

TR-I-0185

「は」と「が」の処理
- Subject と Object の機械識別

A Machine Discrimination of Subject and Object in
A Japanese - English Translation

幸山 秀雄
Hideo Kohyama

1990.10

概要

日英機械翻訳システムにおいて「は」や「が」を含む文節を Subject と識別するか Object と識別するか重要な問題である。Subject と Object を間違えて翻訳すれば常識はずれな文を作り出すことが多い。本レポートでは「は」と「が」を含む文においてどういう場合に Subject と Object との解釈の多義が起こるかを明らかにし、その解決方法を検討する。次に動作主に対して考察を行い、常識はずれな文をどう翻訳処理するかについて検討する。最後に市販されている日英機械翻訳システムが「は」と「が」を含む文をどのように翻訳しているか試してみた。

ATR Interpreting Telephony Research Laboratories
ATR 自動翻訳電話研究所

1. はじめに	1
2. Subject か Object か	1
2.1 「は」 + 「が」 + 動詞	2
2.2 「は」 + 「が」 + 形容詞/形容動詞	3
2.3 尊敬表現での多義解消	4
2.4 多義の起こる原因	5
3. 機械識別の方法	7
4. 動作主の考察	9
4.1 常識はずれな文の処理	10
5. 市販機械翻訳システムの「は」と「が」の処理	12

1 はじめに

機械翻訳システムにおいて、「は」や「が」を含む文節をSubjectと翻訳するかObjectと翻訳するかは重要な問題である。翻訳ミスを通りにわけるとすると次のようになる。

- 構文的、意味的に整合性のない翻訳文
- 構文的、意味的に整合性はあるが間違っている翻訳文

2番目は機械翻訳システムの利用者が気づきにくいという点で1番目より悪質である。SubjectとObjectとを間違えて翻訳することは2番目にあたることが多い。またこの翻訳結果は常識はずれなものになることが多く、利用者のシステムに対する信用をなくすことになりかねない。

本レポートでは、まず、助詞の「は」や「が」を含む文節を翻訳するときに、どういう場合にSubjectとして訳され、またObjectとして訳されるかを分析整理する。なお本レポートのSubject、Objectという用語は 直訳した英語文上でSubjectかObjectという意味で使う。

2 SubjectかObjectか

Subjectを機械識別するには、その名詞が述語の動作主(または感情主)になれるかで識別できる場合がある。¹またObjectを機械識別するには、その名詞が述語の対象になれるかで識別できる場合がある。しかし2つの「は」や「が」を含む文では、どちらの文節も動作主(感情主)になることができるものがある。このような解釈の多義がおこる場合、SubjectかObjectかを機械識別するのは難しい。

SubjectかObjectかの解釈の多義がおこるのは、「は/が」+「は/が」+動詞/形容動詞/形容詞の構文とそのレパートリである。動詞で問題となるのは「は」が「が」「を」「に」を代行する場合と「が」が「を」を代行する構文であり、「が」が目的格「を」になれる条件が問題を明確にするカギとなる。

～は(→がをに)+～が(→を)+動詞

この構文のレパートリを次に示す。「は」や「が」の文節が省略されたもの、助詞が省略されたもの、文節が入れ替わったものがある。

「は」+「が」+述語
「が」+「は」+述語
「は」+「は」+述語
「が」+「が」+述語
「φ」+「φ」+述語
ZERO+「は」+述語
ZERO+「が」+述語
ZERO+「φ」+述語

¹本レポートで扱う{は/が}+{は/が}+動詞構文において直訳すれば英語動詞構文になるものは「動作主と解釈されること」=Subjectと考えて良い。但し{は/が}+{は/が}+感情形容詞/形容動詞構文では感情形容詞/形容動詞を英語動詞として翻訳できない場合も考えられる。その様な場合においては「Subjectになる」という表現は正しくない。しかし、本レポートでは便宜的に「Subjectになる」と言うことにする。

2.1 「は」 + 「が」 + 動詞

次の(1)の例文は一方の文節が述語の動作主となり、もう一方はそうでない場合である。この構文で一方だけが動作主になる場合は、それをSubjectとしてよい。

(1) 私は彼の話しが理解できない。

私は = Subject, 彼の話しが = Object

また(2)の文のように両方が動作主になれる場合は文脈情報などを使わなければ決定は不可である。

(2) 太郎は花子が理解できない。

(前文の例: 太郎を誰が理解できないのですか? / 太郎は誰を理解できないのですか?)

