

中文情态词在机器翻译中的错误类型分析

王丽娜¹, 张倩¹, 张大鲲¹, 郝杰²

(1. 东芝(中国)研究开发中心, 北京 100600; 2. 58 集团, 北京 100015)

E-mail: {wanglina, zhangqian, zhangdakun}@toshiba.com.cn; haojie01@58ganji.com

摘要: 语言本身的复杂性给机器翻译带来了巨大的困难, 对机器翻译结果的详细分析有助于有针对性地提高翻译系统的质量。本文在分析了现代汉语中情态词的分类基础上, 选择了四组测试集(包含和不包含情态词的中文短句和长句), 在中英、中日两个翻译方向上, 对不同的四个翻译系统的翻译结果进行人工标注和错误分析, 详细分析了情态词、句子长度以及语言对因素对翻译错误类型的影响。从情态词的角度, 进一步提出了用于机器翻译系统的改进方案。

关键词: 机器翻译; 情态词; 错误分析

Error Analysis of Chinese Modal Words in Machine Translation

WANG Lina¹, ZHANG Qian¹, ZHANG Dakun¹, HAO Jie²

(1. Toshiba (China) R&D Center, Beijing 100600; 2. 58.com, Beijing 100015)

E-mail: {wanglina, zhangqian, zhangdakun}@toshiba.com.cn; haojie01@58ganji.com

Abstract: Complexity of language brings great difficulty in machine translation. Detailed analysis on translation output can help in improving machine translation system's quality. This paper analyzes types of Chinese modal words, firstly selects four sets of Chinese sentences, two of which are short sentences with and without modal words, and the other two are long sentences with and without modal words, then annotates and analyzes error types of Chinese-to-English and Chinese-to-Japanese output from four machine translation systems. We analyze the effect of modal words, sentence length and language pair on translation error types, and propose some methods to improve the machine translation system's quality.

Key words: Machine translation; Modal words; Error analysis

1 概述

随着计算机技术的发展, 机器翻译系统, 尤其是在线机器翻译系统(如 Baidu, Bing, Youdao 等)在人们的日常生活和工作中发挥着越来越重要的作用, 然而机器翻译系统译文的质量目前还没有达到人们的期望。为了有针对性的提高机器翻译的译文质量, 对翻译结果的错误分析逐渐被人们所重视, 系统性的错误分析不仅有助于我们比较系统之间的性能, 也为提高译文质量的方向提供了数据和理论上的支持。

影响机器翻译译文质量的因素有很多, 其中语言本身的复杂性、灵活性和多变性给机器翻译带来了巨大困难。在现代汉语中, 有一类情态词(包括否定词)的翻译正确与否直接影响到整个句子的翻译质量。情态(modality)的概念最早由西方语言学家提

出^[1]。情态涉及主观性, 表达的是一种主观态度和愿望, 它的表达方式多种多样, 每种表达方式又受到多种因素影响。因此, 机器翻译对于情态词的处理有很大的难度。然而, 情态在句子中的作用却相当重要, 比如:

(1) 大卫去学校了。(2) 大卫一定去学校了。(3) 大卫可能去学校了。这三个句子表达了同一个命题(proposition): 大卫去学校, 但是由于情态词(一定, 可能)的加入, 却表达了说话人对同一事件所持有的不同态度。句(1)陈述了一个客观事实, 表达说话人对这一事件的真实性确定无误, 句(2)和句(3)都增加了说话人的主观态度, 其中句(3)对事件真实性的确信程度最低, 其翻译结果也完全不同: (1) David went to school. (2) David must have gone to school. (3) Maybe David went to school. 可见, 情态词

不仅表达了不同的语义，对句子翻译时的语法和词语的顺序也都会产生影响。

传统的统计机器翻译方法不区分词类，对于情态词的翻译没有任何特殊的处理过程，其翻译质量完全依赖于统计模型。而统计模型一般很难捕捉到语言中的情态现象，所以统计机器翻译系统很难把包含情态词的句子翻译正确。针对这一问题，本文选取了四个常见的机器翻译系统，包括开源统计机器翻译系统 Moses^[2]和三个领先的在线机器翻译系统 Baidu¹、Bing²和 Youdao³，从中英和中日两个翻译方向，对比机器翻译系统的翻译质量，详细分析了包含情态词和不包含情态词的中文句子在机器翻译中的错误类型，以及情态词对不同长度句子的翻译质量的影响。通过对机器翻译译文进行词法和句法层次的对比分析，进一步提出了面向中文情态词的机器翻译质量改进方案，包括对源语言句子的翻译难度分析，训练语料预处理和翻译结果后处理以及源语言和目标语言句子在情态词方面的翻译一致性等。

本文的结构如下：第2节介绍中文情态词的相关概念和类别；第3节描述近年来研究者在情态词标注和机器翻译领域的相关工作；第4节对四个机器翻译系统的结果进行详细评分、错误分析以及提出改进方案；最后是总结。

2 中文情态词

2.1 定义和分类

情态是语义范畴的一个概念^[3]，是通过语法形式（情态动词或情态副词），以及句法手段来表现说话者针对句中命题的主观判断和态度。其涉及说话人的态度和意见，本质上反映的是句子意义相对于一组可能世界的有效性^[4]。在现代汉语的情态研究中，主流思想认为可将情态分为三类：认识情态（Epistemic），道义情态（Deontic）和动力情态（Dynamic）。

