

# 立足学科史的共选论探索

许家金 北京外国语大学

**摘要:** 本文立足作者对语料库语言学学科史的梳理, 考察作者提出“共选论”的学科背景, 并扼要介绍作者围绕该理论在语料库语言学、话语分析和符号学方面的实证探索, 同时强调其对外语教学的启示。

**关键词:** 共选论; 学科史; 语料库语言学; 自主知识体系

[中图分类号] H319 [文献标识码] A [文章编号] 2096-6105(2026)01-0017-4

DOI: 10.20083/j.cnki.fleic.2026.0005

## An exploration of the Co-selection Theory grounded in disciplinary history

Xu Jiajin Beijing Foreign Studies University

**Abstract:** Drawing on the author's study of the history of corpus linguistics, this article examines the disciplinary origins of the Co-selection Theory, briefly reviews author's empirical investigations in corpus linguistics, discourse analysis, and semiotics in this light, and highlights its implications for foreign language teaching.

**Keywords:** Co-selection Theory; historiography; corpus linguistics; independent knowledge system

学科体系构建, 知难行难, 经年累月, 往往不得。不论怎样的体系均需与前贤和时贤对话。就语言学和外语教育理论而言, 新生的知识体系还必须直指语言运作机制和外语教育本质。

借由这篇短文, 我希望对自己语言观的形成作一浅析。诚然, 这尚难称为体系构建, 因为距离真正的构建相去甚远。我将从个人长期关注的语料库语言学视角谈谈知识体系的构建。语料库语言学兼具语言学本体和语言教学应用两个维度。在语料库语言学理念萌芽之初, 其目标之一便是回应并解决语言教学问题。下文将围绕语料库语言学中的“共选”(co-selection) 观念展开讨论。这一表述原本已有, 而我对其内涵作了一些拓展。在本文中, 我对共选的探讨最终也将落脚于外语的教与学。

我倾向于认为, 人文学科鲜有从0到1的迭代式创新, 而往往是累积式或螺旋上

升的发展模式。可以说，人文学科的学术发展史研究是知识体系创新之路的基石之一。学科史自然成为与知识体系联系最为紧密的体裁之一。基于对语料库语言学理念百年演进的追溯，我完成了《语料库语言学史》一书，该书有幸成为“外国语言学及应用语言学学科史丛书”中的一册。过去十余年间，我致力于语料库语言学学科史的整理，初心之一便是希望逐步明确个人的语料库语言学观念，英文可称作“my version of corpus linguistics”。在文献搜集和梳理过程中，我主要遵循寻根问祖、中流砥柱、同道中人、薪火相传的思路。目前这项工作已初见成果。有些史料也已通过网络对外发布，例如，我所整理出的“语料库经典文献”及“语料库语言学大事年表”等均可通过搜索引擎检索访问。另举两例，以显示我国最早具有语料库属性的研究与语言教育的关联。1922年，陈鹤琴主持完成《语体文应用字汇》。该汉语字频表后来成为朱经农、陶知行编写语文教材《平民千字课》（1923）的用字依据。外语教育方面，20世纪八九十年代的全国英语大纲词表便是基于杨惠中、黄人杰（1982）主持建成的JDEST（上海交通大学科技英语）语料库编制而成。

图1扼要绘制出我所梳理的语料库语言学重要学派及其理论发展脉络。在图1学术谱系基础上，我思考如何进一步凝练“语料库是理论”与“语料库是方法”两派学说的共通性。通过反复阅读与仔细推敲，我逐渐发现，两者的底层学科逻辑均围绕“共现”（co-occurrence）或“共选”（co-selection）展开。最后，我选定Sinclair（2004：20）提出的“共选”这一表述，并对其内涵作了拓展。我将语料库语言学视作一种功能语言学，即认为语言是人类用以表义的符号选择。也正因此，我采用了“共选”这一表述，而非“共现”这一描述性概念。起初，我将语料库语言学各派的这一共核理念概括为“概率语境共选论”（许家金2020a），后觉表述累赘，压缩为“语境共选”。我及合作者围绕“语境共选”开展的系列语言研究，以及部分图文多模态研究，已触及符号学层面。我因此愈发觉得“共选”这一表述更具包容性和抽象度。

