

# 派生文法に基づく日本語－ウイグル語機械翻訳 — 動詞接尾辞の変換 —

小川泰弘<sup>†</sup> ムフタル・マフスット<sup>‡</sup> 外山勝彦<sup>§</sup> 稲垣康善<sup>†</sup>

<sup>†</sup>名古屋大学工学部情報工学科 <sup>‡</sup>三重大学工学部情報工学科 <sup>§</sup>中京大学情報科学部情報科学科

<sup>†</sup>〒464-01 名古屋市中種区不老町 <sup>‡</sup>〒514 津市上浜町1515 <sup>§</sup>〒470-03 豊田市貝津町床立1011

yasuhiro@inagaki.nuie.nagoya-u.ac.jp, muhtar@inagaki.nuie.nagoya-u.ac.jp

toyama@sccs.chukyo-u.ac.jp, inagaki@nuie.nagoya-u.ac.jp

あらまし日本語とウイグル語は共に膠着語と分類され、構文的に類似した点が多い。したがって、日本語からウイグル語への機械翻訳を行うためには、日本語入力文の形態素解析が終了した段階で、各単語をウイグル語の対応する訳語に置き換えるという手法が考えられる。本稿では、特に両言語において文の中心的役割を果たす動詞句の翻訳手法について述べる。従来の文法では、日本語には活用があるが、ウイグル語には活用がないとされてきた。しかし、派生文法を用いると、日本語も活用を用いることなく記述でき、単純で精度の高い機械翻訳が可能となる。また、ウイグル語特有の文法現象についても、それが現れる文を生成する手法を示す。さらに、この手法に基づき日本語-ウイグル語機械翻訳システムを作成した。このシステムでは、日本語形態素解析システムとウイグル語整形システムは、それぞれ独立のモジュールとした。これにより他の膠着語への翻訳にも応用可能なシステムとなっている。

キーワード 機械翻訳、派生文法、膠着語、日本語、ウイグル語

## Japanese-Uighur Machine Translation based on Derivational Grammar — A Translation of Verbal Suffixes —

OGAWA, Yasuhiro<sup>†</sup> Muhtar, Muhsut<sup>‡</sup> TOYAMA, Katsuhiko<sup>§</sup> INAGAKI, Yasuyoshi<sup>†</sup>

<sup>†</sup>Dept. of Information Eng., Nagoya Univ. <sup>‡</sup>Dept. of Information Eng., Mie Univ.

<sup>§</sup>Dept. of Computer Sci., Chukyo Univ.

<sup>†</sup>Furo-Cho, Chikusa-Ku, Nagoya, 464-01, JAPAN

<sup>‡</sup>1515 Uehama-Cho, Tsu, 515, JAPAN

<sup>§</sup>101 Tokodate, Kaizu-Cho, Toyota, 470-03, JAPAN

yasuhiro@inagaki.nuie.nagoya-u.ac.jp, muhtar@inagaki.nuie.nagoya-u.ac.jp

toyama@sccs.chukyo-u.ac.jp, inagaki@nuie.nagoya-u.ac.jp

Abstract Since Japanese and Uighur languages are *agglutinative languages*, they have many syntactic similarities. Thus we can translate from Japanese into Uighur languages by replacing Japanese words with corresponding Uighur words after Japanese morphological analysis. In this paper, we discuss a translation method of verbal clauses that play an important role in sentences. By using the derivational grammar which formalizes Japanese verbs syntax without conjugation, we propose a simple and systematical translation system. The system consists of two independent modules and so we can apply this approach to other agglutinative languages.

key words machine translation, derivational grammar, agglutinative language, Japanese, Uighur

## 1 はじめに

日本語-ウイグル語機械翻訳においては、その構文的類似性を利用して、形態素解析が終了した段階で各単語を逐語訳するという手法が考えられる [3]。しかし、文の中心的役割を果たす動詞は、日本語では活用するが、ウイグル語では活用しないと考えられてきたため、日本語の活用を考慮した翻訳手法とする必要があった。

