Oct 28th, 8:00 AM - Oct 29th, 6:00 PM

Introducing Corpus Linguistics as a novelty to English teachers and learners: current language practices of tools, corpora and online resources

Juliana Çyfeku
"Fan S. Noli" University Faculty of Education and Philology Department of Foreign Languages Korce, Albania, jcyfeku@unkorce.edu.al

Venue: UBT
Event: 12th International Conference on Language and Culture
Thematic Issue - *WORKSHOP*
Duration: 60 minutes

Topic: *Introducing Corpus Linguistics as a novelty to English teachers and learners: current language practices of tools, corpora and online resources*

KEYNOTE SPEAKER: **Dr. Juliana Çyfeku**
jcyfeku@unkorce.edu.al
"Fan S. Noli" University
Faculty of Education and Philology
Department of Foreign Languages
Korcë, Albania

Abstract

The present paper represents how Corpus Linguistics has developed and is employed as a methodology in foreign language teaching and learning. It highlights its many relevant, innovative, creative and engaging applications to both axes of the linear process of teaching and learning. It bridges theory into practice so that learners understand the importance of this discipline and are willing to apply corpus-based tools in the language learning classroom. It introduces among a range of online resources Compleat Lextutor (Cobb, 2000) pointing out the most practical, dynamic, easy-to-use and accessible tools such as concordance, hypertext, frequency and vocabprofile. They can be utilized either in a computer lab or out of class and are intended for a range of learners.

Provided the flexibility of the corpus-driven approach students can autonomously use the databases and choose among the tools in dependency to their proficiency level. They can explore, generate or design materials, tasks and activities that can eventually inform or guide towards Data Driven Learning (DDL) on the web and consequently yield in reinforcing vocabulary. Moreover, the present work paper conceptualizes the corpus-driven approach dissemination accomplished under the language educator's instructional guidance with a mere focus towards student learner-oriented autonomous conduct. Alongside, they are addressed to exploit the numerous site's video tutorials; utilize the selected tool/s in understanding the authentic language use as displayed in corpora lines; explore and address linguistic patterns of use, grammatical co-occurrence patterns so that gain consistency of native-like use of the genuine patterns of language use.

Key words: Corpus Linguistics, corpus, tools (concordance, hypertext, frequency, vocabfrofile), online resources (Compleat Lextutor, Cobb 2000), DDL.

1. The rationale

It is time that learning foreign languages becomes a process that indulges learners to learn effortlessly, captivating knowledge by scrolling authentic input in real time, diverse contexts and evolving technological devises. That is made possible throughout Corpus Linguistics (CL), a

new methodology in foreign language teaching and learning. It provides a vehicle for bringing natural language into the classroom in a way that evolves learners through interactive hands-on activities interacting with the 'real' language (Reppen, 2010). Originated with the advent of powerful use of computers in the late 60s, CL served the scope of storing electronically governmental data in corpus databases to later be used in storing educational documents.

CL is specifically defined as a research methodology with a mere focus on empirical investigation of language use and its contextual and geographical variation, providing a plethora of results that have much greater generalizability and validity in feasible natural occurrences (Biber, Reppen & Friginal, 2010). Corpora (referring to enormous bodies of electronically stored texts) and CL have supported the view that language is systematic and as such can be spotted in corpora lines, described, analyzed and taught by employing empirical, frequency- and pattern-based approaches (Friginal, 2018). Finding an immense worldwide application to the teaching of English as a foreign language, this new methodology offers learners relevant and meaningful data-frequency implications and distributions of actual patterns of English vocabulary and grammar structures as they are precisely used in natural occurrences in oral and written discourse. It relies on real-world (linguistic) data supporting so many interrelated theories in gaining satisfactorily outcomes and achievements especially in English language learning and teaching. From the Latin word for "body", the word *corpus* (corpora, plural) has been used to refer to a large collection of *authentic texts* stored electronically on a computer. *Corpus Linguistics* is the study of language using corpora. It is not a theory of language. Corpus Linguistics is a tool/method to study language and investigate linguistic phenomena. Specifically, *'language'* refers to the naturally occurring examples in spoken or written way supporting what Sinclair (1991) defines as *'genuine communication of people going about their normal business'*.

Corpus Linguistics is an extremely versatile method with a whole range of applications in social science research, the digital humanities, as well as in practical areas such as marketing, journalism, language learning and teaching, language testing, textbook writing and so on.

