

英语六级与考研英语仔细阅读语篇难度对比

——来自语料库的证据

周海涛

中国矿业大学外国语言文化学院, 江苏 徐州

收稿日期: 2023年6月20日; 录用日期: 2023年8月2日; 发布日期: 2023年8月15日

摘要

本文采用语料库方法对近年来大学英语六级、考研英语(一)、考研英语(二)的仔细阅读语篇难度进行了比较。研究收集了近年来三种考试的共计144篇仔细阅读文本, 然后从词汇复杂度、句法复杂度、易读性和主题覆盖四个方面分别进行了分析, 以考察三种考试仔细阅读语篇难度上是否有差异以及有怎样的差异。研究发现: 总体而言, 英语六级与考研英语(一)阅读语篇难度基本一致, 且明显大于考研英语(二); 英语六级阅读选材比考研英语更加广泛, 对考生非语言知识储备要求更高。本研究对大学生英语学习和备考六级或考研英语具有参考意义。

关键词

仔细阅读语篇, 难度, 英语六级, 考研英语(一), 考研英语(二)

A Comparison of Text Difficulty Levels of Intensive Reading Tasks between CET-6 and NETEM

—Evidence from Corpus-Based Data

Haitao Zhou

School of Foreign Studies, China University of Mining and Technology, Xuzhou Jiangsu

Received: Jun. 20th, 2023; accepted: Aug. 2nd, 2023; published: Aug. 15th, 2023

Abstract

This paper reports on a corpus-based comparative study of the text difficulty levels of intensive reading tasks between College English Test Band 6 (CET-6) and two versions of National Entrance Test of English for MA/MS Candidates (NETEM-1 and NETEM-2) in recent years. A total of 144 texts

of the three kinds of examinations in recent years were collected, and then analyzed from four aspects: lexical complexity, syntactic complexity, readability and thematic range to investigate whether and to what extent intensive reading texts of the three exams differ in text difficulty levels. It is found that there is a close resemblance between CET-6 and NETEM-1 on the whole in terms of text difficulty levels, and that both are significantly more difficult than NETEM-2. Besides, the reading texts of CET-6 cover a larger thematic range compared to that of both NETEM-1 and NETEM-2, which means that CET-6 requires more extra-linguistic knowledge of examinees than the other two do. This study can be of certain significance in that it might, hopefully, provide reference for college students in English learning and preparation for CET-6 or NETEM.

Keywords

Intensive Reading Texts, Text Difficulty, CET6, NETEM-1, NETEM-2

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

全国大学英语六级考试(以下简称“六级考试”或“CET-6”)是由教育部主办、教育部教育考试院主持和实施的一项大规模标准化考试,其目的是参照《大学英语教学指南》设定的教学目标对我国大学生英语综合运用能力进行客观、准确的测量,同时也为用人单位了解我国大学生英语水平提供参考依据[1]。六级考试在每年的6月和12月各举行一次,考纲词汇为5318词,每次六级考试采用三套平行试卷,每套试卷题目都由写作、听力理解、阅读理解和翻译四个部分构成,其中阅读理解部分涵盖词汇理解(选词填空)、长篇阅读(段落匹配)和仔细阅读(单选)三种题型,总分值占整张试卷约35%。仔细阅读部分要求考生阅读两篇英文文章,每个篇章后有五个问题,考生根据对篇章的理解从每题的四个选项中选出最佳答案,其目的是考核学生在不同层面上的阅读理解能力,包括理解主旨大意和重要细节、综合分析、推测判断以及根据上下文推测词义等[1]。

全国硕士研究生入学英语考试(以下简称“考研英语”)是我国为高等学校和科研院所招收硕士研究生而设置的具有选拔性质的全国统一入学考试科目,其目的是科学、公平、有效地测试考生对英语语言的运用能力,保证被录取者具有一定的英语能力,并有利于各高等学校和科研院所在专业择优选拔[2][3]。从2010年开始,我国的考研英语分为了英语(一)和英语(二)两种考试。英语(一)是适用于报考学术型硕士和临床医学等9类专业型硕士的考生,英语(二)是适用于专业学位硕士。此外,对于诸如金融、资产评估等专业型硕士招生专业,由招生单位自行确定选用英语(一)或英语(二)作为考试科目。考研英语在每年12月底全国硕士研究生入学考试初试时举行,英语(一)和英语(二)试题结构大致相同,都分为完形填空、仔细阅读、新题型、翻译和写作四部分,考纲词汇均为5500词。其中,英语(一)和英语(二)的仔细阅读部分考察形式相同,均为要求考生阅读四篇英文文章并作答每篇文章后的5道单项选择题(四选一),考核学生理解主旨要义、具体信息、概念性含义、进行有关判断、推理和引申、根据上下文推测生词的词义等能力,该部分分值占全卷40%,占客观题总分值的三分之二[2][3]。

