

面向机器翻译的汉英小句复合体转换生成能力调查

邢富坤
浙江外国语学院/ 浙江杭州
xingfukun@126.com

徐佳宁
青岛大学/ 山东青岛
461096572@qq.com

摘要

小句复合体由小句组合而成，不同语言在小句的组合模式上存在差异，该差异对机器翻译有何影响尚不清楚。本文以汉英机器翻译为例，选取多语体的汉语小句复合体及专家译文，从话头共享关系和共享类型两方面对主流机器翻译系统以及ChatGPT开展调查。结果显示，与专家译文相比，机器翻译的小句复合体转换生成能力存在较大不足，表现为机器翻译在话头补足、转换、提炼等方面的能力较弱，小句组合模式单一且带有明显的汉语原文痕迹，译文的地道性受到较大影响。

关键词： 小句复合体；转换生成能力；机器翻译；ChatGPT

Investigation of the Clause Complexes Transfer and Generation Capability from Chinese to English for Machine Translation

Xing Fukun
Zhejiang International Studies University
xingfukun@126.com

Xu Jianing
Qingdao University
461096572@qq.com

Abstract

Clause complexes are combined of clauses and the combination patterns differ across languages, and these differences play a crucial role in machine translation quality. However, our understanding of the exact impact is limited. To shed light on this issue, this paper focuses on Chinese-English translation. It examines Chinese clause complexes and expert translations in various genres, and evaluates mainstream machine translation systems and ChatGPT in terms of naming sharing relationships and sharing types. The findings reveal a significant disparity between the ability of machine translation systems and expert translators to transfer and generate clause complexes. Machine translation systems exhibit weaker performance in completing, converting, and refining naming information. They also demonstrate a limited range of clause combination patterns and retain many traces of the original Chinese text, which greatly compromises the quality of the translation.

Keywords: clause complexes, transfer and generation capability, machine translation, ChatGPT

1 引言

©2023 中国计算语言学大会

小句复合体由具有紧密逻辑语义关系的小句序列构成，是小句之上，篇章之下的语言单位。近年来小句复合体研究取得重要进展，宋柔（2022）对小句复合体的语法特征和性质进行了系统描写，为揭示小句复合体层面的语言规律提供了重要支持。小句复合体具有语种普遍性，其内部的话头共享机制及共享类型普遍存在于汉语、英语等语言之中，但话头共享类型的使用分布在不同语种间存在显著差异（宋柔，2022），这种差异给小句复合体结构分析与转换生成带来很大困难。下面是一个专家译文与机器译文在小句复合体层面的对比。

例1

原文：（选自2022年国务院政府工作报告）

各方面要围绕贯彻这些重大政策和要求，
细化实化具体举措，
形成推动发展的合力。

专家译文：

All of us involved

must adopt detailed and effective measures

||

to meet these major policy require-

ments

and create synergy for driving growth.

翻译引擎A：

All parties should focus on implementing these major policies and requirements,
refine and implement specific measures,
and form a joint force to promote development.

上面的汉语原文、专家译文和机器译文都通过换行缩进形式标注出小句间的话头共享关系和共享类型。汉语原文是一个由3个标点句组合而成的小句复合体，每个标点句一行，其中第2和第3行都缩进了第1行的“各方面要”之后，表示这两行与第1行的“各方面要”之间存在话头共享关系，第2行和第3行分别与共享的话头“各方面要”组合，可生成完整的小句，分别是：①各方面要围绕贯彻这些重大政策和要求。②各方面要细化实化具体举措。③各方面要形成推动发展的合力。由于共享的话头位于第一个小句的起始位置，因此第2、3小句的话头共享类型都属于分支模式。

与汉语原文相仿，使用换行缩进对专家译文和机器译文进行话头共享的标注。在专家译文中，其话头共享关系和共享类型与汉语原文存在差异，表现为：话头共享关系存在差异，译文第2行中的measures作为话头被后句共享，但其对应的原文“举措”在汉语原文中并非话头，不存在共享关系；话头共享类型也存在差异，汉语原文的话头共享类型只有分支模式一种，而专家译文中则是第2个小句共享“All of us”，第4个小句共享“All of us must”，这两个小句的共享类型都是分支模式，而第3个小句则共享了第2个小句的末尾成分“measures”，这种共享前面小句的非开始成分做话头的类型属于新支模式，因此专家译文包含分支模式和新支模式两种话头共享类型。机器译文在话头共享关系和共享类型层面与汉语原文完全相同，第2、3小句都共享了第1句的“All parties should”，且话头共享类型都是分支模式。

