

统编小学语文教材选文可读性影响因素分析

韩景

福建师范大学

DOI:10.12238/mef.v8i15.16187

[摘要] 可读性是衡量文本难度的重要方式,本文研究如何用汉语的语言特征衡量文本的可读性,可读性的衡量包括定量指标和定性指标。其中定量指标包括字数、词数、字频、词频、句子数、平均句长等,定性指标包括抽象词、句子类型、句子结构等。本文通过对1-6年级每册各选取3篇文章,从定量、定性指标对文章进行统计分析,发现大部分指标呈线性上升趋势,但在每个年级中几乎会有一篇文章与其他文章在指标数据方面的不同,因此,不能从单一的某一指标判断文本的可读性,综合各种指标,分析一篇文章的简单点和难点,为教材编写和教师教学提供参考意见。

[关键词] 可读性; 阅读难度; 教材

中图分类号: G722.4 文献标识码: A

Analysis of Influencing Factors on the Readability of Selected Texts in Unified-Compiled Primary School Chinese Textbooks

Jing Han

Fujian Normal University

[Abstract] Readability is a key method to measure text difficulty. This paper explores how to assess the readability of texts using the linguistic features of the Chinese language. The evaluation of readability encompasses quantitative indicators and qualitative indicators. Specifically, quantitative indicators include character count, word count, character frequency, word frequency, sentence count, and average sentence length; qualitative indicators include abstract words, sentence types, and sentence structures. In this study, three articles were selected from each textbook volume for grades 1 to 6. Statistical analysis of these articles was conducted from the perspectives of quantitative and qualitative indicators. The findings reveal that most indicators show a linear upward trend. However, in each grade, there is nearly one article that exhibits differences in indicator data from other articles. Therefore, text readability cannot be judged based on a single indicator alone. Instead, by synthesizing various indicators to analyze the difficult points and easy points of an article, this paper provides reference suggestions for textbook compilation and teachers' teaching practices.

[Key words] readability; reading difficulty; teaching materials

引言

语文教材是以选文为主体的教科书,在语文学科建设及教材编制的发展过程中,选文的选择与编排一直是教材编写者与一线教师研究的热点。选择什么样的文章?文章该置于哪个学段哪个单元?大大小小的问题涵盖了从教材编写到语文教学的全过程。

笔者在选择教材编写研究的论题时,首先对于《义务教育语文国家课程标准(2022版)》(以下简称“课程标准”)进行了阅读,发现课程标准在教材编写建议中提到,教材选文要“难易适度,适合学生学习”。换句话说,也就是一篇选文要适合所在年级的学生学习,不能过难或过易。语文教材的选文多选自原著,篇

幅较长、词语较难,为了适合学生的学习,教材编写者需要对其进行修改,降低难度。

可读性是衡量文本难度的重要方式。国内外学者目前更多的是分析外语和对外汉语文本影响可读性的因素,但这两种文本因面向学生群体不同,编写逻辑和教材也不同,因此,外语和对外文本的影响因素不能完全照搬到国内教材的分析上来。本文通过研究国内外相关论文成果,讨论分析影响文本可读性的指标,最终将分析视角定位到文本的字、词、句。

1 可读性的内涵

可读性(readability)是指一篇文本易于理解的程度。可读性越高,说明文本越简单;反之,说明文本越困难。Klare认为,

可读性可以指以下三个特点中的任意一个手写材料或者印刷品的易辨认性、作品的趣味性造成的易读程度、作品的写作风格带来的理解难易程度。所谓“写作风格”实际上就是指作者所使用的字词的特征^[1]。因此,Harries&Hodges(1995)、Fry(2002)直接将“可读性”定义为“写作中遣词造句的方式引起的阅读的难易程度”^{[2][3]}。

国内学者也给出了易读性的定义。孙汉银指出,易读性是研究一定水平的读者阅读书面材料时,影响读者成功理解文章内容的所有因素的总和(包括相互影响)^[4]。李绍山认为:易读性是文本易于阅读和理解的程度或性质^[5]。目前,国内对易读性的研究多沿用此定义。吕中舌将李绍山给出的定义做了更加深入的阐释,并把易读性的研究内容总结为两点:第一是对影响文本易读性因素的探讨,第二是测定文本对于特定读者是否过难的方法^[6]。这一定义是从中国学生的英语教学角度提出的,切实考虑到了可读性的应用范围,在对外汉语教学中也具有较强的可操作性。

