

导读

《语料库语言学群言集》一书收录了对 14 位国际知名语料库语言学专家的书面访谈，外加该书第一主编 Vander Viana 的一篇论文作为结尾，全书正文共计 15 章。另有 Ronald Carter 所写的序言及三位编者的前言。

该书编者在采访 14 位语料库研究学者时，请每位学者围绕 10 个给定的问题作答，其中七个问题为共答题，另三个问题则因各人专业特色而异。

1. “语料库十问”之共答七问：语料库语言学的本体

14 位学者都必须回答的七问题是：

- 1) Where do you place the roots of Corpus Linguistics? And to what do you attribute the growth of interest in the area? (您认为语料库语言学的学术源头在哪里？您认为这个领域持续受人青睐的原因是什么？)
- 2) Is Corpus Linguistics a science or a methodology? Where would you situate Corpus Linguistics in the scientific or methodological panorama? (语料库语言学是一门学科还是一种方法论？在学科和方法论的大背景下，您会将语料库语言学置于怎样的地位？)
- 3) How representative can a corpus be? (语料库应当具有怎样的代表性？)
- 4) How far should an analyst rely on intuition? (分析者应在多大程度上倚赖个人直觉？)

- 5) What kind of questions should an analyst think of? (分析者应关注哪些研究选题?)
- 6) What are the strengths and weaknesses of corpus analysis? (语料库分析的强项和弱点有哪些?)
- 7) What is the future of Corpus Linguistics? (语料库语言学的未来将会怎样?)

这些问题涉及语料库语言学的核心学科属性,从学科起源、语料库自身特点和研究者素质谈起,一直延伸至语料库语言学的未来发展方向。

问题 1 和问题 2 探讨的是语料库语言学在语言学史上的位置,包括它自身的学科特性和方法论意义。受访的 14 位学者中,有些认为语料库是计算机时代的产物,语料库语言学的历史可以追溯到 20 世纪 60 年代美国布朗大学的美国书面英语语料库项目。然而,更多的学者倾向于将采集真实语言素材、运用量化统计方法描写和分析语言的历史追溯到 20 世纪上半叶,特别是以 Bloomfield 和 Boas 等为代表的美国结构主义学者所开展的语言描写研究。应当说,就语料库语言学的研究理念而言,语料库语言学并非是在机读语料库出现之后才产生的,但机读语料库和相关语料库分析工具的出现大大促进了语料库语言学的发展。

问题 3 关注的是语料库作为语言资源 (resource) 的这一属性。语料库的海量规模往往为人津津乐道,然而除去“体量巨大”这一优势,语料取样代表性的价值更为关键。可是“代表性”(representativeness)这一概念着实难以界定。不同学者对取样代表性的认识分歧不小,受访者们对这一问题的回答,很大程度上是各自学术背景的投射。从事语言教学研究、翻译研究和语法研究的学者,对教学型语料、翻译或多语语料以及通用语言事实的需求各不相同,因此对语料库的代表性理解自然不同。可见,语料库取样代表性有其相对性,很大程度上取决于研究目的。

问题 4 和问题 5 涉及的是语料库语言学的研究 (research) 属性, 或者说学科性。其中问题 4 谈到语料库语言学对研究者的要求。面对海量数据以及量化分析结果, 研究者是做一个客观描述者, 还是结合个人语感以及相关语言学理论对数据加以分析和阐释, 这其中隐含了语料库语言学领域内部“基于语料库”(corpus-based) 和“语料库驱动”(corpus-driven) 两种学术路径的迥异观点(参阅梁茂成、李文中、许家金 2010: 第 6 章; 梁茂成 2012)。前者不排斥个人知识和语言学理论介入语料分析, 后者则主张少用或不用已有的理论。目前多数研究是按基于语料库的思路开展的。然而, 若能有更多人坚持按语料库驱动的思路开展研究, 则更有可能得到原创性的语言学发现。

