

# 考研英语一与考研英语二阅读理解A部分文本复杂度对比研究

张 玲, 时真妹

大连理工大学外国语学院, 辽宁 大连

收稿日期: 2022年11月18日; 录用日期: 2023年1月9日; 发布日期: 2023年1月16日

## 摘 要

自从2010年以来, 全国硕士研究生入学英语考试划分为考研英语一(学术型)和考研英语二(专业型)。本研究以Bachman和Palmer提出的任务特征框架为理论基础, 利用Coh-Metrix和Vocab Profile对考研英语一和考研英语二阅读理解A部分的文本复杂度进行了对比研究。本研究从词汇特点、句法特点、语篇特点和易读度四个维度对两种英语阅读考试的文本进行了研究, 并运用SPSS对以上四个方面分别进行独立样本t检验, 旨在发现两者文本的难易度是否存在差异。结果显示, 考研英语一阅读文本和考研英语二阅读文本在平均词长、篇章长度、语义重叠以及易读度四个方面存在显著差异, 其余部分并不存在显著差异。

## 关键词

考研英语一, 考研英语二, 文本复杂度, 输入特征

## A Comparative Research on Text Complexity of Reading Part A in NETEM I and NETEM II

Ling Zhang, Zhenmei Shi

School of Foreign Languages, Dalian University of Technology, Dalian Liaoning

Received: Nov. 18<sup>th</sup>, 2022; accepted: Jan. 9<sup>th</sup>, 2023; published: Jan. 16<sup>th</sup>, 2023

## Abstract

Since 2010, NETEM (National Entrance Test of English for MA/MS Candidates) has been classified into NETEM I (academic-oriented) and NETEM II (professional-oriented). Based on the framework of task characteristics proposed by Bachman and Palmer, this research uses Coh-Metrix and Vocab

**Profile to conduct a comparative study on the text complexity of Reading Part A in NETEM I and NETEM II. This research studies the two English reading texts from four dimensions: lexical characteristics, syntactic characteristics, discourse characteristics and readability. SPSS is used to conduct independent-samples T test on the four aspects respectively, aiming at finding out whether there are differences in the difficulty of the two reading texts. The results show that there are significant differences in average word length, text length, semantic overlap and read ability between NETEM I and NETEM II, while no significant differences can be found in other aspects.**

## Keywords

NETEM I, NETEM II, Text Complexity, Characteristics of Input

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. 介绍

### 1.1. 研究背景

由于全球经济衰退和冠状病毒的爆发, 失业率在过去几年一直在上升。加上本科招生的扩大, 求职者之间的竞争变得越来越激烈。因此, 渴望攻读硕士学位的大学生人数正在上升。全国研究生入学考试(NGSET)每年举行一次, 以选拔有资格攻读硕士学位的非英语专业本科生。NGSET 包括四个子测试, 即外语测试、政治测试和两个与专业相关的测试。第一个考试叫做 NETEM (全国硕士研究生入学英语考试)。NETEM 旨在客观、有效地判断考生的语言能力。考核标准是考生必须是非英语专业的合格毕业生, 并达到一定的水平, 以保证高校和科研院所遴选出具备研究生学习所需知识和学术能力的人才。

自 2010 年以来, 全国硕士研究生入学英语考试被分为考研英语一(学术型)和考研英语二(专业型)。学术型硕士与专业型硕士的培养方向和培养方法完全不同。学术型硕士以理论和研究为重点, 为科研机构培养高校教师和科研人员。专业型硕士以专业实践为导向, 重视实践和应用, 培养在专业领域受过正规、高水平培训的高层次人才。越来越多的研究人员更加关注这两项测试的重要性, 并对这两项测试的有效性或难度进行了一些相关研究。这两项测试都包括英语使用(完形填空)、阅读理解和写作三个部分, 其中阅读部分在考试成绩中占比第一。阅读能力作为语言能力的重要组成部分, 不仅反映了学习者的语言能力, 也反映了语言学习者的综合水平。从语篇复杂度的角度看, 英语一和英语二的阅读部分有很多共同点, 但也有很多不同之处。作者对两个阅读测试的文本复杂度进行比较研究, 旨在为考生和测试设计人员提出建设性的建议。

### 1.2. 研究目的和研究问题

本研究旨在探讨考研英语一与考研英语二中阅读文本复杂度的异同, 通过对二者文本的对比分析, 找出哪一项阅读测试难度更高。借助 Coh-Metrix、VocanProfile 和 SPSS 三种计算工具, 对文本复杂度进行了精确的定量分析。简而言之, 本研究旨在回答以下几个问题:

- 1) 考研英语一阅读测试与考研英语二阅读测试在输入特征上有哪些异同?
- 2) 考研英语一阅读测试和考研英语二阅读测试在文本复杂度上是否存在显著差异?
- 3) 考研英语一阅读测试的文本和考研英语二阅读测试的文本哪一个复杂度更高?

