

机器翻译和学习二语的学生翻译在汉译英中面对多词单位时的母语负迁移的对比分析

朱 赫

悉尼大学, 悉尼, 澳大利亚

收稿日期: 2022年3月16日; 录用日期: 2022年3月25日; 发布日期: 2022年4月7日

摘 要

西交利物浦大学语言技术研究中心打算设计一个创新的应用程序, 以实现自动编辑机器翻译的输出, 并提高汉英双语者撰写的英语文本的语言纠正工具的准确性。作为本次西交利物浦大学的试点研究, 本研究拟采用错误分析的方法, 探讨机器翻译(MT)和学习第二语言的(L2)学生在汉英翻译中关于L1 (或者说源语言)的负迁移是否存在明显差异; 如果存在差异, 则找出负迁移的分布特征。这是一项基于语料库的研究, 从15篇在线新闻报道中选取40个多词单位(MWU)。然后, 采用定性和定量两种方法。实证研究占主导地位, 描述性研究作为转移类型的补充。研究对象是有道在线翻译和必应在线翻译, 以及西交利物浦大学的10名高年级学生。在收集了机器翻译和翻译测试的输出数据后, 邀请一位母语为英文人士协助作者识别负面词汇迁移。此外, 还在《近代汉语》语料库中确定了相关汉语词汇的频率。最后, 本研究得到了一些有用的发现: 在汉英翻译中, 机器翻译和学习第二语言的学生在MWU方面确实存在明显的差异, 这可以归因于MT和学生在翻译过程中的各自优势和劣势。最后, 本研究对相关研究人员提出了一些启示。

关键词

负迁移, 翻译, MT, MWU, 二语习得学生

A Contrastive Analysis of L1's Negative Transfer in the Chinese-English Translation between Machine Translation and L2 Students in Terms of MWU

He Zhu

The University of Sydney, Sydney, Australia

Received: Mar. 16th, 2022; accepted: Mar. 25th, 2022; published: Apr. 7th, 2022

文章引用: 朱赫. 机器翻译和学习二语的学生翻译在汉译英中面对多词单位时的母语负迁移的对比分析[J]. 国外英语考试教学与研究, 2022, 4(1): 13-29. DOI: 10.12677/oetpr.2022.41003

Abstract

XJTLU Research Centre for Language Technology intends to design an innovative application which could achieve to post edit the output of machine translation automatically, as well as improve the accuracy of language correction tools for English texts written by Chinese-English bilinguals. As a pilot study of this XJTLU's project, the present study intends to adopt the methodology of error analysis to explore whether there is significant difference concerning the L1 (or to say, source language)'s negative transfer in Chinese-English translation between machine translation and L2 students; and if there is, then find what is the distribution feature of the negative transfer. This is a corpus-based study: 40 MWUs are selected from 15 pieces of online news reports. Then, both qualitative and quantitative methods are employed. The empirical study is dominant and descriptive study serves as additional for the transfer types. The subjects are Youdao online translation and Bing online translation, as well as 10 senior students from Xi'an Jiaotong Liverpool University. After collecting data from the output of MT and translation tests, an English native speaker is invited to assist the author to identify negative lexical transfer. In addition, the relevant Chinese words' frequency is identified in the corpus of *Texts of Recent Chinese*. Finally, the study gets some useful findings: significant differences concerning negative transfer in terms of MWU do exist in the Chinese-English translation between machine translation and L2 students, and that can be ascribed to the respective advantages and disadvantages of MT and students in translating process. At last, some implications are drawn from this study for relevant researchers.

Keywords

Negative Transfer, Translation, MT, MWU, L2 Students

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

翻译的本质是不同语言之间的信息转换，是现代世界中企业、组织和国家之间的重要沟通工具。近年来，一种新型的翻译出现了。机器翻译(MT)，为有翻译需求的人提供了一个新的选择，与传统的通过专业翻译机构或自由译者的方式相比，似乎是一个节省时间和金钱的好方法。的确，对于不太重要的项目，人们不需要完美的翻译，只需要对内容有一个大致的了解，可以利用 MT。然而，实际上，对于重要的项目，人们普遍认为人工翻译是必要的，而且更可靠，更利于确保目标文本的高质量。毕竟，现阶段 MT 的译文输出不能避免一定数量的词汇和语法错误。事实上，上述错误不仅存在于 MT 的输出中，也存在于 L2 学生的写作和译文中。就像 MT 容易提供直译、逐字翻译、并且有明显的源文本痕迹一样，L2 学生的一些错误也含有源语言影响的痕迹，这使人们可以考虑语言迁移。语言迁移是一种学习活动，通过这种活动，学习者以前获得的知识会影响到后面的语言学习活动，可以归结为跨语言的影响。具体来说，在中国的语境中，中国英语学习者在写作和翻译中出现的所谓 Chinglish 就属于这种情况。当从中文翻译成英文时，学生会产生一些错误，例如语法错误和错误的句子结构，可能会影响到语义的传递，这与专业翻译不同。毕竟，那些专业译者往往是目标语言的母语者，他们完全熟悉目标语言的特点、语法、表达方式等，从而使他们能够提供最适当和准确的翻译。

因此，与专业译员或母语者的表达方式相比，无论是机器翻译的输出还是二语习得学生的写作或翻译，就目标语言本身的质量而言，都不是完美的。就错误本身而言，MT 和学生在某种程度上应该彼此有些共性。这些错误可能是语法错误、错误的结构或原始含义的丧失。事实上，有一些语言检查应用程序，如 Grammarly，可以纠正上述的一些错误。然而，这类应用程序无法准确定位或纠正由自动翻译工具或非母语人士产出的英语文本中的大多数问题。在意识到这一差距后，西交利物浦大学语言技术研究中心正试图设计一个新的应用程序，它可以实现自动修改机器翻译的译文，以及提高语言纠正工具对汉英双语者所写英语文本的准确性。

因此，作为西交利物浦项目的试点研究，本研究打算采用错误分析的方法，探讨机器翻译和 L2 学生在汉英翻译中关于 L1 (或者说，源语言)的负迁移是否存在显著差异？如果存在，那么负迁移的分布特征是什么？

此外，有两点需要提及：第一，正向迁移的情况也会被考虑在内。其次，考虑到之前的一些尝试是为了识别第一语言(L1)对目标语言的负迁移[1]，但以语法而非词汇为出发点，从而使基于词汇的高语言错误率没有得到解决，本研究只关注词汇迁移。此外，根据 Granger [2]的建议，词汇方法主要集中在搭配上，而不涉及其他语义特征，因此本项目的对象将限制在多词单位(MWU)上。

以下是本研究预计要解决的研究问题。

- 1) 就预选词而言，MT 中与 MWU 相关的词汇迁移率是否普遍高于学生的翻译？
- 2) 词性迁移率与嵌入词的频率之间是否有关联？
- 3) 就预选词而言，机器翻译中与 MWU 有关的负面词汇迁移率是否普遍高于学生的译文？
- 4) 与 MWU 有关的负面词汇迁移率与嵌入词的频率之间是否有关联？

2. 文献综述

2.1. 语言迁移在第二语言学习(SLA)中的地位

正如 Long 和 Richard 在 Odlin [3]撰写的《语言迁移》序言中指出的那样，语言迁移是应用语言学、SLA 和语言教学领域的一个关键问题。许多语言学家都认为，语言迁移可以成为 SLA 研究中的一个核心问题，尽管如此，并不是一切都很顺利。

2.1.1. 语言迁移的定义

语言迁移的概念起源于心理学。事实上，迁移是学习心理学中的一个重要概念，它被定义为一种学习活动，学习者以前获得的有关学习技能的知识将对他们以后的学习或训练活动的产出产生影响。有人认为，迁移既可以是积极的和消极的。

