

英语母语者不同学术文体的计量研究

周婉婷

浙江财经大学外国语学院, 浙江 杭州

收稿日期: 2023年12月29日; 录用日期: 2024年2月21日; 发布日期: 2024年2月28日

摘要

文体的句法复杂性一直受到学者们的关注, 依存距离是评估句法复杂性的指标之一。以往的研究发现, 依存距离、句子长度和文体三者密切相关。然而, 很少有研究系统地考察英语母语者不同学术文体下, 依存距离和句子长度的分布情况。本研究基于LONCESS语料库, 采用定量方法研究了议论文和散文的句长分布和依存距离分布, 以及两者之间的关系。结果表明: 1) 两种文体的句长分布遵循不同的分布模型。2) 两种文体的依存距离分布受体裁的影响有一定的规律性。3) 关于句长和平均依存距离之间的关系, 线性回归模型能较好地拟合议论文体, 复合曲线模型能较好地拟合散文文体。总之, 研究结果表明, 两种文体的句子长度和依存距离的概率分布遵循长尾效应, 表明人类的工作记忆是有限的。

关键词

定量研究, 依存距离, 句长, 文体, 句法复杂度

A Quantitative Study of Different Academic Genres of Native English Speakers

Wanting Zhou

School of Foreign Languages, Zhejiang University of Finance and Economics, Hangzhou Zhejiang

Received: Dec. 29th, 2023; accepted: Feb. 21st, 2024; published: Feb. 28st, 2024

Abstract

The syntactic complexity of genres has always been of interest to scholars, and dependency distance is one of the indicators for assessing syntactic complexity. Previous studies have found that dependency distance, sentence length and genre are closely related. However, few studies have systematically examined the distribution of dependency distance and sentence length in different genres of native English speakers. In this study, based on the LONCESS corpus, a quantitative method was used to investigate the distribution of sentence length and the distribution of dependen-

cy distance in argumentative and prose genres, as well as the relationship between the two. The results show that: 1) The sentence length distributions of the two genres follow different distribution models. 2) The dependency distance distributions of the two genres are affected by the genres with a certain regularity. 3) Regarding the relationship between the sentence length and the average dependency distance, the linear regression model fits the argumentative genre better, and the composite curve model fits the literary genre better. In conclusion, the results show that the probability distributions of sentence length and dependency distance for both genres follow the long-tail effect, indicating that human working memory is limited.

Keywords

Quantitative Research, Dependency Distance, Sentence Length, Genre, Syntactic Complexity

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

文体主要分为记叙文和非记叙文[1]。记叙文主要描述事件和行为,而非记叙文主要表达主张和观点,包括说明文、议论文等。不同的体裁意味着单词和短语的组合方式不同,分句和句子的构成方式也不同[2]。不少学者已证实[3]-[8],句法难度因文体而异。依存距离作为测量句法难度的重要指标之一,通常受到体裁和句子长度的影响。例如,基于BNC语料库,Wang和Liu[9]报告称,句子长度越长,平均依存距离也越长。然而,少有研究探索英语母语者不同文体的句法复杂度。因此,本研究旨在从句长和依存距离两个维度探讨英语母语者在议论文和散文的句法特征,扩大了依存语法的应用范围。同时,为量化母语者的句法复杂度提供了新思路。

2. 文献综述

依存距离被认为是衡量句法复杂度的标准之一,那么平均依存距离也是预测句法难度和写作水平的基本指标[10][11][12]。可能影响依存距离的因素包括句子长度、体裁、句法标注等[13][14]。

不少学者考察了不同语言中句子长度对平均依存距离的影响。例如,Oya[15]在三组不同的语料库(日本高中使用的英语教科书、日本学习者撰写的英语作文和Parc700依存树库)中提取了少于10到30个单词的句子。研究表明,10个单词及以上和20个单词以下句子的平均依存距离有显著差异。Jiang和Liu[14]对比了中英文语料中10到30个词的句子长度。研究发现,平均依存距离与句子长度之间存在正相关。与前人研究一致,Wang和Liu[9]也发现句子长度对依存距离有显著影响,但影响较小。值得注意的是,他们将句长的范围扩大到5~40个单词。

体裁也是影响依存距离的重要因素之一[9][16][17][18]。一些学者致力于量化非学术文体的平均依存距离,尤其是新闻体裁。例如,Wang和Liu[9]报告了BNC中十种文体的平均依存距离,发现信息文体的平均依存距离与想象文体相近或略大于想象文体。Chen *et al.* [17]计算了Freiburg-Brown语料库中12种文本类型在句子长度控制下的平均依存距离和分层依存距离。Oya[13]的研究发现,非虚构文体和科技文体的平均依存距离(分别为3.37和3.20)长于虚构文体(2.65)。以往研究的结果在不同程度上证明了句子长度和文体对平均依存距离的影响。然而,大多数的研究只控制了一个变量(句长或文体),少有研究同时考察不同文体中句长和依存距离的分布和关系。

3. 研究设计

3.1. 研究问题

本研究对英语母语者两种文体的句法复杂度进行了定量研究，主要考察了句长分布、依存距离分布以及句长对依存距离的影响。研究结果可以揭示不同体裁的句法特征，并为依存距离最小化提供证据。

