

文章编号: 1003-0077(2026)03-0041-12

大模型文本简化的篇章结构变化分析与评测

朱君辉^{1,2}, 杨尔弘^{1,2}, 朱浩楠^{1,2}, 杨麟儿^{1,2}

(1. 北京语言大学 国家语言资源监测与研究平面媒体中心, 北京 100083;

2. 北京语言大学 信息科学学院, 北京 100083)

摘要: 该文基于修辞结构理论,从篇章层面对大语言模型(GPT-4o)在中文新闻文本简化任务中的表现进行了系统评估。研究以篇章树与语义关系描述篇章结构,在新闻文本上构建了篇章结构标注数据集,探索性地提出了九项用以表征文本简化性质的篇章结构量化指标,对比、分析了大模型与专家简化文本在篇章结构上的差异。研究表明:大模型简化文本的篇章树相对扁平。进一步分析显示,大模型在删除、合并操作中易导致篇章结构层面的简化过度,出现核心信息遗漏与照应失衡的现象。该文提出了引入篇章结构信息的评测与优化思路,并探讨了其在模型结构约束与自我校正中的应用潜力。

关键词: 文本简化;修辞结构理论;大语言模型;差异研究

中图分类号: TP391

文献标识码: A

Discourse Structure Variation for Large Language Model Evaluation in Text Simplification

ZHU Junhui^{1,2}, YANG Erhong^{1,2}, ZHU Haonan^{1,2}, YANG Lin'er^{1,2}

(1. National Language Resource Monitoring and Research Center Print Media Language Branch,
Beijing Language and Culture University, Beijing 100083, China;

2. School of Information Science, Beijing Language and Culture University, Beijing 100083, China)

Abstract: Based on Rhetorical Structure Theory (RST), this study evaluate the discourse-level performance of the large language model GPT-4o in Chinese news text simplification. Discourse structures are represented through rhetorical trees and semantic relations. A discourse-annotated dataset of news texts is constructed, and nine quantitative indicators are proposed to describe the properties of text simplification at the discourse level. The comparison between expert simplifications and model outputs shows that GPT-4o produced relatively flat discourse trees. Further analysis revealed that deletion and merging operations by the model often caused excessive simplification, leading to information loss and imbalance in referential coherence. The study proposes the incorporation of discourse structural information into text simplification evaluation and optimization, and highlights its potential for model constraint and self-correction mechanisms.

Keywords: text simplification; Rhetorical Structure Theory (RST); Large Language Models; comparative analysis

0 引言

阅读是信息获取的重要方式。在全民阅读素养提升与信息普惠战略推进的背景下,文本可及性(text accessibility)逐渐成为语言资源建设与语言

智能应用的重要议题^[1]。由于受教育程度、阅读能力和信息渠道差异的影响,部分读者在阅读复杂文本时仍面临障碍,尤以新闻、科普、公文类文本为甚。如何使不同阅读能力群体都能平等获取信息,成为当前自然语言处理(Natural Language Processing, NLP)和语言服务技术共同关注的问题。

收稿日期: 2025-10-20 定稿日期: 2025-12-31

基金项目: 国家语委重大科研项目(ZDA145-17);教育部人文社科青年基金(23YJCZH264);北京语言大学优秀博士学位论文培育计划项目;北京语言大学研究生创新基金(25YCX125)

文本简化(Text Simplification)作为一种重要的自然语言生成任务,旨在通过词汇替换、句法调整或篇章重组等方式,生成语义等价但更易理解的文本^[2-4]。它不仅能够辅助语言学习者、低识字群体和老年人群体提高理解效率,也可作为机器翻译、摘要生成等任务的数据增强手段。早期研究多采用基于规则和句法转换的方法,近年来则发展出神经网络驱动的端到端简化模型。然而,现有自动简化研究主要集中于词汇和句法层面,对篇章层面的结构调整与连贯性关注不足^[5-7]。

这一局限的根源在于,以往的简化模型高度依赖大规模训练语料,而篇章层面的标注资源稀缺,使模型既缺乏训练样本,也缺乏结构化的评测依据。篇章简化的目标不仅是降低语言复杂度,更在于优化信息组织与语义衔接,使文本在逻辑上更加清晰、条理更加顺畅。已有研究表明^[8-10],人类专家在简化文本时,常通过删减冗余细节、补充背景信息、强化因果逻辑等策略,重塑语篇结构以增强可读性。这类操作依赖于对篇章语义关系的整体掌握,而传统自动简化系统难以在缺乏篇章标注资源的条件下实现类似的结构调控。

近年来,以大语言模型(Large Language Models, LLMs)为代表的生成式人工智能技术展现出巨大潜力,为文本简化研究提供了新的契机^[11-13]。与以往依赖大规模标注数据的模型不同,LLMs 具备强大的语言生成与语义重组能力,为篇章层面的自动简化研究提供了新的可能。研究者开始重新关注篇章简化这一长期受限于数据资源的问题:一方面,大模型的生成结果可以揭示其在篇章理解与组织方面的潜在能力;另一方面,小规模、高质量的篇章结构标注语料可以反过来作为“结构引导”资源,辅助模型在生成阶段学习篇章逻辑与语义关系。

目前,对大模型篇章简化能力的研究仍以人工主观评分为主^[12-13],依赖语感判断“连贯”与否,缺乏可量化、可复现的评测体系。要揭示大模型在篇章结构处理中的具体问题,就必须建立结构化的量化指标与语料资源。修辞结构理论(Rhetorical Structure Theory, RST)提供了一种有效的篇章分析框架,可通过篇章树形结构描述语篇内部的逻辑与语义关系,在计算语言学中已被广泛应用于摘要生成、篇章连贯性评估和文本质量分析。