太郎は = ?, 花子が = ?

可能表現(「れる」や「できる」)や願望表現(「たい」)ではSubjectかObjectかの解釈で多義がおこる場合がある。²可能/願望表現でない場合は「が」の文節がSubjectになり、「は」の文節がObjectになる。³

・ 「は」の文節 = Object、「が」の文節 = Subjectとなるもの

(3) 太郎は花子が殴る。

(4) 太郎は花子が見る。

(5) 太郎は花子が追う。

(6) 太郎は花子が描く。

(7) 太郎は花子が呼ぶ。

(8) 太郎は花子が叱る。

(9) 太郎は花子が送る。

・ SubjectかObjectかで解釈の多義が起こるもの

(10) 太郎は花子が見(れる / たい)。

(解釈: 太郎は花子を見(れる / たい)。 / 太郎を花子が見(れる / たい)。)

(11) 太郎は花子が追(える / いたい)。

(解釈: 太郎は花子を追(える / いたい)。 / 太郎を花子が追(える / いたい)。)

(12) 太郎は花子が描(ける / きたい)。

(解釈: 太郎は花子を描(ける / きたい)。 / 太郎を花子が描(ける / きたい)。)

² 難易表現の「にくい」「がたい」なども多義になる。ex 太郎は花子が理解しがたい。

³ 存在/所有を表す動詞「ある」「いる」は当てはまらない。

- (13) 太郎は花子が呼(べる / びたい)。
 (解釈: 太郎は花子を呼(べる / びたい)。 / 太郎を花子が呼(べる / びたい)。)
- (14) 太郎は花子が叱(れる / りたい)。
 (解釈: 太郎は花子を叱(れる / りたい)。 / 太郎を花子が叱(れる / りたい)。)
- (15) 太郎は花子が送(れる / りたい)。
 (解釈: 太郎は花子を送(れる / りたい)。 / 太郎を花子が送(れる / りたい)。)

・ (10)～(15)は「を」をとる動詞であるが、「に」をとる動詞では可能 / 願望表現であっても「が」の文節が Subject になる。⁴

- (16) 太郎は花子が会(える / いたい)。
 (解釈: 太郎に花子が会(える / いたい)。)
- (17) 太郎は花子が話(せる / したい)。
 (解釈: 太郎に花子が話(せる / したい)。)
- (18) 太郎は花子が約束(できる / したい)。
 (解釈: 太郎に花子が約束(できる / したい)。)

(16)～(18)の構文では「太郎は」を Subject として解釈することはできない。

2.2 「は」 + 「が」 + 形容詞 / 形容動詞

感情形容詞の構文は微妙なものが多い。構文として微妙なのは他動詞的に解釈できるかという点にある。またこの様な表現では小説などを除き第3者は主格を取れないという考えもある。

次の文の「太郎は」は Subject にも Object にも解釈できる。

- (19) 太郎は花子が嫌いだ。
 (20) 太郎は花子がいやだ。
 (21) 太郎は花子が苦手だ。
 (22) 太郎は花子が必要だ。

形容詞 / 形容動詞で微妙な構文として次のものがある。

⁴(15)は「に」をとることも可能である。(解釈: 太郎に花子が(～を)送れる。)(16)～(18)のように動詞の取れる基本文型を制約として使えず解釈が決まらない。(「は」 + 「が」 + 動詞の構文では「が」が代行できる助詞は、可能 / 願望表現の場合に「を」を代行するだけなので、名詞の意味と文型から「は」の代行可能な助詞が決まる)決まらないのは、基本文型が複数あることにより、「は」が「を」と「に」を代行可能となるためである。「を」と解釈すれば(10)～(14)と同じく多義になるが、「に」と解釈すれば(16)～(18)のように解釈が一意に決まる。「送る」は日本語基本動詞用法辞典[7]によれば次の文型を持つ。