认识情态表达了说话人对句子命题发生或存在的可能性程度的主观判断，Palmer^[5]认为认识情态“本质上是对命题的真实性做出判断”，道义情态则“涉及对动

作、状态或事件的影响表达一种指令”，表达了说话人使句子命题的发生成为可能的影响和作用力。认识情态和道义情态都涉及了说话人的主观性，都与命题的可能性有关；但是认识情态和道义情态有根本的区别，认识情态涉及了说话人对命题发生或存在可能性的判断，道义情态则涉及了说话人对命题发生或存在可能性的控制或影响。第三类为动力情态，表达了主语的能力、效能、动力源以及主语的意志和愿望。比如，下面三句话包含了三种情态：（1）他**可能**在家。（认识情态）（2）他**必须**回家。（道义情态）（3）他**要**回家。（动力情态）

2.2 表现形式

在现代汉语中，情态的主要表现方式有两种，即情态动词和情态副词。

语言学家朱德熙^[6]制定了一套形式标准来定义情态动词，根据该标准，可以区分的情态动词总共有 27 个^[7]：能、能够、可以、会、可能、得(dé)、敢、肯、愿意、情愿、乐意、想、要、应、应该、应当、该、许、准、值得、配、别、甬、好、一定、得(děi)、必须。由于汉字的歧义性，有些情态动词在不同的语境中所表达的情态意义也不尽相同，这种一词多义的现象增加了情态词研究的难度。然而，情态动词的多义现象是有规可循的，为了更好的研究情态动词的意义和用法，彭利贞^[8]将情态动词分为单义情态动词和多义情态动词。因此，根据相关研究结果，我们可以将情态动词按照事件可能发生的程度进行分类，如表 1 所示：

表 1. 情态动词分类及示例

类型 \ 程度	认识情态 (Epistemic)	道义情态 (Deontic)	动力情态 (Dynamic)
可能性 (possibility)	可能：会	可以：能	能
盖然性 (probability)	应该	应该：应当	能够
必然性 (necessity)	肯定：一定	必须	必须：得

相对于情态动词，情态副词的数量相对较多。根据苏秋菊^[9]的统计，现代汉语中情态范畴的副词约有 434 个。本文选取较为常用的 111 个情态副词，对其情态意义进行归类和分析，情态副词示例图如表 2 所示：

¹ <http://fanyi.baidu.com/>

² <https://www.bing.com/translator>

³ <http://fanyi.youdao.com/>

表 2.情态副词分类及示例

程度 \ 类型	认识情态 (Epistemic)	道义情态 (Deontic)		动力情态 (Dynamic)
		指令	评价	
可能性 (possibility)	大概; 也许; 似乎	可; 不妨; 要不	正好; 居然; 到底	宁可; 宁愿; 只好
必然性 (necessity)	的确; 确实; 肯定	万万; 务必; 最好		
情感 (sensitivity)	究竟; 多么; 何等			

此外, 情态的表达方式还有很多种, 比如能性补语^[10]、语气助词、情态名词等等。由于这些表现手段较为灵活, 无法形成系统的标准, 因此本文的分析范围限定在情态动词和情态副词内。

2.3 否定

对于机器翻译系统来说, 还有一种常见的翻译错误是极性 (polarity) 翻译错误。句子的肯定或否定如果错译, 将会导致整个句子的意思完全改变, 直接影响到翻译质量。

本文在分析情态动词和情态副词对于翻译结果的影响的同时, 也将五个常见否定词加入到分析列表, 分析其对机器翻译译文的影响。除去情态动词和情态副词, 本文分析的五个否定词为: 不; 没; 没有; 未曾; 莫。

3 相关工作

3.1 情态词标注

情态词涉及到说话者或命题主语的态度和观点, 属于主观化的范畴, 在自然语言处理和机器翻译领域, 对于情态词的处理首先集中在情态词标注方面。由于不同语言的表达方式不一样, 大部分研究者集中在语言独立的自动标注系统的创建上。情态表达方式种类繁多, 涉及意义宽泛, 因此, 研究者们通过建立情态词列表^[11-12]或者限定情态类型^[13-14]来缩小情态标注的范围。Pakray^[11]建立了一个包含英语情态词和否定词的数据库和四个标注模块, 在此基础上实现对情态词和否定词的自动识别, 其标注精确率可以达到 0.6262。和 Pakray 的标注方法相似, Prabhakaran^[13]创建了一个包含英文情态词列表的自动标注器, 但其对涉及到的情态范围进行了限制, 只处理五种情态: ability (能力)、effort (尝试)、intention (意图)、success (成功) 和 want (希望)。Quaresma^[12]在葡

萄牙语 (Portuguese) 的情态标注研究中, 研究了三个情态词: poder (may/can), dever (shall/might) 和 conseguir (manage to/succeed in/be able to)。Quaresma 和 Prabhakaran 均利用支持向量机 (SVM) 对情态词的意义进行分类, 两者创建的自动标注器在标注精确率、召回率和 F 值方面取得了较好的效果。