“共选”首先是一种语言观。形式上，它指两个及以上符号的并置，包括但不限于词和词的共选、词汇与语法的共选、语法和语法的共选、语言与非语言符号的共选、非语言符号与非语言符号之间的共选，以及上述层面的综合共选。概率上高强度搭配的共选单位构成典型的表义单位。共选机制可有效区分词义、词类、语法范畴、语体、话语语用功能、艺术风格类别等，即所谓“范畴化”。两个或多个符号的有意义组合有明确的概率性。这种概率性共选的潜能在大语言模型时代得到充分展现。大语言模型“预测下一个词元”（next token prediction）的底层机制，正是符号间的概率共选。大语言模型的成功在一定程度上说明，“共选”对语言乃至非语言符号表征的意义与知识生成机制具有很好的解释力。

在形成共选语言观的同时，我也注意到语言学习中常见的词不达意、句不合式、话不中用，抑或母语习得之易、二语习得之难，皆可归结为具体情境中符号共选是否適切的问题（许家金2023）。为进一步服务教学，便于外语教学实践者更好理解和

接受共选理念，我尝试对共选作了通俗化表述，将其概括为“四用原则”：真材实用、优选常用、单词连用、情境活用（许家金 2022）。具体而言，即建立真实语言使用数据库，通过概率统计从中选取常用词汇。词汇学习应将经常性共选的词语搭配置于中心位置。同时，词语与语体及其他诸多场景因素之间的共选也应纳入学习范围。基于这一思路，我策划出版了19册面向不同行业职场词汇手册[如《新时代职业英语酒店英语词汇手册》（许家金 2020b）]，均遵循“四用原则”编写。

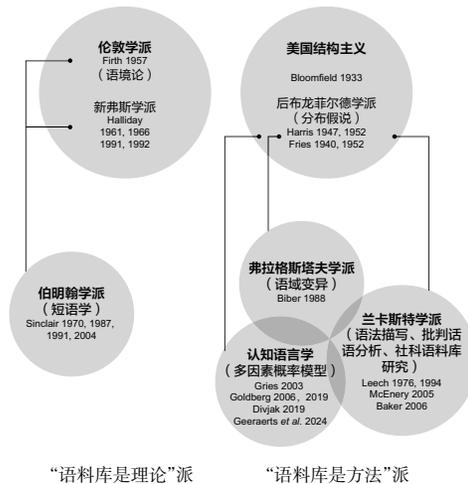


图1 语料库语言学学术脉络简图（许家金 2026）

如今，我们可以充分利用语料数据和大语言模型，特别是多模态大模型，研发融合共选理念、更为学生喜闻乐见的教学材料。在移动互联网与大语言模型时代，“数字多模态写作”（digital multimodal composing）能力对提升传播效果的价值尤其突出。若能结合共选理念，我们有望设计出更多多模态共选教学材料和教学活动。

沿着这一方向，我与合作者也初步开展了一些实证研究，涉及英语绘本中叙事矛盾冲突的推进、评价意义的建构、学术语篇图形摘要的图文意义共建，以及具有跨学科性质的西方绘画主题与流派研究等（参见Huang & Xu 2025, forthcoming; Li & Xu 2025; Ren & Xu 2025）。

综上，我对知识体系构建的理解是：应在充分梳理学科史的基础上，秉持继承与批判的辩证思维，对经典与传统保持充分尊重，一步一个脚印，从模仿式创新起步。语言学与外语教育的大部分理论源自西方。我们的知识体系构建的第一阶段往往难以摆脱西方窠臼；第二阶段通常表现为西方理论的本土化结合。第三阶段有时会进入一种“不中不西、又中又西”的理论建构状态。第四阶段，可能会出现对舶来知识体系批判继承后的本土自主创新。这或可对应沈家煊（2024）所谓“抛”与“捡”的关系。我国语言学研究还应在回答本土问题与时代命题的基础上审视所建构理论的本体价值。我以为理论建构应怀有创制“元理论”（metatheory）的志向。殊

