

# ArtDEAP 艺术学术英语语料库的创建

西安广播电视大学 韩喜春 晏远方

**摘要：**ArtDEAP 艺术学术英语语料库是北京外国语大学 DEAP 学术英语语料库的一个子库，是研究西方艺术学术领域的重要资源。本文主要介绍了艺术学术英语语料库的建设过程，及其在应用语言学、语料库语言学和词典编纂、教材建设方面的应用前景。语料的采集体现了语料库的平衡性和代表性原则，为后续研究者在艺术学术英语这个细分领域提供详实的语料研究基础。艺术和文化作为一个国家最内核的精神凝聚动力，其发展和传播的力量代表一种国家的软实力，所以对艺术学术英语真实语料的进一步开发，在艺术学术英语教学的相关教学资源建设，例如，相关术语词典编纂和艺术学术论文写作等的语料支持方面将具有很高的微观实用价值。

**关键词：**学术英语、语料库、艺术、建库过程

## 1. 引言

语言是国家形象构建和文化软实力输出的重要的战略资源已经成为语言学界的共识。语料库语言学采用接近穷尽的语言使用现象的罗列方式，用不可再分的语言颗粒度进行语用描写，深入描写本身就表明了阐释性，描写的框架一般采用 John Sinclair 的短语学。这种分析和解读使得语料库不仅是一个文本库，同时也是一个知识库和规则库。学习者对语料库特定语境下的信息进行仔细的观察、思考、分析，然后，独立完成对某一个语言现象的解构与认知，这是语言学习的便捷途径。语料库不但反映出语言使用的真实情况，更重要的是它能反映出语言使用的频率。通过专门语料库上检索出来的成千上万个语例反映出一种专用语言的特点和趋势，以此来论证或支持某项理论或某个观点的科学依据。

相对于大规模通用语料库，面向特殊用途领域的专门语料库有着较强的针对性和专业化程度，可以更好地满足不同领域研究者的研究和需求。范谊(1995)指出一篇专业文献最大的信息载体是专业词汇，而不是普通词汇。这些都使得专门语料库在语言研究和语言应用领域成为无可替代的语言聚宝盆。艺术语料库属于专门用途语料库的一支，该类语料库对于艺术语言翻译、艺术语言的理解和分析、艺术英语教材编写、词典编撰等领域有着十分重要的作用。国内艺术领域的学术英语语料库建设非常薄弱。以艺术语料库为主题词和篇关摘来搜索硕博士论文库，发现只有 20 篇相关论文。而且，关于艺术语料库的研究 90%集中在文学译本的话语研究。不难理解，语料库在翻译研究中的应用本身是语料库语言学的一个大的

研究方向（许家金 2017；甄凤超、张霞 2014；梁茂成 2012）。虽然也有社交平台文本和相声台本语言的研究，但是研究数量极少，更不要说关于纯艺术、舞蹈、音乐、摄影、工艺美术等方面的研究了。

基于以上考虑，本研究在“中国外语教育基金专用英语语料库建设项目（DEAP）”的统一指导下建设艺术学术英语语料库。语料库的建设为后续下游研究提供详实的研究资源。其次，艺术学术英语语料库是亿词级专用英语 ESP 语料库的一个子库，在建库标准和建库流程与其他领域，如医学、法学、政治学、语言学等子语料库一致，这就为不同领域子库间的对比研究提供了语料基础和平台。对艺术英语语料库建库标准和流程的清晰说明，是规范艺术英语语料库、免于语料重复建设而造成资源浪费的必要手段。有了专用语料库建库标准和流程，不同领域的语料对比性研究便于开展。本研究主要介绍了艺术学术英语语料库的建库目标、建库流程和应用前景。

## 2. 建库目标

艺术学术英语语料库搜集国际艺术学术期刊公开发表的论文，建设规模约为 500 万词。公开发表的艺术学术论文可以通过数据库进行下载；单篇学术论文的正文字数大约为 7 千词，500 万词规模的语料库大约需要 700 篇。本研究最终建设的艺术学术英语语料库包含 39 种国际艺术学术期刊论文 708 篇，总词数为 5624890，标准形符类符比为 44.29。

## 3. 建库方案

### 3.1 学科领域

按照教育部人文社科网站的人才培养的方向目录给出的艺术学分类，共有 10 个二级学科，包括艺术美学、艺术心理学、音乐、戏剧、戏曲、舞蹈、电影、广播电视文艺、美术、工艺美术、摄影等艺术领域。