### 1.3. 研究意义

首先,在理论层面上,本研究以 Bachman 和 Palmer 提出的任务特征框架为理论基础并根据作者研究内容做出了相应的调整。原有的框架对分析测试的特征是有帮助的并且全面的,但在实践中,进行比较分析并不灵活。因此,为了更好地服务于本研究,作者在原有框架的基础上进行了一些修改。本研究将重点探讨语篇输入特征,希望对未来阅读的内容效度和文本复杂度的相关研究有所启发。此外,本文还采用了 Coh-Metrix、SPSS、Web Vocab Profile 等计算工具对两个阅读的文本难度进行了比较分析,使研究结果更具说服力,为以后的阅读文本定量分析提供了参考。

其次,在实践层面上,比较分析考研英语一和考研英语二阅读部分的内容效度和文本复杂度,对考生、考试设计人员和教师都有帮助。一方面,本文详细介绍了两种测试的特点,并提供了如何有效学习的可行建议,使应试者更好地了解测试,灵活调整自己的学习方法,提高学习效率。另一方面,本研究为如何进一步提高两项测试的内容效度提供了一些参考和建议。同时,从事考研英语阅读部分教学的教师也可以解决学生在阅读理解方面的困难,这对今后的语言教学是有益的。

第三,在实证研究层面上,本研究填补了考研英语一和考研英语二比较研究的空白。虽然有一种普遍的观点认为,考研英语一比考研英语二更难,但这种主观假设缺乏经验证据和定量分析。本研究旨在验证这一假设的真假,并引起人们对两种测试的比较分析的关注,并鼓励未来从其他角度对考研英语一和考研英语二进行更多的对比研究,如考研英语一和考研英语二写作测试的对比研究。

## 2. 文献综述

### 2.1. NETEM I 和 NETEM II 阅读测试概述

考研英语是一项全国性的大型英语入学考试。它会对每个考生以及整个教育系统产生巨大的影响。自 2010 年以来,考研英语被分为两种类型:考研英语一和考研英语二。考试时间均为三个小时,涵盖英语运用、阅读理解和写作三个部分。然而,考研英语一和考研英语二各部分的分值比例具有显著差异。考研成绩报告以百分制计算分数,其中阅读部分在考研英语的所有题型中的分值排名第一。

在考研英语一中,阅读测试包含了三个部分:阅读 A 部分、阅读 B 部分和阅读 C 部分,分别占比 40%、10%和 10%。阅读 A 部分由四篇文章和二十道选择题组成,每篇文章后面有五个问题。根据考试大纲,四篇文章的文本长度应该在 1600 字左右。阅读 B 部分包含一篇文章,文本长度大约在 500 到 600 字,有三种备选题型,每年只出现一种题型。阅读 C 部分,从一篇英语文章中有五个划线的长难句,学生需要将其从英文译成中文[1]。

考研英语二阅读测试也包含了阅读 A 部分、阅读 B 部分和阅读 C 部分,分别占比 40%、10%和 15%。阅读 A 部分由四篇文章和二十道选择题组成,每篇文章后面有五个问题。根据考试大纲,四篇文章的文本长度应该在 1500 字左右。阅读 B 部分包含一篇文章,文本长度大约在 450 到 550 字,包含两种备选题型,每年只出现一种题型。阅读 C 部分要求学生阅读和理解一篇约 150 字的英文文章,并将其翻译成中文[2]。

### 2.2. 文本复杂度的定义

文本复杂度指的是文本在语言特征上呈现出的复杂程度。内容效度指的是测试是否或者在多大程度上考了考试大纲规定要考的。在实际测试工作中,内容效度的保证和阅读文本的复杂度密不可分。文本复杂度分析是考试内容效度验证的一个重要环节,而构建文本复杂度的分析指标是关键所在。分析文本的复杂度有助于考试内容效度的验证,测试任务的设计以及阅读教学的有序进行。20 世纪 70 年代,文