A corpus is a principled collection of texts underlying a planned and intended operation not a randomly collection of texts. Researchers and corpus compilers have a research question on their focus, prior to the process of compiling it or conducting research through it. These electronic texts are equivalent to researcher databases. In linguistics, a corpus is defined as a collection of information, done in a systematic way consisting of naturally occurring categories of texts. This process of storing language information was done manually prior the age of computers. Later on with the advent of personal computers and the digitalization of much of our everyday spoken and written language, CL has become a practical, accessible and innovative approach to examining languages and their evolving use. A linguistic corpus is, by definition, computerized and searchable by computer programs (Friginal & Hardy, 2014 a). The following definitions elicited from seminal publications from various corpus linguists provide us with a clear view of the importance of applying corpora in language teaching and learning:

> A corpus is 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.
>
> *(Sinclair, 2005)*

> A corpus can be briefly defined as systematically designed electronic collection of naturally occurring texts […]. Researchers compile corpora and search for existing constructs of written or speech patterns identified as relevant and measurable. A corpus provides the opportunity to measure tendencies and distributions across registers (and genres) of language.
>
> *(Friginal et al., 2017)*

Several ESP and EAP studies report that students using corpora, corpus tools and corpus-based materials and literature make them effective in awareness-raising exercises where specific characteristics of spoken and written registers are noticed for self-correction and additional motivation in search of language misuse turning the latter into a better use of language production (Gavioli & Aston, 2001). Classroom activities and tasks in association to these types of language learning exercises can be generated under the pitfalls of Corpus Linguistics as an approach evolving within instructional technology. The latter comprises the use of elements such as internet, a computer, laptop or Ipad, and free accessible corpora, used to support teaching and learning of a particular subject, aspect or issue. Corpus approaches eventually merge innovations in instructional technology and educational computing with a range of computational tools as well as promote the inclusion of different perspectives on language learning inside and outside the language classroom. In today's global advancement in technology in all human endeavours mobile technology, individualized instruction and data visualization are estimated as complimentary constituents of CL where digital learners may easily fit, get involved by adopting and appreciating corpus-based approaches to genuinely learn English.

Among a range of CL tools and softwares, the present study introduces the most widely used software in learning languages where English is considered as a mediator to transfer and/or transmit information/thought from one language to another and that is *Compleat Lexical Tutor* (Cobb, 2000). It represents a true armoury of corpus- and frequency-based tools for English and French language learners and teachers, and researchers in linguistics. Since vocabulary learning is a central concern in most language classrooms, the overwhelming focus of the tools is on lexical development, English for Academic Purposes (EAP), and Computer Assisted Language Learning (CALL). The use of this software and its variety of tools (Figure 1.) such as: Concordance, Frequency, Vocabprofile, Hypertext, Corpus Grammar, Concord Writer…etc. LexTutor brings together a multiplicity of research and a plethora of over 25 corpora.

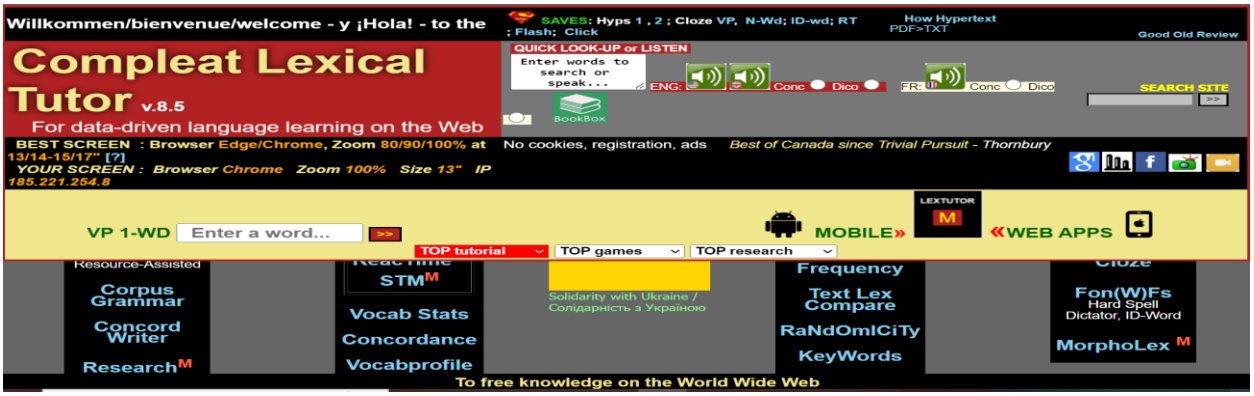


Figure. 1 Compleat Lexical Tutor Interface (Cobb, 2010)

