日本語-ウイグル語機械翻訳のための格助詞の変換処理

ムフタル・マフスット[†] 小川 泰弘[†] 稲垣 康善[†]

日本語とウイグル語は共に膠着語であり、語順がほぼ同じであるなどの構文的類似性が見られる。そのため、日本語-ウイグル語機械翻訳においては、日本語文を形態素解析した後、逐語訳を行うだけでもある程度の翻訳が可能となる。これは、名詞や動詞などの自立語の文中での役割が助詞、助動詞といった付属語によって示されており、そうした付属語においても、日本語とウイグル語との間で対応関係があるからである。特に名詞に接続する格助詞は、文中での他の語との関係を決めるという、言語構造上重要な機能を持っている。そのため、格助詞を正しく翻訳できなければ、違和感のある翻訳文になるだけでなく、ときには致命的に誤った意味となる翻訳文を生成することがある。そこで、本論文では、日本語-ウイグル語機械翻訳における格助詞の取り扱いについて論じる。まず、計算機用日本語基本動詞辞書 IPAL を用いて動詞と格助詞の使われ方を調べるとともに、それぞれの格助詞の機能に対応するウイグル語格助詞を決定する。さらに、この調査結果から作成した動詞の格パターンを利用して複数の格助詞の訳語候補の中から、適切な訳語を選択する手法を提案する。また、本提案手法に対する評価実験では、環境問題関連の新聞社説 3 編の日本語 138 文を対象にし、我々が本論文で提案するアプローチに基づいて実験を行った。その結果、99.3%の正解率を得ることができた。

キーワード: 機械翻訳, 日本語, ウイグル語, 格助詞, 格パターン, 膠着語

Translation of Case Suffixes on Japanese-Uighur Machine Translation

MUHTAR, MUHSUT[†], OGAWA, YASUHIRO[†] and INAGAKI, YASUYOSHI[†]

Japanese and Uighur languages are *agglutinative languages* and they have a lot of syntactic similarities. Thus we can translate Japanese into Uighur sequentially by replacing Japanese words with corresponding Uighur words after morphological analysis of Japanese sentences. However, we should translate case particles correctly in order to prevent wrong reading, because they have important roles on both languages. In this paper, we propose a new approach to translation of case particles. For that purpose we researched the verb dictionary made by IPA and classified the use of Japanese case particles. Our approach selects a correct Uighur case particle using the combination pattern of verb and case particles. We also show the performance evaluation of the system.

KeyWords: *Machine translation, Japanese, Uighur, Case particle, Case pattern, Agglutinative language*

[†] 名古屋大学大学院工学研究科計算理工学専攻, Department of Computational Science and Engineering, Nagoya University

1 はじめに

日本語とウイグル語は共に膠着語である。膠着語には、概念などを表し単独で文節を構成することが可能な自立語と、単独で文節になることはなく、自立語に接続して、その自立語の文中での役割を示したり、自立語に新たな意味を付加する付属語の区分がある。膠着語では、付属語がよく発達しており、言語構造上重要な役割を果たす。これらの特徴は、日本語とウイグル語だけでなく、韓国語、トルコ語、モンゴル語などのアルタイ語系に属する言語に共通するものと考えられている(馬淵 1986)。このグループに属する言語間の機械翻訳については、グローバル化の流れの中で多言語間機械翻訳の重要性が高いにもかかわらず、これまでほとんど行われておらず、日本語と韓国語との翻訳について研究されているのが目立つに過ぎない。そのような状況の中で、ムフタル、小川らは、日本語-ウイグル語機械翻訳の研究を開始した。

ムフタル、小川らは、これらの言語に共通する特徴を有効に利用した日本語-ウイグル語機械翻訳の研究を進めている(ムフタル 1993)(小川、ムフタル、杉野、外山、稲垣 2000)。その特徴の一つは語順の自由度である。日本語は語順が比較的自由であると言われ、例えば、(1)「私の本を買った」と(2)「本を私を買った」は、いずれも日本語として正しい表現である。これは、日本語では文節の役割が付属語によって示されるためである。この性質は同じ膠着語であるウイグル語にも見られ、(1)の直訳となる“mén kitapni setiwaldim”という表現も、“mén”(私)と“kitap”(本)を入れ替えて(2)の直訳とする“kitapni mén setiwaldim”という表現も、いずれもウイグル語として可能である。そのため、日本語文をウイグル語へ翻訳する場合、日本語の語順そのままに翻訳が可能である。そこで、ムフタルらは、日本語文の形態素解析結果を逐語訳することを基本とした日本語-ウイグル語機械翻訳システムを開発している。特に(小川他 2000)では、動詞句の翻訳に焦点を当て、派生文法(清瀬 1991)を利用することで動詞付属語を含めた動詞句に対して自然なウイグル語訳を与えることを可能としている。

ところで、日本語からウイグル語へ語順そのままの翻訳が可能なのは、名詞付属語、特に格助詞によって文節の役割が明示されているからである。これも、日本語とウイグル語に共通する特徴の一つである。しかし、このことは、格助詞を正しく翻訳できなかった場合は翻訳文が意味不明なものになることを意味する。

そこで、本論文では、日本語-ウイグル語機械翻訳の中での格助詞の取り扱いを検討する。格助詞は日本語だけでなくウイグル語にも存在し、例えば(竹内 1991)では、格語尾と呼ばれている。日本語の格助詞とウイグル語の格助詞には対応関係が見られるが、いわゆる多義性の問題が存在し、日本語の格助詞に複数のウイグル語格助詞が対応する場合がある。本論文では、単に格助詞を翻訳するだけでなく、こうした格助詞の多義性も考慮して適切な格助詞の翻訳を行う手法を提案する。

日本語と他の膠着語との間の機械翻訳に関する研究では、日韓機械翻訳が盛んである(崔, 金 1996; 李, 中嶋, 安居院 1989; 金, 大駒 1996; 朴, 李, 李, 寛 1997)。これらの研究の多くは、日本

語と韓国語の語順の類似性や、格形式の類似性を利用し、逐語訳を基本とする翻訳が進められており、比較的品质の良い翻訳を実現しているが、その一方で、語彙の多義性の解消が重要な課題であることが指摘されている(崔, 金 1996). 多義性に関する研究については、(李他 1989; 金, 大駒 1996; 朴他 1997) などがあり、動詞の格パターンと意味解析を利用する手法(李他 1989), 入力文の前後に出現する単語との接続関係を利用する手法(金, 大駒 1996), 連語パターンを用いる手法(朴他 1997) などが提案されている。