本的表面特征是学术界关注的焦点,例如平均句长、平均词长和易读度等。其中 Flesch-Kincaid 可读公式(Flesch-Kincaid Grade Level)得到了广泛的应用。进入 21 世纪,由于计算机技术的广泛应用,文本复杂度可以通过一些工具量化文本特征,从而使文本复杂度的分析更为高效、直观。大体上来说,文本复杂度可以从三个维度来进行分析。

1) 文本复杂度的定性维度(Qualitative):读者通过专心的阅读后才测量出的文本难度,比如文本的意义、写作目的、语言结构、语言特点、所要求的背景知识等。

2) 文本复杂度的定量维度(Quantitative):在计算机软件的辅助之下量化文本特征,如词汇长度、词频、句子长度、文本连贯性等。

3) 读者与任务难度:前两个因素主要与文本本身难度相关,但在判断某个文本是否适合读者时,还应考虑到读者的特点(如动机、背景知识和经验等)以及阅读任务的特点(如阅读目的、任务的复杂度及问题难度等)。

### 2.3. 影响阅读复杂度的因素

阅读的过程是阅读文本和读者之间不断互动的一个过程。在这期间,读者的语言知识和背景知识都在和阅读文本中所提供的信息交互影响。在这一过程中,存在着读者和文本两大因素变量影响着阅读的复杂度。

#### 2.3.1. 读者因素

许多研究已经证明,读者是在主动地与阅读文本进行互动,而不是被动地接受信息。读者是影响阅读过程和阅读理解的关键因素之一。Bartlett (1932)指出,读者所掌握的知识必然会对文本内容的记忆程度和处理文本的方式产生影响。Alderson (2000)提出读者的知识可以进一步分为两类:背景知识和语言知识;前者涉及整个世界、文化和相关话题的知识,后者涉及词汇、语法、文本结构和修辞手法等语言点的知识。这两种不同的知识可以相互补充[3]。一方面,文本变量可以激发读者潜在的认知活动,促进或阻碍阅读理解。另一方面,如果读者熟悉某一主题,阅读过程就会变得更快,弥补语言能力的不足。在本文中,作者关注的是文本本身的内在特征,所以关于读者的因素就不做详细的讨论了。

#### 2.3.2. 文本因素

##### 1) 词汇变量

当被问及阅读理解中最大的障碍是什么时,大多数学生可能会将其归咎于阅读过程中遇到的不认识的单词。许多研究都证明了词汇难度与文本复杂度之间的密切关系。Freedle 和 Kostin (1993)探讨了单词长度对阅读理解的影响,并证明了如果文本中包含更多的多音节单词,文本难度会增加[4] [5] [6]。Milton (2009)声称,当文本中出现更频繁的单词时,文本的可读性会增加。Laufer 和 Sim (1985)证明了对文本的成功理解在很大程度上取决于词汇的广度和深度,词汇的影响超过了句法结构等其他因素的影响[7]。Ure (1971)提出词汇密度指的是文本中实词的比例。实词越多,说明文章所包含的信息越多,文章难度也就越大。因此,词汇量是文本难度的有效预测因素[8]。

##### 2) 句法变量

除了词汇变量外,影响阅读理解的另一个因素是句法变量。Berman (1984)进行了一项研究,以证明句法复杂性是否会对以色列的英语大学生产生影响。他发现句子结构的复杂度会增加读者分析句子或识别句子所包含的成分的难度。Graesser, McNamara 和 Kulikowich (2011)也指出,长句子往往需要更多的工作记忆,从而使阅读更困难。根据前人的研究,可以得出结论,句法变量是文本难度的一个有效指标[9]。

### 3) 语篇变量

语篇变量涉及语篇长度、语篇体裁、语篇主题和衔接手段等多个方面。对于文本长度, Terzi (1984) 认为不同的文本长度可以触发不同的阅读技巧, 较长的文本往往增加语言特征的敏感性。Nuttall (2002) 提出较长的文本更适合读者采用略读和扫读的阅读策略, 以发现文本的主旨。文本体裁对文本难度也有影响[10]。在文本主题方面, Alderson 和 Urquhart (1985) 探讨了读者的背景知识对阅读理解的影响。结果表明, 读者的阅读表现与他们的背景知识密切相关[11]。Hale (1988) 认为读者所从事的学科能够促进阅读理解。他发现, 与读者的学科相关的文本对他们来说更容易处理或解码。因此, 考试设计人员应该精心选择阅读材料, 以尽可能消除主题偏差。在衔接和连贯方面, 使用一些衔接手段可以降低语篇的难度。Britton 和 Gulgoz (1991) 发现连贯有助于阅读理解。他们调整句子的顺序, 使文本遵循从旧到新的模式, 这意味着文本中的隐性引用对读者来说是显性的。因此, 这种明确的文本连贯使得阅读过程比以前容易得多[12]。Horiba (2004) 通过比较两种说明性文本(一种是高度连贯的, 另一种是较不连贯的), 探讨了语篇连贯对记忆能力的不同影响。研究发现, 第一篇文本的连贯有利于阅读理解, 使读者对文本内容建立更加连贯的心理表征。

