

文本难度调控的研究与实践

——从可读公式、多维特征到智能改编*

郭凯¹ 金檀² 陆小飞³

提要: 为外语学习者提供难度合适的文本,需根据文本的复杂程度进行选取与改编,即文本难度调控。本文对文本难度调控的研究与实践进行综述。文本难度调控的发展历程可分为“可读公式”与“多维特征”两个阶段。在回顾发展历程基础上,本文指出,文本难度调控已逐步向“智能改编”阶段过渡,并以“英语文本指难针”为个案讨论这一过渡的研究背景与实践价值。结合我国外语教学与研究情景,本文最后为未来智能改编的研究发展提供可操作性建议。

关键词: 文本难度调控;可读公式;多维特征;智能改编

Abstract: The selection of reading texts of appropriate difficulty for second and foreign language learners involves assessing texts for linguistic complexity, which if necessary, should be adjusted by modifying the texts, a process referred to as text adaptation. This study reviews the history of and current trends in text adaptation research and practice. The authors start with a brief introduction of the two stages in the historical development of research into text adaptation, namely, “readability formula” and “multidimensional analysis”. The authors then argue that text adaptation research is shifting towards the “data-driven” stage and, by using the case of *Eng-Editor*, illustrate the conceptual foundation and practical implications of such a transition. The authors conclude by putting forward a research agenda for applying the data-driven text adaptation approach in the context of foreign language teaching and research in China.

Key words: text adaptation; readability formula; multidimensional analysis; data-driven adaptation

1. 引言

文本难度,在广义上是指一篇文本的难易程度;在狭义上是指一篇文本在语言特征上所呈现出来的复杂程度,即文本复杂度(text complexity, 参见 Mesmer *et al.* 2012)。对于外语学习者来说,使用与其语言水平相适应的文本将有助于其语言能力的发展(Crossley *et al.* 2012; Hiebert & Mesmer 2013),使用明显低于或高于其语言水平的文本都将不利于甚至阻碍其语言能力的发展(Brookhart 2003; Kontovourki 2012)。因此,需要将文本的复杂程度与外语学习者的语言水平进行适度匹配,以取得良好的学习效果。为此,在为外语学习者选取学习文本时,需要对文本的复杂度进行量化计算以判断其难度情况,并在此基础之上对文本进行适当改编以调整其复杂程度,这一过程称为“文本难度调控”。

外语教师在文本难度调控中扮演着十分重要的角色(Bunch *et al.* 2014)。为了逐步提高

* 本研究为广东省哲学社会科学“十三五”规划项目“移动互联网英语文本资源智能化评估研究”(编号:GD17WXZ01)研究成果之一。感谢有米科技首席科学家李百川博士提供有关机器学习的专业意见。

外语学习者的语言能力,外语教师在准备教学素材时,需要根据学生现有水平和实际教学情景,适度调控文本难度,通常是降低真实文本的难度,以适应外语学习者的语言水平。然而,将不同难度的文本与不同水平等级进行匹配,是一项非常困难的任务,即使对于教学经验丰富的外语教师来说,也是极具挑战性的(Sung *et al.* 2015)。目前,在文本难度调控过程中,外语教师能够获得的帮助或可参考的依据是极少的,他们往往只能凭借教学经验或个人直觉来选取和改编真实文本(Crossley *et al.* 2012; Young 1999)。

本文将从研究与实践两个层面,对文本难度调控的发展历程进行综述,回顾文本难度调控在不同发展时期的理论基础、实践应用和阶段特点,并侧重探讨文本难度调控目前的发展状况以及未来的发展方向,以期为我国文本难度调控研究提供理论参考,并为我国外语教师提供实践帮助。

