

## 基于 Wmatrix 在线语料库的中国高铁英文报道之高铁隐喻研究

尧文群\*, 姚美玉, 任忆欣

华东交通大学 江西南昌

**【摘要】**中国高铁的发展令世界瞩目,国内主要媒体通过英文报道向国内外客观如实地展示了中国高铁发展历程。本研究采用 Wmatrix 在线语料库统计和分析了三十余年中国高铁英文报道中以高铁为主题的中国高铁英文隐喻,统计了前十种高频率中国高铁英文隐喻,并借助 Wmatrix 的语义域赋码功能探讨中国高铁隐喻的跨域特点,研究发现中国高铁英文报道中隐喻出现的比例较高,出现海陆空交通概念跨域映射,以及部分高铁隐喻形成隐喻群体现象。本研究一方面有助于拓展隐喻的语料库研究,另一方面有助于推动中国高铁英文报道隐喻语言特点研究。

**【关键词】**中国高铁英文报道; 高铁隐喻; Wmatrix 在线语料库

**【基金项目】**华东交通大学 2024 年大学生创新创业训练计划项目

**【收稿日期】**2024 年 11 月 13 日

**【出刊日期】**2024 年 12 月 20 日

**【DOI】**10.12208/j.ssr.20240056

### HSR metaphors in English reports of Chinese HSR—— a study based on Wmatrix online corpus

Wenqun Yao\*, Meiyu Yao, Yixin Ren

East China Jiaotong University, Nanchang, Jiangxi

**【Abstract】**With Chinese HSR development drawing more global attention, the main domestic media issued many reports to showcase the development of Chinese HSR in English. This study calculated and analyzed Chinese HSR English metaphors in news reports for the past 30 years, including the top 10 frequently used Chinese HSR English metaphors and the discovery of Chinese HSR cross-domain features by applying Wmatrix semantic tagset tool. The study results show that there are large number of HSR metaphors in CRH English reports crossing the concepts of people, water, air and road transportation. This study is helpful for future corpus research on metaphor and the metaphor language study in Chinese HSR English reports as well.

**【Keywords】**Chinese HSR English report; HSR metaphor; Wmatrix online corpus

#### 1 引言

二十世纪末以来,中国政府高度重视高铁事业的发展,经过多年的技术研究,工程实践和安全运营,中国高铁发展迅速,取得令世界瞩目的成就。当下中国是世界上高铁运营里数最长,在建规模最大,商业运营速度最高,高铁技术最全面,运营场景和管理经验最为丰富的国家。中国高铁不仅成为国民出行的主要交通方式和国家经济发展的重要驱动力,而且构成中国对外宣传实力与文化的重要组成部分,各媒体争相报道中国高铁的发展,涌现了许多中国高铁英文报道。当我们浏览中国高铁英文报道,可

以发现报道出现了相当数量的英文高铁隐喻,这些英文高铁隐喻通过跨域映射展示高铁特色,产生良好的媒体宣传效果。

采用语料库方法对自然语篇的隐喻进行实证研究成为隐喻研究的发展新趋势(孙毅,2013)<sup>[1]</sup>。英国 Lancaster 大学 Paul Rayson 等研发的 Wmatrix 在线语料库可以自动对文本进行语义域赋码,提升隐喻研究的科学化和智能化,成为近年来隐喻研究的有力辅助工具(刘超健 王军 曹灵美,2018)<sup>[2]</sup>(孙亚,2012)<sup>[3]</sup>(Tognini-Bonelli,2001)<sup>[4]</sup>。本研究结合 Wmatrix 在线语料库,以研究小组收集的 747 篇

\*通讯作者:尧文群。

在线媒体网络平台刊出的中国高铁英文报道为自建语料, 通过统计和分析, 探讨中国高铁英文报道中的高铁隐喻特点。

## 2 研究过程

### 2.1 上传语料

本研究选择中国日报, 国新办等权威媒体在线平台上的中国高铁英文报道为文本来源, 共收集了 747 篇中国高铁英文报道。为避免格式不符产生的乱码现象, 首先对 747 篇中国高铁英文报道进行清洁处理, 去除标题和插图, 将英文报道转换成纯文本格式上传到 Wmatrix 语料库, 形成一个中国高铁英文报道的自建语料库。Wmatrix 系统自动完成语料库的词性标注和语义域标注, 生成相应词表。数据显示本语料库有 239332 个形符和 4374 个类符。

