

英语新闻依存距离特征

李哲先

上海海事大学外国语学院, 上海

收稿日期: 2025年12月4日; 录用日期: 2026年1月4日; 发布日期: 2026年1月13日

摘要

依存距离(DD)作为量化句法复杂度的指标,已在跨语言和跨体裁比较研究中得到广泛应用,但目前缺乏对单一语类,特别是英语新闻中不同主题和内容类型句子的依存距离分析。本研究选取了Choubey (2020)构建的News Discourse语料库,使用Python的Spacy库进行依存句法分析,比较了英语新闻中不同主题(政治、商业、灾难和犯罪)和内容类型(事件、背景、评价、传闻等)句子的标准化依存距离(NDD)。研究发现:1) 结果显示,商业、政治主题新闻句子的标准化依存距离差异较大,而犯罪和灾难新闻句子的标准化依存距离差异较小。2) 讲述主要事件和未来展望的句子NDD值最高,而传闻类句子的NDD值最低。本研究通过将依存距离指标应用于新闻体裁内部,填补了现有研究空白,并提出NDD或可作为衡量新闻翻译记忆负荷的有效量化指标,对翻译实践与教学具有启示意义。

关键词

标准化依存距离, 新闻主题, 内容类型

Dependency Distance Features of English News

Zhexian Li

College of Foreign Languages, Shanghai Maritime University, Shanghai

Received: December 4, 2025; accepted: January 4, 2026; published: January 13, 2026

Abstract

Dependency distance (DD), a quantitative index of syntactic complexity, has been widely applied in cross-linguistic and cross-genre comparative research. However, analyses within a single genre remain scarce, particularly with respect to how dependency distance varies across topics and content types in English news. Drawing on the News Discourse corpus compiled by Choubey (2020), this study conducts dependency parsing using spaCy in Python and compares normalized dependency

distance (NDD) across four news topics (politics, business, disaster, and crime) and multiple content types (e.g., event, background, evaluation, and rumor). The results show that sentences in business and political news display greater variation in NDD, whereas sentences in crime and disaster news exhibit relatively smaller differences in NDD. In addition, sentences describing the main event and future outlook yield the highest NDD values, while rumor sentences yield the lowest. By applying dependency distance to an intra-genre analysis of news discourse, this study helps fill a gap in the literature and suggests that NDD may serve as a useful quantitative proxy for estimating memory demands in news translation, with implications for translation practice and pedagogy.

Keywords

Normalized Dependency Distance, News Topics, Content Types

Copyright © 2026 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

在翻译研究中,越来越多的学者采用语料库方法,将翻译复杂度转化为可度量的指标,从而推断译者在翻译过程中的记忆负荷。以往的研究主要集中在词汇层面的分析,如词汇多样性、密度和难词比例等。然而,翻译的记忆负荷不仅受词汇因素的影响,还与句法结构密切相关。当句子的依存距离增大时,译者需要在工作记忆中维持更多尚未闭合的句法关系,导致记忆负荷加重。因此,若要解释翻译中的认知压力,仅凭词汇指标并不足够,句法结构,特别是依存结构,也是构成译者记忆负荷的重要因素。

依存距离指标为研究者提供了衡量句法复杂度的新途径,其中平均依存距离(Mean Dependency Distance, MDD)能够有效反映句法关系在工作记忆加工中的难易程度。句法依存距离越长,译者进行翻译转换的记忆负荷越大,依存距离较短的句子的记忆负荷相对越低。尽管关于依存距离的跨语言和跨体裁比较已有一定的研究成果,但针对同一体裁内部功能性差异的研究仍显不足。识别不同体裁文本的特点有助于译者预测其翻译过程中可能遇到的记忆负荷压力,进一步研究同一体裁内不同主题或承担不同内容类型的句子之间的依存距离差异,也有助于预测翻译过程中哪些部分可能导致更大的记忆负荷。

