

## 摘要

教材在英语教学中具有重要地位,它不仅是教师课堂上传输学生英语知识的重要来源,也是学生获取英语知识的一个重要通道。作为英语教学的重要组成部分,英语词汇在英语教学中也占有举足轻重的位置,它甚至会直接影响英语学习者的学业或学术成就。因此,英语教材应确保能够提供学生学业需求和成就的代表性词汇或词表。然而,通过文献检索发现,当前学界有关英语教材词汇、词频、词表的研究大都集中在中小学或者大学英语通识教育教材上,鲜有对高校英语专业教材的研究,对高校英语语言学导论教材的研究更少,从对比的视角对中外高校英语语言学导论教材在词汇、词频、词表上的研究则近乎阙如。这与语言学导论在英语专业课程体系中的核心地位不符,也与不少英语专业学生在学习语言学时感觉晦涩生词很多,导致难学的客观现实不符。因此完全有必要对语言学导论教材的词汇概貌进行研究。

本研究基于语料库语言学理论及其所属词表和词汇概貌理论,使用词表对比分析法对中外英语语言学导论教材之词汇概貌(具体包括词汇覆盖度或词频阈值)和 AntWordProfiler 初始生成词表里的关键词(包括通用词汇、学术性词汇及非词表词汇里的高频词)进行比较研究。为了方便研究的展开,本研究建立了总词数(tokens or running words)约为 150 万的英语语言学导论教材语料库(EILTC),它由两个子语料库组成:其中一个子语料库为“中国学者所编英语《语言学导论》教材语料库(CEILTC)”,它由 6 本中国学者所编的且在国内被广泛使用的英语《语言学导论》教材组成;第二个子语料库则为“外国学者所编英语《语言学导论》教材语料库(FEILTC)”,它由 6 本英美学者所编语言学导论教材组成,而且其中有四本曾被中国相应出版社引进出版。除此以外,每本教材又独立组成一个三级语料库,也就是本论文中一共有 12 个三级语料库。在自建语料库基础上,本研究以 Antony Lawrence 博士开发的 AntWordProfiler 为分析软件,以通用词表(GSL)和学术词表(AWL)、基于 BNC 和 COCA 的 25,000 词词表、英语专业四、八级词表为参照词表,对各级语料库之词汇概貌和 AntWordProfiler 初始生成词表进行了比较研究。本论文的主要研究目的在于,透过对各级语料库,重点对两个子语料库之间的比较,不仅在于分析词汇概貌,还从词汇覆盖度视角评估了中国学者所编语言学导论教材的文本难度及可读性,继而通过对 AntWordProfiler 初始生成词表里的关键词(包括通用词汇、学术性词汇及非词表词汇等)的分析,尝试建立一个适合中国英语专业大学生的语言学导论教材学术英语词表。

本研究主要发现有:1. 从学术词的覆盖度来看,中国学者所编的语言学导论教材比外国学者所编语言学导论教材高,这与先前的研究结果有差异,因此本研究认为:针对特定学科性质的文本,仅从学术词的覆盖度考察文本的难度或可读

性不一定合理，还应当考察非词表词汇的覆盖度。2. 中外 12 本语言学导论教材的非词表词汇里具有很多共同词汇，这些词汇充分体现了语言学导论教材所特有的一些词汇特征，可以归属到语言学学术词汇。3. 当词频阈值设为 95% 时，且先加上四个额外词表的覆盖度，大约 5,000 词族就能覆盖我国学者所编语言学导论教材，而外国学者所编语言学导论教材则需要 7,000 词族。这说明，为了达到对中国学者所编英语语言学导论教材词汇覆盖度的 95%，除了专有名词、语气词、缩略词、合成词之外，学生还需要学习掌握大约 5,000 个词族。而为了达到对外国学者所编英语语言学导论教材词汇覆盖度的 95%，除了专有名词、语气词、缩略词、合成词之外，学生还需要学习掌握大约 7,000 个词族。4. 英语专业四、八词表对中国学者所编语言学导论教材(CEILT1 例外)在形符上能达到 95% 以上的覆盖率，而对外国学者所编语言学导论教材在形符上则达不到 95% 的覆盖率，根据这一数据来判断，外国学者所编语言学导论教材整体比中国学者所编语言学导论教材更难。同时这一发现也表明，我国英语专业大学生在掌握了英语专业四、八级词表后是基本能够读懂我国学者所编的语言学导论教材(CEILT1 除外)的，但若要想读懂外国学者所编的语言学导论教材，则还需要继续学习更多语言学学术词汇。5. 本研究建立了一个语言学导论教材的学术英语词表(EILTWL)，该词表包括基础词表、专有名词词表、缩写词表、词缀表等四个二级词表，其中基础词表包含主要类符数为 1,411，而词族约为 795 个。经过自我验证发现，在基于经典词表基础上，语言学导论教材的学术英语词表-基础部分(EILTWL-Basic) 对中国学者所编语言学导论教材语料库的覆盖率为 6.69%，对外国学者所编语言学导论教材语料库的覆盖率为 6.41%，对中外学者所编语言学导论教材语料库的覆盖率为 6.5%，具有较好的覆盖度和信度。

这些研究发现有助于提升语言学导论教材的设计和开发；帮助教师选择难度合适的语言学教材以及改进语言学的教学策略，完善语言学课程测试，标准化招生考试中的语言学试题编制；促进学生语言学自主学习，提高学生语言学学术水平以及丰富语言学词典编纂等方面具有一定启示和参考价值。

