Modern Linguistics 现代语言学, 2023, 11(6), 2624-2629
Published Online June 2023 in Hans. https://www.hanspub.org/journal/ml
https://doi.org/10.12677/ml.2023.116355

Hans汉斯

语料库语言学未来的发展路径研究

——以“基于语料库”的研究范式为视角

杨超帆

天津财经大学人文学院，天津

收稿日期：2023年5月11日；录用日期：2023年6月20日；发布日期：2023年6月30日

摘　要

语料库语言学为语言学领域继纯理论研究之后开启了一项全新的探索方式。人们不再只关注“内省”，凭感觉和意识归纳推理语言学理论，逐渐开始在数据收集、归纳与推理的实证研究上为语言学贡献自己的力量。为了探究语料库语言学未来的发展路径，本文就语料库语言学的起源与学科定位之争及其两种研究范式——“基于语料库”的研究方法和“语料库驱动”的研究方法和语料库语言学未来的发展进行研究。主要使用质化研究与量化研究相结合，以质化研究为主的方法，最终发现两种研究方法的哲学基础、学科属性、研究目的及操作方法等有很大的不同，而以语料库作为唯一数据来源的“语料库驱动”的研究方法很有可能会被语料库语言学所摒弃。

关键词

语料库，语料库语言学，研究范式，发展路径

Research on the Future Development Path of Corpus Linguistics

—From the Perspective of “Corpus-Based” Research Paradigm

Chaofan Yang

School of Humanities, Tianjin University of Finance and Economics, Tianjin

Received: May 11th, 2023; accepted: Jun. 20th, 2023; published: Jun. 30th, 2023

Abstract

Corpus Linguistics opens up a new way in the field of linguistics after pure theoretical research. People no longer only focus on “introspection”, inductive linguistic theories based on feeling and

文章引用: 杨超帆. 语料库语言学未来的发展路径研究[J]. 现代语言学, 2023, 11(6): 2624-2629.
DOI: 10.12677/ml.2023.116355

consciousness, and gradually begin to contribute to linguistics in the empirical study of data collection, induction and deduction. In order to explore the future development path of corpus linguistics, this paper studies the origin and disciplinary positioning of corpus linguistics and its two research paradigms: the corpus-based approach and the corpus-driven approach, and the future development of corpus linguistics. Qualitative research is mainly used and combined qualitative research with quantitative research, it is finally found that the philosophical basis, disciplinary attributes, research objectives and operation methods of the two approaches are very different, and the corpus-driven approach with corpus as the only data source is likely to be abandoned by corpus linguistics.

Keywords

Corpus, Corpus Linguistics, Research Paradigm, Development Path

1. 引言

20 世纪 60 年代，随着世界上第一个语料库——布朗语料库(the Brown Corpus)的建成，语料库语言学开始逐步进入到人们的视野当中，到 80 年代，语料库语言学开始在语言学及相关领域中普及开来。之后，在布朗大学的语料库团队、夸克(Randolph Quirk)团队、利奇(Geoffrey Leech)团队相互交流影响之下形成的“基于语料库”的研究范式也与“新弗斯”学派(the neo-Firthian School)中的一些学者所倡导的“语料库驱动”的研究范式之间形成了两相对立的阵营。

2. 两种研究范式的起源

2.1. 基于语料库的研究范式

根据梁茂成(2012) ([1], p. 326)在《语料库语言学研究的两种范式：渊源、分歧及前景》一文中对“基于语料库”的研究范式所作的起源分析，“基于语料库”的研究范式并不单单是由某个语言学家所提出的，而是由三个团队：布朗语料库团队、Quirk 团队、Leech 团队之间的相互交流和影响所形成的。

布朗语料库的建成标志着语料库语言学的发展进入萌芽时期。这一语料库由亨利·库塞拉(Henry Kučera)和纳尔逊·弗朗西斯(Nelson Francis)两位学者在 1963 年带领团队开始建设，并在次年完成(Léon, 2005) [2]。由于当时正处在结构主义语言学和转换生成语法相互称霸的时代，布朗语料库的出现并未获得国际语言学界的太多关注，但它仍为之后语料库语言学的发展奠定了坚实的基础。在二十世纪八十年代之后，布朗语料库才得到了语料库语言学界的关注。