### 2.2 生成主题语义域

本研究关注中国高铁英文报道的语义域和主题

语义域信息。鉴于上传语料为书面文本, 在 Wmatrix 中选择 BNC Sampler Written 为参照语料库, 运行 Wmatrix 系统, 生成主题语义域列表 (LL≥6.63, df=1, p<0.01)。与参照语料库对比, 语料获得 143 个主题语义域。

表 1 展示了白皮书排名前五名的主题语义域, 排名 1“Vehicles and transport on land”的主题语义域契合语料的高铁核心主题。表中 O1 和%1 指语义域在研究语料库中的频率和相对频率, O2 和%2 指该语义域在参照语料库中的频率和相对频率, “+”说明该语义域在研究语料库中比在参照语料库中有更高的使用频率, LL 为主题语义域的对数似然比值, 表明该语义域使用的显著性程度 (Kövecses, Z., 2005) [5]。图 1 是语料库排名前 100 的主题语义域云图, 字母越大表明主题出现频率越高。表 1 排名第一“Vehicles and transport on land”位于图 1 的右下方。

表 1 前 5 个主题语义域

Item	O1	%1	O2	%2	LL	LogRatio
1 M3	10112	4.23	2171	0.22+ 22234.13	4.24	Vehicles and transport on land
2 N3.8	4636	1.94	1282	0.13+ 9387.86	3.87	Speed: Fast
3 Z2	8741	3.65	14502	1.50 + 3923.00	1.29	Geographical names
4 Z99	11496	4.80	22165	2.29 + 3781.65	1.07	Unmatched
5 M7	4258	1.78	5888	0.61 + 2582.23	1.55	Places

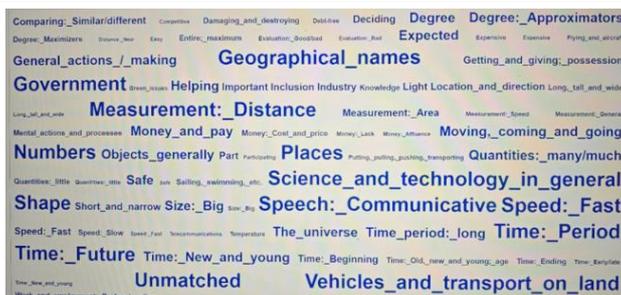


图 1 语料库排名前 100 的主题语义域云图

### 2.3 确定目标域

Wmatrix 为语料词语自动标注的语义域大致对应概念隐喻的源域或目标域 (Pragglejaz Group, 2007) [6], 为提取隐喻提供了便利的基础。本研究分析语料库中的中国高铁英文隐喻现象与规律, 将目标主题词设为“train”和“trains”。点击打开排名第一的主题语义域“Vehicles and transport on land”, 在词项搜索栏内分别键入 “train”和“trains”, 系统出现了 1063 个

“train”索引行(图 2)和 992 个“trains”索引行(图 3)。

### 2.4 识别隐喻

在主题话题检索词的所有索引行中寻找中国高铁的候选隐喻载体词, 并采用隐喻识别程序 MIPVU 进行判断。MIPVU 程序 (Steen et al. 2010) [7]。主要步骤包括: 1) 确定词语的基本意义和语境意义; 2) 判断该词语的基本意义及语境意义是否存在差别; 3) 考察该词语的基本意义和语境意义的指称对

象之间是否存在相似性关系。如果存在相似关系, 则这个词语可以认定为隐喻载体词 (陈朗, 2022) [8]。基本意义指更具体的意义, 如能容易成像、五官能感觉到的意义; 语境意义则是在当下语篇中上下文语境下表达的具体意义 (刘阳, 赵秀凤, 2017) [9]。本研究在判断基本意义和语境意义时参考了《麦克米伦高阶英语词典》和在线词源词典 (Online Etymology Dictionary)。例如在索引行“...in preparation for the train’s debut in 2025...”中, 词项“debut”的基本义是“首场演出”, 在本索引行中的语境义是中国高铁的“第一次亮相与运行”, 体现了两个不同语义域之间的相似性映射, 因此判定为高铁隐喻。

