“我来读文献”【线上论坛】要点整理

2016 年 2 月 23 日下午 4:00-6:00，第 22 期“我来读文献”活动【线上论坛】成功举行，论坛嘉宾河南师范大学的刘国兵老师、对外经济贸易大学的江进林老师与参加活动的研友们通过 QQ 群讨论组实时交流。我们整理了研友们提出的主要问题以及嘉宾老师的解答，供大家学习参考。

本期活动文献：

江进林、许家金，2015，基于语料库的商务英语语域特征多维分析[J]，《外语教学与研究》，47（02）：225-236。

本期活动页面：

http://iresearch.unipus.cn/info/paperreview.php

问题 1. 为什么选择这篇作为本期的研读文献？

刘国兵老师:

我大概花了半个月的时间浏览去年发表的文章，语料库方面的太多了。所以我就定了这样几个原则：1. 与语料库语言学相关；2. 使用了较新的研究方法；3. 读完论文，对大家在选题方面有所启发，最后就定了这篇。

问题 2. 为什么当初要做这个选题呢？

江进林老师:

我做这个选题的初衷是想研究商务英语的特点。因为我所在的对外经济贸易大学以商务英语为特色，我的出发点是如何把学校的特色和自己的研究方向结合起来，于是想到用语料库的方法研究商务英语。查阅了相关文献后发现，这方面的研究很少，于是做了商务英语和通用英语的比较研究。

问题 3. 文献中语体是如何界定的，语体的定义在某种程度上是否和 genre 有交集？

刘国兵老师:

关于英文文献中的 style、genre、register 以及中文文献中的“语体”、“体裁”、“语域”、“文体”、“语类”等，江老师在第二页加了注释：近义术语，有人严格区分，有人混而用之。为避免术语混淆，本文不做区分，统一采用“语域”这一术语。多维分析法的主要研究目标是为了探究“语域变异”（register variation）。这也是本文采用“语域”这一术语的另一重要原因。

问题 4. 能否对文献中提到的多维分析法作进一步说明？

刘国兵老师：

虽然 Biber 提出多维分析方法已有几十年，但基于该方法所进行的研究还不多。国际上不太多，国内更少。其实 1984 年，Biber 在他的博士论文中就提出来了多维分析的方法，到 1988

年把博士论文做了修订出版，对该方法做了系统阐释。Biber 提出的方法涉及到很多，大概六十多个语言特征，方法也较为复杂。但他开发的程序，我们一般没法直接获得。所以这篇文章中提到的 MAT(Nini，2014)是一个非常好的工具，解决了我们的燃眉之急。在这之前我请教了许家金老师，我们之间发了几次邮件专门讨论这个问题。许老师回复的内容，转述如下："MAT（Multidimensional Analysis Tagger）这个软件，我上传到北外语料库团队网站了。大家可以从这里下载 http://www.bfsu-corpus.org/channels/tools，不用翻墙到 Google site 上去找。使用这个软件前，需要事先安装 Java Runtime Environment，有兴趣的老师可以试试。"

问题 5. 文中所用商务英语语料库是否可以免费使用？

刘国兵老师:

我们可以在北外语料库平台在线使用：http://111.200.194.212/cqp/，账号密码都是 test，有关语料库平台的使用说明可参考：

http://www.bfsu-corpus.org/static/corpus_tools/CQPweb_guide.pdf

国内外很多语料库在这个平台上都有链接，大家需要时不妨登录试试。

江进林老师补充：

对外经济贸易大学的大型商务英语语料库尚未经过系统的清理，基本上是生文本，暂时不能购买。大家可以用本文使用的商务英语语料库，网址刚才刘老师已公布。

问题 6. 多维分析一定需要一个对比语料库吗？是否有 MAT 软件培训呢？

江进林老师:

关于多维分析法，以往研究通常会对两个语料库进行比较，目的是发现其不同之处。据我所知，没有对 MAT 软件的培训，不过该软件的操作比较容易掌握。打开 MAT 软件后，仅有三个窗口，点击一下均会跳出对话框，选择读入语料即可。结果会出现在原语料所在的文件夹中。