在现代汉语方面, 由于中文动词没有形态上的变化, 只能根据上下文来判断时态, 因此, Xue^[14]提出了一种远程标注方法, 涉及语义时态、事件类型和情态。该方法在词对齐的双语语料中, 先标注英语句子的时态、事件类型和情态, 然后再利用词对齐将标签转移到相对应的中文词上。Xue 做了两组实验, 第一组实验通过对微博、新闻和广播语料的标注, 测试了标注置信度, 结果显示新闻语料的标注一致性最高, 微博语料标注一致性最差。在另一组实验中, Xue 训练了一个条件随机场模型 (CRF), 验证了事件类型和情态与汉语时态有一定的内在联系。

除了能适用于特定语言和特定范围的情态词标注研究, Nissim^[15]提出了一种跨语言的数据驱动的分层标注模型, 目的是解决语言独立的问题。Nissim 指出, 以往的标注体系都是语言独立的, 对标注类型也没有统一的标准。因此他提出了一种用于情态标注的跨语言、自适应的、分级别和分层次的标注系统。其中, 跨语言是指依赖于大范围的类型学的驱动方法, 自适应是指能够应用于各种语言情态表达方式, 分级分层是指对事件真实性和说话者态度进行单独的标注, 对这两个大类分别进行更深层次的分类。Nissim 构造了一个标注框架, 该框架覆盖了形态学、词汇学和句法学等, 可以根据不同的语言自行调整。该标注方法对于语言多样性研究和情态的翻译问题提供了帮助。

3.2 情态词标注在机器翻译中的应用

在关于情态词的机器翻译研究中,学者们通常以情态词的标注为基础,结合情态词的语义语法等信息,对机器翻译系统进行改进。和 Pakray 类似, Baker^[16]等创建了一个公开的英语情态词和否定词列表,并做了两个自动标注器,一个基于字符串的标注器,一个基于结构的标注器,其中基于结构的标注器在标注 LDC 标准语料方面取得了更高的准确率。在之后的英语到乌尔都语(Urdu)的机器翻译中, Baker 采用了句法增强的机器翻译系统作为基准,在该系统中加入此前标注器标注出来的语义信息,包括情态、否定和命名实体信息,翻译结果总体上取得了 0.5 个 BLEU 值的提高。和 Baker 研究方法相似, Singh^[17]也创建了一个情态词词典,该词典中的每个情态词的每种含义均包括一个到多个的映射和至少一对源语言和目标语言例句。然后使用条件随机场进行学习,选出最好的特征集对机器翻译系统进行实验和评测,在英语到印地语(Hindi)的机器翻译中,翻译准确率最高达到了 63.93%。

在日英机器翻译中, Ideue^[18]和 Murata^[19]对情态词的翻译进行了特殊处理。Ideue 对日英翻译中的否定、肯定和疑问进行改进,在基于短语的机器翻译模型中加入一个特征函数,对源语言和目标语言中的情态词进行识别和计算,进而保证输入和输出句子中情态词的一致性。通过实验表明,该方法在肯定、否定和疑问的翻译准确率上取得了一定程度的提高。和 Ideue 持有相同的观点, Murata 认为,情态词的翻译就是在目标语言中把主要动词短语的表现形式给体现出来,也就是要保证输入和输出的情态一致性。在具体操作中, Murata 首先总结了日语中情态的一些表现形式,然后使用支持向量机对这些表现形式进行分类, Murata 使用了两组特征,其中一组是以日语输入句子中句末 1—10 个字符为特征。实验结果表明,在日语中,句末的字符串对于情态词的准确翻译有很大影响。同时, Murata 还对各个系统中出现的各种时态和情态的错误进行了详细的分析,对以后各个系统提高翻译准确率有很大的帮助。

通过对以上研究的分析,我们发现,情态词的研究在各个语言的机器翻译中均得到了很大的重视,情态词翻译正确与否影响到整个句子的翻译质量。研究者们在不同语言的机器翻译中,也试图从情态的方向对机器翻译系统进行错误分析和改进,尤其是对于源语言输入和目标语言输出在情态上保持一致方面。然而,由于情态词翻译的复杂性,目前对于情态词的研究还没有统一的结论,本文针对现代汉语中的常用情态词,从中英、中日两个翻译方向,对不同的四个翻译系统结果进行分析,总结出常见的翻译错误类型,并进一步提出可能的改进方向。

4 实验结果和错误分析

为了测试情态词在中英和中日方向对于翻译结果的影响,我们选取四个机器翻译系统进行评测,四个系统分别为 Moses、Baidu、Bing 和 Youdao。Moses^[21]是一个开源的统计机器翻译系统,由英国爱丁堡大学、德国亚琛工业大学等 8 家单位联合开发,现在已被业界广泛应用于构建基本的机器翻译引擎。在测试数据方面,我们首先构造了常用的中文情态词列表,共包含 142 个情态词,然后从语料中过滤出四组中文句子,分别为句子长度小于等于 15 个汉字的短句两组:包含情态词的短句和不包含情态词的短句;句子长度大于 15 小于等于 30 个汉字的长句两组:包含情态词的长句和不包含情态词的长句。每组测试数据均包含 115 句中文句子。之后对各个系统的翻译结果进行人工评价以及错误分析,得出具体的统计数据。

