许家金谈语料库语言学的本体与方法*

北京外国语大学 许家金

1. 你怎么看语料库研究对语言学的贡献和意义？

基于电子语料库开展研究，在我国历史不长，若以人作比，不过刚过而立。而如今语言学天地里，语料库研究的学科身份，已日渐清晰、稳固。究其原因，语料库语言学理念符合人们据实论理的经验思维。只是过去囿于技术，人们无法考察海量语言事实罢了。

语料库研究之于语言学，至少有4个方面的意义。我们不妨概括为：用、量、聚、器。

1）“用”的意思是尊重语言事实、关注用法。用法是语法的源泉。过去的用法是今天的语法，今天的用法是明天的语法（Bolinger 1980：30；Newmeyer 2003；Bybee 2005，2006；Meyer & Tao 2005）。Wittgenstein（2009：86）更是明确指出意义不在别处，而在使用中。

2）“量”的含义，一方面指语法是语言实践长期频繁互动的结果。可以说，重复是人们言语交际的主旋律。老子所说的“万物并作，吾以观复”也有这层意思。语法化观点也认为，语言形式在经年累月的重复使用中，会出现实词虚化、音节简并等变化。这是上文有关尊重用法的继续。另一方面，“量”的含义是指，鉴于言语交际中重复是主导性的，则可以利用量化和概率统计手段来描摹（如描述统计）、验证（如显著性差异检验）、探索（如回归分析、因子分析）语言事实，这是语料库研究题中应有之义。换言之，语料库语言学视角下的语言事实是概率性的（probabilistic）（参阅Halliday 1991）。语料库研究观察和解释的是，多大概率人们会使用某一语言事实，而不是断言式地判定，某种说法合法不合法。语料库研究的“用”“量”观，本质上是描写语言学的立场；而“聚”和“器”这两项特征，观照的是如何去考察语言使用，用什么样的技术手段去观察人力所不能及的大量语言事实。

* 本文撰写得到国家社科基金项目“基于双语语料库的汉语复杂动词结构英译研究”（编号12CYY060）和教育部“新世纪优秀人才支持计划”（编号NCET-12-0790）资助，特此致谢。

3）“聚”试图揭示语言成分的聚集共现机制。使用中的语言在很大程度上是线性的。语素和语素、词语和词语、语句和语句、篇章和篇章之间一定有某种看不见的力量，使它们有效地聚合到一起，从而传情达意。这方面的研究，“扩展意义单位模型”（extended unit of meaning model，Sinclair 2004）可以较好阐释词汇短语聚集达意的机制。而类似于Biber（1988）提出的“多特征多维度分析模型”（multi-feature multi-dimensional analysis）将大量语言特征（少则五六十，多则上百）在不同功能维度（如，信息与互动维度、抽象概括与精细阐述维度）的聚集情况，可以量化方式分析出语体间的聚集或排斥，这是一种宏观层面的语言特征的聚集共现现象。

4）“器”与“工欲善其事，必先利其器”中的“器”同义。语料库语言学为研究语言事实的概率特征，提供了强大的工具箱。其中的各种工具可以对特定语言现象的用法、用量和聚集情况进行自动或半自动的分析。计算机应用于大规模语料的量化分析给语言研究带来的变化，就好比人类对星空的探索——从肉眼到光学望远镜，再到射电望远镜——所带来的变化。“聚”和“器”对语言学的贡献，既有技术带来的变革，也有对语言内在机制重新认识引发的变革。

2. 你如何描述语料库研究的本体论与方法论及其关系？

1）语言描写的短语学本体及方法

语料库语言学，作为语言学内部的新生研究领域，其本体理应立足于语言结构、意义、功能及其机制的探究。一言以蔽之，语料库语言学的本体在于语言描写，是典型的实证语言学（empirical linguistics）。坐拥大量语言素材，掌握“语言望远镜”这样的技术手段，可以鸟瞰语言的宏观面貌。这样的新型语言描写，可谓“计算机时代的田野调查”（许家金 2011）。语料库语言学发展至今，另一本体关切是对意义的追寻。意义研究是个宏大的议题，是许多学科的研究焦点。即便在语言学范畴内，意义依然是复杂、精密、难以琢磨的。伦敦学派创始人J. R. Firth在1951年的论文“意义的多种建构模式”（Modes of meaning）中将语义概括为“意义谱段”（the spectrum of meaning）（Firth 1951/1957：203）。其中包含高层级的意义建构模式（如语言使用者的身份特性和文化语境），以及低层级的意义建构模式（包括语音、音系、词汇、语法等层级）。Firth及后继者用力较勤的是对低层级的韵律意义和词汇意义的探究，其中最为专注的是词汇意义方面的探索。对意义研究方法论，Firth（1951/1957：194）提出一个专门术语——“搭配识义”（meaning by collocation），并借助大量诗歌、名人通信阐释其意义研究理念。

