On the Practice of Pedagogic Corpus Linguistics With Secondary School Iranian English Language Learners

Majid Elahi Shirvan

University of Bojnord, Bojnord, Iran

Negin Maddah

University of Zanjan, Zanjan, Iran

The gap between the theories of second language acquisition and their practice in materials development has been a controversial issue in the field of applied linguistics. Recently, the findings of corpus linguistics have narrowed the gap. However, their practice, especially in the realm of vocabulary acquisition, has encountered pedagogic problems in the classroom because the authentic nature of the words in corpus-based examples was beyond the level of learners' comprehension which paves the way for the practice of pedagogic corpus linguistics. Therefore, the purpose of this study was to practice pedagogic corpus linguistics, and mainly data-driven learning, with the first year secondary school English language learners in Iran in terms of their vocabulary acquisition. To do so, 28 students participated in the study. Examples from the British National Corpus for the main words of each unit of their course book, *Prospect 1*, were presented to them in a course of 16 sessions. Those which were beyond their understanding were eliminated and those comprehensible for them, the pedagogic ones, were selected based on their reports regarding the comprehensibility of each corpus-based example.

Keywords: corpus linguistics, pedagogic corpus linguistics, *Prospect*, concordance line, authenticity

Introduction

For a long period of time there has been a gap between the theory and practice in applied linguistics and materials development. Many researchers believe that the findings of most of the studies in applied linguistics have not been practiced yet in the development of materials (Allwright & Bailey, 1991). Thus, the need for efforts to bridge the gap between research and practice in the area of applied linguistics and materials development was emphasized (Tomlinson, 2013). Using authentic materials is one of the ways of bridging this gap and one of the ways of using authentic materials in the area of applied linguistics is the application of corpus linguistics. Therefore, in the realm of corpus linguistics, efforts have been made in order to strike a balance between theory and practice in materials development. A corpus is a collection of authentic, naturally occurring texts, including transcripts of spoken data, stored in a computer (Reppen, 2010) which are "machine-readable and a representative of a particular language or language variety" (McEnery, Xiao, & Tono, 2006, p. 5). Providing authentic materials for learners and facilitating the process of learning vocabularies can be considered as advantages of using corpora in language classrooms. However, some concordance lines extracted from unfamiliar contexts for learners might be beyond their level of understanding. Thus, the need for pedagogic corpus linguistics is felt. It is defined as the corpus of familiar and simplified texts at a proper level,

Majid Elahi Shirvan, Ph.D., English Language Department, University of Bojnord.
Negin Maddah, M.A., English Language Department, University of Zanjan.

supplying students with explanatory examples of the language as much as they need (Tomlinson, 2011). Analysis of the new course books for learning English at Iranian secondary schools, *Prospect 1*, indicates the gap for the practice of vocabulary learning through corpus linguistics. Since the use of corpus linguistics for the selection and presentation of the highly frequent words in English language in their real context of use causes some problems for the secondary school students of English language in Iran due to their low level of language proficiency, the need for the application of pedagogic corpus linguistics seems essential. Thus, the purpose of this study is to enrich the exposure of the Iranian English language learners at secondary schools to English language words in their real context of use with a pedagogical purpose.

Review of Literature

The gap between theory and practice in applied linguistics and materials development is one of the important topics that most researchers have the general consensus on (Tomlinson, 2013). With the increase of attention to lexis since the 2000s, one of the challenging topics in this area is how to present vocabularies in the classroom (Flowerdew, 2015). Furthermore, the concept of authenticity plays a decisive role in this area. Using texts which might not be pedagogically or educationally designed is considered one of the principles within the domain of authenticity. In other words, to present the language use outside the classroom, language samples regardless of the pedagogic purpose of their acquisition by learners were postulated to be more practical for learners (Badger & MacDonald, 2010). Therefore, using authentic texts to present a description of vocabularies in foreign languages necessitated the concept of corpus linguistics (Sinclair, 1987; Rundell, 2002).

Since the mid 1990s, in most of the ESL classrooms, dictionaries based on corpora with information of words' meaning and use for language learners, have been used, e.g., *Oxford English Dictionary* (Scott & Tribble 2006) because "corpora are particularly well suited for observing repeated sequences and patterning in language, independently of traditional linguistic categorization" (Mauranen, 2004, p. 102). Regarding the benefits of corpus linguistics, Granger (2002) explained that corpus linguistics has a great capacity to open up new viewpoints in language teaching methodology. The use of corpus in the classroom provides language learners with authentic materials and helps them to have enough exposure to the target language in order to be familiar with its features in a special context and co-text of use (Tomlinson, 2011). Furthermore, Gavioli and Aston (2000) mentioned that the investigation of corpora by learners provides them with the ability to discover the meanings and grammar themselves and helps them to benefit the meaningful learning. This way, individuals' perspectives of the reality of the language will be constructed. In addition, learners will be enabled to put into question the frameworks presented by teachers or reference materials.

Regarding materials development, to create authentic in-house materials, English language teachers study large amount of language data to make the process of language learning more possible and easier with the help of corpus linguistics. Xie (2015) explained that the application of corpus-based materials (e.g., Bennett, 2010; Reppen, 2010; Scheffler, 2011) with its large capacities and benefits for language instruction has been possible for teachers. While textbooks are in the process of integration with corpus linguistics (Carter & McCarthy, 1997), the need for corpus based activities and materials is still felt in the classrooms (Özdemir, 2014). In other words, the gap between corpus linguistics and language teaching is still great. In that, the application of corpora and concordances in language teaching environment is not enough in extent (Mukherjee, 2004). Understanding the advantages of using corpus for language learners, the significance of corpus linguistics' task to develop its application in schools will be highlighted (Ludeling & Kyto, 2008). As Mukherjee (2004) and Chamber (2005)

call this development to "popularize" corpus usage in schools.

