双语癌症平行语料库构建*

马鹤桐　　任建松

（中国医学科学院/北京协和医学院医学信息研究所/图书馆　北京 100020）　　（国家癌症中心，中国医学科学院/北京协和医学院肿瘤医院　北京 100021）

王序文　方　安　孙月萍　李　姣

（中国医学科学院/北京协和医学院医学信息研究所/图书馆　北京 100020）

〔**摘要**〕　介绍美国国家癌症研究所提供的英文版癌症综合信息库以及国内外有关平行语料库构建相关研究、技术、流程等。通过研究源数据结构，进行数据分析与处理，设计双语语料对齐策略，完成癌症领域中英双语平行语料库构建并就相关设计机制及语料质量进行讨论。

〔**关键词**〕　癌症综合信息库；医学平行语料；癌症；

〔**中图分类号**〕R－056　　〔**文献标识码**〕A　　〔**DOI**〕10.3969/j.issn.1673－6036.2019.09.017

The Building of Bilingual Parallel Corpus of Cancer

MA Hetong, Institute of Medical Information/Medical Library, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100020, China; REN Jiansong, National Cancer Centre, Cancer Hospital Chinese Academy of Medical Sciences, Beijing 100021, China; WANG Xuwen, FANG An, SUN Yueping, LI Jiao, Institute of Medical Information/Medical Library, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100020, China

〔**Abstract**〕　The paper introduces the English version of Physician Data Query (PDQ) provided by U. S. National Cancer Institute and study, technologies, processes, etc. of the building of relevant bilingual parallel corpus at home and abroad. By studying the structure of source data, it conducts data analysis and processing, designs alignment of bilingual corpus, and explicates ways to build bilingual parallel corpus of cancer in English and Chinese. It discusses relevant design mechanisms and the quality of the corpus.

〔**Keywords**〕　Physician Data Query (PDQ); medical parallel corpus; cancer

1　引言

在疾病领域癌症获得了较多社会关注与资本投入。想要了解甚至治愈癌症，一手资料信息以及知识库具有巨大的信息传播和数据挖掘价值。目前医学领域绝大部分一手资料以及前沿资料信息，顶尖的学术成果及报告依然以英文为主要载体，因此，

〔**收稿日期**〕　2018－12－09

〔**作者简介**〕　马鹤桐，硕士，助理研究员；通讯作者：李姣，研究员。

〔**基金项目**〕　中国医学科学院医学与健康科技创新工程服务“一带一路”战略先导科研专项“卫生信息服务研究”（项目编号：2017－I2M－B&R－10）；国家重点研发计划“精准医学本体和语义网络构建”（项目编号：2016YFC090 1901）

英文权威文献资料及信息是我国众多医学专业开展基础和临床研究的重要参考。同时，我国国内医学从业人员交流方式及文字载体都采用中文，更提升了从英文到中文的自然语言处理的需求。基于以上背景，医学领域，特别是癌症等热点疾病领域的知识库、语料库建设更显得尤为重要。

机器翻译作为自然语言处理的重要分支，是实现从英文到中文翻译的重要途径，想要获得高质量的翻译水平，首先需要高质量的平行语料[1]。为此，我们选取医疗领域内语料质量好，并聚焦在前沿热点的癌症领域，以美国国家癌症研究所（National Cancer Institute，NCI）发布的癌症综合信息库（PDQ®）作为英文语料。美国国家癌症中心在 1971 年的法案中确定建立医生信息查询库（Physician Data Query，PDQ），旨在收集、分析、传播所有在癌症预防、诊断和治疗方面有价值的数据，包括建立国际癌症研究数据中心以实现收集、分类、存储及传播癌症信息的目的[2]。PDQ 描述了癌症、药物、基因术语及其他相关知识，其内容准确全面，收录前需经过同行评议，内容是较为权威的癌症信息库[3]。面向不同受众，如医生和病患，PDQ 提供了不同版本，并被翻译为不同语言[4]。同时 PDQ 保持每周更新的频率，为癌症专家学者及其他受众提供最新信息[5]。中文语料则是由中国国家癌症中心组织翻译团队进行翻译[6]。