从上述介绍可知,六级考试和考研英语有着一些共同点:都是全国性的大规模标准化考试;都与考生的求职或是升学录取相关,作用重大;受试群体都是英语为非母语的中国大学生;都考查的英语阅读、翻译和写作能力且仔细阅读题型命题形式相同;考纲词汇量十分接近。当然,六级考试和考研英语,以

及考研英语的两种试题之间又存在着一定的差异：英语六级考试的定位是检验大学英语教学成果和学生实际英语应用能力的测验型考试，而考研英语则是考核学生是否具有进入研究生阶段学习所需英语水平的选拔性考试，这意味着二者的题目难度势必有所不同；英语(一)主要是适用于报考以学术研究导向的学术型硕士的考生，而英语(二)主要是适用于报考以实践应用为导向的专业型硕士的考生，两种类型硕士生培养目标不同，对考生的英语水平要求相应地存在一定差异。值得注意的是，若仅就仔细阅读这一题型而言，六级考试与考研英语以及考研英语的两种试题的命制形式完全相同，因而具有较高的可比性。并且，考虑到大多数高校学生在校期间既会参加六级考试，又会参加考研，研究两种考试的仔细阅读语篇难度对广大考生十分具有意义。

通过文献检索可以发现，国内已有一些关于大型英语考试阅读语篇难度对比的研究。詹先君曾利用语料库技术，采用定量与定性分析方法，对考研英语与托福考试的阅读文本进行了比较研究[4]；高博等运用类似方法，对英语专业专四和专八考试的阅读语篇难度进行了对比研究[5]；江进林与韩宝成曾使用 Coh-Metrix 工具对近年六级考试、托福考试和雅思考试的共计 163 篇阅读语篇进行了难度对比[6]；陈灿将考研英语与雅思考试的阅读语篇进行了对比[7]；付霞辉将考研英语(一)与托福考试的阅读语篇进行了对比[8]。

总体而言，大型英语考试阅读语篇难度对比研究在路径方面已经比较成熟，即大多都是借助语料库方法，从文本的形式或内容方面进行分析。但是，现有研究以国内英语考试与国际英语考试进行对比居多，对于国内的六级考试与考研英语以及考研英语的两种试题进行对比研究的相关成果较为匮乏。因此，本研究在参考和借鉴同类研究原理的基础上，试图对比六级考试与考研英语以及考研英语的两种试题仔细阅读部分的语篇难度，以对现有研究作以补充，同时为广大高校中上述考试的考生提供一定理论参考。

2. 研究设计

2.1. 语料库建立

本研究从近年来三种考试的真题中选取仔细阅读篇目共计 144 篇自建小型语料库，语料包括 2017~2020 年间 48 篇六级考试仔细阅读文本、2010~2021 年间考研英语(一) 48 篇仔细阅读文本、2010~2021 年间考研英语(二)仔细阅读文本 48 篇。所有文本都经过了人工校对，标点符号全部统一为英文状态。英语六级考试阅读语篇中一些超纲词汇的中文释义均已删除，所有阅读篇目都保存为单独的纯文本文件。

2.2. 研究工具与步骤

本研究所用工具包括：自建的小型语料库、语料库检索软件 Wordsmith 7.0、文本易读性测试软件 BFSU-Huge-Mind Readability Analyzer 2.0、在线文本词汇复杂度分析器 Web-based Lexical Complexity Analyzer (LCA)、在线文本句法复杂度分析器 Web-based L2 Syntactic Complexity Analyzer (L2SCA)和统计软件 SPSS 22.0。数据收集方法为：运用 Wordsmith 7.0 和 LCA 分析语料，获得形符、类符、类型比、平均词长、词汇密度等词汇复杂度数据；运用 L2SCA 分析语料，获得句法复杂度数据；运用 BFSU-Huge-Mind Readability Analyzer 2.0 分析文本的易读性数据；运用 Wordsmith 7.0 提取语料的主题词表。分析方法为：根据研究问题的不同，利用 SPSS 22.0 对所收集的数据进行均值计算、方差分析等统计处理，然后对三种考试的仔细阅读语篇在词汇复杂度、句法复杂度、易读性、主题范围这四个方面的特征进行对比分析。