As observed from the site's interface, vocabulary is the target language feature that corpora deal with. Provided the fact that English Foreign Language (EFL) learners strive hard to enhance and be native-like apprehending and using English, Compleat Lexical Tutor has reported to be a rich resource of vast amount of genuine linguistic data. Being fond of digital devices and guided from the curiosity of tackling information in all domains of human development, it fosters to the sharp learner's eye that learning can be fast, immense, fun, motivating and challenging through the main approach that prevails the software programme that is data-driven learning (DDL) on the web. The developer and designer of this website, Tom Cobb, Associate Professor at the Department of Didactics of Languages at Quebec University, Montreal Canada, maintains and updates this website which is consulted by 1500+ learners, teachers, and researchers every weekday worldwide and is regularly cited in research publications and conferences in the research area of Corpus Linguistics, Computational Linguistics, Applied Linguistics, Lexicography, Stylistics, Translation studies and many others.

2. Methodology

The present study employs corpus tools which refer to specific electronic interfaces that help language learners find linguistic patterns in corpus texts. The typical approach that makes use of the existence and compilation of online corpora is Data Driven Learning (DDL) on the web. This approach was pioneered and the term coined by Johns (1991). It is an approach to foreign language learning that allows learners, teachers and researchers to use data independently so that they undertake guided discovery tasks. The impetus in relevance to this pedagogical approach relies upon the *information-knowledge* data paradigm that equips the interested on such pattern-based and/or pattern-driven approach to grammar and vocabulary as well as a lexico-grammatical approach to genuine foreign language use. So as Tognini & Bonelli (2001:84) explicate "descriptions aim to be comprehensive with respect to corpus evidence". In addition Biber (2009: 279) further asserts that [… pattern grammar studies attempt to uncover new linguistic constructs- the patterns- through inductive analysis of corpora].Indeed the pattern grammar studies are instructive since they allow learners to note the lexical associations of each discovered word pattern through corpus analysis. Research in the field, among which spotting out a 2015 Thomas task-based study, acknowledges John's approach and reconfirms the data-driven principle of that he labels as *'cutting out the middle man'* explicating so the direct learning and study of language from the instant use/s of language as viewed from the corpora lines. Besides, it revolutionizes traditional learning from resources such as sourcebooks, dictionaries, grammars, manuals ...etc.

Essentially, DDL is a consistent approach in Second Language Acquisition that embraces the various research remarks elaborated from the theory of language hypothesis and usage-based learning (UBL). On support of the noticing hypothesis theory, Friginal (2018) states:

> [… a language learner needs to notice some feature of input to have the best chance of acquiring it. DDL works by drawing the learner's attention to the target feature by providing numerous of examples in context. One of the tenets of UBL is that we learn language through analyses of massive amounts of form-function pairings in authentic contexts.]
>
> p.224

So the language learner has the ownership to formulate a statement, put forward a research question, test hypothesis and elicit evidence from the corpus-data to answer and clarify the sceptical views, relying so on his/her intelligence, knowledge and intuition to give correct answers. In reality corpora are created so that learners bridge the gap between learners' current language ability and the expected ability.