本論文では、品質の高い日本語-ウイグル語機械翻訳システムの構築を目指して、動詞の格パターンを利用した、格助詞の翻訳手法を提案する。まず、計算機用日本語基本動詞辞書 IPAL(IPAL 1987) を用いて両言語の格助詞間の対応関係について詳細な調査を行うとともに、動詞の格パターンを獲得する。さらに、それを利用した格助詞の変換処理を実現し、評価実験を行った。評価実験に使用する日本語-ウイグル語機械翻訳システムは(小川他 2000)で提案されたシステムに、本論文で提案する格助詞変換処理のモジュールを加えたものである。この方法では、あらかじめ獲得した格パターンと格助詞の対訳の情報を、必要に応じて日本語-ウイグル語の対訳辞書のウイグル語動詞に付加する。実際の翻訳の過程は、まず、翻訳対象である日本語入力文を形態素解析し、それぞれの形態素をウイグル語に逐語訳する。この段階で、すべての単語にデフォルトのウイグル語訳が与えられる。次に、ウイグル語動詞に付加された格パターンと、入力文中に出現した格パターンとを比較し、デフォルト訳では不自然な訳語となる格助詞を適切な他の訳語に置き換える。最後に、訳出のウイグル語形態素を接続してウイグル語文を生成する。

本論文では、ウイグル語における同じ格助詞の音便形を、すべて一つに統合して議論する。例えば、格助詞 “gé” は、音便変化により “ga”, “ké”, “ka” などの形もとるが、本論文では、すべて “gé” と表記する。なお、実際の翻訳システムでは、最後のウイグル語文生成の段階で音便形に従って変化させる。また、ウイグル語には、日本語には存在しない人称接尾辞がある。例えば、同じ「買う」でも、「私が買う」 “mén setiwalimén” と「彼が買う」 “u setiwalidu” では、下線部に示すように、それぞれ別々の人称接尾辞が接続する。しかし、本論文では、いくつかの例文を除いて三人称に統一して議論する。

ウイグル語には、アラビア文字に似た 32 の文字があり、文は右から左へと書かれる。それとは別に、ローマ字表記を用いる場合もあり、本論文では、便宜上、ローマ字表記を用いることにする。不足する文字の代わりに、ç, é, ğ, ħ, k, ö, ş, ü, z を用いる。ウイグル文字とローマ字表記の対応に関しては、付録 A を参照されたい。

本論文の構成は以下の通りである。まず 2 章では、日本語-ウイグル語機械翻訳における格助詞の重要性とその問題点について指摘する。3 章では、計算機用日本語基本動詞辞書 IPAL(IPAL 1987) における格助詞の使用状況と、対応するウイグル語訳語の分布に関する調査結果を示す。4 章では、本論文で提案する日本語-ウイグル語機械翻訳における格助詞の変換処理について述

べ、5章で本手法に基づく実験結果を示す。6章は本論文のまとめである。

2 日本語とウイグル語の格助詞

日本語やウイグル語のような膠着語では、助詞の機能がよく発達している。特に、格助詞は、ある語の文中での役割を決めたり、他の語との関係を決めるなど、文の構造を決める上で重要な役割を果たしている。例えば、「先生」、「教室」、「生徒」、「本」、「大きい声」、「読ませた」の六つの概念語があるとする、そのままでは、どれが主体か、どれが被動態かなど、それらの文中での役割や相互関係は明確にならない。しかし、以下のように格助詞の「が」「で」「に」「を」を用いることによって、それらが明確になる。

先生が 教室で 生徒に 本を 大きい声で 読ませた
(1) (2) (3) (4) (5)

日本語とウイグル語の格助詞は機能的には同じであり、日本語のある格助詞の一つの機能に相当する格助詞は、ほとんどウイグル語にも存在する。例えば、上の文に対するウイグル語訳は次のようになり、格助詞がそれぞれ対応していることが分かる。

oqutkuçio sinipde oquğuçilargé kitapni yuquri awazde oquğuzdi
(1') (2') (3') (4') (5')

これは、日本語-ウイグル語機械翻訳を行う上で、重要なポイントの一つである。なぜなら、例えば日本語を英語に翻訳しようとするとき、多くの場合、日本語の格助詞の機能を、英語の文中の語順、動詞の語彙的機能、あるいは英語の文全体の文脈などで表現しなければならない。そのため、構文解析などが必要になり日本語格助詞の取り扱いが複雑となる。それに対し、ウイグル語に翻訳する場合には、語順をそのままにし、さらに格助詞にも対応する訳語を与えるだけで翻訳が可能となる。よって、形態素解析が終了した段階で、格助詞を含む各単語を逐語訳することによって翻訳が可能となる。

しかし、日本語とウイグル語の格助詞は、必ずしも一対一に対応する訳ではない。

例えば、「仕事を片付ける」「ixlarni yeğixturidu」, 「ゴミを捨てる」「éhlétni tökidu」のような文では、格助詞「を」は“ni”に翻訳される。しかし、「大学前を通る」「univirsititning aldidin ötidu」, 「橋を渡る」「köwürükidin ötidu¹」といった文では、「を」は“din”に翻訳される。一般に、前者の「を」の機能は対象を示すものであり、後者の機能は場所を示すものであると言われる。つまり、日本語の「を」に対応するウイグル語の格助詞は、「を」の機能によって異なるのである。また、場所を示す「を」は、「通る」「渡る」などの移動動詞と一緒に出現する場合が多いことから、「を」を含む名詞句に係る動詞に依存して「を」に対応するウイグル語の格助詞が決まるとも言える。

1 直接的に関係はないが、ここでは、「通る」と「渡る」の両方のウイグル語訳は“ötidu”になる。

先に述べたように、日本語とウイグル語では格助詞が文章を読解する上で重要な役割を果たすため、日本語-ウイグル語機械翻訳においても、正しく翻訳する必要がある。そのため、日本語の格助詞に対応する複数のウイグル語格助詞の中からどれを選択するかという決定は、日本語-ウイグル語機械翻訳において大きな課題である。

そのためには、まず日本語の格助詞の機能を調べ、それに対応するウイグル語の格助詞を決めなければならない。一般的な助詞の機能や対応関係は(ムフタル, 外山, 稲垣 1993)などで詳しく調べられており、ここでは格助詞だけに絞る。

3 動詞辞書調査に基づく日本語-ウイグル語の格助詞の対応付け

前章で述べたように、日本語とウイグル語には共に格助詞が存在するが、一般的に、それら是一対一に対応するとは限らない。日本語の一つの格助詞が複数の意味(機能)を有しており、それぞれの意味に対して、別々のウイグル語の格助詞が対応するからである。その逆も真である。そのため、日本語からウイグル語へ翻訳する際には、日本語格助詞のそれぞれの機能に対応して、適切なウイグル語の助詞を訳語として選定する必要がある。