しかし派生文法 [7][8] によれば、日本語も活用しないとされる。派生文法は日本語の膠着語としての性質に着目した文法であり、派生接尾辞や統語接尾辞を利用して、活用を用いることなく日本語の用言の変形を体系的に扱うことができる。また、同じ膠着語であるウイグル語を記述することも可能である。翻訳の対象となる2つの言語を共通の文法で記述すれば、単純でかつ精度の高い機械翻訳が可能になると考えられる。そこで、本稿では、日本語とウイグル語の両言語で中心的役割を果たす動詞句を共に派生文法で記述した上で、逐語訳による機械翻訳を行う手法を提案する。特に、人称接尾辞などウイグル語特有の文法的性質の記述と処理方法についても示す。

さらに、形態素解析システム MAJO [1][2] を利用して日本語-ウイグル語機械翻訳システムを作成した。MAJO は派生文法に基づいて日本語の形態素解析を行うシステムである。今回の機械翻訳システムでは MAJO で使用する辞書を、日本語-ウイグル語辞書に交換した。また MAJO の出力結果を整形する別モジュールを作成した。このモジュールはウイグル語特有の性質を再現するものであり、これにより自然なウイグル語の訳文を生成できる。このように、形態素解析部分とウイグル語整形部分を独立に作成したため、派生文法で記述できれば、他の膠着語との機械翻訳へも応用可能であると考えられる。

なお、本稿で使用する派生文法は音韻論的手法の一種であり、音素単位で解析するため、日本語の表記の一部にローマ字を用いる。このため日本語、ウイグル語の両者ともローマ字で表記され混同しやすいため、日本語の単語は「」、ウイグル語の単語は“ ”で囲んで区別する。

## 2 日本語-ウイグル語逐語訳

日本語とウイグル語は語順がほぼ同じであるなどの点で構文的類似性が高い。そのため日本語-ウイグル語機械翻訳においては、構文解析を行わず、日本語入力文の形態素解析を行った段階で各単語を対応するウイグル語単語に置き換える手法が考えられる。図1にその例を示す。

しかし、従来の文法では、日本語の動詞は活用すると考えられてきたため、その活用を処理する必要があった。図2に動詞句の翻訳例を示す。ここで「作られる」を翻訳した際には、活用語尾「ら」に対応するウイグル語の単語は存在しない。しかし、活用語尾には対応するウイグル語の単語が常に存在しないわけではなく、「作る」を翻訳した場合には、活用語尾「る」に“-ydu”が対応している。このように活用語尾に関して、ウイグル語との対応は、きめ

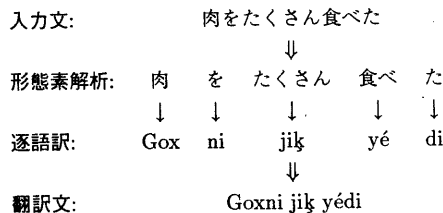


図1: 日本語-ウイグル語逐語訳

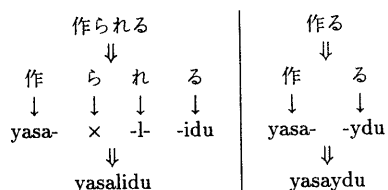


図2: 動詞句の翻訳

細かい処理が必要である。

これに対して [3] では、推移グラフを導入して活用形を処理した。また、推移グラフは動詞と助動詞の複雑な接続を処理することが可能である。しかし、推移グラフは開始ノードが未然形や連用形に対して異なり、また同じ名前のついた推移が数多く存在して複雑であった。

本稿では、日本語は活用しないとする派生文法を利用することにより、動詞句を簡潔に翻訳できることを示す。

## 3 派生文法による日本語の記述

従来の文法では、日本語の用言は活用することを前提にしており、学校文法では、その活用形は未然形、連用形、終止形、連体形、仮定形、命令形の6つに分類されている。

一方、日本語は言語の分類において膠着語であるといわれている。膠着語とは、文法的機能を表す接辞が実質的観念を表す語幹に結合することによって単語を形成するという性質をもつ言語の総称である。

ゆえに、膠着語である日本語が活用することは矛盾であると捉えられ、活用を前提としない派生文法 [7][8] が提案されている。派生文法による日本語の記述方法を、その特徴が表れる動詞句の形成について以下にまとめる。

### 3.1 連結子音と連結母音

動詞の不変化部分を語幹と呼ぶ。一段活用動詞「見ル」「食ベル」などの場合は、不変化の部分「見」「食ベ」が語幹であり、その際、語幹は母音 i か e のいずれかで終わる。また、五段活用動詞の場合には、「書カ」「書キ」「書ク」「書ケ」「書コ」のように、末尾が変化するとされる。これは音韻論的に考えれば「kak-a」「kak-i」「kak-u」「kak-e」「kak-o」であり、「kak」を語幹として取り出すことができ、語幹は子音で終わる。

