

Corpus-based Chunk Application in College English Writing

Bianqi Sun

Foreign Language Department
Beijing Institute of Graphic Communication
Beijing, China
e-mail:sunbianqi@bigc.edu.cn

Abstract—In the light of Swain's output hypothesis, the article first analyzes the possibility of chunk application in college English writing; introduces the types and pragmatic functions of chunks and the extraction from corpus data. Based on the corpus research, the strategies and approaches are discussed about how corpora facilitate lexical chunking in English writing. Finally, concrete advice is given on chunk application to achieve accuracy and fluency in writing.

Keywords—*chunk; college English writing; corpus; output hypothesis*

I. INTRODUCTION

In the context of economic globalization and technological integration, the Ministry of Education in the national long-term education reform and development plan (2010 - 2020) put forward the task of internationalization of higher education. China needs to cultivate a large number of talents with international horizons to participate in international affairs in all aspects. To meet the needs of national economic development and foreign cultural exchange, it is necessary to improve our international competitiveness and to enhance the soft power of language and education, especially in the field of writing, and enable the students to write a variety of business letters and academic literature.

The current problem is that few people are qualified for the advanced writing competence. As a college teacher, in teaching practice I find that, after two years of college English education, writing still remains the biggest headache. Although some students have mastered the knowledge of vocabulary, grammar and discourse, in the writing of national CET (College English Test) Band-IV, only 51% of them score 8 points in a total of 15, while other students score less than 8 points. Variety of errors and mistakes lead to the low score, among which, errors of collocation are considered the most serious ones.

Fortunately, learner corpus analysis has been a very active field of research since the emergence of computerized learner corpora in the early 1990s [1]. Computer learner corpora are systematic "electronic collections of spoken or written texts produced by foreign or second language learners" [2]. According to "Chinese Learner English Corpus" (CLEC) by Chinese scholars Gui Shichun, Yang Huizhong [3], Chinese students generally have a poor performance in the acquisition

of lexical and grammatical collocation and chunks. Students complain that the large number of polysemous words and synonyms are quite confusing, let alone the bundles with them. Differentiation of the feature of the words and lexical items is frustrating. So, to score higher points in the writing, students need intensified and attentive learning of collocation and chunks. As a teacher, necessary practice should be carried out in the classroom teaching, along with the lecture of collocation and chunk knowledge.

It is a relief that corpus can be exploited in eliminating such errors. Through a series of language teaching and learning activities, reforming the conventional way of writing training, teachers can do a lot to improve writing skills of the learners with the aid of corpus. Much has been discussed on the corpora application in vocabulary learning and writing separately, though, few articles specialize in the acquisition of chunks in the context of writing. Furthermore, this paper will further explore how English language teaching may benefit from corpus data to improve language learners' written output, and how to raise the student's awareness of the connection between grammatical, sociolinguistic and pragmatic uses of phraseological items of the English language. Through the discussion, we expect to propose some feasible corpus-based approaches in the sector of college English writing.

II. THEORETICAL BACKGROUND OF CHUNK

The concept of chunks was first proposed by Becker [4] and Bolinger [5] in the mid- 1970s. Chunks, different from the phrases, are loosely integrated blocks of functions and forms, with both vocabulary and grammatical features of language structure. Different English terms were used to indicate such concepts, such as formulaic sequences, chunks, unanalyzed chunks of speech, formulas, lexical phrases, lexicalized sentence stems, multi-word units and so on [6]. Thus we can say chunks are usually fixed or semi-fixed pattern or structure of multi-word units. Multi-word directly linked the phenomenon to language teaching, emphasizing their cohesive function in discourse, and divided into four types [6]:

A. Polywords.

Fixed phrases composed by more than one word may be either a typical structure of the English-type specification, such

as by the way, a piece of a cake etc., or an atypical form like as it were, all in all etc.

B. Institutionalized expressions.

It can be expression in the form of a sentence type in spoken language, fixed or semi-fixed form, fixed combination of pragmatic functions of words, or complete sentences; it can also be fixed form at the beginning of the sentence, including proverbs, aphorisms and social formula language, like How do you do? Where there is will there is a way, once upon a time, to begin with, and so on.

C. Phrasal constraints.

It refers to a phrase framework constituted by certain fixed words. You can fill in the appropriate word or phrase as needed. Such as, a ... ago, which can generate the phrase a couple of days ago / a long time ago; the ...-er, the ...-er structure that can generate the sooner, the better.