2.1 Compleat Lexical Tutor online tools: Concordance, Frequency, Hypertext, Vocabprofile

Concordance and Frequency as seen from Figure 1, refer to a piece of software that can be found and instantly used online form the interface of the originated website of Cobb (2000) or can be installed on a computer to search all kinds of language across different genres or targeted text type. *Frequency* is the base measure to examine whether the targeted word is frequent or infrequent in language use. A Frequency Indexer is a tool used to the frequency of the targeted search word in a text, or is used to compile a frequency list displayed in a range starting from the most frequent to the least frequent word. It gives additional computational information on the percentage of each word and how many times it occurred in a text.
Provided the enormous language examples, DDL approach fosters vocabulary enhancement where learners independently investigate to find out what insights to get beyond the concordance lines. *Concordance programs* create word lists that can be arranged in either alphabetical order or in word frequency order and that in dependence to what lexical significance language searchers aim to elicit. They can be utilized to identify the different usages and *frequency data* of a content word (noun, verb, pronoun, adjective, and adverb), examine word collocations, and explore key words in context indexes to precisely define the semantic meaning/s that is rendered from the genuine source of occurrence/s.
Hypertext basically links a text with various furnished activities such as linking to a dictionary interface, providing examples of language use and images for comprehension or speech discourse. It provides adequate space to search, acquire, investigate, process and assess vocabulary use with a mere target on academic texts especially in reading and vocabulary learning acquisition. Cobb's video tutorial remarks point out that academic success is largely a matter of reading dealing with vocabulary, and the latter when handled with a specific computational approach assists learners to view and recall vocabulary use in additional contexts of future use.
Vocabprofile is a frequency-based analytical tool that sorts the vocabulary of a text into four categories: K1, K2, AWL, or Non-list. Precisely a K1 is a list of the first thousand most frequent words in English, representing about 85% of the speech; K2 is a list of the second thousand; the AWL is an academic word list; and non-list words which consist of words that are not in the lists which may be infrequent or non-academic. Lextutor's tools focus on categorizing words by relationships and frequency. In addition the output from Vocabprofile tool classifies words by lists and also shows 'type-token ratio 'analysis by list and family. This measure indicates the number of different words in the text (types) divided by the number of words on which they are based (tokens). So that learners increase the variety of words used specifically in writing tasks, they are encouraged by a higher type-token ratio number.

3.Implementing corpus web-based tools for English language learning and teaching

The scope of creating these online software tools is not using them separately, since they are intertwined in various activities. As a start you need to know how to operate with each of them, and what novelties and specifics each brings to use as explained in the methodological section of the present study. The most interesting part by using *Compleat Lexical Tutor* website relies on the fact that the creator himself, Tom Cobb provides plenty of free video tutorials for the site implication and the tools it consists of. He eventually updates information through various activities and lesson plans on how to implement the site's novelties in our daily EFL classrooms. Besides there is launched plenty of evidenced video operations how worldwide practitioners, outstanding Tom Cobb, Tony McEnery, and Tatyana Karpenko-Seccome, who evidence online tasks and lessons effectively implemented in their English teaching and learning practices (videos easily found on google. The site itself as a start, gives plenty of abrupt info to check the meaning of unknown words at any time when encountered in unknown contexts either when choosing online reading texts or while using any methodological course books in English classes .

- ❖ Using this site you can generate tasks and activities for vocabulary acquisition in terms of spelling, pronunciation, semantic meanings, contextualization, image comprehension and other additional readings for future contextual inferences and applications (see Figure 2).
- ❖ You can observe simple phonetic, morphological, semantic, etymological, pattern and collocation vocabulary searches *vs* complex book dictionary entries.

Specifically by just typing a word at the Compleat Lexical Tutor interface search box (on the top right side of the interface) you:

- Get the prompt pronunciation of the targeted word either in American or British English;
- Search for the English definition;
- Can find additional linguistic information on the word's etymology;
- Find the word's reference in various dictionaries searching all at once;
- View the word's usage in various parts of speech;
- Expand the vocabulary by searching for the word's synonyms;
- Learn the word's function and meaning/s through online image/s;
- Enhance a deeper understanding of text exploitation by getting the word's collocations throughout the online tool – Concordance.

