

アイヌ語から日本語への漸進的な翻訳処理に関する基礎的考察

桃内佳雄<sup>\*</sup>・大友雄介<sup>†</sup>・越前谷博<sup>\*</sup>

Fundamental Studies of Incremental Translation  
from the Ainu Language into Japanese

Yoshio MOMOUCHI<sup>\*</sup>, Yusuke OTOMO<sup>†</sup> and Hiroshi ECHIZEN-YA<sup>\*</sup>

あらまし

アイヌ語と日本語は構文的に類似した点が多く、アイヌ語から日本語への翻訳処理において、アイヌ語文を構成する各単語の逐語翻訳の結果得られる日本語の単語列で、そのおよその意味を理解することができると思われる。しかし、アイヌ語文と日本語文で語順が一致しない場合がある、また日本語の格助詞に対応するいくつかの要素がアイヌ語にはない、などの相違点も存在する。従って、逐語翻訳によって得られた日本語の単語列が必ずしも自然な日本語文となるとは限らず、より自然な日本語文を得るためには、逐語訳からさらに進んだ処理が必要である。本報告では、アイヌ語文を構成する各単語の逐語翻訳を出発点として、より自然な日本語文を得るためには、さらにどのような処理が必要になるか、また、それらの処理をどのような翻訳解析モデルとして構築したらよいかということについての基礎的な考察を試みる。

1. はじめに

アイヌ語と日本語は構文的に類似した点が多く、アイヌ語から日本語への翻訳処理の出発点として、アイヌ語文を構成する各単語の逐語翻訳の結果得られる日本語の単語列で、そのおよ

<sup>\*</sup> 北海学園大学工学部電子情報工学科  
Department of Electronics and Information Engineering, Faculty of Engineering, Hokkai-Gakuen University  
<sup>†</sup> 北海学園大学大学院工学研究科電子情報工学専攻  
Division of Electronics and Information Engineering, Graduate School of Engineering, Hokkai-Gakuen University

その意味を理解することができると思われる。しかし、アイヌ語と日本語の文で単語の語順が一致しない場合がある。アイヌ語には日本語における格助詞に対応するいくつかの要素がない。また、日本語にはない名詞の所屬形が存在する、などの相違点も多く存在する[1]。従って、逐語翻訳によって得られる日本語の単語列が必ずしも正しい日本語文となるとは限らず、より自然で、正しい日本語文を得るためには、逐語訳からさらに進んだ処理が必要である。本報告では、アイヌ語文を構成する各単語の逐語翻訳を出発点として、より自然な日本語文を得るために、そこから先、どのような処理が必要になるか、また、それらの処理をどのような翻訳解析モデルとして構築したらよいかということについての基礎的な考察を試みる。

漸進的な自然言語解析は、例えば、文の解析において、単語が入力されるごとに、その単語が入力されたところまでの部分的な構造と意味の解析を行い、その結果を段階的に積み上げて行きながら、文全体の処理を終えたときに、その文全体の構造と意味の解析を終えているという解析手法である[2]。解析過程で、曖昧な部分や情報が不完全で決められない部分などがある場合、その状態をそのまま保持して、より後からの情報を得て、その解決を図りながら解析を進めて行く。このような解析手法を文の翻訳処理にも適用することができるであろう。つまり、単語が入力されるごとに、その単語が入力されたところまでの部分的な翻訳処理を行い、その結果を段階的に積み上げて行きながら、文全体を処理し終えたときに、その文の翻訳処理を終えているという翻訳手法として構成することができる。

従来の多くの機械翻訳システムが採用している構文トランスファー方式では、まず、解析処理で原言語の文の構文(意味)構造を抽出し、その後、変換(トランスファー: transfer)、生成という処理を経て目的言語の文が出力される[3]。解析が終わって変換、変換が終わって生成という処理単位ごとの段階的な処理過程となっている。漸進的な翻訳手法では、解析から生成までの処理が融合した形で翻訳処理が進められる。ほぼ線形的に発話される音声の翻訳システムにおいては、漸進的な翻訳手法が有効であると考えられ、今後、アイヌ語の音声からの翻訳システムの構築を考えると、本考察の結果の自然な拡張が可能となることが期待される。

