PAPER • OPEN ACCESS

The Application of Corpora on the Lexical Approach

To cite this article: Fei Yuan and Shichunzi Zhang 2021 *IOP Conf. Ser.: Earth Environ. Sci.* **693** 012058

View the article online for updates and enhancements.

The Application of Corpora on the Lexical Approach

Fei Yuan[1]* and Shichunzi Zhang[2]

[1]Nanchang Institute of Technology, Nanchang 330099, China
[2]Shanghai Urban Construction Vocational College, Shanghai 201415, China
Email: 615509812@qq.com

Abstract. The linguistic scholars has used word frequency for research for a long time. With the rising of corpora study, researchers pay a lot attention to words and their frequency information, and even develop the teaching approach and teaching book. The Lexical Syllabus, The Lexical Approach and Data-Driven Leaning introduced by this paper are the representative of the development of corpora. From the background of this research, this paper gives the definition of some relative concepts, and discusses the arrangement of lexical syllabus and teaching content, and uses examples to explain the practice of The Lexical Approach.

1. Introduction

The utilization of word frequency has a long history during language teaching development. Before the existing of computer in 1946, the frequency list had been produced based on large language samples which used to compile reference books and instruct language teaching. For example, Thorndike published *The Teacher's Word Book* in 1921 and *A Teacher's Word Book of 20000 Words* in 1932 for English language teaching; *The Production Vocabulary of Chinese* by Heqin Chen in 1928 also influenced Chinese language teaching a lot.

The focus of word frequency reflects the concern of the frequent linguistic unit and language information. On the early time, the function of frequency list was defining the demands of vocabulary teaching and reading & writing abilities at different stages. While in the English language teaching, sentence was mainly involved in the basic teaching unit for a long time which was affected by the trend of focusing on the study of syntax on linguistic field. However, since 1980s, with the beginning of corpora building, the boom of vocabulary and frequency study thrived. John Sinclair (1987) and his team from University of Birmingham started a research on lexical grammar based on COBUILD corpora since 1980, and developed a series reference books and textbooks.

On the above background, this research is going to describe the characteristics of English vocabulary grammar and develop a teaching approach with the help of frequency information supplied by corpora.

2. The Basic Concepts of The Lexical Approach

2.1 The View of Language Focusing on Lexical Item
The view of language focusing on lexical item means taking lexical item not syntactic item as the basic learning unit of language. Literature described the lexical items as following:
   a. Language consist of grammaticalized lexis, not lexicalized grammar.
   b. Grammar as structure is subordinate to lexis.
   c. The grammar/vocabulary dichotomy is invalid; much language consists of multi-word "chunks".
   d. Collocation is integrated as an organizing principle within syllabi. (Lewis 1993: vi)

e. English as a lexical language. (Willis 1990: 22)

f. …we define lexical phrases as form/function composites, lexico-grammatical units that occupy a position somewhere between the traditional pokes of lexicon and syntax…(Nattinger & DeCarrico 1992: 36)

The concept of lexical item is much larger than syntactic item. On the structure, lexical item concludes individual words and the conventionalized phrase. Conventionalized structure is the basic factor of fluent expression (Pawley & Syder 1983). Besides, discrete form is also a part of lexical item, such as "it is + adj. +that…". The formation of lexical item is flexible and open, which can be the minimum writing form—word, and the continuous word chunk or phrase, and also can be the non-continuous structure. They could work on the construction of bigger grammar unit that means the language is consisted by grammaticalized lexical items. These basic unit of language built by several words is in the majority, which is vital to language description and learning and used to be neglected in traditional language teaching. This bottomed-up view of language weaken the statue of sentence which was the original unit for language analysis, that's because the sentence is always the coherence of lexical items. Therefore, the conflict between word and sentence has been disappeared.

Lexical item has pragmatic function excepting the form features. Many lexical items have language style tendency, for example, word "kind" compared with "type" is used more in formal or academic style, which could be proved by searching BNC corpora of Brigham Young University built by Mark Davies. The subjunctive mood syntactic structure of "if I were you" was traditionally seen as a discrepancy of fact, but as a lexical item, it imply a suggesting verbal behaviour (Nattinger & DeCarrico 1992:13). Two examples illustrate that lexical item is not only a structure unit but also connect with particular language usage. These functional features uniting with lexical item like the two sides of a coin.

The relationship of a lexical item interior and between lexical items incarnates as collocation. In another word, high-level linguistic unit is the bond of low-level lexical items and the strength of collocation makes items to be a unity. This is the power source of language expanding.