そこで、一段活用動詞のように母音で終わる語幹を母

表 1: 動詞と接尾辞の接続例

活用形	子音幹の例	母音幹の例	接尾辞
未然形	kak-ana-i	tabe-na-i	-(a)na-i
連用形	kak-imas-u	tabe-mas-u	-(i)mas-(r)u
終止形	kak-u	tabe-ru	-(r)u
連体形	kak-u	tabe-ru	-(r)u
仮定形	kak-eba	tabe-reba	-(r)eba
命令形	kak-e	tabe-ro	-e / -ro,-yo

音幹と呼び、五段活用動詞のように子音で終わる語幹を子音幹と呼ぶ。

派生文法においては、動詞の変形は動詞の語幹に接尾辞が接続したものと考える。そのため従来の文法でいう活用形の語尾や助詞、助動詞をすべて接尾辞として扱う。それらを従来の活用の形に対応させると表1のようになる。

ここで、終止形「kak-u(書く)」「tabe-ru(食ベル)」の場合、接尾辞はそれぞれ「-u」「-ru」である。派生文法ではこれを「-(r)u」とまとめて表記する。子音 r の有無は動詞語幹の末尾に依存して決まる。例えば「kak-」に「-ru」が接続した場合、子音が連続するため接尾辞の先頭の r が欠落する。そのような子音を連結子音と呼ぶ。

一方、否定の助動詞「ナイ」が接続する場合を考えてみる。これは未然形に接続する接尾辞である。派生文法では「書カナイ」「食ベナイ」は表1に示すように「kak-ana-i」「tabe-na-i」と解析する。ここで、否定を表す接尾辞は「-(a)na-」の形で表され、母音 a は母音が連続する場合に欠落する。そのような母音を連結母音と呼ぶ。

以上から、派生文法においては、動詞接続は以下の2つの規則を基本として行われる。

**接続規則 1:** 子音幹に連結子音を持つ接尾辞が接続する場合、連結子音を削除する。

**接続規則 2:** 母音幹に連結母音を持つ接尾辞が接続する場合、連結母音を削除する。

### 3.2 統語接尾辞と派生接尾辞

前節で否定の接尾辞を「-(a)nai」ではなく「-(a)na-i」と表記した。これは「kak-ana-katta(書カナカッタ)」のように、さらに他の接尾辞が接続することが可能だからである。これは、動詞語幹に「-(a)na-」が接続することにより、新たな語幹が派生したと見ることができる。そのような語幹を二次語幹と呼び、二次語幹を派生する接尾辞を派生接尾辞と呼ぶ。日本語の派生接尾辞には他に「-(s)ase-」

「-(r)are-」「-(i)mas-」「-(i)ta-」があり、それぞれ使役、受身・可能・尊敬、丁寧、希望の意義を表す。また「-(i)aw-」も共同の意義を表す派生接尾辞と見ることができる。

二次語幹に対する接尾辞の接続に関しても接続規則 1 および 2 は適用される。例えば表1では「kak-imas-fu」において、連結子音 r が削除されている。

表 2: 派生接尾辞の対応

役割	日本語	ウイグル語	日本語例	ウイグル語例
使役	-(s)ase-	-ğuz-	書 k-ase-	yaz-ğuz-
受身	-(r)are-	-(i)l	書 k-are-	yaz-il-
可能	-(r)e-	-(y)ala-	書 k-e-	yaz-ala-
丁寧	-(i)mas-	-	書 k-imas-	yaz-
否定	-(a)na-	-ma-	書 k-ana-	yaz-ma-
希望	-(i)ta-	-ğu-	書 k-ita-	yaz-ğu-
共同	-(i)aw-	-(i)x-	書 k-iaw-	yaz-ix-

一方、派生接尾辞に対して、「-(r)u」のように新たな語幹を派生しない接尾辞を統語接尾辞と呼ぶ。統語接尾辞は動詞形の形成の役割を果たす。ここで、動詞形とは終止形、連体形、連用形、命令形の四形のことである。動詞に複数の接尾辞が接続する場合には、統語接尾辞が最後に接続する。

