E-012

英語-ウイグル語機械翻訳システムの基本検討

～日本語とウイグル語の文法構造の比較～

An English – Uyghur Machine Translation System
– Comparative Study on Japanese and Uyghur Grammar –

カデル・プラト† Polat Kadir　**山田剛一**† Yamada Koichi　**絹川博之**† Kinukawa Hiroshi

1. はじめに

ウイグル語はアルタイ語系に属し、中国新疆ウイグル自治区で暮らすウイグル族が使用している言語である。我々は英語-ウイグル語機械翻訳システムの構築を目指しており、現在はシステムの基盤となる文法規則について検討を行っている[1]。

日本語とウイグル語は同じ膠着語であり、形態素、構文などの多くの面で類似点が見られる[2][3][4]。

英語から日本語への翻訳の研究には長い歴史がある。日本語とウイグル語の文法的な類似点と相違点を明確にすることによって、英日機械翻訳の成果を英語-ウイグル語機械翻訳に応用できると考えられる。本稿では日本語とウイグル語の文法構造を比較検討し、その内容をまとめる。なお、日本語とウイグル語の形態素および構文の対応関係に伴う意味対応については別に詳しく検討を行う。

2. 日本語とウイグル語の文法構造の比較

2.1 形態素レベルでの比較

日本語と同じように、ウイグル語の語は、詞（単独で文の構成要素となるもの）と接辞（詞に付加的な情報を加えるもの）に分類される。形態素レベルにおいては、両言語とも詞に接辞が連接することによって様々な文法的な働きを示すといった膠着語の特徴を持つ。ウイグル語の語構成の例を以下に示す：（語幹：詞、その他：接辞）

- oy（語幹）-ler(複数)-imiz(人称)-ge(格)（訳：私たちの家へ）
- electron（語幹）-lash(動詞化)-turul（使役の受身）-ghan(形容詞化)-likh(名詞化)-tin(原因)（訳：電子化されたから）

日本語とウイグル語は詞や接辞の単位、接辞の順序などが全体的によく似ているが、いくつかの違いも存在する。以下では両言語においての相違点を中心に、名詞、動詞などにおける詞・接辞の対応関係について述べる。

2.1.1 名詞

ウイグル語の名詞には人称があり、名詞が所有の対象となる場合のみ、一人称・二人称・三人称の接辞が接続する。しかも、その接辞には単数複数による形態の違いがある。kitab(本)の例を表1に示す。

2.1.2 動詞

日本語の動詞には活用変化があり、動詞語幹に語尾が接続し、さらに助動詞や助詞などが接続するのに対して、ウイグル語では、活用変化がなく動詞の語幹に接辞が直接接続するという点が異なる。

表1 名詞と人称接辞の例

	単数	複数
1人称	kitab-im	kitab-imiz
2人称	kitab-ing	kitab-inglar
3人称	kitab-i	kitab-i

ただ、「活用」という概念にとらわれずに考えれば、両言語の動詞の接辞接続には類似性があり、日本語の使役、受身、否定、希望などの助動詞にそれぞれウイグル語の接辞が対応している。動詞「書く」の使役形とそのウイグル語での対応の例を表2に示す。

表2 動詞語形変化の対応

日本語			ウイグル語	
語幹	語尾	助動詞	語幹	接辞
書	く		yaz	
書	か	せる	yaz	ghuz
書	か	される	yaz	ghuzul

日本語と異なるもう一つの点として、ウイグル語の動詞は、主語の人称の変化に対応した人称接辞が接続し、その人称接辞には単数、複数、時制などの意味も含まれている(表3)。

表3 動詞の人称変化

	数	現在形	過去形
1人称	単	yaz-imen	yaz-dim
	複	yaz-imiz	yaz-dukh
2人称	単	yaz-isen	yaz-ding
	複	yaz-isi-ler	yaz-ding-lar
3人称	単	yaz-idu	yaz-di
	複	yaz-idu	yaz-di

ウイグル語の述部における文法範疇の順序は次の2点を除いて日本語と同じである（表4）：

i. 日本語の動詞は活用変化するのに対して、ウイグル語の動詞は活用変化しない。

† 東京電機大学大学院工学研究科情報通信工学専攻

表4　述部(動詞)の文法範疇の順序

動詞		態	可能	相	丁寧さ	否定	時	不定	人称接辞	終助詞
語幹	語尾									
書く	○	せる	かける	ている	ます	ない	た	だろう	×	ね
yaz	×	ghuz	ala	iwat	idu	ma	di	ghandu	○	he
		随意							必須	随意

花子は　新しい　先生　について　太郎 に　詳しく　聞いて　みた。
↓　↓　↓　↓　↓　↓　↓　↓　↓
Hanako　yengi　muellim　toghrisida　Taro din　tepsili　sorap　bakhti.