2 相关研究

2.1 语料构建流程

目前，国际上平行语料构建的基本流程与技术已经相对成熟，如就不同目的来说，Dalianis 等人完成了部分平行语料的构建[7]，Koehn 则是以机器翻译为目的构建了一份平行语料，为后继研究者们提供了初始的研究思路和设计流程[8]。就不同构建语种来说，Mohamed 等人实现了小语种的 Sinhala 与 Tamil 的自动平行语料构建[9]。就不同构建流程来说，Tyers 等人通过数据爬取、收集、以及分句、句子对齐等工具成功构建了巴尔干半岛语言的平行语料[10]，Cettolo 等人以类似的方式但更简便的流程设计完成了 TED 演讲资料的语料库构建[11]。Hailong Jin 等人提出通过数据处理、跨语种知识图谱建设、数据质量提升以及应用来构建跨语种知识图谱[12]。

2.2 句子对齐策略

在句子自动对齐方面，目前主要有 3 种相对成熟的方法，第 1 种是统计方法，如根据词量和句子的长短予以判断；第 2 种是语言学方法，如根据语言学特征、语态、句型等给予判断；第 3 种是前两种方法的混合方法。McEnery 等人通过比较两个句子的长度以计算他们之间对齐的相似度，以实现中英双语的句子对齐[13]，Chuang 等人则是利用标点符号对齐句子[14]。基于语言学的方法常常不仅采用语言学特征得以实现，也会加入统计学的元素共同实现对齐，最典型的的方法就是 Gale 等人提出的基于锚点的对齐方法[15]，Wu 等人的方法则是基于句子长度实现对齐[16]，Moore 等人提出了将句子长度与统计概率结合的方法[17]，Varga 等人则是基于长度的方法，插入双语字典以获得更好的准确率[18]，此外，Sennrich 提出根据机器翻译与目标文本的比对取代直接对齐方式[19]。

2.3 我国语料库构建状况

我国尽管在平行语料库的构建方面已从设计转入建设阶段[20]，如以开发基于语料库的翻译教学及翻译软件为目的，于 2009 年启动建设的英汉医学平行语料库[21]；以促进医学英语语言学和教学研究为目的，聚焦于医学著作的英汉双语的平行语料库[22]。但在我国医学领域，平行语料库的建设尚处于初步建设阶段，且在癌症领域相对空缺，没有较为完善覆盖癌症大量权威、专业信息知识的知识库或语料库，仍需要大量平行语料库提供相关信息及数据，并作为其他功能的支撑数据，实现数据挖掘与分析的目的。

2.4 构建语料库的重要性

构建癌症领域语料库对于癌症信息传播具有重要意义。语料库根据癌症的不同种类，分别建设不同子库，从癌症筛查、诊断、预防、治疗等方面分

别详细描述癌症信息及相关知识。建设癌症资源知识库可以有效帮助患者、医生、医学科研人员获取大量精准并具有指导性的信息与知识，将最基础及权威的理论背景传达给更多用户，方便其即时理解并获取前沿信息，如先进药物的生产、成分、使用方法、副作用及应用效果或已有药物的最新用法等内容，均可通过知识库实现更为广泛的传播。构建跨语种癌症信息库对于机器翻译、知识库建设、知识图谱建设及决策支持具有重要意义。由于医学领域内容具有快速更新的特点，医学从业人员尤其是一线的临床医生与科研人员需要尽可能实时跟进领域更新，导致以英文为主要载体的文献资料自动翻译需求的增加。机器翻译的效果甚至可以作为机器智能水平的衡量[23]，帮助医学从业人员实现辅助机器阅读和机器理解，从而更好地制定医疗决策。同时机器翻译离不开平行语料库的构建，而跨语种知识库也为更多用户通过熟悉的语言了解疾病知识提供了保障。另外，跨语种语料库的建设在相当程度上，有助于实现语料溯源的需求，使得双语医学著作、大型医学语料的多语种自动更新成为可能。

