

「韻」に基づく並列句の解析

6C-2

森口 修 鈴木克志 丸山冬樹
三菱電機㈱ 情報電子研究所

1. はじめに

自然言語処理が困難であるという原因はその曖昧性による
ところが大きであるが、並列(等位)接続はその中でも特に曖
昧性が強い。そのため、これまでも多くの試みがなされて
きた⁽¹⁾⁽²⁾⁽³⁾⁽⁴⁾⁽⁵⁾。これらの曖昧性を解消するためには、
そのすべての可能性を評価し、最適と思われる解を導出する
ことも一つの方法である。

一方、機械翻訳においては、翻訳結果に対して一意性と即
答性が要求される。すなわち、ただ1つの最適な結果を短い
時間で出力する必要がある。

本論文では、そのために有効であると思われる「韻」の情
報について触れ、それを積極的に認識し、並列スコープの決定に
役立てるという方法について述べる。

2. 韻を踏んだ並列接続

並列接続の中でも、取り扱いが困難とされているものに、
非名詞句⁽²⁾の並列がある。そこで、技術論文⁽⁶⁾約1000文か
らそれらを抽出したところ、表1のような結果が得られた。

表1. 非名詞句の並列パターン

分類	パターン	数
数量 表現	NがVしてQ年、NがVしてQ年	1
	NQ台、NQ台(、NQ台、NQ台)	3
	Q個のNをQ、Q個のNをQ	1
	N用でQmm、N用ではQmm	1
	NQ階、NQ階、NQm ² 、NQm	2
格 助詞	NとしてNを、NとしてNを	1
	VしているNを、またVしているNを	1
接続 助詞	NをNに、NをNに	3
	NはNに、NはNに	1
	どのNに、つまりどのNに	1
	N時はQに、N時はQと(なる)	1
	Nの場合は、Nとして、Nの場合はNとして	1
	NがVした場合、あるいはNがVした場合	1
	NをVするため、また、NをVするため	1

N:名詞, V:動詞, Q:数詞

これを見て解るとおり、非名詞句の並列スコープはほとんどが
韻を踏んでいるとみなせる。逆にいえば、これまで取り扱
いが困難とされていた非名詞句の並列表現の解析には、「韻」
を手がかりとすることが有効であるといえる。また、書き手
が意識的か無意識的にかかわらず、人間である読み手は韻
が踏まれた文に対して敏感である。したがって、機械翻訳に
おいて、韻の情報を並列スコープ決定の手がかりとすることは正
当であるといえる。なお、「韻を踏む」とは「ひびきをあわ
せる」ことであるが、ここでは「パターンが類似した、文節の1
次元配列の並列」とする。もちろんパターンの中には音(読み)も
属する。

3. 並列スコープ解析フェーズの位置

結論を先にいえば、並列スコープ解析フェーズは図1のように、
形態素解析と構文解析の間に位置するのが最適であると考え
る。その理由は以下のとおりである。

もし、形態素解析前に行なうとした場合、品詞や付属語の
情報が得られない。表1を見てわかるように、品詞や付属
語の情報は「韻」を解析する上で大きな手がかりとなる。

もし、構文解析後に行なうとした場合、構文解析時は並列
スコープが決定されていない上に、(1)のようなDCGが必要となり、
構文の曖昧性や解析時間を増加させてしまう。ここで、coo
は並列結合子⁽⁵⁾である。

(1) X ← X coo X.

これらの理由で、並列スコープ解析フェーズは図1のように形態
素解析と構文解析の間に配置し、名詞句または文形式以外の
並列を吸収することとする。

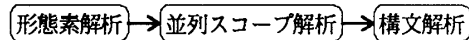


図1. 並列スコープ解析フェーズの位置

4. 韻解析の仕様

4.1 入出力仕様

並列スコープ解析の一部である韻解析フェーズの入力は文節ADSの
リストであり、出力は文節ADSのネィティングリストである。すなわちリス
トのネィティングによって並列スコープおよび並列要素を指定する。こ
こで、ADS(A dependency structure)とは、日英機械翻訳システム
AMELTRAN-J/Eにおける自然言語構造の内部表現形式であり、
文節ADSとは、各文節ごとに作られるADSである。韻を踏んだ
原文に対する韻解析前半部の入出力の例を図2に示す。この
例では、並列スコープは「構成を図3に、特徴を表1に」であり、
並列要素は「構成を図3に、」と「特徴を表1に」の2つで
ある。

入力例: 「システム構成を図3に、特徴を表1に示す。」

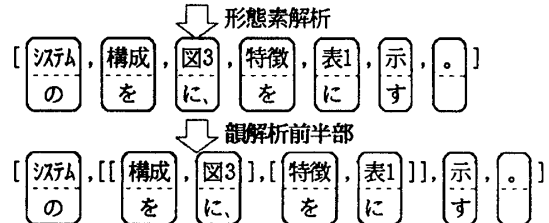


図2. 韻解析前半部の入出力仕様

4.2 韻解析前半部

このフェーズは文節の配列の特徴から並列スコープを決定する。アルゴリズムの概要は以下のとおりである。

入力例: 前方←「~ABCDcooA'C'D'~」→後方

- ①並列結合子cooを見つける。
- ②cooの直前の文節Dの特徴を見いだす。
- ③cooの後方でDと似た特徴を持つ文節D'を見つける。
- ④cooの直後のA'からD'までの配列パターンP'を見いだす。
- ⑤Dより前方かつDまでの配列パターンPと、P'との類似度を最大とするような文節Aを見つめる。
- ⑥並列スコープを[[ABCDcoo],[A'C'D']]とする。

