

機械翻訳向け前編集のための日本語係り受け構造の曖昧性検出方式[†]

平井章博^{††} 梶博行^{††} 芦沢実^{††}

機械翻訳用の前編集の作業効率を高めるために、原文中に存在する形態素、係り受け構造等の曖昧性を検出する機能が求められている。必要以上の多量の曖昧性の検出は、かえって前編集の作業効率に悪影響を及ぼすため、曖昧性検出においては、検出すべき曖昧性の検出の漏れを抑えながら、不要な曖昧性検出（前編集しなくても機械翻訳システムが正しく解釈する箇所に関する曖昧性検出）を少なくすることが重要な課題となる。本論文では、日本語の曖昧性検出のうちで重要な係り受け構造の曖昧性検出について、この課題の解決策として、①機械翻訳システムと同一の解析を行って機械翻訳システムの採用する係り受け構造の最優先解を求め、これに対して係り受けの非交差条件と格の非重複条件を満たす範囲に別解の検出を抑える、②係り受けの飛び越しを抑制する傾向のある読点や語句を利用して、不要な別解の検出原因となる範囲を別解探索範囲から除く、③語句ごとの固有の構文的性質を利用して不要な別解の検出を制限する、の3点を特徴とした曖昧性検出方式を提案している。そして、本方式に基づいて、係り受け構造の曖昧性検出システムを開発し、特許抄録を対象として評価している。この結果、本方式は、必要な曖昧性検出の漏れを実用上問題ない量に抑えながら、不要な曖昧性検出を、可能な構文構造をすべて検出する従来方式の1/2未満に削減でき、上記課題を解決できることが確認できた。

1. ま え が き

広範囲の文章に対して、機械翻訳システムの精度を高めるためには、原文を機械翻訳システムに適するように修正しなければならない。このような原文の修正を前編集という。この前編集の作業効率を高めるために、原文中に存在する形態素、係り受け構造等の曖昧性を検出する機能が求められている²⁾。ここでいう曖昧性とは、機械翻訳システムにとって、複数の解が存在することを意味する。ところが、必要以上の多量の曖昧性の検出は、かえって前編集の作業効率に悪影響を及ぼす。したがって、曖昧性検出においては、前編集しないと機械翻訳システムが誤って解釈する箇所に関する曖昧性（以下、これを検出すべき曖昧性と呼ぶ）の検出漏れを抑えながら、前編集しなくても機械翻訳システムが正しく解釈する箇所に関する曖昧性検出（以下、これを不要な曖昧性検出と呼ぶ）を少なくすることが重要な課題となる。

従来、前編集支援に関連する、曖昧性を検出するシステムとして、富田のインタエディット・システム³⁾、長尾らの制限文法に基づく文章の作成支援システム⁴⁾、有田らの前編集支援システム¹⁾などが開発されている。しかしながら、富田のシステムは、可能な

構文構造をすべて検出する方式の採用を前提として、検出した構文構造を効率良く扱うことを目的としたものであるし、長尾らのシステムは、文章作成者の意図を全面的に確認することを前提として、高速な応答の実現を目的としたものであるため、これらのシステムでは、不要な曖昧性検出を抑制する問題は扱われていなかった。また、有田らのシステムでは、ヒューリスティックな条件を適用して、曖昧性検出を抑制しているが、その条件の詳細と評価は公表されていない。

そこで、本論文では、日本語の曖昧性検出のうちで重要な係り受け構造の曖昧性検出について、この課題の解決策として、①機械翻訳システムと同一の解析を行って機械翻訳システムの採用する係り受け構造の最優先解を求め、これに対して係り受けの非交差条件と格の非重複条件を満たす範囲に別解の検出を抑える、②係り受けの飛び越しを抑制する傾向のある読点や語句を利用して、不要な別解の検出原因となる範囲を別解探索範囲から除く、③語句ごとの固有の構文的性質を利用して不要な別解の検出を制限する、の3点を特徴とした曖昧性検出方式を提案している。そして、本方式に基づいて、係り受け構造の曖昧性検出システムを開発し、特許抄録を対象として評価している。

以下、本論文では、第2章において、最優先解情報を活用した係り受け構造の曖昧性検出の精度向上方式、第3章において、この方式に基づいた、係り受け構造の曖昧性検出システム、第4章において、この方式の評価について述べる。

[†] Ambiguity Detection in Japanese Dependency Structures for Pre-Editing by AKIHIRO HIRAI, HIROYUKI KAJI and MINORU ASHIZAWA (Systems Development Laboratory, Hitachi, Ltd.).

^{††} (株)日立製作所システム開発研究所

2. 日本語の係り受け構造の曖昧性の検出方式

2.1 日本語の係り受け構造とその曖昧性

(1) 日本語の係り受け構造

日本語は、語順の自由度が高いため、文節間の修飾関係に注目した係り受け構造が、日本語の構文をとらえるのに適している。したがって、本論文においては、係り受け構造で日本語の構文を表現する。この係り受け構造は、次の規則¹⁰⁾を満たすものである。

(a) 係りの後方性と唯一性：文末以外の文節は後方の一つの文節に係る。

(b) 係り受けの非交差性：係る側の文節から受ける側の文節に矢印を引いたとき、それらの矢印が交差することはない。

(c) 格の非重複性：一つの文節が同一の格の文節を複数受けることはない。

また、受ける側の文節が用言である係り受けを連用の係り受けと呼び、受ける側の文節が体言である係り受けを連体の係り受けと呼ぶこととする。

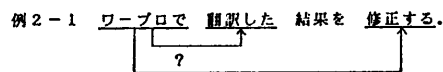
(2) 係り受け構造の曖昧性

機械翻訳システムにとって、一つの文節の係り先の解として、二つ以上の文節が存在する場合、係り受け構造の曖昧性があると呼ぶ。次に、図1を参照しながら、係り受け構造の曖昧性の例を説明する。

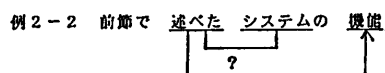
(a) 係り先が用言の場合：図1の例2-1では、「ワープロで」の係り先の解として、「翻訳した」と「修正する」の二つの用言が存在し、係り受け構造の曖昧性がある。

(b) 係り先が体言の場合：図1の例2-2では、「述べた」の係り先の解として、「システム」と「機値」の二つの体言が存在し、係り受け構造の曖昧性がある。

このような係り受け構造の曖昧性がある場合、機械翻訳システムの採用する第1解が誤る可能性が高くな



(1) 連用の係り受けの曖昧性



(2) 連体の係り受けの曖昧性

図1 係り受け構造の曖昧性の例

Fig. 1 Examples of ambiguities in dependency structures.

