

The Application of Corpus in The Teaching of Conditional Clause

Wenyu Guan

School of Foreign Studies, Zhaoqing University, Guangdong province, 526020, China

wenyu2013@qq.com

Abstract

This paper, in line with the new trends in grammar teaching research, discusses the impact of intentional attention and consciousness arousal on language learning. It outlines the characteristics of corpora and their application in grammar teaching. Taking the teaching of conditional clause as an example, it utilizes the Michigan Corpus of Academic Spoken English (MICASE) to design grammar exercises, demonstrating through examples the significant role that corpora play in grammar instruction.

Keywords

Corpus, conditional clause teaching.

1. Introduction

Grammar teaching is an important component of language instruction[1]. Since the adoption of the grammar-translation method in grammar teaching, the deductive approach has been the privilege of many grammar textbooks and self-study grammar books [2]. In fact, in classroom teaching, the introduction of a particular grammatical item usually involves both deductive and inductive approaches. The former explicitly presents learners with grammatical rules, which are then confirmed through specific examples. The latter allows learners to infer or summarize the grammatical rules themselves after observing a series of linguistic examples. Numerous studies have shown that the traditional demonstration-practice model no longer has support and encouragement[3]. Some experts believed that understanding all the grammatical rules would naturally lead to communication; later influenced by functional-notional syllabuses and communicative syllabuses, systematic grammar teaching was neglected. Nowadays, people can use large computer corpora to describe and analyze linguistic structures, bringing renewed attention to grammar teaching and defining it from a more comprehensive and objective perspective[4].

2. Theoretical Background

Larsen-Freeman [5,6] proposed three levels of grammar teaching and divided grammar teaching into three stages: demonstration, practice, and communication. Furthermore, as corpora are used in the field of linguistic research, there is an increasing awareness of the complexity of language, and it is recognized that while grammatical rules taught by books or teachers can be helpful to students, they are often too simplistic and generalized, even misrepresenting real language, which is not conducive to students' flexible mastery of language forms. Consequently, many alternative teaching methods and approaches have been proposed, including the inductive method, observational method, focused attention, and consciousness-raising teaching methods. These methods differ in various aspects but share a common point to some extent: they do not explicitly demonstrate grammatical rules to learners but instead guide them through certain methods or approaches to observe and discover the rules and characteristics of language use on their own.

Larsen-Freeman challenges the traditional view of grammar, introducing the concept of "grammaring." [6] She argues that grammar is not merely linguistic knowledge; the ability to correctly use grammatical structures is also a skill that requires generative training. Grammar is not a rigid set of structural rules but a meaningful, dynamic system, and thus, grammar teaching should be conducted in a meaningful, focused, and interactive manner. This kind of training involves learners in certain communicative tasks, which necessitate the use of specific structures for task completion. This encompasses the relationship between form (morphosyntax/form), meaning (semantics/meaning), and usage (pragmatics/use). She points out that the goal of grammar instruction is to enable students to master both knowledge and skills, helping them to accurately, meaningfully, and appropriately grasp English grammatical structures [5]. She proposes various teaching methods to achieve this objective. The first method is to draw learners' attention to the characteristics of specific grammatical structures, with a more proactive approach being to increase input quantity; the second is consciousness-raising (C-R), which involves learners deriving general grammatical rules from language materials through C-R activities.

Schmidt [7] posits that the role of education is not necessarily to provide explicit instruction to students, but rather to guide their attention and awaken their linguistic awareness. He emphasizes that conscious noticing to language forms is a necessary condition for language learning. He also points out six major factors influencing such attention: The first is frequency, meaning that under equal conditions, the language form that appears most frequently is most likely to be noticed and integrated into the interlanguage system; The second is salience, which refers to the prominence of language forms in output under equal conditions, with more prominent forms being more likely to attract attention; The remaining four factors include instructional guidance, individual differences in learners' information processing abilities, preparation, and task requirements. Among these six factors, frequency, salience, instructional guidance, and task requirements can be controlled during the teaching process.

Research by many scholars abroad also acknowledges the significant role of consciousness-raising activities and focus-on-form instruction in promoting second language development [8,9]. They state that consciousness-raising approaches refer to "consciously directing students' attention to the grammatical features of the target language," essentially allowing learners to engage with linguistic materials of the target language, encouraging them to rationalize and recognize certain rules of the language system. Rutherford and Sharwood-Smith's [10] research on C-R and Universal Grammar indicates that C-R has potential advantages in aiding learners in acquiring linguistic competence because the realization of second language acquisition depends on learners' ability to compare the grammatical structures of their native and target languages. C-R helps learners consciously attend to the grammatical structure characteristics of the target language, thereby facilitating this comparison to enhance second language learning capabilities.