3　研究方法和内容

3.1　总体流程

根据对相关研究的了解及对源数据的调研，设计了构建平行语料的研究方法，包括以下步骤：（1）收集双语种 PDQ 数据内容。（2）数据分析与处理。包括基于元数据的数据分析和数据处理分析，其中元数据分析包括标签锚点的策略设计与数据内容的统计分析；数据分析处理包括实现数据对齐、内部去重、外部去重 3 个步骤并完成分析。（3）设计语料库规则，构建双语平行语料库，见图 1。

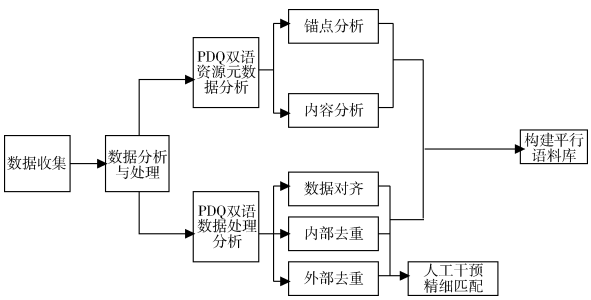

图 1　研究方法流程

3.2　数据收集

本文选取 PDQ 2013 – 2014 年发布的癌症信息英文总结专业版，并收集对应的中文版本，分别作为不同数据集的平行语料。其中 22 篇英文版，22 篇中文版。PDQ 专业版数据集从治疗、预防、诊断、筛查等方面描述了不同癌症的具体信息，中英文 PDQ 均为 XML 格式，内含大量标签及嵌套文本内容。本文涉及到的癌症种类分别是肺癌、肝癌、结直肠癌、乳腺癌、食道癌、胃癌，其中英文句子数、中文句子数、英文单词数和中文单词数分别为 22 091，12 365，291 095 以及 170 688。

3.3　数据分析与处理

3.3.1　元数据分析　本文统计了基于主题文件的标签，希望借此对中英双语料深入了解，对中英文件中非共有的标签及内容进行删除，找到共有的标签，进行内容提取和对齐，并尝试以该方法作为基准，对其他标签的内容进行提取与处理。据统计共有标签 54 种，其中英文版独有元素 9 种，中文版独有元素 7 种；对每种癌症的中英文标签及数量同时进行比较，见表 1。标签内容与出现次数均一致代表在源双语语料中，双语内容可被认为对齐程度较高。经过统计，中英文标签内容与数量均对齐者为 58.59%，显示中英文资料并没有完全对齐，数据需要进一步处理。分析双语语料发现，不少标签内都包含文本句对，是可以提取的主要内容来源，但多数标签在文件中出现次数不一致，且不只包含文本，内容涉及到大量字符、字母、数字等干扰性内容。只有 Para 标签有 ID 属性，并全部由正文句对组成，具有唯一标识的功能。因此，对 Para 标签中的内容进行深入分析，以了解其作为对齐和基准语料操作上下文语料的可行性。Para 标签代表“段落”，其中包含大量可对齐的句对，均为癌症信息的主要描述性内容，是构建平行语料的主要组成元素。因此将 Para 标签中的内容提取出来并进行统计，para 代表段落数中英文保持完全一致，说明 Para 标签可以作为双语平行语料构建的锚点。

表 1 PDQ6 种癌症中英文标签统计

标签/元素统计	共有元素	英文版独有元素	中文版独有元素
标签/元素内容统计	AltTitle，ColSpec，DateFirstPublished，DateLastModified，Emphasis，entry，ExternalRef，GeneName，ItemizedList，ListItem，LOERef，MainTopics，MediaLink，OrderedList，Table，TBody，TermRef，TFoot，TGroup，Para，ProtocolRef，Row，SectionType，SectMetaData，SpecificDiagnosis，Strong，Summary，SummaryAudience，SummaryDescription，SummaryLanguage，SummaryMetaData，SummaryRef，SummarySection，SummaryType，SummaryURL，Superscript，THead，Title	Caption，Citation，ListTitle，Note，Reference，ReferenceSection，ScientificName，Subscript，SummaryTitle	em，font，listh2，span，summaryh2，sup，u
标签/元素数量统计	38	9	7