在定义方面，迁移这一术语产生了令人困惑的问题。由于迁移在传统上与行为主义有关，而行为主义将其定义为习惯形成的结果，意味着学习者在学习新语言时，其主要语言的消失，这在语言学家中引起了很大的反感和不满。许多语言学家认为，习惯形成不能充分说明母语的影响。同时，一些研究表明，迁移并不只是起到干扰的作用，还能带来优势和好处。此外，迁移也不仅仅是母语影响的问题[3]，因为其他一些以前学过的语言也会产生影响[4]。因此，仅仅使用“语言迁移”这个词是不够的。Corder [5]、Kellerman 和 Sharwood [6]建议放弃这个概念，或者只在相对有限的情况下利用它。他们主张用“跨语言影响”这个词来代替。他们说：“……‘跨语言影响’这一术语，在理论上是中立的，允许人们将‘迁移’、‘干扰’、‘回避’、‘借用’等相关方面等现象归入一个标题下，从而允许讨论这些现象之间的相似性和差异(1986, p.1)”。

即使如此，由于 Odlin 的《语言转移》[3]的巨大影响，“语言迁移”这个词如今仍被采用。在回顾

了关于这个问题的相关文献后, Odlin 对“迁移”是这样定义的:“迁移是指目标语言与之前(也许是不完全)习得的任何其他语言之间的异同所产生的影响。”考虑到缺乏一个完全适当的迁移定义, 本文仍采用“语言迁移”一词。然而, 这个词在这里是在广义上使用的, 而不是在狭义上的行为主义内涵。

2.1.2. 语言迁移的表现形式

在这一研究领域, 可以发现迁移的多种表现形式; 但是, 有四种表现形式经常被学者们讨论: 负面迁移、正向转移、回避和过度使用。

1) 错误(负面迁移)

错误, 也被称为“负迁移”, 发生在旧的学习习惯妨碍了学习新习惯的时候。在 SLA 领域, L1 的语法习惯会对 L2 的语法习惯的习得过程产生干扰, 这应该属于主动抑制。主动抑制的意思是, 之前学习或记忆的东西会影响到后来的学习。具体来说, L1 和 L2 之间的差异导致学习困难, 从而产生错误。

2) 促进作用(正向转移)

当 L1 和 L2 之间有相似之处时, 就会发生积极的转移, 这可以促进快速学习。然而, 促进并不意味着完全避免错误, 而只是相对减少错误的数量。

3) 回避

如果母语和目标语言之间存在着明显的差异, 学习者就会容易避免使用困难的语言结构。

例如, 一些中国学生会用英语写作时避免采用复杂的语法结构。根据 Kellerman [6]的说法, 回避的原因有三个: a) 学习者知道或预料到有问题, 并且至少对目标形式有一些模糊的概念; b) 学习者知道目标形式是什么, 但发现在特定情况下很难使用; c) 学习者知道要说什么和怎么说, 但不愿意真的说出来, 因为这将导致他们违反自己的行为规范。

4) 过度使用

在 SLA 中过度使用某些类型的语法形式是对语言内部过程过度概括的结果。例如, 常规的过去式变形可能会扩展到不规则动词; 回避也可能是另一个原因。日本的二语学习者由于避免使用定语从句等, 可能会过度使用简单的句子。

2.2. 与对比分析理论有关的语言迁移

对比分析(CA)是语言迁移的基本理论。对比分析模型是由 Fries [7]首次提出的。他认为学习者的母语在他们学习目标语言的过程中起着决定性的作用, 他们的母语的形式和结构会下意识地转移到二语习得中, 特别是对于那些新的二语学习者。Fries [7]出于对教学法的考虑, 主张对学习者的目标语言和母语进行详细描述, 具体如下:“最有效的材料是那些基于对目标语言的科学描述, 并与学习者的母语的平行描述进行仔细比较的材料。”

然而, 人们普遍认为, 1957 年 Lado 撰写的《跨越文化的语言学》一书的出版标志着 CA 理论的建立 [8]。Lado 受到 Fries 关于有效教材的假设的启发, 并在双语研究和测试研究中观察到, 他根据这种假设制定了对比分析(CA)理论:“……接触到外国语言的学生会发现其中的一些特点很容易, 而另一些则非常困难。那些与他的母语相似的元素对他来说是简单的, 而那些不同的元素则是困难的” [8]。

从那时起, 语言迁移在二语习得中就享有重要地位, 当时研究的重点是对母语和目标语言的结构进行对比分析。简而言之, CA 的主导思想是: 二语习得的主要障碍是母语结构对第二语言结构的干扰; 换句话说, 学习的困难应该归因于语言习惯的那些区别, 而二语习得被认为是一种克服这些困难的活动。通过关注母语和目标语言之间的差异, CA 希望能够预测在第二语言学习过程中可能出现的困难。

2.3. 误差分析

错误分析(EA)一直被用于语言学领域,尽管它在20世纪70年代后失去了一些人气[9]。正如Corder [5]所建议的,语言学家可以通过这些错误看到语言是如何习得的。此外,Dodigovic [10]所持有的观点,即学习者是一个有缺陷的二语,也通过对二语学习者口语输出中错误的检查得到支持。James [11]认为EA可以客观地描述学习者的interlanguage (IL)。IL是指第二语言学习者或正在学习目标语言的外语学习者所使用的语言类型。通过处理第二语言学习者的错误,可以达到三个目的:1) 确定学习者在语言学习中使用的策略;2) 确定学习者错误的原因;3) 获得关于语言学习中常见困难的信息。正如Corder [12]所建议的,进行错误分析通常有五个步骤:

- a) 收集学习者语言的样本
- b) 识别错误
- c) 描述错误
- d) 解释错误
- e) 对错误的评估

具体来说,首先,关于收集样本的过程,存在着自然样本和诱导样本的区别。前者意味着数据是由第二语言学习者自发产生的,可以为研究者提供更自然和真实的数据,但它很难获得。因此,大多数研究者选择采用诱导性样本[9]。说到数据收集方法,James [11]指出了两种不同的方式:一种是在一段时期内连续收集同一学习者群体的数据,另一种是在一个时间点上从不同人群中收集数据。许多研究者选择横断面的方法,因为它很方便。此外,就数据规模而言,有三种类型的样本:大量样本包括相当数量的二语学习者的语言输出,这个样本中的错误可以代表整个人群的表现;一个特定的样本包含来自有限数量的二语学习者的错误;一个偶然的样本只涉及一个学习者的错误。大多数SLA研究者倾向于使用后两种方式[9]。

在识别错误之后,可以采用语言策略或表面策略来描述这些错误[9]。语言学策略是由Dulay [13]和Krashen [14]提出的,将错误分为形态学、句法和词汇。在这个项目中,分析的范围是母语的词汇迁移。表面策略涉及四种结构:省略、增加、错误的形式和错误的顺序。而第五种是混合策略,由James [11]添加补充。

在描述了错误之后,需要对错误的原因进行调查。在评估了珍贵的错误来源类别后,James [11]提出了4个原因:母语影响,即语际错误;目标语言原因,即语内错误;基于交际策略的错误;以及诱发性错误。本研究将重点关注由母语迁移引起的语际错误。

错误分析的最后一步是错误评估,主要是指这些错误对被调查者造成的影响。正如Ellis [9]所言,有关错误评价的目的是为了改善语言教学法。“错误的严重性”这一术语表明了一个错误的严重程度。Khalil [15]提供了评价错误严重性的三个一般标准:可理解性(有错误的句子能被理解的程度),可接受性(错误的严重程度),以及刺激性(由于错误的数量而引起的人们的情绪反应)。

尽管错误分析在20世纪70年代后因其在范围和程序上的劣势而不受欢迎,但它从未被许多二语习得研究者放弃,以探索特定的研究问题[9]。举例来说,在认知心理学领域,有越来越多的研究加强了对第二语言形式、输入增强和错误纠正等方面的关注。特别是在本研究中,对汉语词汇转移的调查就是借鉴的EA的方法。

2.4. 机器翻译(MT)的发展

互联网为机器翻译(MT)提供了完美的平台,使其能够触及到普罗大众[16]。免费的在线MT始于1998年Altavista搜索引擎的Babelfish [17]。它远非完美,但使用起来很简单,而且往往比没有翻译更好。然

而,经过几年的发展,MT可能还没有准备好用于传播用途,也不会产生可出版质量的输出,但会提供有用的文本要点,甚至为用户提供足够的信息来执行手头的任务[16]。

与此同时,一种新型的自动化翻译进入了实践生产阶段:统计机器翻译(SMT)。语言翻译毕竟可以被认为是一种破译活动,而且,如果有足够的数据,源语言中的 n-gram 与目标语言中的 n-gram 相联系的概率,即使不能超越语言规则,那么至少也能发挥同样的作用。谷歌翻译在 2007 年将其引擎转向 SMT,标志着这种趋势。MT 已经证明了自己在社会上的作用,人们的兴趣和投入的资金也在继续增长[16]。本研究中的有道翻译和必应翻译都处于这种情况。