While, the language portrayed to students in textbooks is not a rich representation of real or authentic language outside the classroom (Gilmore, 2007), considering students' provision of linguistic expertise to have the ability to take part actively in real situations, e.g., using the informal kind of English suitable in most of the everyday conversations, as one of the basic goals of language teaching materials, the weakness of language textbooks in this area is obvious. As the highlighted gap within the domain of authenticity in materials for language teaching is observed, fundamental changes still remain to be done (Gilmore, 2007). Corpus linguistics can be considered one way of filling this gap. However, some weaknesses within the domain of authenticity may cause problems for applying corpus in the language classrooms. Considering the definitions of authenticity, "the language produced by native speakers for native speakers in a particular language community" (Porter & Roberts, 1981; Little, Devitt, & Singleton, 1989) or "the language produced by a real speaker/writer for a real audience, conveying a real message" (Marrow, 1977; Porter & Roberts, 1981; Swaffar, 1985; Nunan 1988, 1989; Benson & Voller, 1997), the level of difficulty in applying the native like or real language in the classroom based on corpus may be higher than students' level of comprehension and this may cause feelings of frustration, confusion, and demotivation in language learners (Guariento & Morley, 2001). Thus, the need for simplification (Widdowson, 1978) of concordance lines to help learners' engagement moves us to the concept of *pedagogic corpus linguistics*. In other words, since the examples of corpus linguistics might unavoidably include some sentences that are not understandable for learners, using more familiar and easier examples of texts known as pedagogic corpus is needed (D. Willis & J. Willis, 2007). Furthermore, according to Breen (1983), the role of learners in interpreting authentic texts, corpus, the importance of tasks and classroom interactions may be regarded as other justifications for the use of pedagogic corpus linguistics. To explain more, the learners' interpretations of texts according to their own knowledge may be different from the exact meaning of text. Thus, the difficulty level of text should be considered to avoid frustration and misunderstanding. Furthermore, authenticity of tasks relies on the extent to which tasks provide learners with communication and meta-communication strategies and these occur when the language is understood by learners. Finally, the role of classroom in the construction of authenticity is decisive. Interactions of teachers with students and students themselves make the classroom a social authentic place which needs special pedagogical conditions to facilitate the process of learning. Another justification for the use of pedagogic corpus is the limit around the first principle of authenticity. As mentioned earlier in this paper, the first principle of authenticity is the use of texts that are not designed for the purposes of education. Since the level of difficulty and motivation of texts is different in language learners, teachers should pay a special attention to the selection of authentic texts that are appropriate for their students' level. Thus, using pedagogic corpus linguistics is applicable here (Badger & MacDonald, 2010).

The relationship between corpus linguistics and language teaching is considered a dynamic relationship. The impact of corpus linguistics by its resources, methods, and insights on language teaching is great which indicates that the application of corpus linguistics in language teaching has lots of advantages. Furthermore, it is worth mentioning that this relationship is two way (Ludeling & Kyto, 2008). Using corpus tools and methods in a language teaching and learning context is known as pedagogical corpus application. As Leech (1997) has mentioned two distinct types of pedagogical corpus applications are direct and indirect use. In that, the indirect use of corpora is decisive in choosing what to teach and how to teach it. However, in direct use teachers and learners will be engaged in the process of teaching and learning a language (Flingestone, 1993, p. 98). Despite

the indirect approach which emphasizes on the influence of corpus on syllabus design and on materials designers, the emphasis of direct approach is on teachers and learners. In that, direct approach is more teacher and learner centered (Ludeling & Kyto, 2008). In other words, language learners and teachers are not dependent on the researchers to provide patterns of language for them. They are in the role of discoverers who work with corpus based materials themselves in an "autonomous" way (Bernardini, 2002, p. 165). In the same vein, John (2002) to support the direct method made the suggestion to "confront the learner as directly as possible with the data, and to make the learner a linguistic researcher" (p. 108). In addition, the application of direct method in using pedagogic corpus linguistics is in line with the theories of authenticity mentioned above and will emphasize the authenticity of learner in the process of interaction. This method which is based on the interaction between learners/teachers and corpus based materials known as "data driven learning" or DDL (Johns, 1986, 1994, as cited in Ludeling & Kyto, 2008).

Discussing direct method, some researchers have provided a number of examples of different ways language learners can benefit from general corpora (Ludeling & Kyto, 2008). For example, the description of the usefulness of "corpus-aided discovery learning" with the BNC which was conducted by Bernardini (2002, p. 165). Besides, Kettemann (1995) have emphasized the significance of applying DDL in ELT classrooms to benefit learners with the high sense of motivation and exploration. It is worth mentioning that the application of DDL method is not limited to English language teaching or English language corpora. It is also a useful method in teaching other languages and corpora of other languages (Ludeling & Kyto, 2008). For instance, the works of Whistle (1999), Dodd and Jones (1997), Brodine (2001), Coffey (2007), etc., are considered here.

Based on the explanations about the significance of applying corpus linguistics in ELT and its weakness around the concept of authenticity and regarding the advantages of using pedagogical corpus application and direct method, in this study we are to provide English language learners of secondary schools in Iran with pedagogic corpus application through direct method since the investigation of the newly developed book for secondary school language learners, *Prospect 1*, indicates the weakness in the vocabulary teaching.

Method

Participants

All the examples of corpus linguistics were piloted in one classroom with 28 students in Bojnord, Iran. The students' age ranged from 12 to 14. All of the students were male and their native language was Persian. They did not have any background in English language before the beginning of the course.

Materials

Prospect 1 is the newly developed book used in secondary schools in Iran. The students' book consists of eight lessons and four review parts at the end of each two lessons. There is a photo dictionary at the end of the book. This photo dictionary includes all the words of each lesson with related pictures. All of the words targeted by corpus examples for corpus linguistics were selected from the photo dictionary at the end of the book.