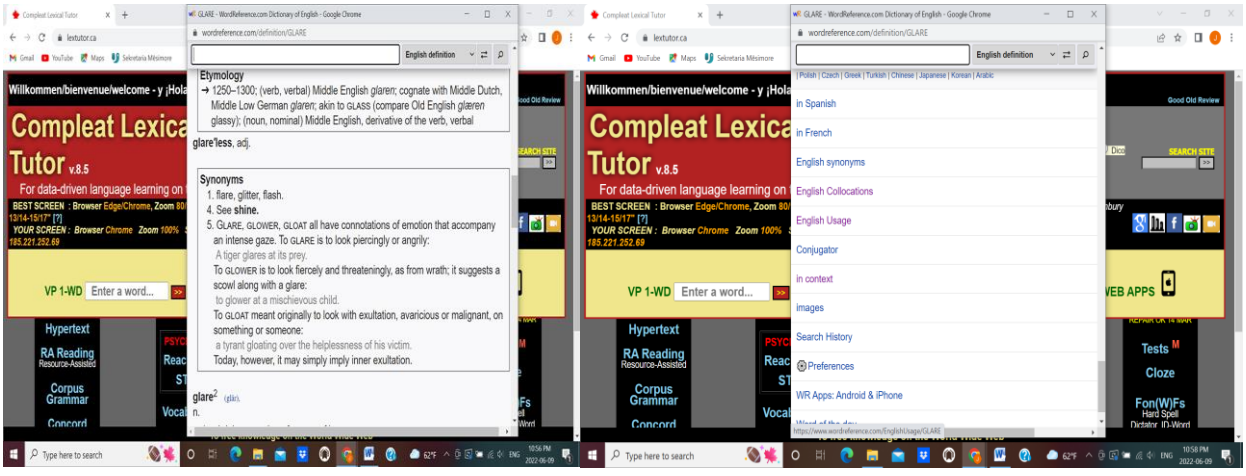


Figure. 2 Etymology & synonym word search entry **Figure. 3** Complex Word search entries

3.1 Hypertext and Concordance

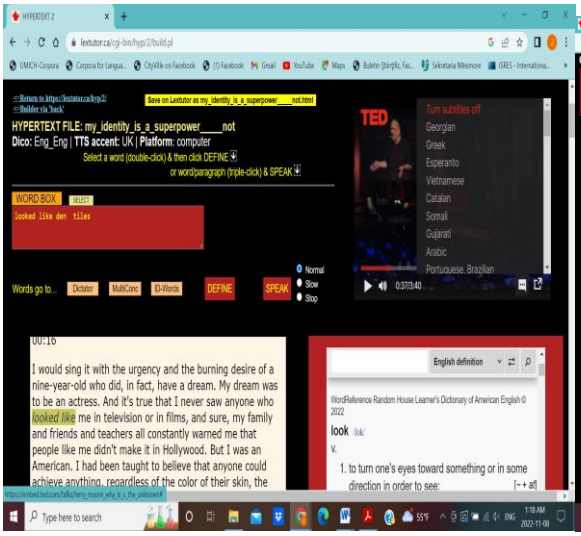
Throughout the use of Hypertext teachers can choose a text (fiction, film, newspaper, magazine, science, literature, sport, entertainment..etc) and further manipulate data for linguistic purposes on a concordancer to get a deeper text analysis. These tools give EFL students the ability to search, access and analyze new vocabulary when encountered in their course books; or in cases when linking an online text from various resources in diverse subject areas.

Once you click the Hypertext tool you get its interface with three additional vocabulary activities serving the scope to test, reinforce and check vocabulary such as:

- ✓ to send a word as a Dictator
- ✓ to send the word to Concordance
- ✓ to find word in a jumble ID-word search box

Proceed by inserting the reading or listening text in, (if willing to analyze a TV show, radio broadcasting.., etc), which is linked to dictionary and speech, choose a dictionary (French-English; English-English, English- German…etc); choose TTS (Text-to-Speech) Voice: US, BT, French (pronunciation) that is used from students for multi-modal learning; and then click the button 'BUILD'. As noted from Figure 4, the text displays in the search box, to later operate by clicking on the unknown words or the difficult ones to get a profound dictionary use of them.

Once completing the dictionary word comprehension activity, manipulate the data using the three other tools to yield into the selected words' dissemination by practicing dictation; spelling the words you hear by practicing listening and punctuation; use of words across concordance (Figure 5) lines (grasp the word use in the blank spaces); or you may test the word's memorization by building the online activity I-D word builder (a query to find the word in the jumble matching both the form and the meaning of the word). The intertwining of *Hypertext* and *Concordance* display in forms of descriptive activities and amusing check-up games by using the targeted words multiple times out of a corpus so that students as sharp observers against what technology offers end up noting and as a result enhancing the word use in numerous contexts. Noticeably, the following figures give a panorama of how EFL lessons report to be interactive, challenging, motivating, making a difference from the daily practice routine. By means of digital gadgets, online resources, educational websites and their profiled tools both senses of the axe can complement one another to yield to satisfactory results in teaching and learning English.