针对上述问题,本文从篇章结构视角出发,构建了覆盖原文、专家简化与大模型简化三类版本的汉

语平行篇章简化树库,并基于修辞结构理论提出九项篇章结构量化指标,以系统评测大模型在简化过程中的结构变化与语义关系变动。本文通过对比分析大模型与专家在简化中的篇章结构差异,揭示大模型在结构保留、关系调控与语篇连贯性方面的表现特征,并进一步探讨篇章结构指标在大模型性能评测与优化中的应用价值。研究结果显示,大模型生成的简化文本在整体结构上趋于扁平化,虽能有效删减冗余信息,但在处理复杂语义关系与维系逻辑连贯性方面与专家仍存在显著差距。

本文的研究目标在于建立一套基于修辞结构的篇章简化效果评测框架,实现对大模型篇章级生成质量的客观量化评估,为后续自动化评测体系的构建与模型微调提供依据。本文的主要贡献包括:

(1) 构建了首个汉语平行篇章简化树库,为篇章层面简化研究提供数据基础。

(2) 提出九项篇章结构指标,量化简化过程中的结构变化特征。

(3) 揭示了大模型在篇章简化中的结构性偏差及潜在改进方向。

(4) 探索了篇章结构信息在自动简化评测与模型优化中的应用潜力。

1 数据与篇章结构标注

1.1 语料来源与构成

为系统分析大语言模型在篇章结构简化中的表现,本文构建了覆盖原文、专家简化文本和大模型简化文本三种版本的汉语新闻平行语料。基础数据来自 Chinese News Simplification Data(CNS)^[13],该数据集由汉语报刊文章原文及简化版组成,涵盖政治、文化和经济等各个主题。CNS 数据集包含 100 篇平行新闻文本,其中原文 50 篇、专家简化文本 50 篇。专家简化的文本来源于吴成年为初级汉语水平的学习者编写的《新编读报纸学中文》系列教材,主要从篇幅控制与词汇难度两个方面对原文进行了简化。

在此基础上,为获得第三类简化版本,本文使用 GPT-4o 模型生成大模型简化文本。鉴于不同模型在篇章简化中的输出差异较大,本文通过应用程序编程接口(API)进行文本生成,以确保简化过程的可控性与可复现性。提示词(Prompt)仅限定任务目标而不过多干预生成策略,以避免人为偏向。具体提示词为:

Q: 请你将提供的复杂文本替换成初级汉语水平的学习者可以理解的简单文本。你要保持原文的意义与趣味性不变,但使文本可读性更高。###
示例:复杂文本:{ },简单文本:{ }

请根据以上示例完成下面这篇文本的简化,下面是需要你改写的复杂文本:{ }

实验通过应用程序编程接口(API)请求的方式发送实验参数和提示词,在生成过程中,模型采用如下参数设置:Temperature 设为 0.7, Top-*p* 设为 1.0,最大生成长度(max tokens)设为 2 048。除上述参数外,其余参数均保持为 API 官方文档所推荐的默认设置。生成后文本经人工审阅,确保无遗漏或异常重复。最终得到 150 篇新闻文本(原文 50 篇、专家简化 50 篇、大模型简化 50 篇)。

新闻体裁具有结构稳定、语义层次清晰的特点,约 50% 的新闻故事呈现程式化结构^[14],因此适合进行篇章结构标注与对比分析。篇章简化数据集的语料基本信息如表 1 所示。该数据集既能反映不同简化策略下的结构差异,也为大模型篇章结构评测提供了真实数据基础。

表 1 汉语新闻篇章简化数据集基本信息

语料	篇数	字数	词数	句数	段落数
新闻原文	50	59 238	37 764	1340	670
专家简化文本	50	34 066	21 835	849	445
大模型简化文本	50	27 895	17 375	868	462

1.2 句子对齐与篇章单元划分

为实现不同版本文本的结构可比性,本文首先对三类文本进行句级对齐与篇章单元划分。

句子对齐采用“语义一致性+编辑距离”双重判定策略:先基于语义匹配规则进行初步配对,再结合编辑距离与人工审校进行修正,确保语义对应的精确性。

在篇章层级,篇章单元划分以句子为最小分析单位,即每个句子被视为一个独立的最小语篇单元(Elementary Discourse Unit, EDU)。这一设定旨在聚焦句级简化对整体篇章结构的影响,避免在句内再行拆分引入不必要的复杂性。划分过程中遵循以下原则:

(1) 每个句子作为一个完整的语义单元独立标注。

(2) 对应句间若出现拆分或合并,则在对齐层

记录结构变化。

(3) 引述语、插入语等保留原句边界,不单独成段。

划分完成后,对原文与简化文本进行结构比对,确保相同成分节点层次间的一致性,为后续语义关系标注提供对齐基础。

1.3 修辞结构标注体系

篇章语义关系的识别与标注遵循修辞结构理论^[15]的核心思想。RST 假设语篇可由一系列语义关系构成的树形结构描述,每个节点由“核(Nucleus)”和“卫星(Satellite)”组成,两者通过特定的语义关系连接,如“因果”“条件”“对比”“背景”等。RST 的语义关系集是一个开放的系统,研究者可根据语料特征扩展或合并关系类型。本文结合对上述语料的篇章分析,对经典 RST 语义关系集做了相应调整与细化,最终得出 34 种语义关系,如表 2 所示。标注单元为句级 EDU,每个 EDU 根据其在篇章树中的层级关系被赋予对应的核-卫星标签。关系判定标准以语义功能为依据:若某单元对整体意义起主要作用,则标注为“核”;若起补充、解释或限定作用,则标注为“卫星”。

表 2 标注体系采用的语义关系

关系类型	语义关系
背景	背景关系/Background、环境关系/Circumstance、来源关系 */Source
条件	条件关系/Condition、假设关系/Hypothesis
因果	原因关系 */Cause、结果关系 */Result、后果关系 */Consequence、动机关系/Motivation、目的关系/Purpose
评论和解释	评论关系/Evaluation、解说关系/Elaboration、解释关系/Interpretation、解答关系/Solution、引论关系、评论-多核 */Comment-mn
论证	详述关系/Elaboration、详述-多核 */Elaboration-mn、证据关系/Evidence、证明关系/Justify、让步关系/Concession、举例关系 */Example、方式关系/Means
重述和总结	重述关系/Restatement、总结关系/Summary、总结关系 */Overview
递进	递进关系 */Progression、递进-多核 */Progression-mn
比较	类比关系 */Analogy、对照关系/Antithesis、对比关系/Contrast
并列	补充关系 */supplement、顺承关系/Sequence、并列关系/Joint