### 3.2 来源期刊

ArtDEAP 艺术学术英语语料库的语料来源为发表在国际艺术学术期刊上的论文。通过 EBSCO 国际数据库获取。研究者在 Web of Science (WOS) 搜索没有发现艺术学学科分类的期刊资源，因为它是关于自然科学研究成果的数据库。通过国内博士论文的数据库里参考文献交叉验证的办法，确定了理论性和学术性都比较强的杂志，其中有几本是交叉学科的杂志，例如，为二级学科艺术心理学门类收集的文章就使用了心理学杂志“plos one”。最后，确定了每个学科下 2-6 种期刊，10 个二级学科共 39 种期刊（见表 1）。

表 1 ArtDEAP 语料来源期刊

| 序号 | 二级学科                 | 期刊名称  |
|----|----------------------|---|
| 1  | Art Aesthetic 艺术审美   | 1 Journal of Aesthetics & Art Criticism<br>2 Metaphor and symbol<br>3 Journal of Visual Art Practice<br>4 Israel Affairs  |
| 2  | Art psychology 艺术心理学 | 1 plos one<br>2 National Identities   |
| 3  | Music 音乐             | 1 Journal of research in music education<br>2 Ethnomusicology Forum<br>3 Instrumentation Science & Technology<br>4 Musicology and Cultural Science  |
| 4  | Drama 戏剧             | 1 Comparative Drama<br>2 Research in African Literatures<br>3 Philosophy of Science<br>4 Studies in Philology<br>5 Journal of Modern Jewish Studies<br>6 International Studies in the Philosophy of Science                               |
| 5  | Costume Opera 戏曲     | 1 Journal of the Royal Musical Association<br>2 Hispanic Research Journal<br>3 Theatre History Studies  |
| 6  | Dance 舞蹈             | 1 arts education policy review<br>2 Rethinking History<br>3 Central European History<br>4 Teaching Artist Journal   |
| 7  | Film 电影电视            | 1 Journal of Cinema and Media Studies<br>2 Cinema Journal<br>3 Journal of Popular Film & Television<br>4 International Journal of Cultural Policy<br>5 Studies in French Cinema   |
| 8  | Fine Art 美术          | 1 Art bulletin<br>2 Art history<br>3 Visual Anthropology Review   |
| 9  | Crafts Art 工艺美术      | 1 international journal of arts and design education<br>2 Journal of Arts Management, Law & Society<br>3 Fashion Theory: The Journal of Dress, Body & Culture<br>4 Computer Graphics Forum<br>5 Civil Engineering & Environmental Systems |
| 10 | Photography 摄影       | 1 Photogrammetric Record  |

## 2 Visual Studies

### 3 Third Text

### 3.3 文献类型

ArtDEAP 艺术学术英语语料库的语料共选取了 39 种学术期刊上的 3 种文本类型，它们是：研究论文、文学评论和书籍评论（只有 1 篇电影评论）。研究论文的篇数占 90.6%，文学评论占 6.4%，书籍评论占 3%。

表 2 ArtDEAP 语料库构成

| 文本类型          | 文本数量 | 库容      |
|---------------|------|---------|
| 研究论文 RA       | 642  | 5196428 |
| 文学评论 CR       | 45   | 383132  |
| 书籍/电影评论 BR/FR | 21   | 45330   |
| 合计            | 708  | 5624890 |

### 3.4. 元信息

按照期刊的收录时间为最近 3 年的原则，最终确定了 39 种国际艺术学期刊作为语料搜集来源（见表 1）。元信息是关于语料库的全局信息，包括语料库名称、版本、建库单位等，同一语料库内每篇语料的元信息都是一致的。包括文章 ID、文本类型、标题、期刊名、作者、出版年、卷号和 DOI 等。为了快捷保存元信息，一篇文章保存成网页版和 PDF 版同样重要，因为一些更详细的元信息可以到网页版找到。例如，字数、数据库、DOI 等。先把文章的 PDF 版导入到 Endnote，确保应有的元信息都填写完成，再导出为 TXT，然后保存到 Excel，比较方便。Endnote 还可以查重，防止一篇文章保存两边遍，加快了整理节奏。