## 2. アイヌ語-日本語対訳テキストベースの作成

本考察における基本的な資料として、次の著作におけるアイヌ語-日本語対訳を参考にした。  
 (B1) 中川裕, 中本ムツ子: エクスプレス アイヌ語, 白水社, 1997。  
 (B2) 中本ムツ子, 片山竜峯: アイヌの知恵 ウバシクマ[1], 片山言語文化研究所, 1999。  
 (B3) 田村すず子: アイヌ語, 言語学大辞典セレクション: 日本列島の言語, 三省堂, 1997。  
 B1とB2については、著作中のアイヌ語-日本語対訳を参考にして、アイヌ語-日本語対訳テキストベースを作成した。それぞれのテキストベースを、EXP, UPAと名づけた。それ

これらのアイヌ語-日本語対訳テキストベースの基本的な構成要素は、付加コードを付与した、アイヌ語、逐語訳、品詞列、日本語の4つの組である。品詞列はアイヌ語に対応する品詞列である。例を以下に示す。

例1 exp 0100601: usey e=ku rusuy ya? (アイヌ語) あなたは飲みたいか? (日本語)  
 exp 0100602: お湯 あなた=飲む たい 辞類か? (アイヌ語) あなたは飲みたいか? (日本語)  
 exp 0100603: 名詞 人接=他動詞 助動詞 終助詞? (アイヌ語) あなたは飲みたいか? (日本語)  
 exp 0100604: お湯を飲みたいか? (日本語)

例2 exp 0100605: テキストベースの識別記号 (exp, upa) が記号として付与されたアイヌ語の例として、01 章、節、単元、など (文) の識別記号として付与されたアイヌ語の例として、006 片、レコード (文、節) 番号の識別記号として付与されたアイヌ語の例として、01 アイヌ語 ([01])、02 逐語訳 ([02])、03 品詞列 ([03])、04 日本語 ([04]) の識別記号として付与されたアイヌ語の例として、

アイヌ語には書き言葉がなく、話し言葉に対応してアルファベット表記あるいはカタカナ表記が行われているが、本報告では、アルファベット表記を対象とする。従って、B1、B2、B3に依拠して、アルファベット表記のアイヌ語文の翻訳処理について考察することになる。

アイヌ語における品詞の種類は次のように設定した [1, 8, 9]。  
 名詞 (普通名詞、固有名詞、代名詞、位置名詞、形式名詞、数詞)；  
 動詞 (完全動詞、自動詞、他動詞、複他動詞)；  
 連体詞；  
 副詞；  
 接続詞；  
 助動詞 (格助詞、副助詞、接続助詞、終助詞)；  
 人稱接辞 (人接)；  
 接辞；

アイヌ語では形容詞という独立の品詞がなく、日本語の形容詞で表現される様態はたいていの場合自動詞によって表現される [3]。[例: poro: 自動詞: 大きくある (大きい)、大きくなる]。

### 3. アイヌ語-日本語逐語翻訳とその基本的な問題点

アイヌ語-日本語逐語翻訳は、基本的には、アイヌ語に対応する、対訳辞書中の日本語をその訳として割り当てる翻訳と考える。逐語翻訳の基本規則は次のようにまとめられる。

語・アイヌ語-日本語対訳辞書中の対応する日本語訳を割り当てる。アイヌ語-日本語対訳辞書の基本的な構成は次の辞書を参考にして作成する予定である。

(D1) 中川裕：アイヌ語千歳方言辞典、草風館、1995。

(D2) 田村すず子：アイヌ語沙流方言辞典、草風館、1996。

(D3) 萱野茂：萱野茂のアイヌ語辞典、三省堂、1996。

・人称接辞には、人称の代名詞に対応する逐語訳をあてる。格助詞は付加しない。

・名詞の所属形の逐語訳には、「の」を付加する。

・名詞のうち位置名詞の逐語訳には、「の」を付加する。

③・動詞、助動詞の逐語訳には、終止形をあてる。

逐語訳の具体的な例と基本的な問題点についてみてみよう。アイヌ語-日本語対訳辞書におけるアイヌ語-日本語対訳情報として、ここでは、簡単のために次のような情報だけを考える。