【人・組織】 {が / は} (【人・所】から) 【人・所】 {に / へ} 【人・物・信号】 を送る
【人】 {が / は} 【人】 を 【人・所】 {へ / まで} 送る など

- (23) 太郎は花子が憎い。
- (24) 太郎は花子が憎らしい。
- (25) 太郎は花子が欲しい。
- (26) 太郎は花子が不思議だ。
- (27) 太郎は花子がかわいい。
- (28) 太郎は花子がうるさい。
- (29) 太郎は花子が痛ましい。
- (30) 太郎は花子がたのもしい。
- (31) 太郎は花子がかゆい。
- (32) 太郎は花子がまばゆい。
- (33) 太郎は花子が物足りない。
- (34) 太郎は花子がわずらわしい。

表現としては微妙でも通じる文である。(23)～(34)は「太郎は」をSubjectと解釈したい気がするが、Objectの解釈はないとも断言できない。もっともこれらの文は適切な表現か(「と思う」などが省略されていると解釈するなど)という議論もあるのかもしれない。

2.3 尊敬表現での多義解消

SubjectかObjectかの多義性があるもので、尊敬表現をしているものは多義性を解消できる場合がある。[5]

- (35) 太郎先生は花子がお嫌いだ。
- (36) 太郎は花子先生がお嫌いだ。

2.4 多義の起こる原因

「は」 + 「が」 + 動詞構文でどちらの文節も動作主になれる意味素性をもつ場合において、まず、動詞の取れる基本文型が複数ある場合に「は」の代行可能な助詞が複数になる。例えば、「送る」は下の α と β の文型を取れるが、「花子は太郎が送る」という表現では「は」の代行可能な助詞は α の文型では「に」であるのに対し、 β の文型では「を」である。よってこの表現の解釈は「花子に太郎が送る」と「花子を太郎が送る」の2つが存在する。¹

(α) (人 / が)(物 / を)(人・所 / に) 送る

ex) 太郎が登録用紙を花子に送る。

(β) (人 / が)(人 / を)(所 / に・まで) 送る

ex) 太郎が花子を家まで送る。

次に動詞の可能 / 願望 / 難易表現により「が」が「を」を代行可能となる。それにより動詞が取りえる基本文型が複数ある場合にはさらに他義になる。「花子は太郎が送りたい」という表現ではそれぞれの助詞の代行により、次の解釈が可能である。

解釈 1: 花子が太郎を送りたい。

解釈 2: 花子を太郎が送りたい。

解釈 3: 花子に太郎が送りたい。

久野 [2] は目的格「を」が期待される所に「が」が現れる構文を次のように分析している。

- 能力を表す形容詞 / 形容動詞 (上手、苦手、下手、得意、ウマイ)
- 内部感情を表す形容詞 / 形容動詞 (好き、嫌い、欲しい、コワイ)
- 動詞 + タイ
- 可能を表す動詞 (デキル、レル / ラレル)
- 自意思による感覚動詞 (解る、聞こえる、見える)
- 所有、必要を表す動詞 (アル、要る)

「は」 + 「が」 + 形容詞 / 形容動詞構文ではその対象をあらわすために「が」が使われ、対象と感情主との間で対立が起こることにより多義が生じる。[1]