研究问题主要有以下三个：

- 1) 句子长度在议论文体和散文文体中的分布情况如何？是否遵循某些分布模型？
- 2) 不同句子长度的依存距离在不同体裁中的分布情况如何？是否遵循某些分布模型？
- 3) 不同体裁下平均依存距离和句子长度之间有什么关系？

3.2. 语料库

LOCNESS 语料库由鲁汶天主教大学英语语料库语言学中心(CECL)编制，共计 324,304 字的英语母语文章。其中，有 436 篇文章是由美国大学生写的，90 篇文章是英国大学英语完成的。本研究从中选取了英国母语者的 33 篇议论文和 33 篇散文。所有写作的学生均在考试中或一小时左右完成了至少 500 字的作文，且不借助任何其他材料(见表 1)。

为了后续的研究，笔者进行人工删除乱码后，使用 BFSU_Sentence_Segmenter_1.0 对文本进行分段，将其划分为不同的句群。本研究以词为测量单位，使用 Python (3.12.0) 计算句长，表 2 展示了各句长的详细数量。

Table 1. The description of the corpus

表 1. 语料库概况

Genre	Tokens	Texts	Participants
argumentative	19,019	33	British university students
literary	58,889	33	British university students

Table 2. Number of sentence length in two genres

表 2. 两种文体下句长的数量

Sentence length	Argumentative	Literary
5	2	21
6	5	31
7	11	44
8	4	54
9	24	67
10	18	75
11	9	84
12	23	100
13	26	87
14	31	98
15	30	121
16	38	120

续表

17	26	110
18	37	112
19	29	102
20	24	79
21	33	88
22	34	91
23	21	91
24	26	80
25	23	77
26	23	72
27	15	66
28	24	52
29	15	54
30	19	37
31	19	40
32	17	53
33	14	40
34	16	45
35	11	30
36	13	36
37	9	28
38	8	26
39	7	18
40	4	16
Total	688	2345

3.3. 研究方法

本研究采用 Liu [19]提出的计算方法来收集每个句子的平均依存距离。句子的依存距离计算公式如下所示：

$$MDD (the\ sentence) = \frac{1}{n-1} \sum_{i=1}^{n-1} |DD_i| \quad (1)$$

这里的“ n ”是指句子的总字数， $|DD_i|$ 是指句子中第 i 个句法链接的绝对依存距离。图 1 展示了我们语料库中一个样本句子的依存结构。每个箭头都连接着一个主句和一个从句。弧线表示这个句子包含四个依存关系对，有四种不同的依存关系。在(suffer, section)这对依存关系中，section 被称为依存词，suffer 是支配词。为了测量句子的平均依存距离，所有文章都用自动技术 Wordless 进行了处理。表 3 列出了例句中各依存词和支配词之间的距离。根据公式(1)，例句的平均依存距离计算如下： $(4 + 1 + 2 + 1 + 0)/4 = 2$ 。

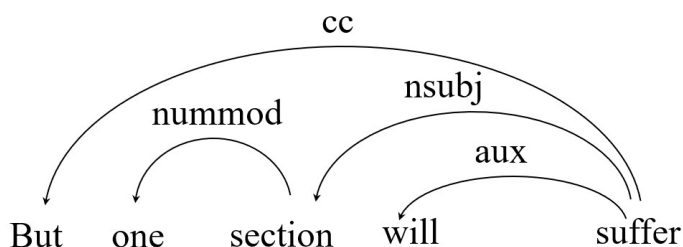


Figure 1. Dependency structure of sample sentence *But one section will suffer*
图 1. 例句 *But one section will suffer* 的依存结构

Table 3. Dependency distance of sample sentence *But one section will suffer*
表 3. 例句 *But one section will suffer* 的依存距离

Governor	Dependent	Dependency type	Dependency distance	Absolute dependency distance
suffer	But	cc	4	4
section	one	nummod	1	1
suffer	section	nsubj	2	2
suffer	will	aux	1	1
suffer	suffer	ROOT	0	0

3.4. 数据分析

首先, 本研究使用 Altmann-Fitter 考察句长和依存距离是否遵循某些分布模型。在 Altmann-Fitter 的处理结果中, 有三个重要指标用于衡量模型的拟合效果: 概率 $P(x^2)$ 、差异系数 C 和拟合判定系数 R^2 。当 $P(x^2) > 0.05$ 时, 拟合结果最佳。 $C < 0.02$, 结果良好; $C < 0.01$, 结果非常好。 $R^2 > 0.8$, 结果可接受; $R^2 > 0.9$, 结果非常好[20] [21]。其次, 回归相关分析用于考察句长与平均依存距离之间的关系, 旨在找出相关系数显著($p < 0.05$)且相关性强($R^2 > 0.8$)的最佳模型。

4. 结果

4.1. 句长频次的概率分布

研究结果显示, 议论文的句长频次分布符合波利亚概率分布。如表 4 所示, 模型拟合良好, $P(x^2) = 0.0972 > 0.05$, $C = 0.0621$, $R^2 = 0.7486 > 0.5$ 。然而, 散文的句长频次符合右截尾负二项模型。如表 5 所示, 该模型拟合效果较好, R^2 值大于 0.8。

