النقل السلبي، مدونة المتعلمين، متعلمي اللغة الانجليزية كلغة أجنبية.

Corresponding author: BENCHEIKH Youcef, e-mail: bencheik.y.30@gmail.com

1. INTRODUCTION

The aim of the present paper is two-fold: (a) to discover the most frequent adjectival collocations used by Algerian undergraduate students of EFL in writing descriptive essays, and (b) to investigate the potential factors influencing their production of the identified collocations. Furthermore, this paper is organised into three sections: (1) The literature review section provides a thorough review of collocations as a salient feature of an idiomatic use of language, (2) The methodology section discusses the corpus compilation and analysis process, and (3) The results section analyses students' collocational patterns and provides plausible explanation of the sources of these collocational errors.

2. Literature Review

The present section tackles three central points. First, corpus as a new orientation in second language research is discussed. The notion of collocation is then defined along with researchers' different classifications. This section ends with the discussion of the factors that may influence students' production of collocation.

2.1 Defining Corpus

A corpus is defined as 'a collection of naturally occuring examples of language , consisting of anything from a few sentences to a set of written texts or tape recordings, which have been collected for linguistics study'(Huntson, 2002 p 02). Form and purpose are two defining features of corpora, thus they can be of the following types (ibid, pp 14-16):

1. **Specialised Corpus :** **A** corpus of texts of a particular type, such as newspaper editorials, geography textbooks, academic articles in particular subject, lectures, casual conversations, essays written by students…etc.
2. **General Corpus :** A corpus of texts of many types. It may include written or spoken language, or both, and may include texts produced in one country or many.
3. **Comparable Corpus :** Two (or more) corpora in different languages (e.g. English and Spanish) or in different varities of a language (e.g. Indian English and Canadian English).
4. **Parallel Corpus :** Two (or more) corpora in different languages, each contaiining texts that have translated from one language into the other.

5. **Learner Corpus :** A collection of texts –esays, for example-produced by learners of a language.

6. **Pedagogic Corpus** : A corpus consisting of the language a learner has been exposed to.

7. **Historical or Diachronic Corpus :** A corpus of texts from different periods of time.

8. **Monitor Corpus :** A corpus designed to track current changes in a language.

The fifth type is the focus of the present study. Granger (2009, p14) defined learner corpora as ''electronic collections of foreign or second language learner texts assembled according to explicit design criteria.'' The nature of the language produced by learners compared to that of the native speaker necessiate from the analyst of the above mentioned corpora more caution than that of native corpora.

In addition, a corpus informs us more about how language is used than a native speaker can. Thus, it is mandatory in language classes. Another important element in corpus analysis is the use concordance programmes such as AntConc. Huntson (2002, p 39) defined concordance as:

> Program that searches a corpus for a selected word or phrase and present every instance of that word or phrase in the centre of the computer screen, with words that come before and after it to the left and right. The selected word, appearing in the centre of the screen, is known as the node word.

Figure N°1 Screen Shot of best+noun Collocation

Hit	KWIC	File
2	traditional places it is considered as the best destination. In the North of Sahara and	E.10.txt
4	whenever I needed her she is the best example of a good friend. What I	E.19.txt
14	between the City-filled concert, and The best place to relax and enjoy the scenery.	E.07.txt
18	drinking tea, for them this is the best time ever. It has such beautiful sights	E.06.txt

AntConc programme will be used to extract the most frequent adjectives and adjective-noun collocations in the corpus.

2.2 Classification of Collocation

Collocation is a fundamental component for an idiomatic use of language. Therefore, "The fewer collocations students are able to use, the more they have to use longer expressions with much more grammaticalisation" (Lewis, 2000 p 16)**.** The term colocation was first introduced by Plamer (1933), then Firth (1957) elaborated it more. Different classifications of collocation were proposed by researchers such Sinclair (1991), Benson et al (1997), Howarth (1998a), to name but few.

To begin with, Sinclair (1991) classified collocations into: (1) the upward collocations and (2) the downward collocations. Benson et al (1997) classified collocations into: grammatical and lexical .The first category contains eight main patterns while the second contains only seven patterns. Contrary to grammatical collocations, lexical collocations do not contain function words such as preposition and /or infinitive. Howarth (1998b) categorised collocations into : (a) free collocations, (b) restricted collocations, (c) figurative idioms, and (d) pure idioms. Moreover, Biskup(1992), Farghal & Obediedat (1995), Granger(1998), and Howarth (1998b) claimed that restricted collocations is the most important category to teach and/or learn. This pattern falls between free combinations and pure idioms on the collocation continuum.