D. Sentence builders.

As a means of discourse organization, it is used mainly in written language. It is in the form of fixed or semi-fixed phrases with syntactic features, appropriate words or clauses can be added as needed, for example, "It is universally acknowledged that ...", "There is no doubt that ..."

As can be seen from the above classification, the form of chunks, different from the idioms, is relatively free. In contrast, Idiomatic structures are fixed expressions, whose forms seldom change, and the meaning can hardly be guessed from the separate words. In college textbooks, English idioms or phrases are usually listed as fixed expressions for students to learn, while the chunks, because of the loose combination and the large numbers, are never given proper attention to. In addition, the lexical chunks show a notable dual-feature: lexical and sentential. On the one hand, discontinuous building components are flexible in meaning expression; on the other hand, the flexibility increases the difficulty of pragmatic mastery. Just because of this, the learning of chunks leaves a big space for teachers to explore for learners' fluency.

An important part of language acquisition is the ability to understand and output blocks of the words as a whole. These language chunks form the original data of how people understand the pattern. Lexical approach emphasizes that language learners should learn and use the language in chunks, and language production process is not constrained by the rules of syntax, but a process of direct extraction of integrated word units from memory.

Over the past 25 years, corpora, corpus tools and corpus evidence have not only been used as a basis for linguistic research but also in the teaching and learning of languages. Computer-aided discourse and language acquisition research also proves that chunks can be used as an ideal unit in language teaching. This is because: First, the speed of processing for the information stored in the human brain is limited, thus, shortcuts must be found to improve the efficiency of information processing in the brain. In Chinese college writing test, within 30 minutes time, students are required to write an article with no less than 120 words, the average length usually exceeds 150

words. In a way, writing can be a slow, painful process even in our mother tongue, but when it is in a foreign language the problems can be magnified. So students usually have no time for careful thinking, and direct retrieval of chunks could save them much time. If the word block is taken as a memory unit, the sentence creation process can greatly reduce the psychological burden and improve the efficiency of the sentence output, which enhances language fluency. Second, the word block, as a combination of form and meaning, can greatly minimize the number of syntactic and semantic mistakes, because the fewer language components in sentence-making leads to a higher quality of sentence output. Third, because the word block is stored and extracted as an integral part, students with poor language level will confront less tension, and accordingly build up their self-confidence in foreign language expression. Fourth, cognitive learning of chunks conforms to the law of language acquisition. Chunk input as a whole can greatly increase the memory capacity with an ideal effect. Finally, word chunk is a combination of form and pragmatic function. In language output, learners do not have to consciously resort to the relevant context, which effectively avoids pragmatic failure in communication.

III. CHUNK LEARNING IN COGNITIVE AND OUTPUT THEORY

A. cognitive basis of collocation acquisition

From the cognitive perspective of collocation analysis, it is pointed out that verbal blunders experienced three stages: contrastive analysis of the 1950s-60s to error analysis of the 70s, then the transfer analysis in the 1980s. As we can see, second language acquisition has shifted from external environmental factors to a transitional system of internal factors related to the psycholinguistic cognition of learners. Study shows that, in the past two decades, the study of language transfer is no longer considered mechanical transfer of native language structure, but an implicit cognitive mechanism in the second language acquisition. The purpose of the learner is to establish conceptual representation through direct contact with chunk form and meaning. Mental lexicon is stored in the brain as the basis of the mental representation of lexical knowledge. Each L2 learner possesses a high-capacity thesaurus, which is well organized and efficiently extracted when necessary on the basis of their own second language thesaurus. By establishing direct links to the second language in sound and meaning, the implicit concept reconstructing system is formed, and the learner gradually increases the automation of his second language vocabulary extraction. This automatic prevention of mother tongue interference constitutes a firewall. Concerning the construction theory, many scholars[7] [8] [9] [10] tried to expound the processing of sentences in their own ways, which, to some extent, contributed to the language acquisition from different perspectives, such as construction grammar, cognitive grammar, cognitive semantics, frame semantics, etc.

B. Swain output hypothesis

In college English teaching, output has always been put in the first place. Last year, paragraph translation was introduced into the national CET (college English test) Band-IV examination, replacing the long-existing "Cloze" item.

Together with part of writing, written output account for 30% of the total score. From this, we can see language output is further emphasized in college English teaching.