2. 文本难度调控发展历程

2.1 “可读公式”阶段

20世纪70年代,文本难度缺乏客观、统一的评判标准,为了向学习者提供难度适合的学习文本,研究者开始寻找测量文本复杂度的量化指标、探究有效的计算方法,这可以视为文本难度调控的早期研究阶段。在这一阶段中,由于技术手段的限制,研究者主要通过设定公式来计算少量能够区分阅读难度的文本表面特征(如平均词长和平均句长等),进而对文本进行难度定级,故而这一阶段称为“可读公式”阶段(Klare 1974-1975; Koslin *et al.* 1987)。这一阶段初期的典型代表为Flesch-Kincaid可读公式(Flesch-Kincaid Grade Level)(Kincaid *et al.* 1975),其本质为一个多元回归方程。其公式内容为:

$$\text{Grade Level of Text} = 0.39 * \text{ASL} + 11.8 * \text{ASW} - 15.59。$$

其中,自变量包括平均句长(average sentence length, ASL)和平均词长(average number of syllables per word, ASW),前者使用每个句子的单词数来计算,后者使用每个单词的音节数来计算。在这个等式中,11.8是赋予词汇难度的权重参数,0.39是赋予句子难度的权重参数。通过该公式的计算,可将文本难度划分为10个不同难度等级(Grades 1-10)。

随着信息技术的不断发展,研究者对于文本难度的量化计算方式不再局限于小规模的人工计算,而逐渐转为采用以计算机为辅助工具的大规模计算。尤其是对于词汇特征的计算,由原来的将文本中的单词与等级词表进行匹配,转变为将单词的平均词频与文本语料库进行匹配。这一阶段后期的典型代表为Lexile阅读框架(Lexile Framework for Reading)(Stenner & Fisher 2013)。Lexile阅读框架的应用十分广泛,且影响深远,被用于美国共同核心国家标准(Common Core State Standards)中的阅读文本难度分级。Lexile阅读框架主要依据对目标文本的单词频率和平均句长等指标的分析,将其套入可读公式,进行对数处理和标准差分析,从而得到目标文本的难度值,使美国共同核心国家标准拥有了一套完整的文本难度评估体系,由低到高共包含12个难度等级(Grades 1-12)。与初期阶段的可读公式相比,Lexile阅读框架虽然在技术手段方面有所提升,但其所聚焦的文本特征仍停留在词汇和句子两个层面,且核心理念仍为以线性公式来计算文本难度。

由此可见,在可读公式阶段,研究者将线性发展作为学习者语言能力发展的主要理论假设依据,因而采用线性模型建立文本难度调控公式,并将词汇和句子的核心特征作为主要测量指标。与此同时,研究者对不同水平阶段划定了参考取值范围,通过公式计算得出结果数值之后,根据取值范围判断文本难度等级。由于公式简单、使用便捷,在文本难度调控的早期阶段,

可读公式得到广泛应用 (DuBay 2004)。然而,可读公式仅涵盖语言表层特征,并未涉及文本类型、文本含义及连贯程度等诸多要素,而这些要素可能会对文本难度判断产生不同程度的影响。例如,在科普类文章中,通常由于需要介绍某个生僻概念而重复出现某一单词,依照可读公式,这可能会增加文本的词汇难度值进而影响文本的整体难度值。但实际上,这一生僻单词并不会影响学生对文本总体内容的理解,生词复现对于学生学习反而有所帮助。因此,众多学者认为可读公式缺乏阅读能力构念代表性,在实际应用中存在一定问题 (McNamara *et al.* 2014),应继续探索更为可靠的文本难度调控方法。

2.2 “多维特征”阶段

从 20 世纪 90 年代到 21 世纪初,由于自然语言处理、计算语言学和二语习得等学科的进一步发展,文本难度调控开始吸收语义、修辞及连贯性等语言深层特征,对文本进行更为全面的难度定级研究 (McNamara & Kintsch 1996)。由于这一阶段的文本难度调控研究包含诸多特征且具备广泛意义的构念代表性 (Ozuru *et al.* 2009),故而称为文本难度调控的“多维特征”阶段。在这一阶段,除了纳入考量范围的语言特征不断丰富之外,伴随着语料库技术的逐渐成熟,研究者开始使用大规模典型文本语料库,对不同等级的文本复杂度参考范围进行有效验证。这一阶段的典型代表包括 Reading Maturity Metric (Landauer *et al.* 2011), TextEvaluator (Sheehan *et al.* 2014) 和 Coh-Matrix (Graesser *et al.* 2011)。

