

## 大模型驱动的英汉新闻状语功能成分序偏好对比

马义然 2022212784

**摘要：**本文基于大模型标注的英汉新闻可比语料库，尝试从状语性功能块角度探讨了英汉新闻语体的成分序差异，分析其典型位置偏好及分布模式。研究发现：(1)英语新闻更倾向于“核心信息优先”的线性叙事，功能块多后置，而汉语新闻则更偏好“背景优先”的整体铺陈模式，功能块常前置；(2)在 SVO 结构中，英汉新闻均表现出功能块分布差异，但汉语前置倾向更为显著，英文后置倾向则相对温和；(3)多功能块共现时，英汉新闻在顺序选择上均展现出较高灵活性，语序的调整受信息焦点和语用功能的共同驱动。研究揭示了语序既具有系统性偏好，也体现出动态适应性，为英汉信息结构对比研究提供了新的实证支持。<sup>1</sup>

**关键词：**英汉对比；成分序；新闻语体；信息结构；大语言模型

### Contrastive Analysis of Constituent Order Preferences Within Adverbial Roles in English and Chinese News:

#### A Large-Language-Model-Driven Approach

**Abstract:** Based on comparable English-Chinese news corpora annotated by Large Language Model (LLM), this paper attempts to explore the differences in constituent order of English-Chinese news from the perspective of functional chunks with adverbial roles, and analyze their typical positional preferences and distribution patterns. It is found that: (1) English news prefers linear narrative of “core information first”, and functional chunks are mostly post-positioned, while Chinese news prefers overall presentation mode of “background first”, and functional chunks are often pre-positioned; (2) In SVO structure, both English and Chinese news show differences in the distribution of functional chunks, but the tendency of Chinese pre-positioning is more significant, while that of English post-positioning is relatively mild; (3) When function blocks are co-occurring, both English and Chinese news show high flexibility, and the order adjustment is driven by information and pragmatic purposes. The study reveals that word order has both systematic preference and dynamic adaptability, providing new empirical support for contrastive study of English-Chinese information structure.

**Keywords:** English-Chinese Comparison; Constituent Order; News Discourse; Information Structure; Large Language Model (LLM)

### 1. 引言

成分序(constituent order)即词或词组作为功能单元在语法结构体内部的次序(Jacobson, 1996;安玉霞, 2006),是语序(word order)的一类。英汉语句在传递时间、地点和方式等状语性信息时,成分序存在一定差异,成为语言学习、交流和使用的难点。其困难主要表现在:(1)二语学习者难以准确掌握成分位置(李锡江 & 刘永兵, 2020;黄一清, 2010),如例(1)a、例(1)b,在传达同样的新闻信息时语序差异较大,呈现出明显偏好。(2)翻译人员需调整语序以维持信息传递效果(章璐 & 王富银, 2019),如例(1)b、例(1)d,译者应充分满足目标语使用者的语序预期。对于成分序偏好,相关文献只作了一定定性判断(袁毓林, 2006; 吴为章, 1995),难以总结出成分的实际使用规律。

(1)a 端午小长假,山东枣庄台儿庄游人如织,徜徉在大运河畔。(ToRCH2019-A02-7)

(1)b Tourists flocked to Tai'erzhuang in Zaozhuang, Shandong, during the Dragon Boat Festival holiday, wandering along the banks of the Grand Canal. (经过语序调整的译文)

(1)c I spoke with Mr. Cole, Ms. Davids and Mr. Luján this month about the legislation and its prospects. (CROWN2021-A01AA)

(1)d 这个月我与科尔先生、戴维兹女士和卢汉先生谈了这项立法及其前景。(经过语序调整的译文)

<sup>1</sup> 代码、标注结果和完整实验数据已公开于: [https://github.com/rexera/constituent\\_order](https://github.com/rexera/constituent_order)

针对语序与状语语序问题,学界进行过长久深入的探讨,主要包括:(1)语序在定语、状语、从句等语法单元中的体现(胡裕树 & 陆丙甫,1988);(2)影响语序与状语语序的语法和认知因素(Wasow,2003);(3)语际语序异同及其是否存在一定固定规律(Hawkins,1994; 荣晶,2000)。

前人研究呈现出以下特点:(1)分析角度上,倾向于从具体语法单元入手,重视句法分析,而对语义信息传达观照不足(李金满,2010; 曾静涵,2015; Müller, 2021);(2)研究方法上,常常通过感性认知推断归纳出语序差异,但受限于语料规模和主观性,难以提炼出实证支撑的普遍规律(郭中,2012; 双文庭,2016; Zhang & Wang, 2022);(3)研究焦点上,部分研究重视语序差异的影响因素,而对语序本身的分布模式关注较少(Futrell, Levy & Gibson, 2020)。胡裕树和陆丙甫(1988)指出,成分序可以理解为“功能块”之间的关系和排列,不同语言在表达相同语义功能时,会根据各自的句法及语义规则,在“功能块”排列上作出或同或异的选择。而随着 GPT-4(OpenAI, 2023)等大语言模型的发展,对语料中“功能块”的自动标注使得高效分析大规模的语序模式成为可能。鉴于此,本文基于大语言模型的自动标注方法,尝试从状语性功能块角度对英汉新闻语体中的成分序分布规律展开研究,以期从数据驱动揭示英汉成分序偏好差异。

## 2. 研究背景

### 2.1 英汉语序差异的类型学表现

人类语言普遍在句法位置和语序类型之间存在关联性(Dryer, 1986)。如英汉尽管 SVO 结构一致<sup>2</sup>,但在修饰成分顺序上存在差异:英语多将修饰成分置于中心词之后,汉语则倾向于前置(汪康懋 & 肖研,1981)。英汉在传达时间、地点和方式等信息时在句内位置选择上各有偏好,这种成分序差异在定语、状语、从句等不同语法单位中普遍存在(安玉霞,2006)。纵观学界论述,对英汉语序差异的划分较为粗略,对于全面认知语序差异帮助有限。虽然学界普遍认为英汉语序存在类型学差异,但在诸如状语位置灵活选择上仍然基于直观的类型学推论,较难提取典型语序分布模式(丁志斌,2014)。

针对语序与句法的关系,李金满(2010)发现汉语虽为 SVO 语序,但在关系从句、比较结构等方面呈现 OV 倾向,形成在全球语言中不典型的混合语序特征。金立鑫(2019)通过语序指数测量也证实了汉语的 VO 特征体现在动宾结构中,而在定语、介词位置等方面表现出 OV 特性。此外,汉语的信息推进依赖“话题链”,信息随之层层展开,允许状语出现在句首,而英语更严格遵循主谓宾的逻辑顺序(荣晶,2000)。上述论断多基于内省式思辨,尚缺乏语料的实证检验。

### 2.2 语义与信息结构动因

人类语言在词序演变上存在“单向变化”规律,在提升句法结构解析处理效率上有普遍倾向(Hawkins, 1994),因而英汉在语篇信息推进中存在共性,如 Liu & Guo(2014)和郭骅(2022)分别在广播和学术语篇中结合 Halliday(1985)的主位和述位(theme and rheme)理论,发现了一系列英汉共有的主位推进模式。然而,英汉在演化中逐渐产生分野:英语“开门见山”的结构将重心前置,而汉语则多以“画龙点睛”的结构在句末揭示信息重心(贾德霖,1990)。英语语序受到句法结构、移位操作和“线性修饰”原则等因素的约束,信息通常按照“已知-未知”或“弱-强”的顺序排列(Chamonikolasová, 2009; Lee, 2019)。此外,存在“依存局部性”“外置”等认知

<sup>2</sup> 英语句型通常划分为 SV、SVO、SVC、SVA、SVOO、SVOC、SVOA 七类 (Quirk, Greenbaum, Leech, & Svartvik, 1985), 其中 SVO 句型最为常见。而对于中文存在不同解释:一种观点认为主谓宾(SVO)是现代汉语的基本语序(王力,1980);另一种观点认为中文作为字本位语言,以 SVO 框架概括有所失当,应使用“主位-述位”(赵元任,1968; 张伯江,2011),本文不作进一步讨论。为了便于对比,我们权且认为英汉在 SVO 结构上一致,是最普遍的句型。

处理原则以降低句法复杂性和认知处理负荷(Huddleston & Pullum, 2002; Futrell, Levy & Gibson, 2020), 如 Hannay(1991)提出的信息管理策略, 包括“全新模式”“话题模式”“反应模式”等框架, 强调了其与成分频次、认知固化的关联性。汉语由于缺乏形态变化, 语序承担了更多的语法功能, 受到句法结构、信息结构、认知象似性等因素的多重影响, 满足整体性和语境化的语义需求(安玉霞, 2006), 具有信息焦点的句末偏向、高度省略已知信息, 以及“旧-新”信息模式等特点(张炼强, 1997; 荣晶, 2000)。对于语序的灵活性, 学界存在争议。部分约束句法研究和配价研究认为“语序-功能”对应关系相对固定, 信息位置的安排受到信息流动、修饰语范围或语义角色控制(Ross, 1984; Xing, 2015)。但也有学者强调词序的动态性, 如在汉语从属结构中常呈现“头后置”模式, 受语用需求和语义兼容性驱动(Yurayong & Sandman, 2023)。鉴于此, 有必要基于大规模的英汉可比语料, 从语序偏好差异入手, 深入分析新闻语体中成分序差异的语义和信息结构动因。

### 2.3 语序分布的模式化趋势

在前人的英语语序研究中, Stolterfoht(2013)发现状语位置并非随机, 往往受到信息结构原则调控而产生位置偏好。Biber 等人(1999)也发现, 不同类型的状语在句中的分布具有一定倾向性, 语体差异也会显著影响语序分布偏向。Zhang 与 Wang(2022)基于依存树库探讨了汉语时间状语从句的位置分布及其句法和语义特性。总体来看, 前人的研究揭示了英汉成分序在语料中的模式化倾向。传统语料研究多依赖手工标注 (Ai, Zhu, Gao, & Yin, 2016), 因数据量有限, 难以提炼出更广泛的语序分布规律。Travis(1984)通过参数化句法理论中的“领域邻接条件”解释了动词与状语的排列约束, 但缺乏实证验证; Futrell、Levy 和 Gibson(2020)提出的“依存局部性”模型尽管能一定程度上通过依存长度预测词序, 却有无法显式提取规律的不足。随着大语言模型的应用, Kojima 等人(2022)提出的零样本思维链(0-Shot CoT)方法启发了大模型在大规模语料自动标注上的可能性, 为成分序差异的探索性数据分析提供了技术手段。为更全面揭示英汉成分序分布规律, 本文选择英汉新闻语体可比语料库, 使用大模型自动标注进行分析探讨。