Table 4. The Polya distribution of sentence length of argumentative genre
表 4. 议论文句长的波利亚概率分布

	X^2	$P(x^2)$	C	DF	R^2
Value	42.7394	0.0972	0.0621	32	0.7486

Table 5. The Right truncated negative binomial distribution of sentence length of literary genre
表 5. 散文句长的右截断右截尾负二项模型

	X^2	$P(x^2)$	C	DF	R^2
Value	108.1125	0.0000	0.0461	32	0.8926

为了更好地观察句长频次的概率分布,我们用折线图来展示两种文体的理论值和预测值的分布。从图2和图3可以看出,议论文和散文的句子长度分布呈现出相似的趋势:分布呈正偏态,曲线右侧较长,左侧较短。达到峰值后,句子长度越长,频率越低。具体来说,议论文的频率峰值在15~23之间,当句子长度为16时达到峰值。散文的最高频率介于[15,19]之间,当句子长度为15时达到最高值。比较两个模型,散文的预测值和实际值之间的差异大于议论文。

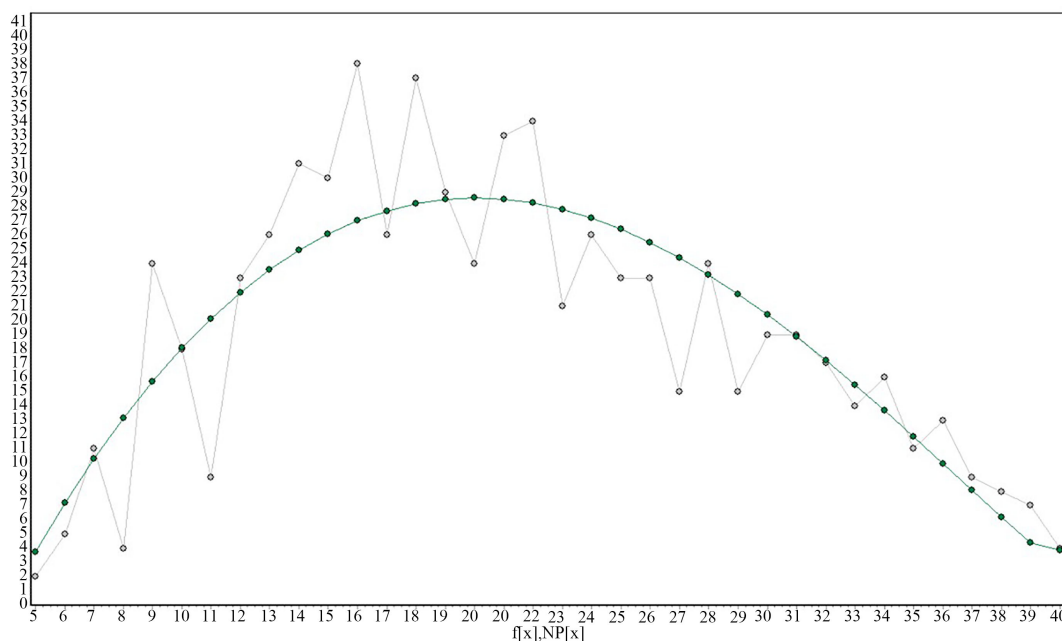


Figure 2. Distribution of sentence length in argumentative essays

图2. 议论文的句长分布图

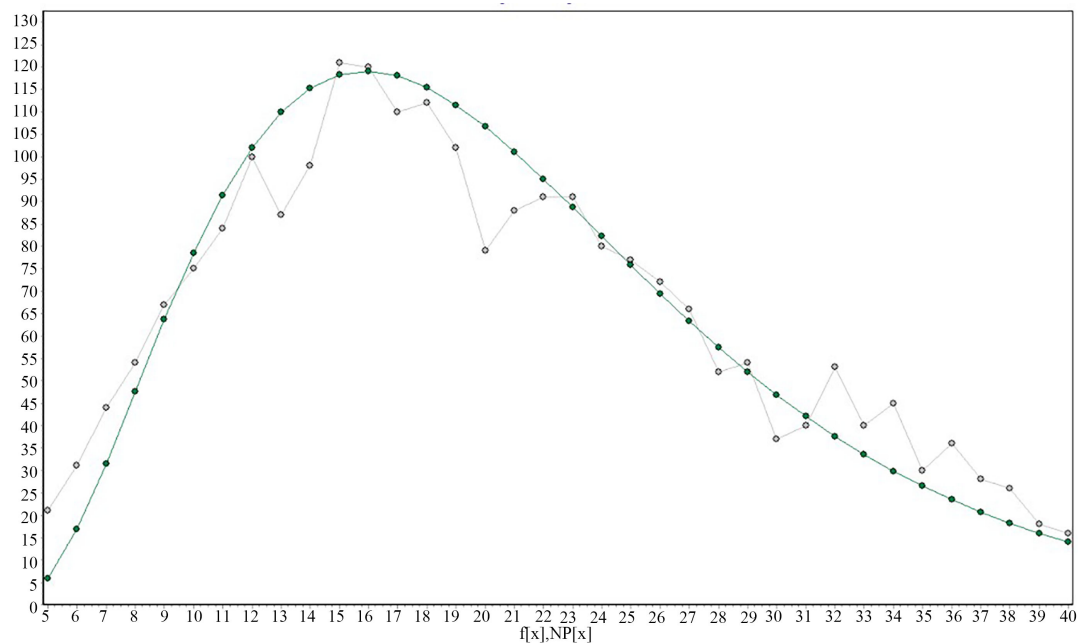


Figure 3. Distribution model of sentence length in literary essays

图3. 散文的句长分布图

4.2. 依存距离频次的概率分布

人类语言中依存距离分布的规律性已被许多学者证实。从本质上讲，语言呈现出指数函数和幂律的混合分布。语种、句长和文体都会对分布产生影响。本研究使用 Altmann-Fitter 进行自动拟合，发现两种文体的依存距离符合以下概率分布：Right truncated modified Zipf-Alekseev ($a, b; n = x - \text{最大值}, \alpha$ 固定)、Right truncated Waring (b, n)、Right truncated Negative Binomial ($k, p; R = x - \text{最大值}$)和 Waring (b, n)。表 6 列出了两种体裁在不同分布模型下的 R^2 值。

就议论文而言，Waring (b, n)的拟合效果($R^2 = 0.9978$)优于其他三种分布模型(b, n) ($R^2 = 0.9976$)。Baayen [22]认为 Waring 是重要的词频分布之一。Jiang 和 Liu [14]也发现 Waring 分布适合英语语料($R^2 = 0.994$)。就散文而言，Right truncated modified Zipf-Alekseev ($a, b; n = x - \text{max}, \alpha$ 固定)模型表现最佳($R^2 > 0.99, C < 0.02$)。总体而言，上述研究结果证明了不同文体的依存距离概率分布也是具有一定的规律性。

Table 6. Fitting models of the dependence distance across two genres

表 6. 两种文体的依存距离拟合模型

	Right truncated modified Zipf-Alekseev	Right truncated Waring	Right truncated Negative Binomial	Waring
Argumentation	0.9965	0.9976	0.9628	0.9978
Literary	0.9993	0.9948	0.9566	0.9950

为了更好地观察依存距离的概率分布，图 4 和图 5 展示了议论文体的 Waring 分布和散文文体的 Right truncated modified Zipf-Alekseev 分布的拟合情况。从图中可以看出，两种文体的依存距离概率分布呈现长尾效应。进一步考察依存距离的频率，研究发现相邻依存距离(依存距离为 1)的频次最多。其中，在议论文中，相邻依存距离占总依存距离分布的 47%。在散文中，相邻依存距离的比例达到了 49%。可见，依存距离越长，出现频率越低。