2.5. 多词单元

Sinclair [18]指出,单词容易系统地聚类。当集群的模式化相对有规律时,这些词就会建立起具有一定意义的词串。因此,这样的词串被称为多词单元(MWU) [19]。语言学家(例如 Alexander [20]; Moon [21])将 MWU 分为四种语言学分类:“短语动词”、“固定短语”、“习语”和“谚语”。从分析语言生产的角度来看,MWU 被认为是“公式化表达”、“词汇短语”或“词汇块”。

MWU 中的语言迁移

MWU 储存在长期记忆中,很容易被激活[22]。因此,它们是准确、流利和高效的语言产出的关键因素之一,在 SLA 领域占有重要地位[23] [24] [25]。此外,除了成语,所有的 MWU 在某种程度上都是构成性的[26]。也就是说,MWU 中每个孤立的的部分的意义都能对整个 MWU 的意义做出贡献。这种组合不是任意的,所以各部分的知识对整个 MWU 的习得起着有利的作用[27] [28]。MWU 也可以被认为是可迁移的,因为它们在语义和句法上都是模块化的[29]。许多学者(如 Rafee, Tavakoli & Amirian [30]; Adel & Erman [31]; Karabacak & Qin [32])意识到 MWU 在二语习得中的重要性,进行了相关研究,结果显示,与母语者相比,二语学习者使用的 MWU 往往较少,而且学习者很可能过度使用部分 MWU,从而对其他 MWU 使用不足。

但是,很少有研究涉及母语对第二语言 MWU 习得的影响。然而,Peromingo [33]进行了一项相关研究,通过分析几个英语学习者语料库中的论证性写作,考察母语对二语学习者产生的正确和不正确英语 MWU 的影响。

通过分析几个英语学习者语料库中的论证性写作。结果显示,二语学习者很容易过度使用那些与母语相似的 MWU。这项研究只是量化了二语学习者使用 MWU 的频率,且只发现两种语言的相似性与学习者使用某些 MWU 的频率之间有一定的联系,缺乏对二语 MWU 不正确使用的例子的展示和提供其根本原因。

此外,基于来自中国一所中外大学的 100 名不同学生的 100 篇写作的学习者语料库数据, Ma [34]的研究发现,在学习者语料库中发现的迁移错误数量从多到少是由中文多义词、MWU 和搭配引起的。此外,水平较低的学习者比水平较高的学习者更容易出现词性转换错误。然而,本研究并没有获得二语学习者在写作时打算写的确切含义,所以当有多个可能存在时,本研究不能确定词汇错误所对应的正确中文含义。换句话说,如果不确定与英语措辞相对应的中文含义,本研究中关于词汇迁移的解释似乎就是主观的。

与使用学生的写作作为研究数据相比,学生的翻译文本作为数据可以在很大程度上减少研究者的主观性。Luo [35]首次尝试研究翻译实践中的负迁移现象。在她的研究中,研究工具是交付给 54 名英语专业学生的翻译试卷。结果显示,受试者在潜在含义上犯的迁移错误最多。搭配使用中的迁移排在第二位,第三位是语义领域的迁移错误。然而,一个局限性是,本研究的样本只限于英语专业的学生。本研究的样本仅限于英语专业的学生。此外,在这项研究的基础上,Dong [36]进行了对英语专业和非英语专业学

生在汉英翻译中第一语言的负迁移进行了对比分析研究,旨在探讨英语专业和非英语专业学生在汉英翻译中 L1 的负迁移是否存在显著差异。研究表明,英语专业和非英语专业的学生在翻译错误方面确实存在明显差异,这在一定程度上可以归因于他们所采取的不同的翻译策略。这些发现对人与翻译机之间的翻译活动中源语言负迁移的对比分析有一定的启发。

总的来说,现有的文献还不能全面描述机器翻译(MT)输出中的 MWU 使用或错误,也没有澄清 MT 输出和学习第二语言的学生的输出在这方面是否有任何相似之处。然而,Peromingo [33]和 Paquot [37]的研究确定了一般翻译过程和二语写作在 MWU 生成方面的惊人相似性,Dong [36]的研究提供了研究语言迁移的对比性方法,表明了人类和翻译机器之间的翻译活动中对源语言的负面迁移进行对比性分析的可能性。因此,若想要深入研究,便要探索 MT 的英文输出中中文 MWU 的词性迁移证据。

3. 研究方法

3.1. 使用的中文和英文语料库

本项目是一个基于语料库的研究。本项目所使用的中文语料库是由从中国的三个政府网站中选出的总共 15 篇文章(共 61,804 个字)组成。这三个网站分别是中华人民共和国商务部(www.mofcom.gov.cn)、中华人民共和国外交部(www.fmprc.gov.cn)、中国国家统计局(www.stats.gov.cn)。选择政府网站的原因是:与其他类型的网站相比,这类网站的语言更加正式和主流。这些文章的体裁和主题没有区别:都是新闻报道。

英语语料库由有道翻译的输出形成,有道翻译是中国常用的翻译程序,是基于网易开发的互联网语料库和搜索引擎。源文本是上述的 15 篇中文文章。这个英语语料库的总字数为 38,438。

3.2. 步骤

在研究开始时,研究者首先仔细阅读了中文语料库,旨在挑选出一些有代表性的中文 MWU 表达,然后在各自的文件中进行标注。在这个阶段,研究者的选择是基于过去的个人经验和阅读同一体裁的中文文章的知识,这有助于判断什么样的 MWU 表达是典型的,因而值得研究。基本上,选择 MWU 的整个过程是一种随机的行为,而在翻阅了整个语料库之后,有十个中文动词在这个阶段被认为是特别值得关注的,以便使这个项目具有方向性和针对性。这些词是“达成(da cheng)”,“扩大(kuo da)”,“加强(jia qiang)”,“联系(lian xi)”,“占据(zhan ju)”,“带来(dai lai)”,“恢复(hui fu)”,“表示(biao shi)”,“促进(cu jin)”和“推动(tui dong)”。之所以选择这些词,是因为这些词在语料库中出现的次数都超过了两次,所以这些例子更具有典型性,可以增强研究的针对性。

根据中文语料库中的标记结果,研究人员又查阅了分析器建立的英文语料库,找出有道翻译产出的相应表达,对其进行标记,并将其输入 Excel 电子表格,供后期评估和比较。此外,为了使机器翻译的输出结果更具代表性和普遍性,我们采用了第二台翻译工具——Bing 网站翻译,它是由微软开发的,采用了与有道类似的翻译技术。研究者用它对拥有这些中文 MWU 的源句子进行了重新翻译,翻译结果也被输入 Excel 电子表格。

在项目的第三阶段,研究者准备了翻译测试,并邀请 10 名学生参加测试。这些测试要求学生用英语来表达所选的中文 MWU 的含义。受访者被告知,他们不需要逐字翻译,也不允许使用任何字典或翻译工具;他们需要做的只是尽力通过英语来保持相应的意思。受访者是西交利物浦大学(XJTLU)的 10 名四年级学生,这是一所由利物浦大学和西安交通大学在中国联合开办的中外大学。他们都是来自英语文化与传播系(ECC)的中文母语者。结果也被录入 Excel 电子表格。

在第四阶段,该研究旨在评估上述收集到的所有结果。考虑到研究者无法确认输出的某些英语

MWUs 表达的适当性, 所以使用了厦门大学的汉英平行语料库帮助寻找参考英语表达。研究结果被录入 Excel 电子表格。之后, 我们邀请了一位来自西交利物浦大学 ECC 系的英语母语者教师对 Excel 表格中的每一个结果进行评估和评分。“1”代表正确, “0”代表不正确。最后, 研究人员分别计算了学生结果和翻译机结果的正确率。正如 Biber、Conrad 和 Reppen [38]所建议的, 为了使学生和翻译机输出的 MWU 表达的正确率具有可比性, 有必要将原始率“规范”到一定的单词基础上。本研究采用了以下公式:

$$\text{正确率} = (\text{正确结果的数量} / \text{总结果的数量}) * 100\%。$$

在最后阶段, 研究人员通过 SUBTLEX-CH-WF (可从网站 www.ugent.be/pp/experimentele-psychologie/en/research/documents/subtlexch/subtlexchwf.zip)来识别每个中文词的频率。以帮助发现在第三阶段计算的准确率与上述频率之间的相关性。