## 3. 研究设计

### 3.1 研究问题

针对新闻语体, 本文尝试从状语性功能块入手, 探究英汉成分序在以下三方面的语内偏好与语际差异:

- (1)不同类型功能块的语序偏好;
- (2)SVO-功能块组合中的成分序趋势;
- (3)多重功能块同时出现时的顺序倾向。

### 3.2 语料来源与标注

本文使用 CROWN2021(Sun, Xu et al., 2022)新闻子库和 ToRCH2019(李佳蕾, 孙铭辰, & 许家金, 2022)新闻子库, 这两个子库包含了广泛的新闻语体, 剪表性强; 都采用了 BROWN 家族语料库范式, 具有可比性。CROWN2021 和 ToRCH2019 分别是平衡美式英语和平衡现代汉语语料库, 由北京外国语大学在 2021 年和 2019 年编制, 库容各约 100 万词。本文使用的英文子库共 90,131 词, 中文字库 88,539 词, 相关统计数据见表 1。本文基于状语语义功能, 参考 Quirk 等人(1985)对状语的划分, 将成分功能块分为<时间>、<地点>、<方式>、<原因>、<结果>、<条件>、<目的>、<让步>八类, 分别标注为<time>、<place>、<manner>、<cause>、<effect>、<condition>、<purpose>和<concession>, 主谓宾分别标注为<S>、<V>和<O>, 便于控制在英汉中相同的 SVO 结构。对于从句, 处理为多个 SVO 结构。对于无法归类到以上功能块的字词或词组规定不标注, 允许出现空缺。

综合效率与可靠性，采用目前综合性能领先的 GPT 4o 模型 API 接口<sup>3</sup>，结合少样本提示(Few-Shot Prompting)策略引导模型按标签体系标注<sup>4</sup>。实际标注示例和抽象版本见例(2)，标注统计结果见表 2 - 表 4。

- (2)a <time>端午小长假</time>, <place>山东枣庄台儿庄</place><S>游人</S><manner>如织</manner><V>徜徉</V><place>在大运河畔</place>。
- (2)b <S>I</S><V>spoke with</V><O>Mr. Cole, Ms. Davids and Mr. Luján</O><time>this month</time><purpose>about the legislation and its prospects</purpose>.
- (2)c <time><place><S><manner><V><place>
- (2)d <S><V><O><time><purpose>

表 1 英汉新闻语体子库原始数据统计

子库	文本数	形符	类符	类符/形符比
CROWN2021-A	220	90131	12502	0.139
ToRCH2019-A	44	88539	12341	0.139

表 2 英汉新闻语体子库标注数据统计

子库	行数	总标签数	行平均标签数	总功能块标签数	行平均功能块标签数
CROWN2021-A	2649	17865	6.74	5846	2.21
ToRCH2019-A	1735	26162	15.08	8389	4.84

表 3 英文新闻语体功能块分布

标签	频数	频率
<time>	1326	0.23
<place>	1129	0.19
<effect>	1098	0.19
<purpose>	765	0.13
<manner>	540	0.09
<cause>	476	0.08
<concession>	356	0.06
<condition>	156	0.03

表 4 中文新闻语体功能块分布

标签	频数	频率
<manner>	3293	0.39
<place>	2362	0.28
<time>	2050	0.24
<effect>	235	0.03
<concession>	168	0.02
<purpose>	112	0.01
<condition>	93	0.01
<cause>	76	0.01

<sup>3</sup> [https://api.openai.com/v1/fine\\_tuning/jobs](https://api.openai.com/v1/fine_tuning/jobs), 模型名: gpt-4o-2024-08-06

<sup>4</sup> 对于大模型自动标注的有效性和局限性，学界已有探讨。尽管其能达到甚至超越众包标注员能力(Alizadeh et al., 2024)，扩展性强，但仍然存在偏见、幻觉等问题，且在专业知识上能力有限，而提升数据多样性、引入模型自校正机制等方法能够缓解这些问题(Tan et al., 2024)。鉴于此，本文引入了少样本提示，加入原句与手工标注结果的对照 3 例，并在提示词中对每一类功能块作出举例说明。尽管无法在逐句标注质量上达到人类专家水平，但该方法能够保证宏观统计层面的可用性。标注过程平行进行两次，以模拟双标注员策略，独立完成同一批次数据的标注后计算标注一致性，经计算高于 0.95。

### 3.3 统计分析

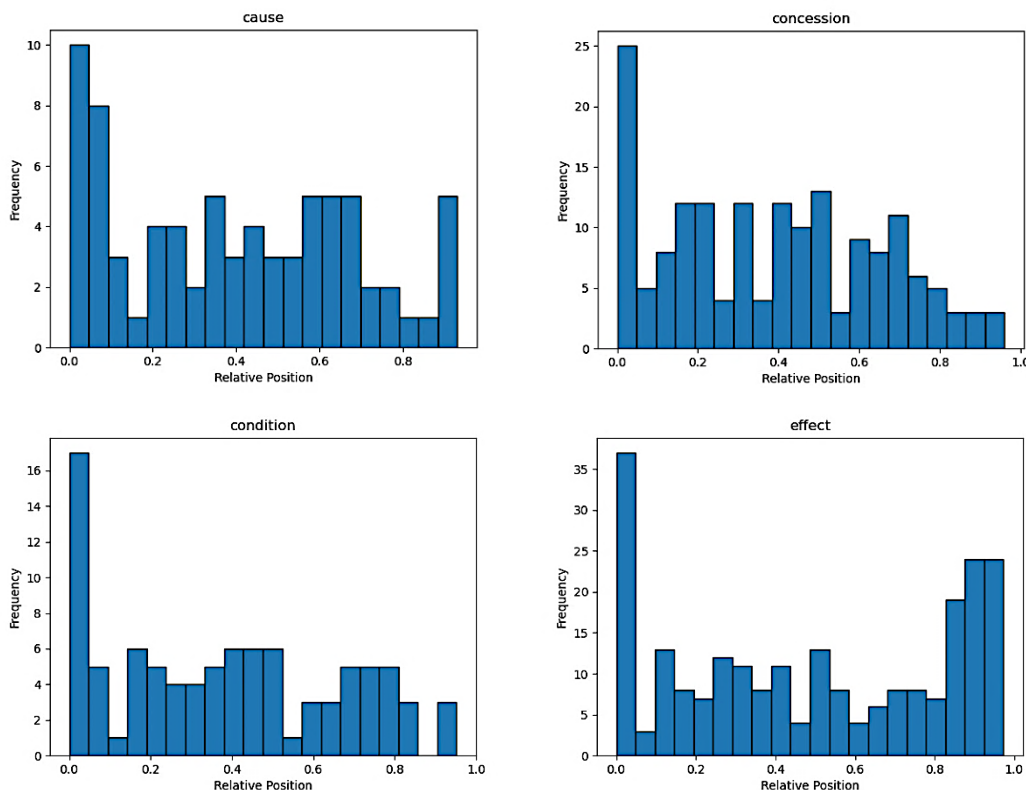
本文基于统计方法，探讨英汉新闻语体中功能块的分布特性及其语序规律，以期揭示两种语言在成分序上的偏好差异。构式语法(Goldberg, 1995)指出，构式与其内部成分存在认知固化，源于在语言高频输入中形成的范畴化与模式化。而成分序与信息结构之间的紧密关联性同样可以近似视为一种偏好的固化与泛化。在这一过程中，成分序与其语义功能逐步适配；在统计层面，实现相同信息功能时，成分序偏好与均匀分布假设呈现出显著差异。基于上述视角我们可以认为，若功能块在句中的相对位置与均匀分布假设相比，在语内、语际呈现显著差异( $p < 0.05$ )，则说明其位置偏好在语内具有一定的系统性和稳定性，在语际存在明显差异。为此，本文结合条件概率分析、卡方检验、独立样本 t 检验与马尔可夫链转移矩阵分析，试图探讨语内语际的：(1)不同类型功能块句内位置偏好；(2)功能块 - SVO 框架组合规律；(3)多重功能块排列模式及分布倾向。