「は」 + 「が」 + 動詞 / 形容詞 / 形容動詞の構文では次のことが言える。²但し「は」と「が」の文節は共に動作主 (感情主) になれる文節に限る。

(A) 「を」格を持ち可能 / 願望 / 難易表現である構文はどちらの文節も Subject となり得る。

(B) 感情などを表す形容詞 / 形容動詞は「は」と「が」のどちらの文節も Subject となり得る。

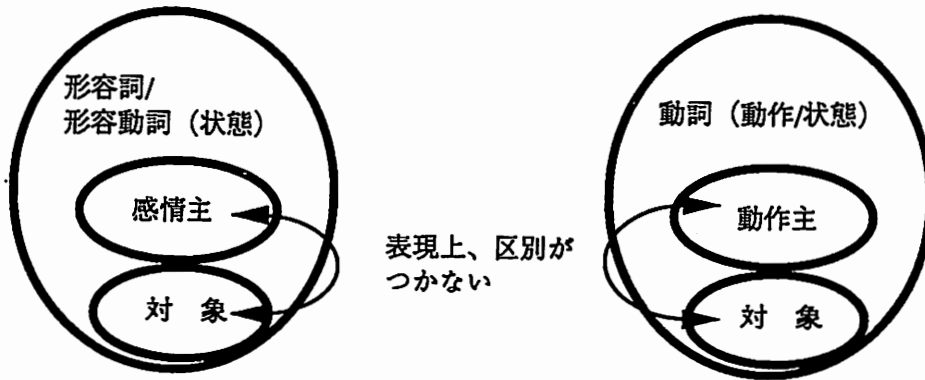
(C) 「を」格を持ち可能 / 願望 / 難易表現でない構文は「が」の文節が Subject である。

¹ 「は」 + 「が」 + 動詞構文において、可能 / 願望 / 難易表現を除き、「が」が他の助詞を代行できることはない。例えば「花子は太郎が送る」という表現で「花子は太郎に送る」または「花子は太郎を送る」とは解釈できない。

² 「花子は妹がある。」の文の様に存在 / 所有を表す動詞「ある」「いる」には当てはまらない。

(D) 「を」格を持たない動詞の構文は可能/願望/難易表現であっても「が」の文節がSubjectである。

「は」 + 「は」 + 述語構文や「が」 + 「が」 + 述語構文では、どちらの文節が助詞の代行を行なうのか識別できないため、(C)(D)の制約を使えなくなることには注意する必要がある。