Procedure

The method used in the process of piloting the material is based on the direct application of pedagogic corpora or data-driven learning. As explained in the review of literature in DDL method learners are exposed directly with reference materials (Johns, 2002). They travel between the lines of concordances and explore the

meanings and functions of words by their own. To name a few from the advantages of applying DDL in the application of corpora in the classroom we can point up that DDL authorize language learners to understand and explore things autonomously (Ludeling & Kyto, 2008). Furthermore, the interaction among students and also between the teacher and a student creates an authentic atmosphere which is the authenticity of the classroom according to Breen (1983).

First, all of the students read the examples of corpus based reference works for each specific word individually. While reading, they paid attention to both meaning and function of each word in sentences. Students' attempt to identify keywords in this part is known as assembling process introduced by Willis (1998, as cited in Tomlinson, 2011). Then, they analyzed the sentences to comprehend specific words. Finally, in the extension and consolidation process, they tried to find out the meanings and functions of specific words in different contexts of use. In this phase, each individual student marked the sentences whose meaning he could not understand. Then, students worked in pairs reading, discussing and evaluating the meanings of sentences. Partners in each pair helped each other to understand the sentences that were not understandable for each individual. Then, they both marked the examples that were beyond their level of understanding. Finally, in groups of four, they read sentences and exchanged information regarding the meaning of the sentences they had already marked as beyond their understanding and reached consensus about the sentences that were not comprehensible for them. Thus, in this phase the examples that were not understandable for learners were omitted from the examples which had already derived from the corpus. Therefore, all the remaining sentences which were meaningful for the students and in line with their understanding were considered pedagogic corpus examples, having pedagogical use in the classroom. It is noteworthy that in each phase of reading sentences by students, in pairs or in groups, the interaction of students with each other was helpful in understanding the sentences that were considered difficult for individuals. This was repeated for all the key words of each unit of the course book during October 2014 to May 2015, 16 sessions. All of the students explored the meanings and functions of words among the examples of British National Corpus known as BNC.

The third edition of BNC, xML, was made available in 2007. It is a collection of 100 million words of spoken and written language. It is collected from the wide range of sources and considered as a wide cross-section of British English from the late 20th century. BNC includes 90% written and 10% spoken part. The written part derived from regional or national newspapers, specialist periodicals and journals for all ages and interests, academic books and popular fiction, published and unpublished letters and memoranda, school and university essays. Besides, the spoken part includes orthographic transcriptions of unscripted informal conversations and spoken language collected in different contexts, ranging from formal business or government meetings to radio shows and phone-ins. BNC is encoded based on the Guidelines of the Text Encoding Initiative (TEI) to be a presentation of the parts of speech tagging and different structural properties of texts.

Examples

As illustrated in the following example, the topics of lesson one and two are *My Name* and *My Classmate*. One of the important words of these two lessons mentioned at the end of *Prospect 1* is *student*. As explained in the method section, at first students exploring corpus based references individually. The results of their exploration are the sentences in the following which were comprehensible for them.

• Full-time student officers are available to give advice and make representation to the University on behalf of members;

- Each student may choose to concentrate on either Graphics or Photography;
- Towards the middle of the course bombing was introduced and the student crew started to work together;
- Hendrique was reputed to have been the best student ever to graduate from Balashikha;
- Here again, student counselors in college may help in talking you through such problems;
- At first, the student is completely puzzled.

Then, students read sentences in pairs and finally in groups of four. The sentences in the following are comprehensible sentences for students in groups of four.

- Full-time student officers are available to give advice and make representation to the University on behalf of members;
- Each student may choose to concentrate on either Graphics or Photography;
- Here again, student counselors in college may help in talking you through such problems;
- At first, the student is completely puzzled.

In lesson 3, *My Age*, one of the words is *Saturday*. The comprehensible sentences for individual students are as follows.

- On the Saturday of my third weekend at the Palace Hotel I looked their address up in one of the telephone directories in my local library;
- It was Saturday afternoon and he was standing, as usual, on the cart he hired as his platform;
- You simply cannot describe how important Saturday is to us all;
- On a Saturday in November 1980, I entered Micheline's cluttered flat;
- The show runs from Tuesday until Saturday;
- The team have to pick themselves up and perform on a Saturday;
- That was where Ken and Zara went on Saturday night;
  The following are sentences comprehensible for students in groups.
- It was Saturday afternoon and he was standing, as usual, on the cart he hired as his platform;
- You simply cannot describe how important Saturday is to us all;
- On a Saturday in November 1980, I entered Micheline's cluttered flat;
- The show runs from Tuesday until Saturday;
- The teams have to pick themselves up and perform on Saturday;
- That was where Ken and Zara went on Saturday night.

The title of lesson 4 is *My Family Tree*, the comprehensible examples for the word *teacher* in this lesson are as follows.

- Ramsey hoped now to be a teacher in a university;
- The question is whether grammar ought to be taught as a separate formal study, like harmony or counterpoint, or whether a teacher can assume that it will be picked up through practice;
- He explained that he was training to be a teacher in Bechar, a large town a little north of where I had begun walking;
- Discussion of the film afterwards can aid the teacher in evaluating the effectiveness of the film;
- Teacher: Should we put the big ones together over here?;
- They specify what makes a good person and a good teacher;
- Keep the test papers safe or pass them on to the class teacher (or the student) so that weak areas can be diagnosed;

- One teacher described the effect of an air raid on the students;
- It is housed in a semi detached house consisting of nine classrooms, teacher's room, resources room, small bar, and garden;

Sentences considered comprehensible after a group discussion are in the following.

- Ramsey hoped now to be a teacher in a university;
- He explained that he was training to be a teacher in Bechar, a large town a little north of where I had begun walking;
- Discussion of the film afterwards can aid the teacher in evaluating the effectiveness of the film;
- Teacher: Should we put the big ones together over here?;
- They specify what makes a good person and a good teacher;
- Keep the test papers safe or pass them on to the class teacher (or the student) so that weak areas can be diagnosed;
- One teacher described the effect of an air raid on the students;
- It is housed in a semi detached house consisting of nine classrooms, teacher's room, resources room, small bar and garden.