Reading Maturity Metric 主要通过对词汇和多项文本结构特征进行测量,实现文本难度调控 (Landauer *et al.* 2011)。其中,词汇层面的主要指标为词汇成熟度 (word maturity),其通过模拟语言学习者的词汇意义发展情况建立数学模型,是对潜在语义分析 (latent semantic analysis) 的一种应用。TextEvaluator 通过主成分分析 (principal component analysis),选取学术词汇、句法复杂度、词汇具体度、词汇陌生度、互动或会话风格、叙述程度、词汇连贯性和论证共 8 个维度对文本难度进行调控 (Sheehan *et al.* 2014),基于这 8 个维度对文本难度差异判断的准确程度可以达到 60% 以上。Coh-Matrix 主要基于基础计数、文本主构易懂性、指称连贯性、潜在语义分析、词汇多样性、关联词使用、情景模式、句法复杂性、句法组构密度、词汇和文本可读性等 11 大类,总计 108 项指标,对文本难度进行调控 (Graesser *et al.* 2011)。

由此可见,在“多维特征”阶段,研究者认识到单一维度的难度测量对于提高学习者语言能力的作用不大,因而选用了多个维度的特征指标对文本难度进行量化计算。同时,研究者使用了不同级别的基准文本作为参照数据,采用数据挖掘技术,以机器学习作为模型拟合的主要手段^①。然而,需要指出的是,无论是在可读公式阶段还是多维特征阶段,对于文本难度的调控都忽视了外语教师的主体作用,未将外语教师置于整个过程的核心地位,而是寻求数学公式或机器工具的绝对支配和权威地位,外语教师只是处于被动接受的位置。不可否认,无论是公式还是机器,都只能为文本选取与难度定级提供一个统一标准,而不能根据实际情景做出相应调整,只有外语教师才有能力根据具体的教学情景和学生的现有水平作出判断。此外,在以上两个阶段中,难度调控只注重对文本难度的最终效果评定,而不能在调控过程中对外语教师改编文本提供有效支持。因此,本文认为,文本难度调控应该且正在过渡到一个新的发展阶段——“智能改编”阶段。

^① 一般将文本特征数值化,以基准文本的级别作为目标值,采用机器学习的分类算法,如支持向量机和随机森林等训练模型,拟合出一个从特征到目标值的复杂非线性函数。

2.3 “智能改编”阶段

经历前两个阶段之后,研究者发现虽然“可读公式”和“多维特征”可为文本难度调控提供客观依据,但在难度判断之后,真实文本还需要外语教师来进行确认和改编。因此,研究者逐渐认识到了外语教师在文本难度调控过程中的主体地位和能动作用(Jin & Lu 2018)。由于外语教师在此过程中通常仅仅依靠个人的主观经验(Crossley *et al.* 2012; Young 1999),有必要深入了解外语教师在改编过程中的具体做法。Green & Hawkey (2012)探究了雅思考试命题员文本改编的主要方式和手段。研究者邀请了4位经验丰富的命题员参与研究,他们首先对自己过往的命题经历进行反思,然后为雅思阅读理解模块命制题目。通过观察,研究者发现,命题员命题核心环节主要包括资源选取、仔细阅读、文本改编和题目设计等。其中,在文本改编这一环节,他们通常使用合并、删除、替换、扩充以及插入等方法对文本难度进行调控。然而,不同教师在改编程度上存在较大差异,例如在调整低频词比例时,4位命题员分别将真实文本中的低频词比例降低了2%、2.8%、6.5%和3.9%。由此可见,造成这些差异的主要原因之一,就是外语教师在文本改编过程中缺乏量化反馈和有效帮助(Jin *et al.* 2016)。

因此,Jin & Lu (2018)提出应该实现文本难度调控过程中的“智能改编”,即基于外语教师在改编过程中的实际需求,将机器工具作为调控过程中的有效辅助,实现人机合作的调控模式。在智能改编的过程中,机器工具可为外语教师提供“难度等级判断”“量化改编建议”“后续教学反馈”三个环节的数据驱动支持,帮助外语教师完成对真实文本的难度改编,并最终应用于实践(Lu & Jin 2017; 金檀,郭凯 2018)。如图1所示,在“难度等级判断”环节中,机器工具可为外语教师提供真实文本的核心特征测量指标,并依据所建立的难度等级体系划定文本的难度情况。在“量化改编建议”环节中,机器工具可依据难度调整目标为外语教师明确需要改编的对象,并提供量化参考范围。在“后续应用反馈”环节中,机器工具可根据外语教师的实际需求,为其提供改编文本在后续使用过程中的可行建议。第一和第二个环节主要关注文本改编过程中的具体操作,而第三个环节主要关注文本改编结束后的结果应用。

