

アイヌ語・日本語機械翻訳のための基礎的考察

桃内佳雄 大友雄介 越前谷博
北海学園大学工学部

1. はじめに

アイヌ語と日本語は構文的に類似しており、アイヌ語から日本語への翻訳において、語から語への単語直接翻訳によっても、そのおよその意味を理解可能な翻訳結果が得られる。しかし、単語直接翻訳は、そのままでは自然な日本語文とは言えず、自然な日本語文を得るためには多くの問題点を解決する必要がある。本考察では、アイヌ語文の単語直接翻訳をより自然な日本語文へと変換するための基本的な変換処理について基礎的な考察を進める。また、慣用句などの複数の語からなる複合的な単位の自然な翻訳のための基礎として、複合翻訳単位の分類についても考察する。本考察の以下の記述において、アイヌ語の表記は「アコロイタク AKORITAK アイヌ語テキスト 1, (社)北海道ウタリ協会 [1]」に準拠して行う。ローマ字表記で、ほぼ単語を単位として分かち書きを行い、アクセント記号は省略している。また、アイヌ語に関する記述は文献 [1, 2, 3, 4, 5, 6] を全面的に参考にしていく。

2. アイヌ語・日本語対訳コーパスの作成

3章から5章の考察における基本的な資料として、文献 [2, 5, 6] におけるアイヌ語・日本語対訳を参考にした。また、文献 [5, 6] 中のアイヌ語・日本語対訳を基礎にして、アイヌ語・日本語対訳コーパス (EXP, UPA) を作成した。それらのアイヌ語・日本語対訳コーパスの基本的な構成要素は、付加コードを付与した、アイヌ語文、アイヌ語品詞列、単語直接翻訳、日本語品詞列、日本語文の5つの組である。その一部を以下に示す。

exp0100601	: usey e=ku rusuy ya ?
exp0100602	: 名詞 人接=他動詞 助動詞 終助詞 ?
exp0100603	: お湯 あなた=飲む たい か ?
exp0100604	: 名詞 代名詞=他動詞 助動詞 終助詞 ?
exp0100605	: お湯を飲みたいか?

3. 単語直接翻訳と基本的な問題点

アイヌ語・日本語単語直接翻訳は、基本的には、アイヌ語の語に対訳辞書中の対応する日本語をその訳として割りあてる翻訳と考える。単語直接翻訳(直

接翻訳と略記)の基本規則は次のようにまとめることができる。

- ①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、2章で述べたアイヌ語・日本語対訳コーパスと文献 [3, 4, 7] を参考にして進めている。
- ②人称接辞(人接と略記)には、人称の代名詞に対応する直接翻訳をあてる。格助詞は付加しない。
- ③名詞所属形の直接翻訳には「の」を付加する。
- ④位置名詞の直接翻訳には「の」を付加する。
- ⑤動詞、助動詞の直接翻訳には終止形をあてる。

単語直接翻訳の具体的な例と基本的な問題点についてみてみよう。また、ここでは、簡単のためにアイヌ語・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、次のような情報だけを考える。

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
?	?	?	?

<1>usey e=ku rusuy ya ?

名詞 人接=他動詞 助動詞 終助詞

↓ ↓ ↓ ↓ ↓
お湯 あなた=飲む たい か ?

名詞 代名詞=他動詞 助動詞 終助詞

この単語直接翻訳は、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られる。

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?

<人称接辞の化><語形処理>

お湯を 飲み たい か?