As to the output theory, Swain asserts if learners want to express themselves fluently and accurately, they need to have both comprehensible input and output, for the process of output also serves as an opportunity for learners to reflect how well they have mastered the target language. Besides, the output feedback can test their own hypotheses and in turn consolidate their knowledge. The comprehensible output (CO) hypothesis states that learning takes place when a learner encounters a gap in his or her linguistic knowledge of the second language. By noticing this gap, the learner becomes aware of it and may be able to modify his output so that he learns something new about the language [11]. Although Swain does not claim that comprehensible output is solely responsible for all or even most language acquisition, she does claim that, under some conditions, CO facilitates second language learning in ways that differ from input and enhance input due to the mental processes connected with the production of language.

According to her, there are three functions of the output hypothesis:

1) *Noticing functions*: The learner realizes what they do not know or only partially know. They know what they want to say but are unable to communicate it. This is done through practice, verbally communicating in the second language.

2) *Hypothesis-testing function*: It is when the learners provide statement realizing that the grammar is not always correct and they receive feedback in order to improve. This enables the learners to reformulate their statements. Interaction within the classroom with teacher and peers can assist the learners to improve their grammar.

3) *Metalinguistic function*: The learners reflect upon the language learned and this enables them to control their output and internalize their linguistic knowledge. After the first two functions, the students should be able to internally reflect on what they have learned.

Under this theory, we find that corpus serves as a good source for learners to practice their output. The fact that most educational institutions have access to the internet has promoted the use of the web as corpus, and large corpora such as the British National Corpus and the COBUILD Corpus and Collocations Sampler are now accessible, free of charge, online and can be usefully incorporated into a process writing approach to help develop students' writing skills.

Compared with the conventional teaching mode, the corpus-based teaching mode enables students to find the underused and overused phrases or words, or even the misused grammatical structure. The findings strengthen the understanding of self-testing hypotheses, and knowledge of their own language gap pushes them to recognize and consolidate the inherent language knowledge and skills.

IV. ADVANTAGES OF CHUNK APPLICATION IN ENGLISH WRITING

A. *Strengthen the fluency of English writing*

In college English teaching, writing is one of the five basic skills to examine the students' comprehensive ability to use language, the other four skills being listening, speaking, reading and translation. Writing not only requires students to have a certain vocabulary, but also requires students to be proficient in target language. Therefore, teachers should, in the process of writing teaching, cultivate the students' fluency in use of word blocks. With reference to the cognitive lexical interpretation, the focus of chunk teaching approach will be put on the actual application. As chunks embody the dual functions of lexicon and syntax, skillful use of chunks in writing not only shortens the time of sentence making, but also reduces the grammatical errors, and the latter being the biggest headache for Chinese college students. Through constant practice of the chunks, the writing process will be free from syntactic constraints. The stored chunks in the brains will make it possible for students to write error-free English. Moreover, proper use of chunks has the function of discourse cohesion, as a result, the textual logic and reasoning of the writing will improve. In the course of teaching, teachers should emphasize the importance of chunks, and raise students' awareness of learning strategies, thus promote students' English writing ability.

B. *Improve the authenticity of English writing*

Before going on to discuss about how to output genuine English, we take a look at a technical term "interlanguage". There is a continuum from the source language through successive learning stages to the acquisition of the target language. At every stage there is an interlanguage. Richards claims that interlanguage is the type of language produced by second- and foreign- language learners who are in the process of learning a language. In language learning, learners' errors are caused by several different processes [12]. These include:

1) *Borrowing patterns from the mother tongue.*

2) *Extending patterns from the target language.*

3) *Expressing meanings using the words and grammar which are already known.*

L2 acquisition theory and research show that Chinese English learners mainly use the interlanguage in the learning process, learning from the vocabulary to grammar. When writing articles, they rely more on grammar rules to construct sentences instead of the chunks. However, to reach the level of native speakers, foreign language learners must master and store a lot of word blocks. Teachers should provide students enough authentic input material, in the meantime, focusing more on the output of chunks. The contact and absorption of the chunks will lay a solid foundation for a standard expression. Repeated practice of chunks will ensure a good command of language, thereby enhancing the student's target language output in a certain degree.

Exactly, the corpus offers the best solution to this problem. Johns' idea is to put the learner at centre stage and make him/her "a linguistic researcher" [13] who takes on an active role in discovering patterns around and bundles of selected

lexical items, often related to problems that are found in learners' writings. By exposing to the "real language" of the corpus, learners assimilate the authentic English quickly.