4. 研究结果

本部分包含几个图表, 客观地介绍了本研究中数据分析的统计结果和发现。在中文语料库中, 共有 107 个 MWU 被研究者标记。其中, 根据方法论部分提到的 10 个预选词, 研究者共分离出 40 个涵盖这些词的 MWU, 并将这些 MWU 分为 10 组。每组中, MWU 的数量不超过 5 个, 也不低于 3 个。

为了回答第一个研究问题: “就预选词而言, MT 的词汇迁移率是否普遍高于学生的翻译?”, 我们分别统计了 MT 和学生产生的词汇迁移的数量, 然后计算它们的比率。从表 1 可以看出, 在某些词上, MT 的迁移率较高, 而在其他词上, 学生的迁移率较高。在这个方面, 没有人占据主导地位。然而, 如下表 1 所示, 平均而言, MT 的词汇迁移率普遍高于学生的译文。

Table 1. The rate of lexical transfer of the 10 selected words

表 1. 10 个预选词的词汇迁移率

Group of MWUs	The rate of lexical transfer	
	student	MT
达成...	50%	37.50%
扩大...	96%	80%
加强...	75%	100%
联系...	45%	62.50%
占据...	40.00%	50%
带来...	80%	90%
恢复...	53.30%	100.00%
表示...	35%	12.50%
促进...	50%	100.00%
推动...	55%	75.00%

SUBTLEX-CH-WF 是建立在最近的中文文本语料库上的, 在检查了 SUBTLEX-CH-WF 的数据后, 发现了每个中文词的排序(见表 2):

Table 2. The rank of word frequency in the corpus of Texts of Recent Chinese
表 2. Texts of Recent Chinese 语料库中的词频排序

Word	达成	扩大	加强	联系	占据	带来	恢复	表示	促进	推动
Frequency rank	2630	6207	5589	561	6331	706	1303	1556	12025	9604
Words/million	26.53	7.99	9.36	179.48	7.78	133.9	65.79	52.49	2.74	4.02

那么，第二个研究问题“词汇迁移率和嵌入词的频率之间是否有关联？”主要可以通过绘制一个将词频和词汇迁移率联系起来的表格来回答：

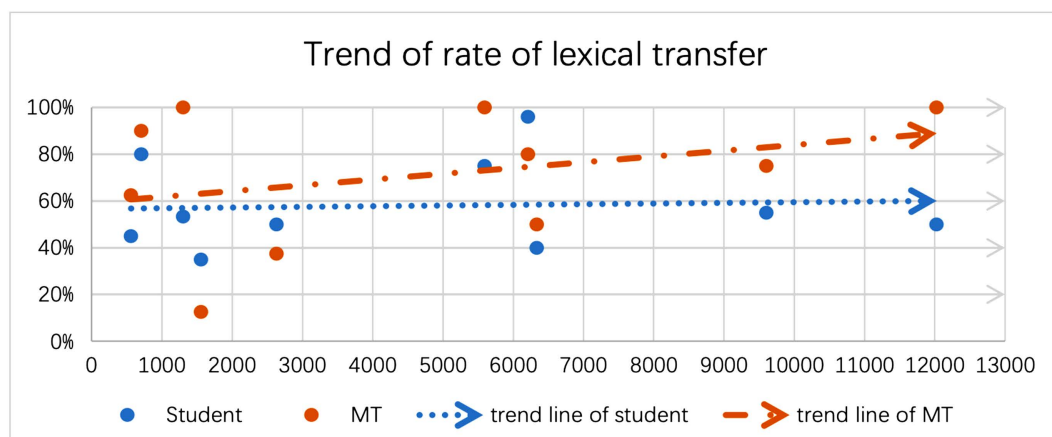


Figure 1. The distribution feature of the rate of lexical transfer regarding to word frequency
图 1. 根据词频分类的词汇转移率的分布特征

从图 1 可以看出，无论是学生还是 MT 的表现，如果人们只是单独关注每个词，那么词频和词性迁移率之间似乎没有什么强关系。然而，当分析者把数据作为一个整体来考虑时，这里就有一种直线的相关关系了。也就是说，随着词频的增加，MT 的词汇迁移率呈上升趋势；而学生的情况则没有明显变化。换句话说，随着词频的减少，MT 产生了更多的词性迁移。总的来说，词频和词汇迁移率之间的预期模式存在于 MT 而非学生组。

针对第三个研究问题“就预选词而言，MT 的负面词汇迁移率是否普遍高于学生的翻译？”，我们分别统计了 MT 和学生产生的负面词汇迁移实例的数量，然后计算了它们的比率。如表 3 所示，令人惊讶的是，结果是反直觉的：是学生组在每个词上的负面词汇迁移率较高。

Table 3. The rate of negative lexical transfer of the 10 selected words
表 3. 10 个预选词的词汇负迁移率

Group of MWUs	The rate of negative lexical transfer	
	student	MT
达成...	10%	0
扩大...	20.83%	12.50%
加强...	27%	0
联系...	10.00%	0

Continued

占据…	16.70%	0
带来…	20%	11.10%
恢复…	25%	0.00%
表示…	14.30%	0
促进…	40%	25.00%
推动…	18%	17%

同样，参照汉语语料库中的词频排序，第四个研究问题“负面词汇迁移率与嵌入词的词频之间是否存在关联？”可以通过绘制词频与负面词汇迁移率的关联图来回答。

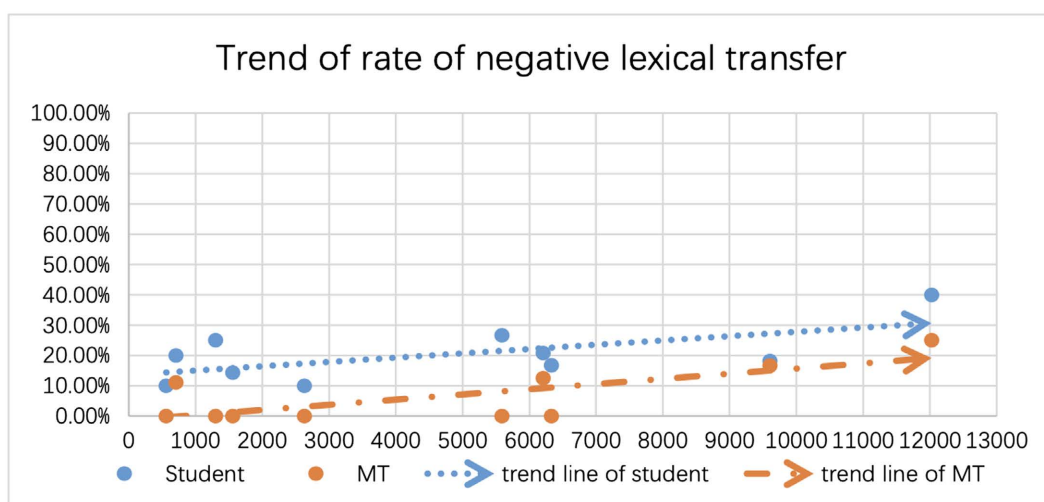


Figure 2. The distribution feature of the rate of negative lexical transfer regarding to word frequency

图2. 根据词频分类的词汇负迁移率的分布特征

从图2中可以看出，无论是学生还是MT，随着词频等级的增加，负面的词汇迁移率都有下降的趋势。换句话说，在嵌入频率较低的词的MWU中发现了相对较多的词汇迁移错误，这不仅适用于学生，也适用于MT。此外，对于每个词，学生的比率都比MT高。换句话说，一般来说，学生比MT更容易产生词性迁移错误。

5. 讨论

5.1. 由MT产出的词汇迁移

根据研究结果，在10个组别中，第(3)、(4)、(5)、(6)、(7)、(9)和(10)组的MT迁移率要高于学生。这些中文词语分别是“加强、联系、占据、带来、恢复、促进和推动”。本节将以第4组和第9组为典型例子进行研究。