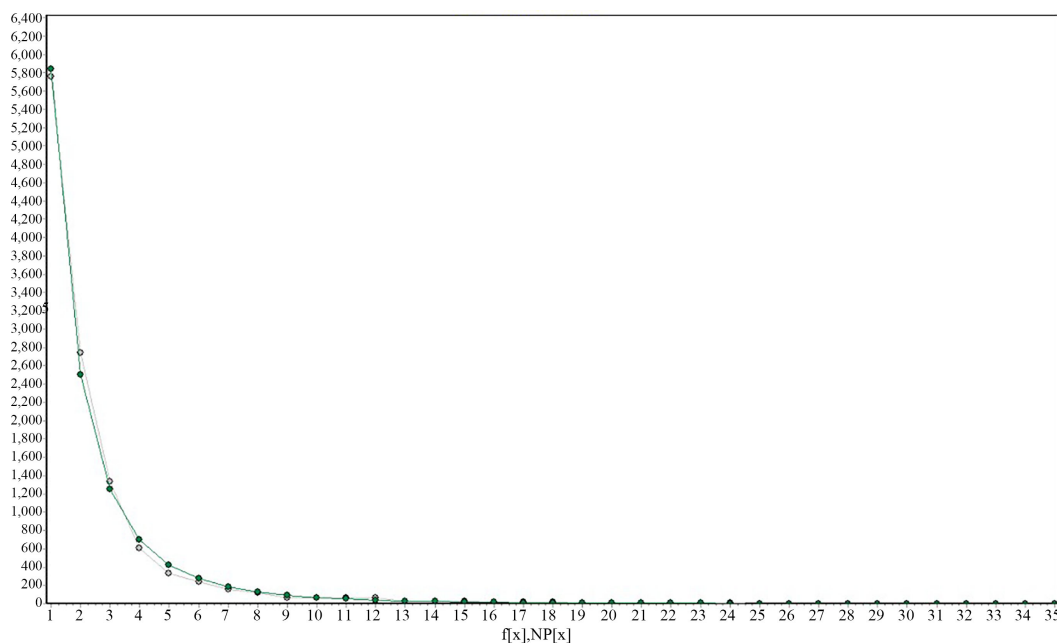


Figure 4. Waring model of argumentative genre

图 4. 议论文的 Waring 模型

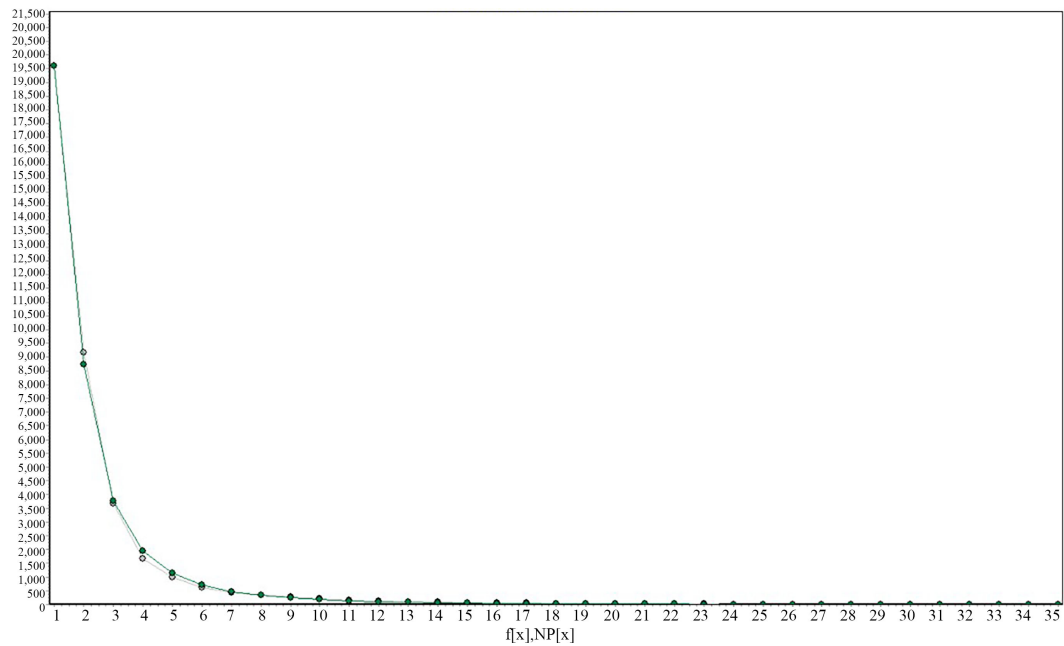


Figure 5. Right truncated modified Zipf-Alekseev model in literary genre
 图 5. 散文的 Right truncated modified Zipf-Alekseev 模型

4.3. 不同文体下句长与平均依存距离的关系

图 6 显示，两种文体的平均依存距离均随着句长的增长而增大。这证明了前人研究的结果[9]，即句子越长，平均依存距离越大。对比两种文体，研究发现除了 21 到 30 的句子长度范围外，散文的平均依存距离都比议论文体长(见表 7)。