単語直接翻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φ(ゼロ)化、そして動詞の語形処理(活用処理)という処理が必要となることが理解されるであろう。また、これらの処理は、文法的、語彙的な知識と文脈に依存して進められる。アイヌ語文の単語

直接翻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。どのような変換処理が必要となるであろうか。

以下の記述で品詞の名前を次のように略記する。
 [代名詞：代名，固有名詞：固名，位置名詞：位名，他動詞：他動，自動詞：自動，格助詞：格助，終助詞：終助，副助詞：副助，接続助詞：接助，助動詞：助動，接続詞：接続，連体詞：連体]

4. 単語直接翻訳から日本語文への基本的な変換処理

3章で考察した変換処理も含めて，現時点で，単語直接翻訳を自然な日本語文へ変換するための基本的な変換処理を次のようにまとめている。

- [1] 付加処理：格助詞付加，の付加
- [2] 削除処理：人称接辞φ化，位置名詞φ化
- [3] 変形処理：語形処理，の格化
- [4] 連結処理：所属形連結
- [5] 並べ替え処理：語順変化

格助詞付加，人称接辞φ化，語形処理については前章で例を示した。他の処理について例を示す。

○の格化，の付加

<2>ku=kor totto noya ham uk.

私=持つ 母 よもぎ 葉 採る。

代名=他動 名詞 名詞 名詞 他動

<格助詞付加>

私が=持つ 母 よもぎ 葉 採る。

<の格化>

私の 母 よもぎ 葉 採る。

<格助詞付加>

私の 母が よもぎ 葉 採る。

<の付加>

私の 母が よもぎの葉 採る。

<格助詞付加>

私の 母が よもぎの葉を 採る。

<語形処理>

私の 母が よもぎの葉を 採った。

<の格化>は，「“ku=kor”の直接翻訳」+「人称接辞格助詞付加」の結果である“私が=持つ”を“私の”と変換する処理である。また，<の付加>は，“よもぎ”と“葉”の名詞並びを“よもぎの葉”と変換する処理である。“ku=kor”がつねに“私の”に変換されるわけではなく，また，名詞並び“A B”がつねに“AのB”に変換されるわけではないところに問題がある。

○所属形連結

<3>ku=tekehe piro hi ta ,
 私=の手 傷つく ときに ，
 代名=名詞[所属形] 自動詞 名詞 格助詞

<所属形連結>

私の手 傷つく ときに ，

<格助詞付加>

私の手が 傷つく ときに ，

<所属形連結>は，人称接辞と名詞[所属形]，あるいは名詞と名詞[所属形]を連結する処理とする。代名詞や名詞に対する格助詞「の」の付加とは考えない。アイヌ語における普通名詞の多くは概念形と所属形の2つの形を持つ。所属形は対象をある特定のものに密接に所属するものとして表現する。上の例では，“手”に対する概念形は“tek”で，所属形は“teke”または“tekehe”である[3]。

○語順変化

<4>ku=sinki kusu somo k=arpa.

わたし=疲れる から ない わたし=行く。

代名=自動 接助 副詞 代名=自動

<格助詞付加>

わたしが=疲れる から ない わたし=行く。

<人接φ化><語形処理>

疲れる から ない わたし=行く。

[ない保持] <格助詞付加>

疲れる から (ない)わたしが=行く。

<人接φ化>

疲れる から (ない) 行く。

<語順変化><語形処理>

疲れる から 行か ない。

否定の副詞“somo”が，動詞の前におかれていて，自然な日本語への変換では語順の変化が必要となる。

○位置名詞φ化

<5>Nupurpet or un e=arpa ya ?

登別 の所 へ あなた=行くか ?

固名 位名 格助 代名=自動 終助 ?

<位置名詞φ化><格助詞付加>

登別 へ あなたが=行くか ?

<人接φ化><語形処理>

登別 へ 行くか ?

