

A Corpus-Aided Approach in EFL Instruction: A Case Study of Chinese EFL Learners' Use of the Infinitive

Ying Zhang¹ & Lin Liu²

¹ Foreign Languages Teaching and Research Department, Qufu Normal University, Rizhao, China

² School of Foreign languages, University of Jinan, Jinan, China

Correspondence: Zhang Ying, Foreign Languages Teaching and Research Department, Qufu Normal University, No. 80, Yantai Road, 276826 Rizhao, Shandong, China. Tel: 86-187-6337-6766. E-mail: qufuyzh@163.com

Received: March 14, 2014 Accepted: May 6, 2014 Online Published: June 13, 2014

doi:10.5539/elt.v7n7p152 URL: <http://dx.doi.org/10.5539/elt.v7n7p152>

Abstract

English language corpora, containing the widest possible range of varieties of English, provide empirical data concerning language usage, helping to redefine the notion of 'standard' to which language learners should aspire. This paper takes as its theoretical framework an approach to corpus-aided discovery learning in which the central role of corpora is seen as that of providing rich sources of autonomous learning activities. Here, by investigating Chinese EFL learners' use of the infinitive 'to' in the Chinese Learner English Corpus (CLEC), the suggestion is put forward that availability of different corpora and software tools and the ability to combine them in different ways may help learners develop an understanding of the patterned quality language and be conducive to more appropriate use, as learners are not just to observe patterns, but also to develop hypotheses as to their variability. Finally, a corpus-aided approach is proposed to provide insight into new perspectives to revolutionize EFL instruction.

Keywords: English language corpora, CLEC, discovery learning, EFL instruction

1. Introduction

A corpus is a body of text or speech that provides a representative sample of a language. The availability of large, online native corpora provides a straightforward tool for making a comparison. Such native corpora as the American National Corpus (ANC), Corpus of Contemporary American English (COCA) and the British National Corpus (BNC) have plenty of examples of fictions, magazines, newspapers and academic writings that demonstrate the frequent patterns and changes in the spoken and written varieties of English. The people recorded in the corpora come from different regions of the countries and incorporate a range of ages, social classes, and gender. While the learner corpora are collections of authentic texts produced by non-native speakers such as the Chinese Learner English Corpus (CLEC) which consists of one million words of written compositions by 5 types of learners: senior middle-school, tertiary college English (band 4), tertiary college English (band 6), tertiary majors in English (1st and 2nd years), tertiary majors in English (3rd and 4th years) and is annotated with grammatical tags (automatically) and error tags (manually). Inevitably, corpora are becoming increasingly popular within linguistics to evaluate existing natural language systems, investigate the occurrence of linguistic features and the production of probabilistic models of language. Besides, access has become fairly easy on standard small computers, user-friendly software is available for most normal tasks, websites are accumulating fast, and corpora are almost part of the pedagogical landscape (Sinclair, 2004).

2. Benefits of Corpus Analysis in EFL Instruction

Corpora have a distinct advantage in enabling learners to achieve language awareness and sensitivity. Corpora are capable of supplying a comprehensive description of language. The large amount of storage of texts gives enough resources to shed light on remarkable aspects of language. Such national corpora as ANC, BNC and COCA are electronically stored and processed and available on-line, which can be used to do statistical analysis and hypothesis testing, checking occurrences or validating linguistic rules within a specific language territory. Being provided access to authentic interaction, learners are highly motivated when making a close observation of how the target language is used in certain contexts.

The convergence between teaching and text corpora facilitates EFL learners' autonomous learning. Johns' (1991)

work on data-driven learning (DDL) has been proved extremely influential and ground-breaking in showing the relevance of corpus analysis techniques to the wide and varied audience of language teachers and learners around the world. Teaching is to be learner-centered and learners are encouraged to discover the foreign language, taking responsibility for their own learning, i.e. to elicit autonomous findings by employing concordance lines from a reference corpus, which helps to develop learning capacities and establish a non-authoritarian learning environment. In turn, autonomy and responsibility are conducive to increased motivation to learn and consequently to increased learning effectiveness. Through the analysis of large corpora of authentic language with the help of sophisticated concordance software, learners do no longer have to rely on the intuitions of prescriptive scholars but can inductively draw their own conclusions, which seems to be a highly desirable goal in the age of “learner autonomy” (Kettleman & Marko, 2002). Thereby, doing corpus analysis can develop linguistic awareness and encourage learning autonomy.

