

Research on Uyghur-Japanese Machine Translation Method Based on Grid Additive Components

Simayijiang·Aimaier

Xinjiang Jiashi No. 1 Middle School, Jiashi, Xinjiang, 844300, China

Abstract

This paper based on the Uyghur-Japanese grammatical case suffixes, mainly discusses the machine translation methods of Uyghur and Japanese, and validates the methods and its feasibility and effectiveness by experiments. Firstly, an analysis carries on the grammatical case forms of Uyghur and Japanese; secondly, comparatively studies the Uyghur and Japanese cases; finally, constructs a corresponding rules database of Uyghur and Japanese case suffix forms, and then design a corresponding algorithm for Uyghur and Japanese case suffixes.

Keywords

case suffix; machine translation; Uyghur; Japanese

基于格附加成分的维吾尔语—日语机器翻译方法研究

斯马依江·艾买尔

新疆伽师县第一中学, 中国·新疆 伽师 844300

摘要

论文基于维吾尔语—日语名词格附加成分, 探讨维吾尔语—日语机器翻方法, 对该方法进行验证性实验, 验证该方法的可行性及有效性。首先, 对维吾尔语格附加成分及日语格附加成分进行分析; 其次, 对维吾尔语和日语的格附加成分进行比较研究; 最后, 构建维吾尔语和日语格附加成分对应规则库, 设计维吾尔语—日语格附加成分对应算法。

关键词

格附加成分; 机器翻译; 维吾尔语; 日语

1 引言

维吾尔语和日语都是阿尔泰语系的成员, 是黏贴语。黏贴语中的助词是很发达, 语言构造上起着重要作用。笔者认为黏贴性这个特点不仅仅是维吾尔语和日语具有的特点, 而且韩语、土耳其语, 蒙语等阿尔泰语系中每一个语言的共同特点^[1]。维吾尔语—日语格附加成分研究是维吾尔语—日语机器翻译系统中的一项基础性工作。到目前为止, 在维吾尔语—日语机器翻译中格附加成分相关研究较少。论文有效地利用维吾尔语和日语相近的特点, 针对维吾尔语—日语机器翻译系统进行研究。

例如, (1) [مەن مەكتەپكە باردىم]和(2) [مەن مەن باردىم]是在维吾尔语中正确的语句, 这些特点在日语中也同样存在。无论是将(1)直接翻译成“私は学校まで(行き)ました”也可以换成“**مەن**”(我)和“**مەكتەپ**”(学校)将(2)翻译成“学校まで私は行きました”, 同样正确, 表达意思一样。

将维吾尔语句子翻译成日语时, 以词汇为单位进行翻译, 但是词序不用调整。论文中维吾尔语句子的分析结果进行逐词翻译方法, 研发了维吾尔语—日语机器翻译系统。在(2)中, 将重点聚焦在动词的角色角度来翻译。

由于维吾尔语与日语词序相同, 主要难点就是格附加成分的分析, 这亦是维吾尔语和日语具有的一样特点之一。然而, 这就说明将格附加成分不能准确翻译时, 翻译内容模糊不清。

我们首先要讨论维吾尔语—日语机器翻译的格附加成分问题。格附加成分不仅存在于维吾尔语中, 日语亦同样存在。虽然在维吾尔语的格附加成分和日语的格附加成分中能找出互相对应关系, 有存在多义性关系, 有些句子中多个日语的格附加成分和维吾尔语的格附加成分相互对应^[2]。论文不仅讨论格附加成分翻译, 还讨论格附加成分的相互多义性关系, 提出适宜的格附加成分的翻译方式。

建立以高效的维吾尔语—日语机器翻译系统为目的, 提

出格附加成分方式翻译动词的形态。先利用计算机建立维吾尔语常用维吾尔语—日语动词词典^[3]，对两种语言格附加成分对应问题进行详细研究，并获取动词格形态变化对应表，且使用该表实现格附加成分的变化处理。

在维吾尔语—日语格附加成分相互对应，维吾尔语中有些格附加成分。例如，像格附加成分“**كە**”，根据语境变化**كە، قا، غا**等，这说明向格附加成分在句子中表示语言环境。在日语中不存在，但维吾尔语中存在的人称尾词，如[私はご飯を食べました。](我吃过了)“**تاماق يېدىم مەن**”与[彼はご飯を食べました。](他吃过了)“**ئۇ تاماق يېدى**”。从这两个句子可以看出来，日语中如同下划线部分表示同样，各自有不相同的人称尾词。