注:“*”号标记为本文新增的语义关系。

1.4 标注流程与质量控制

为了增强语料标注的可信度,本文邀请两位具有语言学与计算语言学背景的研究者共同进行篇章结构标注。标注过程包括以下三个阶段。

(1) 独立标注阶段:两位标注者分别完成 150 篇文本的初次标注。

(2) 一致性比对阶段:对存在差异的标注进行交叉检查与讨论,修订歧义部分。

(3) 终稿确认阶段:通过反复协商确定最终的篇章树结构。

标注工作使用 MickO'Donnell 设计的修辞结构理论分析软件 RST TOOL 3.42^① 完成,该工具支持树形结构可视化、关系自动检查及一致性比对功能。为确保标注一致性,对每篇文本的主要节点与关系进行 Cohen's Kappa 检验,平均一致性系数达到 0.82,表明标注质量较高。

1.5 树库概况与特征分布

本研究构建的篇章结构树库共包含 150 篇新闻文本,覆盖政治、经济、文化等多个主题,包括原文、专家简化文本与大模型简化文本三类版本。所

有文本均以句子为最小分析单位,经过 RST 标注,生成.rs3 格式的篇章树文件,并附带“核-卫星”关系与句级对齐信息,可为后续的篇章结构量化分析提供结构化数据基础。

从规模统计看,语料整体平均篇幅约 20 句,每篇平均词数约 513 词,整体包含 6 250 个句级节点和 3 162 条修辞关系标注。篇章树的平均深度约为 4.8 层。语料在篇章层次与结构复杂度上呈良好的均衡性,能够反映新闻文本的典型组织特征。

从语义关系分布看,原文共出现 34 种语义关系,总频数 1 465 次,专家简化版中共出现 33 种语义关系,总频数 931 次,GPT-4o 简化版中共出现 32 种语义关系,总频数 794 次。并列、顺承、详述、评价、总括是出现频率最高的关系类型,占全部关系实例的约 49%(见图 1,只列出前 21 种)。此外,还涵盖方式、类比、目的、重述等多种关系类型。

总体来看,该树库在篇章深度、节点规模及语义关系类型上分布合理、覆盖充分,能够较为全面地反映汉语新闻文本的篇章结构特征。后续章节将在此基础上,基于九项篇章结构量化指标,对不同简化版本文本的结构复杂度与语义连贯性进行系统评测。

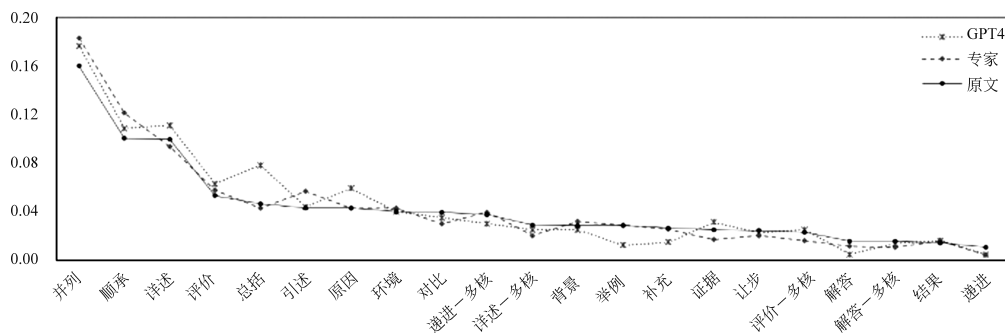


图 1 三种语料的修辞结构关系分布

2 篇章结构量化指标设计

2.1 指标设计的理论基础

Van Dijk 和 Kintsch^[16] 的篇章理解模型指出,当篇章呈现出更深的嵌套关系时,篇章结构的复杂性增加,理解的难度也随之加大。在 RST 篇章树中,句子的“深度”指句子到根节点的距离,代表着它在树结构中的嵌套程度,较浅的深度(靠近树根)通常表示该句子更“重要”或更突出,意味着它是文本主旨或主要内容的一部分^[17]。篇章树越深,嵌套关

系越复杂,篇章理解的难度越大。

此外,RST 的核心理念认为,文章的每一部分都是相互关联的,之间的语义关系分为“核-卫星”关系(Nuclei-Satellite)和多极核心关系(Multi-nuclei)两大类。其中,核心关系承载文本的主要信息,辅助关系则是对核心信息进行补充、修饰的次要信息。“核-卫星”关系又称“单核关系”,例如,详述、证据关系,某个句子充当核心信息单元,另一个句子则

① 标注工具网址为: <http://www.wagsoft.com/RSTTool/>。

在标注过程中,如涉及一句话以上的整个段落乃至更大单位的删除,仅计为一次编辑并在后面标记为“*”段落,提取篇章语义关系时也仅提取整个段落承担的关系。

作为卫星,提供补充细节或对比信息,以增强核心句子的阐释力。多极核心关系也称为“多核关系”,例如并列、顺承关系,连接的各部分重要程度相等。

2.2 指标体系的构建

在已有理论研究的基础上,为了更好地量化文本在不同篇章层级的结构特征,本文从篇章树与语义关系两个维度设计了相应的 9 个测量指标。

2.2.1 结构层指标

通过计算最大篇章树深度、平均篇章树深度、篇章树深度方差、浅层句子比率和平均每层级句数 5 个指标衡量篇章结构的复杂性。每个叶节点对应一个句子,较高层次的节点将这些较低层次的表述整合成更复杂的结构,如从句子到段落再到篇章,构成了完整的篇章树^[18-19]。最大篇章树深度表示篇章中所有句子到根节点的距离中最大的距离,平均深度表示篇章树中所有句子到根节点的平均距离。较小的平均深度通常表示较为简洁的逻辑结构,而较大的平均深度表明文本的逻辑嵌套复杂性较高。篇