Learner corpora allow us the possibility of investigating learners’ distinguishing features. Describing learner language is a primary objective and a most important approach to the study of second language acquisition (Ellis, 1997). Corpora are eligible for collective comparisons in terms of the frequency of given words and phrases, the internal and external structures of phrases and the composition of sentences containing key words. Therefore, corpora make it easier to study the features of the learner language and to illustrate how and in what aspects they differ from the native speakers’ typical features.

Text corpora, providing empirical data concerning language usage, compensate for the lack of authenticity of EFL teaching materials and the limitedness of the teachers’ language sensitivity. Thus, a corpus-based EFL teaching makes the teaching objective much more specifically targeted, and the teaching syllabus together with the wordlist much more reliable.

3. Literature Review

The views on corpus use in the classroom have been discussed widely abroad. Researches abroad are mainly centered on how to use corpora to solve the practical teaching problems (Fox, 1998; Aston, 2001; Houston, 2006; Baker, 2006; Granger, 2007), how to use corpora to reflect on classroom teaching and to evaluate the effectiveness of corpora-aided classroom teaching (Burnard & McEnery, 2000), and how to harness corpora to develop quality coursebooks (McEnery & Tono, 2006).

However, in China, the actual use of corpora in language learning settings has for a long time remained somewhat backward. It was not until the late of 1990s that researchers have become interested in the corpus linguistic approach. Zhen Fengchao (2005) retraced the development of data-driven learning. Liang (2008) and He (2010) introduced the application of corpus tools in foreign language teaching and researches. Gui and Yang (2003) performed a systematic study of CLEC. Pan (2012) attempted to apply corpora data to language teaching. Despite that, the effectiveness of the corpus-aid teaching is not satisfactory. Current corpora, primarily targeted at language researches, are usually large-scaled with a wide range of genres and subjects, which can hardly meet with the teaching objectives. Moreover, present researches mainly focus on the typical features of the learner language. Such issues as the construction of the specialized corpora, the advantages of corpora in designing quality teaching materials are hardly seen. Hence, the purpose of the study is to construct a systematic corpus-aided approach in EFL teaching by making a comparative analysis of the typical outputs of the native speakers and the Chinese English learners in respect of the use of the infinitive. Particularly, this research will focus on those ideas helping us rethink language pedagogy from a corpus perspective. Hopefully, it may provide insight into new perspectives to revolutionize EFL instruction and to evaluate teaching materials.

4. Corpus Analysis in EFL Instruction—A Case Study of Chinese English Learners’ Use of the Infinitive

The analysis in this article uses a mixed method approach, combining qualitative and quantitative data. The primary source of qualitative data is COCA. To investigate the typical features of the infinitive, the researcher first proposes hypothesis and then collects data to test the hypothesis with the help of concordance and frequency list. Specifically speaking, how the infinitive is used is manifested through the abundant authentic data extracted from COCA. The primary sources of quantitative data are LOCNESS and CLEC. With the help of AntConc3.2.2w, the researcher first summarizes the distinguishing features of the Chinese English learners in using the infinitive, followed by an analysis of such features, and then makes a comparative study between CLEC and LOCNESS concerning the use of the infinitive.