り、前編集の必要性も高くなる。このため、前編集必要箇所の発見を容易にするものとして、このような曖昧性を検出する機能が強く求められている。

2.2 係り受け構造の曖昧性検出方式

(1) 本方式の基本的な考え方

最優先解が正しいか否かは機械的に判断できない場合が多いため、一般に、不要な曖昧性検出を抑制しようとするとき、検出すべき曖昧性も検出されなくなってしまう可能性が生じる。そこで、不要な曖昧性検出を少なくすることと、検出すべき曖昧性の検出漏れを実用上問題ない程度に抑えることの両方を満たすための工夫が必要になる。本方式は、その工夫として、翻訳に利用する機械翻訳システムの最優先解の情報を活用して、最優先解に対して一定以内の近さにある構文構造の別解だけを検出するようにし、検出精度を向上させようとするものである。こうすれば、抑制される検出は最優先解から遠いものになるため、最優先解の正しい確率がある程度高ければ、検出漏れは少ないはずであるし、しかも、検出範囲を小さくしているので、不要な検出も少なくなる。機械翻訳システムとして、高精度なシステムを利用する場合は、本方式は非常に有効となる。

(2) 本方式の詳細

本方式では、係り受け構造の最優先解を決定した後に、係り受け構造の別解を検出し、個々の文節の別解の有無により、その文節の係り先の曖昧性の有無を判断する。ここでいう係り受け構造の最優先解とは、機械翻訳システムが採用する係り受け構造の第1解に対応する。例えば、筆者らが開発している日英機械翻訳システム^{4),5)}では、意味的な制約や統計的に有意な制約による例外を除いては、なるべく近い文節間に係り受けが成立するように最優先解が決められる。

本方式では、以下に述べるような別解検出の制約を設けて、不要な別解の検出を抑制する。

(a) 係り受け構造の最優先解から遠く離れた別解の検出の抑制

本方式では、不要な曖昧性の検出を抑えるために、係り受け構造の最優先解から遠く離れた別解の検出を抑制するというヒューリスティックな制約³⁾を設定する。係り受け構造の最優先解から遠く離れた別解の検出を抑制するために、具体的には、

(i) 係り受け構造の最優先解との非交差性

(ii) 係り受け構造の最優先解との格の非重複性

を満たす範囲に別解の検出を絞る。最優先解の精度が

高ければ、正しい係り受け構造が、最優先解から遠く離れたものである可能性は少ないので、この制約は有効に機能するはずである。なお、例外として、直後に読点がある文節に関しては、(i)の条件は満たさなくても良いものとする。

図2はこの制約による別解検出の抑制の例を示したものである。図2の例2-3において、「ディスクに」と「衛星から」が「送られた」に、「送られた」が「データを」に、「データを」が「書き込む」に係るのが最優先解であるとする。「衛星から」が「書き込む」に係るという別解は、「ディスクに」が「送られた」に係るという最優先解と交差してしまうので、本方式では検出しない。「ディスクに」が「書き込む」に係るという別解は、他の最優先解と交差しないので検出する。なお、図2中の→は最優先解の係り受けを表し、---→は検出対象となる係り受けの別解、---×→は検出対象とならない係り受けの別解を表す。表1~3においても、これらの記号は同じ内容を表す。

また、図2の例2-4において、「訳文を」が「修正した」に、「修正した」が「結果を」に、「結果を」が「表示する」に係るのが最優先解であるとする。この場合、「訳文を」が「表示する」に係るという別解は、最優先解との非交差性を満たしているが、「結果を」が「表示する」に係るという最優先解と格が重複してしまうため、本方式では検出しない。

このように、最優先解から遠く離れた別解の検出を抑制することにより、不要な曖昧性の検出を少なくすることができる。なお、この抑制により、検出漏れが生じて、係り受け構造の最優先解において、誤った係り受け関係があれば、そのうちの一つについては、必ず、その別解が検出される(付録I参照)。したがって、すべての曖昧性検出の結果が不要なものになるまで、前編集と曖昧性検出を繰り返すようにすれば、検

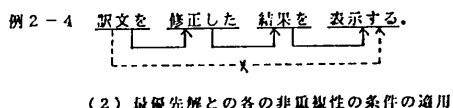
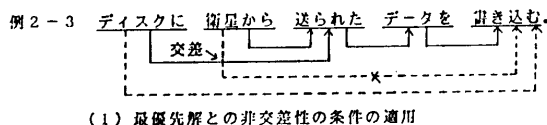


図2 最優先解から遠く離れた別解の検出の抑制の例
Fig. 2 Examples of the inhibition of detecting alternative dependency structures far from the highest-priority structure.