## 4. 文本收集

表 3 艺术学期刊收集列表

| 序号 | 篇数 | 二级学科  | 二级学科英文 (缩写)        | 学科范围   | 期刊  |
|----|----|-------|--------------------|--------|---|
| 1  | 22 | 艺术美学  | art aesthetic ARA  | 音乐美学、  | 1 Journal of Aesthetics & Art Criticism 0101-0122 |
| 2  | 26 |       |                    | 影视美学   | 2 Metaphor and symbol 0201-0226                   |
| 3  | 42 |       |                    | 建筑美学   | 3 Journal of Visual Art Practice 0301-0342        |
| 4  | 7  |       |                    |        | 4 Israel Affairs 0401-0407                        |
| 5  | 11 | 艺术心理学 | Art psychology ARP | 绘画、书法、 | 1 plos one 0101-0111                              |

|    |          |              |                   |                             |  |
|----|----------|--------------|-------------------|-----------------------------|--|
| 6  | 38       |              |                   | 音乐心理学                       | 2 National Identities 0201-0238                                  |
| 7  | 35       | 音乐           | music MUS         | 音乐史<br>作曲与作曲理论、表演艺术         | 1 Journal of research in music education 0102-0135               |
| 8  | 14       |              |                   |                             | 2 Ethnomusicology Forum 0201-0214                                |
| 9  | 13       |              |                   |                             | 3 Instrumentation Science & Technology 0302-0313                 |
| 10 | 12       |              |                   |                             | 4 Musicology and Cultural Science 0401-0412                      |
| 11 | 12       | 戏剧           | drama DRA         | 戏剧史、理论                      | 1 Comparative Drama 0101-0112                                    |
| 12 | 20       |              |                   |                             | 2 Research in African Literatures 0201-0220                      |
| 13 | 9        |              |                   |                             | 3 Philosophy of Science 0301-0309                                |
| 14 | 7        |              |                   |                             | 4 Studies in Philology 0401-0407                                 |
| 15 | 21       |              |                   |                             | 5 Journal of Modern Jewish Studies 0501-0521                     |
| 16 | 3        |              |                   |                             | 6 International Studies in the Philosophy of Science 0601-0603   |
| 17 | 4        | 戏曲           | costume opera CSO | 戏曲史<br>戏曲理论<br>戏曲表演         | 1 Journal of the Royal Musical Association 0101-0104             |
| 18 | 5        |              |                   |                             | 2 Hispanic Research Journal 0201-0205                            |
| 19 | 17       |              |                   |                             | 3 Theatre History Studies 0301-0317                              |
| 20 | 10       | 舞蹈           | dance DAN         | 舞蹈史<br>舞蹈理论<br>舞蹈编导、舞蹈表演    | 1 arts education policy review 0101-0110                         |
| 21 | 14       |              |                   |                             | 2 Rethinking History 0201-0214                                   |
| 22 | 3        |              |                   |                             | 3 Central European History 0301-0303                             |
| 23 | 3        |              |                   |                             | 4 Teaching Artist Journal 0401-0403                              |
| 24 | 23       | 电影<br>广播电视文艺 | film FIL          | 电影史<br>电影理论<br>电影艺术、广播电视文艺  | 1 Journal of Cinema and Media Studies 0101-0123                  |
| 25 | 32       |              |                   |                             | 2 Cinema Journal 0201-0232                                       |
| 26 | 10       |              |                   |                             | 3 Journal of Popular Film & Television 0301-0310                 |
| 27 | 13       |              |                   |                             | 4 International Journal of Cultural Policy 0401-0413             |
| 28 | 16       |              |                   |                             | 5 Studies in French Cinema 0501-0516                             |
| 29 | 60       | 美术           | fine art ART      | 美术史、理论<br>绘画艺术、雕塑艺术         | 1 Art bulletin 0101-0160   |
| 30 | 26       |              |                   |                             | 2 Art history 0201-0226  |
| 31 | 11       |              |                   |                             | 3 Visual Anthropology Review 0301-0311                           |
| 32 | 29       | 工艺美术         | crafts art CRA    | 工艺(服装)美术史<br>工艺美术理论<br>环境艺术 | 1 international journal of arts and design education 0101-0129   |
| 33 | 26       |              |                   |                             | 2 Journal of Arts Management, Law & Society 0201-0226            |
| 34 | 45       |              |                   |                             | 3 Fashion Theory: The Journal of Dress, Body & Culture 0301-0345 |
| 35 | 36       |              |                   |                             | 4 Computer Graphics Forum 0401-0436                              |
| 36 | 3        |              |                   |                             | 5 Civil Engineering & Environmental Systems 0501-0503            |
| 37 | 11       | 摄影           | photography PHO   | 摄影史、理论                      | 1 Photogrammetric Record 0101-0111                               |
| 38 | 6        |              |                   |                             | 2 Visual Studies 0201-0206                                       |
| 39 | 13       |              |                   |                             | 3 Third Text 0301-0313   |
|    | 708 (合计) |              |                   |                             |  |