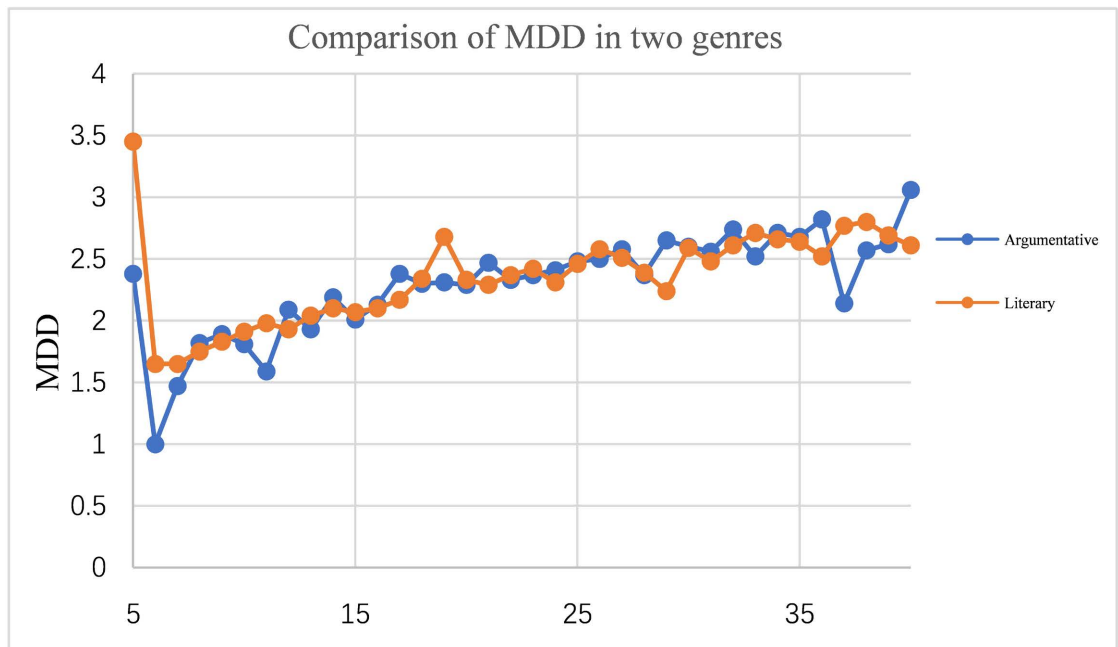


Figure 6. Comparison of mean dependency distances across genres with different sentence lengths
 图 6. 两种文体下不同句长的平均依存距离对比

Table 7. The dependency distance with different sentence length across two genres
表 7. 两种文体下不同词长的依存距离

Sentence length range	Argumentative essays	Literary essays
5	2.38	3.45
5~10	1.73	2.04
11~20	2.12	2.17
21~30	2.48	2.42
31~40	2.64	2.65