2 影响教材文本可读性的影响因素

2.1 汉字

在对中文文本可读性的分析中,汉字数、字种、字频、难字是大多数学者关注的主要因素。

2.1.1 汉字数

汉字数,即一篇文章的汉字总量,是衡量文本长度的直接指标,也是影响学生阅读体验与效果的关键因素。篇幅过长的文本会天然地给学生带来显著的阅读压力。从认知负荷理论来看,长文本所承载的庞杂信息会迅速消耗学生的注意力与工作记忆资源,迫使他们投入更长的时间进行信息解码与主旨提炼。

2.1.2 平均字频

平均字频是指一篇文本中所有字种频率的平均值。在日常生活中,少数的汉字就可以覆盖到各个场景,《通用汉字规范表》收录了8105个汉字,BBC语料库汉字频率统计信息一共5742字。本文采用BBC语料库汉字频率统计信息统计教材中课文的平均字频。

2.1.3 字种

字种数即汉字的种类,指的是一篇文本中所有汉字的种类数目,即一篇文本中,汉字类别与汉字总字数的比例。小学阶段,尤其是小学中、低年级阶段,学生的核心任务之一就是扩大识字量,构建基础字库。此时,一篇课文的字种数,成为衡量其汉字教学价值的关键指标。

2.1.4 难字

难字指一个文本所有字种数中难字种所占的比例。其中本文对难字的认定是《现代汉语常用字表》2500字以外的字。只是用平均字频来统计的话,有可能得出两篇文本字频值差不多的情况,尽管其中有一篇难字相对较多的情况。例如二年级上册的《八角楼上》和三年级上册的《大自然的声音》,两篇文本的平均字频值为0.0506和0.0503,显然《大自然的声音》相对还低

一点。但是在难字的使用上《大自然的声音》使用了12个字,《八角楼上》仅使用5个字。因此,在评估文本可读性时,不能单纯地只使用平均字频一个指标,低频字也是重要影响因素之一。

在此基础之上,我在1-6年级课文中,每册随机抽取3篇课文,统计了以上4个因素随着课文年级递升的变化关系。最终得到的结果如图1所示:

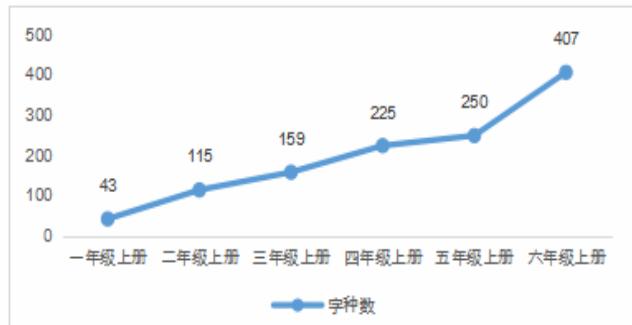
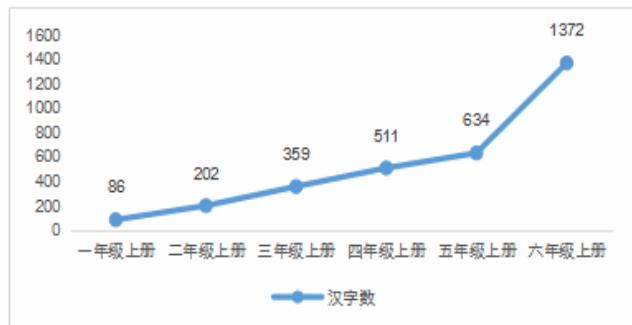


图1 小学1-6年级上册汉字影响因素

从图1可以看到,只有一年级上册到六年级上册。研究者主要是为了先从上册课文中分析4个影响因素变化的趋势,再从下册课文中验证趋势。从表2中可以看到,对于总字数来说,使用总字数最少的为小学一年级,平均总共使用了86字,其中字种数平均为43字,难字为0,平均字频为0.0894。使用总字数最多的为小学六年级,平均总共使用了1372字,其中字种数平均为407字,难字平均为21,平均字频为0.0328。