章树方差计算句子在篇章树中的深度分布情况,较大的方差表示句子层次分布的差异较大,可能导致理解难度增加;反之则意味着句子分布均匀,有助于理解。

浅层句子比率计算深度较浅的句子所占比例,本文将其界定为处于树的前四层,以此衡量篇章中具有关键信息的句子的集中度。平均每个层级句数计算篇章树中每个层级包含的平均句子数量,衡量每个层级的复杂度。平均每个层级句数越多,文本内容越丰富。

2.2.2 语义关系层指标

通过计算核心关系句数、卫星关系句数、关系多样性及逻辑连贯度 4 个指标衡量语义关系的复杂性。关系多样性的计算借鉴词汇多样性,衡量语义关系类型的丰富程度。语义关系的多样性越高,表达方式越多样化,读者的认知负荷越重。核关系占比较高,通常表明文本结构更合理,逻辑性和连贯性更强。

综上所述,篇章树与语义关系共形成 9 个篇章结构量化指标,计算方法如表 3 所示。

表 3 篇章结构量化指标

指标	描述	公式
Largest Depth	最大篇章树深度	$Depth_{\max}$
Mean Depth	平均篇章树深度	$\sum_{i=1}^n Depth(i)/n$
Depth Variance	篇章树深度方差	$\sum_{i=1}^n (Depth(i) - \text{Mean Depth})^2 / n$
Salience Sentence Ratio	浅层句子比率	$\text{if}(\sum_{i=1}^n N_{\text{Sentences}}(Depth \leq 4)/n, 1, 0)$
Average Sentences per Level	平均每层级句数	$\sum_{i=1}^n N_{\text{Sentences}} / N_{\text{Levels}}$
Nucleus Sentences num	核心关系句数	$N_{\text{NucleusRelations}}$
Satellite Sentences num	卫星关系句数	$N_{\text{SatelliteRelations}}$
Relation Diversity	关系多样性	$N_{\text{Type of Relations}} / \sqrt{N_{\text{Relations}}}$
Logical Coherence	逻辑连贯度	$N_{\text{NucleusRelations}} / N_{\text{SatelliteRelations}}$

3 大模型篇章简化评测与结果分析

3.1 评测方法

本文旨在从篇章层面对大语言模型(LLM)在新闻文本简化任务中的表现进行系统评测,重点考察其在结构层次与语义组织方面的简化效果。以人类专家简化文本为基线、原文为参考,比较 GPT-4o 简化文本在篇章结构复杂度、关系多样性与逻辑连

贯性等维度的表现差异,揭示其在篇章层面简化能力的优劣。具体而言,评测包括两部分:

第一部分为结构层指标评测,基于第三节提出的九项篇章结构量化指标,计算原文、专家简化文本与 GPT-4o 简化文本的结构特征。各指标分别从篇章树结构维度(最大深度、平均深度、深度方差、浅层句比率、平均每层句数)和语义关系维度(核心句数、卫星句数、关系多样性、逻辑连贯度)刻画文本的篇章复杂度与逻辑组织特征。数据经标准化处理后,

采用 Python 进行统计分析。对每个指标进行正态性与方差齐性检验;若不符合正态分布,则使用 Kruskal-Wallis H 检验。当显著性水平 $p \leq 0.05$ 时,视为差异显著。若检验结果显著,则进一步进行事后比较 (Dunn's Post Hoc Test),以确定差异的来源。

第二部分为编辑行为分析,在句级对齐的基础上,将简化操作划分为删除、合并和插入三类,统计 GPT-4o 与专家在不同语义关系类型上的操作频率及结果。此部分分析旨在揭示大模型在简化过程中对篇章结构的调整策略及其潜在偏差。

通过上述两步评测,本研究能够定量与定性地

评估大模型在篇章结构简化上的表现,为其在自动简化系统中的改进提供实证依据。

3.2 量化指标评测

经检验,除关系多样性外,其余指标均不符合正态分布,但满足方差齐性条件,因此采用非参数检验 (Kruskal-Wallis H) 进行总体差异分析;关系多样性指标则使用单因素方差分析 (ANOVA) 验证组间差异的显著性。结果如表 4 所示。除逻辑连贯度外,其余八项指标在三组语料间均存在显著差异 ($p \leq 0.05$),说明简化过程确实引起了篇章结构层级与语义关系的系统性变化。

表 4 三种语料可理解性指标差异性检验结果

	Mean			p value	H-statistic	Effect size (r)	post-hoc test
	原文	专家	GPT-4o				
最大结构树深度	10.60	9.08	7.38	0.00	34.86	0.49***	1>2>3
平均结构树深度	5.52	4.75	4.01	0.00	37.1	0.51***	1>2>3
结构树深度方差	6.63	5.58	4.07	0.00	21.44	0.39***	1>3, 2>3
浅层句子比率	0.26	0.33	0.43	0.00	39.25	0.52***	1>2>3
平均每层级句数	4.72	3.64	3.54	0.00	24.44	0.41***	1>3, 2>3
卫星关系句数	11.23	6.77	5.22	0.00	42.1	0.54***	1>3, 2>3
核心关系句数	18.79	12.40	10.80	0.00	27.53	0.44***	1>3, 2>3
逻辑连贯度	0.64	0.60	0.54	0.06	5.36	0.19***	Null
				p value	F-value	Partial Eta Squared	post-hoc test
关系多样性	2.70	2.65	2.28	0.00	13.69	0.16***	1>2, 1>3

在篇章树结构维度上,大模型简化文本整体呈现出更明显的“扁平化”趋势。最大结构树深度从原文的 10.60 下降至专家简化文本的 9.08,在大模型简化文本中进一步降至 7.38;平均深度与深度方差亦呈同向变化。这表明无论是人工简化还是自动简化,都会减少篇章层级嵌套,但大模型的削减更为剧烈。与此同时,浅层句比率显著上升(从原文的 26%提高至大模型的 43%),反映出篇章主旨句集中于浅层节点,层次结构趋于单层化。