4.1 Providing Rich Resources of Autonomous Learning Activities

The native corpora with abundant empirical data are further renewed with the passage of time. Thus they are quite authoritative in displaying language patterns. Take the use of the infinitive as an example. Typical features

of the infinitive are manifested by searching the native corpora. (i) The position of the infinitive or infinitive phrase is flexible when it acts as the subject. For instance, *to be content with little is true happiness. It is better to lose honorably than to succeed with dishonesty. However, it depends whether it should be placed at the beginning or end of the sentence.* For the sake of sentence balance, it should be placed at the very beginning. *To mimic him would be impersonal and possibly be perceived as mockery.* For the sake of coherence. *I cannot discard clothes if they were gifts, no matter how hideous. To do so would make me feel ungrateful for friendship.* For the sake of highlighting subjects when used in pairs. *To love for the sake of being loved is human, but to love for the sake of loving is angelic.* Despite that, “it” can be employed as the formal subject with the infinitive or infinitive phrase postposed when it is too long. *As Mark Twain said, it is better to deserve an honor and not have it than to have it and not deserve it, because dignity is not in possessing but deserving. It is better to lose honorably than to succeed with dishonesty. Losing honorably may signify lack of preparation but dishonest winning signifies lack of character.* Besides, it needs to be postposed when the behavior expressed by the infinitive is to be evaluated. *It makes sense to be optimistic when a goal is far away, and more realistic when it's close at hand. That allows us to prepare for an unexpected setback.* Moreover, “it” can be used together with “take” to draw forth the conditions needed to perform a certain behavior. *Faultfinding expends so much negative energy that nothing is left over for positive action. It takes courage and strength to solve the genuine problems that afflict every society.* (ii) Such infinitive phrases as “*to be sure, to begin with, to make matters worse, to be honest, to tell you the truth, to conclude, to sum up, to summarize, to start with*” can be put at the very beginning of the sentence, just to show the speakers’ attitude towards what has been said or just to supplement or highlight. *The heckling doesn't bother me. To be honest, it's something I look forward to.* (iii) As for the negation of “*too...to*”, attention needs to be paid. (a) *It is never too early to start teaching children a sense of duty.* (b) *He is too smart not to see your point.* a negates the idea expressed by “*too...to*”, while b negates the idea expressed by the infinitive phrase. (iv) The adverbs “*only, but, all*” can be used right before “*too...to*” for emphasis. *When the magazine asked me to provide readers with helpful tips, I was only too happy to share what I have learned.*

As can be seen from the discussion above, corpora are extremely helpful in that they provide the teachers and learners with an abundance of authentic materials which may help the learners sum up and discover the general patterns of language usage. Therefore, it is of great value to take them as teaching resources to facilitate autonomous learning and to optimize EFL teaching.

4.2 Providing Effective Teaching Tools

CAI (Computer-Assisted Instruction) and MAI (Multimedia-Assisted Instruction) put special emphasis on the outside teaching environment by resorting to aural and visual images. While corpora present the teachers as well as learners with such effective tools as concordance, frequency list, cluster list, wordlist, keyword list etc. to help with learners’ autonomous learning. A minimized pedagogical corpus with varied subjects acceptable to the learners can be built with the help of those tools. Thereby teachers can teach materials relevant to the learners and teach the most useful and most frequently used items. The following examples, extracted from the native corpora based on the frequency of occurrence, has just proved how the structure *It [be] [adjective] to [do]* is used.

It is hard to imagine life without the Internet.

It is fair to say she reinvented her life.

It is interesting to note that alcohol use is regarded as a stressor by some teenagers.

It is good to see them get rewarded.

It is nice to have an intelligent conversation like this from time to time.

It is difficult to determine the scope of the problem.

It is reasonable to assume that he was in serious shock.

It is easy to be enthusiastic about creating something new.

It is impossible to know how people choose their paths through grief.

4.3 Making Comparative Studies

It is significant to note that a comparative study between the native corpora and the learner corpora helps to identify the learners’ distinguishing features in using the language. On the one hand, a systematic study of the learner corpus helps to identify the words, phrases or even structures that are overused, underused, mistakenly used and fossilized. On the other hand, from the perspective of foreign language teaching, the distinctive features of the foreign language are what are most frequently used and the differences between the mother tongue and the

foreign language represent the difficulties in foreign language learning. Therefore, it is necessary to take the scientific discoveries on learners' interlanguage into account, which provides further feedback to teachers in designing classroom activities, teaching syllabus as well as in compiling teaching materials.