**关键词：**语料库；词表；词汇概貌；教材；英语语言学导论

## Abstract

Textbook plays an important role in English teaching and learning. It is not only a vital resource for teachers to pass on their English knowledge to the students in class, but also a critical channel for students to acquire English knowledge. As an integral part of English language, English vocabulary also plays an important role in English teaching and learning, and even directly affect English learners' academic performance or achievement. Therefore, English textbook should provide key vocabulary or word lists to satisfy students' academic needs and to ensure their academic success. Through literature search, many studies focus on the word frequency or word lists in the English textbooks for the primary school students, secondary school students and non-English major university students. However, few such kind of research can be found in the study of English textbooks for English major university students. Furthermore, comparative study of the English-medium *Introductory Linguistics*' word frequency or word lists in the way of corpus linguistics has not been found in the present literature. The status quo of research does not conform to the *Introductory Linguistics*' core position in the curriculum for English major university students, and also does not conform to the fact that many students think it's difficult to study linguistics because of the plenty of new and obscure words in *Introductory Linguistics*. So, it is worthwhile to study the lexical profile of *Introductory Linguistics* textbooks.

Under the theory of corpus linguistics and its sub-theories of word list and lexical profile, this thesis adopts the word-lists-based comparative study to analyze the lexical profile (including lexical coverage or lexical threshold) and the keywords (equivalent to the high frequent words in General Service List, in Academic Word List and in the non-list) in English *Introductory Linguistics* textbooks by Chinese and native English editors. The lexical profile and the keywords are initially produced by a software-AntWordProfiler. At the start of this research, English *Introductory Linguistics* Textbooks Corpus(EILTC) with a total of 1.5 million tokens or running words have been created, which comprise two sub-corpora. The first sub-corpus(CEILTC) consists of six *Introductory Linguistics* textbooks compiled by Chinese scholars and widely used in China. The second sub-corpus(FEILTC) consists of 6 *Introductory Linguistics* textbooks compiled by British and American scholars, and four of which imported for publication by their Chinese corresponding publishers. In addition, each book could comprise the third level corpus, and there are 12 third level corpora in this thesis. On the basis of these self-built corpora, Dr. Antony Lawrence's software - AntWordProfiler

has been adopted to produce and analyze the lexical profile and keywords in this thesis, and the following word lists are adopted as the reference or level lists in AntWordProfiler: General Service List(GSL), Academic Word List(AWL), 25,000 words from the BNC and COCA lists, TEM-4 list and TEM-4&8 list. The main research purposes of this thesis try to make a comparative study among the three levels of corpora, focus on the comparative study between the two sub-corpora, analyze the lexical profile and the keywords (initially produced by the AntWordProfiler), evaluate the difficulty and readability of the *Introductory Linguistics* textbooks compiled by Chinese and native English scholars from the perspective of lexical coverage, and create an academic word list of *Introductory Linguistics* for the Chinese university students of English major.

The main findings of this thesis are as follows: 1. The coverage of academic words in the *Introductory Linguistics* textbooks compiled by Chinese scholars is higher than those of native English scholars, which is different from the findings of previous studies. Therefore, this study holds that it is not reasonable to judge the difficulty or readability of a discipline-specialized text only from the coverage of academic words, instead the coverage of non-lists words should be taken into account. 2. Many high frequent words distribute in all of the 12 *Introductory Linguistics* textbooks, which reflect the unique lexical features of the *Introductory Linguistics* textbooks, and these words can be categorized into academic words of linguistics. 3. When the lexical threshold is set as 95%, 5,000 word families plus the lists of proper nouns, interjections, abbreviations, and transparent compounds could cover the *Introductory Linguistics* textbooks compiled by Chinese scholars, 7,000 word families for the *Introductory Linguistics* textbooks compiled by native English scholars. This finding implicates that the students need to learn about 5,000 word families plus proper nouns, interjections, abbreviations, and transparent compounds if they wish to reach the 95% lexical coverage and fully comprehend the *Introductory Linguistics* textbooks compiled by Chinese scholars, 7,000 word families for the *Introductory Linguistics* textbooks compiled by native English scholars. 4. TEM-4&8 word list could cover the *Introductory Linguistics* textbooks compiled by Chinese scholars more than 95%, instead less than 95% for those by native English scholars. This data could lead to the conclusion that the *Introductory Linguistics* textbooks compiled by native English scholars are more difficult than those by Chinese scholars. This finding implicates that the students who have mastered the words in TEM-4&8 list could comprehend the *Introductory Linguistics* textbooks

compiled by Chinese scholars with the exception of CEILT1, instead, they have difficulties to comprehend those by native English scholars. If they wish to reach the 95% lexical coverage and fully comprehend the *Introductory Linguistics* textbooks compiled by native English scholars, they need to learn more academic words of linguistics. 5. In this thesis, an academic word list of *Introductory Linguistics* for the Chinese university students of English major (EILTWL) has been created, which consists of the list of the basic, the list of proper names, the list of abbreviations, the list of affixes. The list of the basic has about 1,411 types and 795 word families. When EILTWL-basic is selected as the level list in AntWordProfiler with the GSL and AWL, EILTWL-basic could cover the CEILTC 6.69%, the FEILTC 6.41%, and EILTC 6.5%, which means EILTWL-basic has a good lexical coverage and reliability.

In sum, these findings firstly could contribute to improving the design and development of *introductory linguistics* textbooks for the editors. Secondly, they could help teachers select suitable linguistic textbooks, improve the teaching strategies of linguistics, perfect the tests of linguistic courses, and standardize the items in the linguistic examination papers for enrollment. Thirdly, they could promote students' autonomous learning of linguistics, and improve their academic performance of linguistic studies. At last, they could provide insights and reference to the compilation of linguistic dictionaries.

**Key Words:** Corpus or Corpora; Word List or Vocabulary List; Lexical Profile; Textbooks; English-medium Introductory Linguistics