3.3.2 数据处理分析 将一一对应的 Para 标签作为对齐基准，根据上下文的顺序将其他标签自动对齐并进行手工校验，以英文为基准对齐中文标签。所有标签对齐后，从中提取大量文本内容，舍弃非文本的字符、字母、数字以及其他没有实际内容的数据。对大量语料进行分析并设计自动分句策略，对数字、缩写及多次出现的内容进行总结，将其归纳作为分句辅助策略。最终确定为主要通过标点实现分句，结合辅助策略共同完成自动分句。分句后未对齐的句对，采用手动对齐的方法，期间允许存在一对多，多对一的关系。完成语料的初步处理后，对初步平行语料进行文件内部去重任务，删除完全重复多次的语料内容，使得每个句子在文件内部保证其唯一性，去除冗余。完成去重后，对整体 22 份平行语料进行整体去重工作，筛选出现在多个语料文件的相同内容，将相同内容单独提出置于一份新建的语料文件中，删除原文件中的重复内容，保留每个句子在整体语料中的唯一性。

3.4 构建平行语料库

根据单词个数对语料进行分类，单词个数为 1 ~2 个的，存储在单词（word）标签内，单词个数 3 ~7 个的，存储在短语（phrase）标签内，单词个数在 8 个及以上的，存储在句子（sentence）标签内。根据以上对语料的处理步骤，设置平行语料策略，构建平行语料库。具体策略包括：中英文语料分开构建、平行语料设置相同的 ID 以便查找、内容包括单词，短语，句子 3 个层面。