问题 7. 假如我基于 Biber 的该方法，假设自己设定的语言特征，通过因子分析，分析自己手上语料的维度，这样做是否有科学性？维度分数应该怎样计算呢？

江进林老师:

这位老师的想法不错，理论上说是完全可以的。不过设定的语言特征最好有一定的理论依据，否则大家可能会质疑：为什么抛弃 Biber 已有的框架，而自己设定语言特征。只要这个问题能回答清楚就没有问题。维度分数的计算请参考目前我们阅读的这篇文章，文献综述部分有步骤；更详细的步骤在 Biber 的书里面有介绍。不过要用自己的语言特征还原整个过程，会需要投入较多的时间。

刘国兵老师补充：

我感觉 Biber 想出来那么多特征，当时做博士论文时一定是绞尽脑汁的。所以，如果完全抛弃他的框架另立炉灶，会比较麻烦。另外，就像江老师提到的，需要一定的理论依据，就是你为什么不用 Biber 的，而自己来设定语言特征。你设定的与 Biber 相比有哪些优越性，这都是要考虑的。

问题 8. Biber 在他多篇文献中对于维度分数的对比使用的是 ANOVA（Analysis of Variance，即方差分析）的统计方法，江老师这个文章里使用的是独立样本 T 检验，我想问下这两种统计方法有区别吗，有优劣之分吗？三组之间进行比较，能否用 T 检验呢？

江进林老师:

原则上说，ANOVA 是对三组及以上的数据进行比较，T 检验是对两组数据进行比较。本文比较的对象都是两组数据，所以使用的是独立样本 T 检验。如果 Biber 使用了 ANOVA，比较的应该是三组数据。这两种统计方法没有优劣之分。

刘国兵老师:

独立样本 T 检验与成对样本 T 检验都是对两个样本的均值进行差异检验。如果要对三个或三个以上样本的均值进行差异检验，就要使用方差分析的方法，三组样本或者更多样本的容量可以相等或不等。如果三组之间进行比较，就是用 T 检验比较完两组，再比较另外两组。其实只要不嫌麻烦，也可以做。唯一的不同时，每次比较之后，准确性就会降低一些。检验的准确性会随着检验次数的增多而降低。因此，在对三组及以上的数据进行比较时，通常我们都选用方差分析的方法。

问题 9. 如果将多维分析法运用于商务英语教学，具体该怎么入手？

江进林老师:

我觉得，多维分析法直接放到课堂上教给学生不太现实，主要还是由教师做一定的研究，把研究结果对教学有启示的地方放到课堂中去。并且，由于多维分析的结果比较抽象，教师需要归纳、总结，最好结合具体的语料（如例句）放到商务英语教学中去，强化商务英语的语言特征，加深学生的印象。

问题 10. 本人对本文献的研究方法部分涉及到软件的使用和数据分析方法并不熟悉，因此感觉数据处理部分有些抽象，理解不透。最近在读到其他文献的时候，也总是遇到类似情况，作为一般研究者，该如何克服这种技术障碍？请问老师有没有什么好的建议？

刘国兵老师:

多维分析对于我们研究者而言，对其数据的解读还需要做一定的功课。因此，对于学生而言，看到之后可能就给吓倒了。我们外语老师绝大多数属于文科背景，所以对软件的操作与使用不是很擅长。因此在看带有数据处理及分析的论文时，总感觉作者对该部分解释的有些不太清楚。

另外，在学术论文中，研究者通常不报告软件的具体操作方法与数据的具体分析方法。比如在该篇论文中，对于 MAT 的操作步骤，还有使用到的独立样本 T 检验等，研究者均没有做详细说明。一是因为篇幅有限，研究者不可能把与此相关的所有知识回顾一遍；二是也没有必要，因为要了解此类信息，很多书上都有详细介绍。对我们研究者来说，碰到不熟悉的情况，如果想进一步了解，我们只能顺着这条线走下去，查阅相关资料，学习补充这方面的知识。其实目前外语研究方面使用的软件，多数还是很友好的，很多软件打开后，操作界面都非常简单，可直接上手尝试，或者看看软件附带的 PDF 文档自主学习。