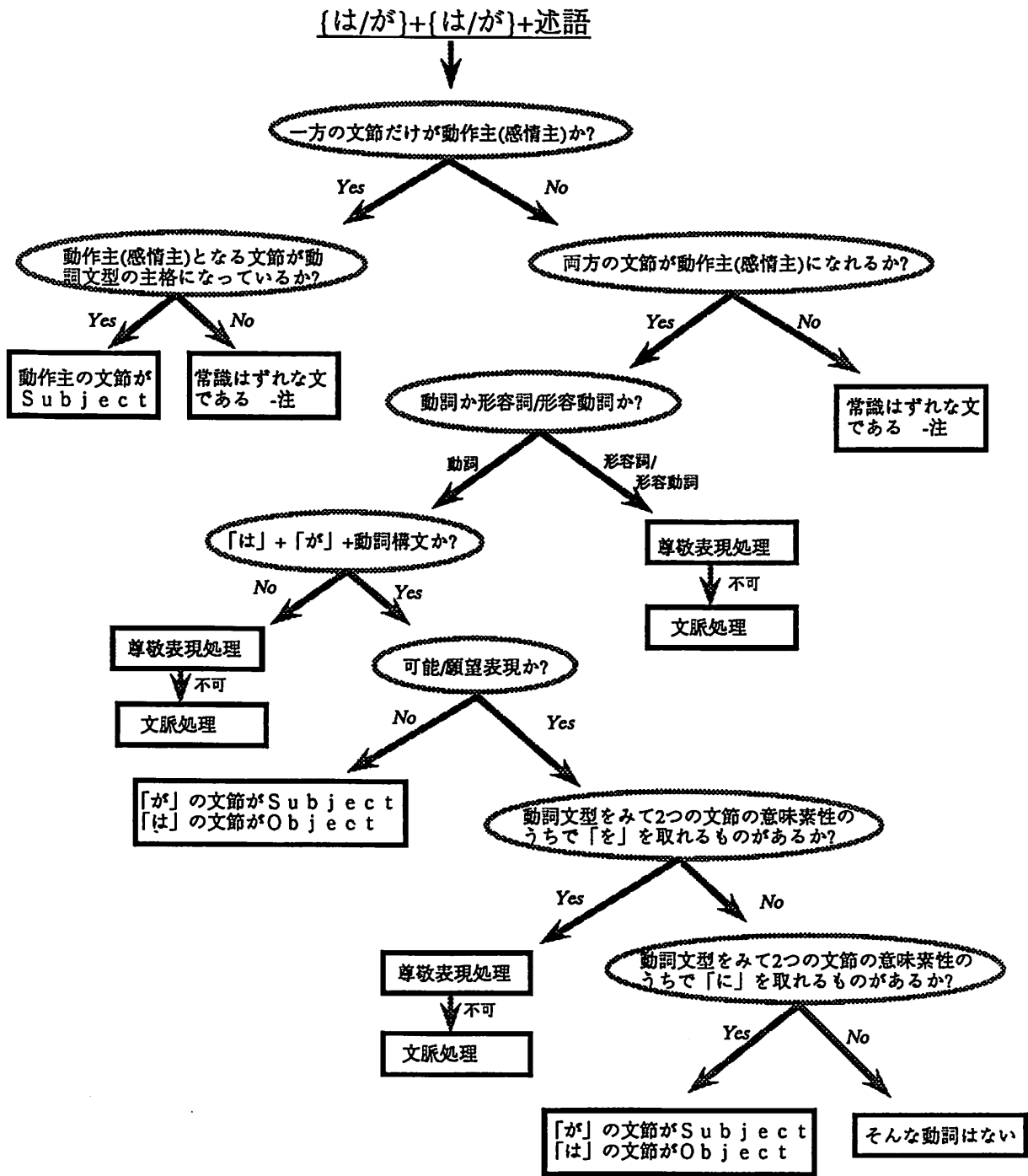
3 機械識別の方法

SubjectかObjectかを機械識別するには、まず解釈の多義の起こる構文であるかチェックしなければならない。そしてSubjectかObjectかを機械識別できるのは次の場合に限られる。

- 動作主(感情主)をチェックすることにより判断できる場合
- 構文から判断できる場合
- 尊敬表現により判断できる場合
- 文脈により判断できる場合

上記以外は完全な解釈は不可である。

SubjectかObjectかを機械識別する処理手順を次ページに示す。



-注 後述する動作主の考察を参照

{は/が}+{は/が}+述語構文における
Subject と Object の機械識別処理の流れ図

4 動作主の考察

SubjectかObjectかを決定するのに動作主という概念は非常に強力である。そこで動作主の概念とは何だろうかと疑問が生じる。言語学辞典での動作主の定義は次のようになっている。