4 研究结果

经过对语料的多步骤处理，完成了美国国家癌症中心发布的癌症综合信息 2013 –2014 年版本中英双语语料的处理和平行语料的构建。语料内容包括肝癌、肺癌、结直肠癌、乳腺癌、食道癌和胃癌 6 种癌症的预防、筛查、治疗等专题信息，最终形成 23 对 46 份 XML 格式的中英文内容分开的平行语料，覆盖 514 对单词、1 236 对短语、5 438 对句子平行对齐的医学癌症信息中英双语平行语料。经过对齐程度验证，发现正确率在 99% 以上，表明该平行语料库对齐准确率较高，可以作为机器翻译、知识库、知识图谱构建的源数据，见图 2。

```
<Phrase id="p21">便潜血检查（FOBT）</Phrase>
<Phrase id="p22">文本引用，或</Phrase>
<Phrase id="p23">便潜血检查（FOBT）</Phrase>
<Phrase id="p24">本总结更新通知（02/21/2014）</Phrase>
<Phrase id="p25">病例对照研究，正在进行中的RCT</Phrase>
<Phrase id="p26">病灶检出率证据</Phrase>
<Phrase id="p27">从NCI得到更多信息</Phrase>
<Phrase id="p28">减少结直肠癌的证据</Phrase>
<Phrase id="p29">对左半结肠的60%-70%</Phrase>
<Phrase id="p30">开会讨论，</Phrase>
<Phrase id="p31">RCT=随机对照试验</Phrase>
<Phrase id="p32">使用本总结的许可</Phrase>
<Phrase id="p33">FOBT联合软式乙状结肠镜</Phrase>
<Phrase id="p34">主要终点为CRC相关死亡率</Phrase>
<Phrase id="p35">特异性从97.8%降低到96%</Phrase>
<Phrase id="p36">该研究出勤率为65%</Phrase>
<Phrase id="p37">虚拟结肠镜（计算机断层[CT]结肠成像）</Phrase>
<Phrase id="p38">局部肿瘤T1-3N0M0的百分比</Phrase>
<Phrase id="p39">关于结直肠癌死亡率降低的证据</Phrase>
<Phrase id="p40">近92%的检查为再水化检查</Phrase>
<Phrase id="p41">对同效终点的效应强度</Phrase>
<Phrase id="p42">新型FOBT：来自非随机对照试验的证据</Phrase>
<Phrase id="p43">关于本总结的意见或评论</Phrase>
<Phrase id="p44">两篇系统性综述总结了这些试验的结果</Phrase>
<Phrase id="p45">iFOBT对CRC的检出率>60%且≤90%</Phrase>
<Phrase id="p46">但筛查不影响全因死亡率</Phrase>
<Phrase id="p47">CT结肠成像中常可见结肠外异常病变</Phrase>
<Phrase id="p48">粪便DNA突变检测</Phrase>
<Phrase id="p49">随访时间为6-13年</Phrase>
<Phrase id="p50">特异性为86%，PPV为23%</Phrase>
<Sentence id="s1">软式光纤乙状结肠镜于1969年问世</Sentence>
<Sentence id="s2">结肠镜的严重并发症发生率为0.5%</Sentence>
<Sentence id="s3">软式乙状结肠镜或FOBT检查结果为阳性则需行结肠镜</Sente
<Sentence id="s4">关于PDQ总结的较佳引用方法为：</Sentence>
<Sentence id="s5">计算不同阈值的敏感性和特异性</Sentence>
<Sentence id="s6">调整阈值增加敏感性的同时伴随着特异性降低</Sentence>
<Sentence id="s7">男性CRC发病率高于女性</Sentence>
<Sentence id="s8">对左半结肠约60%-70%，对右半结肠不确定</Sentence>
<Sentence id="s9">这些图像模拟传统结肠镜的作用</Sentence>
<Sentence id="s10">若年人的不良事件发生率可能增加</Sentence>
<Sentence id="s11">1976年1500px软式乙状结肠镜上市</Sentence>
```

```
Phrase id="p21">Fecal Occult Blood Test (FOBT)</Phrase>
Phrase id="p22">be cited with text, or</Phrase>
Phrase id="p23">Fecal Occult Blood Testing (FOBT)</Phrase>
Phrase id="p24">Changes to This Summary (02/21/2014)</Phrase>
Phrase id="p25">Case-control studies, RCTs in progress</Phrase>
Phrase id="p26">Evidence about lesion detection rate</Phrase>
Phrase id="p27">Get More Information From NCI</Phrase>
Phrase id="p28">Evidence about colorectal cancer reduction</Phrase>
Phrase id="p29">About 60%-70% for left colon</Phrase>
Phrase id="p30">be discussed at a meeting,</Phrase>
Phrase id="p31">RCT = randomized controlled trial</Phrase>
Phrase id="p32">Permission to Use This Summary</Phrase>
Phrase id="p33">Combination of FOBT and Flexible Sigmoidoscopy</Phrase>
Phrase id="p34">The primary endpoint was CRC-specific mortality</Phrase>
Phrase id="p35">Specificity declined from 97.8% to 96%</Phrase>
Phrase id="p36">Attendance in this study was 65%</Phrase>
Phrase id="p37">Virtual Colonoscopy (Computed Tomographic [CT] Colonography
Phrase id="p38">% Localized = T1-3 N0 M0</Phrase>
Phrase id="p39">Evidence about colorectal cancer mortality reduction</Phrase>
Phrase id="p40">Nearly 92% of tests were rehydrated</Phrase>
Phrase id="p41">Magnitude of Effects on Surrogate Endpoints</Phrase>
Phrase id="p42">Newer FOBTs: Nonrandomized Controlled Trial Evidence</Phrase
Phrase id="p43">Questions or Comments About This Summary</Phrase>
Phrase id="p44">Results were summarized in two systematic reviews</Phrase>
Phrase id="p45">iFOBT detects >60% and ≤90% of CRCs</Phrase>
Phrase id="p46">There was no effect on all-cause mortality</Phrase>
Phrase id="p47">Extracolonic abnormalities are common in CT colonography</Ph
Phrase id="p48">Detection of DNA Mutations in the Stool</Phrase>
Phrase id="p49">Follow-up ranged from 6 to 13 years</Phrase>
Phrase id="p50">Specificity was 86% and PPV was 23%</Phrase>
Sentence id="s1">The flexible fiberoptic sigmoidoscope was introduced in 1969</
Sentence id="s2">The serious complication rate of colonoscopy was 0.5%</Sente
Sentence id="s3">A positive FS or FOBT qualified for colonoscopy</Sentence>
Sentence id="s4">The preferred citation for this PDQ summary is:</Sentence>
Sentence id="s5">Sensitivity and specificity were calculated at various cut-points
Sentence id="s6">Increasing sensitivity entailed adjusting cut-points to decrease
Sentence id="s7">The incidence is higher in men than in women</Sentence>
Sentence id="s8">About 60%-70% for left colon; uncertain for right colon</Sente
Sentence id="s9">These images simulate the effect of a conventional colonoscopy
Sentence id="s10">The rate of complications may increase among older persons<
Sentence id="s11">The 60 cm flexible sigmoidoscope became available in 1976</
```