4.1 整体翻译质量

首先,我们从总体上对中英、中日两个方向包含情态词和不包含情态词的短句和长句的整体得分情况进行汇总,图 1、图 2 为合并各个系统翻译结果后的评分统计图。人工评测采用五分制:从五分到一分,译文质量逐渐降低,五分为完全正确,且符合语言习惯;一分为完全错误,跟原文意思相差甚远。

中英方向,包含情态词短句的平均分为 3.66,不包含情态词短句的平均分为 4.05 分,上升了 0.39 分;包含情态词长句的平均分为 3.37 分,不包含情态词长句的平均分为 3.88

分，上升了 0.51 分。

中日方向的机器翻译结果总体得分较低，包含情态词短句的平均分为 2.92 分，不包含情态词短句的平均分为 3.29 分，上升了 0.37 分；包含情态词长句的平均分为 2.74 分，不包含情态词长句的平均分为 2.99 分，上升了 0.25 分。

该结果表明，在中英、中日方向，情态词的加入均会影响翻译系统的翻译质量。从不同长度的句子评分增量来看，中英方向情态词对于长句的影响较大，而中日方向中，情态词则对于短句的翻译质量影响相对较大。

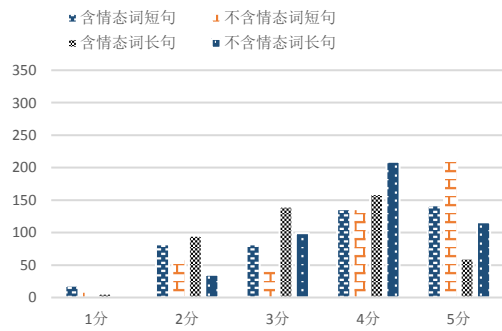


图1 中英人工评分统计图

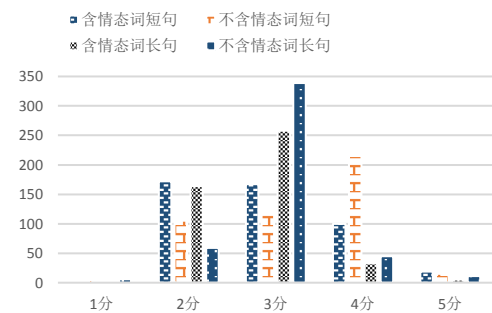


图2 中日人工评分统计图

4.2 各系统比较

从不同的翻译系统角度，我们继续比较中英、中日方向各个系统的翻译质量，分析在不同的翻译系统中，情态词对翻译质量的影响。图3、图4为各个系统中英、中日方向的评分示意图。

在中英、中日方向，从整体的趋势来看，各个系统中包含情态词长句的总体评分最低，不包含情态词短句的整体评分最高，可见句子长度和是否包含情态词均对翻译质量有较大的影响。

中英方向，在四个系统中，Moses 系统评分相差较大的是包含情态词长句和不包含情态词长句，平均分相差 0.64 分；中日方向四个系统中，Moses 的评分同样相差较大，最高的为包含情态词短句比不包含情态词短句平均分低 0.45 分。该结果表明，在句子长度一定的情况下，Moses 系统处理情态词的能力较弱，而其他在线系统对情态词的处理并没有 Moses 系统的差异明显。此外，对于中英方向，Moses 系统的翻译质量在四个系统中排名最低，而对于中日翻译，除 Baidu 外，Moses 系统反而翻译质量较高，我们推测，相对于对中英翻译的处理，Bing 和 Youdao 系统对中日翻译并没有做有针对性的提高，而且 Bing 和 Youdao 两个系统不管是在中英方向还是中日方向，其翻译质量都几乎在同一个级别。

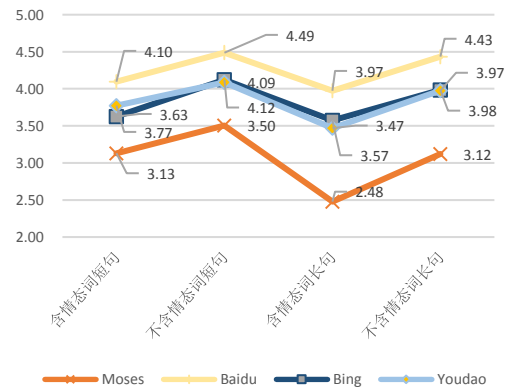


图3 各系统中英评分统计图

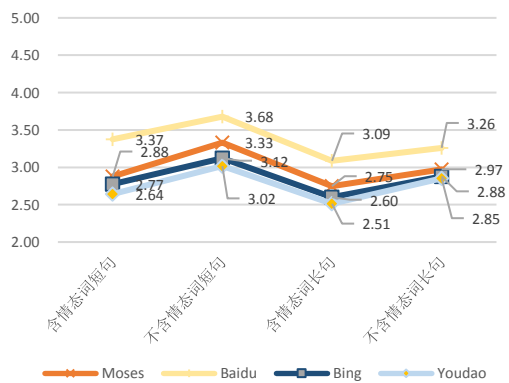


图4 各系统中日评分统计图

4.3 错误类型及分析

在对各系统进行人工评分的同时，我们也对各个系统的翻译结果所包含的错误类

型进行了标注。由于错误类型涉及到各个方面，比如句法层次、词汇层次以及语法层次等等，因此设定一个统一的错误类型标准是非常必要的，对将来改进机器翻译质量也能提供准确的依据。本文采用 Vilar 等^[20]构造的一个错误分析结构，并对该结构进行适当的增减，得到了一个两级的错误类型分类，如表 3 所示：