C. Strengthen the organizational capacity of discourse

Chunk is not just limited to vocabulary, but also covers a wide scope of collocation, either lexical or syntactic. Actually, many of the chunks have the function of discourse conjunction and cohesion. Common modes in English writing comprise argumentative, expository, narrative and descriptive. Whatever the writing styles are, the various structural patterns of chunks can be incorporated in the writing, which help to create a comprehensive and cohesive writing. Moreover, the use of chunks will also increase the power to reason, stimulating them to think and write in a logic way. As a result, with the accumulation of various topic chunks, students can not only improve the speed of writing, but also use the relevant discourse blocks to build a reasonable discourse framework to improve the discourse logic and coherence.

V. CHUNK TEACHING STRATEGIES AND APPROACHES

Knowledge of lexical chunks correlates positively with L2 proficiency. Lewis' lexical approach to language teaching calls for a change of focus. Instead of lexicalized grammar, concentration should be on heavily contextualized, grammaticalized lexis [14]. Direct teaching of grammar rules, very useful for L2 learners though, should be replaced by pattern teaching, focusing on lexical items. Some of the exercises he proposes cover typical phraseological types, such as collocations.

For example, the extraction of 3- and 4-word clusters around the word 'way' from a corpus results in a list of phraseological items such as "on the way to", "in this way", "by way of", "way in which", "in the same way", "out of the way", "in such a way", "in a way", "gave way to" or "as a way of (V-ing)" – items that may be very useful for students in their own academic writing.

Concordance			Concordance Plot	File View	Clusters	Collocates	Word List	Keyword List
Hit	KWIC							File
1	power over us and does in no way justify our servitude under th							LOCNESS
2	ing to Earth in much the same way. Here, he is trying to show th							LOCNESS
3	importance of living. In this way, Caligula is in fact doing the							LOCNESS
4	ats this like a child all the way through. In fact he is very m							LOCNESS
5	de could be contemplated as a way out of this situation, but pas							LOCNESS
6	le intentions really, but the way he goes about it is where the							LOCNESS
7	ization that <quote>. So in a way, Caligula can be seen as under							LOCNESS
8	s too that killing is not the way out. He does not however find							LOCNESS
9	lf be killed, seeing no other way out for himself nor for his pe							LOCNESS
10	rejects this, saying the true way to attain power is the revolut							LOCNESS
11	e pushed forward in a certain way but he is. At the beginning o							LOCNESS
12	she turns to repentance as a way out. <#> Oreste asks for signs							LOCNESS
13	turn away from religion as a way out and decides his actions fo							LOCNESS
14	en Jupiter offers her an easy way out she gladly accepts. Jupite							LOCNESS
15	acts of bad faith as an easy way out. As he says, <quote>. He :							LOCNESS
16	be made possible, and in this way he believes, at the beginning							LOCNESS
17	that this is not the 'right' way. <#> The patricians are older							LOCNESS
18	ng intellectual suicide. In a way he is driving for people to k							LOCNESS
19	is life, in order to find his way. <quote>, he says, <quote> He							LOCNESS
20	although Oreste can clear the way for the Argonites to free ther							LOCNESS
21	cts, therefore, in a positive way on the PP. <#> Similarly, alth							LOCNESS
22	erer's death respected in the way it should be - of a man who d							LOCNESS
23	is ideals. Hugo sees no other way of doing this than to allow h							LOCNESS
24	his background, in a similar way to his father, thus epitomiz							LOCNESS

FIGURE 1. COLLOCATION OF THE WORD "WAY"

In his view, lexical approach is to focus the attention on the word blocks to build language system. The principles of the chunk teaching are summarized as follows: (1) Vocabulary and grammar are not separated; language is not composed of the grammar system, but rather consists of multiple chunks. (2) Teachers should not analyze or translate the target language, but to guide the students to focus on chunk learning. (3) Hypothetical experiments are used to demonstrate the learning process as the output process instead of practice. (4) The vocabulary and grammar should be treated as an integrated part to learn. Based on the above principles, we put forward the following teaching suggestions:

First of all, teachers themselves should first update their pedagogical notion, changing from grammatical teaching to chunk teaching. Used to the conventional way of language learning, the students might find it hard to accept methodology of chunk learning. Thus, how to raise the awareness of the students and arouse their interest becomes a problem. Besides the regular classroom teaching, the corpora-based or corpora-driven chunk retrieval should be carefully guided. It is predictable that the students will meet with tremendous difficulties. Even for professional language practitioners, the initial use of the corpus could be frustrating, let alone the students.