With the help of AntConc3.2.2w, the author has found the distinguishing features of the Chinese English learners in using the infinitive. They are marked by (i) topic-prominence, (ii) pseudo-passive, (iii) co-occurrence of two finite verbs without any cohesive devices, (iv) inability to arrange the position of the infinitive or infinitive phrase when it acts as the subject, which can be seen from the following sentences retrieved from CLEC.

Example 1 The young people are highly necessary to work in the countryside.

Example 2 Firstly, a short passage needs skim.

Example 3 Don't forget remove the weeds.

Example 4 Many students want show their singing in the performance.

Example 5 Some students don't know make full use of it.

Example 6 You can go buy it.

Example 7 To eat more fruits and vegetables than sugar and salt is better.

It is generally believed that Chinese is topic-prominent while English is subject-prominent. In other words, topics are what are discussed in the sentence and are usually put at the very beginning. Subjects are often nominal phrases that have an illocutionary effect on the predicates. Besides, in Chinese, those put at the beginning are relatively random, mainly depended by the thoughts or ideas to be expressed. However, in English, subjects must be in accordance with the predicates in person and number. Example 1 shows that the Chinese English learners' use of the infinitive are interfered with by the different language typology.

Pseudo-passive is another principal feature of the Chinese learners in employing the passive voice, i.e., a passive voice is necessary whereas it is unnoticed, as shown in Example 2. Here it is also important to make the learners know the different ways to express the passive meaning, which are further summarized by resorting to examples from the native corpora. (i) A short passage needs *to be skimmed*. (ii) People hardly ever need *training* to be emotional. We laugh early in life, and we are born crying. (iii) I carry bamboo chopsticks: They're cheap, light, sustainable, heat-resistant, and *easy to clean*. (iv) If you're suffering from headaches, depression or hair loss, your food choices may *be to blame* (v) Office *to let*.

Examples 3, 4, 5, 6 are unacceptable in that two finite verbs are put together without any cohesive device. There is a big difference between Chinese and English when we finish a sentence. Chinese depends much on semantic relations and a series of behaviors can be presented in time sequence. However, English is morphological and the main verb needs to be accordant with the subject in person and number with other verbs being non-finite.

The research has also noted that Chinese learners tend to underuse the infinitive phrases when they act as the adverbials. Indeed, the infinitive phrases can be widely employed to express the adverbial meaning of purpose, condition, result, manner, reason, etc., which can be observed from the following sentences from the native corpora.

To get her help, you will do better.

Always remember, if the deal sounds too good to be true, it just might be.

She'd sigh and shake her head a little, as if to say we all have our burdens to bear.

I'm proud to be an angry mom. Angry mom means a mom who cares enough to stand up for her child's health. You get angry when your boundary has been violated, and the food industry has violated our boundaries with what they are offering our kids.

Moreover, a comparison between CLEC and LOCNESS (The Louvain Corpus of Native English Essays, a corpus of native English essays made up of British pupils' A level essays, British university students' essays and American university students' essays with the total number of words 324,304) has been conducted to see how the Chinese learners differ from the native speakers in terms of the frequency of tenses and aspects of the infinitive per 100000 words. The findings are as follows. "*to have been done*", "*to be doing*" and "*to have done*" are rarely used in both of them. LOCNESS is twice as much as CLEC in using "*to be done*". Thus, when setting classroom teaching objectives, teachers should have a clear mind about what to be imparted. And a data driven courseware can be designed to make the learners discover and explore the language patterns in the process of second language acquisition.

Therefore, a corpus-aided approach is inductive: from language data to grammatical generalization. Wherever possible, frequency data are supplied. It is possible to distinguish central features of the language from peripheral ones.