这 39 种艺术学期刊覆盖了艺术美学、艺术心理学、音乐、戏剧、戏曲、舞蹈、电影、广播电视文艺、美术、工艺美术、摄影等艺术领域，可以较好的代表

艺术学术英语的文本特征。期刊来自三个数据库 Academic Search Complete、Taylor & Francis Group、Wiley-Blackwell，便于搜集数据。从表 1 中我们可以发现，艺术学的学术期刊分散在其他各个学科中。如艺术心理学的期刊来自心理学和认知研究的期刊；艺术美学来自文学批评；舞蹈来自对历史的思考。因此艺术学的期刊兼具艺术学和其他社会学科的交叉融合，艺术学术英语语料库的建设对于寻找和发现艺术学术英语的特点具有重大现实意义。

## 5. 文本命名

ArtDEAP艺术学术英语语料库对所选出的708篇文献采用“二级学科代码+杂志序号+文章序号+文献类型+题目”的方式命名。文献类型有三种，一种是研究论文，用RA表示，第二种是文学评论CR，最后一种是BR。其中，二级学科按照教育部人文社科网站的人才培养的方向目录给出的艺术学分类，共11个二级学科，书法学科因为没有合适的学术文章，故未列出），显示为10个二级学科。例如电影这个二级学科（Film，编码为Fil）中的第一种杂志Journal of Cinema and Media Studies里下载的第一个文本The Melancholic Voice-Over in Film Noir，其存放目录为：艺术语料库FIL0101RA-The Melancholic Voice-Over in Film Noir。

表 4 艺术各二级学科名称、代码及文本编号

| 序号 | 二级学科  | 代码  | 文本编号   |
|----|-------|-----|--|
| 1  | 艺术美学  | ARA | 0101-0122<br>0201-0226<br>0301-0342  |
| 2  | 艺术心理学 | ARP | 0101-0111<br>0201-0238   |
| 3  | 音乐    | MUS | 0102-0135<br>0201-0214<br>0302-0313<br>0401-0412                           |
| 4  | 戏剧    | DRA | 0101-0112<br>0201-0220<br>0301-0309<br>0401-0407<br>0501-0521<br>0601-0603 |
| 5  | 戏曲    | CSO | 0101-0104<br>0201-0205<br>0301-0317  |
| 6  | 舞蹈    | DAN | 0101-0110  |

|    |       |     |   |
|----|-------|-----|---|
|    |       |     | 0201-0214<br>0301-0303<br>0401-0403                           |
| 7  | 电影、电视 | FIL | 0101-0123<br>0201-0232<br>0301-0310<br>0401-0413<br>0501-0516 |
| 8  | 美术    | ART | 0101-0160<br>0201-0226<br>0301-0311                           |
| 9  | 工艺美术  | CRA | 0101-0129<br>0201-0226<br>0301-0345<br>0401-0436<br>0501-0503 |
| 10 | 摄影    | PHO | 0101-0111<br>0201-0206<br>0301-0313                           |

## 6. 文本清理

研究者进行文本的清理工作量很小，因为一开始保存下载的论文时就保存了两份，一份 PDF，另一份仅网页形式。只有一种杂志没有找到网页版的保存方式，就先保存为 PDF，再用 abbyy 软件进行文本转换，转换成 TXT 形式。需要删除论文封面，页眉页脚，纠正文本中的单词错误和段落连接等格式错误，并删除图画，图画删除后在相应位置插入符号 <P>，用以取代原文中的图画信息。清理工作量大，费时，所以一开始就保存成两份形式的文本很重要。

## 7. 艺术学术英语语料库的应用前景和展望

语料库语言学研究有描述、理解、解释、评价四个办法：去描述语料库里出现的语言模式；去理解这些语言模式在语境中怎样起作用；在历史背景中、社会语境中去解释为什么会出现这些话语模式？去评价谁会受益于这些模式或话语？应该改变什么或怎样改变分析者的评价。例如，一个搭配和固定短语在文章中反复出现，那么就可以断定人们用这个表达来思考问题。还有，我们通过一个词语在两个语料库出现的频率差，可以发现这个词语就是摸一个领域的关键词，它提供语料分析的切入点和指路标。然后再检查索引及其搭配，就可以揭示其重要性。