この辞書情報の詳細化、特に多義の問題については後に検討する。

アイヌ語	品詞	日本語	品詞
usey	名詞	お湯	名詞
e	人称接辞	あなた	代名詞
ku	他動詞	飲む	他動詞
rusuy	助動詞	たい	助動詞
ya	終助詞	か	終助詞

<アイヌ語文> usey e=ku rusuy ya?

名詞 人接=他動詞 助動詞 終助詞

↓ ↓ ↓ ↓ ↓

<逐語訳> お湯 あなた=飲む たい か ?

名詞 代名詞 他動詞 助動詞 終助詞

この逐語訳は、自然な日本語文とは言えず、引き続き、次のような処理が必要であると考える。

<逐語訳> お湯 あなた=飲む たい か ?

↓ ↓ ↓ ↓ ↓

<格助詞付加> <格助詞付加> たい か ?

↓ ↓ ↓ ↓ ↓

お湯を あなたが=飲む たい か ?

↓ ↓ ↓ ↓ ↓

お湯を φ あなたが=飲む たい か ?

↓ ↓ ↓ ↓ ↓

<日本語文> お湯を φ 飲み たい か ? (φ:ゼロ化(省略))

逐語訳を自然な日本語文へ変換するために、名詞の格助詞付加、人称接辞の格助詞付加、人称接辞のゼロ化(省略)、そして動詞の語形変化という処理が必要であることが理解されるであろう。また、これらの処理は、主として、文法的な知識(動詞の意味)と文脈に依存して進められる。アイヌ語のさらに多くの文の逐語訳を自然な日本語文へ変換するためには、ここで述べた処理だけでは不十分である。さらに、どのような処理が必要となるであろうか。正し

#### 4. 逐語訳から日本語文への基本的な変換処理

前章で考察した変換処理も含めて、本章では、逐語訳を自然な日本語文へ変換するために必要な基本的な処理について考察する。具体的な例に即して、漸進的な処理ということを含頭におきながらも、その枠組みにそれほど強く制約されない方向で検討を進める。

##### (1) 格助詞付加、人称接辞ゼロ化、語形変化

usey	e=ku	rusuy	ya?
↓	↓ ↓	↓	↓
お湯	あなた=飲む	たい	か?
名詞	人接=他動詞	助動詞	終助詞

このままでは、自然な日本語文とは言えず、次のような処理が必要になる。

お湯      あなた=飲む      たい      か?

##### <格助詞付加> <格助詞付加>

お湯を      あなたが=飲む      たい      か?

##### <人接ゼロ化> <語形変化>

お湯を      飲み      たい      か?

逐語訳から自然な日本語文へと変換するのに、名詞の格助詞付加(主格(が格)、目的格(を格)、に格)、人称接辞の格助詞付加、人称接辞のゼロ化(省略)、そして動詞の語形変化(活用処理)という処理が必要である。

##### (2) の格化、の付加

ku=kor	totto	noya	ham	uk
↓ ↓	↓	↓	↓ ↓	↓
私=持つ	母	よもぎ	葉	採る
人接=他動詞	名詞	名詞	名詞	他動詞

##### <格助詞付加>

私が=持つ      母      よもぎ      葉      採る

##### <の格化>

私が      母      よもぎ      葉      採る

私の 母 よもぎの 薬を 採る。  
 <格助詞付加>

私の 母が よもぎの 薬を 採る。

私の 母が よもぎの 薬を 採る。  
 <の付加><格助詞付加>

私の 母が よもぎの 薬を 採る。  
 <語形変化>

私の 母が よもぎの 薬を 採った。

<の格化>は、「ku=kor」の述語訳+格助詞付加の結果である「私が=持つ」を「私の」とする処理である。また、<の付加>は、「よもぎ」と「薬」の名詞並びを「よもぎの薬」とする処理である。「ku=kor」がつねに「私の」に変化するわけではなく、また、名詞並び「A B」がつねに「AのB」に変化するわけではないところに問題がある。次のような例がある。

soy ta an cikuni ku=kor wa cise onnayke k=omare.  
 ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓

外 に ある 薪を 私=持つ て、家 中 の中、私=入れる。  
 代名詞 格助詞 自動詞 名詞 人接=他動詞 接続詞 名詞 位置名詞 人接=複他動  
 外(に)に(ある) 薪(を) (私) 持つ(て) 家(中)の中(に) 私(を) 入れる。  
 (soy) (ta) (an) (cikuni) (ku=kor) (wa) (cise) (onnayke) (k=omare)  
 「wa (て)」が接続助詞であって、「の」に接続助詞「て」が連接することはない」という規則が働く。

kani isepo kina k=ere.  
 ↓ ↓ ↓ ↓ ↓ ↓

私 うさぎに 草を 私=食べさせる。  
 代名詞 名詞(人接) 名詞(自) 人接=複他動詞

私が うさぎに 草を 食べさせる。  
 (うさぎに 草を)

「食べさせる」の結合値パターンが [(hum) が, (anim) に, (food) を] であることが制約となる。代名詞の「私」と人称接辞の「私」の重複の処理も必要である。最後の語形変化は、「採る=>採った」のタ形への変化であるが、これは文脈(状況)に依存していて、文レベルの処理を超えているが、一つの例としてここに示した。

### (3) 所屬形連結

ku=tekehe piro hi ta.  
 ↓ ↓ ↓ ↓ ↓ ↓

私=の手 傷つく(て)取るときに、



化できない。つまり、日本語では、“先祖”という名詞が「場所を占める」という意味特徴を持っていないので、“の所”をゼロ化することができないといえる。

(6) 語順変化

ku=sinki	kusu	somo	k=arpa	rusuy.
↓ ↓	↓	↓	↓ ↓	↓
わたし=疲れる	から	ない	わたし=行く	何さしたい。
人接=自動詞	接助	副詞	人接=自動詞	助動詞
<格助詞付加>				
わたしが=疲れる	から	ない	わたし=行く	たい。
<人接ゼロ化・語形変化>				
疲れた	から	ない	わたし=行く	たい。

アイヌ語の動詞は、格助詞の付加による格化(格助詞付加)と、語形変化(語形変化)の二つがある。アイヌ語の動詞は、格助詞の付加による格化(格助詞付加)と、語形変化(語形変化)の二つがある。アイヌ語の動詞は、格助詞の付加による格化(格助詞付加)と、語形変化(語形変化)の二つがある。

疲れた	から	ない	行き	たい。
<語順変化>				
疲れた	から	ない	行き	たくない。

否定の副詞“somo”が動詞の前におかれていて、日本語への翻訳では、語順の変化が必要となる。語形変化が“疲れる”なのか、“疲れた”なのかは文脈に依存する。アイヌ語では時制に関する動詞の語形変化あるいは語尾変化はないので、時制を過去と決定するためにはより広い意味での文脈情報が必要である。

以上の基本的な変換処理は、次のようにまとめられるであろう。

- [1] 付加処理 : 格助詞付加、の付加
  - [2] 削除処理 : 人称接辞ゼロ化、位置名詞ゼロ化
  - [3] 変形処理 : 語形変化、の格化
  - [4] 連結処理 : 所属形連結
  - [5] 並べ替え処理 : 語順変化
- これらに、さらに多義語の処理が加わる。
- [6] 多義語処理