观察机器译文的每一个小句，其质量基本不存在问题，但如果观察译文小句组合而成的小句复合体，则机器译文质量较专家译文有较大差距，具体表现为机器译文的同一个主语，后面连续跟了三个具有并列结构的限定性动词短语，这种多个动词短语并列的表达形式在英语中较少使用，带有明显的中式英语痕迹，机器在小句组合模式层面原样照搬了汉语的小句组合模式，未作出适合英语表达习惯的小句组合模式调整，从而影响了机器译文的整体质量。

现有机器翻译研究缺少对小句复合体转换生成能力的关注，小句组合模式的差异对机器译文质量存在哪些影响以及影响程度如何尚不清楚。本文以汉英机器翻译为例，选取多语种的汉语小句复合体及专家译文，从话头共享关系和共享类型两方面对主流机器翻译系统以及ChatGPT的小句复合体转换生成能力开展调查。调查的机器翻译系统包括百度翻译、小牛翻译和有道翻译三个主流翻译系统，并将ChatGPT纳入调查范围，目的是调查ChatGPT这类生成式大语言模型在小句复合体层面的生成能力，从而能够对专家、机器翻译以及生成式大语言模型在小句复合体层面的转换生成能力进行更全面的对比分析。

2 小句复合体理论和机器翻译现状

2.1 小句复合体理论

语言学界对于小句复合体结构本质的共识就是逻辑语义关系，Halliday (1985) 认为小句复合体是通过衔接手段将语义上相互依赖的各小句连成一体，各小句间有一定的逻辑关系。宋柔 (2018) 则从小句复合体的形式规律层面开展研究，对小句复合体的语法结构进行了系统描写，提出了话头 (naming) 和话身 (telling) 等核心概念，认为话头是话语字面上的出发点，话身是对于话头的陈述，并基于话头共享结构界定了汉语的小句复合体，认为小句复合体不仅是小句间具有紧密逻辑语义结构的标点句序列，而且有特定的语法结构，即小句间遵循一定的模式来共享话头和话身。小句复合体的话头话身共享大体分为四种类型，即分支模式、新支模式、后置模式和汇流模式。其中分支模式的特点是一个话头被多个右置的话身共享，是汉语语篇中出现最多的话头共享模式。

宋柔和葛诗利 (2015) 基于小句复合体理论，面向篇章机器翻译对英汉翻译单位和翻译模型进行探究，以NT小句为基本单位，设计了篇章机器翻译的拆分、翻译、装配三步模型 (PTA模型)，并以PTA模型的翻译过程为标注内容，建设英汉篇章NT小句对齐语料库。宋柔与葛诗利团队 (2020) 基于宾州英语树库的华尔街日报建立了一个包含5000个英语句子的英语语料库，并对其进行NT小句划分，结果显示小句复合体理论的话头话身关系类型可以对语料库中99%以上的句子进行描写，进一步证实了该理论的语种通用性。英语中同样存在小句复合体，小句复合体中的小句之间可以通过成分共享进行分析，共享类型与汉语一致，包括分支、新支、后置、汇流四种模式，但模式的使用分布与汉语有较大差异，集中表现在英语中多用新支模式，而汉语多用分支模式，具体统计结果可参考宋柔 (2022)。

小句复合体理论从话头共享角度描写小句间关系，不仅揭示出小句间组合的形式规律，同时也为不同语言间的对比研究提供了重要视角，从已有研究结果看 (宋柔, 2022; 张学贞, 2022)，小句复合体理论一方面可以全覆盖地对汉语、英语和日语进行描写，另一方面三种语言间在话头共享模式层面又各具特点，各不相同。

2.2 机器翻译现状

机器翻译的概念起源于20世纪30年代，历经基于规则的机器翻译、基于统计的机器翻译、基于实例的机器翻译和基于不同方法应用的机器翻译 (胡开宝, 2016)。2016年，谷歌推出基于神经网络的机器翻译，机器翻译效果有了极大提升，并取代统计机器翻译成为谷歌、百度、微软等商用在线机器翻译系统的核心技术。

随着机器翻译的不断发展，人们对机器翻译的需求也不仅局限于原先的句子级别，而是到了句子之上的篇章级别，神经网络机器翻译的发展也带动了篇章机器翻译的发展，目前有关机器篇章翻译的热点主要集中在神经机器翻译领域。虽然篇章神经翻译的对象是篇章，但由于受到计算资源限制，大部分模型仍先依次翻译篇章中每个句子，再拼接得到最终篇章译文。因为需要关注上下语义的连贯，目前篇章神经机器翻译大多注重译文中篇章的共指性和连贯性，从机器的上下文建模、模型分析和模型训练几个方面进行研究，篇章层面的翻译仍存在的问题 (苏劲松, 2020)。