问题 5 试图了解的是语料库语言学擅长解决哪些研究问题。这一问题引发了 14 种不同的答案。不同学术背景的学者向读者们展示了语料库语言学应用的广阔前景。

通常一个研究领域优势突出, 弱点也会比较明显。问题 6 谈的就是这个问题。语料库语言学的长处似乎尽人皆知, 倒是它的局限性更值得一说。有意思的是, 不少作者在面对这个题目时, 并不承认语料库语言学自身有什么明显的缺点或不足。

比如, Leech (见 179 页) 明确指出: “I do not see any weaknesses in Corpus Linguistics *per se*. (我不认为语料库语言学本身有什么缺陷。)” Leech 认为很多时候问题出在对语料库语言学的应用上, 而语料库作为一种数据源自身并没有什么问题。他反对将语料库作为语言学家的唯一数据和证据来源, 他认为, 如能结合语感判断、诱导数据 (elicited data, 包括心理语言学实验数据) 等方法, 可以更有效地推进这一学科的发展。对于有人认为语料库研究只注重对语言事实进行描述, 不注重理论建构的言论, Leech 以 *Corpus Linguistics and Linguistic Theory* 这个杂志的创立为例, 说明其实已经有越来越多的研究在关注语料库数据与

语言学理论的结合。对于 Widdowson 等人对语料库分析脱离语境的批评, Leech 表示这只是因现阶段技术所限, 目前很多语料库研究更关注文本内部词汇语法特征的频数分析, 事实上语料库语言学从来不排除包含更完整语境信息的音视频材料。随着语音识别技术的发展, 语义和语篇自动分析技术的发展, 基于语料库的研究一定会发挥更大的作用。

再如 Gries (见 98 页) 在谈及这一问题时指出, 如果说语料库语言学有什么缺陷的话, 那么问题主要在于研究者们对语料库及语料库方法使用得不够恰当, 而不应归咎于语料库语言学本身。比如, Gries 注意到, 很多已发表的研究中统计方法运用不尽合理。

在谈及语料库的未来走向(问题 7)时, 受访者们都期待有更多、规模更大、语境信息更丰富的语料库以及更先进的语料库工具出现, 期待语料库能解决更多的语言学问题。与此同时, 学者们似乎越来越倾向于淡化语料库语言学的学科地位。比如 Leech (见 181 页) 说道: “... Corpus Linguistics, as a separate area of linguistics, will simply melt away, as a sub-branch of linguistics, because using corpora will be the most obvious natural way to do linguistics. (语料库语言学作为语言学的一个独立的分支领域, 很自然会消解掉, 因为运用语料库从事语言学研究是再显而易见和再自然不过的做法了。)” Sampson (见 226 页) 也有类似的观点, 他提到: “... in a sense I believe that Corpus Linguistics as such has not got a future. ... I would hope that its future is simply to fade away as a concept, because all concerned will take corpora for granted as one important set of tools in any linguist’s toolbox. (照此我认为, 某种意义上, 语料库语言学没有未来。我希望语料库语言学这一概念最终消逝, 因为那时所有有关的语言学家, 都理所当然地将语料库视为其工具箱中的一套重要工具。)”

然而, Louw (见 204 页) 对于上述这种认识表现出深深的担忧, 他觉得上述“工具论”或“方法论”是对语料库语言学庸俗不堪

的 (insultingly mundane)、下里巴人式的 (pedestrian)、混淆不清的 (hybridising)、言不由衷的 (dishonest) 认识。Louw 觉得现在对语料库的看法将会摒弃掉 Firth、Malinowski 和 Sinclair 开创的“搭配生义” (meaning by collocation) 这样具有远见卓识的观念 (参阅李文中 2016)。Louw 慨叹, 语料库语言学的巅峰时代在 20 世纪 80 年代, 那时伯明翰大学正在开展 COBUILD 语料库项目。他希望学者加强对语义韵和词语搭配的研究和应用。