したがって、日本語格助詞の機能を分類し、それぞれに対応するウイグル語格助詞を決定する必要があるが、本論文では格助詞と動詞との関係に着目した。日本語とウイグル語では、格助詞が動詞と密接な結合関係をもち、各動詞がそれぞれ決まった格パターン²を持っている。そのため、日本語-ウイグル語機械翻訳における格助詞の機能の決定と訳語の選定は、動詞と切り離して考えることはできない。そこで、我々は計算機用日本語基本動詞辞書 IPAL(IPAL 1987)中に出現した格助詞の機能を調査し、それぞれの出現において対応するウイグル語格助詞はどれになるかという観点から分析し、対応付けを行った。(IPAL 1987)を用いたのは、日本語の基本的な動詞とそれらの意味や用法が格パターンの観点から説明されており、両言語の格助詞の機能及びそれらの間の統計的な対応関係を調べるのに適していると考えたからである。

(IPAL 1987)には、900に近い日本語動詞があり、各動詞のそれぞれの意味を説明するための例文 3473 文が含まれている。今回の調査では、それらの例文において動詞の格形式パターンとして示されたすべての格助詞に対してウイグル語の訳を与え、それらの統計を取った。ここでは、日本語からウイグル語への機械翻訳を想定しているため、日本語の格助詞が、ウイグル語のどの格助詞にどのように対応しているかを中心に調査を行った。

ただし、日本語の「の」および、それに対応するウイグル語の“ning”は、どちらの言語の文法においても格助詞と分類されているが、「の」は「A の B」のような形でその格助詞的機能を果たしており、「A の B」全体が「名詞的」であるため、(IPAL 1987)では動詞の格パターンの構成要素として出現していない。同様に、格助詞「や」も動詞の格パターンの構成要素にならないため、(IPAL 1987)では取り扱われていない。そこで、本論文では、これらを除いた「が」

2 結合価や格構造、格フレームという場合もある

表 1 格助詞間の統計的対応関係

格助詞	対応するウイグル語とその数							合計
が	ø 3637 97.7%	ning 77 2.0%	gé 5 0.3%					3719
を	ni 1566 74.3%	ni/ø 320 15.2%	din 118 5.6%	ø 34 1.6%	gé 19 0.9%	dé 4 0.2%	fault 47 2.2%	2108
に	gé 1183 73.5%	dé 294 18.3%	din 81 5.0%	ø 16 1.0%	fault 35 2.2%			1609
で	dé 638 97.8%	bilén 14 2.2%						652
と	dép 193 50.1%	ø/dép 46 12.0%	bilén 146 37.9%					385
から	din 289 98.3%	fault 5 1.7%						294
へ	gé 236 100%							236
より	din 14 100%							14

「を」「に」「で」「と」「から」「へ」「より」について、対応するウイグル語格助詞が何になるかを調査した。その結果を表 1 に示す。それぞれの格助詞とウイグル語格助詞との対応関係について以下に考察する。

3.1 格助詞「が」

今回の調査では、「が」の出現数が一番多く、全部で 3719 個あった。その内の 3637 個 (97.7%) がウイグル語の主格を表わす格助詞 “ø” に対応していた。ここで、“ø” は空白を表わすが、ウイグル語では主格が確かに存在し、それを空白で表現しているのであり (Musa 1985), 何かの格接尾辞があって、それが省略されている訳ではない。なお、格助詞「が」については、動詞だけでなく、「物価 が 高い」などのように形容詞の主体に接続する場合もある。今回は、動詞だけを調査対象としたが、形容詞の主体になる場合の「が」に対するウイグル語訳も、多くの場合は “ø” である。

その他、77 個 (2.0%) は “ning” に、5 個 (0.3%) は “gé” に対応していた。“ning” はウイグル語で所有格を表す格助詞であり、日本語の格助詞「の」にほぼ相当する。今回の調査で「が」

の訳語として“ning”が出現したのは以下の二つの場合である。一つは、「我 が 国日本」などの例における所有を表す「が」であり、これをウイグル語に翻訳すると“Bizning Déwletimiz Yapon”（私たち の 国日本）³となる。もう一つは、計算機用日本語基本動詞辞書 IPAL(IPAL 1987)における「は」の扱いに起因するものである。(IPAL 1987)では、「このスカート は サイズが合わない」における「は」を「が」と見なしているが、この文をウイグル語に訳した場合、“Bu yopkaningrazmiri udul kélméydu”（このスカート の サイズが合わない）となり、こうした「は」は、「が」よりも「の」と見なしてウイグル語に置き換えた方が自然だと考えられる。

“gé”は、表1の結果から分かるように、主に日本語の「に」および「へ」に対応するウイグル語の格助詞である。これについても、(IPAL 1987)では「この仕事 は 忍耐が要る」といった文の「は」を「が」と見なしているのため、今回の調査結果において「が」の訳語として出現した。この文に対するウイグル語訳は“Bu hizmetgé çidaşlıq ketidu”（この仕事 に 忍耐が要る）となり、「は」を「に」と見なせば良いと考えられる。

これら“ning”、“gé”への誤りは、係助詞「は」の取り扱いに起因する。本論文では、格助詞の取り扱いについて議論しており、係助詞「は」に起因する間違いについては扱わないことにする。「は」については、ウイグル語に相当する助詞が存在しないため、「は」を「が」「の」「に」「を」などの格助詞に変換した後で翻訳する必要がある。係助詞「は」をいずれの格助詞に置き換えるかの決定手法については今後の課題である。

また、今回の(IPAL 1987)を使用した調査では出現しなかったが、日本語大辞典(梅棹, 金田一, 阪倉, 日野原 1995)によれば、格助詞「が」にはもう一つ、「願望・好悪・能力などを表したり、可能であることを言ったりする語の対象を示す」という機能もある。これは、例えば「水 が 飲みたい」、「日本語 が 読める」のような場合の「が」である。これらの「が」は対象を示しているから、ウイグル語訳する場合には、それぞれ“su(ni)⁴ıçküm bar”(水 を 飲みたい)、“Yapon tilini okuyalaydu”(日本語 を 読める)となり、「が」が対象格を表す“ni”に対応している。その一方で、例えば「飲みたい」については「私 が 飲みたい」といった表現も可能であり、その場合の「が」は主格を表す“ø”に対応する。こうした例から、「が」については、単純に動詞と格助詞の関係からでは訳語を決定できない場合があることが分かる。

「が」の機能をまとめると、多くの場合は“ø”と翻訳すれば良いが、“ning”、“ni”と翻訳すべきものが若干あることが分かる。

3.2 格助詞「を」

「が」の次に多く出現したのは「を」であり、全部で2108個出現した。その1566個(74.3%)がウイグル語で対象格を表す“ni”に、320個(15.2%)が“ni/ø”に対応していた。ここで、“ni/ø”

3 括弧内はウイグル語文に対する日本語への直訳的表現である。以下の場合も同様である。

4 ここで“ni”が括弧内に表記されている理由は、次の「を」に関する議論を参照されたい。

は、格助詞“ni”に相当するが、省略することも可能であることを示している。ウイグル語では、動詞と目的語の関係が明らかな場合、対象を表す“ni”を省略する場合がある。例えば、「水を飲む」を“su içidu”のように表現し、「水」“su”の後に格助詞が接続しない場合がある。“ni/ø”は、こうした場合を表している。これはあくまでも“ni”が省略されているのであり、主格を表す格助詞“ø”が出現している訳ではない。実際、先程の例文も「この水を飲む」となった場合には、“ni”が記述され、“bu suni içidu”となる。機械翻訳においては、“ni/ø”についても“ni”を訳出すれば良いと考えられるから、両者を合わせると、「を」については約9割が“ni”に相当すると考えられる。