- ～動詞によって表される行為を引き起こす者～
- 動詞の表す動作や出来事の変化を直接的に引き起こす意味上の役割を担う名詞(相当語句)のこと

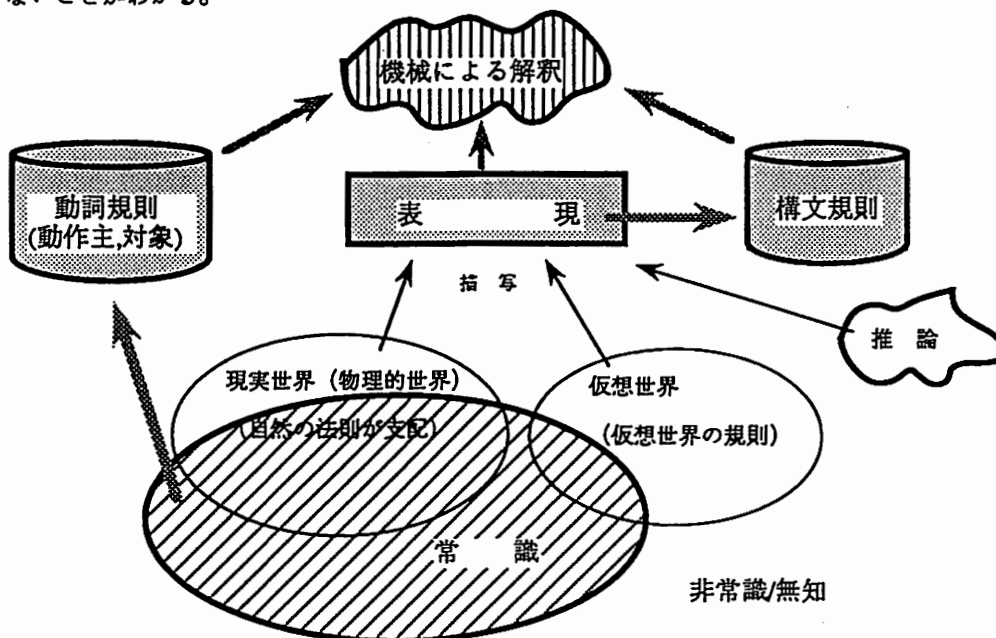
これら辞典の定義は動作主だから、「～動詞によって表される行為を引き起こす者～」であり「動詞の表す動作や出来事の変化を直接的に引き起こす意味上の役割を担う名詞(相当語句)のこと」ぐらいにしか筆者には解釈できない。例えば次の文を考える。

- (a) この本は翻訳する。
- (a') この本は翻訳できない。
- (b) ペンギンは飛ぶ。
- (b') ペンギンは飛べない。

「この本」は「翻訳する」の動作主なのだろうか、又、「ペンギン」は「飛ぶ」の動作主なのかという疑問には言語学辞典は答えてくれない。そのため動作主の筆者なり解釈を以下に示す。

言葉は現実世界や仮想世界を表現という規則性のある記号で描写したものと考える。動作主という概念は現実世界や仮想世界内に存在する規則である。このことは最初に動作主があって表現が決まることを示し、表現の違いにより動作主が変わることはない。

この解釈から現実世界では、「本」が「翻訳する」の動作主になれないし、ペンギンは飛べないことから「ペンギン」は「飛ぶ」の現実世界の動作主にはならない。また動詞に否定がつくとどのような動作主もとれると解釈できるが、この解釈を認めてしまうと動作主の概念自体が無意味なものになってしまうので認めない。(例えば「ペンギンは飛べないことはない」を認めることになる。)このような動作主は便宜的に表現上の動作主と呼ぶことにする。また「カラスが宇宙を飛ぶ」のような文を考えると現実世界の動作主は動詞の取る他のアーギュメントも考慮しなければ判断できないことがわかる。



4.1 常識はずれな文の処理

動詞の取り得る動作主を記述することは基本的な常識を記述することともとらえられる。しかしこの常識をはずれた表現をどう翻訳処理するかが問題となる。(例えば「不思議の国のアリス」を翻訳処理するためにそれ専用の辞書を作るのはナンセンスである。)

前述したaの文は「本」は「翻訳する」の動作主になれないのでObjectとして解釈される。しかしbの文は「ペンギン」は飛べないので、「飛ぶ」の現実世界の動作主になれないがSubjectとして解釈される。これは動詞が持てる文型(動詞が持てるアーギュメント)の違いにより説明される。例えば「飛ぶ」と「翻訳する」は次のアーギュメントを取る。