### 4) 易读度

为了使阅读文本的难度水平适合于目标读者, 必须通过科学客观的方法来衡量可读性。预测文本难度的可读性公式多种多样。例如, Brown (1998) 通过在学术文章中使用完形填空程序验证了传统的可读性公式是否适用于二语学习者, 结果表明传统的可读性公式不能准确预测二语阅读难度。然而, 与 Brown (1998) 不同的是, Greenfield (1999) 在进行相关研究后得出了相反的结论。他要求 200 名日本大学生完成 32 篇学术文章的完形填空。结果表明, 传统可读性公式预测的分数与计算出的完形填空平均分数之间的 Pearson 相关性为 0.85。Greenfield (1999) 认为 Flesch Reading Ease 是预测文本难度的可靠公式。与 Greenfield (1999) 类似, Harrison (1979) 也倾向于传统的可读性公式。同时, 他们也承认, 这些公式只关注单词和句子长度等表面因素, 而忽略了其他重要因素, 如衔接和连贯。因此, Carrell (1987) 认为传统公式在预测 L2 文本难度时并不那么准确。总之, 尽管传统的可读性公式存在一定的局限性, 但它们确实为文本难度提供了依据。本文采用传统的 Flesch Reading Ease 公式和计算工具 Coh-Metrix 提供的针对于二语学习者的公式来探讨考研英语一和考研英语二阅读文本的可读性。

## 2.4. 英语考试的比较研究

### 2.4.1. 国际英语考试之间的对比研究

关于国际英语考试的比较研究很多。一个著名的比较研究是关于 Bachman (1988) 进行的两个英语考试: 托福(TOEFL)和剑桥英语第一证书(FCE) [13]。它被称为剑桥-托福的可比性研究, 其目的是检验两个 EFL 测试组的可比性。通过对两个测试任务特征的详细比较, 可以发现两个测试具有相似的语言能力。

### 2.4.2. 国内英语考试之间的对比研究

一些比较研究集中在国内英语考试。为了明确高考英语(NMET)和大学英语四级(CET-4)之间是否存在难度差异, 李冀、蔡基刚(2012)从听力、阅读、完形填空、写作和翻译等方面进行了比较分析, 最终发现这两个考试在很多方面都缺乏合适的难度跨度。指出难度差距的缺乏是由两个基本原因造成的。一个原因是高考的难度在增加, 而英语四级的难度相对稳定。二是受大学英语四级通过率低的制约, 大学英语四级考试的教学要求过低[14]。陈卓、于鹏(2012)该研究针的研究对象是某高校 148 名在校大学生, 该研究对比了高考英语成绩和四六级的成绩, 并与用 SPSS 进行数据分析。研究结果显示, 四级分数在 425 分以上的大学学生, 他们的高考英语成绩和四级英语成绩呈现正相关。六级分数在 425 以上的学生, 他们的高考英语成绩和四级英语成绩没有明显的相关性[15]。

李丽、高淑玲和刘园园(2015)根据实验数据对大学英语四级考试中两种完形填空的难度进行了比较分析。实验结果表明, 当所选文本和删除词的位置完全相同时, 两种完形填空的难度存在显著差异[16]。