Based on the norm of collocational strength, Hill (2000, pp 63-64) divided collocations into: (a) unique collocations, (b) strong collocations, (c) weak collocations, and (d) medium strength collocations. He stressed that the forth type is the ''the most important for the classroom''. Medium strength collocations are of supreme importance in "expanding learners' mental lexicons".

2.3 Factors Influencing Learners' Production of Collocations

Huang (2001) stated that a handful of studies indentified several factors that influence students' production of collocation. The first factor concerns the semantic field, meaning boundaries, and collocational restrictions of the lexicon (Biskup, 1992 ; Lennon, 1996).These three aspects of the lexicon may ease or complicate the process of collocation production for the students. The second factor concerns L1 transfer both positive and negative.(Teliya et all,1998 & Granger, 1998).Students tend to transfer collocations from their native language into the second or foreign language they are learning. Such a process   may sound

strange depending on the similarities and differences between the two languages. The third factor concerns individual learners' collocational competence (Bahns & Eldaw, 1993 ; Farghal & Obediedat, 1995 ; Granger, 1998, & Howarth, 1998).

3. Methodology

The corpus of the present study consists of 30 descriptive essays written by second year licence students of English at Kasdi Merbah University of Ouargla-Algeria as a classroom assignment. All gathered essays were then converted into text plain format using AntFile Convertor 3.4.4 programme (2014) for ease of computational processing. Titles and quotations were deleted, if any. Howarth (1996) observes that learner writing, especially academic writing is "adulterated"(p140). This feature, 'adulterated', is best explained in Lesniewska's (2006) words "the learner is likely to draw on a range of phrases and expressions which occur in the sources used" (p 102). Table 1 below summarises the collection and analysis procedures.

Table N°1. Collection and Analysis Procedures of the Corpus

| Step	Process
Step 1	Collection of the essays
Step 2	Coding the essays from E.01 to E.30
Step 3	Deletion of the quotations
Step 4	Conversion of the essays into txt format(AntFile 2014)
Step 5	Automatic generation of frequency lists (AntConc 2014)
Step 6	Semi-manual extraction of lexical collocations in the learner corpus
Step 7	Manual checking of the collocational structures in three dictionaries (BBI Combinatory Dictionary/ Oxford Collocation Dictionary/ Longman Collocations Dictionary and Thesaurus)
Step 8	Automatic checking of collocations in British National Corpus (BNC) and Louvain Corpus Native English Essays (LOCNESS)

Step 9	Experts' judgement of the possible sources of collocational errors

4. Results Analysis

Figure 2 shows the ten most frequent adjectives in the corpus. The presence of these adjectives and their collocates and the absence of others can be attributed to the influence of essays prompts, describing a place, a festival, and a person in the present study, on students' choice of words. Takac & Likac (2013, p 390) argued that "different topics and genres of texts influence lexical choice-which accounts for the presence or absence of lexical collocations."

The high number of correct collocations (59%) in the corpus can be best explained by Mahmoud's argument (2005, p121) "the correctly produced colocations could have been acquired through exposure to the language or they might have been positively transferred from Arabic".The results showed that 75 % of the errors are collocator-based, 13% collocate-based, and 9% both collocator and collocate. See figure 3. Examples 1, 2, and 3 show this, respectively.

**Figure N°2.Top Ten Frequent Adjectives in the Corpus**
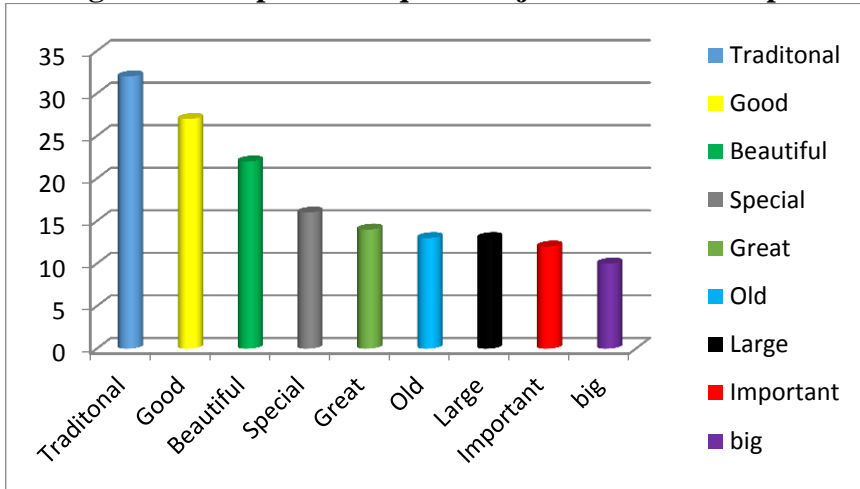
e.g.1. In Touggourt there is no racism where you find *__great combination__ of different kinds of people living together. **(ideal combination)**

e.g.2…., but women like to wear the ***traditional clothes** eg. " Elmelhfa. Elkoftan . Elfargani" while the bride mixes between the two. **(traditional dress)**

e.g.3 ……, and also ***<u>traditional places</u>** related to town edges to remember their ancestors.   **(historical landmarks)(معالم تاريخية)**