The topic of lesson 5 is *My Appearance*. *Tall* is one of the words in this lesson. The followings are comprehensible sentences for individuals from corpus based references.

- Tall, lazily handsome, and boy did he know it!;
- It is got all these white lights all over it and it is real tall;
- She was a beautiful child, too, and tall—not like the rest of us;
- Tall beech trees shed leaves over a disused tennis court, near well-kept lawns and bright hydrangeas;
- I am as tall as the tower;
- The plant is very bushy, and with long stems grows tall;
- He was tall and fat, and without much hair;
- So he made a man, tall and strong like father, to take care of the plants and animals;
- Then the lawyer noticed the tall mirror on the wall between the glass-fronted bookshelves;

Following are sentences comprehensible for students in groups of four.

- Tall, lazily handsome, and boy did he know it!;
- It is got all these white lights all over it and it is real tall;
- She was a beautiful child, too, and tall—not like the rest of us;
- I am as tall as the tower.
- He was tall and fat, and without much hair.

The topic of lesson 6 is *My House*. One of the words in this lesson is *bedroom*. The comprehensible sentences for individuals are in the following.

- In my grandmother's house I had had a big bedroom; here I had to share;
- It was a new me who demanded locks on my bedroom door the next day;
- Combined everything except bedroom, there was one separate bedroom;
- I was not allowed in the master bedroom without supervision;
- She went up to the bedroom to see Anna and they spoke together for a long time.

All sentences were comprehensible for students in groups.

In lesson 7, *My Address*, telephone was one of the words mentioned in the photo dictionary at the end of

the book. The comprehensible examples for individuals are in the following.

- Her heart was beating wildly as it had for days whenever the telephone bell rang anywhere she happened to be;
- A telephone hotline service provided advice and information in response to just over 2,000 enquiries from all over the UK in 1986/7;
- Hung up to hang up the telephone;
- Miss Mair said: "I'll telephone you when I know when I'm going";
- Then the sound of the telephone pierced the silence;
- After a further telephone interview—this time with myself as the interviewee—for the Radio Times, it seemed I was finished;
- If you change your mind for any reason, please telephone back to say so;
- "Then why didn't she telephone?";
- His last work was the reorganization of the telephone system in Malta;
- Michael picked up a telephone;
- Suddenly the telephone rang;
- The telephone rang and Victoria picked it up;
- Putting her hand over the telephone, Sophie asked urgently, "Are you going out with Ian today?";
- For further information telephone;
- And he did say to me well he has spoken to me about it, he did say answering the telephone;
  Comprehensible sentences for students in groups are in the following.
- Hung up to hang up the telephone;
- Miss Mair said: "I'll telephone you when I know when I'm going";
- Then the sound of the telephone pierced the silence;
- If you change your mind for any reason, please telephone back to say so;
- "Then why didn't she telephone?";
- His last work was the reorganization of the telephone system in Malta;
- Michael picked up a telephone;
- Suddenly the telephone rang;
- The telephone rang and Victoria picked it up;
- Putting her hand over the telephone, Sophie asked urgently, "Are you going out with Ian today?";
- For further information telephone.

The topic of the last lesson is *My Favorite Food*. One of the important words of this lesson is *tomato*. Followings are comprehensible sentences on the part of individual students.

- Patulin was suppressive to the damping-off diseases of tomato;
- One large tomato, cut into small pieces;
- Serve with new potatoes and a tomato and onion salad;
- You have got tomato skin on your front tooth;
- Avoid those dishes wrapped in cheese and tomato sauce;
- Len Hatch looked up from the tall tomato plants that he was tending, then looked back to what he was doing;
- It were half a bottle of Evian, 12 grapes, a piece of cheese and a small tomato;
- By the end of the nineteenth century, when tomatoes had become an upper-class luxury and tomato soup was well on its way, a new legend had become attached to the tomato;

- Lacking fresh tomatoes and meat we tried to compensate by piling tomato paste into all our stews and soups and sauces;
- One tomato, diced, plus 100g/4oz tomatoes, halved;
- On the way to the tomato patch they made a detour to Louis's workshop to collect some baskets;
- She moved her hands among the leaves of the tomato plants, liking their roughness and hairiness, their harsh earthy smell, a bit sour, like that of geraniums;
- It was bright and shiny, very fresh and not at all like tomato ketchup;
- Oh, and tomato sauce, I always have curry and tomato sauce!;
- There is corned beef and tomato;
- Seventy nine, so you had a roll tomato and cheese;
- Do not need that many, whereas last year she had so many tomatoes and tomato plants, she did not know what to do with them;

And at least, here are the comprehensible examples for the word tomato after discussing sentences in groups of four.

- One large tomato, cut into small pieces;
- Serve with new potatoes and a tomato and onion salad;
- You have got tomato skin on your front tooth;
- Avoid those dishes wrapped in cheese and tomato sauce;
- One tomato, diced, plus 100g/4oz tomatoes, halved;
- It was bright and shiny, very fresh and not at all like tomato ketchup;
- Oh, and tomato sauce, I always have curry and tomato sauce!;
- There is corned beef and tomato;
- Seventy nine, so you had a roll tomato and cheese.

Results

As illustrated in Table 1, a number of sentences after group discussion were removed from the comprehensible examples of the corpus for individuals. Therefore, the remaining sentences were pedagogical corpus linguistics examples. Table 1 indicates examples of one sample word of each lesson in *Prospect 1*. For instance, the word *student* is from lesson 1. The number of examples comprehensible for individuals was seven. After discussing the sentences in groups the ambiguous examples regarding the level of students were omitted and the number of examples was reduced to five and these five examples were pedagogical corpus linguistics examples for the word *student*.