表2 主题词“train”和“trains”排名前10高铁隐喻

序号	高铁隐喻词项	出现频率	隐喻类别	语义源域
1	high-speed	456	空间	N3.7
2	run	158	人	M1
3	travel/trip	114	人	M1
4	bullet	55	战争	G3
5	Test	42	教育	P1
6	construction	30	建筑	H1
7	pairs	29	人	N5
8	network	21	社会	S5
9	ride	18	体育	K5.1
10	line	16	物件	O4.4

例如:

- 1) “...take a look at the trains which will serve the two cities.”
- 2) “Train D2763 waits to depart at the station...”
- 3) “...ensure the safe running of the train operation control system...”

隐喻词项“serve”, “waits”, “running”都是人的动作, 将它们应用到报道中国高铁中, 是将中国高铁拟人化, 即通过人的日常行为映射到中国高铁, 可以通俗生动地再现高铁的发展与运营, 充分展现中国高铁为大众日常生活提供便利的使命与价值。

(2) 本研究语料的中国高铁英文隐喻经常性地借用了其它交通领域的概念对中国高铁进行描述和解释, 如公路运输, 水路运输和航空运输。

### 3 研究结果与发现

结合 Wmatrix 的语义赋码和 MIPVU 隐喻识别程序, 通过对索引行内的单元词项进行分析, 识别和统计, 发现中国高铁英文报道的自建语料库中以主题词“train”和“trains”的出现的隐喻数量多, 种类多样, 跨域映射丰富。具体特点表现如下:

(1) 根据统计, 本语料库中围绕主题词“train”和“trains”相关的高铁隐喻种类众多, 下面表 2 展示排名前十位的相关高铁隐喻相关信息。

由表 2 可以看出中国高铁隐喻的拟人隐喻出现频率高, 采用人的特征来映射和解释高铁的特征可以降低高铁专业性的难度并增加内容的生动趣味性, 符合近取诸身的隐喻创造传统。

公路运输:

- 1) “Four coaches of D301 fell off...” (马车)
- 2) “A total of ten 8-car trains which can seat around 600 passengers...” (汽车)
- 3) “On the 200kmh trains, the cabs we work in is clean...” (出租车)

水路运输:

- 1) “A high-speed train rammed into a stalled train nearby...” (船抛锚)
- 2) “The second train also veered to...” (船偏航)
- 3) “...purchase 6 fleets of high-speed trains.” (船舰队)

航空运输:

- 1) “...the train shuttles between the three main

cities...”(航天飞机)

2) “...the Revival Express train took off at a full capacity...”(飞机起飞)

3) “...as the first commercial high-speed train makes its maiden voyage from China...”(首次航行)

中国高铁英文报道中的这些高铁隐喻(公路汽车, 水路船舰和空中飞机)通过与其它运输领域相关事物的相似性(诸如外形, 功能, 价值等), 并基于交通方式之间存在的共通性, 通过隐喻对中国高铁事件进行有效的报道与解说。

(3) 研究发现语料中的中国高铁隐喻在频繁使用某个隐喻概念进行域间映射, 出现各种变形或类替代现象, 整体上呈现一定的隐喻群体与专有性(Lakoff, G. & M. Johnson, 1980)<sup>[10]</sup>。以隐喻出现频率排名第二的 run 隐喻为例, 语料库中以“人跑”隐喻“高铁快速运行”。此隐喻词项出现的形式包括“run”, “runs”, “ran”, “running”, 包含动词, 名词和动名词等多种词性, 已经形成高铁运行的专有通用词群体。语料库中成群结伴的高铁隐喻还有“cover”+“covering”+“covered”+“coverage”; “build”+“built”+“building”+“builder”等。

中国高铁是中国高速铁路的缩写, 这个名称就是一个典型的高铁隐喻。研究发现, 本英文报道语料库中形容中国高铁的速度快主要出现如下隐喻表达:

1) “The high-speed trains have been the milestone in the history.”

2) “...adding that the trains will be able to run at full speed...”

3) “...bullet train designed with top speed of 450 kilometers an hour...”