## 4. 研究发现

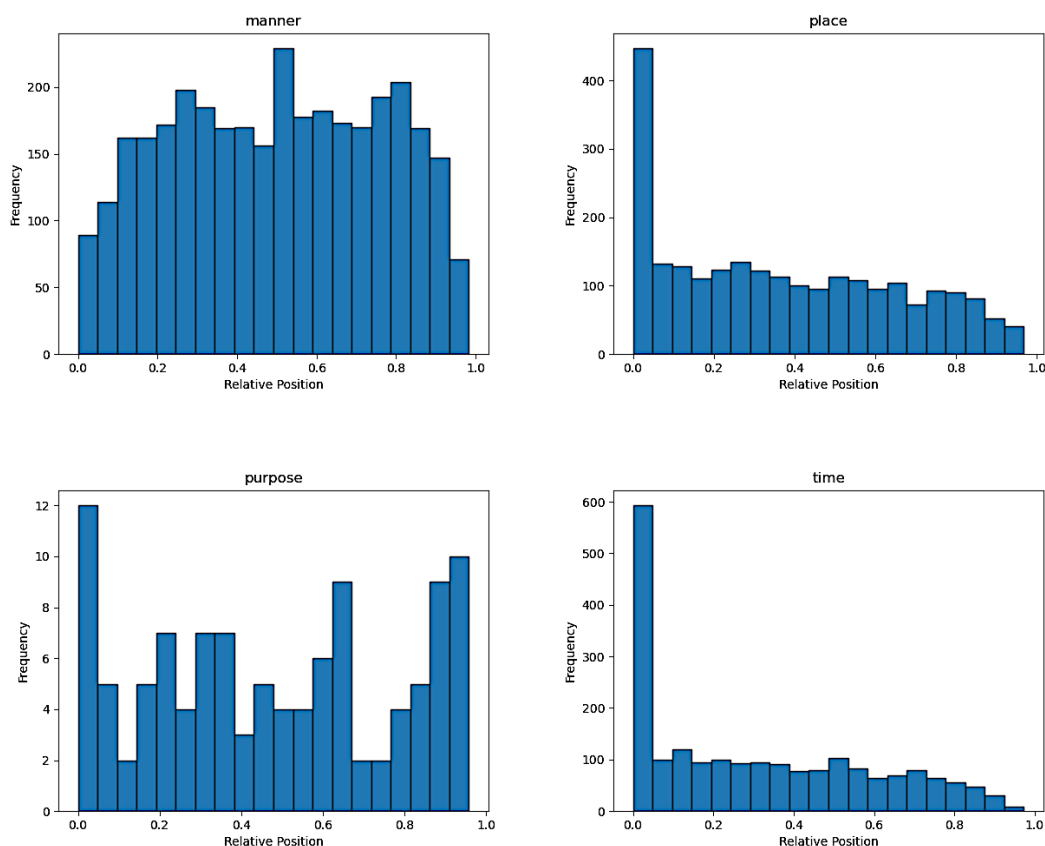
### 4.1 不同类型功能块的句内分布特点

我们首先对于语内的每一类功能块分布作与均匀分布的卡方检验(详见表 5)。中文新闻中，各类功能块在句内呈现整体显著的前置倾向(见图 1)。在针对各类功能块的卡方检验中，<时间>和<地点>高度集中于句首，在叙述中先行交代时间与空间背景。此外，<条件>、<让步>与<原因>也主要位于句首，通过开篇设置引导句子逻辑展开。然而，其他功能块如<方式>、<结果>和<目的>的分布则相对不显著。从直方图可见，其句内位置未表现出明显偏好。这些分布偏好揭示了中文新闻语体在信息组织上的特点：显著前置的<时间>、<地点>、<条件>和<原因>等功能块为叙事搭建清晰的背景框架，而其他功能块则灵活分布，以适应不同语境需求，这与汉语注重整体性和信息铺垫的特点密切相关。

图 1 中文新闻语体中各类功能块相对位置分布

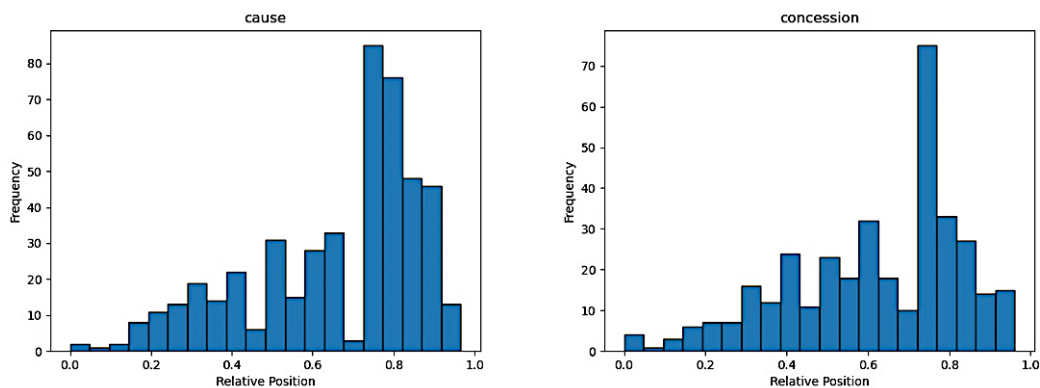


续图 1

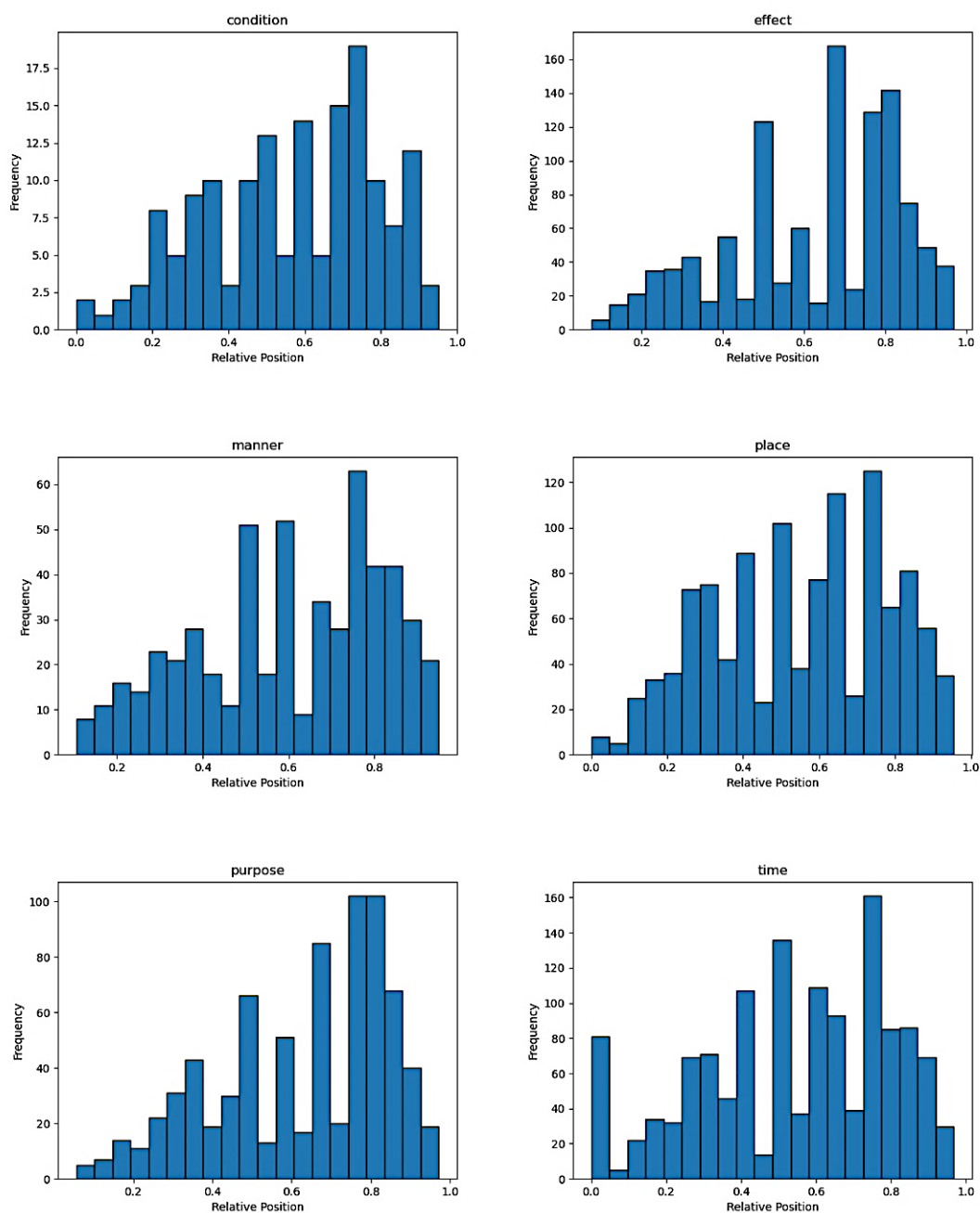


英文新闻功能块的句内分布也存在显著偏好，整体呈现出与中文相反的后置倾向(见图 2)。<结果>、<目的>、<原因>和<让步>四类强调条件或逻辑关系的功能块通常置于句尾，凸显对核心信息的解释性补充。<时间>和<地点>同样显著集中于句尾，体现了英语注重时间和空间信息线性推进的特性。相比之下，<条件>的分布差异性较弱，但仍在统计学意义上表现出一定的句中偏后倾向。这些分布模式展现了英文新闻在信息组织上的鲜明特征，即功能块后置以让位于核心焦点信息，同时保持叙述的线性发展逻辑，使得事件核心在句子前部得到更大的认知优先权。

图 2 英文新闻语体中各类功能块相对位置分布



续图 2



在语际对比中，所有功能块在  $t$  检验下均存在  $p < 0.05$  的显著差异(详见表 5)。<时间>和<地点>的分布差异最为显著，中文通常在开头为叙述设置时空背景，而英文则倾向将其后置，与线性信息推进模式相符。<方式>信息在汉语中处理灵活，而英语则表现出更强的后置偏好。传达逻辑关系的功能块也存在显著差异，汉语强调逻辑铺垫，英语则重视逻辑的后置展开。总体来看，英汉两种语言在功能块的句内分布上表现出稳定而系统的偏好和模式。

表 5 语内功能块卡方检验与语际 t 检验结果

功能块	中文 ( $\chi^2, p$ )	英文 ( $\chi^2, p$ )	语际 ( $p$ )
<time>	612.81, <0.001	18.99, <0.001	<0.001
<place>	332.63, <0.001	31.29, <0.001	<0.001
<condition>	10.18, 0.001	4.14, 0.042	<0.001
<concession>	17.32, <0.001	40.07, <0.001	<0.001
<cause>	5.99, 0.014	94.76, <0.001	<0.001
<manner>	0.13, 0.72	42.59, <0.001	<0.001
<effect>	0.018, 0.89	135.62, <0.001	<0.001
<purpose>	0.28, 0.60	100.24, <0.001	<0.001

#### 4.2 功能块与 SVO 框架的组合模式对比

为了探究功能块与 SVO 的组合关系，我们首先构建了每类功能块分别出现在 S、V、O 前后的条件概率。以<时间>出现在<S>前的条件概率为例，我们将“句中包含<时间>功能块”作为条件事件，将“<时间>出现在<S>前”作为目标事件，则该条件概率可视作“欲实现<时间>功能的条件下，<时间>位于<S>前的概率”。

对中文的分析显示，中文的功能块分布更为集中，明显呈现出关于 S、V、O 高度统一的前置偏好(见表 6)。从差异显著的功能块来看，<时间>、<地点>和<原因>在动词和宾语前的分布尤为突出；差异次显著的功能块，如<方式>和<条件>，在主语前后的分布更为灵活，可能与功能块在句中承担的修饰作用有关，例如<方式>既可以修饰主语，又可以修饰动词，因此在主语前后分布较为均衡。总体而言，中文功能块与 S、V、O 的组合模式呈现出高度统一的前置偏好，而<方式>和<条件>在主语前后更为灵活，这种分布差异反映了中文信息组织中背景优先和叙述灵活结合的特点。