Table 1

Frequency of Corpus and Pedagogic Corpus Examples for the Sample Words of Prospect 1

Lesson	Word	Individually comprehensible corpora	Comprehensible corpora in groups
1, 2	Student	7	5
3	Saturday	7	5
4	Teacher	9	8
5	Tall	9	5
6	Bedroom	5	5
7	Telephone	15	11
8	Tomato	17	9

Table 2 is the illustration of words for one of the lessons of the book, lesson 7. There are two columns in front of each word. One of them is the number of examples of corpus comprehensible for individuals and the other is the number of sentences comprehensible for students in groups of four. For example, there were three examples of corpus linguistics comprehensible for individuals for the word *office* and the number of sentences comprehensible for students in groups of four is three. It means that all three examples were comprehensible for students in groups after discussing the meanings. The information regarding the examples of the other lessons of the book are presented in the rest of the results section.

Table 2

Frequency of Corpus and Pedagogic Corpus Examples for the Words of Lesson 7

	Word	Individually comprehensible corpora	Comprehensible corpora in groups
Lesson 7	Office	3	3
	Call	19	17
	Telephone	15	11
	Street	16	12
	e-mail address	2	2
	Clock	21	19

Table 3 is the illustration of words of lesson 1 and 2 in *Prospect 1*. As seen in Table 3, after discussing the number of examples of corpus linguistics for each individual decreased as some of the sentences were beyond the level of students' understanding in groups. For example, for the specific word *boy* 15 examples of corpus were comprehensible for individuals while 11 sentences were considered comprehensible for students in groups. It means that four of the sentences were omitted to be understandable for all of the students in groups.

Table 3

Frequency of Corpus and Pedagogic Corpus Examples for the Words of Lesson 1 and 2

	Word	Individually comprehensible corpora	Comprehensible corpora in groups
Lesson 1 and 2	Boy	15	11
	Girl	10	10
	Man	8	6
	Woman	7	5
	Teacher	9	8
	Student	7	5
	Library	22	17

Table 4 illustrates the corpus examples of the words for lesson 3 in *Prospect 1*. The number of comprehensive examples for a specific word in BNC corpus and the number of comprehensive examples for students in groups is shown in the second and third column of the table. It is worth mentioning that in some rows these numbers are the same. It means that all of the comprehensive examples of corpus linguistics for that specific word for individuals were comprehensible for students in groups and no sentence was omitted. For example for the word *December,* the numbers in both columns are the same.

Table 4

Frequency of Corpus and Pedagogic Corpus Examples for the Words of Lesson 3

	Word	Individually comprehensible corpora	Comprehensible corpora in groups
Lesson 3	Saturday	7	5
	Sunday	11	11
	Monday	16	12

(Table 4 continued)

	Word	Individually comprehensible corpora	Comprehensible corpora in groups
	Tuesday	23	20
	Wednesday	17	15
	Thursday	11	11
	Friday	18	16
	Date	10	8
	Month	16	15
	Year	12	11
	January	13	12
Lesson 3	February	15	13
	March	8	6
	April	11	9
	May		
	June	13	12
	July	8	5
	August	8	6
	September	11	7
	October	11	9
	November	6	5
	December	6	6

Table 5 is the illustration of the words in lesson 4 of the book. In this table the number of individually comprehensible examples from BNC and those understandable for students in groups is shown. As an example, the word *employee* has 12 examples individually understandable for students while the number of sentences comprehensible for students in groups was eight. It means that four examples that were beyond the level of students' understanding in groups were omitted. However, for some of these words, this number did not change like *waiter* and *postman*.

Table 5

Frequency of Corpus and Pedagogic Corpus Examples for the Words of Lesson 4

	Word	Individually comprehensible corpora	Comprehensible corpora in groups
	Engineer	13	11
	Farmer	13	12
	Policeman	9	7
	Postman	4	4
	Painter	6	5
	Florist	5	5
	Employee	12	8
Lesson 4	Waiter	8	8
	Teacher	9	8
	Doctor	12	12
	Dentist	12	11
	Mechanic	13	12
	Shopkeeper	6	6
	Pilot	6	6
	Nurse	10	9
	Driver	5	5

(Table 5 continued)

	Word	Individually comprehensible corpora	Comprehensible corpora in groups
Lesson 4	Baker Worker Cook	9	7

In Table 6 the words of lesson 5 are shown in one column and the number of individually comprehensible examples and in group understandable ones in other columns. As indicated in the table for some words like *Chador* and *Manteau*, no examples from corpus linguistics were found because they are examples of Persian English words of dressing.

Table 6

Frequency of Corpus and Pedagogic Corpus Examples for the Words of Lesson 5

	Word	Individually comprehensible corpora	Comprehensible corpora in groups
	Tall	9	5
	Short	3	3
	Socks	5	4
	Sandals	8	5
	Sunglasses	6	5
	Belt	3	3
	Gloves	3	3
	Sweater	7	6
	Dress	3	2
	Cap		
	Coat	10	9
	Chador		
	Scarf	13	11
Lesson 5	Suit	4	4
	Jacket	8	7
	T-shirt		
	Shirt	6	5
	Manteau		
	Trousers	8	7
	Shoes	7	5
	Yellow	3	3
	Blue	3	3
	Green	6	5
	Red	5	4
	Black	8	7
	White	1	1
	Gray		
	Brown	2	2
	Pink	2	2

Table 7 is the illustration of the words of lesson 6 in *Prospect 1*. The number of comprehensible examples for individuals and groups is shown in front of each specific word. For example these numbers for the word *working* are 13 and 12. Or 1in both columns for the word *blanket*.

Table 7

Frequency of Corpus and Pedagogic Corpus Examples for the Words of Lesson 6

	Word	Individually comprehensible corpora	Comprehensible corpora in groups
	Living room	9	7
	Bedroom	5	5
	Bathroom	5	4
	Stairs	5	5
	Kitchen	12	11
	Garage	11	10
	Flat	2	2
	Villa	2	2
	Tower		
	Bookcase	9	8
	Clock	10	9
	Computer	10	10
	Mirror	6	5
	Blanket	1	1
Lesson 6	Table	6	5
	Television	8	7
	Radio	6	6
	Comb	5	4
	Cooking	10	9
	Washing	6	5
	Watching	10	9
	Newspaper	10	10
	Playing	4	3
	Fixing	9	7
	Working	13	12
	Reading	10	9
	Studying	10	8
	Eating	12	11
	Drinking	1	1

Finally, Table 8 shows the examples of the words of lesson 8 in *Prospect 1*. The number of examples individually comprehensible and understandable in groups is shown in front of each word. This number decreased for some words after discussion in groups but for some of the words is the same for individuals and in group discussions.