此后，Halliday（1961：276，1966：158）和Sinclair（1966）、Sinclair *et al.*（2004）将搭配识义这一语义研究方法加以发展，并运用统计算法，从大量语料中挖掘语

言使用中惯常搭配形式，从而确定特定词义或短语义。这些成果被成功应用于改造词典编纂，并引发了该领域的变革。在搭配识义理念之上，Sinclair提出“扩展意义单位模型”。这一模型兼顾词汇和语法，外加语义和语用，将单词意义扩展并锁定在短语意义之上。这一开创性理论框架突破了“句法本位”的语言学主流，也超越了“词汇本位”的静态语言观。扩展意义单位的分析方法，明确地与既有语言学理论保持距离，不用（或最低限度使用）已有概念和理论，形成了以短语为核心的语言观（见Sinclair 2004）。以扩展意义单位模型为方法论的短语学语言本体研究，是语料库研究领域具有原创意义的方法论和本体论的综合体。

诚然，扩展意义单位模型和短语学已自成体系，成果可观，但在语言学核心领域并未引起关注。认知语言学中的构式语法与短语学理念相近，但构式语法并非借鉴短语学，两者各自为政，未见有融通的迹象，或许仍需假以时日，或许因为“短语本位”语言观太过前卫，试图摒弃既有的语法范畴。客观来看，短语学视角的语言描写及应用，优势和局限并存，优势前文已有阐释，其局限有如下几点。

第一，短语不是语言的全部。短语学忽视了真实语言中存在大量孤零不靠的单词。Sinclair（2008）宣称语言中“除却短语，别无他物”（The phrase, the whole phrase, and nothing but the phrase[1]），这种说法过于绝对。短语在语言描写中地位重要，但除此之外，还有一些与短语配合使用的单用词语。比如，一些表达起承转合关系的连接词（如and、so、however）、评注性附加语（如用在句首后接逗号的frankly、hopefully）、口语中的语气词（如yeah、Jesus、well、like），话语中单用的小标题（如学术论文中的introduction、results、conclusion等）等。如果说短语是砖块，这些单用词语就是泥浆，两者相辅相成。Sinclair & Mauranen（2006：55-62）将上述单用词语也划归为“组织型词块”（organisation chunks），并细分为信息片段（MF，message fragments），不完整信息片段（M-，incomplete message segments）等。单词则为单词，短语则为短语。这种将单用词语也视为词块的做法有些削足适履。短语学目前在技术上难以解决短语切分的边界问题，也正是对语言构成认识不清造成的。短语边界的困局，还是习语原则和开放选择原则的互动所致。模糊的短语边界，恰是开放选择原则发挥作用之处。目前短语学更注重习语原则，关注线性语言单位的组合关系，忽视在一定语法位置上词汇语义和形态句法的聚合关系，因此“Sinclair的理论在实际应用中并非没有问题”（Hunston 2002：147），在某些情况下完全不起作用，或作用甚微（同上：145）。Hunston & Francis（2000）提出“型式语法”（Pattern Grammar以及“型式语流”（pattern flow）或“搭配叠泉”（collocational cascade，由Gledhill提出）这些概念，对于整合词汇语法，习语原则和开放选择原则，以及解决短语边界问题（参见Hunston 2002：146-148），提供了一些思路。然而，一种语言观若独尊短语，则必定行而不远。