## 4 派生文法によるウイグル語の記述

派生文法で日本語を記述した場合の特徴としては、前節で述べたように、以下の2点が挙げられる。

1. 連結子音・連結母音によって母音幹と子音幹に接続する接尾辞を同じものとして扱う
2. 動詞語幹に派生接尾辞が接続して新たな語幹を派生する

日本語と同様、ウイグル語も派生文法で記述することができるが、この2点はその場合にも見られる特徴である。

例えば、日本語の動詞「作 r-」に相当するウイグル語の動詞は“yasa-”である。使役の意味を表す場合、日本語では派生接尾辞「-(s)ase-」が接続して「作 rase-」となる。同様にウイグル語では、“-ğuz-”という派生接尾辞が接続して“yasağuz-”となる。日本語とウイグル語の派生接尾辞の対応を表2に示す。ここで、日本語の丁寧の「-(i)mas-」に相当するウイグル語は存在せず、翻訳の必要はない。また「-(r)are-」は受身・可能・尊敬の意味があるが、ここでは受身の意味に限定している。

一方、ウイグル語で受身を表す派生接尾辞は“-(i)l-”である。括弧内の i は連結母音であり、日本語と同様、語幹末尾が母音である動詞“yasa-”に接続する場合には欠落して“yasal-”となる。ところが、ウイグル語には欠落する代わりに変化する音素がある。非完了の連体形を表す統語接尾辞“-[i]diğan”は、“yaz-”のような子音幹動詞に接続する場合には、i がそのまま表記され“yazidiğan”となるが、“yasa-”のような母音幹動詞に接続する場合には [i] が y に変化して“yasaydiğan”となる。このような音素を連結半母音と呼ぶ。また連結母音 (i) と区別するため、連結半母音は [i] と記述する。このことから、ウイグル語では次の動詞接続規則を追加する。

**接続規則 3:** 連結半母音は、子音幹に接続する場合は i に、母音幹に接続する場合は y にそれぞれ変化する。

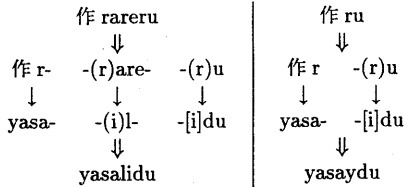


図 3: 派生文法を用いた動詞句の翻訳

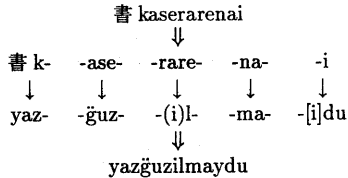


図 4: 逐語訳によるウイグル語-日本語翻訳

## 5 派生文法を用いた逐語訳

派生文法を用いた逐語訳の例を「作 rareru」「作 ru」について示すと、図3のようになる。図2と比較すると、単語が日本語とウイグル語で1対1に対応していることが判る。また接続規則1~3により、ウイグル語の翻訳文も簡単に生成できる。

一方、日本語は語順の自由度が高いといわれるが、動詞語幹に接続する接尾辞の順序には明らかに制約がある。この接尾辞の接続の順序も日本語とウイグル語でほぼ同じである。このため図4に示すように、日本語入力文の形態素解析が終わった段階で、各単語を対応するウイグル語に置き換えれば、機械翻訳は概ね可能となる。

## 6 ウイグル語動詞特有の文法的性質とその処理

日本語とウイグル語は類似性が高いが異なる部分もある。したがって、前節に述べた逐語訳だけでは翻訳できない部分がある。本節では、それらを指摘するとともに、それに対する処理方法を述べる。

### 6.1 人称接尾辞

日本語は動詞語幹に複数(0個も可)の派生接尾辞が接続し、最後に統語接尾辞が接続して動詞句を形成する。それに対してウイグル語では、統語接尾辞の後に、さらに人称接尾辞が接続する場合がある。日本語と同様、ウイグル語でも主語が省略されることがあるが、その場合でも人称接尾辞によって人称の別が示される。人称接尾辞には表3に示す4つのパターンが存在する。接続する人称接尾辞のパターンは、直前の統語接尾辞によって決まる。例えば、非完了の統語接尾辞「-[i]」にはパターン1の人称接尾辞が接続し、以下のように変化する。

- 一人称: Mén yazimén. (私が書く。)
- 二人称: Sén yazisén. (あなたが書く。)
- 三人称: U yazidu. (彼が書く。)