对于第4组，MT和学生之间的主要差距集中在MWU“联系在一起”的表达上。这个MWU的原文是“改革已成为中墨两国的共同追求，也将两国更紧密地联系在一起”（“The reform has been the common pursuit of China and Mexico, which also forms a bond between the two nations”）。有道的译文将其直译为“联系在一起”，其中“link”对应于“联系”，“together”对应于“在一起”。然而，有两个学生把它翻

译成“bring together”。根据《剑桥词典在线》，短语动词“bring together”的意思是“使人们彼此友好”，这符合源文本的语境，尽管它在字典中的似乎与“联系在一起”无关。事实上，这种差距的形成发生在翻译活动的第二和第三步。根据 Vinay 和 Darbelnet [39]的说法，第二步是“检查源文本，评估各单元的描述性、情感性和知识性内容”，第三步是“重构信息的元语言学背景”。在这个案例中，当这两个学生在检查源句时，他们应该发现文本所谈论的是两个国家的关系。因此，他们对信息进行了重构，以接近核心含义，然后利用短语动词“bring together”来表达这个含义。从这一点上看，人类的优势在于人的大脑比机器更善于理解和重构事物。

对于第 9 组，MT 的词汇迁移率达到 100%，而学生的迁移率只有一半。MT 和学生之间的主要差异集中在 MWU 促进就业(“promote employment”)的表达上。有道和必应都采用直译“促进就业”，而两个学生则用“创造就业机会”来转述中文意思。事实上，“创造就业机会”是“促进就业”的具体体现。学生们在这里改变了源语言的语义和观点，使用了一种叫做“modulation”的翻译策略[39]。具体来说，他们把抽象的事物翻译成具体的事物，这就避免了词汇的迁移。同样的策略也被学生们用在第 10 组。在那里，通过 modulation 的策略，推动改善关系(promote the improvement of relations)被翻译成“采取措施改善关系‘和’在关系上做出努力”。同时，其他三名学生在意识到这里的核心动词是“促进”之后，省略了“促进”一词，将其翻译为“改善关系”，而不是“推动”。

5.2. 由学生产出的词汇迁移

根据研究结果，在 10 个组别中，(1)、(2)和(8)组的学生迁移率比 MT 高，分别为：达成、扩大和表示。本节将以(1)和(8)组为例进行讨论。

就第(1)组而言，MT 和学生之间的主要差距集中在 MWU “达成一致”和“达成交易”的表达上。对于“达成一致”，三名学生将其直译为“reach an agreement”；相反，两位 MT 将其翻译为“agree on”。根据 Vinay 和 Darbelnet [39]的观点，这种差距是在翻译过程的第一步形成的：即“确定翻译单位”。在他们的理论中，“翻译单位”应该是“词汇学单位”和“观点单位”的组合，“观点单位”被定义为“话语中最小的片段，其联系方式使它们不应该被单独翻译”。在这种情况下，学生将“达成”和“一致”分别作为两个翻译单位，而 MT 则将“达成一致”作为一个完整的翻译单位。

对于“达成交易”，学生们继续使用直译的策略，而 MT 则采用另一个动词：“have”，来代替“reach”一词。在英语中，“have”是一个高频率的词。它非常常见和灵活，但非常复杂。它可以作为一个辅助动词，也可以作为一个普通动词。它可以和许多不同的名词跟在一起形成搭配，如 have a swim, have a seat, have a rest, have a chat, have a talk, have a wash, have a read, have a drink, have a debate, have a say, 其中“have”这个词失去了自己的词汇意义，而其短语的意义则与搭配词密切相关。因此，由 MT 产生的“have a deal”可以归入这种用法。这至少表明，MT 已经掌握了该词的用法，而学生们是否掌握了这一语言点，还不得而知。

另一方面，它表明在机器翻译的内部，“达成交易”被视为一个 MWU，而学生仍将“达成”和“交易”视为两个独立的单位。对于第 8 组，MT 和学生之间的主要差距集中在 MWU “表示赞同”和“表示反对”的表达上。对于“表示赞同”，有四个学生避免了直接使用英语单词“show”。其中，一名学生使用了另一个词“present”，另外三名学生采用了自由翻译的策略，省略了与“表示”相对应的词，只使用“agree”一词来表达意思，这似乎与上面讨论的 MT 处理“达成一致”的策略相似。事实上，根据 Lv [40]的分类，“表示”这个词属于“dummy verb”，也就是说，其词语搭配的核心意义只是表达的后半部分。换句话说，就“表示赞同”和“表示反对”而言，后半部分“赞同”和“反对”可以单独发挥动词的作用而不失去原有意义。因此，在翻译这些 MWU 时，省略“表示”一词是一个不错的策略。从

这一点来看,大多数学生都能掌握这些 MWU 的基本语义。然而,机器翻译的表现甚至比学生更好:它们也采用了这样的调控策略,尽管仍有三名学生发生了词汇迁移,但是没有进行负面的词汇迁移。

5.3. 学生产出的负迁移

根据研究结果,所有 10 个小组的学生产出的负面迁移率都高于 MT。本部分将(6)、(7)和(9)组作为典型例子。

就第 6 组而言,五个学生中有四个将“带来严峻挑战”字面翻译为“bring challenges”,而根据母语为英文的人士的评价,在英语语境中没有这样的表达。换句话说,这是一个消极的词汇迁移。一个合适的表达是“pose serious challenges”。作为一个单独的词,“pose”在中文中的意思是造成。因此,“pose serious challenges”的表达方式似乎对应于“造成严峻挑战”,这在汉语中是不可接受的。这可能是没有学生意识到“pose”这个词可以在这个表达中使用的原因。另一个可能的解释是,当学生习得“pose”这个词时,他们只是学习了这个词作为一个不及物动词的意义(“to move into and stay in a particular position, in order to be photographed or painted”),而忽略了它作为一个及物动词的意义,它可以与“challenge”一词搭配。同时,MT 的翻译是“pose”,这表明机器翻译把 MWU “带来的严峻挑战”作为了一个整体。

同样,在第(7)组中,“recover market's confidence”是“恢复市场信心”的负向迁移。“Recover”应该被修改为“regain”,而这两个词都有“to get back something that the subject no longer have”的意思,在英汉词典中都有“恢复”的解释。这两个词的区别在于,“recover”指的是“stop suffering from an illness or injury”,而“regain”则与“get something abstract back such as an ability, quality or position”有关。在这里,“confidence”属于一种抽象的东西,所以“regain”在这个语境中应该是合适的。这个错误的出现,说明学生在学习这个词时,没有学会在什么样的情况下可以正确使用“recover”。此外,可能在他们在思考时找不到其他的词来表达“恢复”的意思。毕竟,在《高考大纲》的 3500 个必考词中,“recover”是唯一一个有“恢复”的中文解释的词。

在第(3)组中,加强作风建设(“strengthen the construction of work style”)是另一个学生容易产生错误的 MWU,三位学生在表达中没有使用“strengthen”,而是采用了“enhance”,这被评估为是一种负面的迁移。这两个词在英汉双解词典中都有“加强”的对应含义(见表 4)。

Table 4. Dictionary explanations of the words “strengthen”, “enhance” and “reinforce”

表 4. “strengthen”、“enhance”和“reinforce”的词典释义

Word	Core meaning	Attach meaning 1	Attach meaning 2
Strengthen	加强	变坚固	
Enhance	提高	加强	增加
Reinforce	加强	变坚固	

《牛津学习者辞典》写道:“strengthen”是“the most general of these words and can mean to make sb/sth stronger in any way: physically, emotionally, morally, politically or financially”。另外,这个词也在高考大纲的 3500 个必考词中。从这一点来看,当学生想表达“becoming stronger”的意思时,“strengthen”应该是第一选择。至于“enhance”,一方面,这个词的核心意思是“增加(提高)”,而不是“strengthen”;另一方面,虽然它与汉语中的“加强”相对应,但它只是侧重于提高事物的数量或价值。典型的搭配是“enhance the operation effectiveness”,“enhance the value of land”,“enhance friendship and unity”和