### 2.4.3. 国际英语考试和国内英语考试之间的对比研究

还有一些比较研究旨在探讨国际英语考试与国内英语考试之间的差异。Yang 和 Weir (1988)对雅思和大学英语六级进行了比较研究, 以探索哪种考试更难。在他们的研究中, 426 名学生被要求参加这两项测试, 他们的分数被收集和比较。结果表明, 两种测试在难度水平和效度上存在显著差异[17]。杨旭和裴学敏(2018)从词汇难度、句法复杂性、衔接连贯性和文本可读性四个方面对考研英语一和托福阅读的文本复杂度进行了比较研究。他们的研究基于 Coh-Metrix 3.0, 以 Bachman 和 Palmer 的任务特征框架为理论基础, 利用 SPSS 统计软件从四个方面进行独立样本 t 检验。根据数据分析的结果, 两种测试在词长、句法、深层衔接和相邻段落之间潜在的语义重叠四个方面存在显著差异, 而在其余部分没有显著差异[18]。江进林和韩宝成(2018)使用 Coh-Metrix 对 CET-6、托福和雅思阅读文本的难度进行了分析。他们分析了 163 篇阅读文本的词汇、语法和语义特征。结果表明, 托福和雅思文本的阅读难度高于 CET-6 文本, 且在叙事性、连接性、词语具体性和指称衔接等四个维度上存在显著差异[19]。虞程远(2020)比较了 CET-6、雅思和托福的阅读文本中的词汇复杂度。作者利用在线文本分析工具 Text Inspector, 对于三项阅读中的词汇复杂度(词汇难度和词汇多样性)进行了对于分析。研究结果表明, 词汇难度最高的是托福词汇, 雅思排名第二, 六级排在最后一名。其次, 三者之间的词汇难度和词汇多样性都存在着显著的差异[20]。

从上面可以看出, 关于语言测试的比较研究有很多。研究者从测试难度、文本特征、测试效度等方面对不同测试进行比较。这些研究有助于更好地理解语言测试, 也为改进语言测试提供了有用的启示。然而, 关于考研英语一和考研英语二的比较研究却很罕见。虽然人们普遍认为, 由于考研英语二的历史相对较短, 它比考研英语一更容易, 但没有令人信服的证据表明这两个测试在难度上的差异。考虑到这两个考试的重要性, 考研英语一和考研英语二的比较分析值得更多的学术关注。阅读是最基本的语言能力之一, 而阅读测试是两个测试的重要组成部分, 因此本研究将比较考研英语一和考研英语二阅读的文本复杂度。

## 3. 理论支撑

**Table 1.** A revised framework for the description of reading tests

**表 1.** 阅读测试描述的修订框架

	语篇输入特征
词汇特征	1) 词长 2) 词汇多样性 3) 词的密度 4) 词频
句法特征	1) 句长 2) 句法复杂度
语篇特征	1) 文章长度 2) 文章体裁 3) 文章话题 4) 连贯与衔接
易读度	1) 弗莱士易读度 2) Coh-Metrix L2 易读度

Bachman (1990)提出, 语言测试的一个问题是, 没有一个定义可以清晰明确地定义语言使用任务, 因此, 无论是说明内容相关性还是内容覆盖度都不容易[21]。因此, Bachman 和 Palmer (1996)提出了一个语言任务框架来评估语言能力。该框架由五个方面的任务组成: 环境特征、测试规则特征、语篇输入特征、预期回答特征、语篇输入与预期回答之间的关系[1]。该框架几乎全面地陈述了各种语言测试中经常采用或使用的测试方法的所有特征。虽然 Bachman 的任务特征框架是有效的, 可以为测试描述提供很好的参考, 但它过笼统。因此, 根据本研究的研究目的和研究内容, Bachman 的原始框架进行了修改, 修订框

架见表 1。由于本研究的重点是文本复杂度差异分析, 而不是考研英语阅读测试的整体特征, 因此在修改后的框架中将主要涉及文本输入特征。输入特征是指考试中给出的材料, 考生根据这些材料做出一定的反应。这种材料是根据格式和语言特征来描述的。与阅读测试任务一样, 输入是以语言的形式呈现的, 因此修改后的框架将关注输入的语言, 包括词汇特征、句法特征、语篇特征和可读性。

通过上面的分析, 我们可以看到, 在本研究中并不是对原有框架的所有方面都进行了讨论, 在修改后的框架中加入了一些新的内容。

## 4. 研究设计

### 4.1. 研究材料

选取 2018 至 2022 年考研英语一和考研英语二阅读理解 A 部分的 5 年真题进行对比分析。总共共选取了 40 篇阅读文本进行比较, 其中考研英语一和考研英语二阅读文本各 20 篇。40 篇阅读文本的复杂度将在修订的理论框架的指导下详细分析。

### 4.2. 研究工具

近年来, 随着计算语言学的发展, 一些自动化工具促进了文本特征的分析。本研究采用 Coh-Metrix 3.0、Web VocabProfile 和 SPSS 26.0 进行比较分析两种不同阅读文本的复杂度。