伯明翰大学的 Teubert 和我国一些受 Sinclair 学术思想影响的学者也持有与 Louw 相近的观点, 认 Sinclair 的语料库研究思想为正宗, 其余都是误用或是盗用语料库语言学这个概念 (参阅 Mukherjee 2010: 371)。

在国内外语料库研究学界, 受 Sinclair 学术思想影响的语料库学者虽人数不多, 但影响不小。从学术发展的大势来看, 不同流派或不同学术观点的存在, 正是学术繁荣的迹象。这本《群言集》, 不只收一家之言, 足见其包容性。

我们认为, 运用语料库以及语料库方法开展语言学研究, 要比 Sinclair 的搭配研究和 COBUILD 相关课题来得更早。1959 年开始的“英语用法调查” (Survey of English Usage) 和 1962 年开始的布朗语料库项目早为大家所熟知。从《群言集》的“语料库十问”的第一问中, 大家可以读到很多更早的语料库研究课题。即便在我国, 相关研究开展也有近一个世纪的历史 (参阅 Xu 2015)。

Leech (见 173 页) 将语料库语言学概括为 a methodologically-oriented branch of linguistics (具有很强方法论导向的语言学分支)。他不认为语料库语言学是个简单的方法论, 他通过哈勃太空望远镜的例子说明, 就像天文望远镜可以促成或大大推动天文学或天体物理学这些学科的发展一样, 技术的发展跟学科的建立, 两者互为表里, 不可割裂。

上述七个问题可以说都是根本性话题, 因此, 该书虽是五年前出版,

但丝毫没有过时之虞。虽然每章都是相同的话题，我们却可以读到迥然不同的回答，丝毫不会有重复单调之感。

2. “语料库十问”之个性三问：作者阵容及特色领域

2.1 作者阵容

除了共答题，每位受访者还从各自研究领域出发，就三个个性化题目作了论述。这些个性化论题恰能反映各位受访学者丰富的学科背景，以及该书之兼容并包。

该书作者阵容可谓梦之队。且不以语料库研究成就论，Geoffrey Leech (1936—2014) 一生在英语语法、语义学、语用学、文体学、语料库语言学等诸多方面都作出了杰出贡献。他主要完成或合作完成的代表作有：《英语语法大全》(*A Comprehensive Grammar of the English Language*)、《语义学》(*Semantics*)、《语用学原理》(*Principles of Pragmatics*)、《小说文体论》(*Style in Fiction*) 和“英国国家语料库”(British National Corpus, BNC)。《英语语法大全》，俗称“夸克语法”，长期以来是我国语法教学界的权威基准；《语义学》中列举的七种意义已成为语言学学生的专业常识；《语用学原理》中提出的“礼貌原则”是礼貌研究的重要基石；《小说文体论》是文体学领域当仁不让的第一教材；“英国国家语料库”是近 20 年来语料库领域的业界标杆。Leech 可以称得上是一位世界级语言学家。

Stig Johansson (1939—2010) 是英语语法和对比语言学领域的杰出学者。

Ken Hyland 是学术话语领域最重要的学者之一。

John Swales 是体裁分析的开创者。

若以语料库研究的成就论，受访者中有语料库领域影响最大的分析工具 WordSmith Tools 的作者 Mike Scott，有“网络语料库”(Web

as Corpus) 时代最具影响的在线语料库平台“杨百翰大学语料库平台”(http://corpus.byu.edu) 的创建者 Mark Davis, 有基于语料库的认知语言学研究代表人物 Stefan Th. Gries, 有将语料库与话语分析和社会语言学相结合的代表性学者之一 Paul Baker, 还有语料库翻译学代表学者 Sara Laviosa, 等等。

另外, Stefan Th. Gries 和 Bill Louw 还是 2008 年在 Corpora List (语料库邮件列表) 讨论组上发起著名的“新兵训练营大论战”(Bootcamp Debate 或 Bootcamp Discourse) 的对立双方。他们给《群言集》供稿, 使得这本书能更全面地反映学界声音。