表 6 中文功能块 - SVO 分布条件概率表

功能块	S 前	S 后	V 前	V 后	O 前	O 后
<manner>	0.57	0.43	0.88	0.12	0.93	0.07
<place>	0.72	0.28	0.94	0.06	0.94	0.06
<effect>	0.58	0.42	0.67	0.33	0.72	0.28
<time>	0.81	0.19	0.98	0.03	0.97	0.03
<purpose>	0.61	0.39	0.81	0.19	0.84	0.17
<cause>	0.78	0.22	0.96	0.04	0.92	0.08
<condition>	0.69	0.31	0.93	0.08	0.89	0.11
<concession>	0.77	0.23	0.96	0.04	0.95	0.05

相比之下，英文的功能块分布则表现出更大的灵活性，前后位置的条件概率差异较小(见表 7)。<结果>、<原因>和<目的>在动词和宾语之后的分布稍多于之前；<时间>、<地点>和<方式>的分布则更加均衡，前后差异较小，可能反映了这些功能块既可以用于修饰句子前部的信息，也可以对句尾的动作或宾语进行补充说明，体现了灵活性。总体来看，英文的功能块分布相较中文更为均匀，<结果>、<原因>和<目的>表现出一定的后置倾向，而 <时间>、<地点>和<方式>则分布更加灵活，前后差距较小。这种分布模式反映了英语新闻对功能块位置的多样化安排，为叙述提供了更多的自由度，同时也体现了其逻辑递进和句法核心突出的特点。



表 7 英文功能块 - SVO 分布条件概率表

功能块	S 前	S 后	V 前	V 后	O 前	O 后
<manner>	0.37	0.63	0.40	0.60	0.32	0.68
<place>	0.39	0.61	0.44	0.57	0.37	0.63
<effect>	0.34	0.66	0.33	0.67	0.31	0.69
<time>	0.40	0.60	0.44	0.56	0.41	0.59
<purpose>	0.35	0.65	0.35	0.65	0.29	0.71
<cause>	0.32	0.68	0.31	0.69	0.27	0.73
<condition>	0.45	0.55	0.47	0.53	0.38	0.62
<concession>	0.37	0.63	0.40	0.60	0.33	0.67

其次，我们各提取了两组高频功能块 - SVO 分布模式(详见附录表 1、表 2)。在具体中文语料中，<place><S><V><O> 是一个典型的分布模式，如例(3)a 的多个连续<地点>出现在主语之前，设定了丰富而具体的空间背景。这种排列强化了中文叙述中常见的“从背景到主体”的逻辑结构。<time><place><S><V><O> 模式则展示了时间与地点功能块的联合前置倾向，如例(3)b，时间和地点功能块连续出现于句首，明确了叙述的时空范围。这种组合模式进一步说明中文新闻在功能块排列上的规律性：通过双功能块强化信息背景，为主句的叙述提供充分的铺垫和层次。另一个典型模式是 <place><S><manner><V><O>，如例(3)c 中的<地点>位于句首引出场景，<方式>插入主谓之间，用于修饰动作的方式。这种灵活组合体现了中文新闻在功能块分布上的灵活性，与<方式>在主语前后均有分布的结论相一致。通过这些典型模式可以看出，中文功能块与 SVO 的组合不仅体现出功能块的整体前置偏好，还展现出联合前置特点及一定的灵活性，部分揭示了中文新闻在信息背景交代和叙述逻辑中的规律。

- (3)a <place>从高山雪原到万里海疆</place>，<place>从边陲重镇到都市乡村</place>，<place>食堂、洗澡间、阅览室<place>……<S>基层民警的工作生活环境</S><V>一直以来都是</V><O>党中央、国务院和公安部党委高度重视和关心的问题</O>。(ToRCH2019-A01-38)
- (3)b <time>2018 年以来</time>，<place>福建省德化县</place><S>纪委监委</S><V>先后查办了</V><O>某国企两名财务人员贪污巨额公款和一名职务犯罪公职人员外逃 18 年的案件</O>。(ToRCH2019-A10-33)
- (3)c <place>活动现场</place>，<S>嘉宾</S><manner>以阵阵掌声</manner><V>表达了</V><O>对建设美丽中国与共建美丽地球家园宣示的热烈回应</O>。(ToRCH2019-A09-13)

在英文新闻中，功能块的分布充分体现了其后置补充信息的特点。<S><V><O><time> 是一个高度体现功能块后置特点的分布模式，如例 4(a)中，SVO 首先直接传递关键信息。随后，<时间>徐徐展开，补充具体背景。这种安排使读者能快速抓取核心信息，而时间背景的信息则在句尾逐步展现，避免了过长的句首修饰干扰主干结构。<S><V><O><place> 模式，如例 4(b)，同样以 SVO 迅速呈现事件主干，而<地点>则通过后置逐层补充了事件的具体空间信息。这种句式同样使核心事件叙述直观紧凑，同时空间背景在句尾逐渐展开，既明确了语境又不干扰核心信息。<S><V><O><time><cause> 模式则体现了功能块对逻辑关系的补充作用，如例 4(c)的<时间>和<原因>依次在句尾补充，这种逐步后置尽管与严格逻辑顺序不符，但在保持句子主干简洁性的同时，层

层递进地丰富背景和逻辑,使信息组织易于接受。这些例句清晰地反映出,英文新闻通过功能块后置实现了“快速抛出主谓宾信息,随后逐步细化背景”的表达策略。主谓宾的前置和功能块的后置互相配合,使叙述在逻辑推进中保持紧凑高效,同时为复杂背景信息提供了展示空间。这一特点不仅提升了句子的可读性,也强化了信息的分层组织。

- (4)a <S>Neal</S><V>will assume</V><O>the helm of the nation's immigration court system</O><time>at a time when the Biden administration is wrestling with increasing arrivals at the US-Mexico border and seeking to reform a system that's been plagued with issues</time>. (CROWN2021-A03AA)
- (4)b <S>Cuyjet (pronounced SOO-zhay)</S><V>also has</V><O>a video piece, "For All Your Life Studies,"</O><place>in an exhibition called "In Practice: You may go, but this will bring you back," at Sculpture-Center in Long Island City</place>. (CROWN2021-A43AD)
- (4)c <S>The White House</S><V>announced</V><O>a series of sanctions</O><time>on Friday night</time><cause>for its forced landing of a Ryanair commercial flight and the subsequent removal and arrest of opposition journalist Roman Protasevich</cause>. (CROWN2021-A04AE)

### 4.3 多重功能块的排列顺序与分布规律

我们首先分析了功能块的多重组合模式,统计了多个功能块共现的高频组合(详见附录表 3、表 4)。各从前 10 位来看(见表 8、表 9),中文和英文均以<时间><地点>为最高,说明两种语言在功能块的排列组合上具有一定的相似性,而总体上两种语言均呈现出高度灵活性。在中文中,<地点><方式>和<地点><时间>频数高,而英文则在高频组合中更集中于因果和目的功能块,如<时间><原因>和<结果><时间>。尽管如此,两种语言并未显示出显著差异性,这可能与语料的规模和范围有关。值得注意的是,包含三个功能块的组合在统计中也有出现,但普遍缺乏语际统计显著性,进一步说明了组合模式的灵活性并未集中于特定复杂模式。

受马尔可夫链启发,我们进一步对功能块之间的条件转移关系概率进行分析(详见附录表 5、表 6),构建了转移矩阵热力图(见图 3)。

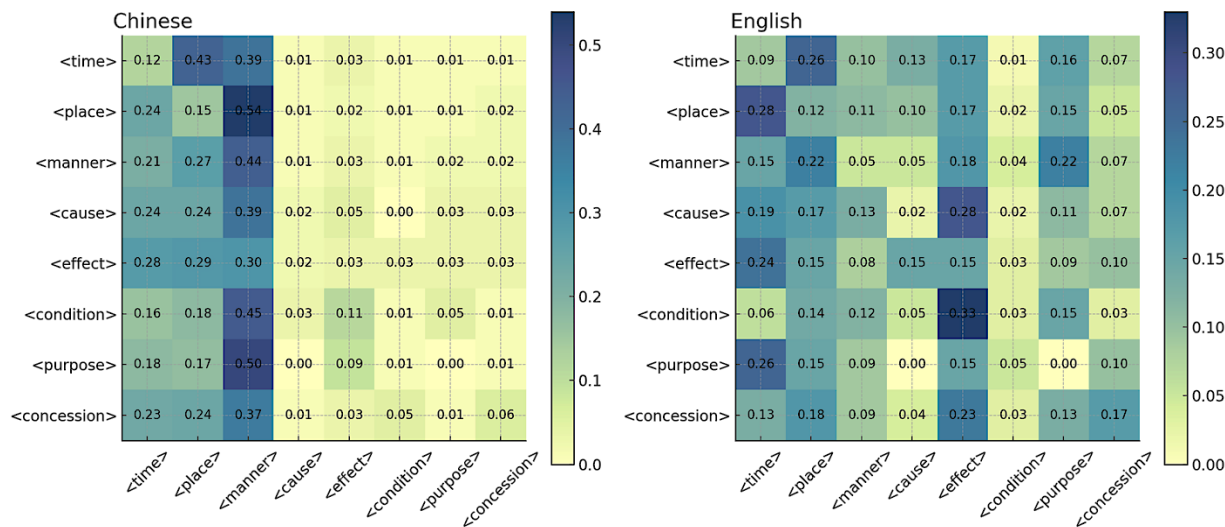
表 8 中文频率前 10 的多重功能块组合

次序	组合模式	频数
1	<time><place>	415
2	<place><manner>	345
3	<place><time>	200
4	<manner><place>	183
5	<time><manner>	155
6	<manner><time>	129
7	<time><place><manner>	55
8	<effect><time>	35
9	<place><manner><place>	30
10	<manner><place><manner>	28

表 9 英文频率前 10 的多重功能块组合

次序	组合模式	频数
1	<time><place>	152
2	<place><time>	135
3	<time><cause>	88
4	<effect><time>	82
5	<time><purpose>	73
6	<effect><cause>	69
7	<purpose><time>	64
8	<time><effect>	63
9	<place><purpose>	58
10	<place><effect>	55