新闻作为一种具有独特且相对稳定结构的体裁,不同栏目报道不同主题,而不同主题及其承担的内容类型在信息组织方式和句法结构上可能存在显著差异。基于此,本文以新闻语篇为研究对象,重点探讨不同主题(如政治、商业、灾难、犯罪)以及内容类型(如事件、前因背景、当前背景、历史背景、传闻、评价、未来展望等)的句子在平均依存距离上的差异,并进一步探讨这些差异对翻译记忆负荷的潜在启示。

2. 文献综述

2.1. 依存距离指标的应用

依存距离(Dependency Distance)作为句法复杂度的衡量指标,广泛应用于心理语言学和翻译学研究中,用以评估句子加工的难易程度。Liu (2008)提出的平均依存距离(MDD)方法,通过跨语言比较揭示了依存距离最小化的普遍规律,指出语言系统倾向于减少长距离依存,以减轻认知负担[1]。翻译学中的认知转向为依存距离的应用提供了新的价值,实证研究表明,当句子的最大依存距离较长时,译文中的非流利现象显著增加[2]。口译研究亦验证了这一规律,交替传译相较同声传译具有更高的记忆负荷[3]。已有历时研究表明,

依存距离的最小化趋势在长句子中表现最明显,在九种依存关系中都存在[4]。这些研究表明,依存距离不仅能够量化句法组织的跨距特征,也为讨论语言加工与翻译活动中的潜在认知约束提供了一个可操作的分析维度。

2.2. 依存距离在体裁和翻译过程中的差异

依存距离不仅在语言变体之间有所不同[5],还因体裁、翻译过程等因素而呈现显著差异。已有研究发现,不同体裁的英语原创文本(如新闻、学术、小说等)在依存距离上存在差异[6],尤其是,依存距离随句长的增加而缓慢上升,这一趋势在不同体裁中较为普遍[7]。此外,翻译英语文本时,源语和译语之间的差异也显著影响依存距离的分布[8]。

目前,依存距离指标主要应用于跨体裁的比较研究,但对同一体裁内部的深入探讨尚显不足。Tsiplakou & Floros (2013)指出,将文本简单地按照体裁进行划分是一种过于粗略的方法,这种做法容易忽视文本内部的语言特征以及语境因素的作用[9]。Mar 等(2021: p. 733)认为,叙事性文本相较于说明性文本,更容易被记忆和理解[10]。心理语言学的实验研究也支持这一观点:已有眼动追踪实验研究发现说明文比叙事文更能增加读者的凝视时间与加工负担[11]。这些研究暗示:即使是同一体裁,以新闻为例,在不同主题的新闻,或新闻文本中承担不同内容类型的句子之间,其依存距离很可能存在差异。

2.3. 新闻语篇中的依存距离差异

在新闻体裁的研究中, Van Dijk (1986: pp. 84-88)提出新闻宏观结构(Schematic Structure) [12],也称新闻的超结构(Super structure)或新闻的格局(schema) [13],遵循着特定的范畴脉络,包括概要(Summary)、主事件(Main Events)、事件影响(Consequence)、评论(Comment)、历史背景(History)、当前背景(context)等不同内容类型,每个内容类型以不同详细程度循环展开(具体取决于文章篇幅和报道规模)。超结构可用于判定话语类型,超结构中的叙事结构包括5个环节:说明(背景交待并引出人物)、复杂化(初始状态的改变)、解决(变化的结果)、评价和说教(结论) [14]。Choubey 等(2020)对 Van Dijk 的原始理论进行了一些细微的修改,增加了传闻这一新类别(D2),且不区分总结主要故事的新闻导语和其他主要事件(M1)句子,构建了 News Discourse 语料库,并在 Van Dijk 理论的基础上进行了新闻内容类型的标注,将新闻文章中的句子分为八种内容类型,展示了叙述新闻故事时句子的共同内容类型[15]。

3. 研究设计

3.1. 研究问题

本研究以新闻体裁为例,比较分属于不同的主题和内容类型的句子是否会在依存距离上存在分布差异。因此,本研究尝试回答以下两个问题:

- (1) 不同主题(政治、商业、灾难、犯罪)英语新闻句子的标准化依存距离有何分布差异?
- (2) 不同内容类型(事件、前因背景、当前背景、传闻、评价等)英语新闻句子的标准化依存距离有何分布差异?