[名詞，位置名詞“or”，格助詞]という並びでは，位置名詞“or”に対応する直接翻訳“の所”をほとんどゼロ化することができるが，例外もある。また，位置名詞“or”には，“の中”という対訳もある。

<6>mosma kur or ta oka sekora

ほかの人 の所 で 暮らす と

連体 名詞 位名 格助 自動 接助

この例で、“の所”をゼロ化して、“人で”とすると日本語としておかしい表現になる。ゼロ化できない。つまり、日本語では、“人”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うことであろう。

以上の処理に加えて、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。多義語の類型をまとめると次のようになるであろう。

・品詞カテゴリが異なる。

kor: 持つ (他動), て (接助)

e: 食べる (他動), はい (間投), 君 (人接)

ku: 私 (人接), 飲む (他動)

・品詞カテゴリが同じで、多義である。

kor: (他動) ①を持つ, を所有する,

②をつける / をかぶる ③ (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 (接助), よ (終助), から (格助)

5. アイヌ語から日本語への漸進的な翻訳処理の可能性

4章での変換処理と多義語処理を漸進的な自然言語解析 [8] の考え方に基づく翻訳処理として実現することの可能性について基礎的な検討を進めている。3章での次の例で考えてみよう。

<1> usey e=ku rusuy ya ?
お湯 あなた=飲む たい か ?
お湯を飲みたいか ?

左から右へ1語ずつ直接翻訳を行いながら漸進的に翻訳を進める過程の概略は次のようになるであろう。

お湯

<格助詞付加保留\$1>

お湯\$1 あなた=

<格助詞付加保留\$2>

お湯\$1 あなた\$2=飲む

<格助詞付加>

お湯を あなたが=飲む

<人接ゆ化><語形変化保留\$3>

お湯を 飲む\$3 たい

<活用処理><語形変化保留\$4>

お湯を 飲み たい\$4 か

<活用処理>

お湯を 飲み たい か ?

曖昧な場合や情報不十分な場合には処理を保留しつつ部分的な構造を作り、変換処理を行いながら漸進

的に翻訳を進めていくおよその枠組みは、この例に則してまとめると次のようになるであろうか。

(1) アイヌ語単語を読んで、対訳辞書を引く。

(2) 日本語単語の品詞に応じて、日本語単語の辞書情報を利用して部分翻訳構造を構成する。

名詞 : 左連接・接続処理。

格助詞の付加を保留する。

動詞 : 結合価パターンに基づいて、次の処理を行う。

・格助詞の付加処理,

・人称接辞のゼロ化処理。

語形変化を保留する。

助動詞 : 動詞への左連接・接続処理。

語形変化を保留する。

終助詞 : 助動詞への左連接・接続処理。

このような漸進的な翻訳処理の枠組みに基づくアイヌ語・日本語機械翻訳システムの試作とアイヌ語・日本語対訳辞書、日本語辞書の構成を進めている。

6. 複合翻訳単位

6.1 複合翻訳単位の分類

アイヌ語と日本語の間の語彙的なそして構造的な違いに依存して、複数の語によって構成される単位間に1対n, m対n, n対1という対応関係が存在する。アイヌ語そして日本語について自然な翻訳結果を得るためにはそれらの対応関係における複数の語を一つの翻訳単位としてまとめて処理する必要がある。以下に例を示す。田村[2,3]からの引用を(T), 中川[4]からの引用を(N)で表す。

<7> yakun somo: もちろんそうだ (T)

接続 副詞

それなら・(ない (否定))

<8> aske uk: ~を招く (N)

名詞 他動

手・~を取る

<7>は慣用句 (T), <8>は連動詞 (N) あるいは連他動詞 (T) と呼ばれている。Moon & Lee [9] は、韓国語から日本語への機械翻訳における複数の語の間の対応関係の表現と処理について考察し、複数の語によって構成されるひとまとまりの単位を、一般化して、MWTU (Multi-Word Translation Units) と呼ぶことを提案している。本考察では、MWTU に対応する日本語訳として複合翻訳単位という複合語をあてることにし、アイヌ語と日本語の間の機械翻訳のための一つの基礎として、アイヌ語における複合翻訳単位について、文献 [2, 3, 4] を参照しながら具体的な例を