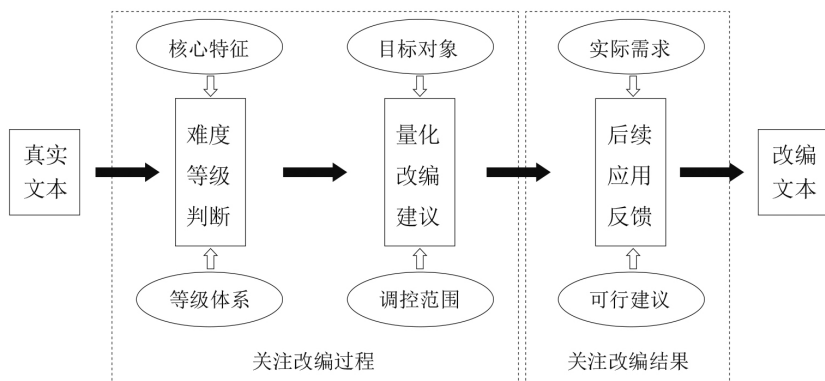


图1 智能改编理论框架图示

“难度等级判断”是指对文本复杂度的核心特征进行量化计算得出文本难度,这是对“可读公式”阶段和“多维特征”阶段的吸收和继承。21世纪以来,由于代表性特征研究发展迅速,可用于调控文本难度的特征不断增多,有时甚至涵盖上百个指标(Crossley *et al.* 2008; Graesser *et al.* 2011)。然而,测量特征的选取并非越多越好,盲目追求指标数量可能会影响文本难度调控的效果。Sung *et al.* (2015)研究发现,在众多文本测量指标中,仅有7项指标对难度测量准

确度的影响程度较高,且综合测量准确度可以达到70%以上。继续增加测量指标,对于提高测量准确度非但没有明显效果,指标超过25个之后甚至会降低综合准确度。因此,在智能改编的文本难度调控中,对于语言特征的选取尤为重要,应通过语料库挖掘技术考察不同文本特征在不同文本类型和各个水平等级的影响作用,进而建立预测模型、确立最优方法(James *et al.* 2013; Sung *et al.* 2015),并形成基准难度等级体系,将目标文本与之匹配,从而对文本难度进行初步判断。由此,帮助外语教师解决“这篇文本是否需要改编”的问题。

“量化改编建议”是指使用文本特征的参考标准明确文本改编的目标对象,通过挖掘基准文本提供改编的量化范围与调控建议。Green & Hawkey (2012) 研究发现,雅思考试命题员在文本改编过程中,主要通过词汇、句子和篇章3个层面的调整控制文本难度。因此,机器工具应基于这3个层面为外语教师提供改编建议,例如文本的低频词汇、平均句长和篇章长度在不同难度等级上的比例范围。划定比例范围,应先根据真实文本改编之后的应用情景确定基准文本并建立基准文本语料库,而后通过数据挖掘和分析得出目标情景下的量化范围。例如,真实文本改编之后将应用于我国初中英语教材编写,则应系统收集我国中小学英语教材典型文本并建立语料库,从而获得我国中小学英语教材在不同阶段的文本难度等级,以及难度情况在词汇、句子和篇章层面的反映,以此为基准提供真实文本的量化改编建议。因此,在智能改编的文本难度调控中,应通过对典型文本语料库的数据挖掘,建立量化参考标准,用以明确改编对象及调控范围。由此,帮助外语教师解决“这篇文本应该如何改编”的问题。