3.2. 文本选择

本研究选用的语料来自 News Discourse 标注语料库(Choubey *et al.*, 2020),涵盖三大媒体(《纽约时报》、新华社与路透社),本研究选取了其中的668篇新闻,覆盖政治、商业、灾难、犯罪四个主题,每主题各167篇,以消除各主题新闻文本的数量对结果的影响。本研究沿用 Choubey 等(2020)已有的不同内容类型的标注[5],语料库基本信息与标注类别如表1所示。

Table 1. Sources of the News Discourse annotated corpus
表 1. News Discourse 标注语料库来源

语料体裁	形符数	主题	来源	数量	时间	标注类别
新闻	370,763	政治、商业、灾难、犯罪	纽约时报(美国)	68 篇(每主题 17 篇)	2008	事件(M)、前因背景
			路透社(欧洲)	300 篇(每主题 75 篇)	2018~2019	(C1)、当前背景(C2)、历史背景(D1)、传闻(D2)、
			新华社(中国)	300 篇(每主题 75 篇)	2018~2019	评价(D3)、未来展望(D4)

在分析过程中,本文采用 Python3.9.0 的 Spacy 库进行依存句法解析,并根据 Liu (2008)公式 $MDD = \frac{1}{N-S} \sum_{i=1}^N |DDi|$ 计算文本的平均依存距离(Mean Dependency Distance, MDD) [8],同时采用 Liu, Zhu & Lei (2022)提出的标准化依存距离(Normalized Dependency Distance, NDD)以消除句长、根距离差异带来的影响: $NDD = \left\lceil \ln \left(\frac{MDD}{\sqrt{\text{Root Distance} \times \text{Sentence Length}}} \right) \right\rceil$ 。随后结合文本所在的栏目(可分为政治、商业、灾难、犯罪四个主题)以及已有的标注信息,对 10,304 个句子与段落进行七大内容类型的分类,原标注中的主事件(M1)和主事件引发的事件(M2)合并为事件(M),分别统计不同主题与不同内容类型句子的依存距离分布情况,并进行描述性分析。

4. 实验结果与分析

4.1. 不同主题新闻的依存距离

研究表明,四类主题的标准化依存距离存在显著差异。商业新闻和政治新闻的标准化依存距离范围较广,分别为(2.06~3.35)和(2.18~3.38),表明这类文本包含较多极值,呈现更多结构复杂的长句和结构简单的短句,形成较大差异。而犯罪新闻与灾难新闻的标准化依存距离分布较为集中,句法复杂度差异较小(图 1)。