图 3 功能块转移矩阵热力图



通过比较转移概率前 10 位的结果(见表 9、表 10)，我们发现中文的转移以<地点>到<方式>为最高概率，达到了 0.55，同时<目的>到<方式>和<条件>到<方式>也位列前 3。相比之下，英文则显示了更高的<条件>到<结果>(0.33)和<原因>到<结果>(0.28)的转移概率。而从热力图可见，中文<时间>、<地点>和<方式>功能块在转移上表现得尤为强势，这表明中文更倾向于通过时间、地点、方式功能块来主导信息的组织和叙述逻辑，反映了中文新闻语体在叙述中的场景感和背景铺垫的显著特点；相比之下，英文的功能块转移分布更加均匀，以<时间>和<结果>稍显强势，表明英文更注重通过逻辑关联来推进信息的线性展开。

总的来说，两种语言在多重功能块的排列和分布上存在一定的倾向性，但这种偏好在整个使用的灵活性当中并不足以形成具有普遍意义的结论。综合功能块组合模式与转移矩阵分析，我们推测，这种灵活性可能反映了两国语言的信息组织并未显著限制功能块的搭配，而是允许其根据语境需求自由组合。

表 10 中文频率前 10 的功能块转移关系

次序	转移关系	转移概率
1	<place> to <manner>	0.54
2	<purpose> to <manner>	0.50
3	<condition> to <manner>	0.45
4	<manner> to <manner>	0.44
5	<time> to <place>	0.43
6	<time> to <manner>	0.39
7	<cause> to <manner>	0.39
8	<concession> to <manner>	0.37
9	<effect> to <manner>	0.30
10	<effect> to <place>	0.29

表 11 英文频率前 10 的功能块转移关系

次序	转移关系	转移概率
1	<condition> to <effect>	0.33
2	<cause> to <effect>	0.28
3	<place> to <time>	0.28
4	<purpose> to <time>	0.26
5	<time> to <place>	0.26
6	<effect> to <time>	0.24
7	<concession> to <effect>	0.23
8	<manner> to <place>	0.22
9	<manner> to <purpose>	0.22
10	<cause> to <time>	0.19

## 5. 讨论

### 5.1 信息组织的功能对位与深层驱动

统计结果表明, 英汉新闻语体在状语功能成分序分布上呈现出显著偏好, 主要差异体现在实现相同信息功能时核心事件和背景信息的呈现顺序上。英文通常采用“核心信息优先”的线性叙事模式, 在第一时间搭建起事件的骨架结构, 随后补充展开时间、地点等背景, 体现了 Dryer(1986)提出的“话题标记性假设”, 即在语序安排上, 英语倾向于将语法主干置于显著位置, 以增强信息的显性表达。这一模式使得叙述的焦点和逻辑主线更加清晰, 从而增强了信息的直接性和条理性, 使得核心内容更易识别和记忆。相比之下, 中文的叙事结构更偏向“背景优先”的整体铺陈模式, 常见的表达方式是通过交代时间、地点或方式等情境信息为叙事提供框架支持, 而核心事件则往往位于句尾作为“点睛之笔”, 以信息的递进性和情境化突出叙事的连贯性和完整性。这一特征也验证了荣晶(2000)提出的汉语“背景-焦点”结构, 即汉语倾向于通过情境铺垫为核心事件提供支持, 并通过渐进式的信息展开引导读者进入叙述逻辑。上述成分序偏好差异从功能块分布的实证角度支持了汪康懋和肖研(1981)对于英汉语序特性的描述, 即英语更加注重信息的集中性, 而汉语则更倾向于语义的整体性。

“核心信息优先”与“背景优先”的差异不仅反映了两种语言的句法结构表层特征, 还揭示了认知模式、语义表达以及文化传统等深层机制上的不同, 具有多维的理论意义和应用价值。

从句法结构角度来看(Ji, 2015), 英语倾向于形合, 句法关系以严格的形式标记和相对固定的规则表现出来, 确保信息的线性推进和逻辑连贯。在实际语料中, <时间>、<地点>等背景信息多后置, 使得 SVO 框架的核心事件在逻辑主线上更加突出。相比之下, 汉语重意合, 句法关系更多地以语义关联和语境推导来体现, 语料中的<时间>和<地点>功能块常位于句首, 构建出完整的叙事背景后才逐步展开事件核心, 突出了汉语对整体语义和上下文联系的依赖。认知与文化传统的差异进一步加深了英汉语序的不同偏好。英语的认知模式强调“物我两分”, 以清晰的主语和谓语为核心, 明确区分事件参与者和背景信息, 反映了英语文化中个体中心和逻辑推演的思维传统, 因而在叙述中优先突出核心事件的行动逻辑, 并将背景信息后置以逐渐渲染事件重要性。与之相比, 汉语的认知模式更注重“天人合一”的整体性观念(贾德霖, 1990; 曾静涵, 2015), 在文化上更注重环境与行为的交互关系, 叙述中倾向于通过先行交代背景信息来展现事件与情境的互动。例如在对重大事件的报道中先营造时间和空间的宏大叙事, 再引出具体的行动或结果(Scollon, 2000)。时间、地点等背景成分在句首的高频分布铺陈背景信息, 为核心事件的呈现构建更为完整的语义网络和语境信息。

需要注意的是, 本文所观察的这些特征在很大程度上受新闻语体本身的限制, 未来研究可以通过拓展语料范围和语体类型, 探讨英汉成分序模式的跨语体一致性与变化性, 为语言对比研究和跨文化交流提供更深刻的理论支持。

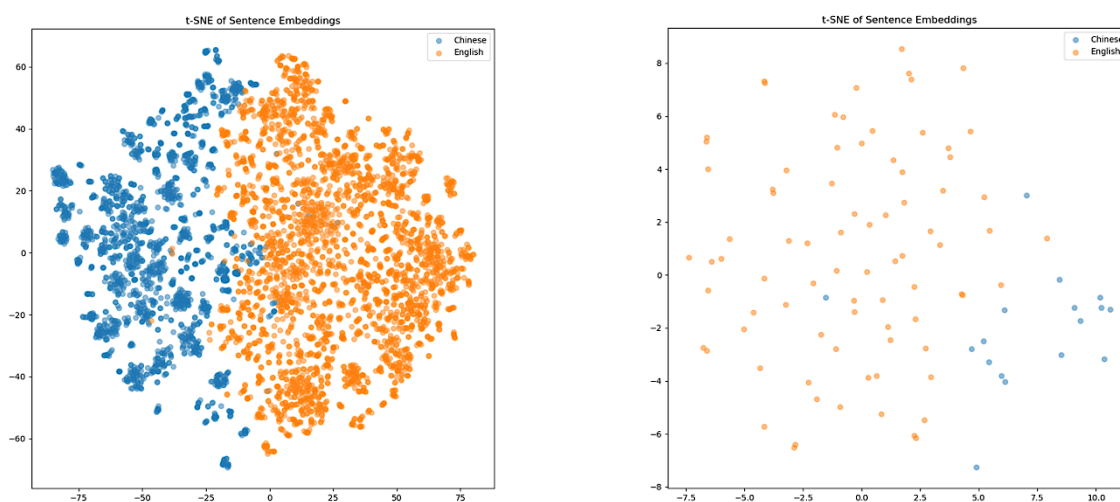
### 5.2 语义对语序的影响探究

前文的结论主要建立在对功能块标签的统计分析之上, 而基本上将语序模式与语句和功能块的具体语义解耦。虽然从功能块标签分布中能够得出两种语言的语序偏好, 且我们预设功能块分布差异与语义关联较小, 主要体现为一种思维构式, 但语料中的语义差异对语序差异的潜在影响亦不可完全忽视。以往对大规模语料的语义分析由于耗时耗力, 难以高效实现。而现今的大模型通过在海量语料上进行训练, 其参数中已蕴含了对语义

的深度表征,使得我们能够将语料嵌入为高维向量,从而利用这些特征向量开展高效的语义研究。为此,我们使用 MiniCPM-Embedding 模型<sup>5</sup>对语料作了选择性嵌入,并对得到的高维特征向量进行分析。

我们首先分别对两个新闻子库以及英汉两种典型语序结构的子集(<time><S><V><O>和<S><V><O><time>)进行嵌入,取中心向量后使用 t-SNE 降维可视化并计算余弦相似度<sup>6</sup>。其中,英汉新闻子库完整语料的余弦相似度为 74.07%,而<time>功能块相关子集的余弦相似度为 62.04%,表明使用的英汉新闻语料在整体语义上存在一定差异,但总体较为相似。而由图 4 可见,在局部语义上语料均有明显差异。结合前文语序偏好相关结论,尽管语料整体语义差异不大,但局部仍然存在差异,因此我们只能认为语义对英汉新闻语体成分序差异影响相对较小,但不可忽视语义差异尤其是局部差异的作用。由此可见,语义特征的分布在一定程度上对语序选择形成了潜在影响,虽然整体影响有限,但在特定情境下可能具有重要意义。

图 4 英汉新闻子库语义特征分布(左)与两子集语义特征分布(右)



我们进一步使用相同方法对八类功能块对应的具体语料内容分别计算,各功能块在整体语义相似度上均接近或超过 85%(见表 12),<时间>、<方式>、<原因>和<让步>功能块在局部分布上仍然具有较高的重合度,而其他类别相较而言存在细微的分布差异(见图 5)。由此可知,功能块本身的语义上差异不大,可以认为功能块语义对功能块分布偏好的贡献较小;然而在功能块层面,我们同样不能认为语义与语序完全无关,功能块的整体语义相似性掩盖了其局部差异,而这些局部差异可能是特定语序选择的潜在原因。

嵌入虽然能够有效捕捉语料的语义特征,但其本身也存在局限性,对微观语义深层关联的刻画能力有限,以及在高维空间中可能遗漏动态语用因素。我们通过变量控制,仅能得出“语义对语序差异影响不大但不能忽视”的结论,而未进一步细化语义对语序的具体作用机制。未来的研究可以在更丰富的语料和功能块范围内,结合动态语义建模,深入探讨语义与语序的复杂交互关系,为英汉语序研究提供更精细的理论框架与实证支撑。