2.3. 研究问题

本研究旨在回答以下问题：1) 六级考试与考研英语(一)、英语(二)的仔细阅读语篇在词汇复杂度、句法复杂度、易读性三个方面，两两之间是否存在显著性差异？2) 三种考试仔细阅读语篇在文本覆盖的主题方面各自具有什么特征，有怎样的差异？

3. 结果与讨论

3.1. 词汇复杂度

词汇复杂度表示文本中词汇的深度和广度[8]。一般而言,语篇的词汇复杂度越高,阅读难度越大。本研究采用类符比、平均词长和词汇密度这三个词汇复杂度的常用指标考察三种考试的阅读语篇难度。

3.1.1. 类符/形符比

形符(token)是指一个语言单位,类似于我们日常说的“词”。类符(type)指不重复计算的形符数,类符/形符比(TTR, type-token ratio)表示类符占形符的比例,是衡量文本词汇复杂度的常用方法[9]。为了避免因为文本长度差异造成的影响,通常采用标准类/形符比(STTR, standardized type-token ratio)来计算词汇密度,STTR越高,说明文本包含的信息越多,难度也就越大。采用 Wordsmith 7.0 获得三种考试仔细阅读篇目的 STTR 数据并以考试类型为因素对 STTR 数据进行单因素方差分析,结果如表 1 所示。

Table 1. One-way ANOVA results of the STTR values for the intensive reading tasks in three tests

表 1. 三种考试仔细阅读语篇 STTR 值单因素方差分析结果

	英语六级	英语(一)	英语(二)	方差 F 值	显著性 p 值
STTR	74.94	75.95	75.93	1.713	0.184

从表 1 可以看出,在标准类/形符比(STTR)这一指标上,三组文本之间不存在显著性差异($p = 0.184 > 0.05$),即三种考试的仔细阅读篇目单位文本长度中所包含的信息量上基本相同。

3.1.2. 平均词长

词汇复杂度的另一个指标是平均词长(mean word length)。词长是指一个单词所包含的字母的数量。英语中常用的词为词长较短的词,一般认为 8 个或 8 个字母以上的词可以算作长词[8]。一般而言,单词越长,使用频率越低,词汇难度越大。语篇的平均词长越长,词汇复杂度就越大,阅读难度就越大。通过 Wordsmith 7.0 获得三组考试仔细阅读语篇的平均词长数据,并以考试类型为因素对该数据进行单因素方差分析(表 2),可见三组文本的平均词长之间存在显著性差异($p = 0.000 < 0.05$)。

Table 2. ANOVA results of mean word length

表 2. 平均词长方差分析结果

	英语六级	英语(一)	英语(二)	方差 F 值	显著性 p 值
STTR	5.0048	5.0192	4.8406	8.959	0.000

Table 3. Tukey HSD post-hoc test results of mean word length

表 3. 平均词长 Tukey HSD 事后检验结果

(I)考试	(J)考试	均差(I-J)	标准误	显著性 p 值
英语六级	英语(一)	-0.01438	0.04687	1
	英语(二)	0.16417*	0.04687	0.002
英语(一)	英语六级	0.01438	0.04687	1
	英语(二)	0.17854*	0.04687	0.001
英语(二)	英语六级	-0.16417*	0.04687	0.002
	英语(一)	-0.17854*	0.04687	0.001

注: *代表 95%的显著性水平。

Tukey HSD 事后检验结果(表 3)显示,在仔细阅读篇目的平均词长方面,六级考试与英语(一)之间无显著性差异($p = 1.000 > 0.05$),二者均差仅为 0.01438 词;英语六级和英语(二)之间有显著性差异($p = 0.002 < 0.05$),前者比后者平均词长多 0.16417 词;英语(一)和英语(二)之间也存在显著性差异($p = 0.001 < 0.05$),前者比后者平均多出 0.17854 词。因此,单考虑平均词长这一指标,六级考试和英语(一)仔细阅读的词汇复杂度无显著差异,且都明显大于英语(二)。