日本語大辞典(梅棹他 1995)によれば格助詞「を」には、動作・作用の対象を示す機能以外にも、「橋を渡る」のように通過点・経過する場所を示す機能や、「空港を出発する」のように起点を示す機能があるとされる。これらの機能を示す「を」は、ウイグル語の格助詞“din”に対応し、上記の日本語文はそれぞれ“köwrükdin ötidu”(橋から渡る)、“ayrudurumdin yolga çıķidu”(空港から出発する)と翻訳される。今回の調査では、そのような例が118個(5.6%)出現した。

しかし、日本語の「を」のうち、対象を表す機能が常にウイグル語の“ni”に対応するのではない。例えば、「彼の無実を信じている」における「を」は信じる対象を示しているが、この文に対する翻訳は“uning aķlıgıgı ixinidu”となり、「を」は“gé”に翻訳されている。ウイグル語の動詞“ixinidu”は、信じる対象を常に“gé”で表す。また、「約束を守る」“wédisidé turidu”のように、「を」が“dé”と翻訳される場合もあった。なお、「守る」には、他のウイグル語動詞に翻訳される場合もあり、その場合には「を」も別の格助詞に対応する。例えば、「彼女は子供を熊から守った。」は“U balılarnı eyıķdin saķlıdidi.”となり、「守った」は“saķlıdidi”に、「を」は“ni”に翻訳されている。

これは、「を」の訳語がその機能だけで決まるのではなく、ウイグル語動詞との組み合わせで決まる場合があることを示している。言い換えれば、ウイグル語の動詞“ixinidu”は“gé”を含む名詞句を、“turidu”は“dé”を含む名詞句を必要とするのである。なお、今回の調査では「を」が“gé”に対応するものが19個(0.9%)、“dé”に対応するものが4個(0.2%)あった。

また、空白“ø”に対応した34個(1.6%)は、“ni”を書かないほうが違和感がない場合である。例えば、「猶予を与えた」“möhlét bérıdu”,「指示を仰ぐ」“yolyoruk soraydu”などでは、“ni”が現れない。これは先程の「信じる」“ixinidu”の場合と異なり、“bérıdu”, “soraydu”の前の「を」が常に“ø”となる訳ではなく、“möhlét bérıdu”, “yolyoruk soraydu”という組み合わせで起きる、一種の慣用表現であると考えられる。

また、日本語のある動詞をウイグル語に翻訳した場合、その動詞に直接対応するウイグル語の動詞がなく、言い回しを適当に変更して翻訳する場合がある。例えば、「その政策は世間から非難を浴びた」のような文をウイグル語に翻訳する場合、「浴びる」に直接対応する動詞がウ

イグル語にないため, 「その政策は世間から非難された」といった表現にして翻訳する必要がある。このような場合, ウイグル語文には「を」に対応する格助詞は存在しない。今回の調査では, こうした直訳不能の場合を失敗として数え, *fault* の欄に記述した。「を」に関しては, このような例が 47 個 (2.2%) 出現した。

以上の結果から, 格助詞「を」については, 多くの場合は “ni” と翻訳すれば良いが, 動詞との組み合わせによっては, “din”, “gé”, “dé” と翻訳すべきものや, 動詞だけでは決定できないものがあることが分かる。

3.3 格助詞「に」

「に」は, 全部で 1609 個出現し, その 1183 個 (73.5%) がウイグル語の “gé” に対応した。これは, 一般に与格と呼ばれる「に」の機能であり, 「政府は Y 氏 に 国民栄誉賞を与えた。」 “*Hökümet Y épéndigé hélq xéripi mukapati bérđi.*” といった文の「に」に相当する。一方, 「に」には時間や場所を示す機能もあり, その場合には「青空 に 気球が浮んでいる。」 “*Kök asmandé yél xari léyléwatidu.*” のように, ウイグル語の “dé” に対応する。今回の調査では, “dé” に対応するものが 294 個 (18.3%) あった。

また, “gé” にも “dé” にも対応しない例として, 「太郎が次郎 に 優る」, 「人工物が自然物 に 劣る」のような比較の場合が 81 個 (5.0%) あった。これらの「に」はウイグル語では “din” となり, それぞれの文も “*Taro Jirođin üstün turidu*”, “*Sün’i nersé Tébi’i nersidin töwén turidu*” と翻訳される。この “din” は日本語の「から」「より」に相当するものである。

また, 16 個 (1.0%) で「に」を “ø” としたものがあった。それは, 「彼は弁護士 に なった。」 “*U adwokatø boldi.*” や「錬金術師は鉄を金 に した。」 “*Kimiyagerqi tömürni altunø qildi.*” といった場合である。「に」が “ø” となるのは動詞「なる」「する」の二つの場合だけである。これは種類としては限られているが, 動詞「なる」「する」は日本語において使用頻度の高い動詞であるため, 翻訳に際して考慮する必要がある。

また, 言い換えないと翻訳できないものとしては, 「彼女は木綿糸を帽子 に 編んだ」などの例があった。この場合, ウイグル語では “*u pahta yipni qalpaq qilip örüdi*”, と表現される。ここで, 「に」に相当する “qilip” は, 動詞「する」 “qilidu” のいわゆる連用形であり, ウイグル語の表現を日本語に直訳すると, 「帽子にして編む」となる。また, 「彼 に は子供が 2 人いる」という表現も “*uning balisidin ikkisi bar*” (彼の子供は 2 人である) のような言い換えが必要である。このような例は, 今回の調査では 35 個 (2.2%) 出現した。

また, 今回の調査では現れなかったが, 「に」には, 上記以外にも, 「本を買い に 行く」のように目的を示す機能もある。これは格助詞「に」が動詞の連用形に接続しているパターンである。この場合は「本を買い に 行く」 “*kitap setiwilgili baridu*”, のようになる。ここで, “gili” はウイグル語の動詞接尾辞の一種であり, 格助詞ではない。

「に」のこれらの機能をまとめると、多くの場合は“gé”と翻訳すれば良いが時間や場所を表す場合には“dé”となる。また、動詞によっては“din”，“ø”となるものが若干あることが分かる。

3.4 格助詞「と」

日本語とウイグル語の格助詞間の対応関係で、ばらつきが最も大きかったのは「と」である。今回の調査では385個出現したが、そのうちの239個(62.1%)がいわゆる引用を表す表現であった。例えば、「合格しようと決心した」の場合の「と」であり、ウイグル語では“dép”に対応する。なお、ウイグル語の“dép”は本来は格助詞ではなく、動詞「言う」“déydu”の連用形であり、「言って」に近いものがある。よって、「…と言う」をそのまま“… dép déydu”とすると、冗長さがあるため“dép”を省略し“… déydu”となることが多い。また、日本語の「…となる」という表現は「…になる」とほぼ同じであり、3.3節で述べたように、この場合も格助詞は“ø”となる。今回の調査では、こうした例が46個(12.0%)あった。