Nunan [11] also notes that C-R reemphasizes grammar teaching but develops it from a higher level. Ellis [12] summarizes three main characteristics of C-R: 1) highlighting specific linguistic features from extensive linguistic material to catch the learner's attention; 2) using authentic texts established for communicative use rather than simple sentences used merely for example; 3) requiring and encouraging learners to master linguistic features through analysis and examination of linguistic material. Willis and Willis [13] further specify seven types of C-R activities: identification/consolidation, semantic or structural categorization, hypothesis testing, cross-linguistic comparison, reconstruction/deconstruction, recall, and reference consultation. In these activities, a large amount of linguistic material is presented to learners in a concentrated and highlighted manner or as a series of contextualized examples, enabling them to observe, categorize, and generalize, ultimately enhancing their linguistic awareness and sensitivity to language.

Corpus and Grammar Teaching A corpus is a specialized database of linguistic information used for research purposes, composed of large volumes of language data collected from real-life situations [14].. Due to the extensive language samples and broad domains it encompasses, the corpus provides highly representative linguistic materials. Researching linguistic features using corpora has become one of their most important applications. For instance, corpora can be utilized to study word frequency, linguistic structures, meaning expression, and pragmatic features within specific domains. Corpora offer authentic, reliable, and rich data resources for conducting various language studies. In terms of linguistic research, the retrieval functions and statistical tools of corpora facilitate syntactic analysis, discourse analysis, conversational analysis, and phonetic studies. As a source of vast amounts of authentic linguistic data, corpora have seen increasingly widespread application in language teaching in recent years. Biber and Reppen [15] have noted that in ESL/EFL classrooms, the primary sources for most students and teachers when addressing English grammar questions come from three origins: grammar reference books, ESL/EFL textbooks, and intuition. They believe these sources provide a wealth of information about grammatical form and structural characteristics, with some examples also serving as methodological guides.

The most powerful tool of a corpus is the concordance. Through corpus indexing software and a vast array of authentic linguistic materials, learners are able to shorten the process of acquiring the target language. This is because computers can help learners organize large amounts of linguistic material, thereby making linguistic structures more easily recognizable. Tribble believes that "the central word index of a corpus turns the invisible into the visible" [16]. By searching for grammatical structures, tenses, function words, etc., we can instantly verify the typical usage of various grammatical structures. This approach to grammar learning presents the grammatical features to be learned within real contexts, enhancing learners' self-discovery and the cultivation of grammatical awareness. It has multiple advantages over traditional isolated grammar explanations and context-detached grammar exercises. Additionally, we can utilize co-occurrence analysis of contextual environments to list all occurrences of a word or phrase form in the corpus, along with their preceding and following contexts, and can set the length of these contexts as needed [14]. In a word, corpora have the characteristics of high information content and repeated prominence of linguistic forms, which align with the factors previously mentioned affecting learners' attention. This indicates that we can use corpora in teaching to promote learners' attention to linguistic forms. To better illustrate how the use of corpora in grammar teaching can enhance teaching and learning, the author has designed the following exercise involving a specific grammatical item—conditional sentences.

Table 1 Seven Types of Conditional Clauses

Structure	Terminology
1: if + pres. , pres.	generic factual
2: if + pres. ,will /be going to	future(predictive)
3: if + past, { would might could }	present hypothetical or counterfactual
4: if + pres. , { should must can may }	explicit inference factual or future with weakened result
5: if + { were were to } , { would could might }	present or future hypothetical or present counterfactual
6: if + { had + en have + en } , { would could might } have	past counterfactual
7: if + pres. , { would could might }	future with weakened result

In a survey on common teaching issues faced by English teachers involved in second language instruction in Los Angeles, it is found that conditional sentences ranked fifth[17]. indicating

that conditional sentences are one of the challenging aspects of grammar instruction. There are various classifications of conditional sentences. Palmer [18] categorizes them into two main types: real conditionals and unreal conditionals. Celce-Murcia and Larsen-Freeman agree with Quirk and Greenbaum's explanation of conditional sentences, defining them as "complex sentences that include a main clause and a subordinate clause, often introduced by the adverbial conjunction 'if,'" and they note that "English conditional sentences express three different semantic relationships: factual conditions, future (or predictive) conditions, and hypothetical conditions". The grammar exercises designed for this paper on conditional sentences follow these perspectives.