Table 8

Frequency of Corpus and Pedagogic Corpus Examples for the Words of Lesson 8

	Word	Individually comprehensible corpora	Comprehensible corpora in groups
	Apple	6	5
	Carrot	4	4
	Orange juice	4	3
Lesson 8	Onion	7	6
	Banana	6	6
	Potato	6	5
	Quince	2	2

(Table 8 continued)

	Word	Individually comprehensible corpora	Comprehensible corpora in groups
	Tomato	17	9
	Zucchini	1	1
	Jelly	4	4
	Water	1	1
	Yoghurt	8	5
	Egg	1	1
	Honey	1	1
	Cheese	7	4
	Fish	2	2
	Coconut	3	3
Lesson 8	Spaghetti	11	9
	Pineapple	5	5
	Ice-cream	5	3
	Bread	4	3
	Rice	6	5
	Dates	1	1
	Kebab	2	2
	Cake	7	5
	Chicken	2	2
	Milk	2	2
	Salad	11	9
	Tea	8	7
	Fruit	5	5

Discussion

Since 2000, the attention to lexis in corpus-based studies in second language acquisition has increased (Flowerdew, 2015). One of the main problems of learners of English as a foreign language is how to learn vocabularies to empower them for reading and writing in academic discourse (Shaw, 1991). Thus, as mentioned in the review of literature, corpus-based studies focusing on lexis like the works of Nation (2001), Sinclair (1991, 1999) and Stubbs (2001) have been conducted. However, the corpora provided in many studies might lose their pedagogic orientation when practiced in the classroom. As observed in this study, the texts examples derived from the corpus were not all pedagogic, considering the participants' comprehensibility. From the perspective of input hypothesis, language acquisition occurs on the condition that the input provided for language learners is as comprehensible as possible. It means that learners should be exposed to the input which is understandable for them. To put it in simple terms, it should be slightly beyond the learners' repertoire of knowledge while not being too difficult and challenging for the learners to understand ($i + 1$). In other words, the development of the language occurs when the input is just one step beyond the understanding level of learners (Krashen, 1985). Considering this hypothesis, in the process of learning vocabularies with corpus-based materials, concordance lines should be just one step beyond the understanding level of learners in a way that is comprehensible for them. Corpus examples might be more difficult for language learners ($i + 2$) and this difficulty might have negative effects on the process of vocabulary acquisition despite being authentic. Thus, the application of pedagogic corpus that is slightly beyond the level of learners ($i + 1$) can provide learners with appropriate comprehensible input.

Regarding authenticity, the requirement to expose learners to authentic texts in order to provide them with direct and immediate input data reflecting authentic communications in the target language is still a controversial issue (Breen, 1985). That is, the use of materials which enable learners understand the target language in an appropriate level to survive and use in real situations, authentic materials, is always suggested. As seen in this study, the examples derived from corpora were used to expose the participants to the authentic texts. However, some of the examples, despite being authentic, were not comprehensible for them. This necessitated the pedagogic corpus linguistics. The pedagogic phase indicated a relative distinction between the texts which are the representations of the target language use and texts tailored to the systems of knowledge of the participants in order to generate comprehension in the target language with the correct interpretation of texts in line with fluent users of the English language as the final goal of exposing to authentic texts. In other words, corpus linguistics and authenticity are intertwined because both are not limited to texts but the learners' proportion is acknowledged. Thus, learners' role within the process of the *pedagogication* of corpus texts highlights the need for the *authenticity of learners* (Widdowson, 1976; Breen, 1983). In other words, in the process of selecting texts, teachers should be aware of the extent to which corpora examples engage learners to interact with the texts. The student's tolerance for difficulty and their degree of comprehension are decisive matters in choosing the material; thus, highlighting the role of learners' participation in the process of text selection. Similarly, corpus materials which may not consider learners' tolerance for difficulty might not motivate learners to engage in the process of learning. Therefore, the use of pedagogic corpus linguistic examples compatible with the learners` level of understanding can be justified here.

Besides the authenticity of texts and learners, the authenticity of the tasks should not be neglected. An authentic task is the one which involves language learners in authentic communication with texts and with *others* in the classroom and facilitate the process of learning for its own sake (Breen, 1985). In other words, the tasks used in the classroom should provide language learners with authentic materials on one hand and create an environment for authentic communication among learners with the text on the other hand. Thus, considering these aspects of authenticity, pedagogic corpus examples can be considered a practical tool for catering the condition of communicating and learning via the authentic tasks developed in the language classroom. To reach comprehensible data, the participants in this study first reflected on the comprehensibility of the corpus examples first, and then they checked this process again in pairs and group work tasks. The pedagogic and authentic aspects of the data they were exposed to would not have been turned into practice if they had not shared it in terms of tasks in the classroom. Pedagogic tasks enable students to produce language (Swain, 1985) and also to develop learners' attention to aspects of meaning and forms in the target language, compression (Willis, 1996; Skehan, 1998). Thus, the concept of authenticity has this potential to be practiced in pedagogic tasks as well as texts (Guariento & Morely, 2001).