表 3: 人称接尾辞のパターン

パターン 1			パターン 2		
	単数	複数		単数	複数
一人称	-mén	-miz	一人称	-m	-k
二人称	-sén	-senlér	二人称	-ng	-nglar
三人称	-du	-du	三人称	-	-

パターン 3			パターン 4		
	単数	複数		単数	複数
一人称	-	-li	一人称	-	-li
二人称	-	-li	二人称	-sén	-sénlér
三人称	-	-li	三人称	-du	-du

表 4: 統語接尾辞

	日本語	ウイグル語	役割
終止形	-(r)u	-[i]-@1	非完了
	-(i)ta	-di-@2	完了
	-(y)ou	-(a)y-@3	前望
	-(u)mai	-may-@4	否定前望
連体形	-(r)u	-[i]diġan	非完了
	-(i)ta	-ġan	完了
連用形	-(i)te	-(i)p	順接完了
	-(r)eba		仮定条件
	-(r)uto	-sa-@2	開放条件
	-(i)teha		却下条件
	-(i)temo	-sa-@2-mu	譲歩
	-(a)zu	-mastin	否定
	-(r)umade	-ġiqe	終格
形	-(i)nagara	-ġaq	同時
	-(i)ni	-ġili	目的
	命令形		
命令形	-e,-ro	-ġin	肯定命令
	-(r)una	-ma-ġin	否定命令

そこで、人称接尾辞はパラメータで表示し、統語接尾辞と一緒に記述する。人称接尾辞を表すパラメータは、4つのパターンに応じて-@1,-@2,-@3,-@4で表現する。文脈などの情報から人称を決定した後、パラメータで表した人称接尾辞を適当なものに置き換える。

### 6.2 終止形と連体形

日本語では終止形と連体形の形は同じであり、例えば統語接尾辞「-(i)ta」は連体形と終止形の両方の役割を果たす。しかし、次の例のように、ウイグル語では連体形と終止形を表す統語接尾辞はそれぞれ別の単語である。

終止形: Mén yazidim. (私が書いた。)

連体形: Mén yazġan kitap. (私が書いた本。)

したがって、終止形と連体形を区別する必要が生じる。

そこで、日本語を形態素解析した際に、「-(i)ta」の後に名詞がくる場合は連体形形成統語接尾辞、「-(i)ta」で文が終わる場合は終止形形成接尾辞とし、それぞれに対応するウイグル語の接尾辞を辞書に登録しておく。

以上を考慮して、日本語とウイグル語の統語接尾辞の対応をまとめると、表4のようになる。

表 5: 「-(a)na-」・「-ma-」の対応の例

日本語	ウイグル語
書 k-ana-i(終止形)	yaz-ma-y-@1
書 k-ana-i(連体形)	yaz-ma-ydiġan
書 k-ana-katta(終止形)	yaz-ma-di-@2
書 k-ana-katta(連体形)	yaz-ma-ġan
書 k-ana-karou	yaz-mas
書 k-ana-ku	yaz-mastin
書 k-ana-kute	
書 k-ana-kereba	yaz-ma-sa-@2
書 k-ana-ito	
書 k-ana-kuteha	
書 k-ana-kattara	yaz-ma-ġan bolsa-@2
書 k-ana-kutemo	yaz-ma-sa-@2-mu

## 7 品詞の不一致

本稿で提案する動詞句の翻訳手法においては、日本語とウイグル語との対応する単語で品詞が一致していることを基本としている。しかし、品詞が一致していない場合も存在する。本節ではそれに対する処理方法を述べる。

### 7.1 否定の派生接尾辞“-ma-”

否定の派生接尾辞は日本語では「-(a)na-」であり、ウイグル語では“-ma-”である。「書 kanakatta」を本手法で機械翻訳すると、次のようになる。

書 k-	-ana-	-katta
↓	↓	↓
動詞	派生接尾辞	形容詞接続 統語接尾辞
↓	↓	↓
yaz-	-ma-	-kän

ここで、“-kän”は形容詞の語幹に接続して完了の意味を表すウイグル語であり、日本語の“-katta”に相当する。しかし、実際のウイグル語においては“-ma-”の後に接続する単語は“-kän”ではなく、“-di-@2”(日本語の「-(i)ta」に相当)である。これは日本語の「-(a)na-」が形容詞語幹を派生するのに対して、ウイグル語の“-ma-”は動詞語幹を派生する接尾辞だからである。このように「-(a)na-」と“-ma-”の品詞が異なるため、適切な翻訳ができないという問題が生じる。

