

目 录

表格目录	vii
图示目录	ix
缩略语	xi
第一章 概论.....	1
1.1 研究缘起.....	1
1.1.1 教材研究的重要性	1
1.1.2 教材文本复杂度研究的重要性	3
1.1.3 文本复杂度自动分级研究的重要性.....	4
1.2 研究意义.....	5
1.3 本书结构.....	7
第二章 文本复杂度概念界定及测量工具.....	9
2.1 语言复杂度.....	9
2.2 文本复杂度与文本难度.....	10
2.3 绝对复杂度和相对复杂度	13
2.4 研究对象.....	15
2.5 文本复杂度维度构成及测量工具	17
2.5.1 词汇复杂度及主要测量工具.....	18
2.5.2 句法复杂度及主要测量工具.....	21
2.5.3 语篇复杂度及主要测量工具.....	25
2.6 小结.....	27
第三章 文本复杂度实证研究综述.....	28
3.1 基于语料库的文本复杂度研究.....	28

3.1.1	教材文本复杂度研究	29
3.1.2	阅读测试和二语写作用文本复杂度研究	35
3.2	文本复杂度自动分级研究	41
3.2.1	文本复杂度自动分级主要方法	41
3.2.2	神经网络在文本复杂度研究中的应用	48
3.3	小结	52
第四章	研究方法	53
4.1	研究问题	53
4.2	研究语料	53
4.3	研究步骤	56
4.4	研究工具和方法	57
4.4.1	文本复杂度特征提取	57
4.4.2	多元统计方法	74
4.4.3	机器学习及模型验证方法	78
4.5	小结	81
第五章	文本复杂度各维度特征的确定	82
5.1	参照语料库的选取	82
5.2	词汇复杂度指标——TAALES 主成分	83
5.2.1	词汇频率和分布	86
5.2.2	学术书面语特征	90
5.2.3	三元词互信息	92
5.2.4	二元词频率和分布	93
5.2.5	动词特异性	93
5.3	句法复杂度指标——TAASSC 主成分	94
5.3.1	名词短语扩展	96
5.3.2	所有格	98
5.3.3	VAC 频率和嵌入式小句	99
5.3.4	小句从属情况	100
5.3.5	VAC 中构式对主动词的关联强度	101

5.3.6	VAC 中主动词对构式的关联强度和构式频率	103
5.3.7	名词主语并列	104
5.3.8	限定词	104
5.4	语篇复杂度指标——TAACO 主成分	105
5.4.1	相邻两句词汇重叠	107
5.4.2	相邻三句和段落间重叠	110
5.4.3	深度衔接	111
5.4.4	并列附加连接	112
5.5	小结	112
第六章	训练语料分级合理性的验证	113
6.1	训练语料的选取依据	113
6.1.1	使用时间长, 业界口碑好	114
6.1.2	由浅入深, 循序渐进	115
6.2	训练语料文本复杂度单元聚类	119
6.3	《新概念英语》文本复杂度册级对比	122
6.3.1	册级间词汇复杂度对比	123
6.3.2	册级间句法复杂度对比	129
6.3.3	册级间语篇复杂度对比	136
6.4	小结	141
第七章	文本复杂度分级模型的构建和评估	142
7.1	模型构建及其评估方法	142
7.1.1	建模方法	142
7.1.2	模型评估方法和指标	144
7.2	不同方法建模性能比较	145
7.3	各类特征建模比较	150
7.3.1	传统可读性公式	150
7.3.2	本研究特征集	153
7.4	模型对不同数据集的泛化能力	160

7.4.1	模型在其他教材语料库上的表现	160
7.4.2	模型在非教材语料上的增值应用	163
7.5	小结	167
第八章	结语	168
8.1	主要发现	168
8.2	主要贡献及启示	173
8.3	局限与展望	175
参考文献	177
附录一	基于语料库的英语教材文本复杂度研究一览表	202
附录二	基于 BNC baby 提取的主成分特征	206

第一章 概论

本章从教材研究、教材文本复杂度研究和文本复杂度自动分级研究三个方面的重要性阐述本研究选题的缘起，在此基础上说明本研究的意义，并介绍本书的整体结构框架。

1.1 研究缘起

1.1.1 教材研究的重要性

教材建设是国家事权，对此目前教材研究者已达成普遍共识。习近平总书记在《培养德智体美劳全面发展的社会主义建设者和接班人》中强调，“要注重教材建设。教材是传播知识的主要载体，体现着一个国家、一个民族的价值观念体系，是老师教学、学生学习的重要工具”。教学改革，改到深处是课程，改到痛处是教师，改到实处是教材（吴岩 2019）。教材是人才培养的主要剧本，抓好教材质量是抓好人才培养四项“新基建”之一（吴岩 2021）。离开高质量的教材，即使有好的专业、好的学科、好的教师，也难以培养出高水平人才。特别是随着全球经济、文化、教育发展和改革的纵深推进，以及教育现代化、信息化建设进程的加速，从国家战略、未来发展和国际竞争的高度重新审视和加强教材建设与发展，提高教材质量，突出教材特色，培养创新人才，已经成为包括我国在内的许多国家的共识（胡军 2019：138）。