2.2 The View of Learning of the Lexical Approach

"A central element of language teaching is raising students' awareness of, and developing their ability to 'chunk' language successfully. Task and process, rather than exercise and product, are emphasized." (Lewis 1993: vi-vii)

The approach mentioned in above statement is to lead students fully aware that language is the coherence of various lexical items with combining structure and function. But it is not enough to own a glossary. Grasping some core lexical items' characteristics of the structure and function, such as style and pragmatics, is very necessary for language learning. According to Communicative Approach, especially Task-based Approach, the learning process is not by rote or mechanically practicing single lexical item, but putting lexical items into actual text which will run through in lively tasks. Actively discovering and solving problems by students is the main path to language learning. Thus, Task-based Approach is the core teaching method.

3. Formulation of The Lexical Syllabus and Arrangement of Course Content

With the development of corpora technic introduced earlier, a fairly number of researchers turned language description to lexical item, and it is more sufficient to linguistic unit representing. Among that, the most important is to quickly extract a large number of real corpus from a bulk of actual language samples, such as index line, single-word glossary and multi-word glossary. This method of description is to focus on lexical item and also consider context, while concern the expanding structure of lexical item, such as some typical type of regular collocation and class association. In addition, the method pays attention to the stylistic variation of lexical item and social cultural context in a more general notion.

Nevertheless, the actual language is much more and complicated, then how to select proper learning material is a question. The scholars of Word-centred Approach advocate "synthetic teaching strategy" (Wilkins 1976), which is separately teaching different chunking of a language, and gradually promoting teaching in a controlled progress, until the students can compose the smaller linguistic units

into a coherent language. But word, phrase, conventionalized phrase, idiom, non-continuous structure and collocation, how to arrange the order of these basic linguistic units, and decide the sequence of learning? Willis and Lewis suggested using the frequency information supplied by corpora technology to divide the linguistic unit and usage into the most common level, the less common level and the not common level to support the student study by stages. The learning sequence of words, multi-word phrases and some structures can be decided by their usage frequency; the common collocations of more than two words could also be sorted according to their co-occurrence rate or collocational strength. Of course, Willis did not only care about the appearance rate of words, but also pay attention to usage weight of lexical items in real language materials. For example, Willis (1990: 48-56) mentioned that the most common usage of preposition "by" in the actual use is in passive structure which brings in the agent of the action not expresses the way of the action usually showed by grammar book; the corpora indicates "any" is not used in negative sentence or question in most conditions; by the language sense of English native speakers, "used to", which means doing something in a habit at past time, is used more frequently and more naturally than "would", but the corpora shows the usage rate of "would" of the same meaning is the three times of "used to". The above information should be followed in teaching.

When Willis practiced word programme, he made a classification for words based on the appearance rate in corpora. For example, based on corpora COBUILD, Willis (1990: 46) found that 700 high frequency English words could cover 70% English texts, and 1500 the most high frequency English words could cover 76% English texts, and 2500 the most high frequency English words could cover 80% English texts. Thus, Willis believed the high frequency English words based on corpora are very significant which should be focused in teaching not some unusual words belonging to specific field. This is the solution of the question of what words are inclusions and what words are exclusions raised by Lewis (1993: 106). In the teaching book *Collins COBUILD English Course* by Willis & Willis (1988) was also carried out this classification. For example, in the Level 1 of the teaching book series, the whole vocabulary was controlled in 700 words, and did not limit the studying sequence of lexical items. Willis applied a "learner's corpus" (Willis 1990: 65) in the teaching book which is a subcorpus extracted from corpora COBUILD, and the words of the subcorpus is about 700. All the language materials, such as sample sentences and exercises, are from the specific subcorpus.

In general, the design of teaching programme, development of teaching book and teaching process are all carried the frequency standard, which is teaching common lexical items and usages first, and teaching uncommon items next. That is to say, the selection of teaching content of lexical programme and relative teaching concepts mainly based on frequency; as to the teaching order is also in the help of rate information. Therefore, not only considering the pure frequency of word but also the frequency of usage, the consideration of word, structure and meaning is an integration.

4. The Lexical Approach as A Communicative Approach

Whether Willis or Lewis were have a clear teaching strategies and guiding advices for The Lexical approach. These advices are communicative in essence which could be named as the typical task-based approach. As the previous description, the start point of The Lexical Approach is to supply a lot of significant inputs centered by lexical items. And then integrating the inputs into learning tasks of listening, speaking, reading and writing. Willis (1990: 72) described the implementing steps of the Lexical Approach as following:

The methodology which exploits this corpus has six components:

Introduction: This gives students initial exposure to target forms within a communicative context.

Task: This provides an opportunity to focus on and realise target meanings. Students may begin to approximate to the target language form or they may use quite different, even ungrammatical forms.

Planning: The teacher helps students to move towards accurate production, often by modelling the target forums of them.

Report: Students have another opportunity to use target forms. Again, however, there is a focus on fluency as well as accuracy.

Listening/Reading: Students have a chance to hear or read the target forms used in a context which has become familiar to them through their own attempts to perform and report the task.

Analysis: This is an awareness raising exercise which gives the learners a chance to formulate generalizations about the language they have heard.