引用以外の「と」の用法としては、動作や状態を共にする相手を表す場合や(「先生と行く」)や、並列を表す場合(「AとB」)，比較の対象を表す場合(「彼と話が合わない)がある。これらの場合は、ウイグル語ではいずれも“bilén”となり、今回の調査では146個(37.9%)出現した。

「と」については、引用を示す場合は“dép”，それ以外の場合は“bilén”と翻訳すれば良いことが分かる。

3.5 その他の格助詞

その他の格助詞「で」「へ」「から」「より」については、今回の調査で、ほとんどの場合はウイグル語の助詞“dé”，“gé”，“din”，“din”にそれぞれ対応することが分かった。よって、機械翻訳の際にも、これらの格助詞は、そのまま対応するウイグル語の格助詞に翻訳すれば良いことが分かる。

4 両言語間の機械翻訳における格助詞の変換処理

前章で述べたように、日本語の一つの格助詞に、複数のウイグル語の格助詞が対応する場合がある。そのため、日本語-ウイグル語機械翻訳においては、概念語の訳語選択の他に、格助詞の訳語選択も適切に行う必要がある。

本論文では、動詞と格助詞の関係を利用することによって、この問題を解決する手法を提案する。格助詞の機能は、格助詞とその直前の名詞だけで決まるのではなく、それらによって構成される名詞句に係る動詞に依存する部分も大きい。例えば、「橋を」という名詞句を翻訳する場

表 2 格助詞のデフォルト訳

日本語格助詞	が	を	に	で	と	から	へ	より
デフォルト訳	ø	ni	dé	dé	bilén	din	gé	din

合、「橋 を 作る」という文であれば、「を」は対象を示していることからウイグル語の“ni”に訳され、全体としては“köwrükni salidu”となる。しかし、「橋 を 渡る」という文であれば、「を」は通過点を示しており、ウイグル語では“din”と訳され、全体としては“köwrükdin ötidu”となる。

そうした点を考慮すると、動詞ごとに、その格パターンとそれぞれの格パターンに対する格助詞の訳語を登録しておき、格助詞を翻訳する際には、その格助詞を含む名詞句に係る動詞の情報を利用して訳語を決定することが考えられる。

ところで、前章の調査から分かるように、日本語の格助詞の主な意味はウイグル語の一つの格助詞に対応する。そこで、一番出現頻度の多いウイグル語の格助詞をデフォルトの訳語として登録し、デフォルト訳では誤った翻訳を出力する格助詞についてのみ、訳語を変更することでも、この格助詞の選択を実現することが可能となる。

そこで、本論文で提案する手法では、表 1において一番出現頻度の多かったものを各格助詞のデフォルトの訳語として登録し、そのデフォルト訳では間違いとなる場合に、その名詞句に係る動詞の格パターンを利用して、格助詞の訳語を置き換える。ただし、「に」については表 1で出現頻度が一番多かった“gé”ではなく、“dé”の方をデフォルトとした。また「と」についても、表 1で出現頻度が一番多かった“dép”ではなく、“bilén”をデフォルトとした。その理由についてはそれぞれ 5.3節、5.4節で後述する。その結果、各格助詞に対するデフォルト訳は表 2のようになる。これにより、すべての動詞について格パターンを登録しなくとも、デフォルトから外れる格助詞が必要となる動詞について格パターンを登録することで、格助詞に対する翻訳精度を向上させることが可能となる。

以下では、次のような条件のもとで議論を進める。

- (1) 翻訳対象とする日本語文に出現するすべての単語について、ウイグル語訳が一つだけ辞書に登録されている。
- (2) 日本語文の形態素解析は正しく行われている。

今回は、格助詞以外の単語に関する訳語の多義性は考慮しなかった。例えば、3.2節で述べた「約束 を 守る」「子供 を 守る」の例では、「守る」に対するウイグル語訳が異なり、それに伴ない「を」に対応する格助詞も変化する。そのような場合も、動詞に対して適当な訳語を一つだけ与え、それに対応する格助詞を選択した。このような前提条件の下で、日本語の格助詞からウイグル語の格助詞への変換処理の手続きを次のようにまとめる。

図 1 提案手法による日本語-ウイグル語翻訳例

Step 1 入力された日本語文を形態素解析する。

Step 2 形態素解析された日本語の各単語を逐語訳する。格助詞にもデフォルトの訳が与えられる。

Step 3 単語に付加された格パターンの情報から訳語置換を行う。

Step 4 各形態素を接続してウイグル語文を生成する。

ここで、日本語-ウイグル語機械翻訳システムには、(小川他 2000)で提案されたシステムを使用した。(小川他 2000)の翻訳システムでは、上記の Step 1 と Step 2 を派生文法(清瀬 1989)に基づく形態素解析システム MAJO(小川, ムフタル, 外山, 稲垣 1999)で行っている。また、Step 4 に相当するモジュールも存在し、ウイグル語文の生成を行っている。そこで、本研究では、Step 3 に相当するモジュールを開発した。

(小川他 2000)の翻訳システムで使用している辞書には、各単語の情報が(日本語単語、品詞名、ウイグル語訳語)の3項組の形式で登録されており、例えば「渡る」については(渡 r, 子音幹動詞, öt-)の形式⁵で登録されている。今回の手法では、格パターンの情報が必要な単語に関してはウイグル語訳にその情報を付加して、(渡 r, 子音幹動詞, öt-{wo/-din})という形式で登録している。ここで、第3項のウイグル語訳語の後には、日本語の格助詞とウイグル語の格助詞をペアで登録する。なお、必要であれば、このペアはいくつでも登録可能である。また、こうした格パターンは動詞だけでなく、「逸脱」「獲得」といったサ変名詞や「やさしい」といった形容詞についても、必要に応じて登録することができる。

本手法を用いて「橋を渡る」がどのように翻訳されるかを、図1に示す。なお、途中の出力は、本来は(日本語単語、品詞名、ウイグル語訳語)という3項組であるが、スペースの都合で品詞名を省略した。本システムでは、形態素解析システム MAJO により入力日本語文の形態素解析(Step 1)と、各形態素の逐語訳(Step 2)が同時に行われる。この段階では格助詞「を」はデフォルトである「ni」に翻訳される。次の Step 3 において動詞「渡 r」のウイグル語訳に「öt-{wo/-din}」の形式で格パターンの情報が付加されていることから、格助詞の訳語の変更処