上述阵容可谓豪华, 称得上全明星团队。若是受访者中还能包括 Sylviane Granger、John Sinclair、Mona Baker、Douglas Biber、Tony McEnergy 则几乎可以称为完美。事实上《群言集》缘起于本书第一编者 Vander Viana 对 Sylviane Granger 的题为“Corpus linguistics, language learning and ELT”的访谈(Granger & Viana 2007)。读者可从网络上下载阅读。John Sinclair 的思想已部分地通过 Bill Louw 的访谈阐述出来。Mona Baker 的语料库翻译学思想也通过其弟子 Sara Laviosa 的访谈传达出来。Douglas Biber 的基于语料库的语域变异研究也通过其合作者 Susan Conrad 的访谈得以展示。Tony McEnergy 的学术观点已有 Geoffrey Leech 和 Paul Baker 两位代言人。

2.2 特色领域

该书编者在前言中通过三张表格, 将《群言集》各章涉及的话题、领域作了极好的概括。因此在这里译出, 以方便读者了解该书内容, 并根据个人兴趣选择阅读。

表 1. 研究所用语料及方法

话题	Topic	受访者
学习者语料库	Learner corpora	Guy Aston
平行和可比语料库	Parallel and comparable corpora	Stig Johansson
适用于译者的语料库	Suitable corpora for translators	Sara Laviosa
通用语料库及工具	Public-domain corpora and tools	Mark Davies
(元信息) 标记	Markup	Guy Aston
(语言学) 标注	Annotation	Geoffrey Leech
树库	Treebanks	Geoffrey Sampson
语料库对比	Corpus comparison	Stefan Th. Gries
对研究者的技术要求	Technological mastery and requirements	Mark Davies
技术发展速度	Speed of technological development	Mike Scott
编程知识	Programming knowledge	Stefan Th. Gries
统计学知识的掌握	Statistics control	Stefan Th. Gries

表 2. 各章节所涉及的语言学研究领域

话题	Topic	受访者
历史语言学	Historical Linguistics	Mark Davies
语用学	Pragmatics	Geoffrey Leech
句法学	Syntax	Geoffrey Sampson
社会语言学	Sociolinguistics	Paul Baker
话语分析	Discourse Analysis	Mike Scott
批判话语分析	Critical Discourse Analysis	Paul Baker
对比修辞	Contrastive Rhetoric	John Swales
隐喻	Metaphor	Tony Berber Sardinha
体裁	Genres	John Swales
语域	Registers	Susan Conrad

(待续)

(续表)

话题	Topic	受访者
语言复杂度	Language complexity	Geoffrey Sampson
专门用途英语	English for Specific Purposes	Mike Scott
写作中的人际特征	Interpersonal features in writing	Ken Hyland
跨语言研究	Cross-linguistic research	Stig Johansson
翻译研究	Translation studies	Sara Laviosa
语料库语言学视角下的语言变化	Changes in Corpus Linguistics	Stig Johansson
发展中国家的语料库	Corpora in developing countries	Tony Berber Sardinha

表 3. 语料库语言学在相关学科中的应用

话题	Topic	受访者
性别研究	Gender	Paul Baker
文学研究的社会功能	Social functions in literary research	Bill Louw
知识的社会建构	Social construction of knowledge	Ken Hyland
对传统文学概念的挑战	Challenges to traditional literary concepts	Bill Louw
文体研究	Style	Geoffrey Leech
大学中的学术语言素养	Literacy at the university	Ken Hyland
学术语言素养及教学	Academic literacy and pedagogy	John Swales
学生自主性	Student autonomy	Guy Aston
教师教育	Teacher education	Susan Conrad
大学文学课程	University literature curriculum	Bill Louw
英语 / 西班牙语外语教材研究	Teaching materials for English/Spanish as a foreign language	Tony Berber Sardinha
专业学术能力训练	Professional practice	Susan Conrad
语料库与传统翻译研究资源	Corpora vs. traditional translation resources	Sara Laviosa