5. Implications

5.1 Developing a Remedial English Grammar for Chinese English Learners

Coursebooks are best seen as a resource of achieving aims and objectives that have already been set in terms of learner needs. It is generally accepted that the role of the coursebooks is to provide the learners with authentic materials of how language is vividly used in different contexts. However, it has to be recognized that there is a striking distance between the curriculum subjects, namely, the content and the order of the grammatical items in the teaching materials and the native speakers' actual use of the language. Any material compiled on the basis of the compilers' intuition is imperfect since it is a violation of the core principles in using language, which to a great extent hinders the development of the learners' communicative competence. Whereas, native corpora afford a much more representative and reliable overall view of how language is used, having an advantage in providing a reasonably solid foundation and a good reference source for selecting authentic, natural and typical teaching materials and grading them in order of difficulty. Besides, the learner corpora help to enable better decisions as to which grammatical structures should be included in quality grammar coursebooks, thereby helping to meet the needs of the learners to the highest degree. Thus, it is strongly advisable to develop a remedial English grammar for Chinese English learners by resorting to both native and learner corpora as well as research findings to examine how specific grammatical items are dealt with, particularly those which relate to learners' learning needs, syllabus requirements, etc., thus making the teaching materials and teaching syllabus much more suitable and scientific.

5.2 Constructing a Corpus-Aided Teaching Model for Chinese English Learners

As can be observed, a corpus-aided approach in EFL teaching provides scientific guidance for EFL teaching practice. This study attempts to propose a corpus-aided teaching model for Chinese English learners — “1 basis +2 analyses +2 effects +2 objectives”. In other words, On the basis of doing corpus analysis, a minimized pedagogical corpus, say a subcorpus from published corpora such as COCA and CLEC, which contains the collections of texts written by Chinese learners themselves as well as texts illustrating a particular text-type or domain of use, is to be constructed by resorting to Error Analysis and Contrastive Interlanguage Analysis (CIA), on the basis of which, a new teaching model is to be constructed to make teaching much more targeted and scientific, with intent to develop Chinese English learners' language awareness and communicative competence.

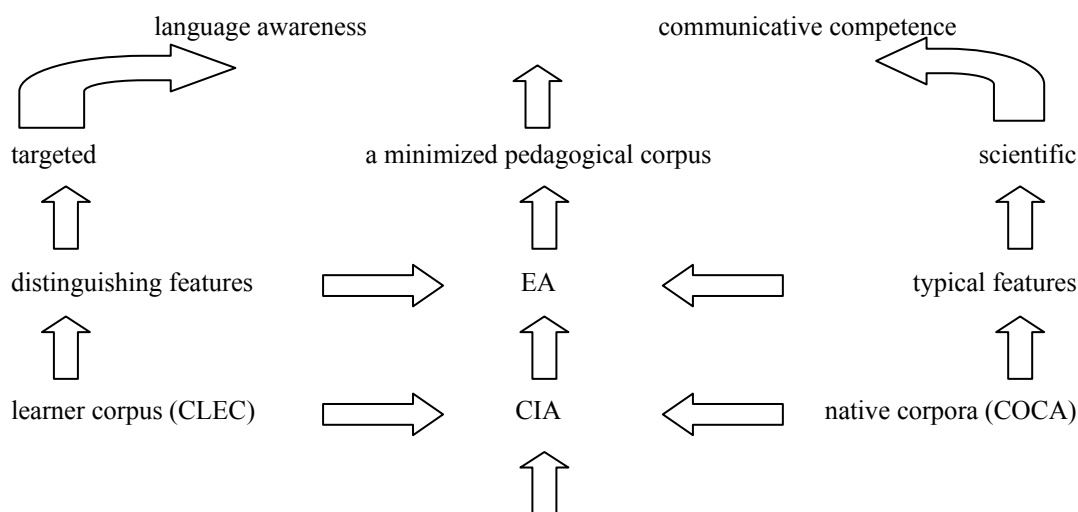


Figure 1. A Corpus-aided teaching model for Chinese English learners

6. Conclusion

All in all, a corpus-aided teaching model is highly productive. Nevertheless, it should in any way abide by the principles of language learning and teaching. 1) A shift from a teacher-centered or materials-centered to a learner-centered perspective for learning in the teaching concept is a prerequisite, which contributes to learners'

autonomous learning. 2) Teaching students in accordance with their altitude is highly advocated. Thus constructing corpora of different difficulty is necessary to meet with different teaching perspectives. 3) Teaching materials are to be useful, informative, interesting and flexible with an eye to promoting learners' language learning motivation and reinforcing autonomous learning competence and communicative competence.