表 3 机器翻译译文错误类型

一级错误	二级错误
Missing words (缺词)	Content words (实词)
	Filler words (虚词)
	Copula words (系动词)
	Sub-sentence (子句)
Extra words (多词)	Content words (实词)
	Filler words (虚词)
	Copula words (系动词)
Incorrect words (错词)	Content words (实词)
	Filler words (虚词)
	Incorrect form (形态错误)
	Incorrect polarity (极性错误)
	Others (其他错误)
Word order (词序错误)	Local range (局部词序)
	Long range (整体词序)
Others (其他错误)	Punctuation (标点错误)

根据该错误类型分类，我们对四个翻译系统中的中英、中日方向的翻译结果进行人工分析，并对错误类型和数量进行统计。

从图 5、图 6 可以看出，在中英、中日方向的整体趋势中，不论是短句还是长句，包含情态词句子中总的翻译错误数量均明显高于不包含情态词句子中总的翻译错误数量。另外，长句中出现的总的错误数量均高于短句中的错误数量。

中英、中日方向，错误数量最多的均为缺词(Missing words)，其次为错词(Incorrect words)，然后是词序(Word order)和多词(Extra words)。各系统在中英、中日翻译中，词汇错误占绝大部分，缺词和错词的现象也比较严重。由于我们使用的 Moses 系统是基于短语的统计机器翻译模型，其在解码过程中倾向于选择代价最小的结果作为最终的翻译输出，因此倾向于输出较少词汇组成的句子，从而导致缺词的严重问题。从缺词现象来看，我们可以估计出其余三个在线系统也是以统计方法为主的机器翻译系统。

从错误类型和错误数量分析，Moses 系统在中英方向出现的错误数量略大于其在中日方向出现的错误数量，而在其他各系统中，中日方向的错误数量均明显高于中英方向的错误数量。该结果表明，在处理中日方向的翻译上，各个系统的翻译效果均没有中英好。此外，在中英、中日方向，Baidu 系统出现的总的错误数量均为最低。

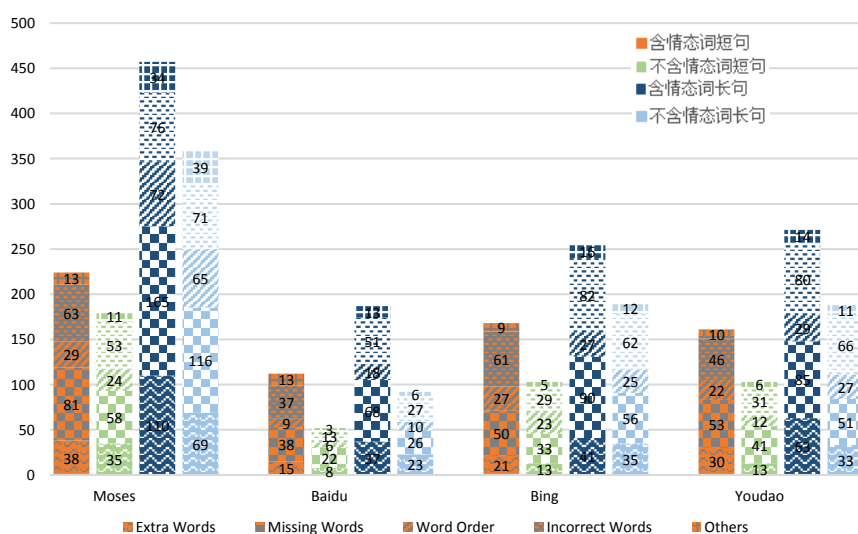


图 5 中英错误分析统计图

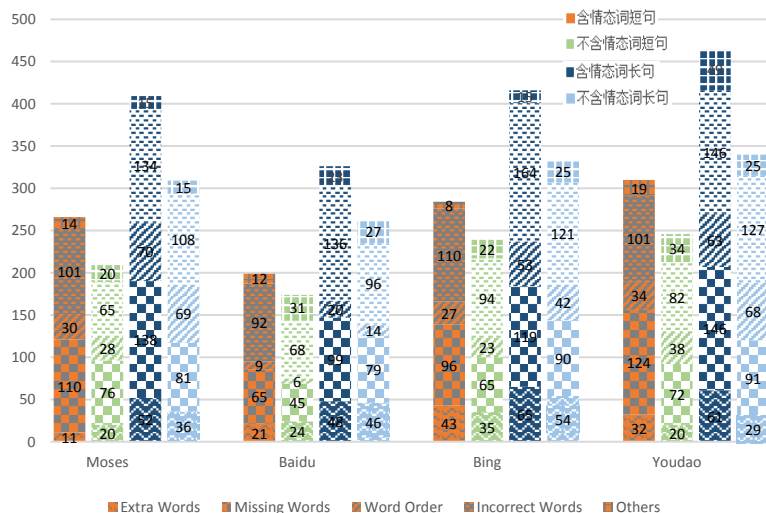


图6 中日错误分析统计图

4.4 具体错误分析

根据以上对各系统的性能和错误类型分析,我们发现四个机器翻译系统之间,在情态词的处理上并没有明显的不同,为了有针对性的提高翻译系统的翻译质量,我们进一步合并了句子长度和各系统的翻译结果,将四个系统中不同长度范围的句子所犯的最主要的四类错误进行汇总,目的是对包含情态词和不包含情态词的句子所出现的主要错误进行详细的分析。