“后续应用反馈”是指根据外语教师的实际需求,提供有助于辅助后续应用的重要信息,并为外语教师提出可行建议。智能改编在这一方面所提供的支持,已不仅仅局限于具体改编的行为和过程,更着眼于后续应用的可能和结果,也就是在改编过程中就应考虑到改编之后的使用问题。想要达到这一目的,首先应该了解外语教师的具体需求和改编文本的主要目的。例如,如果改编文本将要用于课堂词汇教学,那么机器工具就应该为外语教师提供文本中生词的释义和例句等有利于辅助教师进行词汇教学的反馈信息。如果改编文本将要用于课堂词汇测试,那么机器工具就应该为外语教师提供文本中单词的频率和难度等关于试题命制的反馈信息。因此,在智能改编的文本难度调控中,应对一线外语教师的实际需求进行调查,在此基础上,对机器工具的具体功能进行设计和完善,从而在文本改编的过程中解决后续应用可能产生的问题、满足外语教师需求。由此,帮助外语教师解决“文本改编之后怎么使用”的问题。

在“智能改编”这一阶段,研究者聚焦文本改编过程中的人机合作,强调外语教师在难度调控过程中的能动作用,并关注机器工具在“难度等级判断”“量化改编建议”“后续教学反馈”3个方面为外语教师提供实时的、数据驱动的有效支持,依次帮助教师解决“这篇文本是否需要改编”、“这篇文本应该如何改编”和“文本改编之后怎么使用”的问题。笔者接着以“英语文本指难针”为例,讨论智能改编实现工具的研究背景与实践价值。

3. 智能改编工具

为实现智能改编,金檀、李百川(2016)主持研制了“英语文本指难针”(以下简称“指难针”)。这是一项基于数据驱动的在线文本难度调控工具(Jin & Lu 2018)。其研发初期,研究者通过问卷调查的方式,搜集并了解我国外语教师在文本难度调控过程中所需获得的支持类型。分析结果显示,我国外语教师期望获得的支持主要可归为两类:第一,对真实文本的难度判断,希望通过机器自动分析计算快速了解一篇文本的难度情况。第二,对真实文本的特征标注,希望机器能够对一篇文本中的低频词汇、复杂句子和文本长度进行标注,并依据不同目标

等级,提供文本特征的量化参考范围。研究者认为,我国外语教师的这两类需求分别属于智能改编中的“难度等级判断”和“量化改编建议”。

3.1 需求一:难度等级判断

为满足我国外语教师的这一需求,研究者建立了“英语试题语料库”作为“指难针”的基准语料库,系统收集、整理并储存了国内大规模考试的试题素材文本,为我国特定水平等级的阅读文本难度分析和基准设定提供了语料基础。自2012年始建至今,该语料库已系统收集了历年国内大规模考试的试题素材文本三千余篇,包括基础教育阶段的中考、高考英语试题和高等教育阶段的大学英语四、六级与考研英语试题。语料库中所有文本样本均由经验丰富的测试专家和外语教师严格挑选(Jin & Lu 2018)。通过对语料库中样本文本的数据挖掘,研究者建立了与国内大规模考试阅读文本难度相对应的难度等级体系,包括“中考”“高考”“四级”“六级”“考研”5个等级。基于此,外语教师可以了解真实文本的难度情况,并将其与不同阶段水平的学生进行有效匹配。

3.2 需求二:量化改编建议

为满足我国外语教师的特征标注需求,“指难针”首先将需要改编的真实文本与符合目标难度的词表进行对比,对超出词表范围的单词(即超纲词)进行标注。这里使用的是国家教育部颁布的英语课程标准中提供的对应不同学习阶段和能力水平的词表(Jin *et al.* 2016)。其次,“指难针”对真实文本中的句子长度进行计算,并对最长的句子进行标注以引起外语教师的注意。需要说明的是,句法复杂度是一个多维结构,存在多个测量指标(Lu 2011, 2017),研究者之所以选择句子长度而非句法复杂度作为标注对象,是因为句子长度是一个被广泛使用的有效指标,且最为直观,相比之下更易于教师的理解和操作(Gamson *et al.* 2013)。鉴于外语教师在文本难度调控过程中对文本长度的特别关注,“指难针”可以计算并显示文本的总词数。