Thus, in the intensive reading class, teachers should point out the chunks in the reading passage, and tell students what

chunks are. Generally speaking, the fixed phrases are better mastered by the students, because they are all listed in the Phrases and Expression. But the other loosely collocated ones are never specially listed. So to identify the chunks in the passage might be a good practice. From the common errors in college writing, we can see most of them are resulted from the negative influence of the native language. Therefore, textually demonstrating the chunks can effectively minimize or eliminate the errors.

For example, one of the most common errors by Chinese college students is the “There are people do ...” pattern. By accessing the pattern from the corpus, numerous correct patterns are retrieved with the tool of WordSmith or AntConc. In such a way, students might be deeply impressed by the correct expression presented. Moreover, teachers can turn the sentences into blank-filling exercise with the search word blanked out. They could then be asked to complete the concordances properly choosing from a list of words. To develop awareness of collocations and colligations, students could also be presented with complete concordances and asked to study and identify patterns useful for their writings.

By demonstrating the usage of the key word “hold” from the corpus LOCNESS, students will have direct visual understanding about the patterns of “hold” and its lexical and syntactic implications.

Concordance	Concordance Plot	File View	Clusters	Collocates	Word List	Keyword List
Hit	KWVC					
1	he Church and religion (especially Catholic) which holds good considering Sartre's atheism. <#> The whole					
2	tent' in which, in order to exorcise his guilt, he holds it up to everyone else, and exposes it, to then					
3	igula makes himself destiny, <quote>. It is he who holds the power of life and death, not God nor fate. <					
4	s known to have felt very alone, and feels he must hold up the mirror of his face to others - a face whic					
5	ce, but also Camus feels a sense of betrayl. After holding up a mirror to fellow existentialists they may					
6	n pursuit of his freedom. He acts as the saviour - holding up a new "regard" for the people of the town t					
7	t. By accepting guilt and becoming a <F>. Clamence holds himself up as an example of universel guilt and					
8	a question of means between Louis + Hoederer. Hugo holds fast to his belief that the party position canno					
9	d of the nation, by the nation, and as such is the holder and source of legitimate power, having the mand					
10	ntraception and abortion and he retained his tight hold over the financial area. In the environmental fie					
11	ist majority again but Mitterrand has continued to hold himself in the background and has help open confl					
12	in which the President can by-pass parliament and hold a referendum on a planned policy. De Gaulle did t					
13	f the time, many feeling that it was impossible to hold Leibnitz' philosophy in any degree of esteem afte					
14	the others. This view does not seem to hold much credibility in Britain. We want the economic					
15	its own system of legislation then they would also hold on to their own good feeling of identity, which w					
16	liamentary supremacy, and advocates of this theory hold the view that in this respect, a Single Europe do					
17	t the courts would accept this view. It would only hold true if the courts only recognised as valid that					
18	Socialist, Centre Right or Christian Democratic to hold a stable majority. This would lead to coalitions					
19	olor, as obviously not all of the white population holds racist views, but it is quite obvious that it is					
20	reasing rate of violence in society. The opponents holds, however, that prayer in public schools would des					
21	vive the America of the past that is <*>. Quindlen holds that returning to a time when guns were not brou					
22	ts to pray together. <#> This writer and those who hold the same belief must realize that concessions can					
23	ow budget to keep update equipment. Their lucky to hold a game in civic centers, and rarely do you see an					
24	state senator Mike Gunn, amounts to <*>. This view holds no validity. By blanketing all students in Ameri					

FIGURE 2. POLYSEMOUS WORD “HOLD” IN LOCNESS

Besides, the lexical approach does not mean that grammatical teaching should be totally ignored. Grammar is by far a useful way in sentence processing, especially for L2 learners. The ideal way is combine them together in classroom teaching, so that students learn the word blocks and a variety of sentence patterns under correct grammar rules. As a matter of fact, chunks and grammars are vaguely integrated; many of the chunks are grammatical blocks.

Finally, due to the large number of chunks, chunk learning should not be limited to the classroom teaching. How to stimulate the students' interest of self-learning and effectively extend outside-class learning hours is the ultimate aim. Generally speaking, Chinese college students are more dependable, and teacher can be more involved in their learning. Nowadays, offline and online corpus learning becomes another main source. What a teacher needs to do is to help students learn to use corpus tools and introduce relevant network resources. Need to say, popularization of corpora and their pedagogical use also requires some basic training in accessing corpora and in working with collocation lists. Such training is essential because concordance output, at first glance, may seem hard to handle, and because "a corpus is not a simple object, and it is just as easy to derive nonsensical conclusions from the evidence as insightful ones" [15]. Corpora may also be used directly by learners. This application is inspired by the concept of data-driven learning [16] [17] in which L2 learners make their own discoveries about language. Needless to note, using corpus techniques in the classroom has some important advantages. Corpora imbue the learning process with authenticity and allow learners to take control, acting as language researchers in their own right.