Coh-Metrix 3.0 是一个计算工具, 包含 108 个关于文本语言特征的复杂索引集。Coh-Metrix 实现了从文本中自动提取信息的目标。提取的指标可分为 11 类: 1) 描述性指标、2) 指称衔接指标、3) 潜在语义分析(LSA)指标、4) 词汇多样性指标、5) 连接词指标、6) 情景模型指标、7) 文本易读性主成分得分指标、8) 句法复杂性指标、9) 句法模式密度指标、10) 词汇信息指标、11) 可读性指标。为了分析考研英语一和考研英语二中阅读测试的语篇特征, 采用 Coh-Metrix 进行语篇特征分析, 因为该工具能够提供语篇特征的表层和深度数据。它量化了语言特征, 并在词汇、句法、语篇等方面提供了具体证据。

Web Vocab Profile 是一个在线计算工具, 用于分析文本的词汇特征。它根据频率将词汇分为四类: 最频繁的 1000 个和第二频繁的 1000 个英语单词, 英语学术单词(AWL), 以及其他未在三个列表中找到的单词(Off-list)。换句话说, Web Vocab Profile 计算文本中使用的高频和低频词的百分比。

SPSS 26.0 是最流行的统计分析软件之一。本研究采用独立样本 T 检验, 检验考研英语一和考研英语二阅读文本的特征是否存在显著差异。

### 4.3. 研究过程

首先, 作者收集考研英语一和考研英语二阅读 A 部分的 40 篇课文。其次, 基于修改后的理论框架, 选取了 Coh-Metrix 的相关指标, 对两种阅读文本的复杂度进行了比较分析。再次, 将所有从 Coh-Metrix 收集的数据提交到 SPSS 中, 进行独立样本 t 检验, 以确定是否存在显著差异。最后, 作者整理分析所有从 SPSS 中收集到的数据, 最终得出结论。

## 5. 发现与讨论

### 5.1. 词汇特征

在词汇特征方面, 作者从词长、词汇多样性、词的密度和词频四个方面进行了论述。首先, 在音节和字母的单词长度方面, 考研英语一的单词比考研英语二的单词长, 并且在单词的字母数上有显著差异( $P = 0.013 < 0.05$ )而在音节长度上无显著差异( $P = 0.052 > 0.05$ )。其次, 在词多样性方面, 考研英语一的类形符比高于考研英语二, 但差异不显著( $P = 0.0687 > 0.05$ )。第三, 在词密度方面, 考研英语一阅读理

解的实义词占比略高于考研英语二,但在两次阅读测试中均未发现显著差异( $P = 0.635 > 0.05$ )。第四,从词频的角度上看,考研英语一和考研英语二阅读文本的词频没有显著差异( $P = 0.085 > 0.05$ )。词汇特征独立样本 T 检验结果见表 2。

**Table 2.** The result of independent-samples T test for lexical characteristics

**表 2.** 词汇特征独立样本 T 检验结果

T-检验	词长		词汇多样性	词的密度	词频
	音节	字母			
P 值	$P > 0.052$	$P < 0.013$	$P > 0.0687$	$P > 0.635$	$P > 0.085$

## 5.2. 句法特征

就句法特征而言,本文主要包括句子长度和句法复杂度两个方面。首先,考研英语一阅读文本的平均句子长度比考研英语二阅读文本略长。根据独立样本 t 检验的结果( $P = 0.400 > 0.05$ ),考研英语一和考研英语二的平均句子长度没有显著差异。其次,两组阅读测试的句子复杂性基本保持在同一水平,但考研英语一阅读文本的句子简单性略高于考研英语二。从表 3 可以看出,考研英语一阅读文本的句子简单性(50.6625%)高于考研英语二阅读文本的句子简单性(45.7735%),说明考研英语一的句法结构更容易理解。但是,通过 SPSS 软件对句法简单性进行综合比较分析,t 检验结果显示,考研英语一和考研英语二在阅读句法复杂度方面没有显著差异( $P = 0.338 > 0.05$ )。句法特征独立样本 T 检验结果见表 3。

**Table 3.** The result of independent-samples T test for syntactic characteristics

**表 3.** 句法特征独立样本 T 检验结果

考试类型	PCSYNz	PCSYNp
NETEM I (均值)	0.01460	50.6625%
NETEM II (均值)	-0.11645	45.7735%
P 值	0.400	0.388