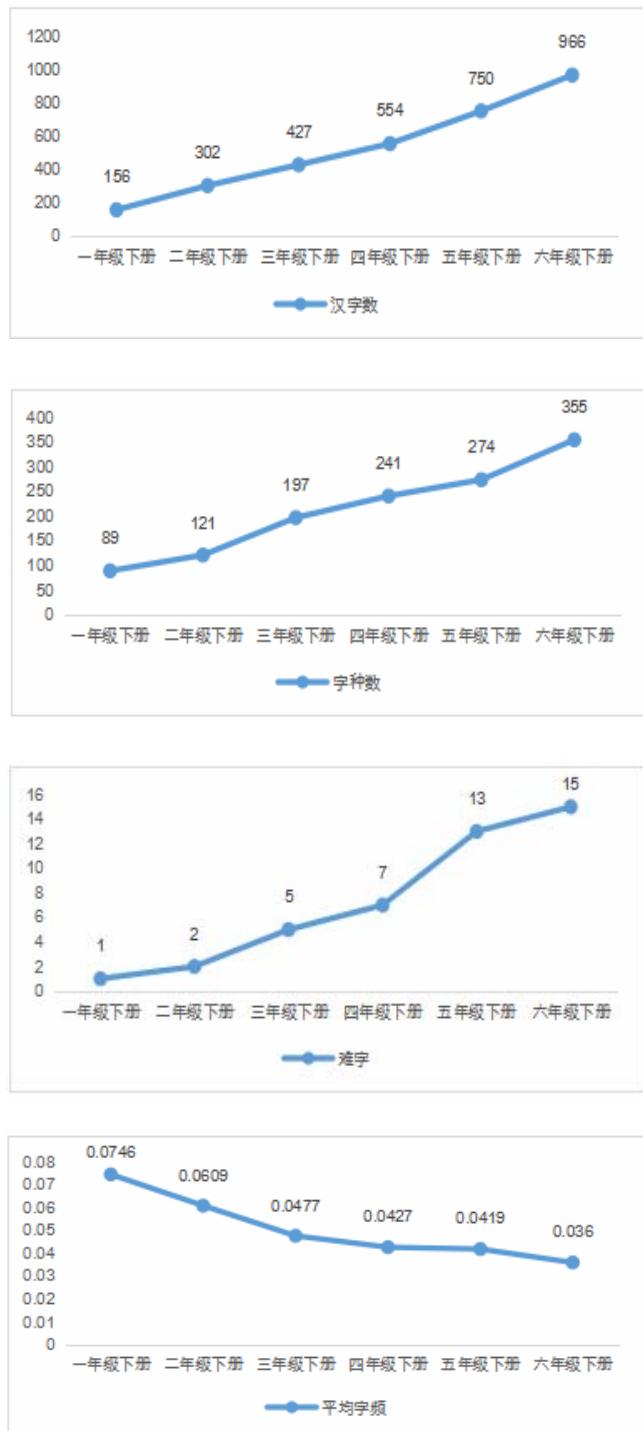


图2 小学1-6年级下册汉字影响因素

为了验证得出来的规律是否准确,笔者同时对小学1-6年级下册也各抽取3篇文章,进行统计,统计结果如图2所示:

从图2可以看到,对于总字数来说,使用总字数最少的为小学一年级,平均总共使用了156字,其中字种数平均为89个,难字平均为1个,平均字频为0.0746。使用总字数最多的为小学六年级,平均总共使用了966字,其中字种数平均为355字,难字为15,平均字频为0.036。从以上4个因素变化趋势来看,跟上册课文一致,总字数、字种数、难字都逐年递升,字的字频值在下降。

从以上4个影响因素在课文中的变化趋势来看,呈现线性变化趋势,即随着年级上升,文本中所使用的汉字数、字种数逐渐增加,字的字频值逐渐降低,而低频字的比例在逐渐上升。但需要注意的是,虽然字种数在增加,这是因为篇幅的影响,篇幅长,字种数自然多,但字种比例在逐渐降低。在高年级,一篇课文的用字丰富程度已经不再是学习的重点,教学的重心、学习的重心也向学习更难的汉字转变。