在语义关系维度上,核心关系句数与卫星关系句数差异显著。在专家简化文本中,核心句占比为 65%;大模型简化文本中,核心句比例上升至 67%,卫星句比例降低。

这一变化意味着大模型在篇章压缩时更倾向于压缩主干信息并弱化修饰与铺垫成分。关系多样性指标显示,大模型简化文本的语义关系类型减少,从

原文的 2.70 降至 2.28,表明语义组织的丰富度下降;逻辑连贯度指标则在三组文本间差异不显著 ($p > 0.05$),说明简化并未破坏整体逻辑的合理性,但篇章的语义关联形式趋于单一。

从篇章理解角度看,文本简化本身即以降低认知负担、突出关键信息为目标,其过程不可避免地伴随着对冗余信息和枝节内容的压缩或删除,较浅的树深与较高的浅层句比例有助于降低读者在篇章层次上的加工负担,提高信息获取效率。但过度扁平化可能削弱篇章的逻辑层级和语义连贯性,使文本缺乏组织弹性。相比之下,专家简化的篇章在简化幅度与结构完整性之间更平衡,大模型在保持语义多样性与层次连贯性方面仍存在不足。这一结果为后续章节关于编辑行为与机制差异的深入分析提供了数据依据。

3.3 编辑行为分析

为进一步评估大语言模型在篇章层面简化任务中的表现,本文从编辑行为角度考察 GPT-4o 在处理篇章结构时的操作模式及其与人类专家的差异。通过对齐原文、专家简化文本与 GPT-4o 输出文本,识别简化过程中的三类编辑行为:删除(Deletion)、合并(Merging)和插入(Insertion)。句子为最小分析单位,并记录每次编辑所涉及的语义关系类型及其在篇章树中的位置,以评估 GPT-4o 在保持信息完整性与语篇连贯性方面的表现。

在提取篇章语义关系时,若句子作为卫星,则直接提取这句话承担的语义关系,如“背景”;若句子为关系的核心,则标注其支配的语义关系并注明,如“背景~核心”^②。分别统计专家与 GPT-4o 简化过程中的编辑类型,结果如表 5 所示。

表 5 专家与 GPT-4o 简化过程中的编辑类型统计

编辑类型\简化	专家	模型
删除	210	316
合并	28	78
插入	22	16

3.3.1 删除

删除(Deletion)是 GPT-4o 最主要的简化操作。统计结果显示,其删除的语义关系类型与专家总体一致,但范围更广。表 6 展示了大模型与专家删除的前五大关系的句子比例(下划线为删除的句子)。专家主要删除详述、举例与评论类句子,占比 35.6%,这些内容多为细节或说明;GPT-4o 的删除比例更高,除详述与背景外,还包括原因(6.0%)、引述(4.7%)、递进(4.1%)等,共计 32 种关系。

表 6 专家与大模型简化删除信息的类型(前五种)

类型	描述	示例	占比/%	
			专家	GPT-4o
详述或举例	举的例子,具体的做法,论证的过程	饮食上他常年坚持简单、按时、按量,食物种类大众化,没有特殊要求,从不大鱼大肉暴饮暴食。他早饭热量要够,蛋白质要足,通常是一个馒头,一个鸡蛋、一杯牛奶,其他食物也要保持类似水平的营养	21.2	25
评价	人物的主观态度或评价	要缓解这种结构性的“用工荒”,还须多管齐下。业内专家认为,加快产业转型升级、改变就业观念、提升服务业吸纳就业空间等制度改革成为破解难题的根本所在	14.4	6.68
背景	与主题相关的背景信息,人物的身份、年龄,故事发展的时间、前提	河北大学自 1965 年开始招收外国留学生,是中国最早开展对外汉语教学的高校之一。近年来,随着中国政府奖学金来华留学自主招生资格的落实,河北大学每年长短期留学生均达 300 余名	11.5	9.8
顺承	某个事件的中间、后续发展,一般与主线关联性较弱	“节目里,我完整地阐述了自己的想法,他们认真地从头听到尾,最后总算同意了。”临行前,贺亮专门教会了父母怎样上微博,收看他的动态消息,“看着我天天更新微博、上传照片,他们也就放心了。”	8.7	5.7
并列	连接在结构上平行、地位相同的元素	大家帮我评理,看我是不是应该果断把父母拉黑? [和朋友去 KTV 玩……][我半夜和朋友去吃宵夜……]	8.2	3.8

值得注意的是,专家的删除操作未出现连贯性受损。经统计,GPT-4o 删除的句子中,9.5%的句子因删除核心语义节点或照应信息而导致语篇衔接削弱,尤其在详述、背景、顺承与后果关系上较为明显,如表 7 所示。这表明 GPT-4o 在判断信息层级时缺乏区分核心与附属关系的能力。

表 7 GPT-4o 简化删除关系类型及错删之处统计
(单位: %)

关系类型	总占比	错删占比
背景	25.0	3.2
评价	9.8	0.9
原因	7.0	—

续表

关系类型	总占比	错删占比
顺承	6.0	0.6
引述	5.7	0.9
递进	4.7	—
详述或举例	4.1	0.6
并列	3.8	0.3
环境	3.5	—
解答	3.5	—
结果	3.2	—
对比	3.2	0.6

续表		
关系类型	总占比	错删占比
让步	3.2	0.3
补充	2.9	—
对照	2.2	—
类比	1.9	—
证明	1.9	—
总结	1.9	—
动机	1.3	—
证据	1.3	—
后果	0.9	0.9

3.3.2 合并

合并(Merging)用于消除冗余并提高语义集中度。对合并的句子承担的语义关系统计结果如表 8 所示。专家的合并多出现在背景、详述或并列关系之间,体现出结构紧凑但逻辑清晰的特征;GPT-4o 则更频繁地执行合并操作,且跨度更大。其合并的句子中详述或举例关系占比最高,同时将顺承、原因、解答、让步和对比关系的句子并入主干语句。顺承关系的频繁合并意味着 GPT-4o 取消了时间顺序紧密连接的相邻句子的独立性,高度浓缩了事件的发展过程。对解答、让步、对比关系的合并表明,GPT-4o 不重视文本中提问、铺垫句的功能,倾向于主要信息先行。这种“过度合并”使信息密度提升,但削弱了时间与因果线索的显式表达,降低了语义层次的可辨性。与专家相比,GPT-4o 在压缩篇幅