第二，短语学对于更大的语言单位（传统意义上的小句、句子、语篇[2]）的意

义建构解释力有限。英国语言学有注重完整语篇的传统。Stubbs（1993：11）和Sinclair（1991：19）都主张“语言研究的单位需要是完整语篇”。他们认为完整语篇优于由样本构成的语料库，完整语篇可以促成我们开发更广泛的语言学研究选题。然而，McEnery & Hardie（2012：153）指出，尽管Sinclair、Stubbs等人如此主张，但在实际操作层面，绝大多数情况下他们都只专注于局部语境，即中心词左右各5个词的索引行，完整语篇极少得到真正的分析和利用。短语学研究者也谈语篇功能，严格来讲，他们谈的是语用功能。语篇概念在形式和结构上是“超句”单位。如果仅仅基于索引行这样的局部语境来谈语篇功能，恐怕更多属于研究者根据直觉和语感补足的语境。Sinclair本人是十分注重语篇研究的，比如他对课堂话语的研究（Sinclair & Coulthard 1975）;《笃信文本》（*Trust the Text*）一书接近一半的篇幅也是有关篇章组织结构的;《线性单位语法》（*Linear Unit Grammar*）也十分注重将口语和书面语的语篇同短语研究衔接起来。目前来看，词汇语法体系对接已在短语层面完成，但短语与语篇的对接还缺少完善的机制（参阅Hoey 2005）。

第三，不太关注词汇短语之外的意义。对于词汇语义趋向和积极消极语义色彩之外纷繁复杂的意义、功能等（如言语行为、话语互动功能、逻辑语义、非命题的语用意义等）并不涉及或涉及很少。前文提到的高层级社会文化意义的建构，短语学更是力有不逮。

第四，效率问题。从词语出发，将最小意义单位定位于短语，细致有余，抽象不足。从语言描写和语言习得效率两方面看，每个语言点都从词语入手，其结果很可能是产生数万条短语型式，这显然有违经济性原则。或许短语学的最佳用武之地是词典编纂和词汇研究与教学。词典是工具书，纵使篇幅一两千页，收词过万，无可厚非。目前，在词典编纂（Sinclair 1987a，1987b）和词汇教学中，例如，词汇大纲（Willis 1990；Lewis 1993；杨惠中、黄人杰 1982；肖忠华、许家金 2008；许家金2009）、通用学术短语表（Simpson-Vlach & Ellis 2010）、通用学术搭配表（Ackermann & Chen 2013）等，都有很多成熟的应用实例。

2）语言描写的话语语用本体及方法

如果意义研究是语料库研究的核心任务的话，语音、词汇、短语、句法这些低层级意义建构维度中，扩展意义单位模型对词汇短语层面确有充分描述，但语音、句法和高层级的语篇、社会语言学层面涉及不多，这些方面仍大有可为。这其中，Biber（1988）的“多特征多维度分析模型”在基于语料库的语篇研究方面作出了突出的成绩。而在更高层级的社会文化意义建构方面，国际范围内基于语料库的批评话语分析（critical discourse analysis）开展得如火如荼。语料库与话语分析具有天然的兼容性。它们在对语言本体的认识方面，都关注真实使用中的语言，都关注连句成篇的较大语言单位，在研究方法上都特别关注上下文语境和社会文化、历史、政治语境对文本意义的解读。可以说语料库研究与话语研究都称

得上是语言本体与方法论的无缝对接。翻开语料库研究期刊和专著，能发现大量借助语料库方法的性别语言、热点事件（恐怖事件、伊斯兰形象、国家形象、气候变暖现象、战争报道等）等方面的研究。这些高层级的意义建构研究，所涉及的意义主要潜藏于文本之中的社会、政治、文化、历史意义，在文献中更多使用的是“意识形态”（ideology）这一表述。

3. 相对于其他语言学领域，语料库研究的方法论特色是什么？

就影响而论，语料库研究的方法论波及更远。语料库研究方法广泛应用于词汇、语法、话语层面的语言描写，在语言习得、社会语言学、语用学、话语分析、翻译研究领域多有采用。再以话语分析领域为例，学术话语、法律话语、媒体话语、日常口头会话、学习者话语、文学作品、网络体裁都广泛借助语料库开展研究。

语料库研究方法的首要特色在于量化。超常规模的语料量，使得研究者可以纵横数百年（如布朗家族语料库、COHA语料库和赫尔辛基语料库等历时语料库），跨越几大洲（比如ICE中近20个英语变体的对比研究），团结巴别塔的后人（如欧盟二十多种语言构成的多语平行语料库）。如今百亿字词的语料库（如“北京语言大学汉语语料库”）已然问世。随着人们日常生活与（移动）互联网越来越不可分离，基于整个互联网进行语言分析变得离我们越来越近。事实上已经有一些此类成果，如Renouf主持创建的WebCorp检索平台，以及Kilgarriff为代表的“网络语料库”（Web as Corpus）的第四代语料库的相关研究及实践，参阅许家金、吴良平（2014）。