In addition, the practice of pedagogic corpus linguistics encompasses the contribution of the actual *social situation* and the role of *classroom* to the concept of authenticity (Breen, 1983). Learning is the basic communicative goal shared by people in the special social environment of the classroom which indicates the authenticity of the classroom. Learning takes place when learners are encouraged to share their own experiences. As experienced by the participants in this study, they reached consensus regarding their comprehension of the corpora examples. Thus, one of the important authentic activities in the classroom environment to pedogogicate the authentic corpus examples was communicating in order to choose the best examples for the comprehension of the words of each unit of the book. According to Breen (1983), students'

working in pairs or groups negotiating meanings and discussing aspects of language shape an authentic interaction in the classroom. As a consequence, this authentic interaction plays a significant and indispensible role in the practice of pedagogic corpus linguistics in the classroom. The pedagogic corpus examples for one class might not be necessarily pedagogic for another class and the legitimacy of the pedagogy depends on the share of knowledge and experience by learners conducted in the tasks and activities in the classroom.

Conclusion

Following the main principles of communicative language teaching, English language teachers in Iranian schools might face some pedagogic problems in terms of providing elementary level students with the authentic example of the words they need to learn for each of their units. On one hand, they are supposed to expose their students with the real context of each word to covey a real message for a real audience (Marrow, 1977), the traditional definition of authenticity, and on the other hand, they are to teach their learners. Thus, there might be a contrast between the two purposes because the contexts of most of the authentic examples are beyond the comprehension level of their learners. Facing this contrast, some teachers might feel disempowered in the classroom; thus, to strike a balance between these two purposes, they might be dependent on their contrived materials (Widdowson, 2007). However, as practiced in this study, Iranian English language teachers can take advantage of the application of pedagogic corpus linguistics to incorporate both purposes in their classroom. That is, providing comprehensible input for their students with authentic texts, practicing authentic tasks with them, considering learners interpretations, understanding and involvement and creating an interactional atmosphere in the authentic environment of the classroom. To achieve this, they can extract concordance lines of the words their learners need at different levels of language learning and then tailor them to the needs of their learners by checking the comprehensibility of the examples on the part of their learners.

Besides, the list of pedagogic corpora examples developed in this study can cater the needs of the first year secondary school English language learners in their practice of vocabulary acquisition. Being more empowered than the past, the Iranian English language learners can follow the footsteps of communicative language teaching, especially familiarizing learners with the authentic nature of the main words of each unit of Iranian secondary school English language course book, *Prospect 1*.

References

Allwright, D., & Bailey, K. M. (1991). *Focus on the language classroom: An introduction to classroom research for language teachers*. Cambridge: Cambridge University Press.

Aijmer, K. (1996). Swedish modal particles in a contrastive perspective. *Language Sciences, 18*(1), 393-427.

Badger, R., & MacDonald, M. (2010). Making it real: Authenticity, process and pedagogy. *Applied Linguistics, 31*(4), 578-582.

Barlow, M. (1996). Corpora for theory and practice. *International Journal of Corpus Linguistics, 1*(1), 1-37.

Bennett, G. R. (2010). Using corpora in the language learning classroom: Corpus linguistics for teachers. Ann Arbor, MI: University of Michigan Press.

Benson, P., & Voller, P. A. (1997). *Autonomy and independence in language learning*. London: Longman.

Bernardini, S. (2002). Exploring new directions for discovery learning. *Kettemann/Marko,* 165-182.

Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., & Quirk, R. (1999). Longman grammar of spoken and written English (Vol. 2). Cambridge: MIT Press.

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers* (Vol. 23). Amesterdam: John Benjamins Publishing.

Breen, M. P. (1985). Authenticity in the language classroom. *Applied Linguistics, 6*(1), 60-70.

Carter, R., & McCarthy, M. (1997). *Exploring spoken English* (Vol. 2). Cambridge: Cambridge University Press.

Carter, R., & McCarthy, M. (2006). *Cambridge grammar of English: A comprehensive guide; spoken and written English grammar and usage*. Stuttgart: Ernst Klett Sprachen.

Chan, T. P., & Liou, H. C. (2005). Effects of web-based concordancing instruction on EFL students' learning of verb-noun collocations. *Computer Assisted Language Learning, 18,* 231-250

Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? *System, 25,* 301-315.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213-238.

Daskalovska, N. (2015). Corpus-based versus traditional learning of collocations. *Computer Assisted Language Learning, 28,* 130-144.

Fillmore, C. J., Kay, P., & O'Connor, M. C. (1988). Regularity and idiomaticity in grammatical constructions: The case of let alone. *Language, 64*(3), 501-538.

Fligelstone, S. (1993). Some reflections on the question of teaching, from a corpus linguistics perspective. *ICAME Journal, 17,* 97-109.

Flowerdew, L. (2015). Corpus-based research and pedagogy in EAP: From lexis to genre. *Language Teaching, 48*(1), 99-116.

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

George, H. V. (1963). *Report on a verb-form frequency count. Monograph 1*. Hyderabad: Central Institute of English.

Gilmore, A. (2007). Authentic materials and authenticity in foreign language learning. *Language Teaching, 40*(2), 97-118.

Granger, S. (2002). A bird's eye view of learner corpus research. In S. Granger, J. Hung, and S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 3-33). Amsterdam: John Benjamins.

Guan, X. (2013). A study on the application of data-driven learning in vocabulary teaching and learning in China's EFL class. *Journal of Language Teaching and Research, 4,* 105-112.

Guariento, W., & Morley, J. (2001). Text and task authenticity in the EFL classroom. *ELT Journal, 55*(4), 347-353.

Holmes, J. (1988). Doubt and certainty in ESL textbooks. *Applied Linguistics, 9,* 21-44.

Hunston, S., & Francis, G. (2000). *Pattern grammar: A corpus-driven approach to the lexical grammar of English* (pp. 2000-2000). Amsterdam: John Benjamins.

Johansson, S. (2009). Some thoughts on corpora and second-language acquisition. In *Corpora and language teaching* (pp. 33-44). Amsterdam: John Benjamins.

Johns, T. (1986). Microconcord: A language-learner's research tool. *System, 14*(2), 151-162.

Johns, T. (1994). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. In T. Odlin (Ed.), *Perspectives on pedagogical grammar* (pp. 27-45). Cambridge: Cambridge University Press.