出漏れは存在しなくなる。こういう意味でも、この制約は有意義なものである。

(b) 不要な別解探索範囲の除外

日本語の文には、読点の付いていない文節は、特定の条件を満たす読点や特定の語句を飛び越えて係る可能性が少ないという傾向が見られる。本方式では、この傾向を利用して、不要な別解の検出原因となる範囲を別解探索範囲から除く。ただし、不要な別解探索範囲の除外の手掛かりとなる読点や語句は、対象とする文章の性質によって異なってくる。ここでは、特許抄録において有効性を確認したものとして、動詞、補助動詞を含む文節(連体修飾する文節は除く)の直後にある読点について述べる。特許抄録では、この読点を飛び越える連用の係り受けの別解を抑制すると、不要な曖昧性検出が少なくなる。これは、この読点が単文の区切りになっている可能性が高く、一つの文節がこの読点を飛び越えて係るようにするためには、その文節に読点を付ける必要性が高くなるからであると考えられる。この例を図3に示す。図3においては、「キーボードから」が「表示される」に係るという別解は、「与えら」という動詞を含む文節の直後の読点を飛び越えることになるので、この制約により別解として検出しない。なお、この別解探索範囲の除外の例外については、3.3節(4)(b)で述べる。

(c) 単語ごとの固有の構文的性質を利用した不要な別解検出の抑制

本方式では、上記の二つの別解検出の制約のほか、単語ごとの構文的性質を利用して、さらに、不要な別解の検出を抑制するための制約を設ける。この制約には、次の3種類のパターンがある。なお、ここで利用する単語の性質は、対象とする文章の性質により、多少異なってくるため、例としては、特許抄録で有効性を確認したものを示す。

(i) 依存関係の強い語句に関する不要な別解検出の抑制

係り受け構造の最優先解において、依存関係の強い語句の間に係り受け構造が成り立っている場合、その

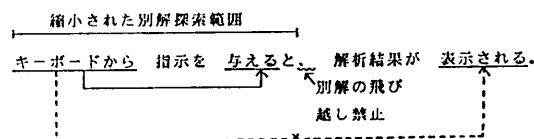


図3 不要な別解探索範囲の除外の例
Fig. 3 An example of excluding the unnecessary search space.

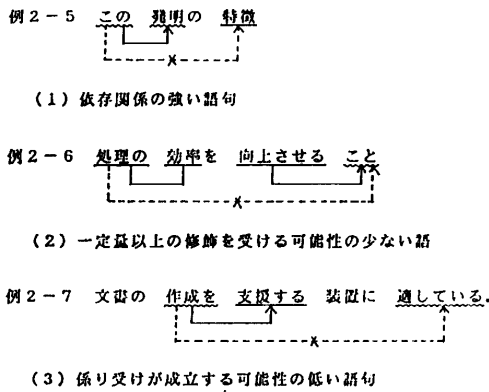


図 4 単語ごとの固有の構文的性質を利用した不要な別解検出の抑制の例
 Fig. 4 Examples of the inhibition of detecting unnecessary alternative dependency structures using the syntactic features of each word.

係る側の文節の係り先の別解を検出しないようにする。これは、このような場合、最優先解が正しい確率が非常に高くなるからである。

例 図 4 の例 2-5 のように、係り受け構造の最優先解において、隣接した、連体修飾要素と形式名詞（後方の語句と並列関係にないもの）の間に係り受けが成立している場合、その連体修飾要素の係り先の別解を一切検出しないようにする。

(ii) 最優先解において最大限の修飾を受けていると見なせる語句に関する別解検出の抑制

ある種の語句には一定量以上の修飾を受ける可能性が少ないという性質がある。本方式では、係り受け構造の最優先解においてそれらの語句にその上限と見なせる数の語句が係っている場合には、その語句に係るような別解を検出しないようにする。

例 図 4 の例 2-6 のように、係り受け構造の最優先解において、形式名詞に一つの語句が係っていれば、その形式名詞に係る別解は検出しない。

(iii) 係り受けが成立する可能性の低い語句に関する別解検出の抑制
 特定の語句間に係り受けが成立する可

能性が小さいという性質を利用して、そのような語句間の別解の検出を抑制する。

例 図 4 の例 2-7 のように、格支配の規則に適合しない別解の検出を抑制する。

3. 係り受け構造の曖昧性検出システム

上述した、日本語係り受け構造の曖昧性検出方式に基づいて、日英機械翻訳システムのための前編集の支援用に、係り受け構造の曖昧性を検出して警告するシステムを開発した。本章では、このシステムの構成と具体的な処理内容について述べる。

3.1 システム構成

本システムは、図 5 に示すように、

- (1) 最優先解決定部
- (2) 曖昧性検出部
- (3) 表示部
- (4) 辞書
- (5) 解析文法
- (6) 検出文法

から構成されている。本システムでは、原文が入力さ

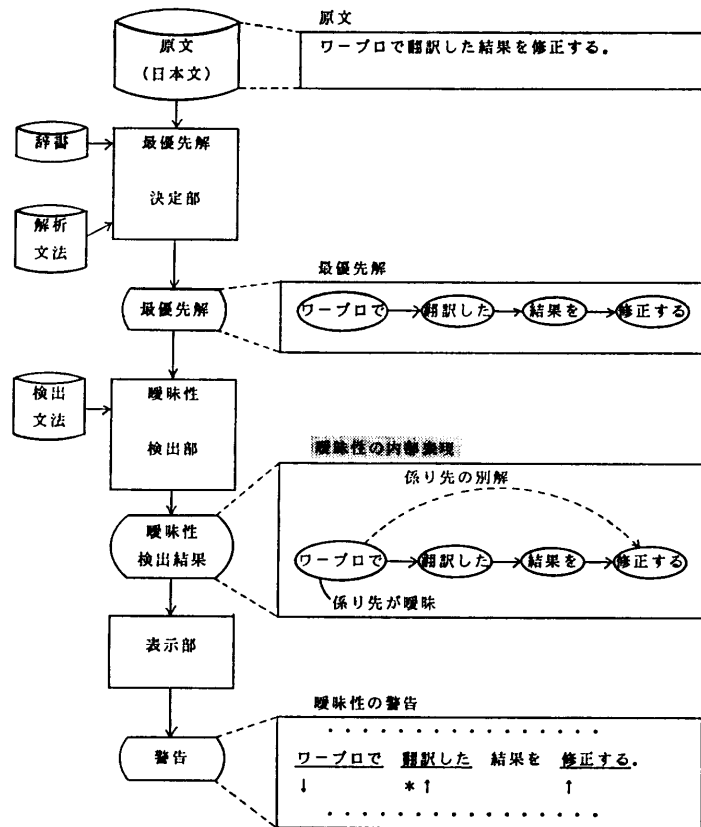


図 5 システム構成
 Fig. 5 The configuration of the system.