以句子为单位机器翻译系统在处理篇章信息时会忽略上下文信息 (李哈佶, 2020)，同时也欠缺相关的技术对小句复合体内部小句间逻辑语义的分析。目前机器翻译较少关注汉英转换在小句复合体层面的生成能力，研究方面也较为欠缺对汉英机器翻译在小句复合体层面生成能力的调查。

美国OpenAI公司于2022年11月底推出ChatGPT模型，虽然该模型并非是专门针对翻译任务而训练的模型，但由于其依托于大语言模型 (Large Language Model, LLM)，具有语言间的翻译能力，在翻译领域有一定的应用 (朱光辉, 2023)。目前已经有研究者 (Wang&Lyu, 2023; Jiao et al., 2023) 对ChatGPT等大语言模型的翻译能力开展调查和评价工作，但评价方法与标准存在一定局限性，表现为现有译文评价基本沿袭已有翻译数据集并基于BLEU值进行译文评价。由于主流翻译模型基本是在该评价框架下进行训练优化，因此利用该框架评价ChatGPT与主流机器翻译系统具有一定的不公平性，难以反映出ChatGPT这类大语言模型独有的翻译能力。更重要的是，BLEU评价方法得出的结果反映的是机器译文与人工译文在Ngram层面的相似度，但由于小句间的话头共享关系具有跨距特点，且有时会跨越有多个小句，因此很难被Ngram捕获，因此现有的译文自动评价方法难以反映出不同译文在小句复合体

层面的差异，进而无法对不同模型的小句复合体翻译能力做出评价。本文针对小句复合体层面的转换生成能力开展人工调查与评价，目的是调查不同模型以及专家在小句复合体层面的翻译情况，进而给出不同翻译主体在小句复合体层级的翻译特点与能力，为优化机器翻译提供支持。

3 调查方法

3.1 翻译引擎选取

本研究选取百度翻译、小牛翻译和有道翻译三个主流翻译引擎和大语言模型ChatGPT作为调查对象，前三个翻译引擎以翻译引擎A、翻译引擎B和翻译引擎C来代表。前三个翻译引擎均使用主流的神经机器翻译理论和技术，通过深度神经网络使计算机先对语料库进行自动学习，再自动生成机器译文。ChatGPT是新兴的生成式大语言模型，通过海量无标注语料的预训练以及高质量标注语料进行有针对性的微调而成，具有较为广泛的能力，本文只对其翻译能力进行调查。

这四类翻译引擎应用广泛，市场上有很大的客户群体，其翻译能力在机器翻译行业具有一定的代表性。本研究对这四类翻译引擎的汉译英机器翻译进行标注、分析和总结，反映出目前整体机器翻译引擎的汉译英在小句复合体层面的机器翻译能力。

3.2 调查语料

本文选择政府工作报告、白皮书、法律以及小说四类文本作为调查语料，其中政府工作报告为2022年全国人大政府工作报告，白皮书为2019年白皮书，包括《新时代的中国国防白皮书》《中国新型政党制度白皮书》《中国应对气候变化的政策与行动白皮书》，法律为中国法律节选，包括《法官法》《个人信息保护法》《国防法》《行政处罚法》《反食品浪费法》。政府工作报告、白皮书、法律这三类信息类文本作为政府官方文件，汉语特点鲜明且语言严谨。小说选取《围城》和《蛙》的节选片段。补充文学类文本为的是增加样本中汉语语言的多样性和代表性。其中2022年政府工作报告的译文为新华网发布的国务院官方权威英译文，白皮书译文为国务院新闻办公室发布的英文版，中国法律节选的英文版出自全国人民代表大会发布的官方权威英译文，《围城》的译文为珍妮·凯莉和茅国权的英译版，《蛙》的译文为中国文学翻译家葛浩文所译的英译版。选取的译文均为该领域的专家译文，译文质量和权威性较高，可作为调查的基准与机器译文进行对比，探究机器的译文生成能力。

3.3 调查设计与分析方法

本文选取了汉语中使用最广泛的分支模式作为调查对象，从调查语料中共选出共44个具有分支模式的小句复合体，共154个小句，其中政府工作报告小句70句，白皮书、法律、小说这三类语体的小句各28句。目前的调查样本的绝对数量不大，但并不会对调查目的造成本质影响，一方面是考虑了语体因素，在多种语体中采样，保证调查具有较大的覆盖范围，另一方面是选用了多个机器翻译系统进行调查，同一个原文提交给不同的机器翻译系统翻译，然后对机器译文进行比较分析，调查不同机器翻译系统间是否存在某些共性特点。未来将进一步扩大调查的语料规模，提高调查结果的可靠性和代表性。调查的基本单位是小句复合体，内容包括汉语原文的小句复合体、原文对应的专家译文小句复合体、机器译文小句复合体以及ChatGPT翻译得到的译文小句复合体。