为满足我国外语教师的参考范围需求,研究者对基准语料库中的文本样本进行了超纲词、最长句和总词数的特征标注,并基于样本文本的数据分析,针对不同水平等级分别设立了此3个维度的量化参考范围(Jin *et al.* 2017),为教师提供原文难度、参考基准、改编提示与改编后难度等信息。例如,用于难度等级为“四级”的文本,超纲词比例应该控制在3%到7%之间。如果3篇真实文本的超纲词比例分别为2%、5%和8%,“指难针”将分别提示“偏少”“适中”“偏多”。基于此,外语教师可以实时了解文本的改编程度和难度情况,并依据不同指标的参考范围进行有针对性的调控。

3.3 需求补充:后续教学反馈

基于上述两类需求,研究者完成了“指难针”试用版的研发,并邀请外语教师使用,辅助其完成文本难度调控。他们表示,“指难针”为改编提供了丰富的参考,使文本难度调控得到了有效支持(Jin & Lu 2018)。与此同时,部分外语教师也对试用版提出了一些疑问,主要集中于改编文本的后续使用问题,希望机器能够针对课堂教学提供反馈意见。例如,有的教师认为,虽然已经能够判断文本难度并将其改编到需要的难度范围,但是将改编文本用于课堂教学时,常常很难把握应该对哪些超纲词进行重点讲解,应该提供哪些信息帮助学生理解。基于这些有价值的反馈意见,研究者发现,应将改编文本在后续使用中的问题纳入考虑范围,也就是智能改编中的“后续教学反馈”。因此,研究者对“指难针”进行了修订。

修订后的“指难针”可为外语教师的课堂教学提供有效的帮助。例如针对词汇教学,“指

“指南针”首先可对每个超纲词在基准语料库中的出现频率进行标注(参见 Gardner & Davies 2014) 并提供语料库中包含该单词的句子。基于此,外语教师可以了解文本中的超纲词在真实语料中的使用情况,在日常教学时可以将其作为例句参考,并对比超纲词在不同水平等级基准语料中的义项与用法。其次,基于标记出的超纲词,“指南针”可以直接为外语教师生成按基准语料库中出现频次排序的超纲词表,从而实现对一篇文本中超纲词的优先程度排序,教师可以有针对性地选择教学过程中的重点词汇。最后,“指南针”还可以提供每个超纲词的英文释义和同义词等信息,外语教师可以十分便捷地对文本中的超纲词进行替换或添加注释,并在课堂教学时将这些信息提供给学生。

由此可以发现,针对我国外语教师的实际需求所研发的“指南针”,可以分别帮助我国外语教师解决在文本难度调控过程中“这篇文本是否需要改编”“这篇文本应该如何改编”“文本改编之后怎么使用”的3个问题,与智能改编理论框架中的“难度等级判断”“量化改编建议”“后续教学反馈”3个方面相符合,可以视为智能改编的典型实现工具。

4. 结语与建议

自20世纪70年代至今,文本难度调控的研究与实践已经历多个发展阶段。在发展初期,即“可读公式”阶段,研究者以线性发展为理论假设,主要选取词汇和句子的核心特征作为测量指标,并建立线性公式对文本复杂度进行判断。随着代表性特征研究迅速发展,文本难度调控逐步进入了“多维特征”阶段,研究者开始选来自多个维度的众多特征指标进行量化计算,并对不同级别的基准文本进行数据挖掘得出参照指数,从而对文本复杂度进行判断。如今,文本难度调控已经逐渐过渡到了一个新的阶段——“智能改编”阶段,这一阶段突出强调外语教师在文本难度调控过程中的主体作用,提倡采用人机合作的调控模式,使机器真正成为教师的有效辅助。

然而,新的时期和阶段也为文本难度调控的发展提出了新的要求和挑战,仔细思考不难发现,仍存在许多问题有待解决。本文针对文本智能改编的未来发展提出3点研究建议。第一,建立文本难度调控的“语料数据联盟”。目前国际上已成立数据联盟(Mandinach & Gummer 2016),我国也可面向英语教学与测试资源建立文本的“语料数据联盟”——长期、全面地收集高质量文本资源,进行多维度语料标注,作为高质量的语料基准。第二,分别从改编过程和改编结果两个角度进一步检验“智能改编”的效度问题,并为其优化提供实证证据(Jin & Lu 2018)。改编过程研究应聚焦教师在文本改编过程中的认知,从而检验智能改编的运行机制、组成要素和反馈维度等方面是否需要进一步优化。改编结果研究应聚焦学生使用改编文本时的学习表现,从而检验智能改编的效果影响和教师的改编行为等方面是否需要进一步优化。第三,发展外语教师的“数据素养”,培养外语教师基于语言的“数据意识”“数据知识”“数据工具”“数据使用”,帮助教师将数据转化为有效信息,并将信息应用于教学实践,从而进一步促进外语教师的专业发展(金檀等 2018)。