3.1.3. 词汇密度

词汇密度(lexical density)指的是实义词数相对文本总词数的比率[9],是衡量文本词汇复杂度的又一常用指标。艾海洋、陆小飞曾将实义词界定为包含名词、形容词、动词(除去情态动词、助动词)以及带有形容词基础的副词(包括既可以作形容词也可以作副词的词,如“fast”以及通过向形容词根加后缀“-ly”形成的副词)[10]。通过陆小飞等开发的在线文本词汇复杂度分析器 LCA 分析得到三组考试仔细阅读篇目的词汇密度数据,然后以考试类型为自变量进行单因素方差分析(表 4)。结果显示,三种考试的仔细阅读文本在词汇密度方面无显著差异($p = 0.095 > 0.05$)。

Table 4. ANOVA results of lexical density

表 4. 词汇密度方差分析结果

指标	英语六级	英语(一)	英语(二)	方差 F 值	显著性 p 值
词汇密度	0.5479	0.5604	0.5471	2.39	0.095

Table 5. Syntactic complexity indicators provided by L2SCA

表 5. L2SCA 句法复杂度指标

类别	指标
第一类单位长度	1) 平均子句长度(MLC)
	2) 平均句子长度(MLS)
	3) 平均 T 单位长度(MLT)
第二类句子复杂度	每个句子中的子句数量(C/S)
第三类从属子句使用量	1) 每个 T 单位中的子句数量(C/T)
	2) 复杂 T 单位比率(即每个 T 单位中的复杂 T 单位数量)(CT/T)
	3) 从属子句比率(即每个子句中的从属子句数量)(DC/C)
第四类并列结构使用量	4) 每个 T 单位中的从属子句数量(DC/T)
	1) 每个子句中的并列短语数量(CP/C)
	2) 每个 T 单位中的并列短语数量(CP/T)
第五类特定短语结构	3) 并列句比率(即每个句子中的 T 单位数量)(T/S)
	1) 每个子句中的复杂名词性短语数量(CN/C)
	2) 每个 T 单位中的复杂名词性短语数量(CN/T)
	3) 每个 T 单位中的动词短语数量(VP/T)

通过对上述三种词汇复杂度指标的分别考察,可以发现,三种试题的仔细阅读篇目只在平均词长方面存在差异。

3.2. 句法复杂度

句法复杂度是指语言产出中句法结构的多样性及复杂性[11]。句法复杂度越大,文本的阅读难度也就

Table 6. ANOVA results of 11 syntactic complexity indicators with homogeneous variance
表 6. 方差齐性的十一项句法复杂度指标方差分析结果

指标	F[2,141]	显著性 p 值
MLS	0.722	0.487
MLT	1.44	0.24
MLC	0.515	0.599
C/S	0.589	0.556
VP/T	0.202	0.817
DC/C	0.374	0.689
C/T	1.068	0.346
DC/T	0.508	0.603
T/S	0.545	0.581
CT/T	2.249	0.109
CP/C	1.148	0.32

Table 7. Welch's test results for three syntactic complexity indicators with varying variance
表 7. 方差不齐的三项句法复杂度指标 Welch 检验结果

指标	统计	df1	df2	显著性 p 值	
CPT	Welch(W)	1.819	2	88.212	0.168
CN/T	Welch(W)	5.774	2	90.860	0.004
CN/C	Welch(W)	5.924	2	90.295	0.004

Table 8. Games-Howell post-hoc test results for the indicators of CN/T and CN/C
表 8. CN/T 和 CN/C 指标 Games-Howell 事后检验结果

变量	(I)考试	(J)考试	平均差(I-J)	标准错误	显著性 p 值
CN/T	英语六级	英语(一)	0.0446167	0.1627721	0.959
		英语(二)	0.4099375*	0.1508373	0.022
	英语(一)	英语六级	-0.0446167	0.1627721	0.959
		英语(二)	0.3653208*	0.1285699	0.015
	英语(二)	英语六级	-0.4099375*	0.1508373	0.022
		英语(一)	-0.3653208*	0.1285699	0.015
CN/C	英语六级	英语(一)	-0.0500208	0.0802964	0.808
		英语(二)	0.1623000	0.0693925	0.056
	英语(一)	英语六级	0.0500208	0.0802964	0.808
		英语(二)	0.2123208*	0.0674972	0.006
	英语(二)	英语六级	-0.1623000	0.0693925	0.056
		英语(一)	-0.2123208*	0.0674972	0.006