“enhance beauty”。在本例中，下面的名词短语是“the construction of work style”，属于物理层面的，所以不适合与“enhance”搭配。通过这个 MWU，我们可以推断出，学生们很容易使用一些超出低级阶段所学的词汇来显示他们的“高”水平，而这些词汇却根本没有被真正掌握。也就是说，他们可能分不清哪个含义是这个词在英语中的核心意义，他们不知道这个词在什么具体的语境中使用是可行的。因此，在这种情况下就产生了负面的词汇迁移。Cui [42]建议教师在教学内容中加强对词汇核心意义的讲解，其次是最常见的附加意义，这样可以帮助减少此类错误。

5.4. 总结：MT 和学生在翻译过程中各自的优势和劣势

通过以上的讨论，分析者总结了 MT 和学生各自的优势和劣势(见图 3 和图 4)。对于 MT 来说，由于其强大的存储能力，它比学生有更多的 MWU 表达，因此它可以更直接地输出。而且，由于它的存储量非常准确，一旦机器翻译在其中搜索到这种表达，输出的结果就很难出现错误。如果这种表达方式是以前没有存储过的临时搭配，那么机器翻译就必须采用翻译的方法。由于它不能真正理解该表达的含义，所以意译的策略就很难利用：那么唯一的办法就是使用直译。在面对词汇的选择时，MT 的表现比学生好。原因有二：第一，机器的词汇量是学生的数倍，所以它有更多的候选词可以选择；第二，当从一系列的同义词中选择时，统计机器翻译(SMT)由于其一定的机制，通常能够选择相对可被接受的一个。在 SMT 中，语言翻译可以被认为是一种破译活动，在给定足够数据的情况下，源语言中的 n-gram 被链接到目标语言中的 n-gram 的概率较高。

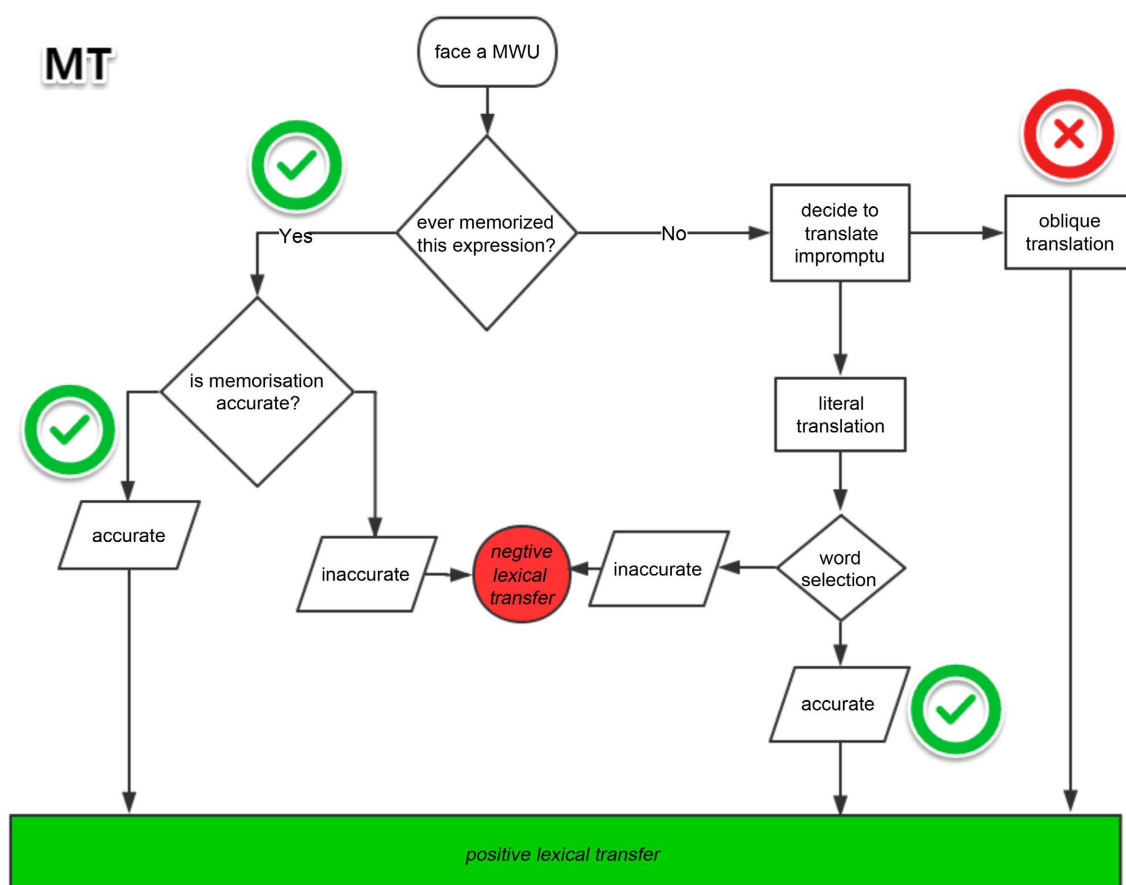


Figure 3. The advantages and disadvantage of MT in translating process

图 3. 机器翻译过程中的优缺点

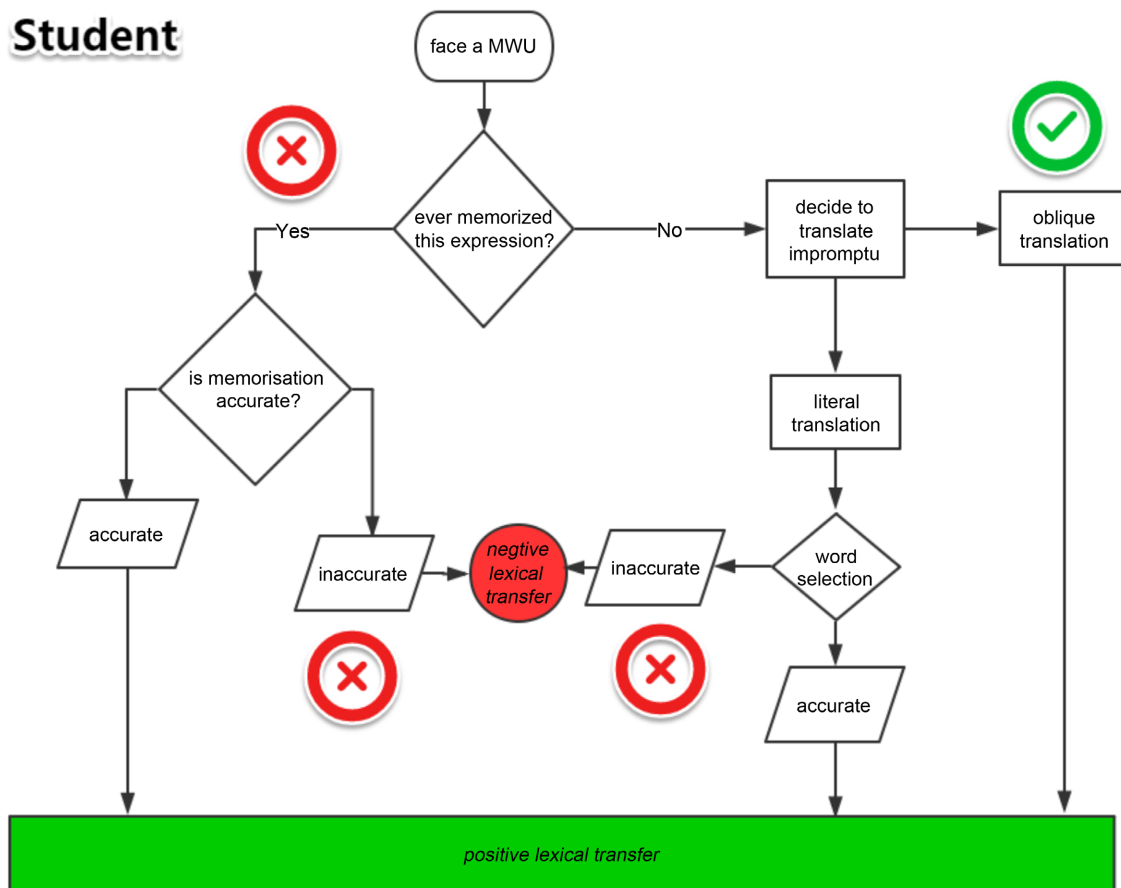


Figure 4. The advantage and disadvantages of students in translating process

图 4. 学生在翻译过程中的优缺点

对于学生来说，唯一的优势体现在意象翻译的活动上，具体来说就是调控翻译。看起来学生翻译的优势数量比机器翻译少，然而，正如 Vinay 和 Darbelnet [39]所说，“调制是一个好译者的试金石”。调制策略的运用表明了译者真正理解了源文本的语义。在“联系在一起”的例子中，在通过参考上下文理解了源表达的语义后，学生使用了另一个短语动词“bring together”，而不是直译的“link together”，这就避免了词汇迁移。另一个例子是“促进就业”，学生能够把它翻译成“创造更多的就业机会”，把表达方式从抽象的变为具体的，这就避免了词性迁移，尽管 MT 的直译“促进就业”也是正确的。