新工具的运用，使得纵横、跨越、团结成为可能。这恐怕是再长寿、再博学的语言学者，凭个人语感，所不能企及的。新工具可以同时观察到成千上万的语言实例，据此概括出来的语言规律是有可能提升我们对语言的认识的。这即Sinclair（1991：100）所谓“若是同时观察很多语例的话，语言看上去会迥然不同”（The language looks rather different when you look at a lot of it at once）。

4. 统计分析手段在语料库研究中的地位与作用如何？

频数统计是语料库统计的基本方式。通过语料库查询工具可以穷尽式地列举出语料库中所有的词汇、短语、词类、特定句法结构等，或者根据需要精确搜查某些（类）词汇或语言单位。一般在列举出语言结构的同时，语料库软件都会给出相应的出现频次。

对两个或多个语言特征共现频率（co-frequency），特别是对词语搭配的统计分析在语料库研究中十分常见。近几十年来，伴随着短语学的发展，有关搭配统

计的算法不下几十种，互信息、对数似然率、dice系数、Δp等算法层出不穷。搭配强度算法是在上述单个语言单位频数统计基础上进行的，通常是分别得到单词A和单词B的使用频数，然后检索A和B共同出现的频数，再综合考虑语料库容量等因素，从而计算出A、B两个词经常在一起出现的概率。

语料库研究中，为了进行对比分析（比如，英国英语和美国英语、口语和书面语、英语和汉语、中国人写的英文与本族语者的英文、男性和女性的语言、不同时代的语言等），需要将某个（些）语言现象在两类文本中出现的频数分别检索出，然后进行统计检验，以判别两项（类）频数是否存在显著性差异。较为常见的统计方法是卡方检验、对数似然率、费舍尔精确检验等。遇到较多文本类型，多变量情况下的频数差异对比，还会用到更高级的统计方法。如今，有越来越多的高级统计方法被引入语料库研究中。其中主要是针对多项语言特征多个分析维度的降维方法。例如，因子分析以及与之相关的多特征多维度分析、聚类分析、层级配置频次分析（HCFA，hierarchical configural frequency analysis）等。

当然，还有一些源于自然语言处理领域的统计方法，比如文本分类（text classification）领域所适用的文本特征相似度的统计方法，如词频*倒排文本频率法（TF*IDF，term frequency * inverse document frequency）、向量空间模型（VSM，vector space model）等。

如前所述，语料库研究中所用统计方法基本都是频数统计。然而，常常也可见到语料库方法应用于二语习得等方面的研究案例。这些情况下，T检验、方差分析、回归分析等也同样适用语料库研究（参阅Oakes 1998；Gries 2009）。

需要强调的是，统计分析不是语料库研究的全部。概率统计是针对大规模语料进行语言描写的手段。对于统计得到的规律性现象，需要从前文所谈的各个意义层级或借助其他领域的理论加以阐释，才是完整的语料库研究。

5. 语料库研究如何做到与相关学科的借鉴融通？

不难发现，如今语料库研究越来越以跨学科面貌示人。除了*Corpora*、*Corpus Linguistics and Linguistic Theory*、*ICAME Journal*、*International Journal of Corpus Linguistics*、*International Journal of Learner Corpus Research*、《语料库语言学》等语料库专业期刊外，现在很多（应用）语言学期刊都登载语料库研究论文。语料库研究因其关注语言事实，采用量化分析方法，那些认同“用法观”（usage-based approach）和实证方法的领域（比如，社会语言学、功能语言学、话语分析、语用学、二语习得），借鉴语料库方法，顺理成章。对于其他一些领域（比如，认知语言学、心理语言学），其研究基本取向与语料库研究相异，但近年来也常见有利用语料库数据和方法开展研究的。就认知语言学而言，其中一部分学

者坚持用法观，认为语言的概念表征有体验基础，是人们长期接触语言输入形成并固化（entrenchment）而来的。某种意义上，语料库方法的引入给认知语言学带来了新的研究方法和思路（参阅Tomasello 2005；Gries & Stefanowitsch 2006；Stefanowitsch & Gries 2006；Glynn & Fischer 2010；Taylor 2012）。而在语言习得领域，不少学者（以Nick Ellis等为代表），兼容并包，综合考察语言习得中的心理表征和输入（产出）数据，试图了解心理语言学数据与语料库语言学数据能否互证。这也是对二语习得领域长期以来有关认知观和社会文化观之争的积极回应。语言习得过程本应是社会和认知两相互动的结果。

