

基于语料库的化工领域中外学者外壳名词使用对比研究

张子航^{1*}, 汤益愉^{2*}, 杨子瀚³, 管 博^{2#}

¹华东理工大学国际卓越工程师学院, 上海

²华东理工大学外国语学院, 上海

³上海外国语大学西方语系, 上海

收稿日期: 2024年6月23日; 录用日期: 2024年8月20日; 发布日期: 2024年8月29日

摘 要

科学写作的一个显著特点是大量使用外壳名词来创建信息密集的文本。本研究采用语料库方法, 对比中英母语者在英语化工期刊写作中运用外壳名词的频率、词汇-语法型式, 旨在分析其在外壳名词使用方面的差异。研究结果表明, 中外化工领域学者在外壳名词以及名词-语法型式上的频率有显著性差异; 两个语料库中词汇语法型式以th-N和N-cl为主, 其中th-N的比重更高; 中国学者偏爱使用非限定性定语从句。

关键词

外壳名词, 词频, 词汇-语法型式, 中外对比

Comparison of the Use of Shell Nouns in English Chemical Journals by Chinese and English Scholars: A Corpus-Based Study

Zihang Zhang^{1*}, Yiyu Tang^{2*}, Zihan Yang³, Bo Guan^{2#}

¹International Elite Engineering School, East China University of Science and Technology, Shanghai

²School of Foreign Languages, East China University of Science and Technology, Shanghai

³School of European and Latin American Studies, Shanghai International Studies University, Shanghai

*共同第一作者。

#通讯作者。

文章引用: 张子航, 汤益愉, 杨子瀚, 管博. 基于语料库的化工领域中外学者外壳名词使用对比研究[J]. 现代语言学, 2024, 12(8): 626-634. DOI: 10.12677/ml.2024.128733

Abstract

A distinctive feature of scientific writing is the extensive use of shell nouns to create information-intensive texts. Using a corpus approach, this study compares the frequency and lexical-grammatical patterns of shell nouns used by Chinese and English native speakers in their writing in English-language chemical journals, with the aim of analyzing the differences in their use of shell nouns. The results of the study show that there are significant differences between Chinese and English scholars in the field of chemical engineering in terms of the frequency of shell nouns as well as lexico-grammatical patterns; the lexico-grammatical patterns in the two corpora are predominantly th-N and N-cl, with a higher weighting of th-N; and Chinese scholars favor the use of non-finite determinative clauses.

Keywords

Shell Nouns, Word Frequency, Lexico-Grammatical Patterns, Contrast between Chinese and English

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

外壳名词(Shell Nouns)是一类特殊的抽象名词, 如 idea、fact、notion、suggestion、finding 等, 它具有语义抽象性和笼统性, 犹如一个“外壳”, 其完整和具体意义只能通过上下文的信息填充和具化[1]-[3], 前面或者后面往往跟随表示具体内容的短语或者从句, 和这些具体信息构成“外壳-内含复合体”。不同的学者曾使用不同的术语来表达这类名词, 如泛指名词、前指名词、概括性名词、标示名词[2]和外壳名词[1] [4]。

经常使用外壳名词不仅是学术论文的突出风格, 也是衡量学习者高级语言能力的重要指标之一。因此, 对外壳名词的使用进行研究, 对于增强中文母语者在英语化工期刊中对外壳名词的使用意识, 从而提高语言表达能力有一定意义。

2. 文献综述

Schmid 指出, 外壳名词主要通过四种词汇-语法型式, 即“N-cl”、“N-be-cl”、“th-N”、“th-be-N”(见表 1)来实现与其具体内容和语义的连接, 其中“N-cl”和“N-be-cl”这两种型式为前指照应, “th-N”和“th-be-N”这两种型式为后指照应[1]。

在 Schmid 的理论基础上, Charles 采用语料库方法研究了“N-that”型式中外壳名词在两个完全无关的学科中构建作者立场的变化[5]。基于 COCA 语料库, 李健雪和蒋林娜通过定性定量相结合的方式对典型词汇-语法型式“N-be-that”中的外壳名词搭配关系、使用特点和语义特征进行了分析和总结, 丰富了外壳名词认知层面的研究, 为类似研究提供了借鉴[6]。此外, Mousavi 和 Moini 分析了 239 篇教育类期刊论文研究文章中的外壳名词[7]。研究结果表明, 一些外壳名词在语料库中出现的频率高于其他名词。