图 2　癌症双语平行语料

5　结语

本文以癌症综合信息库 PDQ（Physician Data Query，PDQ）的英文版以及相应中文翻译版本为研究对象，针对肺癌、肝癌、结直肠癌、乳腺癌、食道癌、胃癌等 6 种癌症构建平行语料库，通过对源文件的结构研究、数据分析与处理，语料对齐策略设计构建平行语料。平行语料格式为 XML，包含单词、短语与句子。结合前期的基于结构与内容的研究成果，发现在 PDQ 双语语料中，虽然存在一些“多对一、多对多”的句对，但大量内容仍旧可以做到一一对应，这些双语平行语料可成为良好的训练语料。在未来的工作中，将设计自动语料构建流程辅以自动语料评价技术，考虑与不同数据格式的兼容性，以便能够提供更加高精度、高准确率的语料知识库，实现辅助机器阅读理解，以及机器翻译的目的。

参考文献

1　柏晓静，常宝宝，詹卫东，等. 构建大规模的汉英双语平行语料库［C］. 北京：全国机器翻译研讨会，中国中文信息学会，中国人工智能学会，2002.

2　NIH – National Cancer Institute. PDQ Retrieved from. 2018［EB/OL］.［2019 – 04 – 07］. https：//www. cancer. gov/publications/pdq.

3　冀玉静. 医师数据查询系统的术语结构、内容模块及获取途径［J］. 医学信息学杂志，2016，37（9）：70 – 73.

4　孙月萍，侯震，侯丽，等，医学知识库语言学特征比较分析［J］. 医学信息学杂志，2018，39（1）：46 – 50.

5　Manrow RE，Beckwith M，Johnson LE，NCI’s Physician Data Query（PDQ®）Cancer Information Summaries：History，Editorial Processes，Influence，and Reach［J］. Journal of Cancer Education，2014，29（1）：198 – 205.

6　中文版癌症综合信息库 2018［EB/OL］.［2018 – 04 – 07］. http：//pdq. cicams. ac. cn/.

7　Dalianis H，H C Xing，X Zhang. Creating a Reusable English – Chinese Parallel Corpus for Bilingual Dictionary Construction［C］. Valletta：Proceedings of International Conference on Language Resources and Evaluation，2010：1700 – 1705.

8　Koehn P. EuroParl：a parallel corpus for statistical Machine translation［C］. Phuket：Machine Translation Summit，2005：79 – 86.

9　Mohamed MZ，Ihalapathirana A，Hameed RA，et al. Automatic Creation of a Word Aligned Sinhala – tamil parallel corpus［C］. Osaka：Proceedings of the 6th workshop on South and Southeast Asian Natural Language Processing，2016：124 – 132.

10　Tyers FM，Alperen MS. South – East European Times：a parallel corpus of Balkan languages［C］. Malta：Proceedings of the workshop on Explitation of multilingual resources

and tools for Central and (South) Eastern European Languages, 2010: 49 – 53.

11 Cettolo M, Girardi C, Federico M. Wit3: web inventory of transcribed and translated talks [C]. Trento: Proceedings of the 16th EAMT Conference, 2012: 261 – 268.

12 Jin HL, Li CJ, Zhang J, et al. XLORE2: large – scale cross – lingual knowledge graph construction and application [J]. Data Intelligence, 2019, 1 (1): 77 – 98.

13 McEnery AM, Oakes MP. Sentence and Word Alignment in the CRATER Project: methods and assessment [C]. Dublin: Proceedings of the EACL – SIGDAT Workshop, 1995: 77 – 86.

14 Chuang TC, Yeh KC. Aligning Parallel Bilingual Corpora Statistically with Punctuation Criteria [J]. International Journal of Computational Linguistics & Chinese Language Processing, 2005, 10 (1): 95 – 122.