针对调查目标，对原始语料进行标注，基本内容包括：汉语原文的标点句切分及话头共享关系标注、专家译文的小句切分和话头共享关系标注以及机器翻译引擎A、B、C和ChatGPT的译文做小句切分和共享关系标注，同时在小句复合体层面将汉语原文与专家译文、机器译文进行对齐。

调查的基本方式是，以专家译文为基准，将专家译文以及四类机器译文的标注结果分别与汉语原文的标注结果进行对比，标注出译文与原文在话头共享关系和话头共享类型层面的异同。

“话头共享关系”是指汉语原文小句中的话头与对应译文小句中的话头之间的对应关系，如果原文和译文的话头一致，则标注为“话头一致”，反之标注为“不一致”。对于话头不一致的情况，进一步调查不一致的类型，分别是话头补足、话头变形、话头转换和话头提炼。“话头共享类型”是指汉语原文小句的话头共享类型与译文小句的话头共享类型间的对应关系，也分为一致

原文第1行中的话头没有显式出现，但从上下文可以推断出隐含话头是“我们”，因此在标注中以大括号的形式补充出话头，译文的第1行与汉语一致，也没有出现话头，标注中同样以大括号的形式补充出话头“*We will*”，原文与译文均未出现隐含的话头，且补充后的话头具有对译关系（忽略助动词，下同），这类情况被归为话头一致。

(2) “话头不一致”的情况主要包含四类：①“话头补足”，即在汉语原文没有出现话头，但话头可在上下文中找到，译文将原文隐含的话头显式补足出来的情况，类标注为“*b*”；②“话头变形”，即英语小句中以其他的形式如代词等表现出汉语隐含的话头的情况，类标注为“*c*”；③“话头转换”，即英语小句中的话头由汉语原文中非话头成分转换而来，如汉语原文中的宾语成分在英语译文中转换为话头的情况，类标注为“*d*”；④“话头提炼”，即英语小句的话头是基于汉语原文提炼总结得到的，类标注为“*e*”。例5给出了话头提炼的情况，其他情况不再一一给出实例。

例5

原文：（选自2022年国务院政府工作报告）

{我们}加强大宗商品保供稳价，
着力解决煤炭电力供应紧张问题。

ChatGPT:

Efforts must be made to stabilize the supply and prices of bulk commodities, //e
and to tackle the problem of coal and electricity shortages. //e

例5中汉语原文的话头为“我们”，但译文的话头“*Efforts*（举措）”则在原文中没有出现，是译者基于原文提炼得到，原文讲到的“加强大宗商品保供稳价”和“着力解决煤炭电力供应紧张问题”都可被认为是“*Efforts*（举措）”的具体表现，因此译者提炼出一个在原文中并不存在但符合原文表达意思的话头，这类情况称之为“话头提炼”。

可以看出，无论话头共享关系，还是话头共享类型，都是发生在小句与小句之间的组合上，与小句中具体词或短语翻译无关，因此是小句复合体层面的语言现象，通过调查机器翻译在这两类语言现象上的表现，可以反映出机器在小句复合体转换生成上的基本能力。

4 调查结果与分析

4.1 调查结果

由于专家译文与机器翻译引擎的译文在小句总数量上不完全一样，为了对不同来源的译文进行统一对比，本文在原始数据基础上，进行归一化处理，下表给出不同情况的绝对数量和其占各自译文小句总数的百分比。其中四类不同的翻译引擎A、B、C、ChatGPT与专家译文话头共享类型对应情况调查结果如表1所示。

	翻译引擎A	翻译引擎B	翻译引擎C	翻译引擎总	ChatGPT	专家译文
一致	87.01%	88.89%	86.58%	87.50%	87.84%	62.94%
独立小句	10.39%	9.15%	10.07%	9.87%	6.08%	11.68%
新支	0.65%	1.31%	2.01%	1.32%	2.70%	19.29%
后置	1.95%	0.65%	1.34%	1.32%	3.38%	6.09%

表 1: 翻译引擎、ChatGPT与专家译文话头共享类型对应情况调查结果

表1给出了四类翻译引擎和专家译文在话头共享类型层面的对应情况，从中可以看出，四类翻译引擎在话头共享类型的对应关系中，“一致”情况具有很大的相似性，其占比均为超过85%，位于85% 89%之间，而专家译文的“一致”情况只有62.94%，低于机器翻译引擎约20个百分点，反映出专家译文在话头共享类型上，与原文有较大差异，而机器译文与原文的差异程度远小于专家译文的差异程度。三种翻译引擎“一致”情况总占比与ChatGPT都是87%多一些，这两类翻译引擎在话头共享类型上都与原文保持了较高的相似程度，而与专家译文则有较大差异。这初步说明机器译文在话头共享类型上更多地受到源语的影响。