“the-N-of”和“the-N-in”是这些名词中出现频率最高的句型。董敏基于《时间简史》平行语料库,重点研究了“be-that-clause”、“this-N”和“to infinitives clause”后的三种补语模式[8]。讨论了相应的翻译技巧,如前置、后置、排序、分离固定结构和重组等。

研究者们还关注了文本类型对外壳名词在词语选择和语法结构方面产生的影响。例如, Flowerdew 对教科书和课堂演讲中的外壳名词进行了对比[2]。Aktas 和 Cortes 比较了发表的期刊论文与 ESL (English as Second Language) 学生写作中的外壳名词的使用频率、词汇-语法型式、功能、差异[3]。

Charles 发现,学术话语中的外壳名词具有学科属性,同一学科类别内的学术话语倾向于使用相对固定的高频壳词,不同学科中运用到的外壳名词或词汇-语法型式往往会有差异[9]。前人对外壳名词的学科特点研究主要针对教育学[7]、物理学[10]、语言学[11]等,而涉及化工领域尤其是基于较大批量语料的实证研究较少。

因此,本研究拟收集中英语母语者在化工期刊上发表的文章,在自建语料库的基础上比较中英母语者在化工领域英语期刊写作中外壳名词使用的差异。

Table 1. The lexico-grammatical patterns of shell nouns from Schmid
表 1. Schmid 的外壳名词词汇-语法型式

Function	Pattern	Abbreviation
Cataphoric	Shell none (N)-postnominal clause (cl)	N-cl
		N-that
		N-wh
		N-to
	Shell noun phrase (NP)-be-complementing clause (cl)	N-be-cl
		N-be-that
		N-be-wh
		N-be-to
Anaphoric	Demonstrative adjective (this, that)-(premodifier)-shell noun (N)	th-N
	Demonstrative pronoun as subject (this, that)-be-shell noun (N)	th-be-N

3. 研究方法

为探究中外化工学者英语期刊中使用外壳名词的特点和差异,本研究通过语料库对比分析方法,试图回答以下问题:

- 1) 英语母语者与中文母语者在化工领域外壳名词的使用频率上是否存在显著差异?
 - 2) 中外学者在常用外壳名词的选择以及所使外壳名词用的词汇-语法型式上有哪些异同和特点?
- 因此,本研究将分为以下三个步骤进行。

3.1. 语料库建立

本研究自建语料库,分别选取了影响因子较高的 14 本中文母语者和 12 本英文母语者在英语化工期刊上的文章作为研究对象(见表 2),体量分别为 600 篇和 575 篇,内容涵盖生物、无机、有机、物理、高聚物五个方向。构建英语母语语料库时通过作者的姓名、国籍以及工作单位等信息确认了其英语母语者身份。

Table 2. The information of journals home and abroad
表 2. 国内外期刊信息

国内期刊	国外期刊
中国化学快报	Chemical Industry and Chemical Engineering Quarterly
催化学报	Industrial & Engineering Chemistry Research
石油勘探与开发(英文版)	ACS Catalysis
能源化学杂志	Angewandte Chemie
高等学校化学研究(英文版)	Science
绿色能源与环境(英文版)	Nature Chemistry
无机材料学报	AIChE Journal
分析化学	Journal of the American Chemical Society
中国科学: 化学	Chem
中国化学工程学报(英文版)	Journal of Catalysis
分析检测(英文版)	Chemical Reviews
纳米研究(英文版)	Accounts of Chemical Research
材料组学杂志	
结构化学	

化工文本中图、表、图形符号、特殊字符等较多, 转换为语料库时文本较乱, 为不影响文字处理及研究结果, 本研究参考闫鹏飞、谢文龙的方法, 遵循最大程度原文原则, 对语料进行了降噪, 尽可能保留期刊文章中的化学物质的信息和专业内容[12]。对于图表部分, 研究保留了表头部分, 删去了图表中的数据和化学物质名称; 在公式符号的处理方面, 未采取直接整体删除的方式, 而是运用 Python 和 PowerGREP, 通过替换特殊字符进行删除和人工核对, 在保证语法准确、句法结构和语义表述相对完整的同时, 替换或简化图形格式及特异字符, 以最大程度地保留专业信息。此外, 通过撰写正则表达式和人工处理结合的方式删除了作者信息、参考书目等非正文文本的内容, 经过一系列的降噪我们最终得到 Native Speaker Corpus (NSC)和 Chinese Speaker Corpus (CSC)两个语料库(见表 3)。