⁵ 派生文法に基づいているため、動詞は音韻論的な語幹が登録されている。

理を行う。ここで、係り受けの判定であるが、本機械翻訳システムは構文解析なしで翻訳することを目標の一つにしており、構文解析を利用した係り受け判定は行っていない。そこで、動詞より前に出現した格助詞の内、格パターンの情報に示された格助詞を置き換える。もしも、同じ格助詞が複数出現していた場合、その動詞に近い方の格助詞だけを置き換える。これは、係り受けにおける非交差性、後方修飾性の原則を考慮したものである。図1の例では、「渡る」“ötidu”の前に出現する「を」は一つだけなので、その訳語をデフォルトの“ni”から“din”に変更する。

5 実験と評価

前章で示した手法に基づき、日本語の格助詞からウイグル語の格助詞への翻訳に関して実験を行った。実験対象として、環境問題を扱った新聞社説など3編の日本語文138文を本システムを用いて翻訳し、生成された文に出現した295個の格助詞について、出現頻度の一番多い訳語を与えた場合と図1の本手法を用いた場合とを比較した。

ただし、以下のような格助詞は評価の対象外とした。

- (1) 「我が国」などの「が」
- (2) 「による」「に関して」のような慣用句の中に含まれる「に」
- (3) 「視野に置く」のような単語の組み合わせの中に含まれる「に」
- (4) 「買いに行く」のような動詞の連用形に接続する「に」
- (5) 引用を表す「と」
- (6) 直訳不能なもの

(1) については「我が」で一つの連体詞と考え、その訳語を“buning”として与えた。これは、所有を表す「が」の用法が限られているためである。よって、今回の実験では、この「が」は格助詞としては現れないため評価対象外とした。

(2) については「による」のウイグル語訳は“gé asasén”である。ここで、「に」は格助詞“gé”に対応している。しかし、残りの日本語動詞「よる」に対応するウイグル語“asasén”は動詞ではなく接続助詞である。また、「に関して」については、これ全体に“toğrisida”というウイグル語が対応している。そこで、このような例については、「について」「に関して」を一語として辞書に登録し、その訳語を、それぞれ“gé asasén”, “toğrisida”とした。

(3) については「視野に置く」でウイグル語の“közdé tutidu”に対応する。「視野」と「置く」に分解すると、この訳は得られないため、「視野に置く」で一つの単語として辞書に登録した。今回の実験では、このような慣用表現に含まれる「に」は評価しなかった。

(4) について、日本語大辞典(梅棹他 1995)では動詞の連用形に接続して目的を表す「に」を格助詞としている。しかし、今回の実験に使用した形態素解析システム MAJO では、この「に」を格助詞ではなく、動詞接尾辞「-i に」と解析する。これは、MAJO が使用している派生文法の分類に従ったものである。また、日本語-ウイグル語翻訳において、この「-i に」にはウイグ

表 3 格助詞の翻訳結果

日本語 格助詞	出現数	デフォルト (正訳率)		本手法 (正訳率)	
が	66	66	(100%)	66	(100%)
を	91	77	(84.6%)	91	(100%)
へ	1	1	(100%)	1	(100%)
に	63	40	(63.5%)	61	(96.8%)
から	13	13	(100%)	13	(100%)
より	2	2	(100%)	2	(100%)
で	30	30	(100%)	30	(100%)
と	28	28	(100%)	28	(100%)
合計	295	257	(87.1%)	293	(99.3%)

ル語の動詞接尾辞 “gili” が対応する。

(5) についても同様に、日本語大辞典 (梅棹他 1995) では、引用を示す「と」を格助詞としているが、派生文法では接続助詞⁶としている。これは、引用の「と」が体言以外にも多くの表現に接続可能であり、その意味で並列の意味の「と」とは品詞として異なると考えられるからである。よって、MAJO の解析結果においては、引用の「と」と並列の「と」には異なった品詞が付けられる。また、3.4節で示した通り、引用の「と」はウイグル語 “dép” に対応する。よって、本研究では、こうした「に」および「と」の区分は形態素解析のレベルで解決する問題として考え、MAJO の形態素解析の段階で格助詞にならない「に」および「と」については、今回の評価対象からは除外した。

なお、格助詞の「に」と動詞接尾辞「-i に」の区別、および、格助詞「と」と引用の「と」の区別は、MAJO だけでなく、例えば (舘, 米澤 1995) で提案されている文法でも行われている。

また、直訳不能な表現については、元の日本語を別の表現に置き換えて翻訳するなどの特別な処理が必要となるが、今回の実験システムでは逐語訳を基本とし、こうした処理を行っていないため直訳不能な表現は翻訳できない。本論文は格助詞の翻訳精度を調べるのが目的であるため、こうした文は評価の対象外とした。

表 3 に本実験の結果を示す。

5.1 「が」

「が」は今回の実験では、66 個出現したが、それらはすべて、「汚染 が 広がっている」 “moħit bulğinxø kengiywatidu”, 「異常気象など が 懸念されている」 “gélitilik ħawa kilimat ħatarlıqlarø éndixé ħiliniwatidu” のように主格の機能を持っており、すべて正しく翻訳できた。

今回の実験では 3.1 節で述べた「水 が 飲みたい」「私 が 飲みたい」のような表現は出現しな

⁶ 派生文法の用語では、「接続助辞」となる

かったが、これは動詞「飲む」や、その派生語である「飲みたい」の格パターンからでは区別ができないため、格助詞と動詞との関係だけで処理する本手法では、訳し分けができない。こうした訳し分けを実現するためには、名詞「水」、「私」の意味素性と動詞との関係を利用する必要があると考えられる。

5.2 「を」

「を」は今回の実験では、91 個出現したが、ウイグル語では対象を示さない用例が含まれており、デフォルトの“ni”に翻訳する手法では、14 個については正しく翻訳できなかった。しかし、動詞の格パターンを利用する本手法によって、すべてを正しく翻訳可能となった。この中には、「再生能力 を 超えて進む環境破壊」“*hasil iktidaridin ھاڭپىپ ilgirleydiğan mohit wéyranqiliği*”のように、「を」を“din”と翻訳したものが 5 個、「産業革命 を 促し」“*sanaét inkilawigé ھéydékqilik kilip*”のように，“gé”と翻訳したものが 6 個であった。

ただし、(IPAL 1987)を調査した段階では発見できなかった以下のような問題が見つかった。「四方 を 海に囲まれた」「生存 を 許されている」といった文では、動詞の受身形に「を」が格が係っている。通常、受身形には「が」格が係り、上記の文でも「を」を「が」と言い換えることができる。ウイグル語の場合も、上記の「を」は「が」のデフォルトの訳である“ø”と翻訳するのが正しい。こうした動詞は、本手法において「囲む」や「許す」の格パターンを登録しても解決できない。なぜなら「四方 を 囲む」や「生存 を 許す」という文においては「を」をデフォルトである“ni”と翻訳するのが正しいからである。そこで、今回は「囲まれる」「許される」を一つの動詞として辞書に登録し、それに「を」を“ø”という格パターンに登録することで翻訳を可能とした。今回の実験では、こうした取り扱いが 3ヶ所あった。