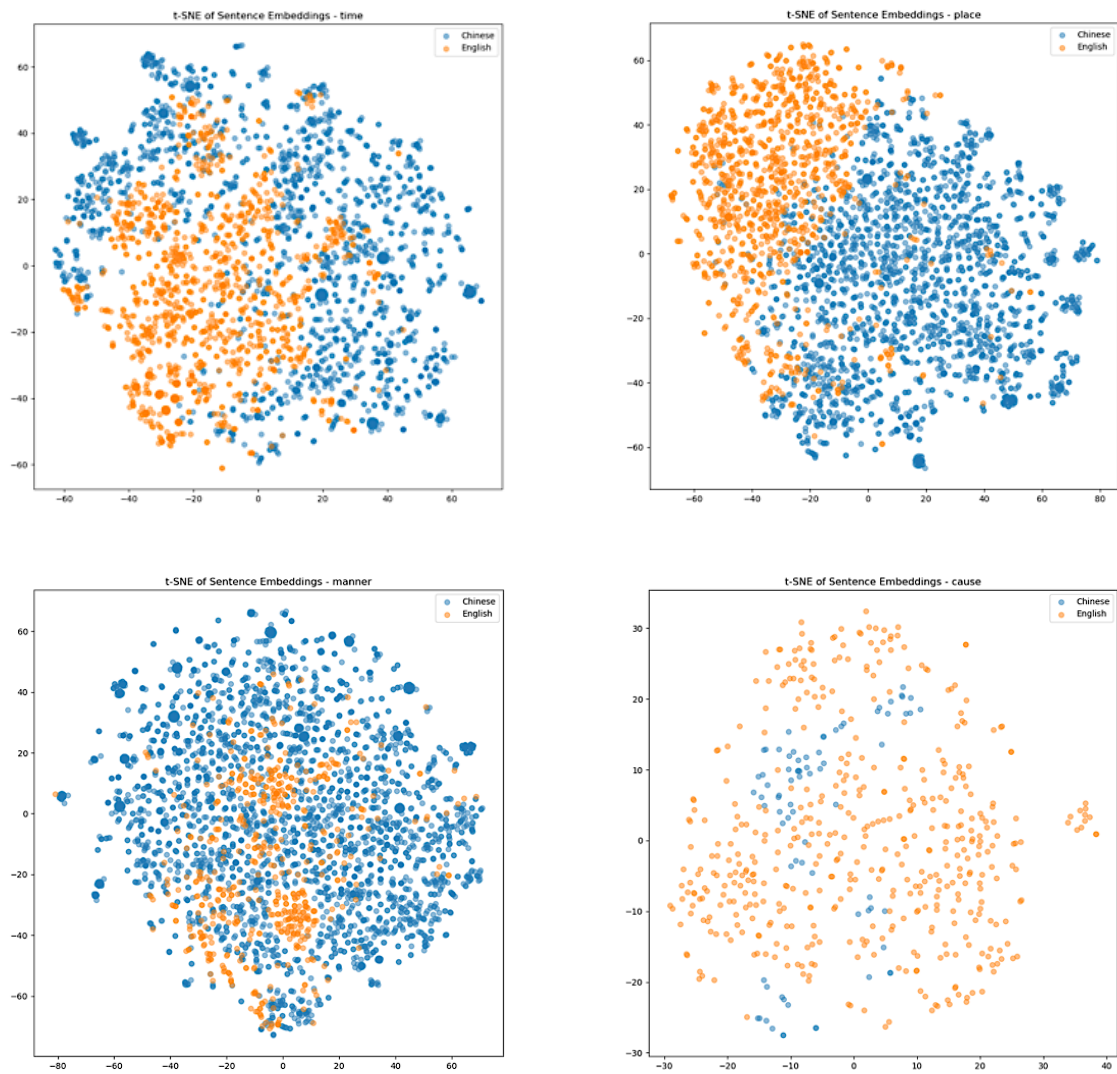
<sup>5</sup> <https://huggingface.co/openbmb/MiniCPM-Embedding>. MiniCPM-Embedding 是面壁智能、清华大学自然语言处理与社会人文计算实验室及东北大学信息检索小组共同开发的中英双语言文本嵌入模型。所有语料嵌入使用 4 张 RTX 4090D 24G 显卡完成。

<sup>6</sup> t-SNE(t-分布邻域嵌入)用于高维数据可视化的非线性降维,保留高维数据点之间的局部邻域关系,展现数据的局部结构特征。而余弦相似度计算基于全局特征向量的方向关系,揭示数据整体相似性,而对局部差异较不敏感。

表 12 功能块语义相似度

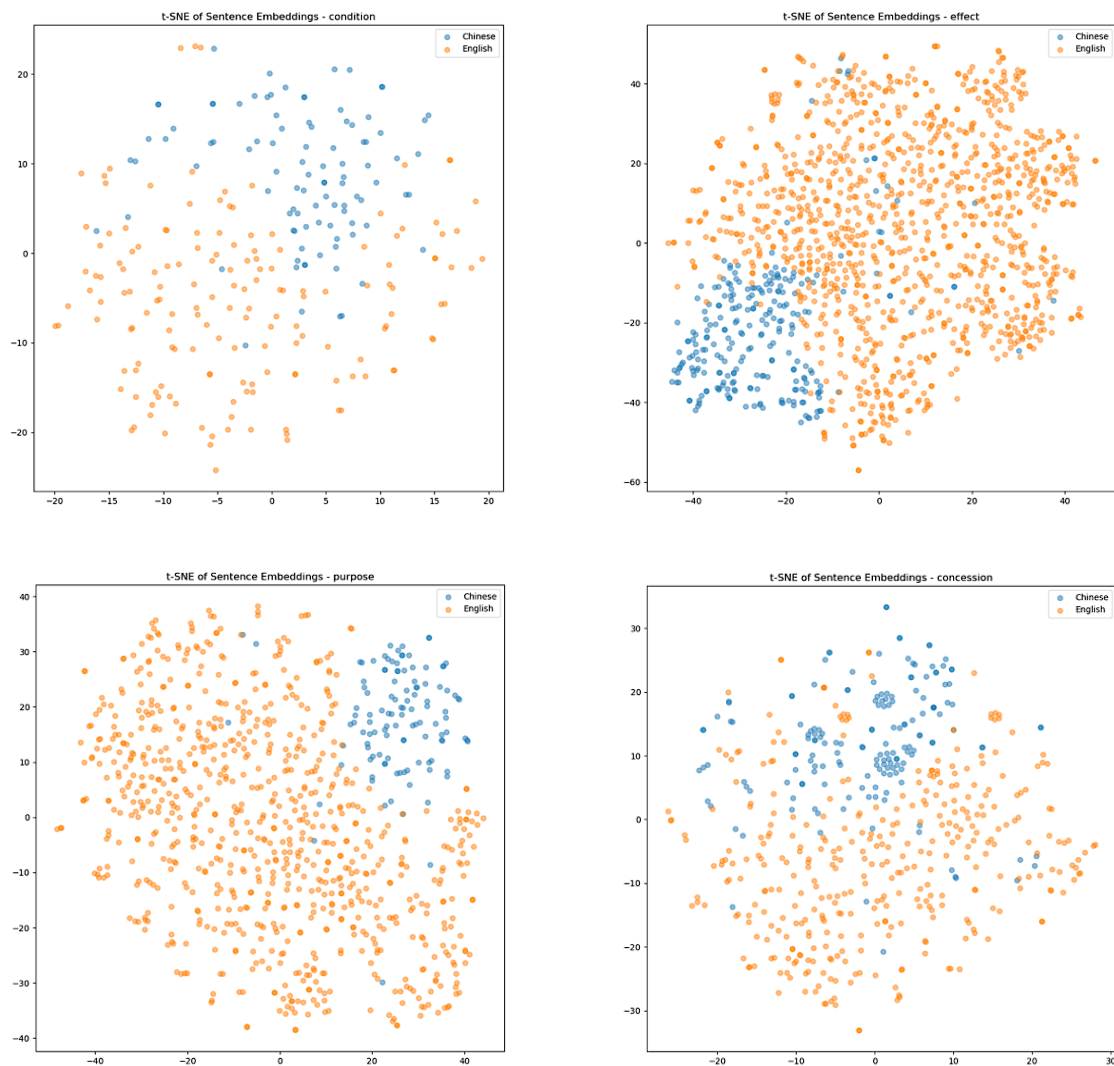
功能块类别	余弦相似度(%)
<time>	91.19
<place>	88.47
<manner>	86.61
<cause>	84.92
<effect>	87.47
<condition>	88.01
<purpose>	85.29
<concession>	86.63

图 5 功能块语义特征分布





续图 5



### 5.3 固定偏好中的灵活适应机制

英汉新闻语体的成分序分布显示出显著的模式化特征，但我们认为这些模式绝非完全固定，而是带有一定灵活性。这种灵活性主要受信息结构、语用需求及具体语境的影响，体现出成分序在英汉两种语言中既有偏好也有适应性的双重特点。这一观察为学界关于语序的争议提供了新的实证支持。例如，在英文语料中，当某些背景成分承载重要的语篇功能时，其位置可能会从后置转变为前置，如：

The public defense of the court as a nonpartisan institution comes at a fraught time for the justices and their credibility. The Court's approval rating has dipped below 50% for the first time since 2017 and down 9-points from a decade high just last year, according to Gallup. <time>This month</time>, <S>the court</S> <V>became embroiled</V> <O>in a dramatic and highly divisive debate</O> <cause>over abortion in Texas</cause> (CROWN2021-A12AB)

这一句中，前文提到了最高法院自 2017 年的“微妙”处境和公众对其公正性的质疑，结尾句中的“this month”则自然引入当下的一个具体案例，作为这种信任危机的直接体现，从抽象时间过渡到目前的具体事实。可见时间成分被前置是因为其在语篇中起到了承上启下的重要作用。这样的调整显示了英语语序对语用需求的灵活适应能力。类似地，中文新闻也表现出灵活的语序安排。如：

三名主要“接头人”进入调查组视线。2016 年 5 月 19 日与赖双平会面的年轻男子名叫刘海文，是铭威公司负责人。该公司于 2016 年 5 月 6 日刚中标新馆的音视频设备采购项目。会面次数最多的是艾玛公司的邵正飞。2015 年至 2016 年，该公司先后中标新旧两馆“图书自助借还系统”。<S>30 多岁的女性</S><V>叫</V><O>杨琴</O>，<place>来自大为公司</place>，<time>2016 年初</time><V>中标</V><O>新馆家具采购项目</O>。令人起疑的是，这几家中标方在收到合同款后几天之内，必有取出大额现金的行为。根据这一疑点，调查组决定先以两条线锁定三个人展开调查。一条线是艾玛公司，另一条是大为公司。(ToRCH2019-A05, 18-23)