另外，语料库研究与自然语言处理和计算语言学之间的对接是天然的。目前语料库研究中很多技术手段，本来就源于自然语言处理领域的成果，例如词性标注和自动句法分析。如果能将命名实体识别、知识本体、情感分析等方面研究也充分引入语料库研究，将会大大推进语料库研究的发展。如今大数据十分热门，试想如果基于互联网用户的地理信息、个人身份数据（性别、职业、年龄、收入水平等）、网络行为数据，结合到微观语言特征的分析之中，这在很大程度上可以将高层级的意义建构也纳入到语料库研究的框架中来。

在国际上，已有学者将语料库研究运用于历史、政治、心理学、社会学、医学、传播学等社会科学各个领域。这方面国内学者显得学科壁垒意识过强，缺少真正的跨学科思维。殊不知，跨学科结合处正是新知识的萌发处。不论哪个学科，但凡研究中涉及文本或话语，都有构建语料库，开展语料库研究的潜力。而量化分析、文本意义及社会文化的解读，可以为解决其他学科的一些研究问题另辟蹊径。这与学术界所谓“话语转向”也有内在关联性。

最后，尺短寸长，语料库研究者需要认识到语料库显然不是无所不能的。首先，目前看语料库比较擅长在词汇和短语层面进行语言描写，如果将整个语言描写和语言学理论建构仅限于此，将无法得到语言的全貌。另外，语言使用中不出现（如省略、回避使用），或缺少明确的形式标记（如隐喻、情感态度、言语行为）的语言现象，并非语料库所长。再者，语料库语言学适合分析重复出现的语言现象。然而，按照语料库频率统计，出现一两次的语言现象其总和在语料库中的比重常常达50%左右。这些低频的语言现象是语言中“沉默的大多数”。语料库研究所着眼的是高频语言现象，适合揭示概率上具有规律性的常见语言特征，可以将不擅长的工作交由其他学科去完成，或者与其他相关学科协同完成。

注释

1. Sinclair这一说法是The truth, the whole truth, and nothing but the truth的戏仿。这句话具有法律和宗教双重含义。在英美等国，证人法庭作证前，需左手置于《圣经》之上，举起右手，跟着法官宣誓I do solemnly swear that I will tell the truth,

the whole truth, and nothing but the truth, so help you God（美国法庭宣誓词）。Truth是多义词，至少有事实真相和真理两层含义。Sinclair戏仿这句话时，似乎想表达短语在语言描写中具有独一无二以及至高无上的地位。

2. 本文涉及discourse这一术语的中文表述时，使用“语篇”时更强调其作为结构单位的性质，使用“话语”则更突出其社会文化属性。

参考文献

Ackermann, K. & Y. Chen. 2013. Developing the Academic Collocation List (ACL) – A corpus-driven and expert-judged approach [J]. *Journal of English for Academic Purposes* 12(4): 235-247.

Biber, D. 1988. *Variation across Speech and Writing* [M]. Cambridge: CUP.

Bolinger, D. 1980. *Language – The Loaded Weapon: The Use and Abuse of Language Today* [M]. London: Longman.

Bybee, J. 2005. The impact of use on representation: Grammar is usage and usage is grammar [R]. The presidential address presented at the annual meeting of the Linguistic Society of America in Oakland, California on January 8, 2005.

Bybee, J. 2006. From usage to grammar: The mind’s response to repetition [J]. *Language* 82(4): 711-733.

Firth, J. 1951/1957. Modes of meaning [A]. In J. Firth (ed.). 1957. *Papers in Linguistics* [C]. London: OUP. 190-215.

Glynn, D. & K. Fischer (eds.). 2010. *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches* [C]. Berlin: Walter de Gruyter.

Gries, S. & A. Stefanowitsch (eds.). 2006. *Corpora in Cognitive Linguistics* [C]. Berlin: Mouton de Gruyter.

Gries, S. 2009. *Quantitative Corpus Linguistics with R: A Practical Introduction* [M]. New York: Routledge.

Halliday, M. 1961. Categories of the theory of grammar [J]. *Word* 17(3): 241-292.

Halliday, M. 1966. Lexis as a linguistic level [A]. In C. Bazell, J. Catford, M. Halliday & R. Robins (eds.). *In Memory of J. R. Firth* [C]. London: Longmans, Green and Co., Ltd. 148-162.