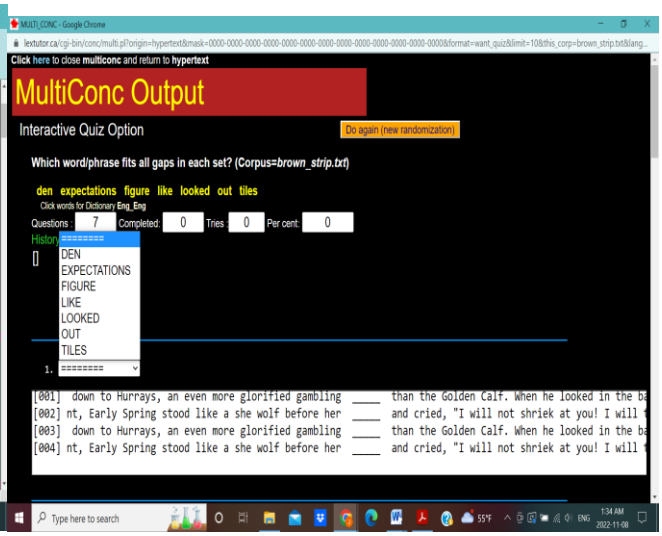


Figure. 4 Interface of Hypertext builder

Figure.5 MultiConc Interactive Quizz Interface

So the highlighted words from the text indicate to be the words that are being examined into specific contexts. A small concordance is given for each of the highlighted word which applies to a range of texts by requiring so the students' attention to find out from the data-driven question which one from the selected words applies to all texts. Recalling so grammatical, structural and semantic knowledge they decide which words go together and which not. The program displays such meaningful input as it urges by gamification for the students' attention to get the valuable insights of the *word's pattern use* in both sides before and after the target word, and afterwards test their skills in English vocabulary. Indeed the scope of presenting the Compleat Lexical Tutor reveals what Karpenko-Seccombe (2021) estimates her students' behaviour towards corpora and concordancers in [… fun to use, engaging and enlightening]. She concludes that these easy-to-use, open-source, accessible corpus-tools assist them work independently and consequently they gain a greater control of the language learning process.

3.2 Vocabprofile and Frequency

Lextutor Vocabprofile has been used as tool in foreign language learning by many researchers. Sapa-asa (2006) reports the website to be an effective way to verify the level of a reading text difficulty and consequently improve and facilitate English vocabulary teaching and learning. It provides an insight on students' lexical richness, on the level of the words giving clues whether they use common or more specific words. Language teachers often step when dealing with reading and text exploitation directing so the possible questions: How hard this text might be for my students? How can I test whether the text I intent to use in the EFL class is in the correct level of my students? Relevantly, what you need to get the instant answers is using the online Vocabprofile to run a profile of the targeted text. In this way you get a clearer text analysis framework prior to conducting tasks to the EFL students. To elaborate results and obtain efficiency against these variables the present study gives explicit information on:

- ✓ How this online tool works;
- ✓ How to analyze the search results;
- ✓ How to operate with the targeted text.

Vocabprofile is a measure of the proportions of low and high-frequency vocabulary used in a written text. It gives information in percentages of how frequent or infrequent in use the vocabulary in the text is. After the text is chosen for vocabulary analysis, copy and paste it in the tool's search box, choose VP classic and click on the Submit box to get the instant results which pursue as the following. *The text is divided into four categories by Frequency:* 1) the most frequent 1000 words of English (level K1), 2) the second most frequent thousand words of English i.e. 1001 to 2000 (level K2), 3) the AWL (Academic Word List, the 750 words that are frequent in academic texts across subjects, 4) and off-list words that are not found on the other lists. The later may include proper nouns, unusual words, specialist vocabulary, acronyms, abbreviations and misspellings. After submitting the text, the search results obtained in percentages show how many times a word occurs in the text and further displays in certain computational figures in different colours representing: blue the K1, green the K2, yellow the AWL, and red off-list words (Figure 6).