的同时牺牲了部分篇章逻辑的层次感。

表 8 专家与大模型简化合并信息的类型

(单位: %)

关系类型	背景	详述或举例	顺承	并列	递进
专家	19.05	14.29	—	14.29	14.29
GPT-4o	6.94	21.84	11.11	8.33	—
关系类型	原因	解答	让步	对比	—
专家	9.52	—	—	—	—
GPT-4o	9.72	8.33	5.56	5.56	—

3.3.3 插入

插入(Insertion)是 GPT-4o 最少使用的编辑行为。在简化过程中,插入信息又被称为“详述简化(elaborative simplification)”,指通过重述、解释的方法进行简化的过程^[20-22]。插入信息的原因有多种,例如提供概念的定义、提供下文的概要,或提供某些作为主要内容先验知识的背景信息。这些编辑工作都有助于减少读者的认知负担,从而降低理解的难度^[23]。

专家简化文本中插入信息的比例为 8.5%,主要用于补充背景、总结与解释,从而增强连贯性,降低读者理解难度;GPT-4o 的插入比例仅 3.9%,且几乎全部为总结类关系(表 9)。这类插入更多体现为对原有信息的概括,而非提供新的解释或衔接信息,说明 GPT-4o 较难实现基于语义推理的“详述型简化”,对现实知识与上下文逻辑的补偿能力有限。

表 9 专家与大模型简化插入信息的类型

类型	描述	示例	占比/%	
			专家	GPT-4o
背景	提供理解文本某部分的先决条件的信息	可能还是现在的 80 后与以前的想法不一样吧,不想顺着老一辈的目光,到了什么年纪就做什么事,而是根据自己现阶段的情况去做理性的选择。→中国的传统观念是结了婚就得要个孩子,但我现在心理上还没准备好,要根据自己的现阶段的情况去做理性的选择	35.29	6.3
总结	通过概括和提炼复杂内容的核心信息,将信息简化为要点	喜爱户外徒步的人都知道,西藏的墨脱县是中国最难走的地方,尤其是反穿墨脱,更是难上加难。[因为……][贺亮遭遇了……][贺亮最终……]→喜欢旅行的人都知道,西藏的墨脱县是中国最难走的地方,但贺亮也一个人到那里走了一次	29.41	81.3
补充	提供新的信息,但不一定是理解要点所必需的	另据了解,《打工服务社》将于 6 月 10 日正式开播,每晚 9 点 08 分开门迎客	17.65	6.3
举例	提供示例,使概念更清晰	王晓初表示,“试点的城市必须足够多,不同的城市都要探索总结经验,再推出更适合的政策。→王晓初表示,“实验的城市必须足够多,考虑到不同特点,比如有的城市出生率并不低,但同样老龄化,因此不同的城市都要探索,要总结经验,再推出更适合的政策。”	11.76	—

续表				
类型	描述	示例	占比/%	
			专家	GPT-4o
解释	提供某个概念的定义或摘要	从 2010 年 10 月 1 日开始,上海展开城镇企业职工“柔性延迟领取养老金”尝试。→上海在延迟退休方面最先进行实验:从 2010 年 10 月 1 日开始,职工可以和用人单位协商延迟退休,即可最多延长五年退休,“男性一般不超过 65 周岁,女性一般不超过 60 周岁”	5.88	6.3

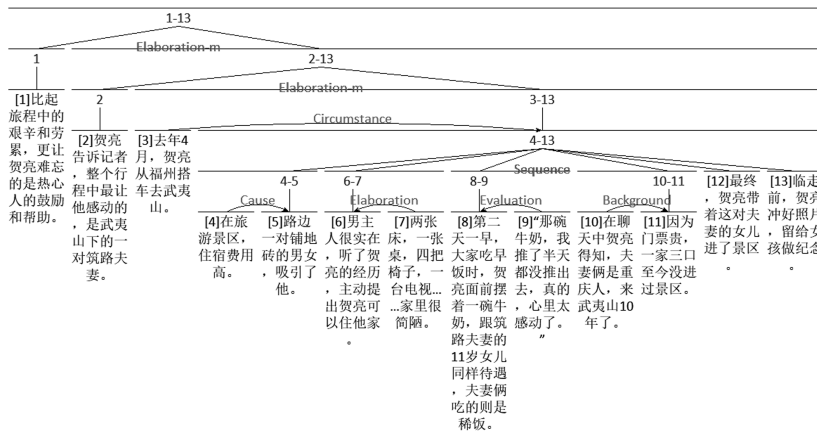
总体来看, GPT-4o 在篇章简化中展现出明显的“压缩化”策略:大量删除、宽幅合并、稀少插入。与人类专家相比,其能有效减少篇幅并保留主要信息,但在区分核心语义与辅助语义、保持逻辑递进和互动等方面存在不足。这些行为特征与前文的量化结果一致,反映出 GPT-4o 在篇章层级简化中仍缺乏对语义关系功能的细粒度控制。下一节将进一步从篇章连贯性与核心语义保留的角度,探讨其结构偏差的形成机制。

3.4 篇章连贯性偏差的成因分析

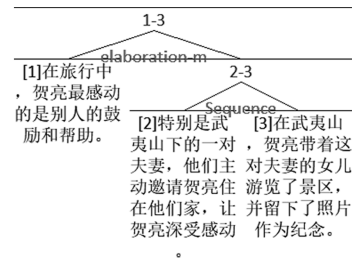
前文的量化指标与编辑行为分析表明, GPT-4o 在篇章简化过程中虽能有效降低层级复杂度,但在语篇连贯性和语义层次保持上仍存在系统性偏差。本节结合具体实例,对其结构性问题的表现形式及成因进行分析,以揭示大模型篇章简化的内部机制。