2. 词汇

在对中文文本可读性的分析中,总次数、词种、平均词频、实词、抽象词是大多数学者关注的主要因素,但哪些因素适合分析教材文本可读性仍要进行分析。

2.2.1 总词数

与总字数一样,从认知负荷理论看,阅读的基本处理单元是“词”而非单个的“字”。因此,总词数直接决定了读者需要识别与理解的语义模块数量。

2.2.2 平均词频

平均词频,是衡量一篇文本用词常用度的核心指标。该数值能够有效反映文本整体词汇的常见性与通用性水平。为此,语言学家李行健与苏新春先生编纂了具有权威性的《现代汉语常用词表》,该词表系统性地收录了56,790个常用词汇,并依据其在大型语料库中的出现频次进行了降序排列,为汉语词汇的常用度研究提供了客观、量化的标准。

2.2.3 词种数

词种数指文本中不重复的词汇数。每一个新词,都需要学生调动工作记忆来解码其含义、理解其用法、并将其整合到不断发展的语境模型中去。因此,尽管词种丰富是语言发展的目标,但在单篇课文中若过度集中,同样会显著增加学生的认知负荷,使其将过多的精力耗费在词汇层面的局部理解上,从而影响对文章整体结构与核心思想的把握。

2.2.4 抽象词

小学低年级阶段,他们的思维正处在从具体形象向抽象逻辑过渡的阶段,认知往往依赖于亲眼所见、亲手所触的实际经验。因此,像“苹果”“跑步”这类具体实词,他们能迅速与脑海中的形象关联,轻松理解;但面对“公平”“系统”“精神”这类抽象词时,由于无法对应到一个确切的事物或动作,就容易产生困惑和隔阂。

在此基础之上,我统计了以上6个因素随着课文年级递升的变化关系。最终得到的结果如图3所示:

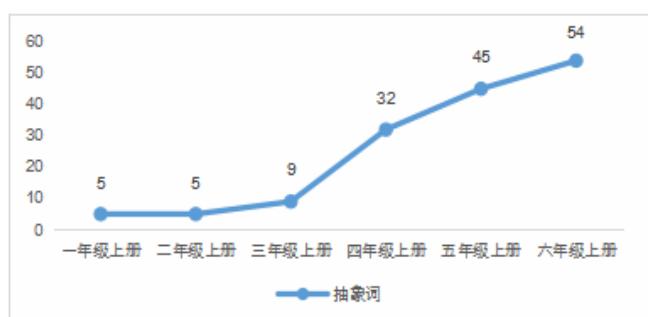
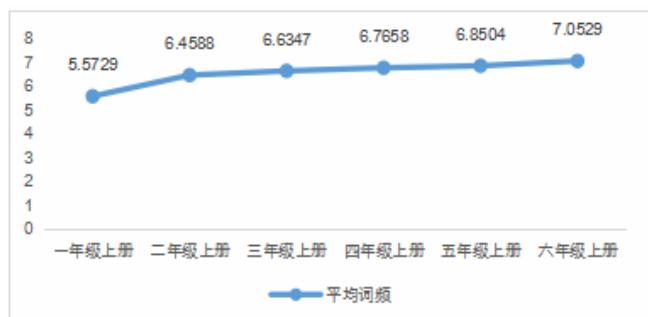
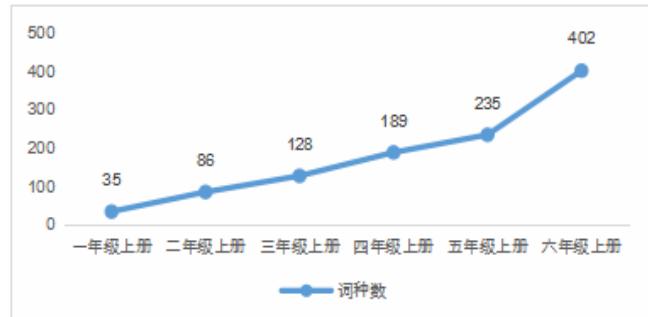
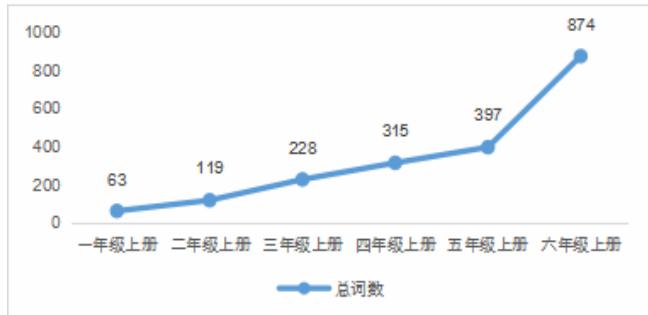