Related research by Hwang [17] analyzed corpora of spoken and written English across different text types and identified that out of 70 forms of conditional sentences present, only seven (see Table 1) accounted for two-thirds of the corpus.

According to Hwang's analysis[17], these seven patterns are the most frequently used in both spoken and written language corpora. Moreover, they appear more frequently in speech (4.2 per thousand words) than in writing (2.7 per thousand words). Additionally, Table 1 lists the possible semantic meanings each pattern may convey, such as expressing factual relations, contradicting current/past facts, or expressing predictions or speculations about the future. These studies provide a reference for analyzing the structural forms and semantics of conditional sentences in the following exercises.

3. Experimental Design and Procedures

3.1. Objectives

The objective is to enable learners to understand and correctly use if-conditional sentences, master how to express factual conditional relationships, future (predictive) conditional relationships, and hypothetical conditional relationships using if-conditional sentences, and accurately distinguish between real and unreal conditional sentences.

3.2. Source and Tool

The corpus source is the MICASE (Michigan Corpus of Academic Spoken English) developed by the University of Michigan (<http://www.hti.umich.edu/m/miscase>), which contains spoken English data from university students. We selected the discussion sections, focusing on segments involving native American university students.

3.3. Procedures

To begin, access the MICASE homepage at <http://www.hti.umich.edu/m/miscase>. Next, proceed to the MICASE Search interface and initiate a query: select "discussion section" as the document type; specify the speaker as "native speaker, American English", and enter the search term "if". Among the results, 394 collocations appear on the page in the form of contextual co-occurrences (due to space limitations, only a portion of the search lines are presented here). Step 3 involves randomly selecting several examples from the results to serve as teaching cases (the number of examples depends on the progress of the teaching schedule). By double-clicking on an example, one can obtain a full display of the paragraph related to that item, which may contain more than one instance of an if-conditional sentence. Practice tip: a) Pay attention to the verbs appearing before and after "if", noting any differences in their tenses. Please arrange them in order of frequency or calculate their proportions. b) Can you differentiate between real conditional sentences and unreal conditional sentences? What is your basis for this distinction? c) Together with classmates, select three real conditional sentences and three unreal conditional sentences each. Discuss them in the context of their surrounding text to determine the relationship they express: Is it a factual conditional relationship? A future (predictive) conditional relationship? Or a hypothetical conditional relationship? Fill in the results in the

table below. d) Pay attention to the tense of the verbs in these conditional clauses and their main clauses, then identify the patterns of tense changes.

4. Findings

The results showed that students, through retrieval and contextual co-occurrence analysis, found that if-conditional sentences appeared frequently in spoken discourse, averaging 6.8 per thousand words; by observing the verbs appearing in the main clause and subordinate clause of if-conditional sentences, it was discovered within the corpus.

The seven most frequently occurring forms are "if + present, present," "if + present, will/be going to," "if + present, would/could," "if + past, present," "if + past, would/might/could," "if + present progressive, present," and "what if + present," in descending order of frequency(see table2).

Table2 Results

structure	fre.	example
1. if + pres. , pres	49%	... if they're duplicated, what happens next?
2. if + pres. , will/be going to	13%	... if they do certain things, there'll be certain punishments
3. if + pres. , would/could	5.4%	they could be iden- if no crossing over occurs sure, they could be identical
4. if + past, pres.	4.3%	so but if you did that that's fine
5. if + past, would/might/could	3.8%	... if I was, George Bush's, son like the first one? I probably wouldn't've even gone to school
6. if + pres. prog. , pres	3.2%	... if you're not doing something, then people really question why you're not doing anything
7. what if + pres.	2.1%	okay, what if we go back to the Play-Doh

Based on the context and in conjunction with the research findings of Hwang[17] as presented in Table 1, it is possible to analyze the semantics expressed by different sentence structures, Structure 1,expressing factual conditional relationships, occurs most frequently, consistent with Hwang's structural classification that indicates present factual conditionals are the most common type of conditional sentences used in spoken language. Following closely behind are future conditionals (Structures 2 and 3), which express speculation about the future. Structure 5 conveys a hypothetical condition contrary to present facts, often featuring verbs such as "was" or "were" within subordinate clauses. Structures 4, 6, and 7 represent factual conditional relationships. As shown in Table2, these structures appear with a frequency that is not insignificant, but they were not listed in Hwang’s study, possibly due to differences in the sources of the corpus.