れると、最優先解決定部が係り受け構造の最優先解を決定し、曖昧性検出部が係り受け構造の曖昧性を検出し、表示部が曖昧性検出結果をプリンタに出力する。なお、最優先解決定部は筆者らが開発している日英機械翻訳システム^{4),5)}の日本語解析部と同一の処理をするものであり、出力する係り受け構造の最優先解はこの機械翻訳システムの決定する第1解と同じである。また、辞書としては、5万語の見出しを有するものを使用している。

3.2 係り受け構造の曖昧性の内部表現

係り受け構造の最優先解は、文節をノードとしたネットワーク構造で表現し、係り受け構造の別解は、最優先解のアークとは異なった種類のアークを用いて表現する。例えば、「ワープロで翻訳した結果を修正する。」という文において、「ワープロで」が「翻訳した」に係り、「翻訳した」が「結果を」に係り、「結果を」が「修正する」に係るのが係り受け構造の最優先解であり、かつ、「ワープロで」の係り先の別解が「修正する」である場合には、図5中の「曖昧性の内部表現」のように表現する。この表現法をとることにより、最優先解と別解を区別して表現できる。

3.3 曖昧性検出処理

本システムの係り受け構造の曖昧性検出処理は、特許抄録を対象としたものであり、不要な曖昧性検出を抑制するための具体的な制約条件は、特許抄録の文型を検討することにより定めたものである。係り受け構造の曖昧性の検出処理の流れを図6に示し、さらに、図6中の各処理の内容を以下に述べる。

(1) 別解探索範囲の認定：不要な別解探索範囲の除外の手掛かりとなる読点や語句を認識し、別解探索

範囲の境界に対応する文節ノードに目印を付与する。なお、本システムでは、表1に示す読点と語句を、不要な別解探索範囲の除外の手掛かりとして扱う。

(2) 依存関係の強い語句間に設定された係り受けの最優先解の認定：係り受け構造の最優先解において、依存関係の強い語句間に係り受けが設定されている場合、その係る側の文節ノードに、係り先の別解の検出不要の目印を付与する。なお、本システムで、認識の対象とした依存関係の強い語句の例を表2に示す。

(3) 最優先解において最大限の修飾を受けている語句の認定：係り受け構造の最優先解において、最大限の修飾を受けていると見なせる文節ノードに、係り

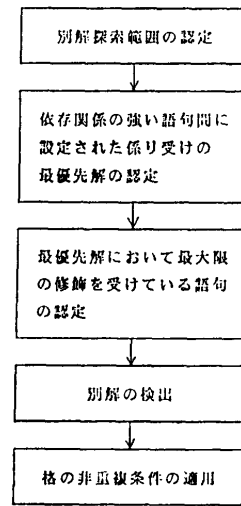


図6 係り受け構造の曖昧性検出処理の流れ
Fig. 6 The process of detecting ambiguities in dependency structures.

表1 不要な別解探索範囲を除外するための手掛かり
Table 1 Markers for excluding unnecessary search space.

No.	対象となる係り受けの種類	手掛かり	例
1	連用	動詞、補助動詞を含む文節（連体修飾の文節は除く）の直後にある読点	キーボードから 指示を 与え <u>ると</u> 、 解析結果が 表示 <u>される</u> 。 -----x-----
2	連用	「ことにより」、 「ことよって」 の直後にある読点	これに 表示盤を 設ける ことにより、 操作性の 向上を もた <u>らす</u> 。 -----x-----
3	連体	直前の語が体言や 並列を表す助詞・ 接続詞でない読点	共通した 部分を、 内蔵する メモリに 登録する 処 <u>理</u> -----x-----
4	連体	連用中止形の 動詞	ROMの アクセスタイムを 利用し 演算処理を 継続する 方 <u>式</u> -----x-----

表 2 依存関係の強い語句の分類とその例
Table 2 Groups of the words and phrases with strong dependencies and their examples.

No.	分類	例
1	機能語に近い役割を果たす語とそれに必須に係る文節	「～に」+「基づいて/従って/応じて/関して」
2	隣接した、動詞と、それに強く要求される助詞を持つ文節	「と」+「いう/呼ぶ」
3	呼応表現	「もし」+「～ならば」、「決して」+「～しない」
4	隣接した、連体修飾要素と形式名詞（後の語と並列にならないもの）	LEDを <u>駆動する</u> ことによる <u>消耗</u>
5	隣接した、連体修飾要素と「間/内/前/後」などの形式名詞に近い語（後の語と並列にならないもの）	<u>動作している</u> <u>間</u> に <u>割り込みを受け付けない</u> <u>装置</u>
6	隣接した、助詞「の」を伴った「属性を表す名詞」と、「ある/ない」や程度を表す形容詞の連体形	<u>拡張性の</u> <u>ある</u> <u>構造</u> / <u>応答性の</u> <u>良い</u> <u>システム</u>
7	隣接した連体詞と名詞（後の語と並列にならないもの）	<u>この</u> <u>発明の</u> <u>目的</u>

表 3 一定量以上の修飾を受ける可能性の少ない語の分類とその例

Table 3 Groups of the words that have a kind of limitation in the number of dependents and their examples.