文本难度调控的发展,应基于面向我国外语教学与测试情景的、高质量的“文本语料基础”,应进一步检验并完善“智能改编”的效度,应认可外语教师在文本难度调控中的主体地位,并进一步提升外语教师的数据素养,从而有力推动外语教师的专业发展。

参 考 文 献

- [1] Brookhart S M. Developing measurement theory for classroom assessment purposes and uses [J]. *Educational*

- Measurement, Issues and Practices*, 2003, 22: 5–12.
- [2] Bunch G C, Walqui A & Pearson P D. Complex text and new common standards in the United States: Pedagogical implications for English learners [J]. *TESOL Quarterly*, 2014, 48: 533–559.
- [3] Crossley S A, Greenfield J & McNamara D S. Assessing text readability using cognitively based indices [J]. *TESOL Quarterly*, 2008, 42: 475–493.
- [4] Crossley S A, Allen D & McNamara D S. Text simplification and comprehensible input: A case for an intuitive approach [J]. *Language Teaching Research*, 2012, 16: 89–108.
- [5] DuBay W H. *The Principles of Readability* [M]. Costa Mesa, CA: Impact Information, 2004.
- [6] Gamson D A, Lu X & Eckert S A. Challenging the research base of the Common Core State Standards: A historical reanalysis of text complexity [J]. *Educational Researcher*, 2013, 42: 381–391.
- [7] Gardner D & Davies M. A new academic vocabulary list [J]. *Applied Linguistics*, 2014, 35: 305–327.
- [8] Graesser A C, McNamara D S & Kulikowich J M. Coh-matrix: Providing multilevel analyses of text characteristics [J]. *Educational Researcher*, 2011, 40: 223–234.
- [9] Green A & Hawkey R. Re-fitting for a different purpose: A case study of item writer practices in adapting source texts for a test of academic reading [J]. *Language Testing*, 2012, 29: 109–129.
- [10] Hiebert E H & Mesmer H A E. Upping the ante of text complexity in the Common Core State Standards examining its potential impact on young readers [J]. *Educational Researcher*, 2013, 42: 44–51.
- [11] James G, Witten D, Hastie T & Tibshirani R. *An Introduction to Statistical Learning* [M]. New York: Springer, 2013.
- [12] Jin T, Li Y & Li B. Vocabulary coverage of reading tests: Gaps between teaching and testing [J]. *TESOL Quarterly*, 2016, 50: 955–964.
- [13] Jin T, Guo K, Mak B & Wu Q. Lexical profiles of reading texts in high-stakes tests: Where are the benchmarks? [J] *International Journal of Computer-Assisted Language Learning and Teaching*, 2017, 7: 34–49.
- [14] Jin T & Lu X. A data-driven approach to text adaptation in teaching material preparation: Design, implementation, and teacher professional development [J]. *TESOL Quarterly*, 2018, 52: 457–467.
- [15] Kincaid J P, Fishburne L R P, Rogers R L & Chissom B S. Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for navy enlisted personnel [R]. *Research Branch Report, 8-75, Millington, TN: Naval Technical Training*. U.S. Naval Air Station, Memphis, TN, 1975.
- [16] Klare G R. Assessing readability [J]. *Reading Research Quarterly*, 1974–1975, 10: 62–102.
- [17] Kontovourki S. Reading leveled books in assessment-saturated classrooms: A close examination of unmarked processes [J]. *Reading Research Quarterly*, 2012, 47: 153–171.
- [18] Koslin B I, Zeno S & Koslin S. *The DRP: An Effective Measure in Reading* [M]. New York: College Entrance Examination Board, 1987.
- [19] Landauer T K, Kireyev K & Panaccione C. Word maturity: A new metric for word knowledge [J]. *Scientific Studies of Reading*, 2011, 15: 92–108.
- [20] Lu X. A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development [J]. *TESOL Quarterly*, 2011, 45: 36–62.
- [21] Lu X. Automated measurement of syntactic complexity in corpus-based L2 writing research and implications for writing assessment [J]. *Language Testing*, 2017, 34: 493–511.
- [22] Lu X & Jin T. *Teachers' Use of Corpus-based Tools: An Online Professional Development Course* [Z/OL]. Beijing: Foreign Language Teaching and Research Press (Unipus), 2017.