Moreover, students encounter frequent use of colloquial expressions such as "gonna" and "wanna" in conditional sentences within the corpus, which are not mentioned in grammar books. Through practice, students observe and discover various forms of conditional sentences in a large amount of authentic material, which is richer and more representative than examples seen solely in textbooks. This also enhances their ability for autonomous learning and learning through observation. The repeated appearance of linguistic forms on the contextual co-occurrence interface draws students' attention to formal characteristics, thereby deepening their impression of grammatical structures and improving their language awareness and sensitivity to grammatical constructions. Through the corpus, teachers can also find reliable and persuasive examples to more conveniently guide students in mastering language rules and understanding linguistic phenomena.

5. Conclusion

Corpus linguistics can reveal the most typical language features, discover the actual usage patterns of languages, and reflect the true nature of language. The analysis results after

statistical examination of language features can significantly reduce the randomness in foreign language teaching. In our country's foreign language grammar teaching, teachers often rely on the personal intuition of native speakers to make judgments about grammatical issues, but such judgments are often limited and unreliable. However, data obtained through actual observation and analysis using corpora are obviously more accurate and scientific. A rich corpus provides a large number of authentic and reliable examples for grammar teaching. Specific language structures repeatedly appear at the contextual co-occurrence interface, promoting intentional attention among learners, awakening language awareness, and helping students accurately, meaningfully, and appropriately grasp grammar knowledge in foreign language classrooms that combine communicative and formal teaching.

References

- [1].Fox,G.1998."Using corpus data in the classroom". In B. Tomlinson (ed.).*Material Development in Language Teaching*. Cambridge:CUP,pp:32
- [2].Fortune,A.1992."Self-study grammar practice:Learners' views and preferences". *ELT Journal*46:160-171
- [3].Skehan,P. 1998. " Second language acquisition research and task-based instruction", In J.Willis&D. Willis(eds.). *Challenge and Change in Language Teaching*. Shanghai:Shanghai Foreign Language Education Press:1-30
- [4] Anping,He2001, *Foreign Language Teaching Syllabus Design--Classroom Teaching Design &Evaluation*, Guangzhou: Guangdong Education Press,pp:86
- [5]Larsen-Freeman,D. 2001. "Teaching grammar".In M.Celce-Murcia (ed.). *Teaching English as a Second or Foreign Language* (3rd edition). Boston:Boston Heinle & Heinle Publishers:251-266
- [6]Larsen-Freeman,D. 2003. *Teaching Language:From Grammar to Grammaring*. Boston:Heinle& Heinle Publishers.
- [7]Schmidt,R. 1990."The role of consciousness in second language learning". *Applied Linguistics* 13.
- [8]Fotos,S.&R. Ellis.1991."Communicating about grammar: A task-based approach". *TESOL Quarterly* 25
- [9]Robinson,P. 1995."Review article: attention,memory,and the ·noticing hypothesis".*Language Learning* 45/2.
- [10]Rutherford,W. E. & Sharwood-Smith,M. 1985."Consciousness-raising and universal grammar". *Applied Linguistics* 6/3.
- [11]Nunan,D.1991. *Language Taching Methodology:A Textbook for Teachers*. New York: PrenticeHall.:66
- [12]Ellis,R.1993."Second language acquisition research:how does it help teachers? An interview with RodEllis". *ELT Journal* 47
- [13]Willis, D&J.Willis. 1998"Consciousness-raising activities in the language classroom".InJ.Willis &D. Willis(eds.). *Challenge and Change in Language Teaching*. Shanghai:Shanghai Foreign Language Education Press.pp:69.
- [14]Anping,He1997, "On the Research of Corpus",*Journal of Foreign Languages* Vol.5,pp:21,23
- [15]Biber,D.& R.Reppen. 1998."Comparing native and learner perspectives on English grammar:A study of complement clauses". In S.Granger (ed.). *Learner English on Computer*. London and New York:Longman.pp:145
- [16]Stevens,V. 1995. "Concordancing with language learners:Why? When? What?" *CAELL Journal* 6/2.
- [17]Xiaolei,Ma.et al translated 2002, *The Grammar Book*, Marianne Celce-Muria&DDiana Larsen-Freeman,Beijing University Publisher, pp:405,414,418
- [18]Palmer, F.R.1978. *The English Verb*. London: Longman. pp:139-149