Table 3. The information of corpora
表 3. 语料库信息

语料库	选取文章数量	字符数
NSC (Native Speaker Corpus)	575	3,787,700
CSC (Chinese Speaker Corpus)	600	2,100,740
总计	1175	5,888,440

3.2. 外壳名词确定

Schmid 指出, 外壳名词语义上都有结构性的、固有的、需要被上下文填补的语义空缺, 但是, 语义标准有时较难判定, 因此判断一个名词是否是外壳名词可以看它能否被用在典型的词汇 - 语法型式里[1]。本研究参考 Schmid、Flowerdew 和 Forest 以及刘芹和王心怡提供的外壳名词的词汇 - 语法型式列表用来判定外壳名词[1] [11] [13]。

Table 4. Shell nouns' CQL language form
表 4. 外壳名词 CQL 语言型式

外壳名词 CQL 语言型式		
词汇 - 语法型式	缩写	CQL 语言表达式
Shell noun + complement clause (N-cl)	N-that	[pos="N.*"] [lemma="that"] [pos="P.*"]
		[pos="N.*"] [lemma="that"] [pos="N.*"]
		[pos="N.*"] [lemma="that"] [pos="DT"]
		[pos="N.*"] [lemma="that"] [pos="JJ"]
Shell noun + be + nominalization (N-be-cl)	N-wh	[pos="N.*"] [pos="WRB"] NOUN, which
	N-be-that	NOUN BE that
	N-be-to	NOUN BE to
	N-be-doing	NOUN BE GERUND
	N-be-wh	NOUN BE wh.*
		NOUN BE DETERMINER .* NOUN
		NOUN BE DETERMINER .* .* NOUN
		NOUN BE DETERMINER .* .* .* NOUN
Demonstratives/which + be + shell	this-be-N	this BE DETERMINER NOUN
	these-be-N	these BE DETERMINER NOUN
	that-be-N	that BE DETERMINER NOUN
	those-be-N	those BE DETERMINER NOUN
Demonstratives + shell noun (th-N)	this-N	this NOUN
	these-N	these NOUN
	that-N	that NOUN
	those-N	those NOUN
N-EP	N-such as	NOUN such as
	N-including	NOUN including
	N-for example	[pos="N.*"] [word=="for example"]
	N-like	NOUN like

考虑到 N-cl 型式中的 N-of 型式还常用来表示从属关系和量化单位, 这一型式识别难度较大, 且数量过多对分析造成一定困难, 研究将 N-of 型式进行了删减。此外, Flowerdew 和 Forests 以及 Schmid 都曾明确指出“名词 + 定语从句”结构中的名词并不是外壳名词, 因此本研究在检索 N-cl 型式时通过限定检索式和人工筛查排除了定语从句(见表 4) [1] [11]。例如:

(1) Foley and Bhan have proposed a method for DRC analysis degree of rate control to superimpose the additional condition that time constants of the analytical rate function are kept constant when taking the partial derivatives of rate with respect to the rate constants/free energies. (NSC)

(2) It was noted that phosphorylation resulted in significant destabilization of the unbound state while hav-

ing no impact on the bound state, a condition that would result in a universal increase for any protein/protein interaction involving the N-terminal domain. (NSC)

在(1)中该词汇语法结构复合表 4 中的[pos="N.*"] [lemma="that"] [pos="N.*"]这一词汇语法结构, 是同位语从句, 发挥了外壳名词 N-cl 语法型式的前置功能, 而(2)是一个定语从句, condition 被从句修饰, 并没有起到外壳名词的作用。