2 维吾尔语和日语的格附加成分比较研究

2.1 维吾尔语格附加成分分析

在现代维吾尔语中，附着在名词后面连接的附加成分**تە، دە، ئا، ئا، كە، گە، قا، قا، تەن، دىن، نى، نىڭ**等，在现代维吾尔语语法中一直被叫作“格附加成分”^[4]。近年来，维吾尔语中的**دىكى، كىچە، گىچە، قىچە، غىچە، چە، تىكى، تەك، دەك، چىلىك**等语言成分也开始被叫作“格附加成分”^[5]。

2.2 日语格附加成分分析

2.2.1 格附加成分

日语的单词本身没有格的变化，表示某一个体言在句中属于哪种格（即决定该词与其他词的关系，亦即该词在句中的地位）的助词叫格附加成分。重要的格附加成分共有9个：**が、の、を、に、へ、と、から、より、で**等，接在体言（日语中具有实质或形式上的意思的独立词，包括名词、代名词、数词等）后面，也可以接在相当于体言性质的词或词组后面^[6]。

2.2.2 分类

主格附加成分**が**，连体格附加成分**の**，连用格附加成分（动词、形容词、形容动词总称为用言）**を、に、へ、と、から、より、で**^[7]。

维吾尔语和日语都是黏贴语，助词的作用极为重要，尤其是格附加成分，提到决定某个词在句子中的功能，或决定该词和另外词的相互关系等方面，指定句子构建起重要作用。

2.3 维吾尔语日语格附加成分比较分析

如果有[ئۇ](他)、[كۈتۈپخانا](图书馆)、[كىتاب](书)、

[ئوقۇ](读)等四个词语，将此四个单词排列，这些一排单词都未能表达任何意义，它们不能成句子。通过利用格附加成分[دا]，[نى]，它们形成了符合维吾尔语语法的句子。这样它们的功能就会更加明显。

ئۇ كۈتۈپخانىدا كىتابنى ئوقۇۋاتىدۇ (他在图书馆读书。)
(3) (2) (1)

维吾尔语的格附加成分在日语中的作用上是一样的，该句子译成日语，结果如下所示，可以显示格附加成分都相互对应。

彼は ϕ 図書館で 本を 読みます
(1) (2) (3)

这在维吾尔语—日语机器翻译系统中是很重要的部分。在进行翻译时，词序按照原样处理，只需要加上格附加成分互相的对应就可以。因此，在做完词素分析时，通过逐词逐句地译出包括有格附加成分在内的各单词即可。

3 维吾尔语—日语格附加成分互相对应关系研究

3.1 维吾尔语日语格附加成分对应规则构建

维吾尔语和日语的格附加成分不一定能一一对应或翻译，例如，在[مەكتەپتىن باردىم]（我从学校过去）“学校から行きました”和[ئۆيىدىن باردىم]（我从家里过去）“家から行きました”这种句子中，将格附加成分[دىن، تەن]翻译为“から”。总而言之，与维吾尔语的[دىن، تەن]互相对应的日语格附加成分是“から”，对日语来讲，多个对应一个关系。

显示一一对应一关系的维吾尔语—日语格附加成分，如表1所示。

表1 维吾尔语—日语格附加成分中一一对一关系

维吾尔语格附加成分	ϕ	نىڭ
日语格附加成分	か	の

由表1可以看出，主格、方向格等两个名词格一个对应一个。

显示多对一的格附加成分，如表2所示。

表2 维吾尔语—日语格附加成分中多对一关系

维吾尔语格附加成分	دىن	تىن	غىچە	گىچە	كىچە	قىچە
日语格附加成分	を		まで			

由表2可知,多个维吾尔语格对应一个日语格。

显示多对多的格附加成分,如表3所示。

表3 维吾尔语—日语格附加成分中多对多的关系

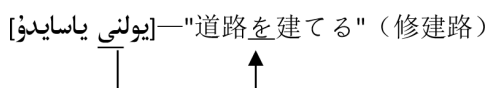
维吾尔语格附加成分	كە،قا،گە،كە	دا،تا،دە،تە	دەك،تەك
日语格附加成分	へ に	で には のよおな のよおに	