至于数据分析与统计方法，对于应用语言学特别是语料库语言学方向的老师，应该了解一些基本知识。这方面的书很多，仅咱们外研社就出了好几本。

（此处，小编前来硬广一下，数据分析与统计方法相关书目包括：《第二语言研究中的数据收集方法》、《第二语言研究中的统计案例分析》、《外语教学定量研究及数据分析》（第二版）等。）

问题 11. 多维度分析或者说这种基于语料库的研究，如果向上追溯的话，是否有理论基础？如果有，可能的理论基础有哪些？

江进林老师:
多维分析的理论基础主要是语言学或二语习得中的相关研究、Biber 及其他人的大量研究。比如 Biber 在他的著作中分析过 67 个语言特征，为什么要提取这些特征，基本上是基于以往的语言学研究或二语习得的研究（如学生口语和笔语的差异性特征）。在我看来，如果使用 Biber 的框架，没有必要向上追溯那么远，因为 Biber 已经分析过了。

问题 12. 对比语料库是本族语者的语料库比较好吗？

江进林老师:
对比语料库不一定非要用本族语者的语料库，取决于研究目的。比如，可以比较《红楼梦》译本和《水浒传》译本，都不是本族语者的语料库。

问题 13. 基于语料库的研究最后得到的数据结果有时候很难找到相应的现成理论解释，该如何处理这个问题呢？因为结果都是语料得出的，是归纳式的思维；而不是基于理论演绎式的思维。所以，最终如果得出的数据结果找不到现成的理论该如何是好？

刘国兵老师:
这个问题说来话长，可能要提到语料库语言学领域中两种研究范式：语料库驱动范式与基于语料库的范式。语料库驱动研究主张，在研究语言时，抛弃一切现成的语言学理论。检索之后观察语料，基于所得到的语料归纳理论。它是一种自下而上的研究方法。因此，如果基于“语料库驱动范式”研究，最后得到的数据结果找不到相应的现成理论去解释，不是有时候找不到，可能压根就没有。需要研究者根据语料，自己去总结和发现。但是基于语料库的研究范式就不同了，研究者头脑中可能已有一些理论。使用语料的目的，就是来验证已有理论。所以两种研究方法是不一样的。

梁茂成、李文中、卫乃兴等老师都发过这方面的文章，专门谈论语料库研究范式，有空大家可以下载看看。另外有一个梁茂成老师的视频，把这个问题讲得非常清楚，欢迎大家有空看看。网址为
http://teacher.unipus.cn/workshop/videoshow.php?VideoHash=5b01e9c6f6b55529d1a81a16e33c3fdf。

所以“得出的数据结果找不到现成的理论”这一情况出现的话，仔细观察数据，努力总结就好了。说不定，在理论上会有所突破的。

另外，我想补充一点：国内语料库方面的研究或者说基于语料库所做的其他领域的研究越来越多，但需要我们注意的一点是，千万不能罗列数据。语料库语言学研究，说到底是语言学研究。基于语料库得出数据，不是研究的根本目的，而是有了这些数据，才可以有效帮助我们洞察之前观察不到的语言现象。所以，基于语料库所做的研究，得出数据之后，最精彩的应该是对数据进行的深度解释。

问题 14. 如果把这篇文献里的方法用于分析对比同一原著不同译本的语言风格，是否可行？这方面不知道老师有什么建议？

刘国兵老师:

把多维分析方法应用于比较不同译本的语言风格，这方面的文献我见到的还不多。如果有老师感兴趣，真的不妨一试。但多维分析目前为止，通常分析的是不同语体、体裁的文章，因为他们之间的语言特征差别较大。但对于同一原著，不同译本之间的语言差别，利用 MAT 软件能够提取出来多少，我没试过，所以不是很确定。但这个想法真的挺好，做出来结果看看，也许会有新的发现。