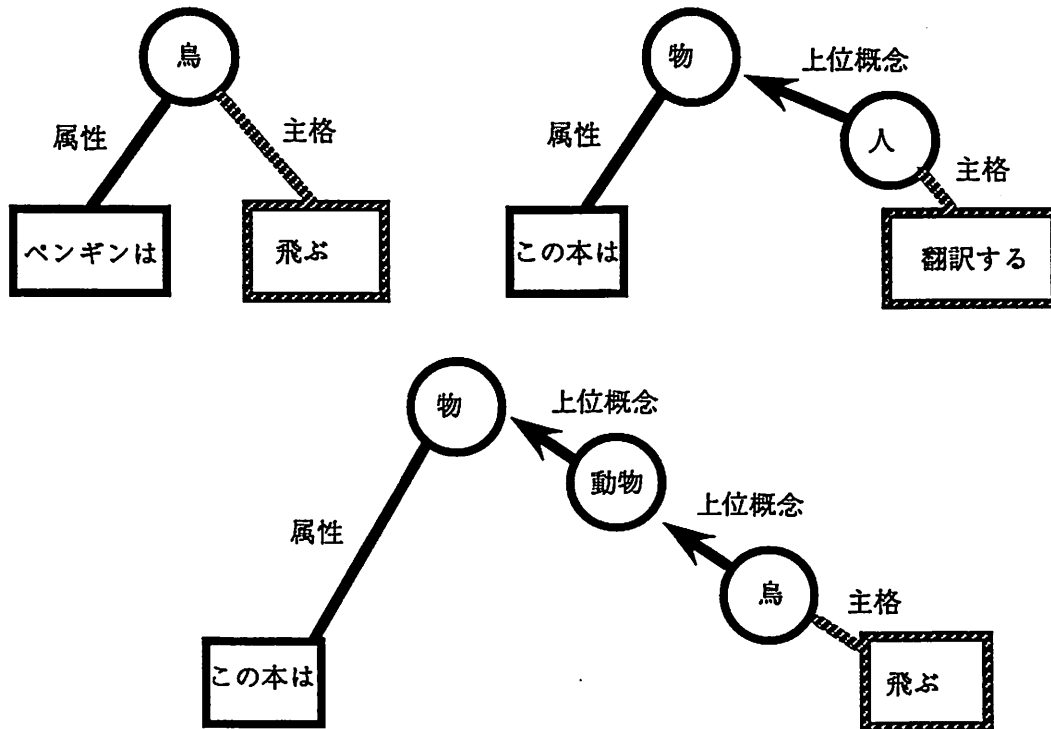
(翻訳するもの名詞群 / が) (翻訳されるもの名詞群 / を) (言語 / に) 翻訳する

(飛べるもの名詞群 / が) (飛べる場所の名詞群 / を) 飛ぶ

「本」は翻訳するもの名詞群には該当しないが翻訳されるもの名詞群に該当する。一方、「ペンギン」は飛べるもの名詞群にも飛べる場所の名詞群にも該当しない。「ペンギンは飛ぶ」は命題内容としてはおかしいが表現として正しい。この問題を解決するために、ペンギンは鳥であり動物であり生物であり物であるという概念(属性)とその上位概念を導入する。「鳥」は飛べるもの名詞群の属性となるため、「ペンギン」は「飛ぶ」の表現上の動作主になれる。他には「飛ぶ」の上位概念として「する」を選び、「ペンギン→動物」は「する」の動作主になれるという方法もある。こういう方法は無茶な感じもするが、生き物である「ペンギン」を飛べる場所として解釈しにくい以上妥当な方法であると考えられる。この上位概念を使った類推により、aの文の「本」もSubjectとして解釈できるようになる。「本」が翻訳するような仮想世界の表現があっても不思議ではない。しかし常識と非常識とで多義の解釈が生まれる場合は常識を優先させるべきである。上位概念をより多くたどった方がより非現実な解釈となり、できるだけ低い上位概念を優先させるべきである。処理方法を整理すると次のようになる。(否定文もこの方法で充分である。)