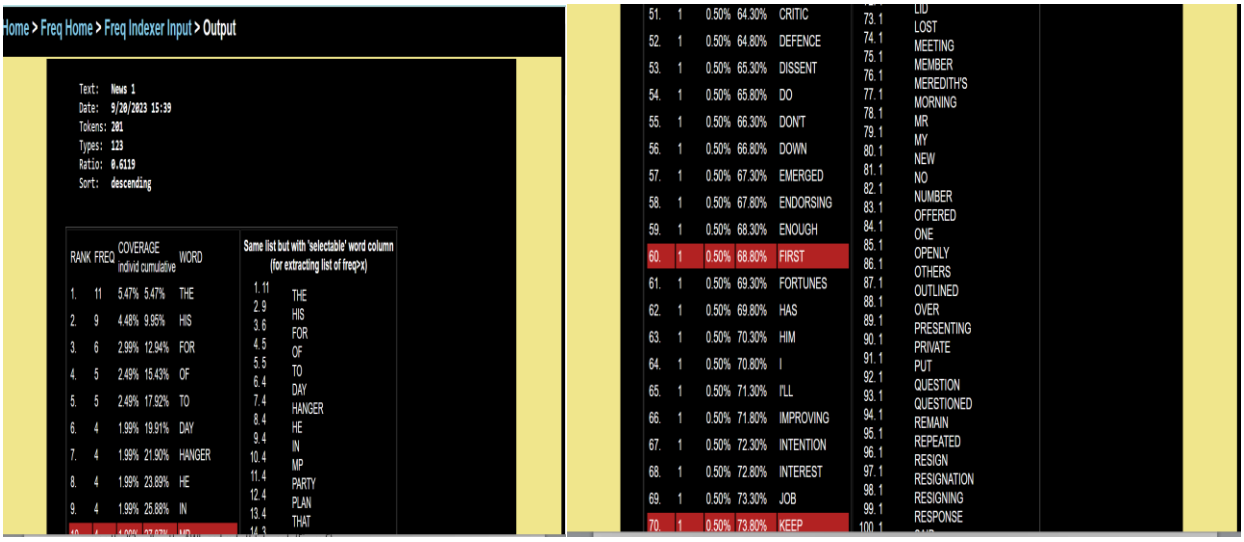


Figure. 6 Frequency Indexer Input

As the computational data show elicited from a 'News' text; the number of tokens is 201 referring to the number of total word counts, the number of types is 123 referring to the number of different words, ratio refers to the level that students can cope with and as research shows it should be less than 5%. In our case the first K1 words show that are easier for them to be reinforced and distinguished into texts; whereas K6, K7, K8 …and so on where the coverage surpasses the ratio of 5% denotes that the words are less frequent in use at the academic level and as such they require students' attention for a better use of them in other formal contexts.

An investigation from McCrostie (2007a) concludes that "Japanese learners have difficulty in identifying high frequency words … and view all words they do not know equally important." He necessarily demands that teachers assist and instruct their learners how to distinguish between these two types of emphatic and most frequently used words. They can simply refer to the Frequency Indexer, a tool that provides even a range of frequency tests in forms of games by the end of which they get the instant scores serving so as a good strategy for word choice in terms of frequency intuition.

4. Results, Conclusion and Refinement

The main focus on the present study concerns vocabulary instruction in a contemporary, challenging, real and virtual learning environment (classroom, computer lab). The knowledge of knowing, using and applying a word goes beyond its dictionary definition as it requires what Corino & Onesti (2019) describe as word's categories: spelling (Phonolology), Parts of Speech (POS), morphology, variant spellings (Semantics), collocations (Phraseology), and specific uses across registers and genres (Stylistics). The implementation of corpora, Compleat Lexical Tutor website and its most relevant, easy- and friendly- to use tools assist all kinds of learner-teacher-researcher commitments. It aims to facilitate the language learning process throughout the grasp of: vast lexical information, patterns of textualization, having an impetus on genre-structuring features of words, the intertwining of tasks and activities leading so to guided *vs* autonomous *interaction* in EFL classroom. Subsequently, what Leech (1997) confirms is that by

learning to interact with corpora […students find themselves learning a great deal about language, and how to study language. They learn about the kind of questions that can be usefully put forward and answered by reference to corpus data p.23].