教学过程的构成要素大体上可以分为教师与学生、教学目标、教学内容（主要指教材）、教学方法、教学组织形式与教学效果等。其中，教学目标通过教材体现和落实，教师根据教材来设计教学方法、组织教学

活动。由此可见，教材作为教学活动的线索，将教学过程中的各个要素紧密串联起来（王郢 2017：19）。教材是任何课程的重要载体，是任何教育体系的支柱（Tomlinson 2012；Guerrttaz and Johnston 2014；赵勇、郑树棠 2006；钟启泉等 2008）。Sheldon（1988）将教材比喻为“外语学习中看得见的心脏”。Hutchinson and Torres（1994）同样认为，教材是外语教学中最基本的要素，是教师组织教学活动的主要依据和学生学习的中心内容。教材是实际意义上的教学大纲，为课堂教学提供最直接的指导和框架，同时也为教学过程和效果的评价提供了依据（Garton and Graves 2014a）。尽管当今信息化时代外语网络资源铺天盖地，但由于我国外语教学缺乏先天的自然语言环境，教材依然是学生输入的主要渠道。正如《大学英语教学指南》（2020）中所提出的，教材是传播新知识、新思想、新观念的重要载体，是教学内容的主要载体，也是实现教学目标的基本保证。教材体现教育理念，制约课程内容，反映教学模式，影响教学效果，联系学习方式，是开展教育、教学和学习的重要条件之一。

外语教材研究旨在析解外语教学材料的规律和特点、考察教材编纂的原则与技巧等（王文斌、李民 2017）。开展外语教材研究是时代所需、国家所需、发展所需。作为外语教育的重要组成部分，教材既关系到立德树人的目标、人才培养的方向，也是推动全国外语教育教学改革、课程创新和教师发展的关键抓手。根据研究对象不同，教材研究可以分为教材内容研究、教材使用研究和教材开发研究（Harwood 2014），其中，教材内容研究是教材研究的主线。这类研究主要有两条路径，一是聚焦特定教材在话题、语言、文化等方面的具体内容展开研究，二是围绕特定教材的内容（文本、语言）开展评价研究。例如，教材中语言素材是否具有真实性，内容的版面设计如何，教材内容的选择和安排是否合理等（如 Breen and Candlin 1987；Cunningsworth 1984；Grant 1987；Hutchinson and Waters 1987；Sheldon 1988）。

外语教材是外语教育学学科体系结构的要素之一，外语教材研究是外语学科教学的重要构成部分，对学科建设意义重大（李民、王文斌 2019）。然而，国内外对外语教材研究却未给予足够重视（Ghorbani 2011；Tarone 2014），关于外语教材编写方面的高质量学术论文尚不多

见,理论专著更是寥若晨星(庄智象 2006: 51)。令人欣慰的是,近年来与外语教材研究相关的专著、论文集以及期刊文章相继出现,逐步成为新兴的研究热点(如 Garton and Graves 2014b; Harwood 2014; Tomlinson 2016; Shawer 2017),甚至有研究指出,外语教育教材研究的时代正在到来(Rixon and Smith 2012)。但总体而言,当下对教材研究的深度和广度都十分欠缺,主观论述多于实证论述,学界对教材的科学研究亟待进一步深入(赵永青等 2014: 32; 文秋芳 2017: 17; 张虹等 2021: 66)。

1.1.2 教材文本复杂度研究的重要性

要对教材进行评价,首先得有相应的教材评价标准。这是实施教材评价的主要依据,旨在说明我们应该从哪些方面去评价教材,不同的标准设计反映教材评价的价值取向。参照国外教材在评价指标拟定中明确的设计原则以及具体的评价标准,我国构建科学有效的教材评价指标体系可采取的措施有:重视教材评价指标体系的开发;制定科学的教材评价指标体系;深刻理解指标之间的关系;关注教材评价指标体系的适用性;提供信度和效度指标。教材评价体系建构不仅是为了确证教材水平,还是为了提升教材质量,其根本指向在于落实人才培养目标,促进学生个体健全发展。