No.	分類	例
1	形式名詞	<u>処理の</u> <u>効率を</u> <u>向上させる</u> <u>こと</u>
2	「間/内/前/後」などの形式名詞に近い語	<u>データの</u> <u>転送の</u> <u>間</u>
3	機能語的な役割を果たす語	<u>正常時には、</u> <u>入力に</u> <u>応じて</u> <u>変化する。</u>

受けの別解を受けない目印を付与する。なお、本システムでは、表 3 に示すような語を、一定量以上の修飾を受ける可能性が少ないものとして扱う。

(4) 別解の検出：係り受け構造の最優先解に対して非交差性の条件を満たし、かつ、上記(1)で目印が付与された、別解探索範囲の境界に対応する文節ノードを飛び越さない範囲で、各文節ノードの係り先の別解を検出する。ただし、以下の例外を設けている。

- (a) 読点が直後に付いた文節ノードの係り先の別解の検出においては、上述の条件を適用しない。
- (b) 別解探索範囲の境界に対応する文節ノードを A とし、係り受け構造の最優先解において、

A を飛び越えて係るような文節ノードを B とする。B や、B の前方にある文節ノードの係り先の別解の検出においては、A を別解探索範囲の境界と見なさない。

例えば、「我々が使用回数を、内蔵するメモリに登録する機能を実現した。」という文の係り受け構造の最優先解において、「我々が」と「使用回数を」が「登録する」に係っているとす。また、直後に読点が存在するために、「使用回数を」が別解探索範囲の境界の文節ノードになっているとする。このよう

な場合、「我々が」は、別解探索範囲の境界（「使用回数を」を飛び越えて「登録する」に係っているので、「我々が」の別解の検出においては、別解探索範囲の境界（「使用回数を」）を飛び越えるような別解、すなわち「実現する」に係るような別解も検出する。

- (c) 上記(2)で目印が付与された文節ノードの係り先の別解は検出しない。
- (d) 上記(3)で目印が付与された文節に係るような別解は検出しない。
- (5) 格の非重複条件の適用：係り受け構造の最優先解に対して格の非重複性の条件を満たさない係り先の別解を消去する。また、2.2 節(2)(c)(iii)で述べた、格支配の規則に適合しない別解を、用言の格支配に関する辞書情報に基づいて検出し、消去する。
- (6) 表示：得られた内部表現を基に、係り受け構造の曖昧性に関する警告を表示する。その表示例を図

```

*** 日本語文表示 (文単位) ***
端末に修正したプログラムのリストを表示する。
----- 係り受け構造の曖昧性 -----
ワープロで 翻訳した 計算機の マニュアルを 修正する。
          *↑
          ↓
ワープロで 翻訳した 計算機の マニュアルを 修正する。
          *↑
          ↓
ノメッセージノ
JPA1101: ①↑の語句が、*↑の語句に係るのか、
↑の語句に係るのかが曖昧です。②*↑の語句に係ると
解釈します。③システムの解釈が誤っている場合は、前
編集用記号【】を用いて係り先を明示して下さい。
    
```

図 7 係り受け構造の曖昧性に関する警告の例
Fig. 7 An example of warning of ambiguities in dependency structures.

7に示す。係り先が曖昧な文節の下に↓が、その係り先の最優先解の下に*↑が、別解の下に↑が表示される。また、①として、検出した曖昧性の内容が、②として、「曖昧な箇所」に対して機械翻訳システムが採用する第1の解釈が、③として、機械翻訳システムの解釈が誤っている場合の前編集方法がメッセージの形で示される。

4. 実験による本方式の評価

第3章で述べたシステムを用いて、電気・情報分野の特許抄録を対象とした実験を行うことにより、本方式の検出精度を評価する。

(1) 検出精度の評価指標

検出漏れの程度を評価するための指標として検出率を、不要な検出の程度を評価するための指標として一文当たりの平均ノイズ数と適合率を新たに設定する。これらの定義を以下に示す。

(a) 検出率

係り先の最優先解が誤っている文節が存在する場合に、検出率を以下のように定義する。

$$\text{検出率} = \frac{\text{ヒットした検出の数}}{\text{係り先の最優先解が誤っている文節の総数}} \times 100 (\%)$$

ここで、

ヒットした検出の数：検出した文節のうち、係り先の最優先解が誤っているものの数。

この定義により、検出率が100%に近いほど、検出漏れが少ないことになる。

(b) 一文当たりの平均ノイズ数

$$\text{一文当たりの平均ノイズ数} = \frac{\text{ノイズ数}}{\text{入力された文の総数}}$$

ここで、

ノイズ数：検出した文節のうち、係り先の文節が正しいものの数。

この定義により、一文当たりの平均ノイズ数が0に近いほど、不要な検出が少ないことになる。

(c) 適合率

$$\text{適合率} = \frac{\text{ヒットした検出の数}}{\text{検出した文節の総数}} \times 100 (\%)$$

この定義により、適合率が大きいほど無駄な検出が少ないことになる。

(2) 実験の概要

(a) 評価用テキスト

電気・情報分野の特許抄録100文を評価に使用する。

特許抄録を用いる理由は、ほとんどの文が長く複雑であるため、係り受け構造の曖昧な箇所が多く、検出率などについて、評価の母数を多くとれ、評価実験の正当性を高めることができるからである。

(b) 本システムの評価

本システムの処理の途中結果として得られる係り受け構造の最優先解の正誤を判断する。この結果に基づき、検出された文節について、係り先の最優先解が正しいものと誤っているものに分類し、評価指標の値を算出する。

(c) 従来方式の評価

本方式の能力を明確にするために、従来採用されていた、可能な構文構造をすべて検出する方式（以下、この方式を全解方式と呼ぶ）について上記(b)と同様のデータを収集し、評価指標の値を算出する。なお、ここで用いた全解方式は、2.1節(1)で述べた規則を満たす係り受け構造の解をすべて求め、係り先の解が二つ以上ある文節を係り先が曖昧であると見なすものである。

(3) 実験結果

(a) 評価用テキストに関するデータ

- 総文数=100文（電気・情報分野の特許抄録）
- 総文字数=13996文字
- 平均文長=13996文字/100文
≒140文字/文
- 構文的に係り先の曖昧な文節の総数
=672文節

(b) 各方式の評価指標に関するデータ

各方式の評価指標に関するデータを表4に示す。なお、一文当たりの平均ノイズ数と適合率について、本

表4 本方式と全解方式の比較
Table 4 Comparison between the proposed method and the existing one.

項目	本方式	全解方式
ヒットした検出数	76個	83個
ノイズ数	259個	589個
検出文節の総数	335個	672個
検出漏れ数 ^{a)}	7個	0個
検出率	91.6%	100%
一文当たりの平均ノイズ数	2.59個/文	5.89個/文
適合率	22.7%	12.4%

^{a)} 検出漏れ数：係り先の最優先解が誤っている文節のうち、検出できなかったものの数。

表 5 本方式における検出漏れの原因
Table 5 Causes of omission of detection
in the proposed method.