在上一部分的数据分析展示了不同文体下，句长和平均依存距离之间呈现正相关，但它们之间的关系还需要进一步分析。为了解决这个问题，本研究使用回归模型进行分析。首先，以议论文的句长和平均依存距离分别作为因变量和自变量。结果显示，议论文的线性拟合公式为 $y = 1.592 + 0.032x$ ($R^2 = 0.647$, $F = 62.348$, $p = 0.000 < 0.05$) (见表 8 和图 7)。该模型显示出中等程度的相关性($R^2 > 0.5$)，这一结果与 Jiang 和 Liu [14] 的研究结果相似，但他们没有考虑文体的影响。

Table 8. Linear regression model of argumentative genre
表 8. 议论文的线性回归模型

Equation	Model Summary					Parameter Estimates	
	R^2	F	df1	df2	Sig.	Constant	b1
Linear	0.647	62.348	1	34	0.000	1.592	0.031

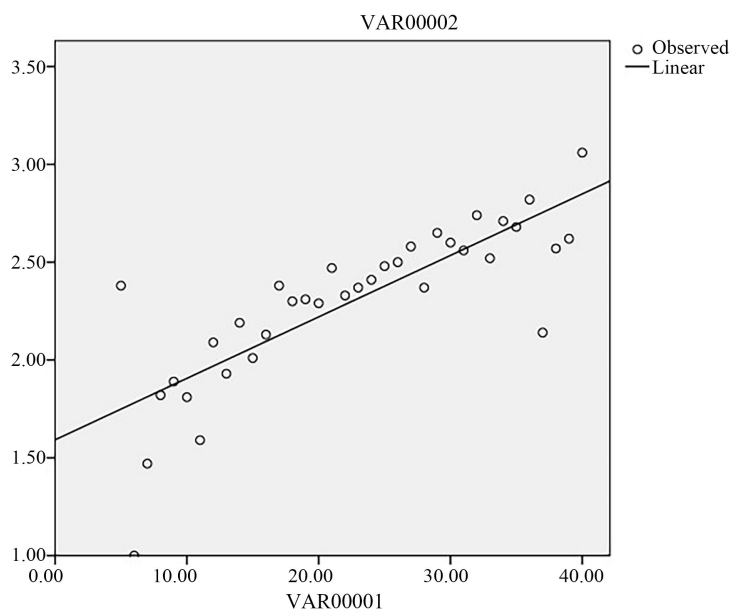


Figure 7. The fitting model of argumentative genre
图 7. 议论文体的拟合模型

其次，本研究考察了散文的句长与平均依存距离之间的关系。研究表明，复合曲线模型的拟合效果最好(见表 9)，但其显著性相对较弱($R^2 = 0.449 < 0.8$, $F = 27.676$, $p = 0.000^*$)。回归方程为 $y = \ln 1.84 + (\ln 1.01)x$ ，拟合曲线如图 8 所示。比较两个回归模型，我们发现 p 值均呈现显著性差异，表明两种文体下

平均依存距离和句子长度是相互关联的。但是，散文的平均依存距离与句长之间的关联较弱($R^2 < 0.8$)。原因可能是散文中存在一个离散点，最终导致结果不理想。

Table 9. Linear regression model of literary genre

表 9. 散文的线性回归模型

Equation	Model Summary					Parameter Estimates	
	R^2	F	df1	df2	Sig.	Constant	b1
Linear	0.449	27.676	1	34	0.000	1.84	1.01

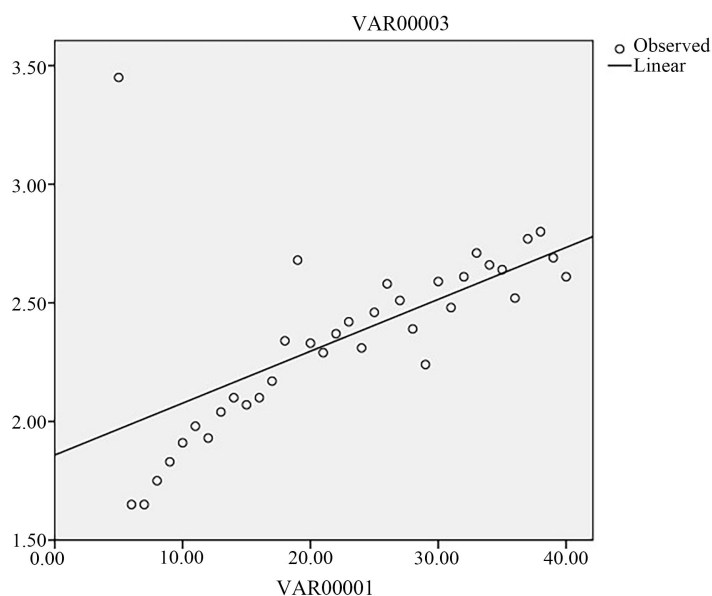


Figure 8. The fitting model of literary genre

图 8. 散文的拟合模型

5. 总结

基于 LONCESS 语料库，本研究对英语母语者两种学术文体的句长和依存距离进行了计量研究。结果表明，句子长度的概率分布模型受到文体的影响：议论文体遵循波利亚分布，而散文文体遵循右截尾负二项分布。虽然两种文体遵循不同的分布模型，但它们都表现出长尾效应。这两种体裁的句子长度在 15~16 个单词达到最高点。在最高点之后，句子数量随着句子长度的增加而减少。