出现最多的一类错误是缺词(Missing words),包含四种二级错误:实词(Content words)缺失,虚词(Filler words)缺失,系动词(Copula words)缺失和子句(Sub-sentence)缺失(图7)。

中英、中日方向中,相对于不包含情态词的句子,包含情态词句子的四类二级错误在总量上均有上升,错误最多的均为实词缺失,其次是虚词、系动词和子句的翻译缺失。

由于日语中不涉及系动词这一概念,中日方向中只有三种二级错误,相比于中英翻

译三种二级错误在数量上均有上升。相对于包含情态词的句子,实词缺失在总错误中的比例增加,虚词缺失在总错误中的比例减少。该结果表明,在中日方向中,情态词的加入对虚词缺失的影响大于对实词缺失的影响。中英方向则相反,情态词对于实词缺失的影响要大于对虚词缺失的影响。

第二类出现较多的翻译错误为错词(Incorrect words),包括五种二级错误,分别为实词选词错误(Content words),虚词选词错误(Filler words),形态错误(Incorrect forms),否定错误(Incorrect polarity)和其他错误(Others)(图8)。

在五种二级错误中,中英方向错误最多的是形态错误,不包含情态词句子出现的此类错误数量为228个,占此类错误的65%,包含情态词句子出现的错误数量为237个,占总错误的48%。两组数据在数量上相当,但在比例上,包含情态词句子所占的错误率比不包含情态词句子的下降了17个百分点,可见情态词对于词汇的形态和句子的时态

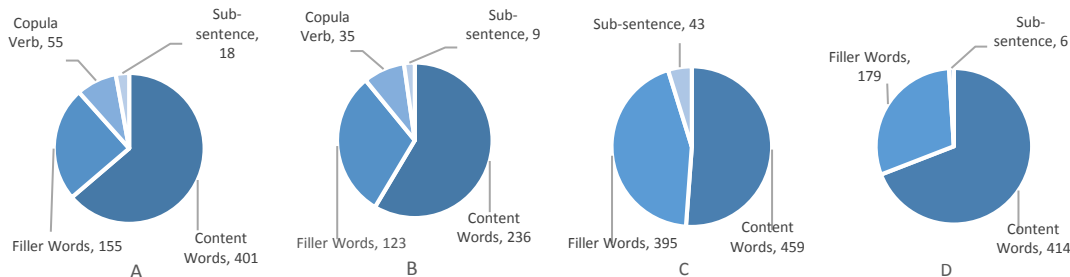


图7 Missing words 错误分布图 (A 中英包含情态词; B 中英不包含情态词; C 中日包含情态词; D 中日不包含情态词)

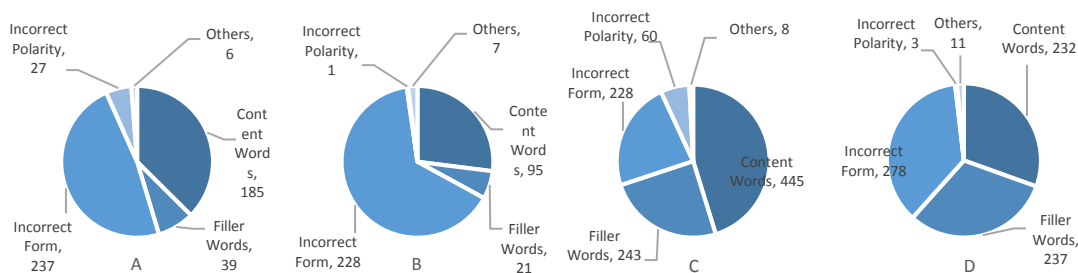


图8 Incorrect words 错误分布图 (A 中英含情态词; B 中英不含情态词; C 中日含情态词; D 中日不含情态词)

等是有较大的影响的。此前，Xue^[14]在对情态词进行标注的研究中，其实验结果也证实了在汉语中，情态和时态是相互关联的。对于否定极性错误，包含情态词句子出现了 27 个极性错误，而不包含情态词句子只出现了 1 个，该结果表明，情态词增加了翻译系统对极性（否定或肯定）处理的难度。

和中英的结果稍有不同，中日方向上包含情态词的句子在实词的选择上错误最多，在数量和比例上均占很大部分；虚词选择错误受情态词影响较小。在形态错误方面，包含情态词的形态错误有 228 个，占此类错误的 37%，不论是在数量上还是比例上，不包含情态词的形态错误均高于包含情态词的错误数量。该结果表明，和中英类似，在日语中，情态和词汇的形态与句子的时态也有内在的联系，情态词的加入对于词汇形态和时态的正确选择有一定的引导作用。在否定极性方面，中日方向的表现更为明显，包含情态词句子出现了 60 个极性错误，而不包含情态词句子只有 3 个。该结果表明，如果能改进情态词中否定词的翻译质量，那么中英和中日方向的翻译效果将会有较大提升。