在外语教学中,教材中的阅读文本是学生学习知识、发展语言技能的重要来源,对学生理解和语言发展起着至关重要的作用。基于苏联著名心理学家维果茨基(Vygotsky 1978)的“最近发展区”(Zone of Proximal Development, ZPD)理论,教材提供给学生的阅读材料如果复杂度恰当,不仅能够保证阅读和学习的顺利开展,而且能够为提高学习者的英语阅读能力搭建“脚手架”。一方面,阅读文本的复杂度必须适当超越学生当前水平,保证对他们来说具有一定的挑战性,这样的教学才会发生在学生的最近发展区内,促使学生主动且有效地学习,从而实现学习效果的最大化;另一方面,教材提供的阅读文本复杂度又不宜远高于或远低于学习者的水平,否则适得其反,不仅会打击学生学习的积极性,而且会挫伤他们的学习兴趣和热情。

英语教材不仅传授英语知识与技能,同时也体现语言认知过程和英语思维方法,其结构应遵循“循序渐进”的语言学习认知规律,而不是知识的机械排列。教材内容和语言材料的选择要遵循从易到难、由简单到复杂逐步过渡的原则,体现一个由浅入深、由少到多的发展过程(林新事 2008: 55; 束定芳、庄智象 2008: 144)。在对阅读准确率和阅读水平提高的关系考察上, Allington *et al.* (2015: 499) 发现, 阅读准确率达到 95% 左右的文本与读者阅读水平的提高有直接因果关系, 而准确率显著低于或高于 95% 的文本不能对读者水平提高产生积极影响。Seals (2013: 3) 的研究也证明, 等级划分合理的教材体系能够有效提高学生的口语阅读流利度和理解水平。因此, 将教材文本复杂度与目标学习者的语言水平进行适度匹配, 对提高学生水平、提升学习效果至关重要。完善的教材体系应当把握语言学习规律, 遵循文本复杂度递进的原则, 规定学习的梯度, 在学习的不同阶段逐渐提高文本的复杂度, 以便与学生不断发展的语言水平相匹配 (Dascălu 2014: 16)。《英语课程标准》(2017) 在“教材编写建议”中明确提出, 教材编写应体现科学性原则, 要依据语言学习的规律, 充分体现不同年龄段和不同语言水平学生的学习特点和学习需要。

目前, 我国英语教材存在的主要问题之一就是教材课文的编排没有依据课文内容的复杂度进行, 各单元之间的复杂度波动明显, 而且不同册教材的课文在复杂度上差别不明显, 因而没有呈现合理的梯度(胡开宝、李翼 2016: 36)。因此, 要提高我国英语教材的编写水平和质量, 必须加强对教材文本复杂度变化趋势的考察, 就“教材是否遵循循序渐进的原则”和“级差递进是否合理”进行科学的、客观的分析与评价。同时, 研究的相关结论可用来反哺教材编写和改编, 指导外语教学, 提高教学效果。

1.1.3 文本复杂度自动分级研究的重要性

长期以来研究者一直在努力构建各种能够对文本复杂度进行自动计算和估测的公式或模型。文本复杂度自动分级就是通过将目标文本转化为一组复杂度特征, 再利用这些特征来进行公式计算或训练模型, 实现

对新文本复杂度的自动预测或定级。然而，究竟哪些文本特征可以表征文本复杂度，这些特征又如何与对应的复杂度级别相关联，这些一直是研究者致力解决的问题。

早期的文本复杂度自动分级研究主要是采用构建可读性公式的方法，先根据经验设计几种语言因素作为构建可读性公式的变量，再通过线性多项式回归等拟合方法得到公式中的其他参数。但是，鉴于这些公式的参数多为文本的表层特征，如句长、词长（字母数或音节数）、词频等，效度和信度一直被学界诟病（Collins-Thompson and Callan 2004; Feng 2010; Petersen and Ostendorf 2009; Si and Callan 2001）。随着人工智能和自然语言处理（Natural Language Processing, NLP）技术的不断发展与进步，文本复杂度研究也得到了越来越多的关注。研究者不再囿于少数几个文本表层特征，而开始借助语言模型和剖析器来探究影响文本复杂度的深层词汇和句法特征，考察的变量大幅增加。然而，总的来说，变量仍然局限于词汇和句法层面，较少关注衔接与连贯、语义重叠等语篇复杂度特征，而这些特征对阅读文本的复杂度判断至关重要。此外，已有研究大都采用线性拟合的方式，虽然线性模型在训练集上的拟合度较高，但很难推广到其他类型文本的复杂度判断，因此通用性和泛化能力有限。

本研究旨在建立一种基于文本多维度语言特征、具有一定泛化能力的文本复杂度自动分级模型，便于对文本复杂度进行自动评估和定级。构建的文本复杂度自动分级模型，不仅兼具客观性和经济性，而且具有较大的实用价值。分级模型的应用场景也很广泛，在教育教学领域既可以帮助教材编写者或教师为学生选择合适的阅读材料，也可以指导学生根据自己的语言水平筛选适合自己的阅读文本，同时也能为教材编写、阅读课程规划和测试设计提供参考。