## 5.3. 语篇特征

**Table 4.** The result of independent-samples T test for textual characteristics (text length)

**表 4.** 语篇特征(篇章长度)独立样本 T 检验结果

考试类型	平均篇章长度
NETEM I (均值)	1699.2
NETEM II (均值)	1623.2
P 值	0.002

从语篇特征来看,可分为语篇长度、语篇体裁、语篇话题和语篇衔接四大类。首先,两种不同的阅读测试在文本长度上存在显著差异( $P = 0.002 < 0.05$ ),英语一阅读文本的长度高于英语二阅读文本长度,因此考研英语一阅读文本的时间更长。其次,考研英语一和考研英语二体裁分布并不均衡,两者都以议论文为主,很少涉及记叙文和说明文。第三,两种测试都重视社会科学的话题,而其他的话题所占比例较小。第四,作者从语义重叠、指称衔接和连接词三个角度探讨了两篇阅读文本的衔接与连贯。在语义重叠方面,考研英语一做得更好,与考研英语二有着显著的区别( $P = 0.002 < 0.05$ )。在指称衔接方面,考

研英语一和考研英语二的指称衔接没有显著差异( $P > 0.05$ ), 但考研英语一的指称衔接水平略高于考研英语二。我们还可以发现, 考研英语一中连接词的发生率高于考研英语二, 但在连接词的六个指标中都没有发现显著差异( $P > 0.05$ ), 这表明考研英语一和考研英语二阅读文本的连接词的发生频率接近。语篇特征独立样本 T 检验结果见表 4、表 5。

**Table 5.** The result of independent-samples T test for textual characteristics (cohesion and coherence)

**表 5.** 语篇特征(连贯与衔接)独立样本 T 检验结果

T-检验	语义重叠	指称衔接	连接词
P 值	$P < 0.05$	$P > 0.05$	$P > 0.05$

#### 5.4. 易读度

在上述三个方面的基础上, 本章还对考研英语一和考研英语二中阅读的可读性进行了比较研究。作者讨论了从 Coh-Metrix 中提取的两个指标, 即 Flesch Reading Ease 和 Coh-Metrix L2 可读性。前者是一个传统的公式, 可以清晰地呈现文本难度的大小。结果显示, 考研英语一和考研英语二阅读文本的可读性并不完全相同, 独立样本 t 检验结果( $P < 0.05$ )证明考研英语一的难度水平显著高于考研英语二。然而, 由于缺乏影响可读性的一些基本因素, 传统的公式并不足以探索阅读文本对于二语读者的可读性。因此, 本文还采用 Coh-Metrix L2 阅读指数, 目的是更好地了解考研英语一和考研英语二的可读性差异。从 Coh-Metrix L2 阅读指数来看, 两种不同的阅读测试在可读性上具有相似性, 并不具有显著差异( $P > 0.05$ )。易读度独立样本 T 检验结果见表 6。

**Table 6.** The result of independent-samples T test for readability

**表 6.** 易读度独立样本 T 检验结果

T-检验	Flesch Reading Ease	Coh-Metrix L2 Reading Index
P 值	$P < 0.05$	$P > 0.05$

## 6. 结论

通过以上分析, 可以得出本研究的结论: 在词汇特征方面, 考研英语一中的单词长度比考研英语二中的单词长, 并且在单词的平均字母数上存在显著差异。在词汇多样性、词的密度和词频三个方面不存在显著差异。在句法特征方面, 考研英语一和考研英语二在句长和句子复杂度上没有显著的差异。在语篇特征方面, 二者皆存在体裁和话题分配不均衡的问题。考研英语一的篇章长度高于考研英语二的篇章长度, 且存在显著差异, 在连贯与衔接上, 除了语义重叠之外, 其他部分并没有显著的差异。在可读性方面, 根据传统的计算公式, 二者存在显著差异, 且英语一的易读度要低于英语二。但在针对于二语学习者的公式下, 尽管英语一易读度依然低于英语二, 但是二者不存在显著差异。

本研究在探索考研英语一和托福阅读文本复杂度特征的基础上, 对两者进行了深入的对比研究, 探究两者之间的差异性。为考试试题设计者, 教师提供参考, 从而使我国研究生入学考试的阅读试题的质量不断优化, 内容效率不断提高。

## 参考文献

- [1] 教育部考试中心. 全国硕士研究生招生考试英语(一)考试大纲(非英语专业) [Z]. 北京: 高等教育出版社, 2022.
- [2] 教育部考试中心. 全国硕士研究生招生考试英语(二)考试大纲(非英语专业) [Z]. 北京: 高等教育出版社, 2022.
- [3] Alderson, J.C. (2000) *Assessing Reading*. Cambridge University Press, Cambridge.