そこで、「-(a)na-」どその後に接続する統語接尾辞を合わせて1語とし、そのウイグル語訳を辞書に登録することで対処した。この対応を表5に示す。先程の例では、「-(a)nakatta」(終止形)に対応する“-madi-@2”を辞書に登録して翻訳を行う。

### 7.2 希望の派生接尾辞“-ġu-”

日本語の希望を表す派生接尾辞「-(i)ta-」は形容詞語幹を派生する接尾辞である。これに対してウイグル語の希望を表す派生接尾辞“-ġu-”は名詞語幹を派生する接尾辞であ

表 6: 人称接尾辞パターン5(名詞幹接続)

	単数	複数
一人称	-(i)m	-(i)miz
二人称	-(i)ng	-(i)nglar
三人称	-(s)i	-(s)i

表 7: 「-(i)ta-」・“-ġu-”の対応の例

日本語	ウイグル語
書 k-ita-i(終止形)	yaz-ġu-@5 bar
書 k-ita-i(連体形)	yaz-ġu-@5 bar
書 k-ita-katta(終止形)	yaz-ġu-@5 bar idi
書 k-ita-katta(連体形)	yaz-ġu-@5 bar
書 k-ita-karou	yaz-ġu-@5 bar-du
書 k-ita-kute	yaz-ġu-@5 bar-liġtin
書 k-ita-kereba	yaz-ġu-@5 bolsa
書 k-ita-ito	
書 k-ita-kuteha	
書 k-ita-kattara	yaz-ġu-@5 bolsa
書 k-ita-kutemo	yaz-ġu-@5 bolsa mu

る。例えば「私は書きたい。」のウイグル語訳は“Mén yazġum bar.”であるが、これを日本語に直訳すると「私は書きたいことがある。」となる。この“-ġu-”は「…したいこと」の意味で動詞を名詞化する接尾辞であり、その後には人称接尾辞が接続する。この人称接尾辞は、名詞に接続する人称接尾辞であり、動詞に接続する場合は異なるパターン(表6)を持つ。この後に「ある」に相当する形容詞“bar”が接続して、“yazġum bar”と表現される。

そこで、否定の“-ma-”と同様に、すべての組合せと対応するウイグル語表現を辞書に登録した。

### 7.3 共同の派生接尾辞“-(i)x-”

通常の日本語文法では、「話し合う」の「合う」は連用形に接続して共同の意味を形成する補助動詞として捉えられる。そのため、「話し合う」を本手法で翻訳すると以下のようなになる(主語は三人称単数と仮定する)。

話し合う			
↓			
話 s-	-i	合 w-	-u
↓	↓	↓	↓
動詞	統語接尾辞	動詞	統語接尾辞
↓	↓	↓	↓
eyt-	-(i)p	yarax-	-[i]-@1
↓			
eytip yaraxidu			

しかし、「話し合う」に相当するウイグル語は“eytixidu”であり、この翻訳結果は誤りである。ここで“-(i)x-”は共同の意味を表す派生接尾辞である。

MAJO への入力:	呼ばれる、 (文字列、品詞名、ウイグル語訳)
MAJO の出力結果:	呼 b 子動詞 ati-
	are 子母辞 -(i)-
TSUE への入力	ru 母終辞 -[i]-@1
	、 句読点 、
TSUE の出力結果:	atilidu ,

図 5: 翻訳システムの動作例

そこで、日本語においても「-i」 と「合 w-」を合成した「-i 合 w-」を共同の派生接尾辞として捉え、辞書に登録した。これにより、「話し合う」に対して「eytixidu」が翻訳されるようになる。

## 8 機械翻訳システムの実現と実験

機械翻訳システムには、日本語形態素解析システム MAJO (Morphological Analyzer of Japanese based On derivational grammar) [1][2] を利用した。MAJO は派生文法に基づいて日本語の形態素を解析するシステムであり、辞書に各単語の情報が(文字列、品詞名、意味)の3項組の形で登録されている。今回作成した機械翻訳システムでは MAJO の辞書を(文字列、品詞名、ウイグル語訳)の3項組で表される日本語-ウイグル語辞書と置き換えて使用した。また、MAJO で使用する接続表の一部を書き換え、終止形と連体形とを区別した。変更部分は辞書と接続表のみであり、MAJO のシステム自体には何ら変更は加えていない。また、MAJO とは別に、連結子音・連結母音や人称接尾辞を処理するシステム TSUE (Translation System into Uighur sEntences) を作成した。TSUE は MAJO の出力を入力とし、整形されたウイグル語の訳文を出力する。図 5 は本翻訳システムの動作例である。MAJO による形態素解析が終了した段階では、ウイグル語の訳語に連結母音や人称接尾辞が含まれる。この出力を TSUE が整形して、最終的なウイグル語の訳文が生成される。