虽然有这样的优势，但学生在记忆的准确性和选词方面还是比较弱。例如，在第(2)组中，当面对“扩大试点”时，其中一个学生错误地使用了“open up”这个短语来表达“扩大”的意思，而不是像其他学生那样使用“expand”或“enlarge”这样的单一词汇。在这里，可以推断出该学生将“扩大”视为两个独立的单位，而不是一个完整的单位，因为“扩大”一词在汉语中对应于“扩”，而“up”和“大”在某种程度上具有相似的语义。然而，事实上，“open up”在英语中的意思是“使某物更容易接近或变得不封闭”，所以扩大和“open up”之间根本就没有对应关系。这个错误的产生应该归咎于记忆的错误。具体来说，这个错误可以用“语义饱和效应”的理论来解释。也就是说，当长时间看一个词时，人们在语义提取活动中会有暂时性的困难[42]。这种情况不会发生在机器上，因此，这个特点是人类和机器翻译之间的区别之一。

到目前为止的讨论可以解释为什么 MT 的词汇迁移率和负面迁移转移率都普遍低于学生的水平。具体来说，MT 在 MWU 记忆的大小、记忆的准确性和选词方面比学生表现得好；而学生只是在意象翻译方面做得不错。在降低词性迁移率和错误率方面，MT 比学生有更多的办法，所以如表所示，MT 的词性

迁移率比学生低。

此外，对于第四个研究问题“负面词汇迁移率与嵌入词的频率之间是否有关联？”，结果显示，在嵌入频率较低的词汇的 MWU 中，发现的词汇迁移错误相对较多，这也可以在图 5 中得到解释。

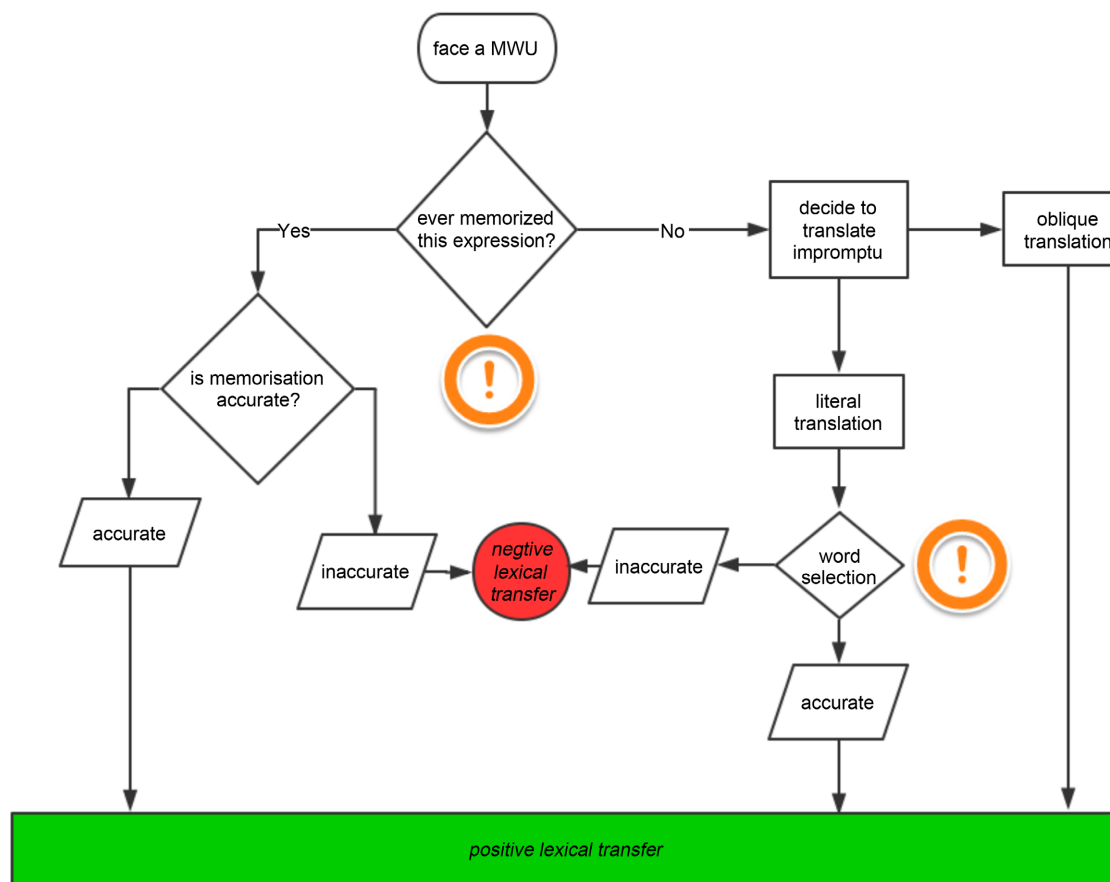


Figure 5. The influenced procedures due to the decreasing of word's frequency

图 5. 受词频降低影响的翻译流程

如图 5 所示，随着词频的降低，译文输出的两个过程受到影响：第一个过程是记忆表达；第二个过程是选词。嵌入高频词的 MWU 很容易成为搭配，如保持联系(keep in touch with)和达成交易(strike a deal)。而嵌入低频词的 MWU 通常是临时组合，如促进人才流动(stimulate the circulation of talents)和推动向前发展(promote the development)。对于前一种类型，机器翻译有机会提前存储这些搭配，可以保证未来输出的准确率；而对于后一种类型，则必须承担迁移错误的风险。就选词过程而言，如上所述，统计机器翻译(SMT)的机制决定了给出的数据越多，输出就越准确。对于低频词，机器可以初步收集的相关数据相对有限。因此，出错的概率会更大。

6. 结论与启示

6.1. 结论

基于文献中的迁移理论和实证研究结果，本研究旨在探讨机器翻译和学习二语的学生在中英翻译中有关这一问题的差异。通过对自建英语语料库中标记的 MWU 以及学生在翻译测试中的答案进行错误分析，作者得出了一些结论：1) MT 输出中的词汇迁移率普遍高于学生的翻译。2) 随着词频的降低，MT

译文中有更多的词性迁移, 而学生的表现则不受词频的影响。3) MT 输出中的负迁移率一般低于学生的翻译。4) 随着词频的降低, MT 和学生都容易产生更多的负迁移。本研究的目的是帮助西交利物浦大学语言技术研究中心的研究人员设计一个应用程序, 以实现机器翻译的输出进行自动编辑, 并提高中英双语者撰写的英语文章的语言纠正工具的准确性。

根据讨论部分的建议, 有三个原因导致 MT 出现转移错误: 1) MT 不善于理解语义, 因此它不能像人类那样在遇到临时性的词汇组合时利用意译的策略; 2) MT 依赖于对某些 MWU 的初步存储记忆, 所以在面对不熟悉的表达时, 它会陷入困境; 3) MT 选择目标词的准确性在很大程度上取决于初步给定数据的数量是否足够, 然而, 随着词频的降低, 相关输入数据的数量也在减少, 因此, 随着嵌入 MWU 的词频率降低, 出错的概率也在增加。