在“不一致”的三种情况中，机器译文与专家译文也表现出较大的差异性。首先，译文是“独立小句”的情况中，占比最高的是专家译文，达到11.68%，而最低的是ChatGPT，为6.08%，二者相差5个百分点，而其他三个专门的翻译引擎则较为相似，基本在10%左右。该数据说明，

相较机器译文，专家译文更多使用独立小句，即译文小句中包含完整的话头和话身，对于汉语原文中使用话头共享关系的小句，在翻译中会将其话头补全，并明确地翻译出来。ChatGPT的译文则更少地使用独立小句，相应地ChatGPT会使用更多的组合小句来进行翻译，这表现出ChatGPT的独特之处，其原因和译文效果需要进行专门评估。

在“新支小句”中，则专家译文的使用比例达到19.29%，远远超过机器译文使用新支小句的比例，翻译引擎A、B、C的比例基本为2%或更少，ChatGPT也只有2.7%。由于在本文选取的汉语原文语料中，话头共享类型都是分支类型，没有新支类型，因此，如果译文出现新支类型，说明在翻译转换中在一定程度上摆脱了汉语原文的影响。从这个角度看，专家译文高比例地使用新支类型，表现出专家译文相对于机器译文的特别优势。同时，由于ChatGPT使用新支类型的比例在一定程度上高于其他三个机器翻译引擎，因此也可以初步判断ChatGPT在小句复合体层面的译文转换生成能力强于已有的机器翻译引擎。

在“后置小句”中，专家译文同样表现出远高于机器译文的水平，与新支小句同理，由于汉语原文中没有后置类型，因此后置类型的使用也能表现出译文脱离原文影响的程度。专家译文受到原文影响的程度最小，ChatGPT居中，另外三个机器翻译引擎则较大地受到了原文话头共享类型的影响。

下面是一个机器译文、ChatGPT译文和专家译文在话头共享类型层面的对比情况。

例6

原文：（选自中华人民共和国国防法）

现役部队是国家的常备军，
主要担负防卫作战任务，
按照规定执行非战争军事行动任务。

翻译引擎A:

The active force is the standing army of the country,
which is mainly responsible for defense operations
and performs non war military operations in accordance with regulations.

翻译引擎B:

The troops in active service are the standing army of the country,
which are mainly responsible for defensive operations
and perform non-war military operations in accordance with
regulations.

翻译引擎C:

The active forces are the state's standing forces,
mainly responsible for defensive operations,
and carry out non-war military operations in accordance with regulations.

ChatGPT:

The active-duty troops are the country's regular army
and mainly hold the responsibility of defensive combat missions,
and execute non-war military operations tasks as required.

专家译文:

As the standing forces of the State,
active-duty forces shall be mainly tasked with defense operations
and perform non-war military operations in accordance with regulations.

例6中汉语原文是一个分支类型的话头共享结构，翻译引擎A、B、C及ChatGPT与原文的话头共享类型一致，呈现出较强的原文影响程度；而专家译文话头共享类型与原文不完全一致，使用了后置类型，表现出专家译文相比于机器译文在翻译转换上的优势。

通过表2可以较为直观地看出四类翻译引擎和专家译文在话头共享关系层面的对应情况，其中“话头一致”情况，机器翻译引擎之间仍有较大的相似性，且占比较高，位于78% 89%之间；而专家译文的“话头一致”情况只有68.02%，低于机器翻译引擎约10-20个百分点，反映出专家译文在话头共享关系上与原文差异较大，而机器译文与原文的差异程度仍小于专家译文的差异程度。将机器翻译引擎分为两大类，在“话头一致”方面三种专门翻译引擎与ChatGPT均为80%以上，这两类翻译引擎在话头共享关系上与原文相似程度也较高，而与专家译文相比有一定的差

	翻译引擎A	翻译引擎B	翻译引擎C	翻译引擎总	ChatGPT	专家译文
话头一致(a)	81.17%	88.31%	78.67%	82.75%	80.54%	68.02%
话头补足(b)	9.74%	1.95%	14.00%	8.52%	1.34%	5.58%
话头变形(c)	6.49%	7.79%	6.00%	6.77%	5.37%	5.58%
话头转换(d)	2.60%	1.95%	1.33%	1.97%	6.04%	19.80%
话头提炼(e)	0.00%	0.00%	0.00%	0.00%	6.71%	1.02%