此外, “名词 + 列举性短语”在语篇中发挥后指功能, 是不可忽视的衔接手段[13]。因此我们在词汇 - 语法型式列表中加入 N-EP 型式。

最后, 在外壳名词最主要的实现型式 th-N 中, “this”, “these”, “that”, “those”这几个词均发挥了 th-N 型式前指的功能, 因此 th-N 也包含 these-N 和 those-N。表 4 为我们最终用以判定外壳名词的词汇 - 语法型式列表及其检索的 CQL 检索式。

3.3. 数据处理

本研究根据检索式检索和人工筛查判定外壳名词后, 统计两库中外壳名词的数量, 并使用对数似然检验, 定量分析中外学者在外壳名词使用频率以及词汇 - 语法型式上是否存在显著差异。

此外本研究统计了两库频率在前五十的外壳名词, 并从频次在前五十中独有的高频外壳名词、各词汇 - 语法型式高频外壳名词、高频名词各个语法型式所占的比重三个维度进行了外壳名词使用差异的定性分析。

Table 5. Log-likelihood ratio
表 5. 对数似然值

	NSC 频率	CSC 频率	对数似然值	
外壳名词总频率	61250	24242	2068.61	+
N-cl	17562	8798	61.34	+
N-be-cl	9364	4639	40.03236107	+
Th-be-N	241	49	50.22220498	+
Th-N	27832	8467	62580.70981	+
N-EP	6251	2296	302.3740134	+

注: “+”表示 NSC 库对 CSC 库的过度使用; “-”表示 CSC 库对 NSC 库的过度使用。

4. 研究结果

中外学者使用外壳名词的对数似然检验结果见表 5。英文母语者外壳名词使用频词显著高于中国学者(p < 0.01; LL = 2068.61), 可见中国学者在化工领域的正文中使用外壳名词的意识较为薄弱。同时, 中国学者使用各个型式的频率均低于外国学者。这一发现与 Flowerdew 的发现相似[14]。

具体来说, 英语母语者使用 Th-N 的频率显著大于中国学者, 说明英语母语者更善于或习惯用 this+ 外壳名词总结前文, 通过检索 this/that/those/these 发现我国学者更倾向于仅用 this/that/those/these 做主语, 这一定程度上反映了中国学者不善于用外壳名词来表达概念范畴, 这可能是专业概念不够清晰和英语词汇缺乏造成的。

此外, 在检索过程中发现, CSC 中 NOUN+, which 结构出现了 6385 次, 占 N-wh 总数的约 80%; 而在 NSC 中, 这一结构只出现了 7409 次, 占 N-wh 总数的约 50%。这种显著的差异表明中国学者更倾向于用 N+, which 来具化前面名词的内容。

4.1. 频次前五十中独有的高频外壳名词

分类梳理结果如表 6 显示, 中国学者语料库前五十高频名词中独有的外壳名词有 22 个。第一类是语域界定词, 这些词的存在可能是因为所选化工期刊的具体研究领域不完全一致, 因此研究价值有限[13]。第二类为学术英语体裁中常见的词汇, 如 advantage、field、phenomenon、reason、research、characterization 和 introduction。而在 NSC 中, 独有的这类词包括 difference、hypothesis、observation、example 和 performance。通过这两组词, 可以发现中国学者并没有使用 hypothesis 和 example 这两个词。此外, 中国学者在使用 example 这个词时, 几乎全部采用 for example 和 as an example 的表达方式, 而外国学者除了这两种表达, 还常用 this example 这一外壳名词语法结构。这反映了中国学者在化工文章中对例子的分析以及对化学反应后续假设的缺乏, 更加注重研究对象的优缺点以及解释各种现象及方法的原因, 较为主观。

Table 6. The first fifty high-frequency shell nouns of the two corpora
表 6. 两库前五十高频外壳名词

语料库	词汇列表					
NSC	study	work	result	method	approach	reaction
	system	case	material	process	review	model
	condition	section	technique	observation	structure	catalyst
	effect	property	application	fact	value	experiment
	example	area	parameter	strategy	site	protein
	group	datum	type	ligand	molecule	step
	time	transformation	paper	solvent	difference	note
	hypothesis	region	specie	surface	absorption	adsorption
	preparation	synthesis	characterization	introduction		
CSC	work	result	study	paper	method	material
	process	article	reaction	catalyst	structure	sample
	system	condition	property	phenomenon	temperature	reason
	value	polymer	compound	field	peak	problem
	time	film	surface	case	experiment	parameter
	effect	model	fact	group	membrane	molecule
	technique	site	advantage	approach	solution	area
	research	acid	strategy	application	layer	ion
	composite	inset	performance	absorption		