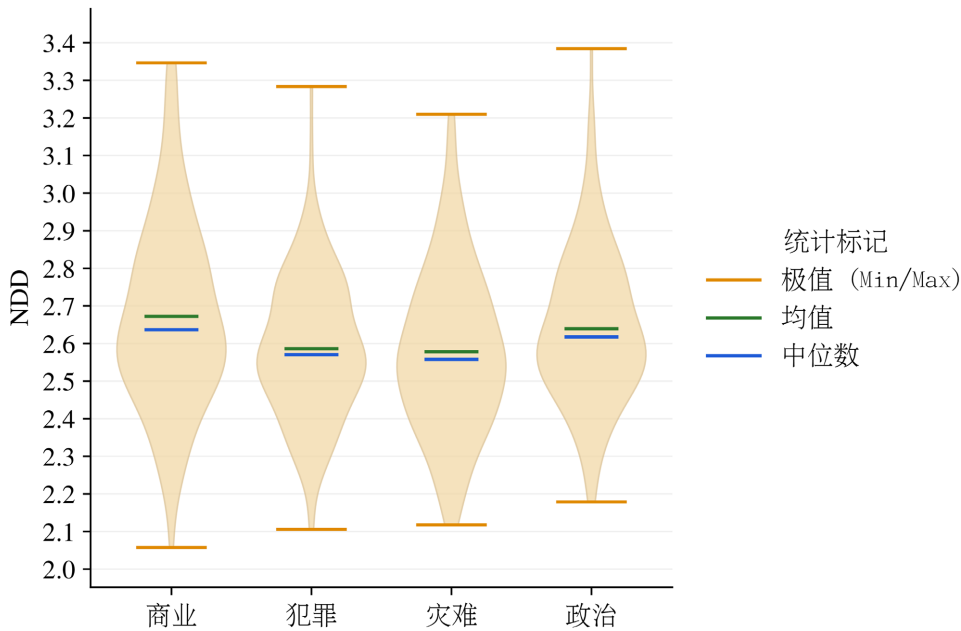


Figure 1. NDD values of news texts across different topics
图 1. 不同主题新闻文本的标准化依存距离 NDD 值

总体而言，四类新闻的标准化依存距离平均值和中位数都集中在 2.6 左右，商业新闻的平均值最高(2.67)，其次为政治新闻(2.64)，而犯罪(2.59)和灾难新闻(2.58)则较为接近且偏低。整体数据表现出右偏特征，均值大于中位数，上端极值较远，揭示出极少数复杂长句对均值的拉升作用。这一现象可能与新闻体裁中高频次的同位语、插入语及复杂句有关。

如表 2 所示，News Discourse (Choubey, 2022)标注语料库中，不同报社刊载的四类新闻揭示出显著的指标差异。与纽约时报和路透社相比，新华社的新闻文本整体在标准化依存距离上的均值较高。具体而言，新华社的商业新闻和政治新闻的标准化依存距离均值约为 2.7，超过了路透社(2.6)和纽约时报(2.41~2.53)。在犯罪与灾难主题的新闻报道中，新华社的标准化依存距离均值略低(约为 2.6)，但依然高于路透社和纽约时报(2.44~2.59)，且差异相对较小。值得注意的是，针对灾难主题的报道，新华社与路透社的新闻文本在标准化依存距离这一指标上几乎没有差异。

Table 2. Mean NDD of four news categories across three newspapers
表 2. 三大报社的四类新闻 NDD 均值

NDD 平均值	商业主题	犯罪主题	灾难主题	政治主题
纽约时报(美国)	2.411	2.458	2.435	2.533
路透社(欧洲)	2.617	2.575	2.594	2.602
新华社(中国)	2.787	2.627	2.595	2.701

4.2. 不同内容类型的依存距离

研究结果表明，承担不同内容类型的句子的标准化依存距离存在显著差异($F(6, 10,297) = 62.55, p < 0.001$)。根据句子所承担的内容类型进行分类后，标准化依存距离的均值范围介于 1.2 至 1.8 之间。具体分析发现，事件(M)的标准化依存距离均值最高(1.71)，其次为未来展望(D4，均值 1.62)，而传闻(D2)的均值最低(1.39) (如图 2 所示)。