注: *代表 95%的显著性水平。

越大。因此, 句法复杂度是从词汇以上层面考察语篇阅读难度的重要指标。本文采用陆小飞所研发的自动句法复杂度分析器(L2SCA), 该在线软件可以实现对上传的语料进行共计五类 14 个句法复杂度指标(表 5)的自动化分析[12] [13]。

使用 L2SCA 获得三种考试仔细阅读文本句法复杂度各项指标数据, 并以考试类型为因素, 进行单因素方差分析。方差同质性检验结果显示, 在 14 个句法复杂度测量指标中, CP/T、CN/T、CN/C 这三个指标方差不齐(p 值分别为 0.033、0.005 和 0.017, 均小于 0.05), 其他 11 项指标满足方差齐性。对满足方差齐性的 11 项指标进行单因素方差分析, 结果见表 6。这 11 项指标对应的显著性 p 值均大于 0.05, 表明在这 11 项句法复杂度指标上, 三组考试的仔细阅读篇目之间无显著性差异。

对方差不齐的 CP/T、CN/T、CN/C 这三项指标进行 Welch 检验, 结果见表 7, 可见三组文本在 CP/T 这一指标上无显著差异($p = 0.168 > 0.05$), 在 CN/T、CN/C 这两个指标上存在显著性差异(p 值分别为 0.004 和 0.004, 均小于 0.05)。采用 Games-Howell 事后检验对 CN/T、CN/C 两个指标进一步分析, 结果见表 8。

由表 8 可知, 在每个 T 单位中的复杂名词性短语数量(CN/T)这一指标上, 英语六级与英语(一)之间无显著性差异($p = 0.959 > 0.05$); 英语六级与英语(二)之间存在显著性差异($p = 0.022 < 0.05$), 均差显示前者比后者平均每个 T 单位中的复杂名词性短语数量多 0.4099, 英语(一)和英语(二)之间也存在显著性差异($p = 0.015 < 0.05$), 均差显示前者比后者平均每个 T 单位中的复杂名词性短语数量多 0.3653 个。仅从这一指标来看, 英语六级难度与英语(一)高度相近, 其中前者略大, 两者难度均显著大于英语(二)。

在每个子句中的复杂名词性短语数量(CN/C)这一指标上, 英语六级与英语(一)之间无显著性差异($p = 0.808 > 0.05$); 英语六级与英语(二)之间无显著性差异($p = 0.056 > 0.05$); 英语(一)和英语(二)之间存在显著性差异($p = 0.006 < 0.05$), 均差显示前者比后者平均每个 T 单位中的复杂名词性短语数量多 0.2123 个。仅从这一指标来看, 三种考试阅读难度从大到小依次为英语(一)、英语六级、英语(二), 其中英语(一)难度显著大于英语(二)。

从句法复杂度指标方差分析结果可以看出, 在句法复杂度上, 三种考试中, 英语(一)和英语六级难度近似相同, 但都大于英语(二)。

3.3. 易读性

易读性(readability), 又称可读度, 是指文本易于阅读和理解的程度或性质。阅读材料的易读性受到词长、句法、概念负载量等多种因素的影响。文本的易读性可以通过易读性公式来表征。Flesh 易读性公式是易读性研究中比较常用的易读性公式[14] [15]。利用文本易读性测试软件 BFSU-Huge-Mind Readability Analyzer 2.0 完成三种考试阅读文本的 Flesh 易读性指数计算, 并以考试类型为因素, 对易读性数据进行单因素方差分析(结果见表 9), 方差同质性检验结果显示方差齐性, 三组文本组间存在显著性差异($p = 0.001 < 0.05$)。进行组间均值事后检验, 结果如表 10 所示。

由表 10 可知: 在仔细阅读篇目的易读性方面, 六级考试与英语(一)之间无显著性差异($p = 0.826 > 0.05$); 六级考试和英语(二)之间有显著性差异($p = 0.001 < 0.05$); 英语(一)和英语(二)之间有显著性差异