1 名詞が動詞文型のアーギュメントにあてはまるならOK

2 名詞の概念を順次シソーラスから得て、動詞のアーギュメントの名詞群のなかの属性、又はその上位概念と一致するならOK



参考までに「飛ぶ」と「翻訳する」の日本語基本動詞用法辞典 [7] での記述を示す。

[人・生き物・物] {が/は} ([所] を) ([所] から) ([所] {に/へ}) 飛ぶ

[人・組織] {が/は} [書類・言語作品・話し] を ([言語] から) [言語] に 翻訳する

この概念の定義では厳密な動作主とその対象は判断できない。しかしだいたいのチェックは可能と考えられる。

5 市販機械翻訳システムの「は」と「が」の処理

現在、市販されている機械翻訳システムの1つで、「は」と「が」を含む構文をどう翻訳するか試してみた。それを以下に示す。

- (37) 花子は嫌いだ。
Hanako is disgusting.
- (38) 花子が嫌いだ。
Hanako is disgusting.
- (39) 太郎は花子が嫌いだ。
Tarou dislikes Hanako.
- (40) 花子が太郎は嫌いだ。
Tarou dislikes Hanako.
- (41) 私は彼の話しが理解できない。
I can not speak and understand him.
- (42) 太郎は花子が理解できない。
Tarou can not understand Hanako.
- (43) 花子が太郎は理解できない。
Hanako can not understand Taro.
- (44) 太郎は花子が見た。
Hanako saw it about Tarou.
- (45) 花子が太郎は見た。
Tarou saw in term of Hanako.
- (46) 太郎は花子が見る。
Hanako sees it about Tarou.
- (47) 太郎は花子が見れる。
If Hanako sees it about Taro.
- (48) 太郎は花子が苦手だ。
Hanako is a hard antagonist about Tarou.
- (49) 太郎は花子がかawaiiそうだ。
Hanako is pitiable about Tarou.
- (50) 太郎は花子がかawaiiそうに思っている。
Tarou thinks that Hanako is pitiable.
- (51) 太郎は花子が不思議だ。
Hanako is strange about Tarou.
- (52) 太郎は花子が不思議に思っている。
Tarou thinks that Hanako is strange.
- (53) 太郎は花子が憎らしい。
Hanako is hateful about Tarou.

- (54) 太郎は花子が憎い。
Tarou hates Hanako.
- (55) 太郎は花子が憎く思っている。
Hanako hatefully thinks about Tarou.
- (56) 太郎は煙草を吸う。
Tarou smokes a cigarette.
- (57) 太郎は煙草が吸える。
A cigarette can absorb about Tarou.
- (58) 太郎は煙草は吸っている。
Tarou is smoking a cigarette.
- (59) 太郎は煙草は吸える。
Tarou can smoke a cigarette.
- (60) 太郎が煙草が吸える。
A cigarette can absorb in term of Tarou.
- (61) 太郎が煙草は吸える。
Tarou can smoke a cigarette.

前処理することを前提として作られたシステムなのでこのような試行は酷ではあるが、次のことが言える。致命的なのは動作主を識別していないことである。また「が」＋「は」＋述語、「は」＋「は」＋述語、「が」＋「が」＋述語の構文を処理していない。

謝辞

本レポートについて意見を述べてくれた樽松社長、森元室長、飯田主幹研究員に感謝します。またレギュラトークで問題を指摘してくれた諸氏に感謝します。

[参考文献]

- [1] 寺村 日本語のシンタクスと意味 I くろしお出版
- [2] 久野 日本文法研究 大修館書店
- [3] 三上 象は鼻が長い くろしお出版
- [4] 三上 日本語の論理 くろしお出版
- [5] 柴谷 日本語の分析 大修館書店
- [6] 益岡 田窪 基礎日本語文法 くろしお出版
- [7] 小泉ほか 日本語基本動詞用法辞典 大修館書店