Advances in Social Science, Education and Humanities Research (ASSEHR), volume 237
3rd International Conference on Humanities Science, Management and Education Technology (HSMET 2018)

Construction and Application of English-Chinese Bilingual Corpus Based on Localization Requirements: A Case Study of Miniature Video Game Corpus

Mingxi Chen[1,a] and Shuying Wu[2,b,*]

[1]The School of Translation Studies, Jinan University, Guangzhou, Guangdong, China

[2]The School of Translation Studies, Jinan University, Guangzhou, Guangdong, China

[a]23333hawking@gmail.com, [b]rabyxizi@foxmail.com

*Shuying Wu

Keywords: English-Chinese corpus, Video Game Localization, Parallel texts

Abstract. Based on the development of localization in the new-era language service industry, the article analyzes the requirements of constructing bilingual corpora. Combined with the current status of representative corpora at home and abroad as well as the purpose of corpus construction, the article takes the bilingual corpus of miniature video games as an example to carry out a research on the corpus construction, exploring the practical application of the corpus and discussing the advantages and disadvantages of its actual operation.

基于本地化需求的英汉语料库构建与应用探究——以小型游戏语料库为例

陈铭熙[1, a]，吴书颖[2,b,*]

[1]暨南大学翻译学院，广州，广东，中国

[2]暨南大学翻译学院，广州，广东，中国

[a]23333hawking@gmail.com, [b]rabyxizi@foxmail.com

*吴书颖

关键词: 中英双语语料库；游戏本地化；平行文本

中文摘要. 本文基于新时代语言服务行业中本地化的发展，分析双语语料库的构建需求，结合国内外语料库代表先例的发展现状和语料库的构建目的，以个人小型游戏英中双语语料库为例对语料库的构建进行剖析，探讨语料库的应用前景及其在实际操作中的优缺点。

1. 引言

过去，在面对量大复杂的项目时，译者缺乏技术支持，工作效率低。近十年，借助内嵌大型数据的语料库来进行翻译实践和研究的方法渐趋流行，而语料库的使用大大提高了翻译效率和质量。Varantola表示，在翻译过程中，50%的时间都用于查找相关资料[1]。而运用计算机辅助翻译手段之一的双语语料库，可以明显提高翻译的速度和质量[2]。

如今，一些新兴翻译领域逐步发展，游戏本地化就是其中之一。在中国，游戏本地化俗称为汉化。市面上的游戏汉化，除了游戏发行公司的官方中文版本，其余大多数是由游戏爱

好者翻译后通过社交媒体传播。言明乐将游戏翻译的内容概括为弹出窗口和工具提示、系统信息、游戏内文本、音频软件和过场动画等[3]，由此可见游戏汉化语料之庞大。将这些术语整理成数据库，即游戏双语平行语料库，对未来的翻译实践和研究具有理论价值和应用意义。

双语平行语料库开放灵活，能够随时添加、瞬时记忆、及时更新、调取灵活。语料库按照用途可以分为通用语料库及专门用途语料库。论及某一专业或学科的特色研究时，就应运用专门的语料库[4]。本文谈论的游戏本地化语料库属于专门用途双语语料库。虽然2003年辛克莱（Sinclair）就已提出专门用途的语料库会是未来语料库语言学的一大发展趋势[5]，但是，目前国内专门用途语料库发展情况并不乐观。

2. 英汉双语平行语料库建立——以小型游戏本地化双语平行语料库为例

专门用途双语语料库一般针对某一类型文本，有着高于通用语料库的构建难度：其一，资料收集难度高。游戏本地化文本作为商用或私用资料，一般只有从事游戏本地化的专业人士才得以获得足量原始语料，而通过公共途径获取的资源有限，成本也相对高昂。其二，相关学科发展处于初始阶段，缺乏研习人才，语料库翻译学在全球仅有不足30年的发展史。

创建双语平行语料库前，需要熟悉必要的软件和原理，谨记以下三个关键点[6]：

（1）注重选材质量，选择合适的本地化材料，这有助于后续的双语对齐；

（2）双语语料需要自动对齐处理，如此才能保证平行语料库建库规模足够大；

（3）为了研究的便利和深入，需要对双语语料做尽量多且精的标注。

2.1 设计语料库

创建语料库，第一步是语料库设计。必须先明确各类设计参数，主要包括：使用目的、语料库类型、语料来源、语料库规模和和翻译语向。本文研究的“冒险类游戏英汉平行语料库”，其创建目的是利用语料库对游戏平台Steam上多个游戏平台Steam上冒险类游戏的中译本进行收集和分析，通过数据分析得出客观研究结论。基于该研究目的，语料库的创建过程可以设计为图1所示。