検出漏れの原因	数(個)
非交差性条件	5
読点による別解検出 範囲の縮小	2
検出漏れ総数	7

方式と全解方式の比を算出すると、次のようになる。

- 一文当たりの平均ノイズ数における本方式と全解方式の比=2.59 : 5.89

$$\approx 1 : 2.27$$

- 適合率における本方式と全解方式の比

$$= 22.7 : 12.4$$

$$\approx 1.83 : 1$$

(c) 本方式における検出漏れに関するデータ

- 検出漏れ数(係り先の最優先解が誤っている文節のうち、検出できなかったもの数)=7個
- 検出漏れ1個当たりの、構文的に係り先の曖昧な文節の数=672 文節/7個
=96 文節/個
- 検出漏れ1個当たりの、文字数
=13996 文字/7個
=1999 文字/個
- 検出漏れの原因:表5参照。

(4) 考 察

(a) 本方式の不要な検出の抑制能力について

本方式では、一文当たりの平均ノイズ数については、全解方式の 1/2.27 に減っているし、適合率については、全解方式の 1.83 倍に上がっており、全解方式と比べて不要な検出が大きく減っている。したがって、本方式を用いれば、不要な検出を大きく減らすことができることが実験的に確認できた。

(b) 本方式における検出漏れについて

本方式では、必要な検出も抑制してしまうことがあるが、①検出率は 91.6% である、②平均して、構文的に係り先の曖昧な文節 96 個に1個の割合でしか検出漏れが生じない、③平均 1999 文字に1個の割合でしか検出漏れが生じない、の3点から、実用上の障害にはならないと考える。

(c) 本方式における各別解検出の制約について

- (i) 係り受け構造の最優先解から遠く離れた別解の検出の抑制

本制約は、関連する2個以上の文節の係り先が最優先解と異なるような別解のうちの多くを検出しないように働く。例えば、図2において、「ディスクに」と「衛星から」の二つの文節の係り先については、この二つの係り先が同時に最優先解と異なってしまような別解は検出されない。このため、最優先解から遠く離れた別解の検出は抑制され、本制約は不要な検出を抑制する能力が高い。また、表5に示したように、本制約のうちの非交差性の条件は、検出漏れを引き起こすことがあるが、その量は少ない。非交差性の条件は、関連する2個以上の文節の係り先が誤っていると、検出漏れを引き起こすことがあるが、精度の高い最優先解においては、関連する文節の係り先が2個以上同時に誤っていることが少ないために、検出漏れが少量に抑えられると解釈できる。

(ii) 不要な別解探索範囲の除去

特許抄録には、以下の性質があるために、表1に基づく別解探索範囲の除去が有効に機能すると考える。

- ①動詞、補助動詞を含む文節(連体修飾する文節は除く)の直後にある読点、および、「ことにより」などの直後にある読点は単文の区切りとなっている可能性が高く、それを飛び越えて係るためには、読点が必要になることが多い。
- ②直前の語が体言や並列を表す助詞・接続詞でない読点は、並列要素以外の名詞句の区切りとなっている可能性が高いし、連用中止形の動詞は動詞句の並列を形成する可能性が高いため、それを飛び越えて連体修飾するためには、読点が必要になることが多い。

なお、上記の読点に関する考えは、昭和21年3月に文部省教科書局調査課国語調査室によって作成された、「くぎり符号の使い方【句読法】案」⁷⁾や、正書法に関する考察⁸⁾などとも矛盾しないものである。

また、表5に示すように検出漏れが少ないのは、上記の性質の存在のほかに、この性質からはずれるものが出現した場合は、精度の高い最優先解において、本制約が働かないように係り受け構造が設定される場合が多いためだと考えられる。

(iii) 単語ごとの固有の構文的性質を利用した不要な別解検出の抑制

特許抄録では、この制約の適用対象となる、形式名詞、形式名詞に近い語、連体詞、機能語に近い役割を果たす語が多く使用されているために、この制約による不要な検出の抑制効果は大きい。

また、表5に示すように、本実験では、この抑制が原因で生じた検出漏れはない。したがって、この制約による悪影響はかなり少ないものと判断できる。本実験で、検出漏れが生じなかった理由は、各語が強く持っている構文的性質を利用したためと考える。

(d) 総合判断

上記(a), (b), (c)の考察により、本方式は、必要な曖昧性検出の漏れを実用上問題ない量に抑えながら、不要な曖昧性検出を少なくするものであると判断する。

また、本方式では、不要な曖昧性の検出抑制の結果として、メモリの使用効率の向上という効果も得られる。本方式では、係り先が曖昧であると判断される文節の数が全解方式の約 $1/2$ (=検出文節の総数の比 = $335/672$) になるので、別解の表現に要するメモリ容量も、その分だけ減ることになる。

さらに、本方式では、最優先解と別解を明確に区別して扱うため、係り受け構造の曖昧性に関する警告において、両者を区別して示すことが可能になる(図7参照)。したがって、ユーザはすべての警告をチェックし、最優先解が誤っている場合だけ前編集すればすむような前編集支援システムの提供が可能となり、すべての曖昧性検出に対して、正しい係り先の指示を要求することを前提としている従来の方式に比べて前編集者の負担を少なくできる。

なお、ここで、翻訳に利用する機械翻訳システムの性能が本実験の場合よりも悪い場合の考察を示す。まず、本実験に用いた機械翻訳システムの正解率(構文的に係り先の曖昧な文節のうち、係り先の最優先解が正しい文節の占める割合)を求める。本実験での構文的に係り先の曖昧な文節の総数は、672個である。また、全解方式では、構文的に係り先が曖昧な文節をすべて検出するので、ヒットした検出の数(83個:表4参照)が、係り先の最優先解が誤っているものの総数に等しい。したがって、本実験に用いた機械翻訳システムの正解率は、 $\{(672-83)/672\} \times 100\% \approx 87.6\%$ となる。ここで、これよりも機械翻訳システムの正解率が低い場合、例として、50%の場合を想定し、全解方式と本方式を比較する。