图3 小学1-6年级上册词汇影响因素

从图3可以看到,对于总词数来说,使用总词数最少的为小学一年级,平均总共使用了63个词,总词数最多的为六年级,平均使用了874个词;词种数方面小学一年级使用得最少,平均为35词,词种数最多的为六年级,平均使用了402个词;平均词频方面,小学一年级最低,平均为5.5729,六年级最高,平均为7.0529;抽象词方面,一年级最少,平均为5个词,六年级最多,平均为54词。

同样为了验证得出来的规律是否准确,笔者同时对小学1-6年级下册也各抽取3篇文章,进行统计,统计结果如图4所示:

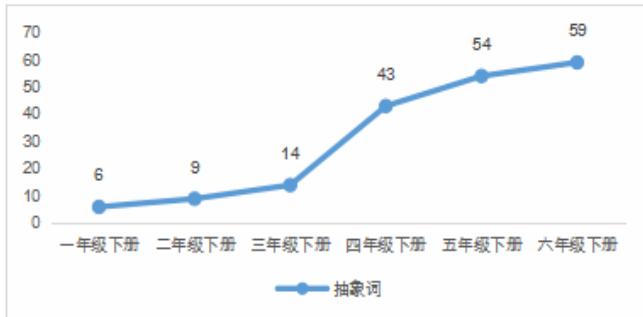


图4 小学1-6年级下册词汇影响因素

从图4可以看到,对于总词数来说,平均使用总词数最少的小学一年级,总共使用了102个词,平均总词数最多的为六年级,使用了623个词;词种数方面小学一年级最少,平均为69词,词种数最多的为六年级,平均使用了320个词;平均词频方面,小学一年级最低,平均为5.5611,六年级最高,平均为6.6848;抽象词方面,一年级最少,平均为6个词,六年级最多,平均为59词。

对以上6个因素统计完之后,可以发现,从以上6个因素变化趋势来看,呈现线性变化趋势,即随着年级上升,总词数、词种数、平均词频、抽象词、连词都逐渐上升。但同样要跟字种数一样注意的是,词种数虽然也在逐渐上升,但在一篇文本中所占

的比例在逐渐下降，因此，也能够说明，一篇文本用词的丰富程度，对于高年级来说，不再是需要学生学习的重点，反而将学习重点放在词的难度上面。

2.3 句子

句子是语言最基本的表达单位，与文本可读性息息相关。在前面字、词的研究当中，都是对文本进行量化指标方面的统计。一篇文本中句子数、平均句长固然会影响文本的可读性，但更重要的因素是文本中句子结构的复杂度，这些因素主要依靠定性分析。

2.3.1 句子数

句子数，即文本中包含的独立句子的总数。一般认为，一篇文本的句子数量庞大，那么这篇文本的阅读难度就比较大。此外，句子数过多，则会使文本流于碎片化，信息之间逻辑联系被弱化，显得跳跃、不连贯，如同“思维断片”，会增加理解与记忆的负担。

2.3.2 平均句长

平均句长是指一篇文本中句子的平均长度。句子越长，说明学生在阅读时，需要处理的信息就越多，认知负荷就越大；反之，句子越短，句子包含的信息就越少，认知负荷就越低。以下是两个句子，一个短句，一个长句。

a. 一只乌鸦口渴了，到处找水喝。

b. 深蓝的天空中挂着一轮金黄的圆月，下面是海边的沙地，都种着一望无际的碧绿的西瓜，其间有一个十一二岁的少年，项戴银圈，手捏一柄钢叉，向一匹猹尽力地刺去，那猹却将身一扭，反从他的胯下逃走了。