表 2: 翻译引擎、ChatGPT与专家译文话头共享关系对应情况调查结果

异，机器译文不仅在话头共享类型上更多地受到源语的影响，在话头共享关系方面也呈现出相似的情况。

“话头不一致”的四种情况，专家译文与机器译文也存在一定的差异性。本文认为这四种情况，即“话头补足”、“话头变形”、“话头转换”和“话头提炼”的话头生成的难度逐级递增，话头生成能力也越来越强。首先，译文为“话头补足”的情况中，三个专门的翻译引擎占比最高，达到8.52%，而专家译文次之，为5.58%，ChatGPT占比最低，仅为1.34%，与三个专门翻译引擎相比，相差7个百分点。该数据说明，机器译文更多地使用话头补足，即在英语译文中对汉语上下文中出现但在小句复合体中省略的主语进行的形式上的补足，专家译文相对较少使用这种方法，而ChatGPT使用占比更少。在“话头变形”的情况中，机器译文与专家译文在占比方面均无较大差异，均在5%-7%之间，表明机器译文和专家译文在翻译译文时均会在一定程度上使用话头变形，即将呈现为名词等情况的原话头，在对应译文以及补充缺失话头时转换为同义名词或代词等情况的翻译方法。在“话头转换”的情况中，专家译文呈现出明显的差异，占比高达19.80%，远超出ChatGPT的6.04%和翻译引擎A、B、C的1.97%，展现出专家译文相比于机器译文，脱离原文句式对话头进行转换的能力。而译文为“话头提炼”的情况中，ChatGPT显示出独特的优势，其占比为6.71%，不仅高于翻译引擎A、B、C的0%，也高于专家译文的1.02%。在其他翻译引擎在话头提炼能力展现出欠缺的情况下，ChatGPT可以做到对语义的理解总结提炼出原文没有明确表达出的内容作为话头，表现出其独特的优势和较高的话头生成能力。

例7给出了话头补足的情况，原文第1行没有出现话头，但可以从上下文中推断出隐含的话头是“我们”，译文中则补充出话头“We will”。由于本数据集大多为信息类文本，缺失主语大多为“We”，补足难度较低，根据数据发现，机器翻译引擎ABC话头补足能力相对不错，为8.52%，高于专家译文的5.58%和ChatGPT的1.34%。虽然在数据上超越了专家译文和ChatGPT，但专家译文使用频率低的原因是话头补足在话头生成方面相对简单，较多使用了较为复杂的话头转换，而ChatGPT则是更多的使用了更为复杂的话头提炼。

例7

原文：（选自2022年国务院政府工作报告）

{我们}加强大宗商品保供稳价，

着力解决煤炭电力供应紧张问题。

翻译引擎A:

We will strengthen supply and price stability for bulk commodities, //b

and work hard to solve the problem of tight supply of coal and electricity. //a

话头变形是指原文话头与译文话头的所指对象一致，但表达形式上有变化，例如原文话头为名词，在译文中转换为另一个同义名词或代词的情况，这类情况翻译引擎、ChatGPT和专家译文差别不大，不详细列举。

话头转换是将原文非话头成分，例如宾语等成分转换为译文的话头，涉及句式的转换，是一类更为复杂的话头生成能力。如例8所示，原文第2行中的“多场专题新闻发布会”不是话头成分，而是宾语成分，但其对应的译文为“A number of special press conferences”充当了译文的话头，这种情况属于话头转换。此外，译文第2句的“major issues”是新支句话头，对应原文的“重大事项”，同样也发生了话头的转换。专家译文中话头转换的占比为19.80%，明显高于翻译引擎的1.97%和ChatGPT的6.04%。专家译文多使用话头转换，可以实现句式的灵活性和多样性，而机器翻译引擎的译文较多地受原文影响，话头转换能力较弱，ChatGPT则位于二者之间，强于机器翻译，而弱于专家译者。

例8

原文：（选自2019年新时代的中国国防白皮书）
 {国防部}围绕深化国防和军队改革、裁减军队员额等重大事项，
 召开多场专题新闻发布会。
 组织近百家中外媒体多次赴部队、军事院校参观采访。

翻译引擎A:

A number of special press conferences were held around major issues //d
 || such as deepening the
 reform of national defense and the military, and reducing military posts. //d

Nearly 100 Chinese and foreign media have been organized to visit the troops and military academies for many times. //d