さらに、このシステムを用いて翻訳実験を行った。EDR 日本語コーパスから動詞句の部分(サ変動詞を除く<sup>1)</sup>)を取り出して翻訳した。意味の曖昧性から1つの日本語動詞に対して複数のウイグル語が対応する場合があるが、今回は考慮せず、対応する1つのウイグル語の翻訳結果だけを出力させている。また約950語の動詞を含む辞書を使用した。対応する単語がなかったものは評価の母集団から除いてある。人称接尾辞はすべて三人称単数と仮定して扱った。実験結果は表8のようになった。これにより本システムが基本的には動作することを確認できた。

## 9 おわりに

本稿では派生文法を用いた日本語-ウイグル語の動詞句の機械翻訳手法について述べた。派生文法を用いれば、逐

<sup>1</sup>サ変動詞は MAJO において名詞と解析されるため

表 8: 翻訳実験結果

評価	個数
文法的に正しくて違和感もない	105
他の言い方が普通	8
文法的には正しいが違和感がある	1
文法的には正しいが意味が通じない	5
文法的には誤りだが意味は通じる	13
文法的に誤っていて意味も通じない	4
合計	136

語訳により、簡潔で高精度な動詞句の翻訳が可能になることを示した。また、ウイグル語特有の文法的性質や、日本語とウイグル語での品詞の違いによって生じる翻訳誤りに対しても、その対処方法を示した。

今回は、1つの日本語の単語に対して複数のウイグル語の単語が相当する場合に、どの単語を選択するかは考慮していない。今後は、こうした曖昧な翻訳語の選択方法、特に受身・可能・尊敬の意味をもつ「-(r)are-」に対する適切なウイグル語の選択方法について検討していく。

また、サ変動詞の中にも日本語とウイグル語で品詞が異なる場合がある。例えば、日本語の「勉強」は名詞であり「する」が接続して動詞となる。一方ウイグル語では「oqu-」が「勉強する」に相当する動詞であり接尾辞「-x」が接続した「oqux」が「勉強」という名詞になる。そのような場合の対処方法は今後の課題である。

さらに動詞句以外の部分への翻訳手法についても引き続き検討していく。

## 参考文献

- [1] 小川泰弘, ムフタル・マフスット, 稲垣康善: “派生文法による日本語動詞接尾辞の形態素解析”, 人工知能学会全国大会論文集, pp.323-326(1996).
- [2] 小川泰弘, ムフタル・マフスット, 稲垣康善: “派生文法に基づく日本語動詞接尾辞の形態素解析”, 情処研報, 96-NL-116, pp.7-12(1996).
- [3] ムフタル・マフスット, 外山勝彦, 稲垣康善: “日本語-ウイグル語機械翻訳における助動詞のパラメータ化による処理”, 信学技法, NLC94-13, pp.47-53(1994).
- [4] 戸部実之: “ウイグル語入門-文法と会話-”, 泰流社(1986).
- [5] 竹内和夫: “現代ウイグル語四週問”, 大学書林(1991).
- [6] 竹内和夫: “現代ウイグル語基礎 1500 語”, 大学書林(1985).
- [7] 清瀬義三郎郎府: “日本語学とアルタイ語学”, 明治書院(1991).
- [8] 清瀬義三郎郎府: “日本語文法新論-派生文法序説-”, 桜楓社(1989).
- [9] 久光徹, 新田義彦: “日本語形態素解析における効率的な動詞活用処理”, 情処研報, 94-NL-103, pp.1-7(1994).
- [10] 西野博二, 鷺北賢, 石井直子: “派生文法による日本語構文解析”, 情処研報, 92-NL-87, pp.43-50(1992).
- [11] 舘武志, 米澤明憲: “日本語形態素解析システムのための形態素文法” 自然言語処理 Vol.2, No.4, pp.37-65(1995).