Implementing the free and easily accessible website Compleat Lexical Tutor, designed from the associate professor Cobb (2000), and provided that it is yearly updated can be effectively achieved in our ELT classrooms by the use of only some laboratory equipment such as: Internet, a digital gadget (PC/ laptop, pad). The present study was implemented at academic level with ELT students. The appeal to the study encapsulates the framework of this setting as addresses the need to be implemented at various pre-university levels in EFL education. Provided that all educational institutions possess a modern infrastructure to transmit knowledge to learners, it is time that computer laboratories are used for foreign language learning as well. Utilizing Compleat Lexical Tutor software, choosing Concordance, Frequency, Hypertext and Vocabprofile among a range of tools language urges practitioners and learners to notice that the web displays basic to advanced learner activities that suit their levels of proficiency. DDL approach stimulates both classroom linguistic inquiry and speculation on the part of the learner. Moreover it assists them to view genuine language in huge bodies of texts across different corpora to patterning of words and form generalizations to account for the pattern explanation. The teacher's role transforms into that of a guidance that highlights the student-autonomous initiated research. Corpus Linguistics and Corpora grounded on DDL approach fosters to the EFL learners what Rutherford (1987) labels as a new style of 'grammatical consciousness-raising' in which the learners themselves discover grammar through language exposure. In sum, the major impetus of the present study is to mirror to both foreign/English language practitioners and learners at all levels of pre-university and university education, *the opportunity* to view, to construct and to generate corpus-based tasks and activities by evolving DDL input. This motivating and novice path would no doubt lead to autonomy of language learning. It is time that are established concrete bridges of institutional cooperation between two levels of education (that of teaching and learning a foreign language), as well as the exchange of the lecturer's experience through teaching and scientific-research in the field of Corpus Linguistics. Resulting as a new methodology unknown for English teachers, the present study aims to transmit the novelty among colleagues, in all chains of the teaching and learning process, inside and outside the institution. The pre-requisites of the study favor among foreign language educators, student learners and researchers:

- Institutional connection and interactive cooperation between these knowledge institutions;
- Contemporary weaving of knowledge (academic context and pre-university context);
- Practicality of using contemporary methodologies in English foreign language learning, through the innovations it presents. (Website and Virtual Tools);
- Transition through digital learning towards the acquisition and reinforcement of English vocabulary learning;
- Design of 'materials, tasks, activities' (practical and interactive exercises) to reinforce, test and expand the vocabulary as defined in the Corpus Linguistics methodology itself.

References

Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. International Journal of Corpus Linguistics, 14(3), 279.

Biber, D. , Reppen, R. , & Friginal, E. (2010). Research in corpus linguistics. In R. B. Kaplan (Ed.), The Oxford handbook of applied linguistics (2nd ed., pp. 548â"570). Oxford, UK: Oxford University Press.

Cobb, T. (2000). The Compleat Lexical Tutor [Website], developed by Tom Cobb of University of Montreal, Canada (Cobb, http://www.lextutor.ca/conc/eng/).

Corino, E, & Onesti, C. (2019). DDL: A Scaffolding Methodology for CLIL and LPS. *Frontiers in Education, Vol 4, Article 7.* Retrieved from: www.frontiersin.org

Friginal, E. (2018). Corpus Linguistics for English Teachers: *New Tools, Online Resources, and Classroom Activities.* New York and London: Routledge, Taylor & Francis Group, 224.

Friginal, E., & Hardy, J.A. (2014a). *Corpus-based sociolinguistics: A guide for students.* New York, NY: Routledge.

Friginal, E., Lee, J., Polat, B., & Roberson, A. (2017). *Exploring English learner language using corpora: Learner talk.* London, UK: Palgrave Macmillan.

Gavioli, L., & Aston, G.(2001). Enriching reality: Language corpora in language pedagogy. *ELT Journal 55(3)*, 238-246.

Johns, T. (1991*). "Chapter 2: Should you be persuaded: Two examples of data-driven learning" (PDF).* Classroom Concordancing. Birmingham: ELR.

Karpenko-Seccombe, T. (2021). Academic Writing with Corpora: A resource Book for Data-Driven Learning: Routledge, Taylor & Francis Group, London and New York.

Leech, G. (1997). Teaching and language corpora: A convergence. In A.Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora (Applied Linguistics and Language Study)* (pp. 1-23). London, UK: Longman Publishing Group.

McCrostie, J. (2007a). Examining learner vocabulary notebooks. *ELT Journal* 61(3):246-255. Retrieved from: https://www.lextutor.ca/freq/train/mccrostie_07.pdf

Reppen, R. (2010). Using corpora in the language classroom. Cambridge: Cambridge University Press.

Sapa-asa, P. (2006). "An Analysis of Vocabulary Used in Essays Written by Upper Secondary School and University Students." Master Thesis of English Program, Faculty of Humanities. Bangkok: Srinakharinwirot University.

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

Sinclair, J. (2005). Corpus and text-basic principles. In M. Wynne (Ed.), Developing linguistic corpora: A guide to good practice (pp.1-16). Oxford, UK: Oxbow Books.

Thomas, J. (2015). *Discovering English with Sketch Engine.* p.22. VERSATILLE.

Rutherford, W. E., (1987) *Second language grammar: learning and teaching:* Longman

Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work.* Amsterdam/Philadelphia: John Benjamins, 84.