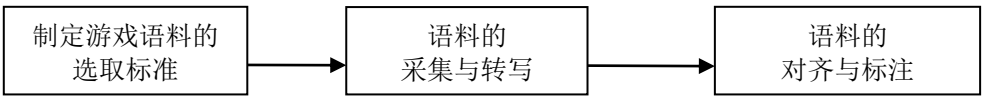

图1 游戏语料库设计流程简图

由于游戏本地化资料收集难度大、成本高，仅仅作为个人研究的游戏本地化语料库，创建容量、文本规模可以稍小。

2.2 语料采集与转写

游戏本地化作为翻译实践的一部分，选择的语料应多为在归化翻译策略指导下的译文，确保译文精准传达原文含义，表达简练，文笔有表现力，内容易于理解[7]。为了确保收录译文的质量，选取译文时官方版本优先，民间汉化次之。因为国内游戏汉化普遍存在质量层次不齐的状况，本文则选择知名游戏论坛下载量较高的最新汉化补丁，并做二次审核。本文主要收集了《古墓丽影：崛起》(Rise of the Tomb Raider)、《伊迪芬奇的秘密》(What Remains of Edith Finch)、《饥荒》(Don’t Starve)以及《求生之路2》(Left 4 Dead 2)的部分游戏文本，包括开始游戏界面、任务列表与人物对话。其中《求生之路2》的中译本来自于3dm游戏论坛，《饥荒》、《古墓丽影：崛起》选自游侠网，《伊迪芬奇的秘密》则选自官方中文版本。

图2 《饥荒》官方与汉化游戏截图

由于游戏文本基本内嵌在游戏中，因此需要人工游戏截图然后转写语料，无需语料降噪。如图2所示，需将游戏中英文版本分别截图进行语料采集，再由人工转写。

若相同的游戏操作有不同的译文版本，则择优保留。例如冒险游戏中常见的操作“Change Character/Player”，这四款游戏出现了两个译文：①转换人物角色；②使用该角色。根据游戏高度互动的性质，作者保留第二个译文。再例如游戏模式中的“Single/Solo Mode”，在没有特殊游戏设定的情况下一般都译为“单人模式”。而《求生之路2》的汉化译本“单人战斗模式”稍显累赘，且“单人战斗模式”对应的术语应为“1V1 Battle Mode”。

2.3 语料整理与加工

由于游戏文本以较短句对为主，容易一一对应，原文与译文匹配度高。使用SDL Trados的WinAlign工具进行语料对齐后，保存为TMX格式文件保存，可以直接应用于翻译匹配，从而形成未经标注的生语料库(Raw Corpus)。除了生语料库外，一般语料库都需要进行标注。标注有多种层次，杨慧中[8]列举了主要几种：格式属性标注，如对段落、字体、字号标注；对识别信息进行标注，如作者、体裁；特殊标注，如错误赋码等。本文研究的主要有两种：自动标注和人工标注。自动标注主要应用于词性赋码，使用工具BFSU Stanford POS Tagger。