These errors can be attributed to : (a) 57% Negative transfer either from Modern Standard Arabic (henceforth, MSA) or Non Standard Arabic (henceforth, NSA), (b) 22% synonymy, and (c) 18% approximation strategy. See figure 4. In the example 3, the student wrongly opted for the combination *<u>**traditional places**</u> instead of the collocation **historical landmarks.**The latter

**Figure N°3.Distribution of Types of Collocational Errors in the Corpus**
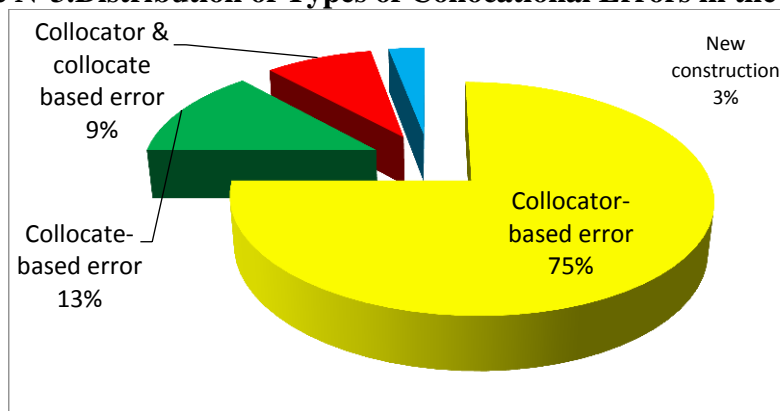
is undisputedly a transfer from Standard Arabic **(*اماكن قديمة)** instead of (**معالم تاريخية)**. An other strategy that students tend to apply is synonym.In example 2, the student chose the noun **clothes** instead of **dress** believing their are the same while they are not. L1 transfer occurs in learners' production of restricted collocations due to their presumption that there exists a one–to-one correspodance between L1 and L2. A worth discussing finding is that students transfer not only from MSA, but also NSA. Such finding supports Mahmoud's (2005, p122) statement "In the case of Arabic-speaking students, there are two varieties of Arabic from which they can transfer : Modern Standard Arabic (MSA) and Non-Standard Arabic (NSA)". Example 4 below shows this clearly.

e.g.4 Also , we **<u>bring a special chef</u>*** to cook the food and men who do everything except "the cleaning "(=njibou tabakh khas)

In addition to all of what is discussed above, one of the potential sources of collocational errors is the incomplete knowledge of collocation in L1 which possibly lead to an error in L2 collocation production.

Figure N° 4. Distribution of Sources of Collocational Errors in the Corpus
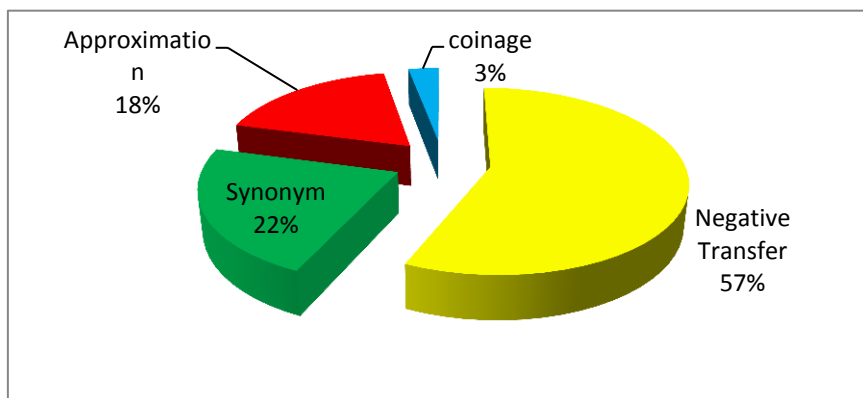
5. CONCLUSION

To sum up, the present study investigated the use of adjective-noun collocations in 30 descriptive essays by Algerian students of English. The results showed that, although the students are able to produce correct collocational structures, they produce erroneous ones too. These errors can be due to the negative influence of native tongue along with other strategies such synonym approximation. This study recommends raising students' awareness of collocations in writing classes. Furthermore, teachers of written expression and comprehension course should integrate corpora into their classes when teaching collocations. The latter can be done through corpus –based activities.