Halliday, M. 1991. Corpus studies and probabilistic grammar [A]. In K. Aijmer & B. Altenberg (eds.). *English Corpus Linguistics* [C]. London: Longman. 30-43.

Hoey, M. 2005. *Lexical Priming* [M]. London: Routledge.

Hunston, S. 2002. *Corpora in Applied Linguistics* [M]. Cambridge: CUP.

Hunston, S. & G. Francis. 2000. *Pattern Grammar: A Corpus-driven Approach to the Lexical Grammar of English* [M]. Amsterdam: John Benjamins.

Lewis, M. 1993. *The Lexical Approach: The State of ELT and a Way Forward* [M]. London: Language Teaching Publications.

McEnery, T. & A. Hardie. 2012. *Corpus Linguistics: Method, Theory and Practice* [M].

Cambridge: CUP.

Meyer, C. & H. Tao. 2005. Response to Newmeyer's 'Grammar is grammar and usage is usage' [J]. *Language* 81(1): 226-228.

Newmeyer, F. 2003. Grammar is grammar and usage is usage [J]. *Language* 79(4): 682-707.

Oakes, M. 1998. *Statistics for Corpus Linguistics* [M]. Edinburgh: Edinburgh University Press.

Simpson-Vlach, R. & N. Ellis. 2010. An academic formulas list: New methods in phraseology research [J]. *Applied Linguistics* 31(4): 487-512.

Sinclair, J. 1966. Beginning the study of lexis [A]. In C. Bazell, J. Catford, M. Halliday & R. Robins (eds.). *In Memory of J. R. Firth* [C]. London: Longmans, Green and Co., Ltd. 410-430.

Sinclair, J. (ed.). 1987a. *Looking Up: An Account of the COBUILD Project in Lexical Computing* [C]. London: Collins ELT.

Sinclair, J. 1987b. *The Collins COBUILD English Language Dictionary* [M]. London: Collins ELT.

Sinclair, J. 1991. *Corpus, Concordance, Collocation* [M]. Oxford: OUP.

Sinclair, J. 2004. *Trust the Text: Language Corpus and Discourse* [M]. London: Routledge.

Sinclair, J. 2008. The phrase, the whole phrase and nothing but the phrase [A]. In S. Granger & F. Meunier (eds.). *Phraseology: An Interdisciplinary Perspective* [C]. Amsterdam: John Benjamins. 407-410.

Sinclair, J. & M. Coulthard. 1975. *Towards an Analysis of Discourse: The English Used by Teachers and Pupils* [M]. London: OUP.

Sinclair, J., S. Jones, R. Daley & R. Krishnamurthy. 2004. *English Collocation Studies: The OSTI Report* [M]. London: Contiuum.

Sinclair, J. & A. Mauranen. 2006. *Linear Unit Grammar: Integrating Speech and Writing* [M]. Amsterdam: John Benjamins.

Stefanowitsch, A. & S. Gries (eds.). 2006. *Corpus-Based Approaches to Metaphor and Metonymy* [C]. Berlin: Mouton de Gruyter.

Stubbs, M. 1993. British traditions in text analysis: From Firth to Sinclair [A]. In M. Baker, G. Francis & E. Tognini-Bonelli (eds.). *Text and Technology: In Honour of John Sinclair* [C]. Amsterdam: John Benjamins. 1-33.

Taylor, J. 2012. *The Mental Corpus: How Language is Represented in the Mind* [M]. Oxford: OUP.

Tomasello, M. 2005. *Constructing a Language: A Usage-Based Theory of Language Acquisition* [M]. Harvard: Harvard University Press.

Willis, D. 1990. *The Lexical Syllabus* [M]. London: Collins ELT.

Wittgenstein, L. 2009. *Philosophical Investigations (Revised 4th Edition)* [M]. Chichester, West Sussex: Wiley-Blackwell.

肖忠华、许家金，2008，语料库与语言教育 [J],《中国外语教育》(2)：50-60。

许家金，2009，词汇中心教学法的交际观：理论溯源与反思 [J],《中国外语教育》(4)：38-45。

许家金，2011，语料库语言学研究——计算机时代的“田野调查”[N]，《中国社会科学报》，7月5日第14版。

许家金、吴良平，2014，基于网络的第四代语料库分析工具CQPweb及应用实例 [J]，《外语电化教学》（5）：10-15，56。

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。

通信地址：100089 北京市北京外国语大学中国外语教育研究中心