1.2 研究意义

我国的英语学习者多达数亿人，人数之多，影响之广，可谓独一无二。因此，构建科学的、系统的文本复杂度分级模型意义重大。本研究将文

本复杂度视作多维度构念，在提取词汇、句法和语篇三个维度特征并进行特征优化的基础上，借助分级合理的训练语料，采用神经网络分类算法，构建具有良好性能和泛化能力的文本复杂度分级模型。简要说来，本研究的意义主要体现在以下三个方面。

首先，在方法论层面，本研究探索了一种适用于文本复杂度模型构建的新方法。该方法通过融合有监督机器学习方法中的多元语言特征与神经网络算法，构建文本复杂度分级模型。具体步骤如下。第一，选取可靠的软件工具，尽可能全面地提取能够表征文本复杂度的语言特征；运用主成分分析法对提取的特征进行降维优化，筛选出少数主成分特征，以最大限度保留原始指标所包含的信息。第二，选取一套公认的、带有复杂度等级标签的语料，在验证其分级合理性后，将其作为模型构建的训练语料。第三，利用神经网络算法进行参数调优，选取最优参数完成文本复杂度分级模型的构建。研究表明，按照上述路径构建的分级模型具备良好的性能与稳健性，且对全新数据具有较强的泛化能力。

其次，在理论层面，本研究加深了对文本复杂度理论内涵的理解。文本复杂度有别于文本难度，本质上是一个多维度、多层次构念。它是文本自身的客观属性，是指影响理解的文本语言特征，这些特征独立于读者和阅读任务而存在，包含词汇、句法、语篇等多个维度。同时，本研究挖掘了表征文本复杂度的关键性特征，可以为教育学、阅读心理学的相关理论和教育教学领域的分级阅读标准的制定提供科学依据。

再次，在实践层面，本研究具有一定的应用价值。第一，为文本匹配适合对其进行阅读的目标读者，既可以帮助教师选择适合学习者水平的阅读材料，也可以帮助学习者找到适合自己水平的阅读材料。第二，为教材文本的选择、编写和修订提供量化反馈和有效帮助。在有参照语料库的情况下，可以借助本研究的特征和建模方法构建分级模型，对教材待选文本或现有文本进行级别预测，对与原先等级不相符的文本进行调整或改编；在参照语料库匮乏的情况下，本研究的训练语料可以为不同复杂度等级文本的语言特征提供一个参考范围，教材编写者可以此为参照，选择合适的文本进行教材编写或改编。第三，为实际教学、阅读课程规划和相关测试设计提供参考。在实际教学中，本研究可以对正在

使用的教材复杂度进行量化考察,教师可以根据教情和学情进行个性化教学。同时,本研究提取的文本复杂度特征和建模方法也可以用来对其他阅读文本和测试材料进行评估和定级。

1.3 本书结构

本书分为八章。第一章为概论,主要从教材研究、教材文本复杂度研究和文本复杂度自动分级研究的重要性三个方面介绍研究缘起,并对本研究的方法论意义、理论意义和实践意义进行简要阐述。

第二章为介绍文本复杂度概念界定及测量工具,就本研究的核心概念“语言复杂度”进行界定,厘清“文本复杂度”及其易混概念,进一步明确本研究的研究对象,并对其所包含的各个维度及现有主要测量工具进行介绍和梳理。

第三章为文本复杂度实证研究综述,主要包括基于语料库的文本复杂度研究和文本复杂度自动分级研究两部分,在总结以往研究的优势与不足的基础上,阐述本研究的切入点。

第四章为研究方法,首先提出了研究问题,接着介绍了本研究的语料和具体流程步骤,最后详细说明了本研究所采用的研究工具和方法,旨在保证研究结果的可信度。

第五章确定文本复杂度各维度特征。在简要说明参照语料库的选取后,分别阐述词汇、句法和语篇各维度数百个指标的主成分降维优化的具体过程,并对最终提取的主成分特征及其所包含的原始指标进行详细解释和说明。

第六章验证训练语料分级的合理性。首先将《新概念英语》(1997年版)确定为本研究的训练语料,并简要交代了其原由,然后基于前面所提取的主成分特征对该教材册级间复杂度变化进行考察,借助聚类 and 方差分析,对该教材的分级合理性进行验证。

第七章构建和评估文本复杂度分级模型。首先,在对几种常见的分类算法进行建模比较的基础上,确定采用性能最优的神经网络算法。其次,采用不同的特征集,分别构建模型,比较性能,旨在证明本研究采用的

多维语言特征集确实优于传统可读性特征和其他单维度特征。最后，借助全新的教材和非教材数据集对本研究模型的泛化能力进行考察，以验证本研究特征和建模方法的通用性和泛化能力。

第八章为结语，主要介绍本研究的主要发现，总结本研究的主要贡献和价值，指出研究不足和未来研究方向。