VI. CONCLUSION

This paper attempts to analyze the use of chunks in college English writing, and explores the functions of both chunk-learning and corpus application. Based on the literature and the practical experience of corpus-based chunk-teaching in the classroom, some suggestions are given as to how and to what a degree corpus can be exploited in chunk acquisition. Improving students' English writing ability is one of the important goals of English teaching. Writing is the process of output of written language. There is no doubt that many factors contribute to a successful writing and appropriate use of chunks play a key role. In fact, corpus-based awareness of chunk learning also benefits other language skills, which in turn further boost students' writing ability. With huge amount of chunk storage, students will greatly decrease the sentiment-making time, and produce genuine English fluently. The negative impact by their mother tongue will be diminished to the least degree, and instead language accuracy and fluency will be achieved. It has to be acknowledged that corpus-based writing practice with chunks is still a new thing, not surprisingly, countless problems will arise in the process, yet "real English" contact and powerful tools have already transform the traditional pedagogy. Small self-made corpus

based on students' own writing is specific and tailor-made, which is best welcomed by students in contrastive study. Word frequency, wordlist and concordance materials offer a powerful shortcut to decompose an article. It has to be admitted that the potentials of direct and indirect pedagogical corpus applications need to be further studied. I hereby propose more pilot experiments be done to probe the existing problems in corpus application, such as the development of multi-model corpus and more powerful corpus tools with user-friendly interface.

ACKNOWLEDGMENTS

This research is supported by the Social Science Research Common Program of Beijing Municipal Commission of Education (No. 18190114/004) and the Major Teaching & Reform Program of Beijing Institute of Graphic Communication (No. 22150114015).

REFERENCES

- [1] S. Granger, *Learner English on Computer*. Addison Wesley Longman, 1998.
- [2] S. Granger, "Learner corpus research: Current status and future prospects," in *Applied Corpus Linguistics: A Multidimensional Perspective*, Ulla Connor & Thomas Upton, Eds. Amsterdam: Rodopi, 2004, pp. 123–146.
- [3] S. Gui and H. Yang, *English Corpus for Chinese Learners*. Shanghai Foreign Language Education Press, 2003.
- [4] J. D. Becker, *The Phrasal Lexicon*. Mass: Bolt J. and Newman, 1975.
- [5] D. Bolinger, *Aspects of Language*. Harcourt Brace Jovanovich, 1975.
- [6] J. R. Nattinger, and J. S. DeCarrico, *Lexical Phrases and Language Teaching*. Oxford University Press, 1992.
- [7] R. Quirk, S. Greenbaum, G. N. Leech and J. Svartvik, *A Comprehensive Grammar of English*. Longman, 1985.
- [8] J. Pustejovsky, *The Generative Lexicon*. MA: MIT Press, 1995.
- [9] R. W. Langacker, *Grammar and Conceptualization*. Mouton de Gruyter, 1999. pp. 171–202.
- [10] A. E. Goldberg, *Construction: A Constructions Grammar Approach to Argument Structure*. The University of Chicago Press, 1995.
- [11] M. Swain and S. Lapkin, "Problems in output and the cognitive processes they generate: a step towards second language learning," *Applied Linguistics*. 1995, vol. 16, pp. 371–391.
- [12] J. C. Richards et al., *Dictionary of Language Teaching & Applied Linguistics*. Longman Group UK Limited, 1992, pp. 186.
- [13] T. Cobb, "Analyzing late interlanguage with learner corpora: Quebec replication of three European studies," *Canadian Modern Language Review*, 2003, vol. 59, pp. 393–423.
- [14] M. Lewis, "Teaching collocation: further developments" in *The Lexical Approach*, M. Lewis, Ed. Hove: Language Teaching Publications, 2000.
- [15] J. M. Sinclair, "Introduction," in *How to Use Corpora in Language Teaching*, J. M. Sinclair, Ed. Amsterdam: John Benjamins, 2004.
- [16] T. Johns, "Should you be persuaded—two samples of data-driven learning materials," *English Language Research Journal*, 1991, vol. 4, pp. 1–13.
- [17] G. Hadley, "Sensing the winds of change: an introduction to data-driven learning," *RELC Journal*, 2002, vol. 33(2), pp. 99–124.