1959 年，以夸克为首的语言学团队开始了“英语用法调查”(the Survey of English Usage)这一语料库的建设，在建设过程中，许多语言学家受到了夸克的影响，在完成项目之后开始了自己的语料库建设，这其中就包括杰弗里·利奇(Geoffrey Leech)，即后来的英国国家语料库(BNC)建设的负责人，夸克与利奇还有其他三人(Sidney Greenbaum、Jan Svartvik 和 David Crystal)被称为语料库语言学界的“五人帮”(梁茂成，2012) ([1], p. 325)。

在完成“英语用法调查”语料库的建设之后，利奇调入了兰卡斯特大学，并开始组建自己的语料库团队，并在 1994 年在团队的共同努力下完成了 BNC 的建设，这一语料库也成为后来语料库领域学者最

常用的文本数据源之一。利奇特别强调语法在“英语用法调查”语料库中的地位，“语法是一项大工程，它凭借的是英语用法调查所积累的研究”(Darnell, 2005) [3]。

三个团队之间的学者在私下有着良好的关系，弗朗西斯不仅筹备了布朗语料库，同时还是“英语用法调查”的项目组成员，Leech & Johansson (2009) ([4], p. 7)在描述 ICAME 由来之时也透露出了夸克还就任布朗语料库的顾问一职。夸克和弗朗西斯之间虽然研究的是同一方向，但他们各自的灵感来源是不同的，对于夸克来说，对口语的兴趣是最高的，而对于弗朗西斯来说，灵感在于计算机的力量(Leech & Johansson, 2009) ([4], p. 5)。话说回来，正是相同与不同的包容兼并才最终形成了“基于语料库”的研究范式。

2.2. 语料库驱动的研究范式

被称为英国历史上第一位语言学教授的弗斯(Firth)一直关注着话语中的搭配问题，并形成了对后人有着重要影响的语境论，这一论点也成为了后来“语料库驱动”研究范式的语言学基础。梁茂成(2014) [5]称搭配研究始终是语料库语言学研究中的重要内容，是研究意义最重要的方法。弗斯去世之后，由他的学生韩礼德(Halliday)和辛克莱尔(Sinclair)继承了他的思想，并形成了“新弗斯学派”。在之后，韩礼德逐渐由研究语境论转向研究他的功能语法，以至于后来只有辛克莱尔等人继续坚持弗斯的语境论，这也就标志着二人在语言学方向上逐渐走向了分歧，并形成了以辛克莱尔为首的新一代“新弗斯学派”(the New neo-Firthian School)，梁茂成(2012) ([1], p. 328)认为二者对于语料库的分歧成为了推动“语料库驱动”研究范式形成的主要原因。他们主张从语料库中建构理论、以全新视角对语言进行描写；他们关注搭配，同时也关注意义，并且认为搭配和意义是密不可分的。

3. 两种研究范式的分歧

托尼尼–博内利(Tognini-Bonelli, 2001) ([6], p. 99)首次对“基于语料库”和“语料库驱动”的研究范式进行了区分。在她看来，“基于语料库”的研究范式利用语料库对已有的理论或假设进行探索，目的在于验证或修正已有的理论；而“语料库驱动”的研究范式是一种以语料库作为实证基础的方法，词典编纂者从中提取数据并在没有事先假设和期望的情况下检测语言现象。斯托约翰(Storjohann, 2005) [7]认为在这样的研究方法之下任何结论或主张完全基于对语料库的观察。

3.1. 哲学基础

总体上来看，两种研究范式都有着来自古希腊的经验主义的哲学基础，这一主义在发展中形成了两类学派：“温和的经验主义”与“激进的经验主义”。前者认为，所有的意识观念均来源于知觉，但同时也承认意识的机能是内在的能力。相比之下，后者的观点则更为激进，公开宣称不仅意念的内容，而且其整个过程都不可能存在内部能力，而只能是习得的。而两种研究范式的哲学基础正好各自对应经验主义的两个学派。“基于语料库”的研究范式更倾向于“温和的经验主义”，后者更倾向于“激进的经验主义”，这是因为，一方面，前者以语料库作为数据源，主张文本长短一致，后者采取的是一种“自下而上”的探索型研究路径；另一方面，前者并不试图推翻已有的语言分析体系，研究中只是将语料库视为众多数据种类中的一种，而后者则主张一切源自语料库(梁茂成，2012) ([1], p. 332)。