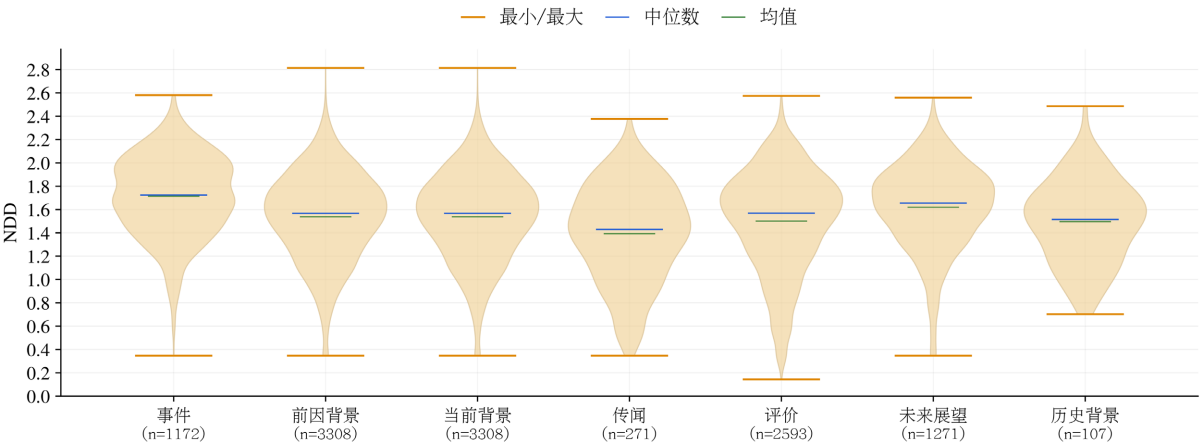


Figure 2. Mean NDD of four news categories across different newspapers
图 2. 不同内容类型的新闻句子的 NDD 值

在介绍与新闻报道主要主题相关的主要事件及其引发的紧随其后的事件(M, 1172 句，均值约 1.71)，或讲述预测性及推演后果的内容(D4, 1270 句，均值约 1.62)时，句子的标准化依存距离的均值和中位数普遍高于其他内容类型。在本研究语境下，承担这两类内容类型的句子在句法依存组织上更为复杂，更

依赖于工作记忆资源的持续调配与支持,或可作为译者加工负担与工作记忆需求差异的一个间接表征。

相较之下,前因背景(C1)主要讲述与主事件紧密相连的近期前置事件及其成因,当前语境(C2)则用于补充当下相关背景信息。两类句子在标准化依存距离上分布更为离散、极端值更多,整体高于历史背景(D1)句子,后者多为发生在主事件之前数月或数年的既往事件。这一差异可能与新闻体裁本身的及时性要求密切相关:对当前事件链的铺陈与解释更易形成依存跨距较大的结构,而历史事件句多以概述性回溯为主,句法组织相对简约。相应地,后者较低的标准化依存距离可能指向较低的翻译加工难度与工作记忆需求,但仍需实验数据进一步验证。

新闻报道还包含反映即时参与者、专家、知名人士及记者或消息来源立场的评价(D3)类句子,其标准化依存距离的均值与中位数处于中等水平。相比之下,传闻(D2)类句子多由匿名或未知来源的不可核实陈述构成,其标准化依存距离均值约为 1.39,低于其他内容类型。该差异或可理解为:传闻类表达往往采用更为简洁直白的句式,依存跨距相对较小;同时,这一结构特征也可能与传闻/轶事在报道中主要承担渲染氛围、补充细节等辅助性功能有关。

以上研究结果发现,承担不同语篇角色语的句子在标准化依存距离(NDD)这一指标上能够显示出较明显的分布差异。当句子用于呈现主事件及其后续发展,或涉及对未来情势的预测与推演时,NDD 往往较高,可能反映出更高的信息整合需求与更大的句法跨距;用于交代背景或引述评价时,NDD 多处于中等水平,呈现相对均衡的句法组织;而传闻类句子通常承担补充细节或营造氛围的辅助性功能,其 NDD 较低,依存关系更为紧凑。总体而言,NDD 或可作为句法复杂度的量化表征,并为讨论新闻语篇中核心信息与附属信息的功能分化提供一个可操作的分析维度。

5. 结语

本研究基于 News Discourse 语料库对英语新闻句子的标准化依存距离(NDD)进行了统计比较。结果显示,主题维度上,商业与政治主题新闻句子的 NDD 分布更为离散、差异较大;相比之下,犯罪与灾难主题新闻句子的 NDD 分布相对集中、差异较小。与此同时,在内容类型维度上,讲述主要事件与未来展望的句子通常具有更高的 NDD,而传闻类句子的 NDD 相对最低。总体而言,这些差异提示不同新闻主题与内容类型可能对应不同的句法组织方式与依存跨距特征。

在翻译研究语境下,这些发现为译者识别潜在结构难度提供了可操作的参照:NDD 可作为句法组织复杂度的量化表征,并可能与不同句子的加工要求存在关联。但需强调,NDD 更宜被视为翻译加工难度与工作记忆需求的潜在代理指标,而非对认知负荷的直接因果解释。基于此,翻译教学与训练可引导译者结合主题与内容类型进行结构预判,例如对事件叙述与推演性表达加强信息分块与结构重组训练,对简短陈述类句式则在尽量保持信息顺序的前提下提升处理效率。此外,由于本研究结论主要来自书面语料统计,其推广至口译等在线加工情境仍需谨慎,并有待结合过程性数据进一步验证。