针对第二个研究问题，结果发现依存距离的概率分布显示出受文体影响的规律性。本研究运用四种分布模型对两种文体进行拟合，拟合结果较好($R^2 > 0.9$)。Baayen [22]提出，Waring 和 zeta 是重要的词频分布，Waring 已被证明适用于句法单位[23]。本研究结果表明，议论文的依存距离分布可能更适合 Waring 模型，而散文更遵循右截断 Zipf-Alekseev 分布。这一结果与 Popescu *et al.* [24]的研究部分一致，他们发现 Zipf-Alekseev 函数可以捕捉许多语言单位的物理长度分布。依存距离的一个重要特征是相邻依存距离比例高，这在人类语言中起着重要作用。Wang 和 Liu [9]也得出了类似的结论，即相邻依存距离的百分比保持在 50%左右。总之，本研究的结果再次证实，无论文体和句子长度如何，人类语言都倾向于依赖距离最小化。这可能是由于人类的工作记忆是有限的，如果存储的词语过多，输出就会失败[25] [26]。

针对不同文体下平均依存距离与句长之间的联系, 结果发现在两种体裁中, 句子长度越长, 平均依存距离越长。散文的平均依存距离略长于议论文。这说明, 与议论文体相比, 作者在写散文时运用了更复杂的句法结构, 带来了更大的认知负担。究其原因, 可能是因为受到作文题目的影响。本研究中的散文题目是讨论某本书中复杂的哲学问题, 学生可能会引用书中的内容, 总结不同学者的生平、哲学思想和对他人的批评。然而, 议论文的基本思路是要求学生给出一个逻辑结构, 并以连贯和简单的方式组织自己的观点[27]。在进一步探讨句子长度与平均依存距离之间的联系时, 研究发现回归结果与假定句子长度与依存距离之间是线性关系的理论假设并不一致。线性模型认为, 句子长度越长, 依存距离越长。议论文体下, 线性回归模型能较好地拟合句长和依存距离之间的关系; 但在散文文体中, 句长和依存距离之间的关系需要用复合曲线模型进行解释。这是因为在散文文体中, 一些离散点导致线性函数的拟合结果不太理想。这与 Ferrer-i-Cancho 和 Arias [28]的研究结果不谋而合, 他们在 Catalan 树库发现线性模型无法拟合所有语言和体裁, 因此拟合了几个非线性函数。

本研究的重要贡献之一是扩大了依存距离的应用范围, 但依存距离并不是文体分类的可靠指标, 原因是两种体裁之间的平均依存距离差异并不显著($p > 0.05$)。其次, 本研究补充了以往的研究结果。前人研究发现记叙文对认知的要求最低, 说明文比记叙文对认知的要求更高, 而议论文对认知的要求最高[29][30][31]。结果显示, 散文体裁比议论体裁对学生认知努力的要求更高。本研究证明了, 不同体裁对学生认知努力的要求不同。

基金项目

浙江省哲学社会科学规划项目(22JCXK12YB)。

参考文献

- [1] Brunner, J.S. (1986) *Actual Minds, Possible Worlds*. Harvard University Press, Cambridge. <https://doi.org/10.4159/9780674029019>
- [2] Biber, D. and Conrad, S. (2009) *Register, Genre, and Style*. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9780511814358>
- [3] Scott, C.M. and Windsor, J. (2000) General Language Performance Measures in Spoken and Written Narrative and Expository Discourse of School-Age Children with Language Learning Disabilities. *Journal of Speech, Language, and Hearing Research*, **43**, 324-339. <https://doi.org/10.1044/jslhr.4302.324>
- [4] Casal, J.E., Lu, X., Qiu, X., Wang, Y. and Zhang, G. (2021) Syntactic Complexity across Academic Research Article Part-Genres: A Cross-Disciplinary Perspective. *Journal of English for Academic Purposes*, **52**, Article ID: 100996. <https://doi.org/10.1016/j.jeap.2021.100996>
- [5] Lu, X., Casal, J.E., Liu, Y., Kisselev, O. and Yoon, J. (2021) The Relationship between Syntactic Complexity and Rhetorical Move-Steps in Research Article Introductions: Variation among Four Social Science and Engineering Disciplines. *Journal of English for Academic Purposes*, **52**, Article ID: 101006. <https://doi.org/10.1016/j.jeap.2021.101006>
- [6] Yu, S. (2021) Giving Genre-Based Peer Feedback in Academic Writing: Sources of Knowledge and Skills, Difficulties and Challenges. *Assessment & Evaluation in Higher Education*, **46**, 36-53. <https://doi.org/10.1080/02602938.2020.1742872>
- [7] Jeong, H. (2017) Narrative and Expository Genre Effects on Students, Raters, and Performance Criteria. *Assessing Writing*, **31**, 113-125. <https://doi.org/10.1016/j.asw.2016.08.006>
- [8] Pu, L., Heng, R. and Cao, C. (2022) The Effects of Genre on the Syntactic Complexity of Argumentative and Expository Writing by Chinese EFL Learners. *Frontiers in Psychology*, **13**, Article 1047117. <https://doi.org/10.3389/fpsyg.2022.1047117>
- [9] Wang, Y. and Liu, H. (2017) The Effects of Genre on Dependency Distance and Dependency Direction. *Language Sciences*, **59**, 135-147. <https://doi.org/10.1016/j.langsci.2016.09.006>
- [10] Liu, H.T. (2008) Dependency Distance as a Metric of Language Comprehension Difficulty. *Journal of Cognitive Science*, **9**, 159-191. <https://doi.org/10.17791/jcs.2008.9.2.159>