5.3 「に」

「に」については、デフォルトを“gé”ではなく、“dé”とした。これは、出現頻度では“gé”の方が多くても、“dé”と翻訳するべき場合に、動詞との関連が希薄であると考えられる例が多かったからである。例えば「今世紀中 に 特定フロンの全廃を目指す」という文では、「に」は時間を表わし“dé”と翻訳されるが、「目指す」の格パターンとして「に」を考えるよりは、「今世紀中 に」をひとまとまりと考えた方がよい。実際、「今世紀中 に」という表現においては、多くの場合に「に」は時間を表し、“dé”と翻訳するのが正しい。

一方、「化石燃料 に 替わる新たなエネルギー」や「宇宙全体 に かかわっている」の「に」は与格であり、“gé”と翻訳するのが正しい。この場合、この「に」は「替わる」や「かかわる」に依存していると考えるのが自然である。これは「に」が時間や場所を表す場合には動詞の必須格であることが少ないが、「に」が与格である場合には動詞の必須格であることが多いと考えることができる。

よって、「に」に対するデフォルトの訳を“dé”とし、与格を必須格とする動詞について、「に」を“gé”にするというルールを与えた。なお、「になる」「にする」という表現に対しては、3.3節で示したように、「に」がウイグル語において“ø”となる場合が多いので、「なる」「する」の2つの動詞については、「に」を“ø”とするルールを与えた。

表3では、単純に“gé”をデフォルトとした場合と、“dé”をデフォルトとし、本手法を適用した場合とを比較している。本手法を用いて置換を行ったのは44ヶ所あり、その内の40個が「化石燃料に替わる新たなエネルギー」“miniral yekilğugé orun basidiğan yingi energiyi”のように、“gé”に置換され、4個が「ゴミの捨て場になる海」“éhlétning töküx oruniğ bolidiğan dengiz”のように、“ø”に置換されていた。翻訳失敗した2個のうち、一つは係り受けの解析を間違えたものである。本実験のシステムは、精密な係り受けの解析を行わず、動詞の前に出現した格助詞のうちの最初の語を置き換えているため、格助詞を含む名詞句が省略されている場合などには、その動詞に係っていない名詞句の格助詞を置き換えてしまう。

もう一つは「人間はここに自然から切り離された存在となった」“insan buğ yér tébi’ettin ayirilğan méwjudiyét bilén boldi”という文である。本システムは、「なる」に登録された格パターンの規則を使用して「に」を“ø”に置換する。「存在になった」という文においては「に」を“ø”と翻訳するのが正しいが、今回は「存在となった」であったため、それより前の「ここに」の方を置換してしまった。「ここに」の「に」は場所を示しており、“dé”と翻訳されるのが正しい。こうした誤りを防ぐためには、名詞の意味素性などの導入が必要と考えられる。

5.4 「と」

今回の実験では格助詞の「と」は28個出現したが、「自然と明確に区別され」“tébi’et bilén eniğ pérqléndürölüp”, 「地球環境の悪化と人類の将来への危惧」“yér xari moñit naqarlixix bilén insanlarning kélgüsige bolğan éndixé”のように、すべてウイグル語の“bilén”に対応していた。これは、引用の「と」を形態素解析の段階で区別したためである。引用の「と」は52個出現したが、これらは「地球が水惑星と呼ばれ」“yér xari su séyyarisi dép atilip”のように“dép”となったのが20個、「核戦争という滅亡の危機」“yadro uruxiğ déydiğan ħalakétning kirizisi”のように、「と言う」「となる」の組み合わせで出現し、“dép”が省略され“ø”になったものが32個であった。なお、表3には、格助詞の「と」の出現数のみ記録してある。

ただし、これは形態素解析が正しく行われているという前提に依拠するものであり、実際には「と」のタグ付けは形態素解析においても難しい問題であり、MAJOでも完璧ではない。

5.5 その他の接尾辞

その他の接尾辞については、デフォルトの訳を与えることによって正しい翻訳ができた。そのため、それらの接尾辞の翻訳のための格パターンは特に登録しなかった。

上記のように、改良できる余地と困難な問題とがあるが、提案手法によって、日本語-ウイグル語機械翻訳における格助詞の翻訳の品質向上が達成できたことが分かる。

6 おわりに

本論文では、日本語とウイグル語の格助詞間の対応関係を詳細に調べ、動詞と格助詞の対訳との関連を明確にした。

これまでの研究では、日本語-ウイグル語機械翻訳のみならず、日韓機械翻訳を含んだ研究においても、日本語とそれらの言語との類似点を指摘しているが、助詞に関する一致点と差異を具体的及び統計的に示した例はほとんどない。単語の文字列としての近さを基準にした言語の分類に関する研究の一つ (Batagelj, Pisanski, and Keržič 1992) では、日本語とアルタイ言語系の一つであるトルコ語は、65 個の言語の中で非常に近い関係にあることが示されているが、本論文での調査結果は、両言語の近さ⁷をもう一つの観点から示すことができたとも言える。

さらに、動詞の格パターンを利用して格助詞の訳語を選択する手法を提案し、それを組み込んだ日本語-ウイグル語形態素解析システムを実現した。さらに、実験によりその有効性を示した。本実験は、比較的規模の小さいデータで行ったが、我々が示した変換処理の方法の有効性をよく示せたと考えている。今後は、比較的規模の大きいデータを対象にした実験を行い、品質の高い機械翻訳システムの実現を目指したい。

ただ、本研究では、以下の二つの前提をもって翻訳を行っている。一つは、格助詞の翻訳を行う前に動詞が一意に決定できるというものである。今回提案したシステムでは、動詞の訳語が決定したあと、その動詞に付与された格パターンから助詞に対する訳語を決定している。しかし、実際には動詞にも多義性があり、3.2節で述べた「約束を守る」「子供を守る」の例のようにその訳語を決定しなければ格助詞の訳が決まらない場合がある。しかし、動詞の多義性が解消できれば、本論文の手法が適用できることはもちろんである。

また、それとは逆に、格パターンから動詞の訳語が決定できる場合がある。例えば、「教える」に相当するウイグル語には “oqutidu”, “ügétidu” の二つがある。ここで、「先生が生徒を教える」, 「先生が生徒に水泳を教える」を、ウイグル語に翻訳すると、それぞれ “oqutquçi oquğuçılarni oqutidu”, “oqutquçi oquğuçılargé su üzüxni ügétidu” となる。つまり, “oqutidu” と “ügétidu” は、意味はほとんど同じであるが、格パターンが異なるのである。我々は現在、こうした名詞との組み合わせや、出現した動詞の格パターンにより動詞の訳語選択を実現する手法について研究を進めている。