On the basic of Task-based Approach, Lewis (1993: 106) emphasized to implement the The Lexical Approach in a broader education background. Such as carrying the lexical items into the task of solving problems (comparison, causal argument and fact and reason presenting etc.). During the practice of specific teaching task, the teacher should insist the language description view and reject the prescription of linguistic and grammar. The introduction and explanation to language points should mainly take the frequency as the standard to recognize the importance and serviceability. Willis and Lewis were also fully aware that the input of real example is not enough to guarantee the learning outcome, then combined the language teaching to communicative approach with the advantages of Task-based Approach: using actual language, simulating actual communication, improving communicative language skill.

5. Teaching Practice: Data-Driven Learning

Tim Johns and Chris Tribble are the initiators of language teaching based on corpora. They created "Data-Driven Learning, DDL" or "classroom concordance". It is hard to say the Data-Driven Learning was influenced by Willis and Lewis, from the literature, Tim Johns had practiced "classroom concordance" in 1988, because they were all from University of Birmingham, the learning method was also the production of the environment of COBUILD corpora at the same time. Tim Johns' learning method had a biggest influence who could be called the leading pioneer of Data-Driven Learning.

The basic concept of Data-Driven Learning conforms to the "Learner-centered" idea, in which the classroom exercises or activities are all centered by autonomous learning. Students observe and hypothesize the designed teaching materials by themselves, and then conclude the answers through mutual discussion. This way matches the method of solving problem, self-discovering and interactive discussion of usual communicative language teaching. The following is the typical "single-word, multi-context" exercise of Data-Driven Learning (Johns 2002:108):

Figure 1.

This exercise tests the usage of nouns matching *on*. The significance of the test is that the matching of *on* and nouns lies on the border between vocabulary and grammar, so this exercise practices both the word and collocation even the sentence and context to let the students gain a thorough knowledge of vocabulary and grammar structure.

The other known way of Data-Driven Learning of Tim Johns is using index line information to help the international students of Birmingham University solve the language problems in writing. He called it "Kibbitzer" teaching designing (Johns 2002: 112-113) which is a teaching approach of one-on-one tutoring. In the help of index sample sentences from corpora, students take the vocabulary or grammar question of writing to compare with the teacher's modifying, and finally discover and conclude the reason of their faults. During the procedure, corpora plays a spectator and critic role, which is so called Kibbitzer. John (2002) gave an example: students first used "present some insight about", but the searching result of corpora showed the proper expression was "offer some insight into". The actual corpus proved the teacher's modifying is correct.

Willis and Lewis paid more attention to the discussion of macroscopic teaching principle, so the trace of utilization of corpora is not very obvious. However, the teaching design of Tribble & Jones and Johns made a full use of corpora definitely.

6. Conclusion

General speaking, like the Collins COBUILD Dictionary, the Lexical Approach not has much development. We conclude the reasons as follow:

Firstly, as a new teaching concept, the teaching design of The Lexical Approach is not an innovation, but borrows the corpora research result to Task-Based Approach. Thus, it is not an independent teaching approach, but a Task-Based Approach in essence. The creation of The Lexical Approach is aimed at the description of language, and the arrangement and classification of language content (includes vocabulary and teaching materials). On the other hand, the number of scholars involved into promoting this approach is not big limits its influence. It might because this kind of approach depends on corpora technic too much or the traditional teaching field is too powerful.

Secondly, the view of language in The Lexical Approach is also in dispute. It is unreal to let the whole language education field take lexical item as the basic unit of language learning, but give up syntax language view. However, there is still an academic discussion of whether language could base on lexical items to compose longer language unit even text.

Thirdly, some scholars disagree to decide the language teaching content totally based on frequency, because some language forms which conform to grammar/language sense or prototypical may not appear at a high frequency, so depending on the frequency along to decide the teaching content is deficient. For example, the vocabulary of Level 1 of *Collins COBUILD English Course* was controlled in 700 which controlled the difficulty but the communication used this high-frequent words is very limited.

Besides, from the angle of user's acceptation, The Lexical Approach exists an efficiency problem. The traditional teaching practice and dictionary or grammar book give definition and answer directly, while the Lexical Approach especially the Data-Driven Learning emphasizes learning by discovery with lower efficiency. It is hard to imagine that learning every word needs to search corpora and then conclude the word usage through a lot of index lines.

7. Acknowledgments

8. References

[1]   Nattinger, J. & J. DeCarrico. 1992. *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press

[2]   Johns, T. 2002. Data-driven learning: The perpetual challenge. In B, Kettemann & G. Marko (eds). *Teaching and Learning by Doing Corpus Analysis*. Amsterdam: Rodopi. 107-117.

[3]   Lewis, M. 1993. *The Lexical Approach*. Hove: Language Teaching Publications.

[4]   Tribble, C. & G. Jones. 1990. *Concordances in the Classroom*. London: Longman.