実際には、並列スコープの終端を指示する「など」や、束ねの表現⁽²⁾「間」、「合計」も考慮して並列スコープを決定する。

文節の特徴とは、[1]読み、[2]付属語、[3]接辞語、[4]見出し、[5]品詞、[6]文節タイプ、[7]語彙分類などの、その文節の属性である。

文節の配列パターンとは、{1}文節の特徴の類似性、{2}文節の個数、{3}文節の順番の他、{4}よく現れるテーブル化された表現か否か、{5}並列スコープを決定する特殊な文節か否かであり、文節パターンの類似度とは、これらをパラメータとする評価関数の値である。

4.3 韻解析後半部

ここでは省略された文節を補う処理を行なう。これは構文解析フェーズなどでも可能であるが、韻の情報も省略された文節を補うための大きな手がかりとなるため、このフェーズで並列要素に省略された文節を補う。

例文(2)のような非名詞句をどのような構造で表現するかは、図3のようにいくつか考えられる。

しかし、図3の(a)では1つの述語に同一の複数の格に係ることになり、格解析フェーズでの格のチェックが甘くなってしまう。

また、表1をみてもわかるように非名詞句の並列要素は2つ以上の文節からなることがほとんどであり、この場合(b)の構造では不具合が生じる。

一方、非名詞句の並列では並列する要素のいくつか、または、すべてに文節の省略があり、これを補った(c)の構造が後のフェーズでは最も扱い易い。また、思い切って文を分割することも考えられるが、束ねの表現を厳密に認識しなければならず、また日本語1文を英語1文に翻訳するという原則にも反するので、そこまでは行なわないこととした。

- (2) 「Aを、またはBをVする」

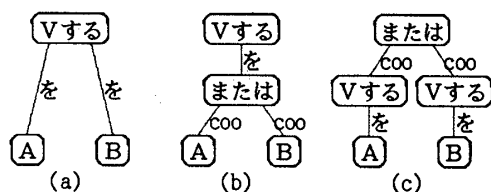


図3. 非名詞句の並列構造

省略された文節を補うための基本的な条件は、

- (1) それを補えば、パターンの類似度が大きくなる。
- (2) それを補わなければ、非名詞句の並列表現となる。

であり、具体的には

- <1> 他の文節配列パターンと比較すると不足した文節がある。
- <2> 補う文節が並列スコープの端に位置していない。
- <3> 補う文節は近くの並列要素から近接される。
- <4> 格助詞や副詞に係るべき述語がその並列要素内にない。
- <5> 並列要素数が多い場合、最後の並列要素のパターンは重視しない。特に「に加えて」や「の他に」などの後に続く並列要素の類似度は採点を甘くする。
- <6> 並列要素の末尾に述部を補う場合、並列結合子がかつまたは読点のみの場合は連用形とし、そうでない場合はそのまま π -する。

からなるルールを用いた処理となる。

<1>~<5>によって文節配列間の類似度を評価する。

<6>はいわゆる連用中止法による重文を作る。

<2>で両端の文節を補われにくくしたのは、両端の文節は近くに係るか全体に係るかという可能性の他に、並列スコープの外に係るという可能性もあるためである。

<4>は同一の述部に同じ格に係るのを防ぐ。

<5>は3つ以上の要素が並列する場合であり、書き手がパターンパターンの類似性を維持できなかった並列要素はしばしば最後に配置されるという傾向があるためである。

5. おわりに

本論文では、「韻」、すなわち、その表層上のパターンの類似性を用いた並列スコープの解析方法について述べた。この手法は特に非名詞句の並列表現の解析に有効であると考えられる。これらの手法は日英機械翻訳システムELTRAN-J/Eに組み込む予定である。

今後の課題は、「韻」の弱い並列表現の解析である。例えば「~は~に、~は~となる」のような文においては、動詞「なる」が「に」格も「と」格も持っているという「格」の情報が必要となる。このように格解析などの意味的な処理を組み合わせることによって、解析能力の向上を図ることが必要である。

参考文献

- (1)長尾,辻井,田中,石川:科学技術論文における並列句とその解析,自然言語処理研究会,36-4,1983.
- (2)首藤,吉村,津田: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.2,1986.
- (3)田村,田中:意味解析に基づく並列名詞句の構造解析,自然言語処理研究会,59-2,1987.
- (4)押金章悟:パターンを用いた日本語の名詞並列句解析,情報処理学会第36回全国大会予稿集,3T-2,1989.
- (5)根本,藤田:名詞句並列表現の解析,情報処理学会第36回全国大会予稿集,3T-3,1989.
- (6)三菱電機技報Vol.64, No.10「昇降機特集」,1990.