第一个句子12个字，学生只需简短的时间就能获得乌鸦口渴和找水两个信息；第二个句子有83个字，学生需要先获取环境的信息，再获取人物特征的信息，最后需要获取人物动作的信息。因此，一篇文本中，平均句子越长，说明文本的信息越丰富。

2.3.3 复句数

复句是指一个句子中所包含的、由两个或两个以上主谓结构构成的独立分句的数量。这些分句并非简单堆砌，而是通过特定的关联词语（如“因为…所以…”“虽然…但是…”“如果…那么…”）或内在的意合关系紧密联结，形成一个意义完整、逻辑严谨的语言单位。因此，一篇文本中复句数的多少影响到文本的可读性。

2.3.4 偏正复句

在复句中，常见的逻辑关系有并列、顺承、因果、转折、条件等。其中因果、转折、条件等关系则构成了偏正复句，其特点是分句间存在明确的“主从”或“偏正”关系——一个分句（正句）是语义核心，另一个分句（偏句）则用来说明原因、提出假设或承认某种让步情况，起到辅助和修饰的作用。

2.3.5 复杂长句

句子结构的复杂性主要体现为多层逻辑嵌套，如一个句子中包含因果关系，因果关系中又包含并列关系，修饰语过多或过长，如句子主干被大量的定语、状语、插入语所包裹，这两个原因使