由表3可以看出,维吾尔语的**غا,قا,گە,كە**等四个格对应日语的两个格**دا,دە,تا,تە**等四个格对应日语两个格**دەك,تەك**等对应日语两个格。

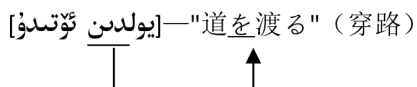
3.2 维吾尔语日语格附加成分对应算法

如上所述,多个维吾尔语格附加成分和一个日语的格附加成分互相对应。因此,在维吾尔语—日语机器翻译系统中,除了进行概念词的翻译以外,还必要进行合适的格附加成分翻译。

利用动词与格附加成分的密切关系,提出解决这些关系的方法。格附加成分取决于依靠着前面的名词,还在很多情况下和名词性动词构成的句子有所关系^[8]。比如,在**[يول]**名词翻译时,如果**[يولنى ياسايدۇ]**,由**[نى]**是表达对象,进行翻译成日语的“を”,全句翻译成为“道路を建てる”。



如果**[يولدىن ئۆتىدۇ]**,由于**[دىن]**表达经过的意思,在日语中可译为“を”,全句翻译成为“道を渡る”。



对每个动词来讲,首先要登记每一个动词和它的格形态和互相对应格附加成分翻译,在进行格附加成分翻译的时,通过包括格附加成分的名词有关的动词信息,可以对应的译词进行选定。

从上述研究中可以看出,维吾尔语的一个格附加成分与日语的格附加成分的主要表达意思互相对应。发现次数最多的日语格附加成分初始默认状况的翻译词要登记,只要必须对初始默认状况错译出来的格附加成分必要进行翻译的词替换,就能实现此格附加成分的选定。

将发现次数最多的作为各格附加成分初始默认状况的翻

译词要登记,当初始默认状况翻译出现错译时,通过此名词句有关的动词性格形态,替换格附加成分翻译的词。

出现以下情况需要进行讨论:

(1)对维吾尔语句进行词素分析。

(2)在翻译对象的维吾尔语句中的全部单词,只需登记一个日语译词到维吾尔语—日语词汇对应库中。

这次,尚未注重格附加成分未有直接关系的单词译词的多义性。即使在此种情况下,与动词互相对应恰当的译词也只提供一个单词,选出与其它对应的每个格附加成分。在这样情况下,将从维吾尔语的格附加成分变化处理成日语的格附加成分步骤总结如下:

步骤1:对翻译对象的维吾尔语句进行分词。

步骤2:从分词过的各维吾尔语单词中提取词干和格附加成分,并对格附加成分供给初始自然状况的译词。

步骤3:从维吾尔语—日语对应库找出互相对应的词干和格附加成分。

步骤4:将词干和格附加成分连接生成日语语句。

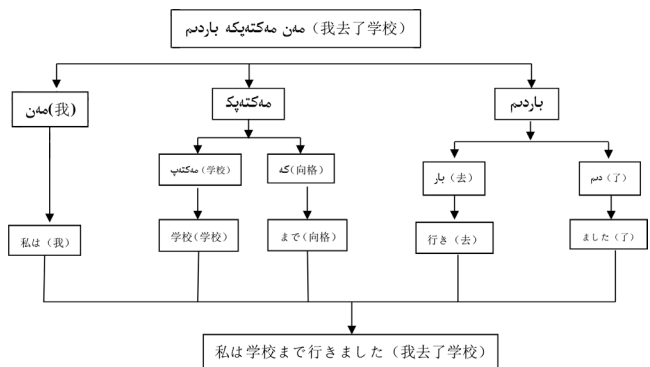


图1 维吾尔语—日语翻译实例

首先输入句子进行分词(步骤1)提取词干和格附加成分(步骤2)同时进行。在此个阶段,格附加成分**[نى]**被翻译成成为初始默认状况的“まで”。

在步骤3中,从已经建立的维吾尔语—日语对应词典中找出相应的词干和格附加成分,如果尚未找到对应的信息。例如,**[ئوينىۋاتىدۇ]**表达的意思是“正在玩儿”,**[ئوينىۋاتىدۇ]**表达的意思是“玩儿”,这个词可以找到,而找不到**[ئوينىۋاتىدۇ]**,则用派生法在词干**[ئوينىۋات]**的基础上可以形成派生新词**[ئوينىۋاتىدۇ]**。在动词**[بار]**的日语译词中,以“行き”的形式格形态的译词,通过格附加成分译词替换处理。在这里,相关修饰与被修饰关联的判断。在动词的前面发现的格附加成分之内,将表现格形态的信息格附加成分替换。如果同样