Johns, T. (2002). Data-driven learning: The perpetual challenge. *Language and Computers, 42*(1), 107-117.

Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. New York: Addison-Wesley Longman Ltd..

Little, D., Devitt, S., & Singleton, D. (1989). *Learning foreign languages from authentic texts: Theory and practice*. Dublin, Ireland: Authentik.

Ljung, M. (1991). Swedish TEFL meets reality. In S. Johansson and A. B. Stenström (Eds.), *English computer corpora* (pp. 245-256). Berlin: Mouton de Gruyter.

Ludeling, A., & Kyto, M. (2008). *Corpus linguistics: An international handbook*. Berlin: Mouton De.

Marrow, A. J. (1977). *The practical theorist: The life and work of Kurt Lewin*. New York: Teachers College Press.

Mauranen, A. (2004). Spoken corpus for an ordinary learner. In *How to use corpora in language teaching* (pp. 89-105). Amsterdam: John Benjamins.

Mayor, M. (2009). *Longman dictionary of contemporary English* (Vol. 5). Delhi: Pearson Education India.

McCarthy, M. (1998). *Spoken language and applied linguistics*. Cambridge: Cambridge University Press.

McCarthy, M. (2001). *Issues in applied linguistics*. Cambridge: Cambridge University Press.

McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. London, UK: Taylor & Francis.

Mukherjee, J. (2004). Bridging the gap between applied corpus linguistics and the reality of English language teaching in Germany. *Connor/Upton,* 239-250.

Nation, I. S. (2001). *Learning vocabulary in another language*. Stuttgart: Ernst Klett Sprachen.

Nesselhauf, N. (2004). Learner corpora and their potential for language teaching. In *How to use corpora in language teaching*. Amesterdam: John Benjamins.

Nunan, D. (1988). *Syllabus design*. Oxford. Oxford University Press.

Özdemir, N. Ö. (2014). Using corpus data to teach collocations in medical English. *Journal of Second Language Teaching & Research, 3*(1), 37-52.

Pereira, L. A. S. (2004). The use of concordancing in the teaching of Portuguese. In *How to use corpora in language teaching*. Amesterdam: John Benjamins.

Porter, D., & Roberts, J. (1981). Authentic listening activities. *ELT Journal, 36*(1), 37-47.

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

Rundell, M. (2002). *MacMillan English dictionary for advanced learners*. New York City: MacMillan.

Swaffar, J., & A. Vlatten (1997). A sequential model for video viewing in the foreign language curriculum. *The Modern Language Journal, 81*(2), 175-185.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. *Input in Second Language Acquisition, 15,* 165-179.

Scheffler, I. E. (2011). *Mitochondria*. New York, USA: John Wiley & Sons.

Scott, M., & Tribble, C. (2006). *Textual patterns: Key words and corpus analysis in language education* (Vol. 22). Amesterdam: John Benjamins Publishing.

Shaw, P. (1991). Science research students' composing processes. *English for Specific Purposes, 10*(3), 189-206.

Sinclair, J. (1987). *Collins Cobuild English language dictionary*. New York City: HarperCollins.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford. Oxford University Press.

Sinclair, J. McH. (1999). The lexical item. In E. Weigand (Ed.), *Contrastive lexical semantics* (pp. 1-24). Amsterdam: John Benjamins.

Sinclair, J. M. (Ed.). (2004). *How to use corpora in language teaching* (Vol. 12). Amesterdam: John Benjamins Publishing.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Smith, S., Chen, A., & Kilgarriff, A. (2008). A corpus query tool for SLA: Learning mandarin with the help of Sketch Engine. In B. Lewandowska-Tomaszczyk (Ed.), *Corpus linguistics, computer tools, and applications: State of the art* (pp. 673-686). Frankfurt: Peter Lang.

Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. Oxford: Blackwell Publishers.

Stubbs, M. (2007). An example of frequent English phraseology: distributions, structures and functions. *Language and Computers, 62*(1), 89-105.

Svartvik, J. (1980). Well in conversation. *Studies in English Linguistics for Randolph Quirk, 5,* 167-177.

Swaffar, J. K. (1985). Reading authentic texts in a foreign language: A cognitive model. *The Modern Language Journal, 69*(1), 15-34.

Swales, J. (2001). Integrated and fragmented worlds: EAP materials and corpus linguistics. In J. Flowerdew (Ed.), *Academic discourse* (pp. 153-167). London: Longman

Tomlinson, B. (2011). *Materials development in language teaching*. Cambridge. Cambridge University Press.

Tomlinson, B. (2013). *Applied linguistics and materials development*. Edinburgh: A&C Black.

Vongpumivitch, V., Huang, J. Y., & Chang, Y. C. (2009). Frequency analysis of the words in the Academic Word List (AWL) and non-AWL content words in applied linguistics research papers. *English for Specific Purposes, 28*(1), 33-41.

Widdowson, H. (1978). *Teaching language as communication*. Oxford: Oxford University Press.

Widdowson, H. G. (1979). The authenticity of language data. In *Explorations in applied linguistics* (pp. 163-172). Oxford: Oxford University Press.

Widdowson, H. (2007). Un-applied linguistics and communicative language teaching. *International Journal of Applied Linguistics, 17*(2), 214-220.

Willis, J. (1996). *A framework for task-based learning*. London: Longman.

Willis, D., & Willis, J. (1996). Consciousness-raising activities in the language classroom. In Willis, J. and Willis, D. (Eds.), *Challenge and change in language teaching*. Oxford: Heinemann ELT.

Xie, Q. (2015). Recent developments in corpus linguistics and corpus-based Research/Department of Linguistics and Modern Language Studies at the Hong Kong Institute of Education. *Language Teaching, 48*(1), 156-160.

Zorzi, D. (2001). The pedagogic use of spoken corpora: Introducing corpus concordancing in the classroom. In G. Aston (Ed.), *Learning with corpora* (pp. 85-107). Bologna: CLUEB.