得学生没办法一下子捕捉到关键信息，增大了阅读的难度。因此，一篇文本中，如果复杂长句的多少会影响到文本的可读性。

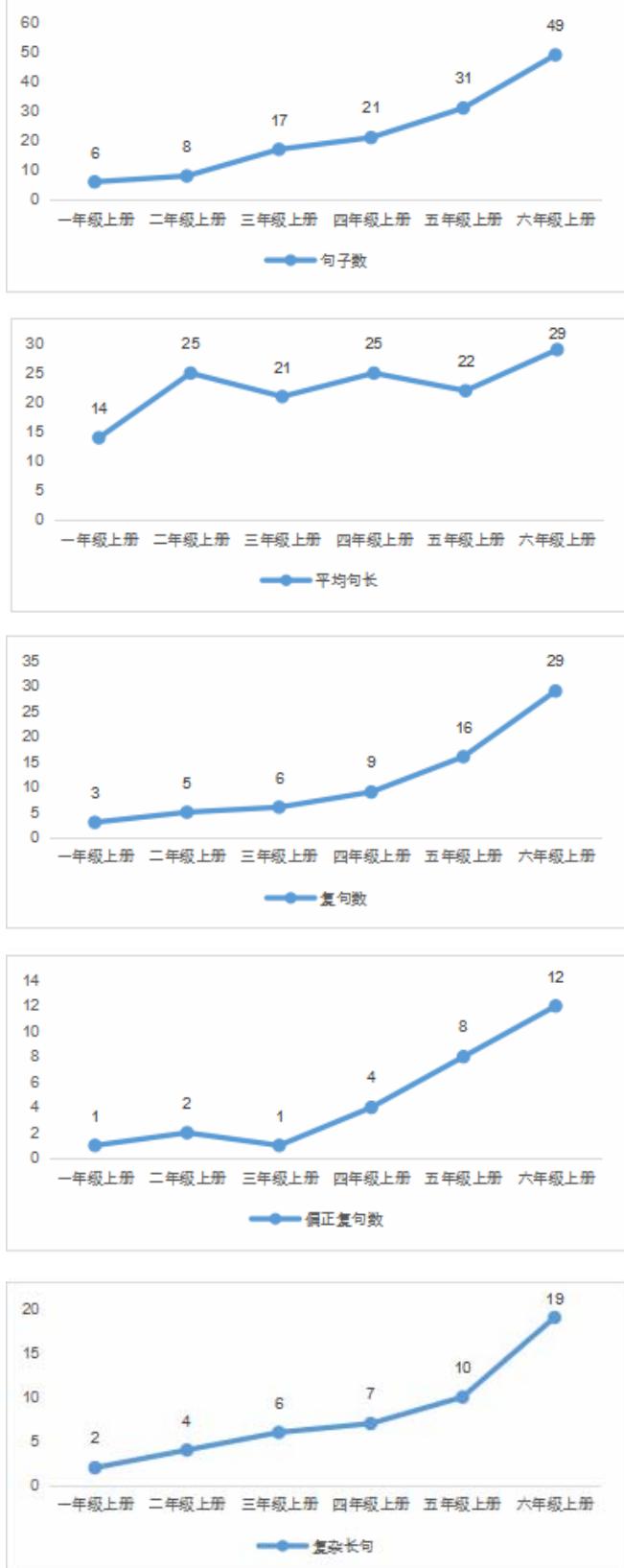


图5 小学1-6年级上册句子影响因素

在此基础之上,我统计了以上5个因素随着课文年级递升的变化关系。最终得到的结果如图5所示:

从图5可以看到,平均句子数最少的是一年级,有6个句子,平均句子数最多的是六年级,有49个句子;平均句长方面,一年级最少,平均句长14个字,六年级最多,平均句长29个字;复句数一年级最少,平均有3个,六年级最多,平均有29个;偏正复句方面,一年级最少,平均有1个,六年级最多,平均有12个;复杂长句方面,一年级最少,平均有2个,六年级最多,平均有19个。

同样为了验证得出来的结果是否准确,笔者同时对小学1-6年级下册也各抽取3篇文章,进行统计,统计结果如图6所示。

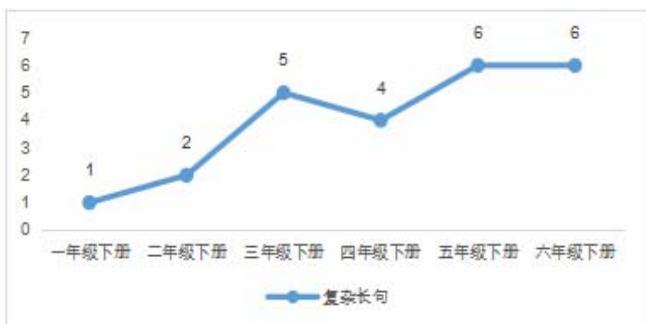
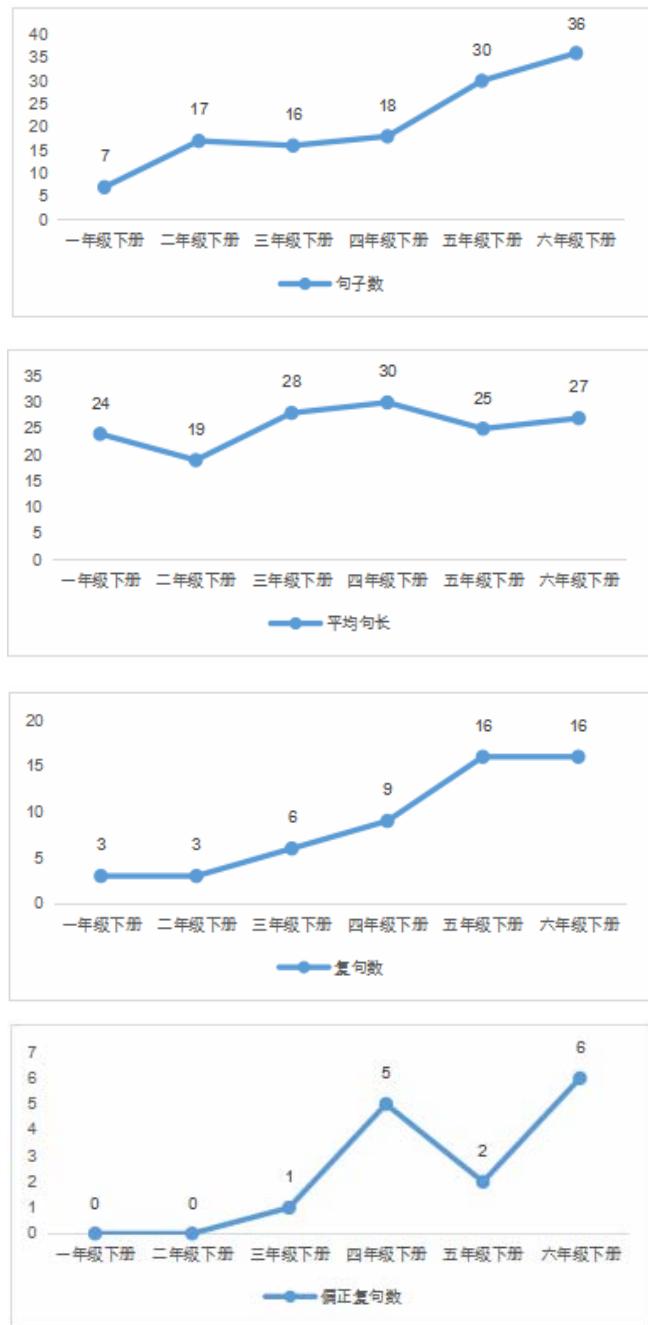