花子は　太郎　に　新しい　先生　について　詳しく　聞いて　みた。
↓　↓　↓　↓　↓　↓　↓　↓　↓
Hanako　Taro din　yengi　muellim　toghrisida　tepsili　sorap　bakhti.

図1．日本語とウイグル語の語順の対応と入れ替え

ii.　日本語の動詞には人称を表す接辞が不要である。一方、ウイグル語の動詞には人称の違いによる人称接辞の接続が必須である。

2.2 構文レベルでの比較

ウイグル語の文の構成単位は、詞に0個以上の接辞からなるものを文節とみなせば、日本語と同じである。本節では日本語とウイグル語の、語順、格助詞の対応関係、文の係り受け構造、人称と単複数の一致について比較検討する。

2.2.1　語順

日本語と同様に、ウイグル語はSOVの語順を持ち、修飾語は被修飾語の前に位置する。そして、語順の入れ替えが比較的に自由という点でも非常に類似している。図1の例で示すように、日本語文の語順を変えた場合、その訳文となるウイグル語文の語順も同じように変化する。

2.2.2　格助詞の対応関係

日本語とウイグル語に共通する大きな特徴のひとつとして格助詞の存在がある。両言語において、格助詞は語の文法的な働きを示す重要な役割を果たしている。日本語の格助詞とウイグル語の格助詞には対応関係があるが、意味的に対応しない例外部分も若干ある。

例えば、日本語の格助詞「を」は、話し言葉以外では省略されることはないが、ウイグル語では、不定のオブジェクトを示すときには目的格をあらわす格助詞を省略しなければならない。

格助詞の対応関係については、文献[3]に詳しい議論がある。

2.2.3　係り受け構造

日本語の係り受け構造においては、係り先の唯一性、係り受けの非交差性、後方修飾性などの制約[5]があり、これらはウイグル語においても同じように存在する。これらの制約を満たす語順の入れ替えに関し、ウイグル語は日本語と全く同様に可能であり、語順入れ替え文も正文となる(図1)。

2.2.4 人称と単複数の一致

日本語の文では、主語の人称あるいは単数複数は動詞に影響を与えないが、ウイグル語は日本語と異なって、名詞や動詞に人称がある（2.1 節を参照）。そのため、構文構成上、ウイグル語では文の主語となる名詞や代名詞の人称と単複数に呼応して、動詞にそれに対応する接辞が接続していることが必要である。

3.　おわりに

我々は、日本語とウイグル語の類似性に着目することによって、英日機械翻訳の成果を最大限利用しながら英語-ウイグル語機械翻訳を実現していきたいと考えている。本稿では、両言語の形態素及び構文レベルでの類似点と相違点について比較検討したが、今後の研究を進めるにあたってはさらに詳しい検討が必要であると考えている。

参考文献

[１]カデル・プラト、山田剛一、絹川博之："英語-ウイグル語機械翻訳システムの基本検討"、情報処理学会第66回全国大会論文集 (2004)

[２]小川泰弘、Muhtar Mahsut、外山勝彦、稲垣康善："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-、自然言語処理120-1 (1997)

[３]Muhtar Mahsut、小川泰弘、稲垣康善："日本語-ウイグル語機械翻訳のための格助詞の変換処理"、自然言語処理Vol.8　No.3　(2001)

[４]小川泰弘、Muhtar Mahsut、杉野花津江、外山勝彦、稲垣康善："派生文法に基づく日本語動詞句のウイグル語への翻訳"、Vol.7 No.3 (2000)

[５]渡邊善之, 松原茂樹, 外山勝彦, 稲垣康善："英日同時翻訳のための漸進的日本語生成", 言語処理学会第6回年次大会発表論文集 (2000)