4.2. 中外学者各构式高频外壳名词

由于 Th-be-N 词汇 - 语法型式数量过少, 研究结果不具有代表性, 并且在 N-be-cl 词汇 - 语法型式中出现的外壳名词的频次并没有显著的差异, 因此本研究只对 N-cl、N-EP、以及 Th-N 这三组词汇 - 语法型式的结果进行了分析。

分析结果表明, 在 N-cl 词汇 - 语法型式中, 中外学者均频繁使用 fact、method 和 process, 但外国学者更常用 case 和 model 这两个词, 如表 7 所示它们在 NS 库中的使用频率均位列前十, 而中国学者对这

两个词的使用频次较低。外国学者在 case 后主要接特殊疑问词(如 which), 而 model 较均衡地接 that、which 和其他特殊疑问词。sample 也呈现出相对均衡的搭配。

Table 7. The first ten shell nouns in N-cl pattern
表 7. N-cl 型式前十外壳名词

语料库	词汇列表及词频									
NSC	fact	system	note	case	reaction	structure	site	model	condition	method
	65.47	48.84	48.57	38.28	37.49	31.68	31.15	30.36	28.78	106
CSC	process	material	structure	fact	temperature	sample	catalyst	method	surface	site
	48.56	44.27	43.31	33.80	33.43	30.47	29.51	29.038	28.56	59

如表 8 所示, 在 Th-N 中, work、result、study、method 和 case 在两个库中均高频出现。观察 Th-N 型式中的同义或近义词, 发现中国学者更倾向于使用 kind (词频 47)而非 type (词频 20), 外国学者则相反, 偏向于使用 type (词频 187)而非 kind (词频 17)。在引用文献时, 外国学者更多使用 review, 而中国学者则较多使用 research。

Table 8. The first ten shell nouns in Th-N pattern
表 8. Th-N 型式前十外壳名词

语料库	词汇列表及词频									
NSC	study	work	result	approach	method	case	reaction	system	process	review
	257.148	249.23	212.53	162.90	142.83	139.13	114.32	102.172	101.38	380
CSC	work	result	study	paper	method	article	phenomenon	reaction	process	case
	441.28	263.25	235.15	184.22	119.48	119.01	46.65	45.71	43.33	71

在 N-EP 型式中, 两个库中的外壳名词高度重合, 在该语法结构中前十位的外壳名词中 property、material 和 application 频次较高。然而, 中国学者高频使用了 advantage 这个词, 而它在外文库中的出现频率并不高。

4.3. 高频外壳名词型式占比

两库中共有 28 个外壳名词在前 50 名, 其中中外学者在使用 fact 时高度一致, the fact that 这一表现形式在两个库中都占 90%左右。有 16 个词在各语法型式中的使用比例差距较大。去除语域界定词后发现, 中国学者更倾向于在 N-cl 词汇 - 语法型式中使用 process、model、technique、effect、area、value 和 time, 而外国学者更倾向于使用 th-N 或 N-EP 型式。此外, 中国学者在 time 的使用中, N-cl 型式占比为 56%, 而外国学者为 41%。其中, time 在 CS 库 N-cl 型式中接近 80%表现为 time when 或 time, which 而外国学者的 time that、time when、time, which 三种表现形式占比相对均衡。

虽然 application 在两库中均以 th-N 型式为主导, 但外国学者确更偏向使用 N-cl 型式, 并且几乎全部以 application 后加特殊疑问词的形式出现, 而中国学者在 application 后多使用 that。

这种差异可能由多方面因素造成。首先, 中文母语者受到其第一语言的语法结构和思维模式影响[15]。其次, 在化学领域, 当需要解释结构或材料时, 中国学者更倾向于使用 N-cl 型式来修饰限定名词, 而外国学者更偏好使用已知的化学结论进行类比。