图6 小学1-6年级上册句子影响因素分析

从图6可以看到,平均句子数最少的是一年级,有7个句子,句子数最多的是六年级,平均有36个句子;平均句长方面,二年级最少,平均句长19个字,四年级最多,平均句长30个字;复句数一年级最少,平均有3个,六年级最多,平均有16个;偏正复句方面,一年级最少,平均有0个,六年级最多,平均有6个;复杂长句方面,一年级最少,平均有1个,五、六年级最多,平均有6个。

通过对上述数据进行系统性分析,可以观察到四个主要影响因素总体上均呈现出明确的线性上升趋势,这符合学生语言能力发展的基本规律。然而,“平均句长”这一指标的变化轨迹则更为复杂,呈现出一种“螺旋式上升”的态势。具体表现为:六年级的平均句长不再像其他影响因素一样居于首位,一年级的也并非始终为最低。

但这一现象并不能否定“平均句长”作为衡量文本可读性核心指标的有效性。其数据的波动,很大程度上源于当前研究方法的局限性:研究者将每个年级三篇文章的平均句长简单相加后取算术平均值,这种方法对极端数值(即某一篇文本平均句长显著过长或过短)异常敏感,容易导致整体均值的“失真”。

3 总结

通过对国内外相关文献进行系统的梳理与评估,可以确认,我们所构建的这套可读性指标体系在整体框架上适用于对小学语文教材的分析与评价。然而,深入的分析过程揭示了一个关键问题:在高年级课文中,个别指标(如平均句长、词汇熟悉度)的数值偶尔会出现“不升反降”的现象,即低于某些低年级课文。这绝非意味着高年级课文的可读性更高、语言难度更低;恰恰相反,这一“反常”现象是对简单化、线性可读性评估模型的重要警示。它深刻地揭示出,文本的复杂度是多维的。一篇文章可能词简单,但是句子蕴含的逻辑就难。因此,可读性评估绝不能依赖于任何单一指标的孤立判断,而必须立足于多维度指标的综合考量与加权分析,以全局的、系统的视角审视文本,才能得出符合教学实际与认知规律的科学结论。此外,深入分析各项指标的具体表现,能够为教学实践提供清晰的指引。这种诊断式的分析,为实现精准教学、有效突破学习难点奠定了基础。

本文的研究重点主要是分析影响文本可读性的关键因素,但在对课文分析的过程中发现,部分指标存在对低年级重要,但在高年级时已不是教学和学习的重要,因此,如果后续研究者综

合衡量文本可读性,应考虑到不同指标在各个年级的权重不同,不能单一地用一个权重去衡量所有年级的文章。

[参考文献]

- [1] Klare G R.The measurement of readability:useful information for communicators[J].ACM Journal of Computer Documentation(JCD),2000,24(3):107–121.
- [2] Harris T L,Hodges R E.The literacy dictionary:The vocabulary of reading and writing[M].Order Department, International Reading Association,Newark,1995.
- [3] Fry,E.Fry' s Readability Graph:Clarifications,Validity,

and Extension to Level 17[J].Journal of Reading,1977,38:243–250.

[4] 孙汉银.中文易懂性公式[D].北京师范大学,1992.

[5] 李绍山.易读性研究概述[J].解放军外国语学院学报,2000(04):1—5.

[6] 吕中舌.可读性理论与英语教材[M].北京:世界知识出版社,2003.

作者简介:

韩景(2001--),女,汉族,河北省廊坊市人,硕士研究生,单位:福建师范大学教育学院,研究方向:课程与教学论。