Table 9. ANOVA results for readability

表 9. 易读性方差分析结果

	英语六级	英语(一)	英语(二)	方差 F 值	显著性 p 值
均值	37.7398	38.9004	23.41	7.365	0.001
最小值	11.50	10.00	23.41		
最大值	57.78	62.08	60.51		

Table 10. Games-Howell post-hoc test results for readability
表 10. 易读性 Games-Howell 事后检验结果

(I)考试	(J)考试	均差(I-J)	标准误	显著性 p 值
1	2	-1.16062	1.97141	0.826
	3	-7.05521*	1.97141	0.001
2	1	1.16062	1.97141	0.826
	3	-5.89458*	1.97141	0.009
3	1	7.05521*	1.97141	0.001
	2	5.89458*	1.97141	0.009

注：*代表 95%的显著性水平。

($p = 0.009 < 0.05$)。结合描述性统计结果可知，六级考试仔细阅读篇目的易读性指数均值(37.7398)与英语(一)(38.9004)较接近，即二者易读程度基本一致；英语(二)的易读性指数均值(44.7950)要远大于六级考试和英语(一)，即三种考试中英语(二)的题目是最易读的，难度最小。

3.4. 主题词分布

语篇的阅读难度不仅与词汇、句法等语言形式上的特征相关，也与文本的内容直接相关。对于一个具有一定背景知识的读者来说，文本所包含的主题越广泛、越多样，不熟悉的内容所占的比例就越大，阅读难度也就越大[15]。因此，除了从词汇、句法、易读性这三个语言形式维度进行考察，本文通过主题词分析，从语义层面考察了三种考试的阅读文本的特征。主题词(keyword)是指在单篇或多篇文本中具有超高复现频率的词。主题词分析在语言教学、语体分析、主题分析以及文本检索中具有重要的应用价值。本文利用 Wordsmith 7.0 软件，把研究所用的三种考试仔细阅读文本语料库作为观察语料库(observed corpus)，并采用 ANC (American National Corpus)词表作为参照语料库(reference corpus)，提取出观察语料库中三种考试仔细阅读文本对应的主题词表(表 11)，从而清楚地显示三种文本在主题覆盖上的差异。

Table 11. Keyword lists of the intensive reading tasks of three tests
表 11. 三种考试仔细阅读主题词表

序号	六级主题词	观察库频数	主题值	英语(一)主题词	观察库频数	主题值	英语(二)主题词	观察库频数	主题值
1	robots	20	184.36	digital	19	80.79	students	46	167.18
2	Antarctic	15	115.86	media	31	68.06	jobs	28	99.18
3	scientists	25	111.09	companies	31	66.70	Facebook	8	89.65
4	government	8	88.06	peer	15	64.42	cent	16	88.44
5	AI	12	83.26	AI	10	64.00	workers	27	69.63
6	researchers	21	69.68	social	33	59.95	companies	27	52.74
7	marine	16	69.62	state	48	53.36	gap	15	47.16
8	organic	16	69.51	fashion	19	50.38	college	25	45.18
9	cities	20	57.12	tax	31	49.91	housing	14	39.02
10	climate	15	56.52	justices	9	45.83	social	26	36.29
11	science	22	53.72	discovery	15	44.95	recession	10	35.94
12	agriculture	13	50.42	factors	5	43.74	generation	17	32.91

Continued

13	academic	17	50.25	boards	12	42.99	UK	9	32.52
14	smartphone	5	48.85	consumers	14	42.75	count	4	31.57
15	stress	18	46.33	court	25	42.02	work	48	28.32
16	sensors	8	41.67	law	33	40.55	women	27	25.76
17	photos	11	40.14	sector	14	34.23	corporate	13	24.03
18	humans	16	39.59	print	13	33.74	economy	17	23.31
19	hunting	13	39.26	ruling	11	31.23	apps	3	23.30
20	debris	9	38.15	sales	15	30.98	robots	5	23.24
21	birds	12	33.62	unions	10	30.57	labor	15	21.50
22	breeding	9	33.21	NHS	6	29.92	zone's	3	18.80
23	choices	12	33.05	Britain	13	28.72	factors	3	18.80
24	social	26	31.96	online	15	28.35	governments	10	18.74
25	decision	21	31.95	pressure	17	26.22	tech	9	18.56
26	farmers	10	27.89	researchers	13	26.01	tasks	8	17.38
27	grades	8	25.18	scientific	13	24.08	economic	18	15.40
28	food	21	24.04	marketers	5	23.33	home	30	15.30
29	krill	6	23.49	google	3	22.88	climate	8	14.50
30	sustainability	5	23.45	prize	10	22.56	psychologist	5	13.56