References

- Aston, G. (Ed.). (2001). *Learning with Corpora*. Houston, Texas: Athelstan.
- Baker, P. (2006). *Using Corpora in Discourse Analysis*. London: Continuum.
- Brown, H. D. (2000). *Principles of Language Learning and Teaching*. Longman, Inc.
- Burnard, L., & McEnery, T. (Eds.). (2000). *Rethinking Language Pedagogy from a Corpus Perspective*. New York: Peter Lang.
- Cheng, X., & Sun, X. (2011). *An Overall Overview of English Materials*. Beijing: Foreign Language Teaching and Research Press (in Chinese).
- Collins Cobuild Dictionary. (2nd ed.). (1995). London: HarperCollins.
- Cunningsworth, A. (2002). *Choosing Your Coursebook*. Shanghai: Shanghai Foreign Language Education Press.
- Ellis, R. (1997). *Second Language Acquisition*. Oxford: Oxford University Press.
- Fox, G. (1998). *Using Corpus Data in the Classroom // Tomlinson, B. Materials Development in Language Teaching*. Cambridge: Cambridge University Press.
- Granger, S. (2009). The contribution of Learner Corpora to Second Language Acquisition and Foreign Language Teaching: A critical Evaluation. In K. Aijmer (Ed.), *Corpora and Language Teaching* (pp. 13-32). Amsterdam/Philadelphia: Benjamins, 2009.
- Gui, S., & Yang, H. (2003). *The Chinese Learner English Corpus*. Shanghai: Shanghai Foreign Language Education Press.
- He, A. (2010). *A Corpus-based Approach to English Language Teaching*. Beijing: Foreign Language Teaching and Research Press.
- Houston, S. (2002/2006). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press. Beijing: Beijing World Publishing Corporation.
- Johns, T. (1991). Should You be Persuaded—Two Examples of Data-driven Learning Materials. *English Language Research Journal*, 1991(4), 1-16. Retrieved from <http://www.docin.com/p-269633683.html>
- Johns, T. (2002). Data-driven learning: The perpetual challenge. In B. Ketterman, & G. Marko (Eds.), *Teaching and Learning by Doing Corpus Analysis* (pp. 107-117). Amsterdam and New York: Rodop, 2002.
- Kettemann, B., & Marko, G. (2002). *Teaching and Learning by Doing Corpus Analysis*. Amsterdam and New York: Rodopi.
- Liang, M., Li, W., & Xu, J. (2010). *Using Corpora: A Practical Coursebook*. Beijing: Foreign Language Teaching and Research Press (in Chinese).
- McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-Based Language Studies: An Advanced Resource Book* (pp. 52-58). London: Routledge.
- Fan, P. (2012). *A Corpus-based Approach in Language Researches and Teaching*. Beijing: China Social Science Press (in Chinese).
- Sinclair, J. et al. (1990). *Collins Cobuild English Grammar*. London Collins.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair, J. (Ed.). (2004). *How to Use Corpora in Language Teaching*. Amsterdam: John Benjamins.
- Yang, H. (2002). *An Introduction to Corpus Linguistics*. Shanghai: Shanghai Foreign Language Education Press (in Chinese).
- Zhang, J., & Zhao, X. (2012). *Corpus-based Studies of Linguistic Features in English Texts*. Shanghai: Shanghai Jiaotong University Press (in Chinese).
- Zhen, F., & Wang, H. (2010). The Application of Learner Corpora to Foreign Language Pedagogy. *Foreign Language World*, 2010(6), 72-76. Retrieved from <http://www.cnki.com.cn/Article/CJFDTotal-WYJY201006014.htm>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).