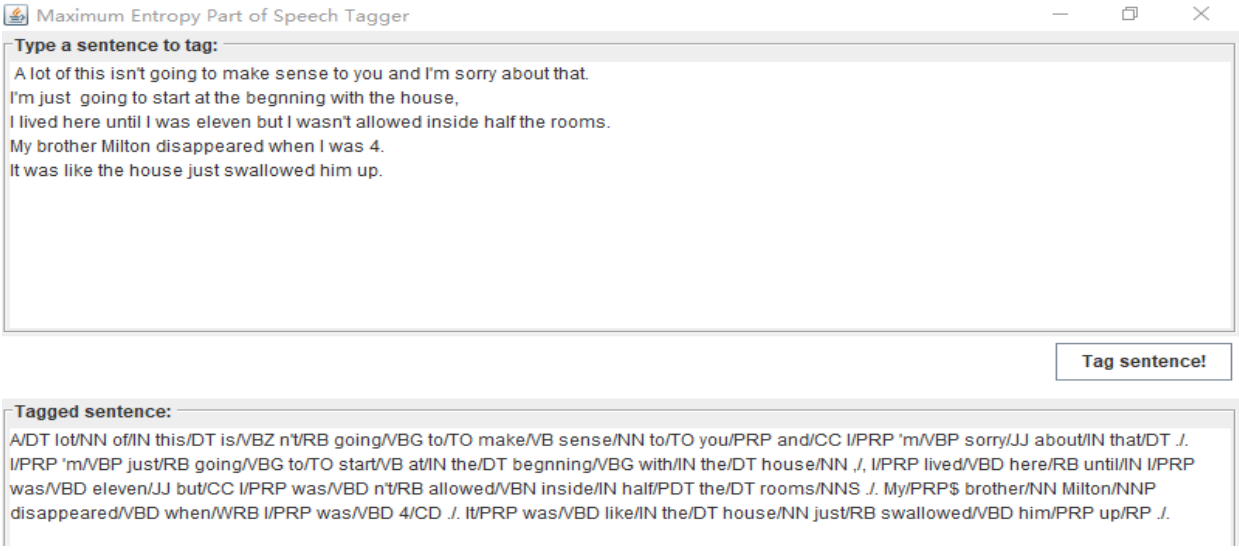

图3 输入游戏《伊迪芬奇的秘密》句子文本进行标注

人工标注主要应用于汉化文本的语误标注（术语错误、时态误用等），应用工具为Microsoft Word以及“快速粘贴于自动填表软件[9]”。

3. 小型游戏本地化双语平行语料库的应用

除了在游戏本地化作为记忆库提供翻译支持外，该双语平行语料库还可应用于翻译共性研究、翻译策略研究以及翻译主体研究。

(1) 翻译共性研究

一方面，可以从宏观层面研究游戏文本词汇密度与变化程度、句子的长度与复杂程度，探讨游戏语言的特点；另一方面，可以从微观层面研究游戏文本的叙事结构、隐喻、习语等，甚至可以跳出文本本身，从游戏设计者的目的、游戏故事背景（意识形态、历史、地域等）去考虑译文翻译的风格。

(2) 翻译策略研究

翻译策略在文本及非文本因素方面都有着密切的联系。在文本层面，需要对游戏文本按段落、句子、词分类，然后考虑翻译策略与文本之间的联系；在非文本层面，可以关注游戏文本比较常见的“宗教”、“民族化”、“国际化”等意识形态角度，以及“归化”、“异化”等策略在译文的动态应用，从而完善对游戏本地化策略的理论认知。

(3) 翻译主体研究

在游戏本地化过程中，作为消费者的异国玩家与游戏文本作者之间孰是主体，孰是被动接受方，一直是个备受争议的话题。除此之外，伴随文本翻译而出现的游戏文化输出、译者、游戏主题、翻译目的之间的互动关系也可作为研究的主题。

4．结束语

以上主要讨论了在大数据时代背景下游戏本地化的发展现状，分析国内双语平行语料库的发展、小型游戏本地化双语语料库的功用以及在语言对比、翻译研究等方面。尽管语料库有着瞩目的优势，但它的不足也应当引起我们注意。首先，建设、维护和更新语料库需要耗费众多的人力、物力及时间，成本高昂，成果却不能立现。其次，即使是一个内存无穷尽的记忆库，它也不能容纳所有同类型的文本，这既不切实际也不甚必要；再者，语料库的语料采集需具有代表性和高质量。但“代表性”和“高质量”却没有权威的评定标准，每个人都有自己的心中的“代表作”和“好作品”，从而得出的结论也并非具有普遍性。在如今的大数据时代，中国是游戏消费市场上的主力之一。然而，无论是游戏官方还是翻译研究者，在游戏本地化，尤其是语料库方面的研究，还亟需完善。

References

[1] Varantola. Translators, Dictionaries and Text Corpora[J], 1997, 12(1).

[2] Aston G.. Corpus use and learning to translate[J]. Textus, 1999, 09(12): 289-314.

[3] Yan Mingle. An Initial Study on Video Game Localization[J]. Read and Write Periodical, 2009, 6(4): 20, 32.

[4] L. Lu, A one-vendor multi-buyer integrated inventory model, *European Journal of Operational Research*, vol. 81, pp. 312-323, 1995.

[5] CAO Hejian. A Corpus-Driven Study of Business English [M]. Beijing: University of International Business and Economics Press , 2008

[6] KE Fei. Parallel Corpus based Approaches to Translation Studies [J]. Foreign Languages and Their Teaching, 2002, 35-39.

[7] GUAN Chaoxin, TAO Youlan. Corpus and Translation [M]. Shanghai: Fudan University Press,2017,69

[8] YANG Huizhong,An Introduction to Corpus Linguistics[M]. Shanghai: Shanghai Foreign Language Press, 2010