

# 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
<b>第四章</b>	<b>研究方法 .....</b>	<b>53</b>
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
<b>第五章</b>	<b>文本复杂度各维度特征的确定 .....</b>	<b>82</b>
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	动词特定性 .....	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
<b>第六章</b>	<b>训练语料分级合理性的验证 .....</b>	<b>113</b>
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
<b>第七章</b>	<b>文本复杂度分级模型的构建和评估 .....</b>	<b>142</b>
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
<b>第八章</b>	<b>结语 .....</b>	<b>168</b>
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
<b>参考文献</b>	<b>.....</b>	<b>177</b>
<b>附录一</b>	<b>基于语料库的英语教材文本复杂度研究一览表 .....</b>	<b>202</b>
<b>附录二</b>	<b>基于 BNC baby 提取的主成分特征 .....</b>	<b>206</b>