- [23] Mandinach E B & Gummer E S. What does it mean for teachers to be data literate: Laying out the skills , knowledge , and dispositions [J]. *Teaching and Teacher Education* ,2016 ,60: 366-376.
- [24] McNamara D S & Kintsch W. Learning from text: Effects of prior knowledge and text coherence [J]. *Discourse Processes* ,1996 ,22: 247-287.
- [25] McNamara D S , Graesser A C , McCarthy P M & Cai Z. *Automated Evaluation of Text and Discourse with Coh-Matrix* [M]. Cambridge: Cambridge University Press ,2014.
- [26] Mesmer H A , Cunningham J W & Hiebert E H. Toward a theoretical model of text complexity for the early grades: Learning from the past , anticipating the future [J]. *Reading Research Quarterly* ,2012 ,47: 235-258.
- [27] Ozuru Y , Dempsey K & McNamara D S. Prior knowledge , reading skill , and text cohesion in the comprehension of science texts [J]. *Learning and Instruction* ,2009 ,19: 228-242.
- [28] Sheehan K M , Kostin I , Napolitano D & Flor M. The Text Evaluator tool: Helping teachers and test developers select texts for use in instruction and assessment [J]. *Elementary School Journal* ,2014 ,115: 184-209.
- [29] Stenner A J & Fisher W P. Metrological traceability in the social sciences: A model from reading measurement [J]. *Journal of Physics: Conference Series* ,2013 ,459: 10921-10930.
- [30] Sung Y T , Lin W C , Dyson S B , Chang K E & Chen Y C. Leveling L2 texts through readability: Combining multilevel linguistic features with the CEFR [J]. *Modern Language Journal* ,2015 ,99: 371-391.
- [31] Young D N. Linguistic simplification of SL reading material: Effective instructional practice? [J] *Modern Language Journal* ,1999 ,83: 350-366.
- [32] 金檀,郭凯.“学者笔谈”系列6:智能改编与“指难针”[Z/OL]. http://www.sohu.com/a/233355215_161093 2018.
- [33] 金檀,郭凯,李百川.外语教师的数据素养与专业发展[R].北京:第三届全国高等学校外语教育改革与发展高端论坛,2018.
- [34] 金檀,李百川.“英语文本指难针”[Z/OL]. www.languagedata.net/tester ,2016.

作者联系方式: 1. 上海交通大学外国语学院,上海 200240
2. 中山大学外国语学院,广东 广州 510275
3. 广东外语外贸大学,广东 广州 510420 [美]宾夕法尼亚州立大学,PA 16802

(上接第34页)

- [31] 肖维青.本科翻译专业测试研究[M].北京:人民出版社,2012a.
- [32] 肖维青.多元素翻译能力模式与翻译测试的构念[J].外语教学,2012b,(1):109-112.
- [33] 杨冬敏.笔译资格考试构念效度研究:测前理论效度的视角[M].北京:科学出版社,2015.
- [34] 邹申.语言测试[M].上海:上海外语教育出版社,2005.
- [35] 庄起敏.对翻译作为英语能力测试手段的分析与思考[J].外语界,2006,(3):55-59.
- [36] 庄智象.我国翻译专业建设:问题与对策[M].上海:上海外语教育出版社,2007.
- [37] 赵玉闪,王志,卢敏.全国翻译专业资格(水平)考试笔译试题命制一致性研究报告[J].中国翻译,2007,(3):53-56,94.
- [38] 席仲恩.“效度”理论对翻译测试的呼唤[J].英语研究,2017,(5):143-150.

作者联系方式: 南方医科大学外国语学院,广东 广州 510515