15 Gale WA, Church KW. A Program for Aligning Sentences in Bilingual Corpora [C]. Berkely: Proceedings of ACL 1991, 1991: 177 – 184.

16 Wu D. Aligning A Parallel English – Chinese Corpus Statistically With Lexical Criteria [J]. Computer Science, 1994, 4 (4): 80 – 87.

17 Moore RC. Fast and Accurate Sentence Alignment of Bilingual Corpora [C]. London: Proceedings of the 5th AMTA, 2002: 135 – 144.

18 Varga D, Halacsy P, Kornai A, et al. Parallel Corpora for Medium Density Languages [C]. Borovets: Proceedings of International Conference Recent Advances in Natural Language Processing, 2005: 590 – 586.

19 Sennrich R, Volk M. Iterative, MT – based Sentence Alignment of Parallel Texts [C]. Riga: Proceedings of the 18th Nordic Conference of Computational Linguistics, 2011: 175 – 182.

20 窦川川, 李婷玉. 关于建立大型标准中医汉英平行语料库的探索性构想 [J]. 中国中医基础医学杂志, 2017 (4): 561 – 563.

21 管新潮, 胡开宝, 张冠男. 英汉医学平行语料库的创建与初始应用研究 [J]. 当代外语研究, 2011 (9): 36 – 41.

22 陈潇潇, 葛诗利. 医学著作英汉双语平行语料库的构建 [J]. 外语艺术教育研究, 2012 (1): 33 – 37.

23 Hirschberg J, Manning CD. Advances in Natural Language Processing [J]. Science, 2015, 349 (6245): 261 – 266.

(上接第 46 页)

26 Ogbolu Y, Fitzpatrck GA. Advancing Organizational Cultural Competency with Dissemination and Implementation Frameworks: towards translating standards into clinical practice [J]. Ans Advances in Nursing Science, 2015, 38 (3): 203 – 214.

27 Guerrero LR, Terry N, Davidson PL. Nih Career Development Awards in Clinical and Translational Science Award Institutions: distinguishing characteristics of top performing sites [J]. Clinical & Translational Science, 2015, 7 (6): 470 – 475.

28 Billett AL, Colletti RB, Mandel KE, et al. Exemplar Pediatric Collaborative Improvement Networks: achieving results [J]. Pediatrics, 2013, 131 (Sup4): 196.

29 Marlene M. Roles for Children's Hospitals in Pediatric Collaborative Improvement Networks [J]. Pediatrics, 2013, 131 (Sup4): 215 – 218.

30 Lannon CM, Miles PV. Pediatric Collaborative Improvement Networks: bridging quality gaps to improve health outcomes [J]. Pediatrics, 2013, 131 (Sup4): 187 – 188.

31 Margolis PA, Peterson LE, Michael S. Collaborative Chronic Care Networks (c3ns) to Transform Chronic Illness Care [J]. Pediatrics, 2013, 131 (Sup4): 219.

32 刘爽, 弓孟春, 朱卫国, 等. 精准医学信息学的关键技术及应用发展方向 [J]. 医学信息学杂志, 2017, 38 (9): 8 – 16.

33 刘爽, 冯时, 徐平, 等. 精准医学应用于罕见神经疾病的现状与前景 [J]. 科技导报, 2017, 35 (16): 37 – 42.

34 Jansen MW, Van Oers HA, Middelweerd MD, et al. Conditions for Sustainability of Academic Collaborative Centres for Public Health in the Netherlands: a Mixed Methods Design [J]. Health Research Policy & Systems, 2015, 13 (1): 36.

35 樊光辉. 聚力推进转化医学创新发展 [J]. 转化医学电子杂志, 2016, 3 (1): 1 – 4.

36 耿波, 刘宁, 陈静, 等. 基于转化医学模式的高校图书馆参考咨询服务 [J]. 中华医学图书情报杂志, 2017, 26 (3): 46 – 49.

37 谢凡. 图书馆在转化医学趋势下的学科化服务 [J]. 现代医院, 2016, 16 (3): 455 – 458.