不知, 结构主义与生成语言学在近一个世纪的人文思潮中产生过广泛影响。语言学也曾一度被视为领先学科 (Greenberg 1973)。鉴于语言承载意义与人类知识的巨大潜能, 语言学本体理论不单能服务于语言学和外语教育研究, 也可能成为跨学科研究的贯通力量。另外, 在构建新理论时, 还必须具备证伪意识, 要充分认识到自身理论在学科与情境上的适用边界。这既是一种谦逊, 也是一种科学精神。

最后, 我借用张忠培 (2018a, 2018b, 2018c, 2020) 的中国考古学文集的四本书名, 表达我对构建中国自主知识体系的理解: 尽到自己的心、说出自己的话、走出自己的路、永远在路上。

## 参考文献

- Greenberg, J. H. 1973. Linguistics as a pilot science [A]. In E. Hamp (ed.). *Themes in Linguistics: The 1970s* [C]. Berlin: Mouton de Gruyter. 45-60.
- Huang, D. & J. Xu. (forthcoming). Extending local grammar to multimodal meaning-making: A case study of conflict resolution in picture books [J]. *Text & Talk*. doi: 10.1515/text-2024-0095.
- Huang, D. & J. Xu. 2025. A multimodal local grammar of evaluation in picture books [J]. *Language & Communication* 105: 259-278.
- Li, W. & J. Xu. 2025. Exploring text-image relations in museum contexts: A multimodal analysis of sub-genre variations across interpretive captions. *Semiotica* 267: 195-223.
- Ren, J. & J. Xu. 2025. Ideational interplay of textual and visual elements in graphical abstracts of biology research articles [J]. *English for Specific Purposes* 78: 156-179.
- Sinclair, J. 2004. *Trust the Text: Language, Corpus and Discourse* [M]. London: Routledge.
- 陈鹤琴, 1922, 语体文应用字汇[J], 《新教育》(5): 987-995。
- 沈家煊, 2024, 《破立集》[M]。北京: 中国社会科学出版社。
- 许家金, 2020a, 多因素语境共选: 语料库语言学新进展[J], 《外语与外语教学》(3): 1-10, 21。
- 许家金, 2020b, 《新时代职业英语酒店英语词汇手册》[M]。北京: 外语教学与研究出版社。
- 许家金, 2022, 高职阶段行业英语词汇表编制与“四用”教学原则的实施[J], 《外语教育研究前沿》(1): 43-49。
- 许家金, 2023, 后经典时代语料库研究方法及其理论启示[J], 《外语教学与研究》(3): 442-454。
- 许家金, 2026 (即将出版), 《语料库语言学史》[M]。北京: 外语教学与研究出版社。
- 杨惠中、黄人杰, 1982, JDEST科技英语计算机语料库[J], 《外语教学与研究》(4): 60-62。
- 张忠培, 2018a, 《中国考古学: 尽到自己的心》[M]。北京: 故宫出版社。
- 张忠培, 2018b, 《中国考古学: 说出自己的话》[M]。北京: 故宫出版社。
- 张忠培, 2018c, 《中国考古学: 永远在路上》[M]。北京: 故宫出版社。
- 张忠培, 2020, 《中国考古学: 走出自己的路》[M]。北京: 故宫出版社。
- 朱经农、陶知行, 1923, 《平民千字课 (第一册)》[M]。上海: 商务印书馆。

## 作者简介

许家金, 北京外国语大学中国外语与教育研究中心教授、博士生导师。主要研究领域: 语料库语言学、话语研究、二语习得、语言对比与翻译。电子邮箱: xujiajin@bfsu.edu.cn