- <https://doi.org/10.1017/CBO9780511732935>
- [4] Freedle, R. and Kostin, I. (1991) The Prediction of SAT Reading Comprehension Item Difficulty for Expository Prose Texts. Educational Testing Service, Princeton. <https://doi.org/10.1002/j.2333-8504.1991.tb01396.x>
- [5] Freedle, R. and Kostin, I. (1992) The Prediction of GRE Reading Comprehension Item Difficulty for Expository Prose Texts for Each of Three Item Types: Main Ideas, Inferences, and Explicit Statements. Educational Testing Service, Princeton. <https://doi.org/10.1002/j.2333-8504.1991.tb01426.x>
- [6] Freedle, R. and Kostin, I. (1993) The Prediction of TOEFL Reading Comprehension Item Difficulty for Expository Prose Passages for Three Item Types: Main Idea, Inference, and Supporting Idea Items. TOEFL Research Reports No. RR-93-44. Educational Testing Service, Princeton. <https://doi.org/10.1002/j.2333-8504.1993.tb01524.x>
- [7] Laufer, B. and Sim, D. (1985) Measuring and Explaining the Reading Threshold Needed for English for Academic Purposes Texts. *Foreign Language Annals*, **18**, 405-411. <https://doi.org/10.1111/j.1944-9720.1985.tb00973.x>
- [8] Ure, J. (1971) Lexical Density and Register Differentiation. In: Perren, G.E. and Trim, J.L.M., Eds., *Applications of Linguistics: Selected Papers of the Second International Congress of Applied Linguistics*, Cambridge University Press, Cambridge, 443-452.
- [9] Graesser, A., McNamara, D. and Kulikowich, J. (2011) Computational Analyses of Multilevel Discourse Comprehension. *Topics in Cognitive Science*, **3**, 371-398. <https://doi.org/10.1111/j.1756-8765.2010.01081.x>
- [10] Nuttall, C. (2002) *Teaching Reading Skills in a Foreign Language*. Shanghai Foreign Languages Educational Press, Shanghai.
- [11] Alderson, J.C. and Urquhart, A.H. (1985) The Effect of Students' Academic Discipline on Their Performance on ESP Reading Tests. *Language Testing*, **2**, 192-204. <https://doi.org/10.1177/026553228500200207>
- [12] Britton, B.K. and Gulgoz, S. (1991) Using Kintsch's Computational Model to Improve Instructional Text: Effects of Repairing Inference Calls on Recall and Cognitive Structures. *Journal of Educational Psychology*, **83**, 329-345. <https://doi.org/10.1037/0022-0663.83.3.329>
- [13] Bachman, L.F., Kunnan, A., Vanniarajan, S. and Lynch, B. (1988) Task and Ability Analysis as a Basis for Examining Content and Construct Comparability in Two EFL Proficiency Test Batteries. *Language Testing*, **5**, 128-159. <https://doi.org/10.1177/026553228800500203>
- [14] 李莫, 蔡基刚. 高考英语考试与大学英语四级考试的对比研究[J]. 山东外语教学, 2012, 33(2): 55-61.
- [15] 陈卓, 于鹏. 高考英语成绩与大学英语四、六级成绩的实证研究[J]. 海外英语, 2019(6): 113-115.
- [16] 李丽, 高淑玲, 刘园园. 大学英语四级测试中两种完形填空难度和区分度对比分析[J]. 新西部(中旬刊), 2015(1): 41-40.
- [17] Weir, C.J., Hawkey, R., Green, A., *et al.* (2009) The Relationship between the Academic Reading Construct as Measured by IELTS and the Reading Experiences of Students in the First Year of Their Courses at a British University. IELTS Research Reports 9, British Council, 97-156.
- [18] 杨旭, 裴学梅. 研究生入学考试英语与托福阅读文本复杂度特征对比研究[J]. 教育现代化, 2018, 5(2): 204-206.
- [19] 江进林, 韩宝成. 基于 Coh-Metrix 的大学英语英语六级与托福、雅思阅读语篇难度研究[J]. 中国外语, 2018(3): 86-95.
- [20] 虞程远. 六级、雅思、托福阅读词汇复杂度对比研究[J]. 外语与翻译, 2020, 27(3): 8-13.
- [21] Bachman, L.F. (1990) *Fundamental Considerations in Language Testing*. Oxford University Press, Oxford.