(i) 全解方式(機械翻訳の正解率が50%の時)

全解方式では、構文的に係り先が曖昧な文節をすべて検出するのだから、検出率は100%となる。また、全解方式の適合率は、50%(=100%-正解率)となる(付録II参照)。さらに、本実験と同じ個数の文節の係

り先が構文的に曖昧であるとする、一文当たりの平均ノイズ数は、 $672 \times (100\% - 50\%) / 100 = 3.36$ 個になる。したがって、検出能力と、不要な曖昧性検出の少なさの両方の観点から、全解方式はかなり良いものとなる。

(ii) 本方式(機械翻訳の正解率が50%の時)

機械翻訳システムの正解率が本実験の場合より半分近くも低くなることにより、関連する2個以上の文節の係り先の最優先解が同時に誤る確率が高くなり、検出漏れが多くなることが予想される。つまり、単純化のために、各文節の係り先の最優先解が誤りとなる事象が、それぞれ独立であると近似すると、二つの文節の係り先の最優先解が同時に誤る確率は、本実験の場合、 $(100\% - \text{正解率})^2 = (100\% - 87.6\%)^2 \approx 1.5\%$ 、正解率が50%の場合、 $(100\% - 50\%)^2 = 25\%$ となる。したがって、正解率が50%になることにより、二つの文節の係り先の最優先解が同時に誤る確率が、17倍近く高くなることになる。このため、関連する2個以上の文節の係り先の最優先解が同時に誤って、検出漏れを引き起こす確率も同様に高くなり、検出率はかなり低くなることが予想される。仮に、検出漏れの数が本実験の場合の17倍になると想定すると、検出率は $(672 \times 50\% - 7 \times 17) / (672 \times 50\%) \approx 64.6\%$ となる。こうなると、曖昧性検出システムの検出結果だけに注目しても十分な前編集ができなくなり、前編集支援の効果がなくなってしまう。

上記(i), (ii)の考察から、機械翻訳システムの正解率が50%というように低い場合には、本方式よりも、全解方式を利用した方が効果的であると判断できる。しかしながら、利用する機械翻訳システムの正解率が低い場合には、修正の手間が多くなるので、翻訳作業に機械翻訳システムを使用すること自体に難がある。すなわち、機械翻訳システムの正解率が低い場合には、曖昧性検出方式の性能の議論以前に、機械翻訳システムの問題の方が大きくなってしまう。

5. むすび

構文構造の曖昧性検出において、必要な曖昧性検出の漏れを抑えながら、不要な曖昧性検出を削減して、前編集者の負荷を軽減させる方式として、①機械翻訳システムと同一の解析を行って機械翻訳システムの採用する係り受け構造の最優先解を求め、これに対して係り受けの非交差条件と格の非重複条件を満たす範囲に別解の検出を抑える、②係り受けの飛び越しを抑制

する傾向のある読点や語句を利用して、別解の検出範囲を小さくする、③語句ごとの固有の構文的性質を利用して別解の検出を制限する、の3点を特徴とした曖昧性検出方式を提案した。この方式の要点は、機械翻訳システムの採用する最優先解を求め、その最優先解に対して一定以内の近さにある係り受け構造の別解だけを検出することである。さらに、本方式に基づいて、係り受け構造の曖昧性検出システムを開発し、特許抄録を対象として評価した。この結果、本方式は、必要な曖昧性検出の漏れを実用上問題ない量に抑えながら、不要な曖昧性検出を、可能な構文構造をすべて検出する従来方式の1/2未満に削減でき、本方式により効率の良い前編集支援が可能になることが確認できた。

謝辞 本研究に当たりご指導頂いた、(株)日立製作所システム開発研究所 堂免信義所長、石原孝一郎主管研究員、渡邊坦主管研究員、ならびに、有益なご意見を頂いた(株)日立製作所ソフトウェア工場 高橋栄部長、吉村紀久雄主任技師に深謝いたします。評価に用いた特許抄録は、(財)日本特許情報機構より提供して頂きました。また、本論文の執筆に当たりご指導頂いた、(株)日立製作所システム開発研究所 相川博之主任研究員、吉原都夫主任研究員、ならびに、研究の遂行に当たって協力して頂いた飯泉紀子さんに深謝いたします。

参 考 文 献

- 1) 有田ほか：日英機械翻訳システムにおけるプリエディットについて，情報処理学会自然言語処理研究，48-7 (1985)。
- 2) 平井ほか：日英機械翻訳用前編集支援システムに関する一考察，第35回情報処理学会全国大会論文集，3S-6，pp. 1243-1244 (1987)。
- 3) 平井ほか：日英機械翻訳用前編集支援システム(1)―構文的曖昧性の検出方式―，第36回情報処理学会全国大会論文集，2U-2，pp. 1229-1230 (1988)。
- 4) 梶ほか：日英機械翻訳のための日本語文の依存構造解析，第30回情報処理学会全国大会論文集，7G-1，pp. 1579-1580 (1985)。
- 5) 梶ほか：意味処理に基づく機械翻訳システム，日立評論，Vol. 69，No. 3，pp. 37-46 (1987)。
- 6) 小泉：日本語の正書法，p. 456，大修館書店(1978)。
- 7) 文部省教科書局調査課国語調査室：くぎり符号の使ひ方【句読法】案，(1946)，in 言葉に関する問答集 12，pp. 71-82，文化庁 (1986)。
- 8) 長尾ほか：制限文法にもとづく文章作成援助

システム，情報処理学会自然言語処理研究，44-5 (1984)。

- 9) Tomita, M.: Sentence Disambiguation by Asking, *Computers and Translation*, Vol. 1, pp. 39-51 (1986)。
- 10) 吉田ほか：機械翻訳のための構文解析方法，情報処理，Vol. 26，No. 10，pp. 1157-1164 (1985)。