第三类翻译错误为词序错误 (Word order)，包括局部词序错误 (Local range) 和整体词序错误 (Long range) (图 9)。

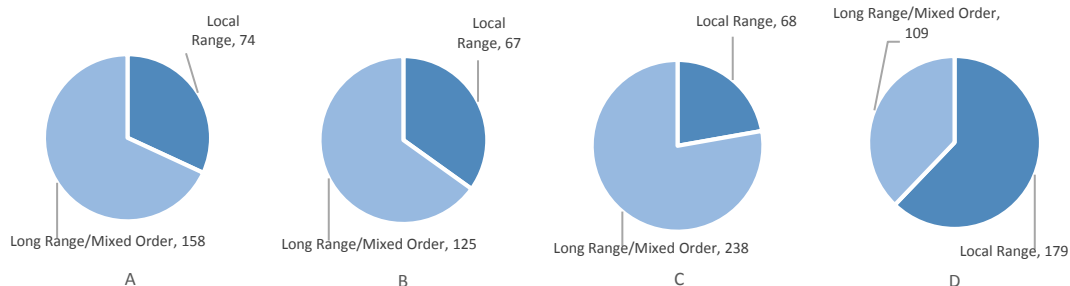


图9 Word Order 错误分布图 (A 中英含情态词; B 中英不含情态词; C 中日含情态词; D 中日不含情态词)

中英方向，局部词序错误稍多于整体词序错误，包含情态词的句子和不包含情态词的句子在这类错误中的比例相似，说明情态词对于词序的影响不明显。

中日方向，不包含情态词句子出现的局部词序错误要远远高于包含情态词句子中出现的局部错误，而整体上，不包含情态词句子词序错误数量低于包含情态词的句子。整体词序如果出现错误，将会导致整句话的可读性降低，因此，在日语中，情态词的处理不当更容易导致词序错误，降低翻译结果的可读性。

第四类翻译错误为多词 (Extra words)，包括冗余实词 (Content words)，冗余虚词 (Filler words) 和冗余系动词 (Copula words) (图 10)。

多词在句子翻译中的影响虽然不如缺词严重，但也对句子流利度产生大的影响，甚至影响对句子的理解。

中英方向，在三种二级错误中，错误最多的是冗余的实词，其次为冗余的虚词和系动词。相对于不包含情态词的句子，包含情态词的句子出现的这三种错误数量均有明显上升。

中日方向，由于日语中不涉及系动词，只有两种二级错误，冗余实词和冗余虚词，

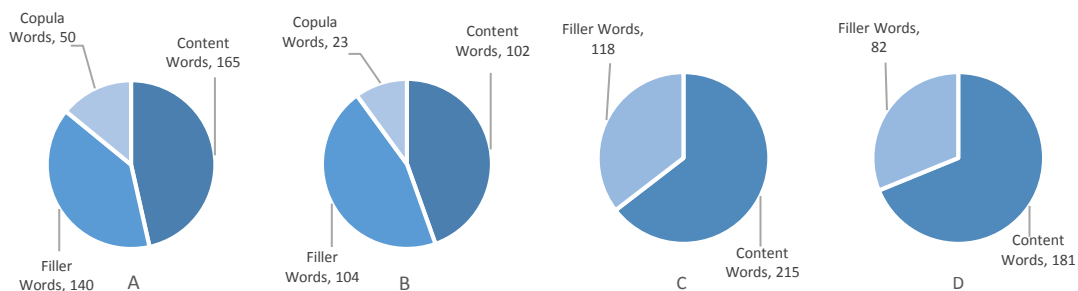


图 10 Extra Words 错误分布图 (A 中英含情态词; B 中英不含情态词; C 中日含情态词; D 中日不含情态词)

相对于不包含情态词的句子, 包含情态词句子中两者的数量也均有上升。结果表明, 在中英、中日方向, 情态词的加入均增加了系统的多余词汇, 对实词的影响均大于对虚词的影响。

4.5 改进方案

从结果分析可以看出, 中文情态词在翻译中的作用比较复杂, 目前主流的机器翻译系统并没有针对情态词做特殊的处理。由于情态词的加入, 增加了翻译的难度, 使缺词, 词序和多词的错误数量明显上升。而对于错词, 由于情态词和词汇形态以及句子时态有一定的内在联系, 包含情态词的句子在翻译时受上下文的影响较大, 对翻译质量的影响不如前三类错误明显。

从详细分析结果我们推测, 人工评测的四个机器翻译系统均为基于统计的机器翻译系统, 由于基于统计的机器翻译系统是建立在对大量的训练语料学习的基础上, 在机器翻译训练阶段, 我们可以尝试对是否包含情态词的句子进行区分训练, 同样, 在解码阶段使用不同的训练模型区分对待包含情态词的句子或者不包含情态词句子, 或者将分别训练的模型利用领域适应方法^[2]——线性插值或者实例加权 (Instance Weighting)——得到新的翻译系统, 从而提高包含情态词句子的翻译质量。

除了对训练语料进行区分之外, 我们也可以对待翻译句子的区分 (是否包含情态词) 进行进一步的扩展, 在翻译之前, 我们可以预先估计待翻译句子的翻译难度, 在翻译的过程中, 根据不同难度的输入句子采用不同的翻译系统, 比如使用不同的翻译模型 (如句法增强的模型), 甚至使用基于规则