由表 11 可知, 英语六级仔细阅读篇目前三十位的主题词涉及科技(robots, scientists, AI, researchers, science, smartphone, sensors)、地理(Antarctic, marine, climate, debris, sustainability)、生物(organic, birds, breeding, krill)、农业生产(agriculture, hunting, farmers)、教育(grades)、政治(government)、日常生活(stress, photos, humans, choices, decision, food)等各个领域。其中, 科技、地理、生物几类自然科学相关的主题词合计占前三十位主题词半数以上, 可见英语六级考试的阅读选文所涉领域较广, 这与考试大纲中“选材广泛, 涵盖人文学科, 社会科学, 自然科学等领域”的相关表述相一致。英语(一)仔细阅读文本前三十位的主题词涉及政治(state, ruling, Britain, unions)、法律(justices, court, law)、经贸(tax, companies, boards, consumers, sales, marketers)、科技(digital, AI, discovery, researchers, scientific, Google)、文化与传播(media, print, prize)、医疗(NHS (“英国国家医疗服务体系”))、时尚(fashion), 还有一些与社会生活非常相关的词如“peer、pressure、social, sector”。由此可见, 英语(一)的阅读选文覆盖的主题较为多样, 主要涉及的一些与社会生活高度相关的领域, 极少涉及自然科学领域相关知识。英语(二)的阅读选文也呈现出与英语(一)高度相似的特点, 其前三十位主题主要涉及经贸(workers, companies, recession, corporate, economy, economic)、教育(students, college)、科技(Facebook, Apps, robots, tech)、工作生产(jobs, work, labor)、住房(housing)、社会群体(workers, women, generation)以及一些与生活常用词(gap, zone's, factors, home), 而几乎不涉及自然科学领域的专业词汇。通过考察高频主题词发现, 无论是考研英语(一), 还是英语(二), 阅读文本的主题词大部分都来源于与社会生产生活密切相关的领域, 如政治、经贸, 文化、科技和日常生活等方面, 有关地理、生物等自然科学领域的主题词极少, 这与英语六级明显不同。由此可以推测, 考研英语命题者在选取仔细阅读文本时有一定的倾向性, 即主要选择涉及与社会生活相关的篇章。尽管不能仅凭语篇的主题覆盖特征断定英语六级和考研英语从内容上讲孰难孰易, 但是根据以上分析可以有一

定把握推测,两种考试在选择阅读篇目上的侧重点不同,英语六级的语篇明显比考研英语涉及的话题和领域更加广泛,对考生语言外知识储备的广度要求应该高于考研英语。

4. 结论与启示

本文借助语料库技术,采用定量与定性相结合的方法,从词汇复杂度、句法复杂度、易读性和主题覆盖四个维度将英语六级、考研英语(一)、考研英语(二)的仔细阅读篇目样本进行了对比分析。研究发现:

1) 在词汇复杂度方面,三种考试的阅读文本在标准类/型符比(STTR)和词汇密度方面均无显著差异,英语六级与英语(一)平均词长几乎相同,且都明显大于英语(二),所以英语六级与英语(一)难度基本相同,英语(二)难度最小。

2) 在句法复杂度方面,三种考试仅在每个T单位中的复杂名词性短语数量(CN/T)和每个子句中的复杂名词性短语数量(CN/C)这两个指标上存在差异。对于前者,英语六级和英语(一)之间无显著差异,且指标均值都大于英语(二);对于后者,英语(一)的指标均值与英语(二)之间有显著差异,前者明显大于后者,英语六级的均值介于二者之间,与二者都比较接近。总体而言,从句法复杂度上讲,英语六级与英语(一)阅读篇目难度十分相近,英语(二)难度最小。