话头提炼是需要通过对语义的理解总结提炼出原文没有明确表达出的内容，并将其作为话头。由于需要语义理解，因此这一类的话头生成较为复杂和困难。例9中，原文的话头是“我们”，但译文话头“Efforts（举措）”是基于原文提炼得到，原文讲到的“加强大宗商品保供稳价”和“着力解决煤炭电力供应紧张问题”都是“Efforts（举措）”的具体表现，因此译者提炼出一个在原文中并不存在但符合原文表达意思的话头。机器翻译引擎的话头提炼方面存在明显不足，在调查的样例中没有出现此类话头；ChatGPT的话头提炼占比则高达6.71%，显著高于专家译文的1.02%，表现出ChatGPT这类生成式大语言模型在话头转换生成方面的独特优势，其能力甚至超过专家译者。深层原因值得进一步调查分析，仅从原文对译文的影响角度看，机器译文受到原文形式的影响最大，没有在原文形式层面出现的内容，很难被翻译生成出来，导致翻译引擎生成提炼式话头的的能力最弱；专家译者基于对原文的理解进行翻译，能够在一定程度上摆脱原文形式影响，生成一些未在原文中出现，但符合原文语义的话头，但这种能力有着较大个体差异，优秀译者此方面能力较强，而一般译者则较弱；ChatGPT这类大语言模型具有较强的语义表示能力，基于深层的语义表示进行转换，从而较好地摆脱原文形式束缚，而能够较为灵活地生成提炼式话头。

例9

原文：（选自2022年国务院政府工作报告）
 {我们}加强大宗商品保供稳价，
 着力解决煤炭电力供应紧张问题。

ChatGPT:

Efforts must be made to stabilize the supply and prices of bulk commodities, //e
 and to tackle the problem of coal and electricity shortages. //e

4.2 翻译引擎与专家译文话头共享类型和话头共享关系的对比分析

为进一步探究不同翻译主体的译文在话头共享类型和共享关系方面的整体差别，本文引入交叉熵进行差异度量。交叉熵是来度量两个概率分布之间差异程度的概念，交叉熵越大，则两个概率分布相差越大，交叉熵越小，则两个概率分布的差异越小（李郝林，2014）。例如两个样本集分别为 $P=\{ \}$ ， $Q=\{ \}$ ，交叉熵的定义为：

$$H(P, Q) = \sum_{i=1}^n (p_i) \log(1/q_i)$$

式中，P为后验概率，Q为先验概率， $H(P, Q)$ 则为概率P和概率Q的交叉熵，本文将专家译文相应的概率作为后验概率，不同翻译引擎的相应概率作为先验概率，分别计算出专家译文和四类不同翻译引擎的话头共享类型和话头共享关系分布的交叉熵，如图1和图??所示，其中数据为0的对数为0，并最终对结果保留小数点后四位。

	翻译引擎A	翻译引擎B	翻译引擎C	ChatGPT
话头共享类型	0.6789	0.6499	0.5972	0.5697
话头共享关系	0.5390	0.5734	0.5990	0.4926

表 3: 翻译引擎与专家译文话头共享类型及话头共享关系分布的对比情况

交叉熵的越大，翻译引擎和专家译文在共享类型的使用上的差异越大，根据表3可以看出翻译引擎A、B、C和ChatGPT的话头共享类型与专家译文的相似程度，其中依托于大型语言模

型的ChatGPT与专家译文的差异最小，相似程度更高；而神经网络机器翻译引擎A、B、C则与专家译文差异较大。在话头共享关系方面也呈现相似的结果，ChatGPT相较于其他三类翻译引擎，与专家译文的差异较小，相似度更高，话头的生成能力与专家译者更为接近。

4.3 话头共享模式对应和话头共享关系对应情况的调查结果

根据数据调查发现，翻译引擎ABC的译文的话头共享模式转换类型大多为与原文分支模式的类型相同的对应转换，转换为独立小句、新支模式以及后置模式的能力相对较弱，本文推测基于神经网络的翻译引擎受原文句式的影响较大，故模式较为单一，英语的地道性相对较弱。

而专家译者在话头共享模式上相对灵活，与原文共享模式相同的类型占比相较于基于神经网络的机器翻译引擎和ChatGPT都较少，转换为独立小句、新支模式和后置模式占比相对较高，表现出共享模式和句式的多样化，可读性相对较强。

ChatGPT在话头共享模式的转换能力上弱于专家译者，但相比于基于神经网络的翻译引擎，其共享模式的转换能力的能力更强，受原文影响较小，与专家译者更为接近。

而通过话头对应的类型和数量来观察基于神经网络的机器翻译和话头生成能力发现，在与原文话头完全对应占比最小的仍为专家译文，基于神经网络的翻译引擎的译文大多为与原文话头对应，呈现出大量译文与汉语原文话头完全对应的情况，在汉语主语缺失时译文也对应翻译为主语缺失的英语译文，汉语痕迹十分明显，译文的地道性、可读性上都较为欠缺，表现出话头生成能力的弱势。

基于神经网络的翻译引擎拥有一定的话头补足能力，但较为复杂的话头转换和话头变形等话头生成能力较弱，更为复杂的话头提炼较为欠缺。ChatGPT前两项能力与专家译文较为接近，话头转换能力优于基于神经网络的翻译引擎但弱于专家译文，但其话头提炼较为优越，弥补了之前的机器翻译在此项能力上的欠缺，甚至在数据上高于专家译文，本文推测ChatGPT在话头生成方面受原文影响和约束程度较小，表现出大语言模型在翻译时话头生成能力的独特优势。