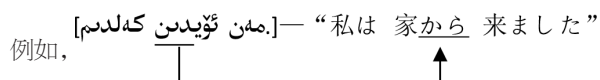
格附加成分发现多个时,只替换与此动词比较近的格附加成分。这次考虑修饰与在被修饰中非交叉性、后方修饰性原则。

4 实验结果分析

4.1 测试方法

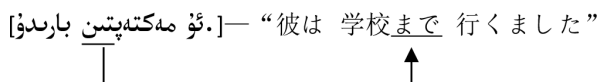
论文研究中,通过包含各自格附加成分的句子进行实验,实验过程中对系统自动翻译的句子和人工翻译的句子进行互相对比。首先要人工翻译然后本系统翻译进行对比。这些句子的来源是日常生活应用的简单句子。

例如, [مەن ئۆيىدىن كەلدىم.] — “私は 家から 来ました”



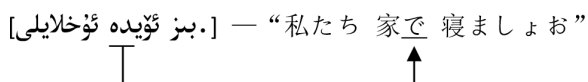
(我从家里过来了。)

[ئۇ مەكتەپتىن باردۇ.] — “彼は 学校まで 行きました”



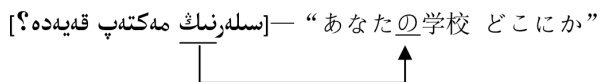
(他从学校过去。)

[بىز ئۆيىدە ئۇخلايمىز.] — “私たち 家で 寝ましょ”



(我们在家里睡觉。)

[؟، سىلەر نىڭ مەكتەپ قەيەدە؟] — “あなた どの学校 どこにか”



(你们学校在哪儿?)

4.2 测试结果分析

维吾尔语中存在主格附加成分,符号“ ϕ ”表示空白。维吾尔语中主格附加成分“ ϕ ”对应日语的格附加成分“か”。维吾尔语的领属格附加成分“نىڭ”与“の”,时位格附加成分“دە”与“で”,从格附加成分“دىن”与“から”,界限格附加成分“گىچە”与“まで”,量似格附加成分“چىلىك”与“ように”相似度和互相对应度比较高。宾格附加成分“نى”与“を”,向格附加成分“گە”与“に”,形格附加成分“دەك”

”与“のように”对应度较低。

5 结语

论文研究了维吾尔语和日语的格附加成分的互相对应关系。论文研究显出也从另一个观念显示这两种语言的相近度,通过维吾尔语—日语对应词典和两个语言的格附加成分对应规则方法,提到了选择格附加成分翻译词的方式,明确了将其合成到维吾尔语—日语的分词系统的目的。另外,利用实验指出了其有效性。本研究虽然以规模较小的语料库进行,却很好地显示了在这里所示的变换问题处理方法的有效性。今后研究将以实现质量较高的机器翻译系统作为目标,继续通过以规模较大的预料数据为实验对象,进行下一步的研究。

参考文献

- [1] 维尼拉·木沙江. 日雄机器翻译系统中粘着性特点的应用 [J]. 新疆大学学报, 2005(33):129-134.
- [2] 维尼拉·木沙江. 日-维机械翻译系统中格助词的研究 [J]. 新疆大学学报:自然科学版, 2008(01):96-102.
- [3] IPLA. 计算机日语 IPA 动词词典(基本动词)[Z]. 日本信息技术促进会, 1987.
- [4] 程适良. 现代维吾尔语语法 [M]. 乌鲁木齐:新疆人民出版社, 1998.
- [5] 武志, 米泽明宪. 有关日语词素分析系统的词素语法 [J]. 自然语言处理, 1985(04):27-29.
- [6] 马和夫(编). 日本语的起源——有关世界的语言学者论集 [M]. 日本:武藏野书院, 1986.
- [7] 金田一春彦, 梅棹忠夫. 日本語大辞典 [Z]. 1989.
- [8] 小川泰弘. 关于日本语动词派生语法的生成 [J]. 自然语言处理, 2007(03):57-77.