的系统等, 其中句子长度也可以作为翻译难度估计的特征之一。

此外, 在训练语料预处理方面, 由于情态动词、情态副词的多样性, 我们可以根据情态动词和情态副词的情态分类, 利用重述方法将复杂的情态动词或情态副词归一化成较为简单的普通动词或者副词, 通过该方法减少情态词的多义性, 使情态词的情态意义更加有针对性, 进而提升翻译结果的准确率。同样, 在对翻译结果进行后处理时, 再利用重述方法进行反向转写。

在保持源语言输入句子和目标语言输出句子的情态一致性方面, 对输入句子进行情态标注, 对输出译文进行基于规则或者基于统计的情态修正, 确保源语言输入句子和目标语言输出句子在情态上的相呼应, 这些都可以作为进一步提高机器翻译系统翻译中文情态词句子质量的候选方案。

5 结论

本文通过对包含情态词和不包含情态词的四组中文句子的机器翻译结果进行人工评价和错误分析, 探讨了情态词对于中英、中日机器翻译系统的影响。结果表明, 不管是中英方向, 还是中日方向, 句子长度和是否包含情态词均对翻译质量有较大的影响, 包含情态词的句子中出现的翻译错误均明显高于不包含情态词的句子, 其中, 四种常见错误类型为: 缺词、错词、多词和词序, 我们进一步对四种错误类型的分布和对两种语言对的不同影响进行了分析, 针对这些统计现象, 我们提出了可能的解决方案, 目的是提高机器翻译系统的翻译质量。

然而, 由于现代汉语中情态的表达方式多种多样, 所涉及的语法现象也没有统一的

理论体系支撑, 情态词的分类界限模糊, 分歧较多, 因此, 本文的研究可能存在覆盖面不足的问题, 错误类型统计也可能不完全。在机器翻译方面, 本文针对主流的翻译系统(基于统计的机器翻译系统)做了主要的分析, 并没有针对其他类型的系统(如基于规则的机器翻译系统)进行分析, 这些都是将来我们需要改进的方向。

参考文献

- [1]. 张喜洪. 现代汉语情态范畴初论[D]. 四川师范大学硕士学位论文, 2008
- [2]. Philipp Koehn, et al., Moses: Open Source Toolkit for Statistical Machine Translation, Annual Meeting of ACL, demonstration session, Prague, Czech Republic, June 2007.
- [3]. J. Bybee & S. Fleischman. Issues in Mood and Modality: Introductory essay [A]. Modality in Grammar and Discourse[C]. 1995
- [4]. Asher & Simpson. The Encyclopedia of Language and Linguistics [C]. Oxford: Pergamon Press, 1994
- [5]. Palmer. Mood and Modality [M]. Lon-don: Cambridge University Press
- [6]. 朱德熙. 语法讲义[M]. 北京: 商务印书馆, 1982
- [7]. 朱冠明. 情态与汉语情态动词[J]. 山东外语教学, 2005(2)
- [8]. 彭利贞. 现代汉语情态研究[D]. 上海: 复旦大学博士学位论文, 2005
- [9]. 苏秋菊. 《现代汉语词典》(第6版)情态副词研究[D]. 河北师范大学硕士学位论文, 2014
- [10]. 李剑影. 现代汉语能性补语的语法化[J]. 辽宁工程技术大学学报, 2009
- [11]. P. Pakray, P. Bhaskar, et al., Workshop on Question Answering for Machine Reading Evaluation Clef Labs & Workshop, 2012
- [12]. Quaresma, P., Mendes, et al., Automatic tagging of modality: identifying triggers and modal value[C]. Proceedings of the 10th Joint ACL SIGSEM-ISO Workshop on Interoperable Semantic Annotation, 2014
- [13]. V Prabhakaran, M Bloodgood, M Diab, et al., Statistical Modality Tagging from Rule-based Annotations and Crowdsourcing [C]. Proceedings of the ACL Workshop on Extra-Propositional Aspects of Meaning in Computational Linguistics, 2012
- [14]. Nianwen Xue, Yuchen Zhang, Buy One Get One Free: Distant Annotation of Chinese Tense, Event Type and Modality[C]. Proceeding of the 9th LREC(International Conference on Language Resources and Evaluation), 2014
- [15]. M Nissim, P Pietrandrea, et al., Cross-linguistic annotation of modality: a data-driven hierarchical model[C]. Proceedings of IWCS 2013 WAMM Workshop on the annotation of modal meaning in natural language Association for Computational Linguistics
- [16]. K Baker, M Bloodgood, BJ Dorr, et al., Use of modality and negation in semantically-informed syntactic MT [J]. Computational Linguistics, 2012
- [17]. AK Singh, S Husain, H Surana, J Gorla, D Misra, Disambiguating Tense, Aspect and Modality Markers for Correcting Machine Translation Errors[C]. Proceeding of Recent Advances in Natural Language Processing (RANLP), 2007
- [18]. M Ideue, K Yamamoto, M Utiyama, E Simita, et al., Modality-Preserving Phrase-Based Statistical Machine Translation[C]. Proceeding of the ICAL, 2012
- [19]. M Murata, K Uchimoto, Analysis of machine translation system's errors in tense, aspect and modality [C]. Proceeding of PACLIC, 2005
- [20]. David Vilar, Jia Xu, et al., Error analysis of statistical machine translation output[C], Proceeding of the International Conference on Language Resource & Evaluation, 2010