5. 总结与讨论

研究结果显示,无论是外壳名词的种类数量还是使用频率,英语母语者都高于中文母语者。这表明,国内化工领域学者在使用外壳名词的意识上稍欠缺于英语母语者。在词汇语法型式方面,两个语料库中词汇语法型式以 th-N 和 N-cl 为主,其中 th-N 的比重更高,这与之前的研究结果一致。高频外壳名词具有鲜明的学科特性,如 reaction、structure、material 等,显示了化工领域注重事实的特点。中外学者在一些词汇和语法型式的选择上存在一定差异,反映了中国学者在学术写作上与英语母语者的不同风格,受到中文母语及英语教学的影响。英语外壳名词的错误使用和过度使用现象也存在,包括过多使用 Noun+, which 这一现象。

本研究也存在一定局限性。受到语料清理的影响,研究结果可能存在偏差。在词频分析方面,本研究考察了两个语料库中前五十个高频外壳名词,考察数量有限,化工学术语篇中外壳名词使用的频率差异和结构性特点仍有待研究。并且,考虑到语料库的时效性,本研究仅分析了近十年文献,没有体现语言随时间的变化。

参考文献

- [1] Schmid, H. (2000) English Abstract Nouns as Conceptual Shells: From Corpus to Cognition. De Gruyter Mouton.
- [2] Flowerdew, J. (2003) Signalling Nouns in Discourse. *English for Specific Purposes*, 22, 329-346.
[https://doi.org/10.1016/S0889-4906\(02\)00017-0](https://doi.org/10.1016/S0889-4906(02)00017-0)
- [3] Aktas, R.N. and Cortes, V. (2008) Shell Nouns as Cohesive Devices in Published and ESL Student Writing. *Journal of English for Academic Purposes*, 7, 3-14. <https://doi.org/10.1016/j.jeap.2008.02.002>
- [4] Gray, B. and Cortes, V. (2011) Perception vs Evidence an Analysis of This and These in Academic Prose. *English for Specific Purposes*, 30, 31-43.
- [5] Charles, M. (2007) Reconciling Top-Down and Bottom-Up Approaches to Graduate Writing: Using a Corpus to Teach Rhetorical Functions. *Journal of English for Academic Purposes*, 6, 289-302.
- [6] 李健雪, 蒋林娜. 外壳名词与 N-be-that 句式搭配研究[J]. 语文学刊: 外语教育与教学, 2014(12): 23-26.
- [7] Mousavi, A. and Moini, M.R. (2014) A Corpus Study of Shell Nouns in Published Research Articles of Education. *Procedia-Social and Behavioral Sciences*, 98, 1282-1289. <https://doi.org/10.1016/j.sbspro.2014.03.544>
- [8] 董敏. 外壳名词及其翻译对等的语法型式分析[J]. 外语教学, 2017, 38(2): 86-90.
- [9] Charles, M. (2003) 'This Mystery...': A Corpus-Based Study of the Use of Nouns to Construct Stance in Theses from Two Contrasting Disciplines. *Journal of English for Academic Purposes*, 2, 313-326.
[https://doi.org/10.1016/S1475-1585\(03\)00048-1](https://doi.org/10.1016/S1475-1585(03)00048-1)
- [10] Dong, M., Fang, A.C. and Qiu, X. (2020) Shell Nouns as Grammatical Metaphor in Knowledge Construal: Variation across Science and Engineering Discourse. *Lingua*, 248, Article ID: 102946.
<https://doi.org/10.1016/j.lingua.2020.102946>
- [11] Flowerdew, J. and Forest, R.W. (2015) Signalling Nouns in English: A Corpus-Based Discourse Approach. Cambridge University Press.
- [12] 闫鹏飞, 谢文龙. MatDEAP 材料科学学术英语语料库的创建[J]. 语料库语言学, 2020(1): 97-106.
- [13] 刘芹, 王心怡. 理工科硕士论文英文摘要中的外壳名词使用研究[J]. 外语界, 2016(2): 52-60.
- [14] Flowerdew, J. (2010) Use of Signalling Nouns across L1 and L2 Writer Corpora. *International Journal of Corpus Linguistics*, 15, 36-55. <https://doi.org/10.1075/ijcl.15.1.02flo>
- [15] Odlin, T. (1989) Language Transfer: Cross-Linguistic Influence in Language Learning. Cambridge University Press.
<http://doi.org/10.1017/CBO9781139524537>