もう一つは、目標言語であるウイグル語に出現する格助詞は、原言語である日本語の方にも必ず現れている、という前提である。しかし、日本語とウイグル語との間の概念の捉え方の違いにより、この前提が成り立たない場合がある。例えば、「このリンゴを3つ買う」 “Bu almidin

⁷ トルコ語とウイグル語は、チュルク諸語の言語である。

3 ni setiwalidu”,「あの封筒 を 10 枚下さい」“Awu konwertdin 10 ni bering” のような文では、日本語文には格助詞「を」だけがあるのに対し、ウイグル語文では、“ni” と “din” が出現している。これらのウイグル語表現を日本語に直訳すると、それぞれ「このリング から 3 つを買う」「あの封筒 から 10 枚を下さい」となる。しかし、ここでの「3 つ」、「10 枚」を取り除くと、「このリング を 買う」“Bu almini setiwalidu”,「あの封筒 を 下さい」“Awu konwertni bering” のようになり、格助詞 “din” は出現せず、「リング」や「封筒」が動作の対象となる。日本語では「リング」や「封筒」がまず動作の対象に設定され、それからその対象の「数」や「量」でその動作が副詞的に修飾されるのに対し、ウイグル語では、「数」や「量」があれば、それが動作の主対象になり、そうでない場合には、日本語と同じ格構造になるのである。また、「原稿 を 3 通送った」“Orginaldin 3 ni éwéttim”,「原稿 を 3 回送った」“Orginalni 3 kıtim éwéttim” の例からも両言語の捉え方の違いが分かる。ウイグル語では、「3 回」は動作対象にならないが、「3 通」は動作対象になるのである。こうした点も、今後の課題として検討していく必要がある。

参考文献

- Batagelj, V., Pisanski, T., and Keržič, D. (1992). “Automatic Clustering of Languages.” In *Computational Linguistics*, Vol. 18, pp. 339–352.
- IPAL (1987). 計算機用日本語基本動詞辞書 IPAL(Basic Verbs). 情報処理振興事業協会技術センター.
- ムフタル・マフスット, 外山勝彦, 稲垣康善 (1993). “日本語とウイグル語の助詞の対応関係とその機械翻訳への応用.” 電子情報通信学会技術研究報告 NLC 93-60, 電子情報通信学会.
- Musa, É. (1985). “Házırkı zaman Uyğur tili.” In *Şinjiang Maarip Neşriyati*, pp. 244–280.
- 小川泰弘, ムフタル, 外山勝彦, 稲垣康善 (1999). “派生文法による日本語形態素解析.” 情報処理学会論文誌, **40** (3), pp. 1080–1090.
- 小川泰弘, ムフタル・マフスット, 杉野花津江, 外山勝彦, 稲垣康善 (2000). “日本語-ウイグル語機械翻訳における派生文法に基づくウイグル語動詞句の生成.” 自然言語処理, **7** (3), pp. 57–77.
- ムフタル・マフスット (1993). “日本語-ウイグル語機械翻訳システムに関する研究.” Master’s thesis, 名古屋大学大学院工学研究科情報工学専攻.
- 金政仁 大駒誠一 (1996). “日韓機械翻訳における拡張翻訳テーブルを用いた韓国語の生成方法.” 情報処理学会論文誌, **37** (9), pp. 1697–1707.
- 清瀬 義三郎則府 (1989). 日本語文法新論-派生文法序説-. 桜楓社.
- 清瀬 義三郎則府 (1991). 日本語学とアルタイ語学. 明治書院.
- 梅棹忠夫, 金田一春彦, 阪倉篤義, 日野原重明 (編) (1995). カラー版 日本語大辞典 第二版. 講談社.

朴哲済, 李鍾赫, 李根培, 笈捷彦 (1997). “日-韓機械翻訳における連語パターンを用いた変換手法.” 情報処理学会論文誌, **38** (4), pp. 707-718.

渕武志 米澤明憲 (1995). “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, **2** (4), pp. 37-65.

竹内和夫 (1991). 現代ウイグル語四週間. 大学書林.

馬渕和夫 (編) (1986). 日本語の起源—世界の言語学者による論集. 武蔵野書院.

崔杞鮮 金泰完 (1996). “日韓機械翻訳システムの現状および分析.” 言語処理学会第2回年次大会発表論文集, pp. 433-443.

李義東, 中嶋正之, 安居院猛 (1989). “語と語の関係を用いた意味解析による日韓単文機械翻訳システム.” 電子情報通信学会論文誌 D-II, **J72-D-II** (10), pp. 1689-1695.

付録

A ウイグル語文字体系

現在使われているウイグル文字は、ほとんどがアラビック文字から取り入れられ、一部が新たに追加された32の文字からなる文字セットであり、アラビア語と同じ右から左へ書かれる。それとは別に、ウイグル語をローマ字表記で表わす場合もあるが特に決った体系がない。我々が用いたローマ字体系とアラビックベースの文字体系の対応関係を表4に示す。

表4 ウイグル文字とローマ字表記の対応

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
ئا	ب	چ	د	ئې	ئە	ف	گ	غ	خ	ھ	ئى	ج	ك	ق	ل	م	
A,a	B,b	C,c	Ç,ç	D,d	E,e	É,é	F,f	G,g	Ğ,ğ	H,h	Ĥ,ĥ	I,i	J,j	K,k	Q,q	L,l	M,m
ن	و	ۆ	پ		ر	س	ش	ت	ۇ	ۈ	ۋ		ي	ز	ژ	ڭ	
N,n	O,o	Ö,ö	P,p	Q,q	R,r	S,s	Ş,ş	T,t	U,u	Ü,ü	V,v	W,w	X,x	Y,y	Z,z	Z,ž	NG,ng

略歴

小川 泰弘: 1995年名古屋大学工学部情報工学科卒業。2000年同大学院工学研究科情報工学専攻博士課程後期課程修了。同年より、名古屋大学助手。自然言語処理に関する研究に従事。言語処理学会、情報処理学会各会員。

ムフタル マフスット: 1983年新疆大学数系卒業。1996年名古屋大学大学院工学研究科情報工学専攻博士課程満了。同年、三重大学助手。現在、名古屋大学計算理工学専攻稲垣研究室特別研究員。自然言語処理に関する研究に従事。

人工知能学会，情報処理学会各会員．

稲垣 康善： 1962 年名古屋大学工学部電子工学科卒業．1967 年同大学院博士課程修了．同大助教授，三重大学教授を経て，1981 年より名古屋大学工学部・大学院工学研究科教授．工学博士．この間，スイッチング回路理論，オートマトン・言語理論，計算論，ソフトウェア基礎論，並列処理論，代数的仕様記述法，人工知能基礎論，自然言語処理などの研究に従事．言語処理学会，情報処理学会，電子情報通信学会，人工知能学会，電気学会，日本ソフトウェア科学会，日本 OR 学会，IEEE，ACM，EATCS 各会員．

(2001 年 1 月 10 日 受付)

(2001 年 4 月 13 日 採録)