该语料库的建立有着极其重要的实际应用价值。随着中国国力强盛，体现国家软实力的文化意识形态需要从边缘地带走向舞台中央，大量的文化艺术人才走出去，宣扬中国优秀的传统文化艺术的时刻即将到来。该语料库可以为各种艺术教材编写提供重要的语言素材，增强艺术英语教学针对性和效果；由于该语料库文献覆盖艺术学不同二级学科子领域，可以满足艺术各个领域术语词典编纂者的不同需求；该语料库收录的均为艺术类的学术文章，对语料库进行深度挖掘标注还可以可以辅助艺术人才进行学术论文的撰写。科研根据教学实际情况去分析学习者的需要，去重新设计课程、制定大纲、选编教材、探索教学方法，这一切的达成也有赖于检索软件的不断进化。Leech 对语料库的发展吃肯定态度，他认为在标注环节的在线协同科教研使得专业圈内感兴趣的特定成员之间彼此磋商，得以确保所标注的语料能为尽可能多的潜在的使用者所使用。语料的语义层和语用层的标注将会成为研究的重点。研究者计划下一步针对艺术学术英语语料库进行语步标注，语步研究主要包括确立研究领域，发现研究点进行研究。语步的标注可以更好地帮助艺术英语领域的写作者进行学术论文的写作范式的训练研究。随着艺术学术英语语料库的不断更新和发展，未来还可能借助自然语言处理技术发挥更大的作用。

## 8. 结语

艺术学术英语语料库建设有利于完善艺术学术英语学科建设，并促进艺术学院创建一流学科的英语教学改革。语料库为当代大学英语课程教改提供了新的资源平台，为专门英语教学提供有益的借鉴。数据库中海量的来自母语人的地道学术数据资源为艺术专门英语教学提供了丰富的语言知识平台。为艺术人才走向国际舞台提供了资源保障。可以据此数据库深入挖掘，开发艺术学术英语的写作教材、艺术学术英语词典、艺术学术英语教学资源库，包括艺术美学、艺术心理学、音乐、戏剧、戏曲、舞蹈、电影、广播电视文艺、美术、工艺美术、摄影等艺术领域的常见语境下的常用语句。这些领域真实性、多样性的高频词汇及搭配信息，为艺术人才提供第一手的学习资源。也为艺术英语师资的提高提供了物质保障。

语料库的建设是一个动态累计的过程，罗马不是一天建成的。随着时间的推移不断扩充和更新，在这一过程中要保证语料库的代表性和平衡性，以及延续性。随着语料库研究的发展，研究手段的深入，语料不断丰富，其作为一种语言资源



将更加弥足珍贵。

### 鸣谢

1. ArtDEAP语料库容量具体计算均使用许家金教授提供的Antconc中词表的tokens数据得出。
2. 感谢吉林师范大学李亮老师在撰写课题申请时提供的技术理解支持。
3. 我从许家金教授于2019年6月1日召集的科研大咖分享会获得的专业领域知识,至今受益无穷。

### 参考文献

- [1] Bernardini, S. Corpora in the classroom[J]. How to Use Corpora in Language Teaching, 2004, (12): 15-36.
- [2] 方秀才, 2012, 基于语料库的英语教学与研究综述:成就与不足——根据 22 种语言学类 CSSCI 来源期刊近 30 年的统计分析[J], 《外语电化教学》(145): 19-24.
- [3] 桂诗春、杨惠中. 中国学习者英语语料库 [M]. 上海: 上海外语教育出版社, 2003.
- [4] 黄立鹤, 2015, 语料库 4.0:多模态语料库建设及其应用[J], 《解放军外国语学院学报》38 (3): 1-7.
- [5] 李德超、王克非, 2010, 新型双语旅游语料库的研制与应用[J], 《现代外语》33 (1): 46-54.
- [6] 李小龙, 2017 基于语料库对《论语》卫礼贤德译本的跨文化研究[D].北京外国语大学.
- [7] 梁茂成, 2012, 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》44 (3): 323-335.
- [8] 苗艳艳. 2017, 基于语料库的毕飞宇、苏童作品风格比较研究[D].南京师范大学
- [9] 宋红波、王雪利, 2013, 近十年国内语料库语言学研究综述[J], 《山东外语教学》(3): 41-47.
- [10] 许家金, 2014, 许家金谈语料库语言学的本体与方法[J], 《语料库语言学》(2): 35-44.
- [11] 许家金, 2017, 语料库研究学术源流考[J], 《外语教学与研究》49 (1): 51-63.
- [12] 杨昊、鞠朝, 2015, 艺术设计英语语料库设计与探索[J], 《湖北函授大学学报》(11): 137-138.