从上述三张表格中我们可以看到，我国（包括中国香港特别行政区和中国台湾地区）学者，乃至整个东亚以及东南亚地区语料库学者的研究并没有走向世界的主流舞台。其中原因之一，还是我国及周边地区和国家的学者迄今在国际上用英文发表论著的数量尚不够多。另外，具有自身独特风貌的语料库研究传统并没有形成。巴西学者 Tony Berber Sardinha 似乎并不能代表发展中国家这个广大的群体。长期在中国香港任教的 Ken Hyland 的学术思想和观念还是源自英国。值得欣慰的是，我国学者在国际上发表论著的上升势头十分强劲。不久的将来，语料库语言学会在世界更广大的地区兴盛起来。

3. 对《群言集》的评价与阅读建议

《群言集》一书的编辑体例应当是当今读者所乐见的。每一章各自独立，读者可根据喜好选读，而不必一气读完。《群言集》基本上是同题讨论，各章相同话题的讨论内容可以相互参照。

从一般阅读经验来看，论文集的质量和可读性往往不及自成一体的专著。然而，这本《群言集》却卓然不群，使读者颇有读一书而览众山之感。

除了通读和根据个人研究兴趣选读相关章节外，书中 Geoffrey Leech、Stig Johansson 和 Stefan Th. Gries 这三位学者的访谈尤其值得推荐。前两位是已故的资深语言学家，在语料库语言学中浸淫达半个世纪，对语料库语言学的认识更为全面，观点更为平衡。Stefan Th. Gries 则是近十年左右崭露头角，当前炙手可热的中年语料库学者。他和同行学者一道，不断将语料库相关的统计方法往前推进，更重要的是，他将扎实的认知语言学背景与语料库量化分析手段进行了充分有效的结合。

《群言集》属于命题作文，各人因为时间安排问题、重视程度的差

异，所作出的回答质量终究还是有高下。当然，上述对访谈内容的评价很容易失于主观。导读作者只是一家之言，生怕自喻专家而误导读者。聪明的读者定然能自辨内容优劣。

北京外国语大学中国外语教育研究中心在编辑《语料库语言学》期刊时，曾策划“专家视点”和“问题共议”栏目，与本书策划的主题十分相似，着重关注中国语料库学者，也算是中国版的《群言集》，是对《群言集》的本土化补充。其中包括：桂诗春语料库语言学答客问、肖忠华语料库语言学答客问、卫乃兴谈语料库语言学的本体与方法、梁茂成谈语料库语言学与计算机技术、李文中谈基于语料库的文化表述研究等。诸位如果对上述话题感兴趣，也可以从期刊网址（<http://ylyy.chinajournal.net.cn>）免费下载阅读。

许家金

北京外国语大学中国外语教育研究中心

2016年5月

参考文献

- Granger, S. & Viana, V. 2007. Corpus Linguistics, language learning and ELT: An interview with Sylviane Granger. *Mindbite* (1): 11–14.
- Mukherjee, J. 2010. Corpus linguistics versus corpus dogmatism – pace Wolfgang Teubert. *International Journal of Corpus Linguistics* 15(3): 370–378.
- Xu, Jiajin. 2015. Corpus-Based Chinese studies: A historical review from the 1920s to the present. *Chinese Language and Discourse* 6(2): 218–244.
- 李文中, 2016, “新弗斯语料库语言学”考辩, 《外国语》(2): 30–38。
- 梁茂成, 2012, 语料库语言学研究的两种范式: 渊源、分歧及前景, 《外语教学与研究》(3): 323–335。
- 梁茂成、李文中、许家金, 2010, 《语料库应用教程》, 北京: 外语教学与研究出版社。