探索し基本的な分類を試みる。

(1) 1対n

アイヌ語の語形成法には、大別して合成、重複、派生の3つがある[4]。そのような方法で形成されたアイヌ語の複合語の日本語訳は複数の語が対応する場合が多い。また、そのような語形成によらない語でも対応する日本語訳が複数の語で構成されるものもある。

kinakar : 山菜を採る (N) (草取りする (T))

keyki : ひざの裏 (T)

(2) m対n

アイヌ語には、目的語である名詞と他動詞とが特別な様式で結びついて独自の意味を持つ連他動詞 (T) (連動詞 (N)) と呼ばれる単位がある。この日本語訳が複数の語から構成されるときはm対n対応となり、1語からなるときはn対1対応となる。

ka(su) opasi : ~のところへかけつける (T)

の上 (位置名詞)・に走る

係り結びによる結合や文末詞もある。

iteki...kunine : ...しないように (T)

hawe an : ~という話なのか? (N)

(3) n対1

連他動詞の日本語訳が一語の動詞であるときn対1対応となる。連他動詞の前の要素である名詞は位置名詞、名詞所属形であることが多い。

・連他動詞の構成: 位置名詞+他動詞

ka, si nukar : ~を見守る (N)

の上 (位置名詞)・~を見る

・連他動詞の構成: 名詞[所属形]+他動詞

kuri ante : ~を笑う, 嘲笑する (T)

の影 (名詞所属形)・~をいさせる

・連他動詞の構成: 普通名詞+他動詞

aske uk : ~を招く (N)

use anu : ~を取り出す, ~を脱ぐ (T)

その他の例をいくつか示す。

tu hot : (二つの二十) 四十 (T)

hine un somo : もちろん (T)

6.2 複合翻訳単位の結合の強さ

要素の語順と結合の強さという視点からアイヌ語の複合翻訳単位は次のように分類することができる。

(1) 語順固定・強い結合 (隣接) 関係

語順固定で、要素が隣接して構成される。

hine un somo : もちろん (T)

(2) 語順固定・やや強い結合関係

連他動詞はその構成要素である名詞と他動詞に

人称接辞の付加が行われ得るので、人称接辞を前あるいは間に置いて隣り合う関係となる。

etok(o) oyki : ~の用意をする (T)

en=etoko oyki : 彼が私の用意をする。

etoko k=oyki : 私が彼の用意をする。

これを次のような一般的なパターンとして表現することができる。@0, @Sは変数を表し、その値は人称接辞に制約される。

< @0=etoko @S=oyki >

(3) 語順固定・弱い結合関係

要素が必ずしも隣接しなくてもよい関係である。

...tasi~nek : ...こそ~さ/よ! (強調)

これらについても、一般的な構成パターンは、任意個の要素の構成に対応する変数を*X, *Yとすれば次のように表現することができる。変数*X, *Yへの代入は節または動詞に制約される。

< *X tasi *Y nek >

7. おわりに

アイヌ語の学習をさらに進めながら、変換処理を詳細化していくこと、多義語処理を漸進的な翻訳処理の中に組み込んでいくこと、実用的な電子化辞書の構成を進めることなどが今後の課題である。

本研究の一部は、北海学園大学ハイテク・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。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工学部 切替英雄先生に感謝いたします。

参考文献

- [1] アコロイタク AKORITAK アイヌ語テキスト1, (社) 北海道ウタリ協会, 1994.
- [2] 田村すず子: アイヌ語, in「言語学大辞典セレクション: 日本列島の言語」, 三省堂, 1997.
- [3] 田村すず子: アイヌ語沙流方言辞典, 草風館, 1996.
- [4] 中川裕: アイヌ語千歳方言辞典, 草風館, 1995.
- [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, 片山龍峯: 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] 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.
- [9] Moon K. & Jong-Hyeok Lee : Representing and Recognition Method for Multi-Word Translation Units in Korean-to-Japanese MT System, Proc. of COLING2000, pp.544-550, 2000.