3.4.1 核心语义识别不足

GPT-4o 的主要偏差来源于对篇章核心语义的识别不稳定。在 RST 树结构中,核心单元(Nucleus)通常承载主要论述或事件信息,是语篇连贯的骨架。专家简化时会有意识地保留这些句子,并围绕核心层重组文本;而 GPT-4o 在删除操作中未能准确区分核心-卫星关系,导致部分核心语句被误删。统计显示, GPT-4o 因删除核心语义句而造成语篇连贯受损的比例高达 93.5%。例如,原文中承担因果解释的句子“因为门票贵,一家三口至今没进过景区”被删除后,事件之间的逻辑支撑消失,读者难以理解后续“贺亮”带夫妻的女儿进入景区的行为的动因(如图 2(a)(b)所示)。这说明 GPT-4o 在篇章层级推理方面缺乏显式的“语义重要度建模”机制,其判断往往基于表层词汇简化而非语义角色关系。



(a) 原文的结构树(部分)



(b) GPT-4o简化版的结构树(部分)

图 2 原文与 GPT-4o 简化版的篇章结构树对比

3.4.2 照应与衔接处理不当

除核心语义的误删外, GPT-4o 在处理照应链与语篇连接词时也存在“漏删”与“断链”问题。部分句子虽属卫星成分,却承担关键的前后照应功能,如提供时间、地点或指代对象。一旦被删,后文的指称关系就会失去依托。例如,虽然原文中的句[1]承担卫星语义关系(环境),但下文中又出现了“买菜”一词,如果删除[1]而保留[2],读者阅读时则会感觉十分突兀。该类失误在语料中占比虽不高(约 4%),

但对读者理解影响显著,表明 GPT-4o 尚未建立起对篇章级指代结构的整体感知。

例:a [1]今天,她像往常一样,来到学校附近的菜市场买菜。[2]Katri说,“学语言的最好方式就是多和别人沟通,在买菜的时候,就有机会跟卖菜的阿姨多交流、多对话。”(原文)

b [1]她说,“学习语言的最好方法就是多和别人交流,在买菜的时候,就有机会和卖菜的阿姨多交流。”(GPT-4o 简化版)

3.4.3 过度并列化

在合并操作中,GPT-4o 倾向于以“并列或顺承”方式压缩句群,将多层逻辑关系平面化。这种处理虽减少了篇幅,但同时削弱了事件的因果、让步、条件等逻辑衔接关系。例如,将“即使……但……”结构简化为单句陈述后,让步关系不再显现,篇章的逻辑张力被消解。与专家“保留主干、隐性支链”的策略不同,GPT-4o 直接删除或融合支链,形成信息密度高而逻辑递进弱的表述模式。

3.4.4 缺乏基于语义推理的补偿机制

从插入编辑行为来看,GPT-4o 的“详述性补偿”能力明显不足。专家在简化时常会插入背景、解释或过渡句,以弥合删改后可能出现的语义空隙;GPT-4o 虽能生成总结性句子,但很少主动补充因果或背景信息。这反映出当前大模型的篇章规划依赖于局部句间连贯,而非全局语义结构建模。

这些偏差反映了当前大模型在篇章层面语义表征与内容规划上的局限。与人类专家相比,GPT-4o 仍主要依赖局部生成概率而非结构化语义控制,缺乏对篇章树中“节点功能”与“关系角色”的显式建模。未来若在简化模型中引入篇章结构知识(如 RST 树嵌入或核-卫星重要度权重),并结合自我校正机制,或可显著提升大模型在篇章级文本简化中的连贯性与逻辑合理性。

3.5 篇章结构信息的应用潜力

研究表明,篇章结构信息不仅能够刻画文本简化过程中层级关系的变化,还可为自动评测与模型优化提供新的思路。现有的自动简化评测方法多侧重于词汇和句法层面,例如词长、句长及词汇覆盖率等指标,难以反映读者在理解过程中感知到的篇章结构复杂度。本文提出的九项篇章结构量化指标,从宏观层面揭示了文本逻辑组织和信息分布的特征,为自动化评测提供了可操作的结构维度。篇章结构指标可作为反映语篇复杂度与逻辑连贯度的重要评测维度。与传统依赖表层统计特征的评测体系相比,结构化指标更接近人类专家的判断方式,可为自动化评价提供更具解释力的依据。

大模型在简化任务中虽能有效压缩文本篇幅,但在缺乏显式结构约束的情况下,往往倾向于采用高频删除与合并策略,从而在跨文本泛化时出现篇章层级过度扁平化、逻辑关系隐化等问题。基于本文的分析结果,可以认为,大模型实现稳定篇章泛化简化的关键,在于引入对篇章结构功能的显式建模与反馈机制。一方面,在生成阶段可通过结构约束型提示(Structure-aware Prompting)或篇章模板引导,明确要求模型在简化过程中保留原文的核心语

义关系与层级组织,如因果、背景与顺承关系,从而避免单纯以信息压缩为目标的生成偏向。另一方面,在生成后阶段,可利用篇章结构量化指标对输出生成文本进行结构一致性检测,当指标显示篇章深度、关系多样性或核-卫星比例出现异常偏移时,触发模型的自我校正过程,对简化结果进行调整。

从这一意义上看,本文提出的篇章结构量化指标不仅可用于事后评测,也可作为大模型文本简化中的结构反馈信号,为构建“结构可控”的简化机制提供基础支持。通过将篇章结构信息融入生成与校正环节,有望在保持简洁性的同时增强文本的逻辑层次与连贯性,从而提升大模型在不同文本场景下进行篇章级简化的泛化稳定性。

4 总结与展望

本文基于修辞结构理论,从篇章结构的视角系统性地评估了大语言模型(GPT-4o)在汉语新闻文本简化任务中的表现。研究构建了首个汉语的简化篇章结构标注语料库,提出了九项篇章结构的量化指标。通过对比分析原文、专家简化文本与大模型简化文本的结构特征和编辑行为,本文揭示了大模型在篇章层面的简化规律及其局限性。结果显示,GPT-4o 能有效降低文本的篇章复杂度,但在识别核心语义、保持逻辑连贯和调控篇章关系方面仍有所欠缺。

