

機械翻訳における日本語の名詞並列
に関する一考察

5B-8

押金章悟¹, 菊池浩三², 川久保賢一³, 加藤晴嗣⁴
富士通(株) (株)富士通静岡エンジニアリング

1. はじめに

日本語における名詞並列表現は幾通りもの解釈が可能であり、これを機械処理する為に、名詞の意味情報を用いた各種の手法が議論されている(参考文献1, 2, 3等)。この手法をより効果的にするためには、①「一見並列に見えるが、並列としての処理をしてはいけないもの(述語に係る「と」の問題)」、②「パターンによって、ある一定の解釈ができる並列表現の問題」を処理しておかなければならない。本稿では、これら①②についての例と処理の評価結果を、大量のデータ分析に基づいて紹介する。

2. 並列に関わるさまざまな問題と対策

名詞の並列表現には、構文や個々の単語の種類によってはさまざまな問題が絡んでいる。これらの例を以下に示す。

2.1. 必須格の「と」

例1) A社とB契約を締結する

例1の場合「A社と」の係り先は「B契約」ではなく、「締結する」である。これを並列の処理にかけると「B契約」を係り先にしてしまう可能性がある。このように「と」を取る述語がある場合、それより前方の「名詞+と」は「と」を取る述語に係りなければならぬ。しかし、表1に示すように「と」についての深層格は何通りもあり、また、同一の深層格でも述語によって名詞の種類が限定されるので、述語をグルーピングして処理する必要がある。

深層格	述語の例	名詞の条件
相手	締結する 接触する ⋮	国/組織 具体物 ⋮
目標	変化する なる ⋮	状態を表す語 特になし ⋮
基準	比較する 比べる ⋮	特になし " ⋮

表1. 必須格「と」の深層格

2.2. 自由格の「と」

例2) A社がB社と製品を開発する

例2は「B社と」の係り先に関する例文である。必須格の「と」の問題と類似しているが、これは「B社と一緒に」の意味の自由格の「と」である。自由格は他動詞述語であればその種類に関わらず何にでも係る。このパターンの場合、「B社」に当たる名詞は組織か動物(主に人間)である。しかし、図1に示すように後方の述語との間に同じ意味を表す名詞がある場合は係り先はその名詞である確率が高い。

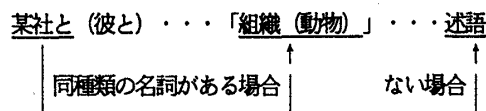


図1. 自由格の「と」の係り先

2.3. 同一表記

例3) エラーを検出するプログラムとメッセージを表示するプログラム

例3で「プログラムと」の係り先は「メッセージ」と2つ目の「プログラム」が考えられるが、当然正解は2つ目の「プログラム」である。一般に、「～N1と…N2～」というパターンの場合で、N1とN2が同一表記の名詞である場合N1とN2が並列であることが多い。ただし「…」の部分には「係り助詞「は」がない/「～場合、」という表現がない」等の条件が必要である。

2.4. 名詞の種類によらないパターン

(1) 指示代名詞

例4) リーダーの教育およびそのフォローを計画したシステムの異常とその対処方法を以下に示す

「AのBとC」というパターンに対しては2通りの解釈が可能であるが、Cに指示代名詞「その」が付いている場合はほとんどの場合「[AのB]とそのC」が正解である。これはA, B, Cの名詞の種類に関わらず言えることである。

(2) 結合の働きのある名詞

例5) AとBの整合性

「AとBのC」のパターンでCの名詞が、「整合性」などの結合の働きのある名詞の場合、A、Bの名詞の種類に関わらず「[AとB]のC」と解釈するのが正解である。このような働きを持つ名詞は、他に以下のものがある。

関係/双方/協力/相互作用/結合/合併会社...etc.

(3) 特定単語に依存する特殊パターン

あるパターン中にある特定の単語が存在したときに他の単語の種類に関わらず一定の解釈ができるものもある。以下にその例を2パターン示す。

(a) 「～と画面の推移(or 流れ)」

例6) 操作方法と画面の推移を図1に示す
起動手順と画面の流れを以下に示す

例6のように「AとBのC」というパターンの場合「BのC」に当たる部分が「画面の推移(or 流れ)」であればAの名詞の種類に関わらず、「Aと[画面の推移(or 流れ)]」という解釈が成立する。

(b) 「～する方法と～の～」

例7) 実験値を解析する方法とそれに関連した理論データを処理する方法とプログラムの位置付け

例7のように「AとBのC」というパターンで、Aに当たる名詞が「方法」であり、かつ、この名詞に連体修飾が付いている場合は、B、Cの名詞の種類に関わらず「～する方法と[BのC]」という解釈が成立する。

3. 評価結果

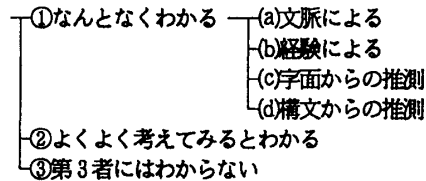
これまで述べてきたようなさまざまな事前処理によって、表2に示す通り、洗いだした問題点の7割強を解決できた。

問題の種類	問題数	解決した数
必須格「と」	150	120
自由格「と」	50	30
他の各種パターン	120	80

表2. 解決した名詞並列の問題点の数

4. おわりに

処理できない問題の中には、書き方が適当でないものもある。これはある程度書き手の責任とも言えるだろうが、あまり強く制限すると現実的でなくなる恐れもある。こういう問題は別にして、今後の課題を「人間が読んだ場合に」という前提で分類すると以下ようになるだろう。



②、③は人間でもよくわからない類の問題なので、機械処理できなくても仕方がない。今後は、①の部分を強化していきたい。しかし、(a)は辞書の機構、プロセッサの機構とも次期検討課題であろう。

(b)、(c)、(d)は、「2.4.名詞の種類によらないパターン」で述べた事柄に該当するが、切り口をいかに上手に発見できるかという問題である。現在の機構でもルール化は充分可能である。エキスパートの知識抽出というAI処理と密接に関わってくると考える。

【参考文献】

- 1) 黒橋 禎夫, 長尾 眞: 長い日本語文における並列構造の推定, 情報処理学会自然言語処理研究会報告, 86-2, (1991)
- 2) 黒橋 禎夫, 長尾 眞: 並列構造の検出に基づく長い日本語文の構文解析, 情報処理学会自然言語処理研究会報告, 88-1, (1992)
- 3) 首藤 公昭, 吉村 賢治, 津田 健蔵: 日本語技術文における並列構造, 情報処理学会論文誌, Vol.27, No. 2, pp. 183-190, (1986)