3.2. 学科属性

两种研究范式所信奉的哲学源头不一样，发展路径自然也不一样。“基于语料库”的研究范式认为，语料库语言学并非一门独立的学科，而是一种研究方法，可以用来验证已有假说和理论(Gries, 2010 [8];

McEnery & Hardie, 2012 [9]; McEnery & Wilson, 2001 [10])。“语料库驱动”的研究范式则认为，语料库语言学是一门独立的学科，摆脱任何已有的语言分类体系和研究框架，从真实数据出发，对语言进行全新的描写(Tognini-Bonelli, 2001) ([6], p. 99)。

3.3. 研究目的

根据梁茂成等(2010) ([11], p. 425)，“基于语料库”的研究方法是实证研究方法的拓展，通过数据归纳和总结得出语言本身之外的一些结论，而“语料库驱动”的研究方法目的在于对语言本身进行描写。前者以语料库的真实文本为数据源，其解决问题所依据的理论不仅有语料库内的文本理论，也有语料库之外的，可以说其使用的手段更加传统；而后者则认为，解决问题要摆脱一切现有的理论，只从语料库中获取信息、发现规律，并进行对语言本身的系统性描写。

3.4. 操作手段

由于研究目的的不同，两者在操作层面上也存在着明显的差别，这明显体现在语料库作为数据源的数量方面。因为“语料库驱动”的研究方法目的在于对语言本身进行描写。所以它所用到的语料库数量一般是一个，而“基于语料库”的研究方法需要两个或以上的语料库作为数据源进行系统性的比对，这包括存在于语料库当中的语言变体及语言文类，所以，对于主题词的分析是“基于语料库”研究方法的主要手段，而“语料库驱动”的研究方法则着重强调对于索引行(concordance lines)的解读，在解读的过程中，不局限于“词”的研究，同时引入“短语”的概念，对词项、搭配、类联接等内容进行分析。

4. 发展与应用

4.1. 语料库语言学的未来发展

自上世纪 60 年代语料库语言学发展以来，从最开始因为乔姆斯基的转换生成语法和语料库语言学产生的冲突导致其发展缓慢，到 80 年代后的迅猛发展，如今已经逐渐成为语言学家和其他领域研究者的主要研究方法和手段。语料库语言学不再是那些熟悉计算机的学者才能掌握的独门绝技，它正成为语言研究相关领域的一块“香饽饽”，被各领域的学者熟知及熟练地运用，这其中包括语言定量分析、词典编纂、作品风格分析、自然语言理解和机器翻译等领域。

在未来的发展过程中，语料库语言学所带给人们的很可能不再是与语言相关的专业领域知识，其中所蕴含的研究方法与理论能够被更多的领域学习并推广，例如在医学领域当中，将一门疾病中的患者归类，标注出他们的身高、体重、年龄、生活习惯、饮食习惯、职业及工作环境等主要数据，依靠语料库语言学的索引等操作方式，以定量定性相结合的混合方法研究疾病成因，这对预防相关有潜在患病风险的人群有着至关重要的作用。

4.2. 语料库的国内应用

在当今，语料库的使用多数还囿于语言文字工作者和外语类研究生的使用中，而对于我国学习英语人口基数最大的中小学学生们来说，语料库的应用始终还未进入到课堂教学之中，曾主持过高考英语标准化改革的桂诗春教授在 2010 ([11], p. 420)年所提倡的语料库进入课堂教学也远远没有成为现实，要想将语料库纳入到国家义务教育的培训体系之中，并真正做到语料库技术与课堂教学的完美结合，需要学科领域内各阶层的共同努力，而这一愿景的实现终究是道阻且长的。如果语料库语言学成功进入课堂，并与教师的教学模式相关联，我们所面临的又一个挑战便是它的使用问题，在这一问题上梁茂成(2009) [12]在其发表的论文《微型文本及其在外语教学中的应用》内有过详细的讨论，在此就不再赘述。

5. “语料库驱动”研究范式的局限性

对于“语料库驱动”的研究方法来说，因为其自身哲学基础的原因，存在一定的局限性：这种研究范式以“激进的经验主义”为哲学基础，完全采用自下而上的方法，否定前人的所有研究，在观察索引行、对语言现象进行分析和归纳时，又难以摆脱现有知识体系的影响(梁茂成，2012) ([1], p. 332)。这样的一种研究范式太过极端以至于难以生存，更与如今学术界倡导的理性主义与经验主义相融合的学术思潮相悖。