3) 在易读性方面,英语六级与英语(一)难度基本相同,但都明显大于英语(二)。

4) 在主题覆盖上,英语六级的选篇题材最为多样,文章广泛涉及人文学科、社会科学、自然科学各个领域;相比之下,英语(一)与英语(二)选篇主要与社会生产生活的各个领域相关,涉及自然科学的科普文章较少。

5) 概括而言,单从语言的形式特征来讲,英语六级与英语(一)难度非常接近,不相上下,英语(二)难度显著小于另外二者;单从语篇内容角度来讲,英语六级阅读选材比考研英语更加广泛,对考生的语言外知识储备要求更高。

本研究的结论对英语学习有一定指导意义。例如,考生可以通过阅读六级语篇的难度感知,估测考研英语(一)或考研英语(二)文章对于自己的阅读难度,从而更有针对性地为考研英语做准备。此外,本研究所发现的英语六级与考研英语阅读选篇的题材特征,有助于指导考生的日常英语阅读练习。例如,计划参加六级考试的考生,平时应注意广泛阅读各种题材的英语文章,扩大自己的知识面;计划参加考研英语考试的同学,可在平时多阅读一些来源于英美国家的原版报刊的文章,加深自己对英语国家社会与国情方面的了解,从而在考试遇到内容相近的篇目时更加得心应手。

参考文献

- [1] 全国大学英语四、六级考试委员会. 全国大学英语四、六级考试大纲[M]. 上海: 上海交通大学出版社, 2016.
- [2] 教育部考试中心. 2022年全国硕士研究生入学统一考试英语(一)考试大纲(非英语专业) [M]. 北京: 高等教育出版社, 2021.
- [3] 教育部考试中心. 2022年全国硕士研究生入学统一考试英语(二)考试大纲(非英语专业) [M]. 北京: 高等教育出版社, 2021.
- [4] 詹先君. 外语测试阅读文本的效标关联效度验证研究——基于语料库的视角[J]. 外语测试与教学, 2015(3): 23-29.
- [5] 高博, 赵海娟. 基于语料库的英语专业等级考试阅读语篇难度对比研究[J]. 考试研究, 2017(6): 76-81.
- [6] 江进林, 韩宝成. 基于 Coh-Metrix 的大学英语六级与托福、雅思阅读语篇难度研究[J]. 中国外语, 2018, 15(3): 86-95.
- [7] 陈灿. 雅思与全国硕士研究生英语入学考试阅读测试对比研究[D]: [硕士学位论文]. 武汉: 华中师范大学, 2012.
- [8] 付霞辉. 托福与考研英语(一)阅读测试对比研究[D]: [硕士学位论文]. 武汉: 华中师范大学, 2015.
- [9] 梁茂成, 李文中, 许家金. 语料库应用教程[M]. 北京: 外语教学与研究出版社, 2010.

-
- [10] Ai, H. and Lu, X. (2013) A Corpus-Based Comparison of Syntactic Complexity in NNS and NS University Students' Writing. In: Díaz-Negrillo, A., Ballier, N. and Thompson, P., Eds., *Automatic Treatment and Analysis of Learner Corpus Data*, John Benjamins, Amsterdam, 249-264. <https://doi.org/10.1075/scl.59.15ai>
- [11] Ortega, L. (2003) Syntactic Complexity Measures and Their Relationship to L2 Proficiency: A Research Synthesis of College-Level L2 Writing. *Applied linguistics*, **24**, 492-518. <https://doi.org/10.1093/applin/24.4.492>
- [12] Casal, J.E., Lu, X., Qiu, X., *et al.* (2021) Syntactic Complexity across Academic Research Article Part-Genres: A Cross-Disciplinary Perspective. *Journal of English for Academic Purposes*, **52**, Article ID: 100996. <https://doi.org/10.1016/j.jeap.2021.100996>
- [13] 陆小飞, 许琪. 二语句法复杂度分析器及其在二语写作研究中的应用[J]. 外语教学与研究, 2016, 48(3): 409-420.
- [14] 陈练文, 李信, 赵超焯. 学术文本易读性和复杂度评价研究[J]. 数字图书馆论坛, 2018(5): 62-66.
- [15] 王艳伟. 专业英语教材与测试阅读文本易读度比较研究[J]. 外语研究, 2021, 38(2): 70-75.