5 总结

本研究对目前机器翻译在小句复合体层面的转换生成能力进行调查，通过对四类机器翻译引擎和专家译文进行话头共享类型、话头共享关系两方面的标注，并将基于神经网络的翻译引擎ABC的译文、基于大型语言模型的ChatGPT译文与专家译文进行对比和分析，不仅可以较为客观地表现出目前机器翻译在小句复合体层面的翻译生成能力，还可以总结出目前机器翻译在小句复合体层面的部分不足。翻译引擎、ChatGPT与专家译文相比，在话头共享模式和话头对应方面的翻译能力仍有提升的空间，而其中ChatGPT作为基于大型语言模型的新兴机器翻译，与前三类神经网络机器翻译相比，在话头共享类型和话头共享关系两个方面与专家译文的相似度更高，甚至在话头共享关系最复杂的“话头提炼”上，不仅弥补了机器翻译引擎在这一方面能力的欠缺，甚至使用数量上高于专家译文，话头提炼需要对文章内容进行总结归纳，需要较高的语义理解能力，而ChatGPT相比于专家译文拥有更强的话头提炼能力，展现出较强的话头生成能力，本文猜测可能是原文对于基于大语言模型的ChatGPT的约束程度要小于原文对于专家译文的约束程度，反映出ChatGPT生成式模型在翻译方面的独特优势，后续可以基于汉语无主语小句的数据集进一步开展ChatGPT话头总结能力的探究。

本文选择了44个小句复合体作为调查样本，在绝对数量上存在一定的不足，未来需进一步扩大调查规模。为弥补样本数量的不足，本文选择了多个语体文本作为调查样本，同时选择了多个翻译引擎进行调查。调查结果显示，专家译者、神经网络机器翻译系统以及ChatGPT在小句复合体层面的转换生成能力存在较大差异，机器翻译系统的复杂话头生成能力较人类专家和ChatGPT都有较大差距，未来需要进一步探究优化方法，提升机器翻译系统在小句复合体层面的译文质量。同时，本文的调查结果也显示出人类专家在小句复合体层面的转换能力具有一定的局限性，容易受到原文的约束，而降低译文的地道性，针对此类问题，需要将小句复合体层面的翻译能力作为专门对象，加以研究并开展针对性训练，为提升人的翻译能力提供支持。

- 本作品已根据《Creative Commons Attribution 4.0 International Licence》获得许可。许可证详细信息：<http://creativecommons.org/licenses/by/4.0/>.

致谢

本文得到宋柔教授的宝贵意见与建议，在此深表谢忱，文中所有错误缺漏全由本文作者自负。

本文是国家社科基金项目“汉英小句级对齐语料库研制与应用研究”（19BYY081）的阶段成果，同时也得到浙江省高等教育“十四五”教学改革项目（jg20220440）的支持。

参考文献

- Halliday M A K. 1985. *An Introduction to Functional Grammar*. Edward Arnold, London.
- Wang L, Lyu C, Ji T, et al. 2023. Document-Level Machine Translation with Large Language Models. *arXiv preprint arXiv:2304.02210*.
- Jiao W, Wang W, tse Huang J, et al. 2023. Is ChatGPT A Good Translator? Yes With GPT-4 As The Engine.
- 葛诗利, 宋柔. 2020. 基于成分共享的英汉小句对齐语料库标注体系研究. *中文信息学报*,34(06):27-35.
- 胡开宝,李翼. 2016. 机器翻译特征及其与专家翻译关系的研究. *中国翻译*,37(05):10-14
- 李晗佶,陈海庆. 2020. 机器翻译技术困境的哲学反思. *大连理工大学学报(社会科学版)*,41(06):122-128.
- 刘满芸. 2016. 翻译技术时代翻译模式的裂变与重构. *中国科技翻译*,29(04):17-20.
- 宋柔, 葛诗利. 2015. 面向篇章机器翻译的英汉翻译单位和翻译模型研究. *中文信息学报*,29(05):125-135.
- 宋柔. 2023. 汉语小句复合体和话头结构. 中国社会科学出版社, 北京.
- 宋柔. 1985. 小句复合体的语法结构. 商务印书馆, 北京.
- 朱光辉, 王喜文. 2023. *ChatGPT*的运行模式、关键技术及未来图景. *新疆师范大学学报(哲学社会科学版)*,44(04):113-122.
- 张学贞. 2022. 基于汉英日三语可比语料库的NT小句复合体对比研究. 青岛大学.