未来的研究可以从以下几个方面进一步拓展:首先,语料的选择应更加多样化,涵盖更多体裁的文本,或引入双语平行语料库,探讨不同语言之间依存距离的差异,以增强研究的广泛适用性。其次,为提高分析的深度和准确性,可以结合多种其他指标,如句法多样性和信息熵等,与标准化依存距离共同验证,从而深化对依存距离与记忆负荷关系的理解。最后,基于当前研究主要依赖语料统计,未来的研究应考虑引入实验数据支持,采用眼动追踪、键击记录等实验方法,提供对翻译过程中文本处理和记忆负荷的更直接、更细致的观察。这将有助于更全面地揭示影响依存距离与译者记忆负荷之间关系的多重因素。

参考文献

- [1] 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>
- [2] 蒋跃, 蒋新蕾. 最大依存距离对口译中非流利度的影响[J]. 外语研究, 2019, 36(1): 81-88.
- [3] Jiang, X. and Jiang, Y. (2020) Effect of Dependency Distance of Source Text on Disfluencies in Interpreting. *Lingua*, **243**, Article ID: 102873. <https://doi.org/10.1016/j.lingua.2020.102873>
- [4] Liu, X., Zhu, H. and Lei, L. (2022) Dependency Distance Minimization: A Diachronic Exploration of the Effects of Sentence Length and Dependency Types. *Humanities and Social Sciences Communications*, **9**, Article No. 420. <https://doi.org/10.1057/s41599-022-01447-3>
- [5] Xu, H. and Liu, K. (2023) Syntactic Simplification in Interpreted English: Dependency Distance and Direction Measures. *Lingua*, **294**, Article ID: 103607. <https://doi.org/10.1016/j.lingua.2023.103607>
- [6] 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>
- [7] 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>
- [8] Xu, J. and Li, J. (2021) A Syntactic Complexity Analysis of Translational English across Genres. *Across Languages and Cultures*, **22**, 214-232. <https://doi.org/10.1556/084.2021.00015>
- [9] Tsiplakou, S. and Floros, G. (2013) Never Mind the Text Types, Here's Textual Force: Towards a Pragmatic Reconceptualization of Text Type. *Journal of Pragmatics*, **45**, 119-130. <https://doi.org/10.1016/j.pragma.2012.11.004>
- [10] Mar, R.A., Li, J., Nguyen, A.T.P. and Ta, C.P. (2021) Memory and Comprehension of Narrative versus Expository Texts: A Meta-analysis. *Psychonomic Bulletin & Review*, **28**, 732-749. <https://doi.org/10.3758/s13423-020-01853-1>
- [11] Kraal, A., van den Broek, P.W., Koornneef, A.W., Ganushchak, L.Y. and Saab, N. (2019) Differences in Text Processing by Low- And High-Comprehending Beginning Readers of Expository and Narrative Texts: Evidence from Eye Movements. *Learning and Individual Differences*, **74**, Article ID: 101752. <https://doi.org/10.1016/j.lindif.2019.101752>
- [12] Van Dijk, T.A. (1986) News Schemata. In: Freedman, A., Ed., *Studying Writing: Linguistic Approaches*, Sage, 155-186.
- [13] 丁和根. 梵·迪克新闻话语结构理论述评[J]. 江苏社会科学, 2003(6): 199-203.
- [14] 王晓军. 范代克的话语科学观研究[J]. 外语学刊, 2009(1): 90-94.
- [15] Choubey, P.K., Lee, A., Huang, R. and Wang, L. (2020) Discourse as a Function of Event: Profiling Discourse Structure in News Articles around the Main Event. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5-10 July 2020, 5374-5386. <https://doi.org/10.18653/v1/2020.acl-main.478>