6.2. 启示

在分析了这三个原因后, 笔者认为 MT 的弱点主要集中在那些临时性的词语组合上。因此, 一方面, 对应用程序的设计者来说, 本研究的启发是: 1) 当中文源文本出现临时词语组合时, 要更加关注英文输出。更具体地说, 通常那些低频动词的 MWU 容易出现临时组合。2) 为了纠正所发现的错误, 应用程序应该比机器翻译拥有更多的原始数据, 这样它就能在处理单词的选择上有更大的能力, 特别是在那些低频动词方面。另一方面, 至于纠正学生错误的功能, 其启示为: 1) 除非是逐字翻译, 否则应用程序不需要太关注低频动词的临时中文 MWU 所对应的英文输出; 2) 由于其较大的数据库, MT 自动翻译后人工编辑功能的资源可以直接借用到应用纠正学生错误的功能上。

参考文献

- [1] Brooke, J. and Hirst, G. (2012) Robust, Lexicalized Native Language Identification. *Proceedings of COLING 2012*, Mumbai, 8-15 December 2012, 391-408.
- [2] Granger, S. (2014) *Learner English on Computer*. Routledge, London. <https://doi.org/10.4324/9781315841342>
- [3] Odlin, T. (1989) *Language Transfer* (Vol. 27). Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9781139524537>
- [4] Ellis, R. (1994) *The Study of Second Language Acquisition*. Shanghai Foreign Language Education Press, Shanghai.
- [5] Corder, S.P. (1992) A Role for the Mother Tongue. In: Gass, S. and Selinker, L., Eds., *Language Transfer in Language Learning*, John Benjamins, Amsterdam, 19-25. <https://doi.org/10.1075/lald.5.04cor>
- [6] Kellerman, E. and Sharwood, M.S. (1986) *Cross-Linguistic Influence in Second Language Acquisition*. Pergamon Press, Oxford.
- [7] Fries, C. (1945) *Teaching and Learning English as a Foreign Language*. Ann Arbor University of Michigan Press, Ann Arbor.
- [8] Lado, R. (1957) *Contrastive Analysis: An Overview*.
- [9] Ellis, R. (2008) *The Study of Second Language Acquisition*. 2nd Edition, Oxford University Press, Oxford.
- [10] Dodigovic, M. (2005) *Artificial Intelligence in Second Language Learning*. Multilingual Matters, Clevedon. <https://doi.org/10.21832/9781853598319>
- [11] James, C. (1998) *Errors in Language Learning and Use: Exploring Error Analysis*. Routledge, London.
- [12] Corder, S. (1974) *Error Analysis and Remedial Teaching*.
- [13] Dulay, H. (1982) *Language Two*. Oxford University Press, New York.
- [14] Krashen, S. (1982) *Principles and Practice in Second Language Acquisition*. Pergamon Press, Oxford. <https://doi.org/10.1111/j.1467-971X.1982.tb00476.x>
- [15] Khalil, A. (1985) Communicative Error Evaluation: Native Speakers' Evaluation and Interpretation of Written Errors of Arab EFL Learners. *TESOL Quarterly*, 19, 335-351. <https://doi.org/10.2307/3586833>
- [16] Garcia, I. (2012) Machines, Translations and Memories: Language Transfer in the Web Browser. *Perspectives: Studies in Translatology*, 20, 451-461. <https://doi.org/10.1080/0907676X.2012.726230>

-
- [17] Yan, J. and Lange, L. (2003) "Going Live on the Internet", *Computers and Translation: A Translator's Guide*. John Benjamins, Amsterdam. <https://doi.org/10.1075/btl.35.15yan>
- [18] Sinclair, J. (1991) *Corpus, Concordance, Collocation*. Oxford University Press, Oxford.
- [19] Schmitt, N. (2000) *Vocabulary in Language Teaching*. Ernst Klett Sprachen, Stuttgart.
- [20] Alexander, R.J. (1984) *Primary Teaching*. Holt, Rinehart and Winston, New York.
- [21] Moon, J. (2001) PDP Working Paper 4: Reflection in Higher Education Learning. Higher Education Academy, York, 1-25.
- [22] Pawley, A. and Syder, F.H. (1983) Two Puzzles for Linguistic Theory: Nativelike Selection and Nativelike Fluency. In: Richards, J. and Schmidt, R., Eds., *Language and Communication*, Longman, London, 191-225.
- [23] Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E. and Quirk, R. (1999) *Longman Grammar of Spoken and Written English (Vol. 2)*. MIT Press, Cambridge.
- [24] Erman, B. and Warren, B. (2000) The Idiom Principle and the Open Choice Principle. *Text & Talk*, **20**, 29-62. <https://doi.org/10.1515/text.1.2000.20.1.29>
- [25] Hyland, K. (2008) As Can Be Seen: Lexical Bundles and Disciplinary Variation. *English for Specific Purposes*, **27**, 4-21. <https://doi.org/10.1016/j.esp.2007.06.001>
- [26] Nation, I.S. (2013) *Teaching & Learning Vocabulary*. Heinle Cengage Learning, Boston. <https://doi.org/10.1017/CBO9781139858656>
- [27] Bogaards, P. (2001) Lexical Units and the Learning of Foreign Language Vocabulary. *Studies in Second Language Acquisition*, **23**, 321-343. <https://doi.org/10.1017/S0272263101003011>
- [28] Boers, F. and Lindstromberg, S. (2009) *Optimizing a Lexical Approach to Instructed Second Language Acquisition*. Palgrave Macmillan, London. <https://doi.org/10.1057/9780230245006>
- [29] Kellerman, E. (1978) Giving Learners a Break: Native Language Intuitions as a Source of Predictions about Transferability. *Working Papers on Bilingualism Toronto*, No. 15, 59-92.
- [30] Rafiee, M., Tavakoli, M. and Amirian, Z. (2011) Structural Analysis of Lexical Bundles across Two Types of English Newspapers Edited by Native and Non-Native Speakers. *Modern Journal of Applied Linguistics*, **3**, 218-236.
- [31] Ädel, A. and Erman, B. (2012) Recurrent Word Combinations in Academic Writing by Native and Non-Native Speakers of English: A Lexical Bundles Approach. *English for Specific Purposes*, **31**, 81-92. <https://doi.org/10.1016/j.esp.2011.08.004>
- [32] Karabacak, E. and Qin, J. (2013) Comparison of Lexical Bundles Used by Turkish, Chinese, and American University Students. *Procedia-Social and Behavioral Sciences*, **70**, 622-628. <https://doi.org/10.1016/j.sbspro.2013.01.101>
- [33] Peromingo, J.P.R. (2012) Corpus Analysis and Phraseology: Transfer of Multi-Word Units. *Linguistics and the Human Sciences*, **6**, 321-343. <https://doi.org/10.1558/lhs.v6i1-3.321>
- [34] Ma, C.C. (2016) *Lexical Transfer from Chinese to English in the Written Production of XJTLU Students*. Project Report.
- [35] Luo, H.X. (2007) *The Negative Transfer of Mother Tongue in Chinese-English Translation: From the Perspective of Lexicon*. Project Report.
- [36] Dong, J. (2010) *A Contrastive Analysis of L1's Negative Transfer in the Chinese-English Translation between English and Non-English Majors*. Project Report.
- [37] Paquot, M. (2013) Lexical Bundles and L1 Transfer Effects. *International Journal of Corpus Linguistics*, **18**, 391-417. <https://doi.org/10.1075/ijcl.18.3.06paq>
- [38] Biber, D., Conrad, S. and Reppen, R. (1998) *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9780511804489>
- [39] Vinay, J.P. and Darbelnet, J. (2000) A Methodology for Translation. In: Venuti, L., Ed., *The Translation Studies Reader*, Routledge, London, 84-93.
- [40] Lv, Y. (2014) The Professional Development of the Foreign Language Teachers and the Professional Foreign Language Teaching Practice. *Theory & Practice in Language Studies*, **4**, 1439-1444. <https://doi.org/10.4304/tpls.4.7.1439-1444>
- [41] Cui, L. (2012) *Learning to Teach in the Beginning Years: Contradiction Analysis of a Secondary English Teacher in China*.
- [42] Lambert, W.E. and Jakobovits, L.A. (1960) Verbal Satiation and Changes in the Intensity of Meaning. *Journal of Experimental Psychology*, **60**, 376-383. <https://doi.org/10.1037/h0045624>