6. Bibliography List:

Anthony, L. (2014). AntConc (Version 3.4.4) [Computer Software]. Tokyo, Japan: Waseda University. Retrieved from https: //www. Laurence anthony.net/software

Anthony, L. (2014). AntFileConverter (Version 3.4.4) [Computer Software]. Tokyo, Japan: Waseda University. Retrieved from https: //www. laurenceanthony. net /software

Bahns J., & Eldaw, M. (1993). Should We Teach EFL Students Collocations? *System*, 21(1), 101–114.

Benson, M., Benson, E., & Ilson, R. (1997). *The BBI Combinatory Dictionary of English: A Guide to Word Combinations*. Amsterdam: John Benjamins.

Biskup, D. (1992). L1 Influence on Learners' Renderings of English Collocations. A Polish/German Empirical Study. In P. J. L. Arnaud, &

H. Béjoint (Eds.). *Vocabulary and Applied Linguistics* (pp 85–93). London: Macmillan.

Cowie, A. (Ed.) (1998). *Phraseology: Theory, Analysis, and Applications*. Oxford: Oxford University Press.

Farghal, M. & Obiedat, H. (1995). Collocations: A Neglected Variable in EFL Writings. *International Review of Applied Linguistics in Language Teaching,* (33), 315– 331

Firth, J. R. (1957). A Synopsis of Linguistic Theory 1930-1955. *Studies in Linguistic Analysis. Special Volume. Philological Society*: 1-32.

Granger, S. (1998). Prefabricated Patterns in Advanced EFL Writing: Collocations and Formulae. In A. Cowie (Ed.), *Phraseology: Theory, Analysis and Applications* (pp 145–160). Oxford: Oxford University Press.

Granger, S. (1998). The Computer Learner Corpus: A Versatile New Source of Data for SLA research. In S. Granger, (Ed.). *Learner English on Computer.* (pp 3-18).London & New York: Addison Wesley Longman.

Hill, J. (2000). Revising Priorities: From Grammatical Failure to Collocational Success. In M. Lewis, (Ed.). *Teaching Collocation: Further Developments in the Lexical Approach* (pp. 47-69). England: Language Teaching Publications.

Huang, L. (2001). Knowledge of English Collocations: An Analysis of Taiwanese EFL Learners. In C. Luke, & B. Rubrecht, (Eds.). *Texas Papers in Foreign Language Education: Selected Proceedings from the Texas Foreign Language Education Conference,* (6), 113-132

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

Howarth, P. (1996). Phraseology in English Academic Writing: Some Implications for Language Learning and Dictionary Making. Niemeyer:Tübingen

Howarth, P. (1998a). The Phraseology of Learners' Academic Writing. In A. Cowie (Ed.), *Phraseology: Theory, Analysis and Applications* (pp. 161–186). Oxford: Oxford University Press.

Howarth, P. (1998b). Phraseology and Second Language Proficiency. *Applied Linguistics,* 19 (1), 24-44

Lennon, P. (1996). Getting ''easy'' Verbs Wrong at the Advanced Level. *IRAL*, (34),11-30

Lesniewska, J (2006). Collocations and Second Language Use. *Studia Linguistica* (123), 95-105

Lewis, M. (Ed.) (2000). Teaching Collocation: Further Developments in the Lexical Approach. England: Language Teaching Publications.

Mahmoud, A. (2005). Collocation Errors Made by Arab Learners of English. *Asian EFL Journal,* (5), 117-126

McIntosh, C, Francis, B & Poole, R. (Eds) (2002). *Oxford Collocations Dictionary for Students of English. Oxford*: Oxford University Press.

Palmer, H. (1933). *Second Interim Report on English Collocations*. Tokyo: Kaitakusha.

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

Takac, V.P & Lukac, M. (2013). How Word Choice Matters: An Analysis of Adjective-Noun Collocation in a Corpus of Learner Essays. *Jesikoslovlje* 14 (2-3), 385-402.

Teliya, V, Bragina, N Oparina, E & Sandomirskaya, I. (1998). Phraseoly as a Language of Culture : Its role in the Representation of a Collocative Mentality. In A. Cowie (Ed.). *Phraseology: Theory, Analysis and Applications* (pp. 23–54). Oxford: Oxford University Press.