进一步的机制分析表明,大模型的篇章结构性偏差主要来源于四个方面:其一,对核心语义单元的识别能力不足,易误删重要信息;其二,对照应关系与篇章连接的感知有限,导致局部衔接缺失;其三,过度并列化,导致文章的逻辑张力减弱;其四,缺乏在删改后自动补偿语义缺口的能力。相比之下,专家在简化中能通过结构重组、插入过渡句等手段维持篇章的逻辑一致性与信息完整性,体现出更高的篇章规划能力。

本文的主要贡献在于提出基于篇章结构的文本简化评测框架,并探讨了其在模型优化中的应用潜力。研究发现,篇章结构指标可作为衡量文本简化质量的重要参数,用于构建结构层面的评分体系;同时,这些指标可作为模型训练中的结构约束信号,引导生成过程遵循合理的层级与逻辑模式,从而实现结构可控的文本简化。本研究为大模型简化过程的细致评估提供了新的视角,但也存在局限性。本文的数据集规模较小且集中于新闻报道文本,其他文体可能具有不同的语篇特点和结构简化需求,后续研究将扩展至面向不同简化需求的多种文体。

参考文献

- [1] 李宇明. 无障碍社会的语言之功[J]. 语言战略研究, 2022, 7(5): 5-6.
- [2] SIDDHARTHAN A. Preserving discourse structure when simplifying text[C]//Proceedings of the 9th European Workshop on Natural Language Generation at EACL, 2003.
- [3] CROSSLEY S A, LOUWERSE M M, MCCARTHY P M, et al. A linguistic analysis of simplified and authentic texts [J]. The Modern Language Journal, 2007, 91(1): 15-30.
- [4] ARFÉ B, MASON L, FAJARDO I. Simplifying informational text structure for struggling readers [J]. Reading and Writing, 2018, 31(9): 2191-2210.
- [5] XU W, CALLISON BURCH C, NAPOLES C. Problems in current text simplification research: New data can help[J]. Transactions of the Association for Computational Linguistics, 2015(3): 283-297.
- [6] PATEL N, NAGPAL P, SHAH T, et al. Improving mathematics assessment readability: Do large language models help? [J] Journal of Computer Assisted Learning, 2023(3): 804-822.
- [7] ZHONG Y, JIANG C, XU W, et al. Discourse level factors for sentence deletion in text simplification [C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2020(5): 9709-9716.
- [8] RETS I, ASTRUC L, COUGHLAN T, et al. Approaches to simplifying academic texts in English: English teachers' views and practices[J/OL]. [2025-09-01]. English for Specific Purposes, <http://doi.org/10.1016/j.esp.2022.06.001>.
- [9] 吴继峰. 多元视角下留学生汉语教材文本改编策略研究[J]. 国际汉语教学研究, 2024(2): 62-73.
- [10] 高雪松, 周小兵, 洪伟. 国际中文教材文本改编: 基本原则与实现路径[J]. 四川师范大学学报(社会科学版), 2024(4): 138-144.
- [11] MARTÍNEZ P, RAMOS A, MORENO L. Exploring large language models to generate easy to read content[J]. Frontiers in Computer Science, 2024(6): 1394705.
- [12] 杨尔弘, 朱君辉, 朱浩楠, 等. 大语言模型的中文文本简化能力研究[J]. 语言战略研究, 2024, 9(5): 34-47.
- [13] LIU F, JIANG Y, LAI C, et al. Teacher engagement with automated text simplification for differentiated instruction[J]. Language Learning & Technology, 2024, 28(2): 163-182.
- [14] 娄开阳. 现代汉语新闻语篇的结构研究[M]. 北京: 世界图书出版公司, 2008.
- [15] MANN W C, THOMPSON S A. Relational propositions in discourse[J]. Discourse Processes, 1986, 9(1): 57-90.
- [16] VAN DIJK T A, KINTSCH W. Strategies of discourse comprehension [M]. New York: Academic Press, 1983.
- [17] CRISTEA D, IDE N, ROMARY L. Veins theory: A model of global discourse cohesion and coherence [C]//Proceedings of the COLING-AACL Conference, 1998.
- [18] 蒋峰, 褚晓敏, 徐昇, 等. 基于主题相似度的宏观篇章主次关系识别方法[J]. 中文信息学报, 2018, 32(1): 43-50.
- [19] 孔芳, 王红玲, 周国栋. 汉语篇章理解研究综述[J]. 软件学报, 2019, 30(7): 2052-2072.
- [20] SRIKANTH N, LI J J. Elaborative simplification: Content addition and explanation generation in text simplification[C]//Proceedings of the Association for Computational Linguistics: ACL-IJCNLP, 2021.
- [21] HEWETT F, ASGHARI H, STEDE M. Elaborative simplification for German-language texts [C]//Proceedings of the 25th Annual Meeting of the Special Interest Group on Discourse and Dialogue. Kyoto, Japan: Association for Computational Linguistics, 2024.
- [22] 洪伟, 赖丽琴. 语境详述能否促进伴随性词汇习得: 来自汉语二语的实证研究[J]. 语言教学与研究, 2023(6): 38-52.
- [23] NUSSBAUMER M. Textbegriff und textanalyse [C]//EISENBERG P, KLOTZ P, eds. Sprache gebrauchen-Sprachwissen Erwerben. Stuttgart: Klett, 1993.



朱君辉(1998—), 博士, 主要研究领域为自然语言处理和智能语言学习。
E-mail: nysyzxzjh@163.com



朱浩楠(2001—), 硕士, 主要研究领域为自然语言处理和智能语言学习。
E-mail: 1930940919@qq.com



杨尔弘(1965—), 通信作者, 博士, 教授, 主要研究领域为语言信息处理、语言监测和语言资源建设。
E-mail: yerhong@blcu.edu.cn