此句的核心信息是该女性的身份和中标行为，因此将 SVO 前置符合焦点优先的表达需求，<时间>和<地点>作为补充信息，置于句尾形成了因果联系。从语篇角度来看，整体呈总 - 分 - 总的叙事结构，三名“接头人”的身份和行为构成平行关系。因此，将<地点>和<时间>调整于核心 SVO 后，这一分布虽然不完全符合中文的整体偏好，但在此处顺应了读者信息接收的逻辑过程，也增强了叙述层次感。这反映出中文在信息组织时的高效适应性，突破了传统的“背景优先”模式，在突出焦点和强化逻辑之间实现了平衡。

学界关于语序是否具有固定化规律的争议，往往聚焦于语法与信息结构之间的关系。一方面，传统研究认为成分序受类型学差异影响，句法能在很大程度上约束信息结构；而另一方面，近年来的研究逐渐强调语序的语用驱动性，指出语序选择更多取决于信息焦点、话题推进和语境适配。本文通过英汉新闻语料的对比分析，发现成分序既表现出稳定的分布规律，也能在语用需求的驱动下展现灵活性。这种灵活性背后反映了语言对认知和信息组织需求的适应性。例如，英语的“依存局部性”原则虽然鼓励线性推进，但在某些场景下，为突出话题或简化认知负担，部分背景信息会被前置。而中文的“话题链”结构虽然注重整体铺陈，但在某些信息焦点需要突出时，也会将背景信息后置。这些现象共同说明了语序的偏好化与灵活性并存，既有一定的系统性偏好，又具有灵活调整的能力。这种双重特性为语序是否“固定化”的争议提供了平衡的解释。语序的选择在很大程度上取决于语用需求、信息焦点以及语境逻辑，这不仅拓展了我们对英汉语序特征的理解，也为进一步研究语序的跨语体规律提供了重要的参考方向。

## 6. 结语

本文基于大规模英汉新闻语料库，结合大语言模型自动标注，尝试从功能块角度探讨了两种语言的成分序偏好差异及其排列规律。研究发现：(1)时间、地点等成分在英汉新闻语体中展现出显著的语序偏好。英语倾向于将这些功能块置于句末，强化核心事件的信息中心地位；而汉语则表现出更强的前置倾向，突出背景信息在叙事开篇中的重要作用。(2)SVO 结构与功能块的组合模式体现了英汉新闻的语序差异，但这种差异的显著程度有所不同，汉语的前置倾向显著，英语的后置特性则相对温和，反映了两种语言在信息组织上的不同重心。(3)多重功能块共现时，英汉新闻语体的成分均表现出较高的灵活性，语序的调整受到信息焦点、语用需求和语篇功能的灵活驱动。上述发现一定程度上揭示了英汉新闻语体在语序偏好上的差异和在信息组织上的不同策略，不仅验证了既有理论关于语序分布的类型学假设，也为语序的语用适应机制提供了新的实证支持。



## 参 考 文 献

- [1] Ai, H., Zhu, G., Gao, M., & Yin, W. (2016). The Expression of Stance in Mandarin Chinese: A Corpus-Based Study of Stance Adverbs. *COLIPS*, 45(2), 188-207.
- [2] Alizadeh, M., Kubli, M., Samei, Z., Dehghani, S., Zahedivafa, M., Bermeo, J. D., Korobeynikova, M., & Gilardi, F. (2024). *Open-Source Large Language Models Outperform Crowd Workers and Approach ChatGPT in Text-Annotation Tasks* (No. arXiv:2307.02179). arXiv.
- [3] Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. K. (1999). *Longman Grammar of Spoken and Written English*. Pearson Education Limited.
- [4] Chamonikolasová, J. (2009). *Word Order and Linear Modification in English*. *Brno Studies in English*, 35(2), 17–28.
- [5] Dryer, M. S. (1986). Primary Chronolects and Primary Topolects. In C. N. Yang (Ed.), *Essays on Chinese phonetics and linguistics* (pp. 147-164).
- [6] Futrell, R., Levy, R. P., & Gibson, E. (2020). Dependency Locality as an Explanatory Principle for Word Order. *Language*, 96(2), 371–412.
- [7] Goldberg, A. 1999. The Emergence of the Semantics of Argument Structure Constructions. In B. MacWhinney (ed.). *The Emergence of Language*. Mahwah: Lawrence Erlbaum.
- [8] Halliday, M. A. K. (1985). *An Introduction to Functional Grammar*. London: Edward Arnold.
- [9] Hawkins, J. A. (1994). *A Performance Theory of Order and Constituency*. Cambridge University Press.
- [10] Huddleston, R., & Pullum, G. K. (2002). *The Cambridge Grammar of the English Language*. Cambridge University Press.
- [11] Jacobson, P. 1996. *Constituent Structure*. In Concise Encyclopedia of Syntactic Theories. Cambridge: Pergamon.
- [12] Ji, M. (2015). Principal Component Analysis of the Information Structure of British English and Chinese Genres. *Glottology*, 6, 209 - 228.
- [13] Kojima, T., Gu, S. S., Reid, M., Matsuo, Y., & Iwasawa, Y. (2022). Large Language Models Are Zero-Shot Reasoners. *Advances in Neural Information Processing Systems*, 35, 22199-22213.
- [14] Lee, D. (2019). The syntax of English: A Comparative Analysis of Word Order Rules Across Registers. *Journal of Linguistics*, 55(2), 251-293.
- [15] Liu, L., & Guo, H. (2014). Thematic Structure and Thematic Progression in English and Chinese Broadcast News Discourse. *Foreign Language Research*, (5), 85-90.
- [16] Müller, S. (2021). Constituent order. *Head-Driven Phrase Structure Grammar: The handbook*, 9, 369.
- [17] OpenAI. (2023). Gpt-4 Technical Report. *arXiv preprint arXiv:2303.08774*.
- [18] Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *AComprehensive Grammar of the English Language*. Longman.
- [19] Ross, C. (1984). Adverbial Modification in Mandarin. *Journal of Chinese Linguistics*, 12(2), 207-234.
- [20] Scollon, R. (2000). Generic Variability in News Stories in Chinese and English: A Contrastive Discourse Study of Five Days' Newspapers. *Journal of Pragmatics*, 32(6), 761–791.
- [21] Stolterfoht, B. (2013). Adverb Position and Information Structure in Processing English. *Questions and Answers in Linguistics*, 1(1), 1-15.
- [22] Sun, M., Xu, J. et al. (2022). *The CROWN2021 Corpus*. National Research Centre for Foreign Language Education, Beijing Foreign Studies University.
- [23] Tan, Z., Li, D., Wang, S., Beigi, A., Jiang, B., Bhattacharjee, A., Karami, M., Li, J., Cheng, L., & Liu, H. (2024). *Large Language Models for Data Annotation and Synthesis: A Survey* (No. arXiv:2402.13446). arXiv.
- [24] Travis, L. (1984). *Parameters and Effects of Word Order Variation*. PhD dissertation, Massachusetts Institute of Technology.

- [25] Wasow, T. (2003). Post-Verbal Constituent Ordering in English. *Determinants of Grammatical Variation in English/Walter de Gruyter*.
- [26] Xing, J. (2015). Adverbial Modification in Mandarin: Functional and Structural Constraints. In *The Oxford Handbook of Chinese Linguistics* (pp. 245-263). Oxford University Press.
- [27] Yurayong, C., & Sandman, E. (2023). Chinese Word Order in the Comparative Sino-Tibetan and Sociotypological Contexts. *Languages*, 8(2), 112.
- [28] Zhang, Y., & Wang, X. (2022). The Syntax and Semantics of Temporal Adverbial Clauses in Mandarin Chinese: A Dependency Treebank Analysis. *Journal of East Asian Linguistics*, 51(2), 147-183.
- [29] 安玉霞. (2006). 汉语语序问题研究综述. *汉语学习*, 44-51.
- [30] 曾静涵. (2015). 多项状语语序的类型学考察与认知分析. *现代语文(语言研究版)*, 125-129.
- [31] 丁志斌. (2014). 英汉语序的类型学特征. *西安外国语大学学报*, 20-23.
- [32] 郭骅. (2022). 英汉学术语篇主位推进模式比较研究. *上海理工大学学报(社会科学版)*, 44(2), 116-123.
- [33] 郭中. (2012). 汉语多类状语共现的语序自由度及其解释. *汉语学习*, 104-112.
- [34] 胡裕树, 陆丙甫. (1988). 关于制约汉语语序的一些因素. *烟台大学学报(哲学社会科学版)*, 55-60.
- [35] 黄一清. (2010). 外国学生简单状语语序偏误考察. *语文学刊*, 7-9.
- [36] 贾德霖. (1990). 思维模式与线性序列——汉式英语语序特色. *外国语(上海外国语学院学报)*.
- [37] 金立鑫. (2019). 汉语语序的类型学特征. *解放军外国语学院学报*, 1-13, 159.
- [38] 李佳蕾, 孙铭辰, & 许家金, 2022, *ToRCH2019 现代汉语平衡语料库*. 北京外国语大学中国外语与教育研究中心.
- [39] 李金满. (2010). 语言类型学视角下的汉英语序对比研究. *当代外语研究*, 45-51, 61.
- [40] 李锡江, 刘永兵. (2020). 中国学生英语时空状语语序分布与母语概念迁移. *外语教学与研究*, 90-102, 159.
- [41] 荣晶. (2000). 汉语语序研究的理论思考及其考察. *语言文字应用*, 25-30.
- [42] 双文庭. (2016). 英汉多项状语并置语序差异的认知阐释——以工具、地点、频度和时间状语为例. *天津外国语学院学报*, 24-27, 80-81.
- [43] 汪康懋, 肖研. (1981). 英汉语序的比较研究. *外语教学与研究*, 64-72.
- [44] 王力. (1980). *汉语史稿(中册)*. 北京: 北京书局.
- [45] 吴为章. (1995). 语序重要. *中国语文*, 429-436.
- [46] 袁毓林. (1999). 定语顺序的认知解释及其理论蕴涵. *中国社会科学(02)*, 185-201.
- [47] 张伯江. (2011). 汉语的句法结构和语用结构. *语言学*, (3), 1-12.
- [48] 张炼强. (1997). 汉语语序的多面考察(下). *首都师范大学学报(社会科学版)*, 39-47.
- [49] 章璐, 王富银. (2019). 基于翻译实践的英汉语序对比研究. *英语广场*, 57-59.
- [50] 赵元任. (1968). 汉语口语语法. *语言学*, (1), 1-20.