另一方面，因为其研究手段的特殊性，完全将语料库本身之外的其他数据源排除在外，如诱发数据等，而通过诱发手段获取数据已经成为学术研究中十分常见的研究手段。如此一来，“语料库驱动”研究范式就很难被其他领域的研究者所接纳，与如今的跨学科理念相悖，也不被研究学界所倡导。

而对于“基于语料库”的研究范式来说，当前的语言研究很大程度上都是在已有的理论框架下展开的，研究者们并不需要很多专门的训练就可以操作基于语料库的研究范式。因为这种研究范式与常见实证研究方法的相似性，可以说，基于语料库的研究范式似乎更能够为学科背景各不相同的学术英语研究者广泛接受(姜峰，2019) [13]。

6. 结语

语料库语言学的发展历史并不长，但为语言学及其他领域的实证研究提供了大量可靠且分类明确的数据源，利用语料库的研究方法进行科学研究也成为当今跨学科领域的主要研究方法之一。在探究了其起源、分歧之后，未来的发展路径也是研究者们关注的焦点，两种研究范式之间的争论为学者们指明了方向。

两种研究范式的相互竞争贯穿于语料库语言学的发展过程之中。“基于语料库”的研究范式以“温和的经验主义”为哲学基础，发展过程中受到了美国结构主义的影响。而“语料库驱动”的研究范式则以“激进的经验主义”为哲学基础。这两者因它们信奉的哲学思想不同而在诸多方面出现了分歧，包括学科属性、研究目的和操作手段等方面。

随着语料库语言学的不断发展，“语料库驱动”的研究范式所展现出来的弊端也越来越多，而与其相对的“基于语料库”的研究范式则呈现出蒸蒸日上的发展态势，在今后的某一天，由于“语料库驱动”的研究范式的局限性和排他性，其很有可能会被语言学界和学术界所淘汰。

参考文献

[1] 梁茂成. 语料库语言学研究的两种范式: 渊源、分歧及前景[J]. 外语教学与研究, 2012, 44(3): 323-335+478.

[2] Léon, J. (2005) Claimed and Unclaimed Sources of Corpus Linguistics. *Henry Sweet Society Bulletin*, **44**, 34-48. https://doi.org/10.1080/02674971.2005.11745607

[3] Darnell, R. (2005) Linguistics in Britain: Personal Histories. *Journal of Linguistic Anthropology*, **15**, 276. https://doi.org/10.1525/jlin.2005.15.2.276

[4] Leech, G. and Johansson, S. (2009) The Coming of ICAME. *ICAME Journal*, **33**, 5-20.

[5] 卫乃兴, 李文中, 濮建忠, 梁茂成, 何安平. 变化中的语料库语言学[J]. 解放军外国语学院学报, 2014, 37(1): 1-9+159.

[6] Tognini-Bonelli, E. (2001) Corpus Linguistics at Work. John Benjamins, Amsterdam. https://doi.org/10.1075/scl.6

[7] Storjohann, P. (2005) Corpus-Driven vs. Corpus-Based Approach to the Study of Relational patterns. *Proceedings of the Corpus Linguistics Conference* 2005. Birmingham, 14-17 July 2005, 1-20.

[8] Gries, S. (2010) Corpus Linguistics and Theoretical Linguistics: A Love-Hate Relationship? Not Necessarily. *International Journal of Corpus Linguistics*, **15**, 327-343. https://doi.org/10.1075/ijcl.15.3.02gri

[9] McEnery, T. and Hardie, A. (2012) Corpus Linguistics: Method, Theory and Practice. CUP, Cambridge. https://doi.org/10.1093/oxfordhb/9780199276349.013.0024

[10] McEnery, T. and Wilson, A. (2001) Corpus Linguistics. Edinburgh University Press, Edinburgh.

[11] 桂诗春, 冯志伟, 杨惠中, 何安平, 卫乃兴, 李文中, 梁茂成. 语料库语言学与中国外语教学[J]. 现代外语, 2010, 33(4): 419-426.

[12] 梁茂成. 微型文本及其在外语教学中的应用[J]. 外语电化教学, 2009(3): 8-12.

[13] 姜峰. 语料库与学术英语研究[M]. 北京: 外语教学与研究出版社, 2019.