这些由“high/top”/“高”到“full”/“满”到“fast”/“快”进行概念转变的隐喻映射体现了人们通过空间丈量概念来理解速度测量概念的智慧思维, 尤其是“high-speed trains”/高速铁路已经成为人人习以为常的名称了。

(4) 中国高铁英文报道内容涉及高铁发展的不同阶段, 涵盖设计、生产、测试、运营、评价和出口的方方面面, 当然这些过程中不可避免地出现了相关地隐喻语言与概念。

中国高铁设计:

1) “...the train will be customized according to MTR requirements.”(定制)

2) “...A new CRH380 train, which is tailor-made for Guangzhou-Shenzhen -Hongkong...”(度身定做)  
中国高铁生产:

1) “...could not be better reflected in its home-grown high-speed trains...”(土生土长)

2) “...the widely recognized birthplace of high-speed trains...”(发源地)

中国高铁测试:

1) “...bullet train traveled at the speed of 453km/h on test runs in Fujian province...”(测试)

2) “June 4, 2022, during a trial operation of the Chongqing-Wuhan section of the high-speed trains...”(检测)

中国高铁运营:

1) “More high-speed trains now shuttle between cities like buses.”(穿梭)

2) “The high-speed trains have raced along Chinese rails...”(飞奔)

中国高铁评价:

1) “... the smart train started from Beijing.”(聪明的)

2) “...Fuxing, the intelligent trains are put into operation.”(智慧的)

3) “The high-speed trains have been a milestone in the history of Chinese railway development.”(里程碑)

中国高铁出口:

1) “Railway High-speed (CRH) trains have become a calling card for China...”(名片)

2) “High-speed trains are just like a huge business magnet, driving huge foreign orders.”(磁铁)

可以看出在本研究的语料库中围绕主题词“train”和“trains”的隐喻出现在中国高铁发展各阶段, 不仅丰富了中国高铁报道语言表达, 而且多角度地展示中国高铁的发展与成就。

#### 4 结语

本研究以近几十年主流媒体发表的在线中国高铁英文报道为研究对象, 在 Wmatrix 语料库系统里创建了语料库; 根据中国高铁的核心主题, 在语料库中搜索主题词项的索引行中的潜在隐喻词项, 并

借鉴 MIPVU 程序加识别高铁隐喻。研究发现中国高铁英文报道中在中国高铁发展不同阶段分别出现了不同概念映射的高铁隐喻, 既有专有性隐喻的高频性, 又有新奇创新隐喻, 体现各媒体对中国高铁价值的高度认可。

### 参考文献

- [1] 孙毅. 基于语义域的隐喻甄别技术初探——以 Wmatrix 语料库工具为例[J]. 解放军外国语学院学报, 2013(7):10-18.
- [2] 刘超健, 王军, 曹灵美. 语料库中的隐喻标注研究[J]. 外语学刊, 2018(2):35-41.
- [3] 孙亚. 基于语料库工具 Wmatrix 的隐喻研究[J]. 2012(5): 7-12.
- [4] Tognini-Bonelli, E. *Corpus Linguistics at Work* [M]. Amsterdam: John Benjamins, 2001.
- [5] Kövecses, Z. *Metaphor in Culture: University and Variation* [M]. Cambridge, New York: Cambridge University Press, 2005.
- [8] Pragglejaz Group . MIP: A method for identifying metaphorically used words in discourse [J]. *Metaphor and Symbol*, 2007, (1) : 1—39.
- [9] Steen, G. J. et al. *A Method for Linguistic Metaphor Identification* [M]. Amsterdam: John Benjamins, 2010.
- [10] 陈朗. 从 MIP 到 MIPVU: 隐喻识别的方法、应用与问题 [J]. 外语学刊, 2022(5): 30-40.
- [11] 刘阳, 赵秀凤. 基于语料库工具 Wmatrix 的能源隐喻研究 [J]. 现代语文, 2017(5):84-87.
- [12] Lakoff, G. & M. Johnson. *Metaphors We Live by* [M]. Chicago: University of Chicago Press, 1980.

**版权声明:** ©2024 作者与开放获取期刊研究中心(OAJRC)所有。本文章按照知识共享署名许可条款发表。

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



**OPEN ACCESS**