## 附 录

表 1 中文频率前 20 的功能块 - SVO 分布模式

排名	模式	频数
1	<time><S><V><O>	17
2	<time><V><O>	15
3	<place><S><V><O>	13
4	<manner><V><O>	13
5	<place><V><O>	12
6	<S><V><O><manner><V><O>	7
7	<S><V><O><manner>	7
8	<place><S><V>	7
9	<effect><V><O>	5
10	<V><O><V><O><manner>	4
11	<place><S><manner><V><O>	4
12	<time><S><V>	4
13	<S><place><V><O>	4
14	<time><place><S><V><O>	4
15	<S><V><O><manner><O>	3
16	<time><S><V><O><V><O>	3
17	<S><V><O><place>	3
18	<time><place><V><O>	3
19	<place><manner><V><O>	3
20	<time><S><V><place>	3

表 2 英文频率前 20 的功能块 - SVO 分布模式

排名	模式	频数
1	<S><V><effect>	127
2	<S><V><O><time>	81
3	<S><V><O><effect>	70
4	<S><V><O><place>	61
5	<S><V><O><purpose>	59
6	<S><V><place>	51
7	<S><V><time>	45
8	<S><V><purpose>	43
9	<S><V><O><concession>	33
10	<S><V><O><manner>	32
11	<S><V><O><cause>	23
12	<S><V><effect><time>	21
13	<S><V><effect><cause>	20
14	<S><V><time><place>	20
15	<S><V><cause>	18
16	<S><V><O><time><cause>	17
17	<S><V><time><cause>	15
18	<S><V><effect><concession>	15
19	<S><V><time><effect>	14
20	<S><V><O><place><time>	14

表 3 中文频率前 50 的多重功能块组合

排名	模式	频数
1	<time><place>	415
2	<place><manner>	345
3	<place><time>	200
4	<manner><place>	183
5	<time><manner>	155
6	<manner><time>	129
7	<time><place><manner>	55
8	<effect><time>	35
9	<place><manner><place>	30
10	<manner><place><manner>	28
11	<manner><effect>	25
12	<effect><manner>	25
13	<effect><place>	24
14	<manner><time><place>	23
15	<concession><place>	21
16	<manner><concession>	20
17	<concession><manner>	19
18	<purpose><manner>	18
19	<place><time><place>	18
20	<condition><manner>	17
21	<place><manner><time>	16
22	<place><time><manner>	16
23	<time><place><time>	13
24	<time><effect>	12
25	<time><manner><place>	11

续表

排名	模式	频数
26	<manner><condition>	11
27	<manner><time><manner>	11
28	<place><manner><place><manner>	10
29	<manner><purpose>	10
30	<concession><time>	10
31	<time><condition>	9
32	<time><cause>	9
33	<cause><manner>	8
34	<time><purpose>	8
35	<purpose><time>	8
36	<cause><place>	7
37	<cause><time>	7
38	<purpose><place>	7
39	<place><cause>	7
40	<purpose><effect>	6
41	<effect><time><place>	6
42	<manner><cause>	6
43	<manner><place><manner><place>	6
44	<manner><place><time>	6
45	<condition><place>	6
46	<time><manner><time>	6
47	<place><purpose>	6
48	<effect><place><manner>	6
49	<condition><effect>	5
50	<time><place><manner><place>	5

表 4 英文频率前 50 的多重功能块组合

排名	模式	频数
1	<time><place>	152
2	<place><time>	135
3	<time><cause>	88
4	<effect><time>	82
5	<time><purpose>	73
6	<effect><cause>	69
7	<purpose><time>	64
8	<time><effect>	63
9	<place><purpose>	58
10	<place><effect>	55
11	<manner><place>	53
12	<place><manner>	53
13	<manner><purpose>	51
14	<place><cause>	48
15	<time><concession>	45
16	<time><manner>	43
17	<effect><concession>	43
18	<effect><place>	40
19	<purpose><place>	37
20	<purpose><concession>	36
21	<manner><time>	35
22	<manner><effect>	31
23	<place><concession>	31
24	<purpose><cause>	27
25	<concession><cause>	25

续表

排名	模式	频数
26	<effect><purpose>	23
27	<manner><concession>	23
28	<purpose><manner>	23
29	<cause><effect>	22
30	<purpose><effect>	21
31	<effect><manner>	19
32	<cause><time>	16
33	<purpose><condition>	14
34	<concession><effect>	14
35	<condition><effect>	14
36	<concession><place>	12
37	<cause><concession>	12
38	<time><condition>	11
39	<cause><purpose>	11
40	<time><place><effect>	11
41	<manner><cause>	11
42	<time><place><manner>	10
43	<effect><condition>	10
44	<cause><place>	10
45	<concession><purpose>	10
46	<place><time><cause>	9
47	<cause><manner>	9
48	<concession><time>	9
49	<time><place><cause>	9
50	<place><purpose><time>	9

表 5 中文功能块转移概率

次序	转移关系	概率
1	<place> to <manner>	0.55
2	<purpose> to <manner>	0.50
3	<condition> to <manner>	0.45
4	<manner> to <manner>	0.44
5	<time> to <place>	0.43
6	<time> to <manner>	0.39
7	<cause> to <manner>	0.39
8	<concession> to <manner>	0.37
9	<effect> to <manner>	0.30
10	<effect> to <place>	0.29
11	<effect> to <time>	0.28
12	<manner> to <place>	0.27
13	<cause> to <time>	0.24
14	<cause> to <place>	0.24
15	<place> to <time>	0.24
16	<concession> to <place>	0.24
17	<concession> to <time>	0.23
18	<manner> to <time>	0.21
19	<condition> to <place>	0.18
20	<purpose> to <time>	0.18
21	<purpose> to <place>	0.17
22	<condition> to <time>	0.16
23	<place> to <place>	0.15
24	<time> to <time>	0.12
25	<condition> to <effect>	0.11
26	<purpose> to <effect>	0.09
27	<concession> to <concession>	0.06
28	<condition> to <purpose>	0.05
29	<purpose> to <condition>	0.05
30	<concession> to <condition>	0.05
31	<cause> to <effect>	0.05
32	<concession> to <effect>	0.04

续表

次序	转移关系	概率
33	<effect> to <purpose>	0.03
34	<cause> to <purpose>	0.03
35	<cause> to <concession>	0.03
36	<manner> to <effect>	0.03
37	<effect> to <effect>	0.03
38	<effect> to <condition>	0.03
39	<effect> to <concession>	0.03
40	<condition> to <cause>	0.03
41	<time> to <effect>	0.03
42	<place> to <effect>	0.02
43	<manner> to <concession>	0.02
44	<effect> to <cause>	0.02
45	<place> to <concession>	0.02
46	<manner> to <purpose>	0.02
47	<cause> to <cause>	0.02
48	<concession> to <cause>	0.01
49	<concession> to <purpose>	0.01
50	<condition> to <condition>	0.01
51	<condition> to <concession>	0.01
52	<time> to <purpose>	0.01
53	<purpose> to <concession>	0.01
54	<manner> to <condition>	0.01
55	<time> to <cause>	0.01
56	<place> to <purpose>	0.01
57	<time> to <concession>	0.01
58	<place> to <cause>	0.01
59	<place> to <condition>	0.01
60	<time> to <condition>	0.01
61	<manner> to <cause>	0.01
62	<cause> to <condition>	0.00
63	<purpose> to <cause>	0.00
64	<purpose> to <purpose>	0.00

表 6 英文功能块转移概率

排名	转移关系	概率
1	<condition> to <effect>	0.33
2	<cause> to <effect>	0.28
3	<place> to <time>	0.28
4	<purpose> to <time>	0.26
5	<time> to <place>	0.26
6	<effect> to <time>	0.24
7	<concession> to <effect>	0.23
8	<manner> to <place>	0.22
9	<manner> to <purpose>	0.22
10	<cause> to <time>	0.19
11	<manner> to <effect>	0.18
12	<concession> to <place>	0.18
13	<time> to <effect>	0.17
14	<concession> to <cause>	0.17
15	<cause> to <place>	0.17
16	<place> to <effect>	0.17
17	<time> to <purpose>	0.16
18	<condition> to <purpose>	0.16
19	<purpose> to <place>	0.15
20	<effect> to <cause>	0.15
21	<manner> to <time>	0.15
22	<effect> to <place>	0.15
23	<effect> to <effect>	0.15
24	<place> to <purpose>	0.15
25	<condition> to <place>	0.14
26	<purpose> to <effect>	0.14
27	<concession> to <time>	0.13
28	<concession> to <purpose>	0.13
29	<time> to <cause>	0.13
30	<cause> to <manner>	0.13
31	<condition> to <manner>	0.12
32	<place> to <place>	0.12

续表

排名	转移关系	概率
33	<place> to <manner>	0.12
34	<cause> to <purpose>	0.11
35	<purpose> to <concession>	0.11
36	<time> to <manner>	0.11
37	<effect> to <concession>	0.10
38	<condition> to <concession>	0.10
39	<purpose> to <purpose>	0.10
40	<place> to <cause>	0.10
41	<purpose> to <manner>	0.10
42	<effect> to <purpose>	0.09
43	<time> to <time>	0.09
44	<purpose> to <cause>	0.09
45	<concession> to <manner>	0.09
46	<effect> to <manner>	0.08
47	<cause> to <concession>	0.07
48	<manner> to <concession>	0.07
49	<time> to <concession>	0.07
50	<condition> to <time>	0.06
51	<manner> to <manner>	0.06
52	<manner> to <cause>	0.06
53	<condition> to <cause>	0.05
54	<place> to <concession>	0.05
55	<purpose> to <condition>	0.05
56	<concession> to <condition>	0.04
57	<manner> to <condition>	0.04
58	<concession> to <concession>	0.03
59	<condition> to <condition>	0.03
60	<effect> to <condition>	0.03
61	<cause> to <cause>	0.02
62	<cause> to <condition>	0.02
63	<place> to <condition>	0.02
64	<time> to <condition>	0.01