- [11] Jiang, J. and Ouyang, J. (2017) Dependency Distance: A New Perspective on the Syntactic Development in Second Language Acquisition: Comment on “Dependency Distance: A New Perspective on Syntactic Patterns in Natural Language” by Haitao Liu *et al. Physics of Life Reviews*, **21**, 209-210. <https://doi.org/10.1016/j.plrev.2017.06.018>
- [12] Ouyang, J., Jiang, J. and Liu, H. (2022) Dependency Distance Measures in Assessing L2 Writing Proficiency. *Assessing Writing*, **51**, Article ID: 100603. <https://doi.org/10.1016/j.asw.2021.100603>
- [13] Oya, M. (2013) Degree Centralities, Closeness Centralities, and Dependency Distances of Different Genres of Texts. *Proceedings of the 17th Conference of Pan-Pacific Association of Applied Linguistics*, Beijing, 21-23 August 2012, 89-90.
- [14] Jiang, J. and Liu, H. (2015) The Effects of Sentence Length on Dependency Distance, Dependency Direction and the Implications—Based on a Parallel English—Chinese Dependency Treebank. *Language Sciences*, **50**, 93-104. <https://doi.org/10.1016/j.langsci.2015.04.002>
- [15] Oya, M. (2011) Syntactic Dependency Distance as Sentence Complexity Measure. *Proceedings of the 16th International Conference of Pan-Pacific Association of Applied Linguistics*, Hongkong, 8-10 August 2011, 313-316.
- [16] Hiranuma, S. (1999) Syntactic Difficulty in English and Japanese: A textual study. *UCL Working Papers in Linguistics*, **11**, 309-322.
- [17] Chen, R., Deng, S. and Liu, H. (2021) Syntactic Complexity of Different Text Types: From the Perspective of Dependency Distance Both Linearly and Hierarchically. *Journal of Quantitative Linguistics*, **29**, 510-540. <https://doi.org/10.1080/09296174.2021.2005960>
- [18] Gao, N. and He, Q. (2023) A Corpus-Based Study of the Dependency Distance Differences in English Academic Writing. *SAGE Open*, **13**, 1-12. <https://doi.org/10.1177/21582440231198408>
- [19] Liu, H. (2009) *Dependency Grammar: From Theory to Practice*. Science Press, Beijing.
- [20] Mačutek, J. and Wimmer, G. (2013) Evaluating Goodness-of-Fit of Discrete Distribution Models in Quantitative Linguistics. *Journal of Quantitative Linguistics*, **20**, 227-240. <https://doi.org/10.1080/09296174.2013.799912>
- [21] Liu, H., Xu, C. and Liang, J. (2017) Dependency Distance: A New Perspective on Syntactic Patterns in Natural Languages. *Physics of Life Reviews*, **21**, 171-193. <https://doi.org/10.1016/j.plrev.2017.03.002>
- [22] Baayen, R.H. (2001) *Word Frequency Distributions*. Springer, Dordrecht. <https://doi.org/10.1007/978-94-010-0844-0>
- [23] Köhler, R. and Altmann, G. (2000) Probability Distributions of Syntactic Units and Properties. *Journal of Quantitative Linguistics*, **7**, 189-200. <https://doi.org/10.1076/jqul.7.3.189.4114>
- [24] Popescu, I.I., Best, K.H. and Altmann, G. (2014) *Unified Modeling of Length in Language*. RAM-Verlag, Lüdenscheid.
- [25] Covington, M. (2003) *A Free-Word-Order Dependency Parser in Prolog*. The University of Georgia, Athens.
- [26] Cowan, N. (2001) The Magical Number 4 in Short-Term Memory: A Reconsideration of Mental Storage Capacity. *Behavioral and Brain Sciences*, **24**, 87-185. <https://doi.org/10.1017/S0140525X01003922>
- [27] Grabe, W. (2002) Narrative and Expository Macro-Genres. In: Johns, A.M., Ed., *Genre in the Classroom: Multiple Perspectives*, Lawrence Erlbaum Associate, Mahwah, 249-267.
- [28] Ferrer-i-Cancho, R. and Arias, M. (2013) Non-Linear Regression on Dependency Trees. *Lecture on Complex and Social Networks*.
- [29] Bain, A. (1967) *English Composition and Rhetoric*. 2nd Edition, Appleton & Company, New York.
- [30] Genung, J.F. (1900) *The Working Principles of Rhetoric: Examined in Their Literary Relations and Illustrated with Examples*. Ginn & Company, Boston.
- [31] Weigle, S.C. (2002) *Assessing Writing*. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9780511732997>