アイヌ語には、多義語が多く存在する。多義語の曖昧さの解消は文脈に依存して行われるのが普通である。多義語の類型と例を次に示そう。



- ・品詞カテゴリが異なる。  
 kor : 持つ (他動詞), て (接続助詞)  
 e : 食べる (他動詞), はい (間投詞), 君 (か) (人称接辞)  
 ku : 私 (人称接辞), 飲む (他動詞)
- ・品詞カテゴリが同じで, 多義である。  
 kor : (他動詞) ①( )を持つ, ②を所有する, ③をつける/をかぶる, ④(子供)を生む
- ・品詞カテゴリは同じであるが, 下位区分カテゴリが異なる。  
 wa : て (接続助詞), よ (終助詞), から (格助詞)

## 5. アイヌ語から日本語への漸進的な翻訳処理の可能性

アイヌ語の文 (発話) を左から右へと読み (聞き) 進みながら, 日本語への漸進的な翻訳を進めてゆく機構の構成について, 基本的な考察を進める。アイヌ語の語を読み, 解析し, その結果である部分的な, あるいは曖昧な構造を保持しながら, 左から右へと漸進的に変換を進めるというしくみで, どこまで翻訳が可能であろうか, その可能性を探る。

### 5.1 漸進的な翻訳処理の基本的な枠組み

例えば, 前出の次のような例について考えてみよう。

<アイヌ語文>	usey	e=ku	rusuy	ya?
<逐語訳>	お湯	あなた=飲む	たい	か?
<格助詞付加>	お湯を	あなたが=飲む	たい	か?
		<人接ゼロ化>	<語形変化>	
<日本語文>	お湯を	φ	飲み	たいか? (φ:ゼロ化)

基本的な辞書の構成を次のように考える。

#### (a) アイヌ語-日本語対訳辞書

[ アイヌ語 : 品詞 品類 : 日本語 品詞 品類 ]  
 [ usey : 名詞 : お湯 : 名詞 ]









い問題である。連体修飾節というより大きな言語単位の構成と被連体語句との意味的關係についての解析が必要である。

## 6. おわりに

アイヌ語から日本語への漸進的な翻訳処理の可能性について基礎的な考察を行った。考察のもとになったアイヌ語のデータは書かれたアイヌ語の初歩的な、そして部分的なデータである。データの範囲と量をさらに広げてゆく必要がある。特に、日本語にはない語順の変換パターンや連語の処理についての考察が一つの大きな課題である。アイヌ語のより多くの様々な言語現象の基本的な処理のしくみを明らかにしながら、それらを漸進的な翻訳処理の枠組みの中にどのように埋め込んでゆくかについても引き続き考察を進めて行きたい。また、アイヌ語-日本語対訳辞書と日本語辞書の構成も進めなければならない。両方とも既存の辞書を参考にして、また、構築するテキストベースを基礎にして、その電子化を進めて行きたいと考えている。

## 謝 辞

本研究の一部は、北海道大学ハイテク・リサーチ・センター研究費による援助を受けて行われました。ここに記して謝意を表します。また、アイヌ語の文法についてご教示をいただいている本電子情報工学科 切替英雄先生に感謝いたします。

## 参 考 文 献

- [1] 田村すず子：アイヌ語，「言語学大辞典セレクション：日本列島の言語」，三省堂，1997。
- [2] 奥村学：漸進的な自然言語解析モデルについて，bit, Vol. 26, No. 8, 共立出版，1994。
- [3] 佐藤理史：機械翻訳，「長尾真編，自然言語処理，第12章」，岩波書店，1996。
- [4] 石綿敏雄・森野孝野：結合国から見た日本文法，「文法と意味」：2章，朝倉書店，1983。
- [5] 中川裕，中本ムツ子：「エクスプレス アイヌ語」，白水社，1997。
- [6] 中本ムツ子，片山竜華：「アイヌの知恵 ウパシクマ」[1]，片山言語文化研究所，1999。
- [7] 萱野茂：萱野茂のアイヌ語辞典，三省堂，1996。
- [8] 田村すず子：アイヌ語沙流方言辞典，草風館，1996。
- [9] 中川裕：アイヌ語千歳方言辞典，草風館，1995。