付 録 I

係り受け構造の最優先解との非交差性，格の非重複性を満たす範囲に別解の検出を抑制しても，係り受け構造の最優先解において，誤った係り受け関係があれば，そのうちのひとつについては，必ず，係る側の文節の係り先の別解が検出されることの理由

誤った係り受け関係の種類に応じて，次の二つの場合を考える。

(1) 係り受け構造の最優先解において，正しい係り先よりも，後方の文節に係る係り受け関係が存在する場合 (図8参照)

「正しい係り先よりも，後方の文節に係る係り受け関係」のうちの任意の一つに着目し，その係る側の文節を i ，受ける側の文節を x ，この係り受け関係自体を (i, x) とする。今，文節 i の係り先の別解がすべて検出されなかったとする。文節 i の正しい係り先 (これを j とする) は，文節 i の係り先の別解の一つに当たるから，別解検出の抑制に用いられる次の二つの条件

- (a) 係り受け構造の最優先解との非交差性を満たす別解しか検出しない。
- (b) 係り受け構造の最優先解との格の非重複性を満たす別解しか検出しない。

のうち，少なくともいずれか一方により，文節 i から文節 j に係るといふ，正しい係り受け関係 (これを (i, j) とする) の，別解としての検出が抑制され

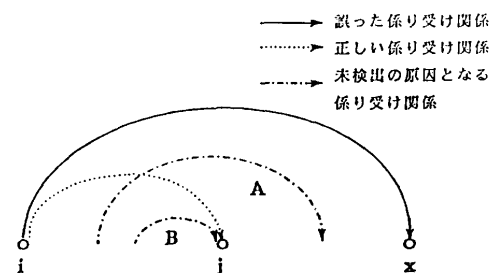


図 8 正しい係り先よりも，後方に係る文節が存在する場合

Fig. 8 Case that a bunsetsu incorrectly depends on a farther bunsetsu than the right one.

たことになる。したがって、係り受け構造の最優先解において、少なくとも、①正しい係り受け関係 (i, j) と交差する係り受け関係があるか、②文節 i と格が同一の文節が文節 j に係っているかのいずれかであることになる。

①係り受け構造の最優先解において、正しい係り受け関係 (i, j) と交差する係り受け関係がある (これを係り受け関係 A とする) 場合

係り受け構造の最優先解は、非交差性の規則 (2.1 節(1)の(b)参照) を満たすように決定されるはずであるから、係り受け関係 A は、(i, x) と交差しない。このことと、係り受け関係 A が (i, j) と交差することと、係りの後方性の規則 (2.1 節(1)の(a)参照) より、係り受け関係 A における、係る側の文節は、文節 i と j の間 (文節 i, j は含まない) にあり、受ける側の文節は、文節 j より後方にあることになる。一方、正しい係り受け関係は、非交差性の規則 (2.1 節(1)の(b)参照) により、(i, j) と交差しないから、文節 i, j 間にある文節 (文節 i, j は含まない) の正しい係り先は、文節 j か、文節 j の前方の文節になる。これらを考え合わせると、係り受け関係 A は、正しい係り先よりも、後方の文節に係っている係り受け関係であることになる。

②係り受け構造の最優先解において、文節 i と格が同一の文節が文節 j に係っている (これを係り受け関係 B とする) 場合

係り受け構造の最優先解は、非交差性の規則 (2.1 節(1)の(b)参照) を満たすように決定されるはずであるから、係り受け関係 B は、(i, x) と交差しない。このことと、係り受け構造の最優先解は、係りの後方性と唯一性の規則 (2.1 節(1)の(a)参照) を満たすように決定されるはずであることと、係り受け関係 B における受ける側の文節は j であることから、係り受け関係 B における係る側の文節は、文節 i と j の間 (文節 i, j は含まない) にあることになる。一方、上記①で考察したように、文節 i, j 間にある文節 (文節 i, j は含まない) の正しい係り先は、文節 j か、文節 j の前方の文節になる。ところが、係り受け関係 B は、正しい係り受け関係 (i, j) との格の非重複性 (2.1 節(1)の(c)参照) を満たさないため、係り受け関係 B における係る側の文節の正しい係り先は、文節 j ではない。これらのことにより、係り受け関係 B は、正しい係り先よりも、後方の文節に係っている係り受け関係であることになる。

したがって、①、②のどちらの場合でも、係り受け関係 (i, x) とは別の「正しい係り先よりも、後方の文節に係る係り受け関係」が存在することになる。しかも、上記の考察から容易に分かるように、その係り受け関係は、係り受け関係 (i, x) の内側に位置するものになる。

以上のことより、もし、係り受け構造の最優先解において、「正しい係り先よりも、後方の文節に係る係り受け関係」が存在するにもかかわらず、そのうちのどの一つについても、係る側の文節の係り先の別解が検出されないならば、任意の一つの「正しい係り先よりも、後方の文節に係る係り受け関係」の内側に、それとは別の「正しい係り先よりも、後方の文節に係る係り受け関係」が存在し、さらに、その内側に、それらとは別の「正しい係り先よりも、後方の文節に係る係り受け関係」が次々と存在することになり、無限個の係り受け関係が、係り受け構造の最優先解の中に存在することになってしまう。ところが、係り受け構造の最優先解の中に、係り受け関係は有限個しか存在しないので、仮定が誤っていることになる。ゆえに、係り受け構造の最優先解において、「正しい係り先よりも、後方の文節に係る係り受け関係」が存在するならば、上記(a), (b)の条件により、別解の検出を抑制しても、「正しい係り先よりも、後方の文節に係る係り受け関係」のうちの一つについては、必ず、係る側の文節の係り先の別解が検出されることになる。

(2) 係り受け構造の最優先解において、正しい係り先よりも前方の文節に係る係り受け関係が存在し、かつ、正しい係り先よりも後方の文節に係る係り受け関係が存在しない場合 (図 9 参照)

「正しい係り先よりも、前方の文節に係る係り受け関係」のうち任意の一つに着目し、その係る側の文節を i 、受ける側の文節を x 、この係り受け関係自体

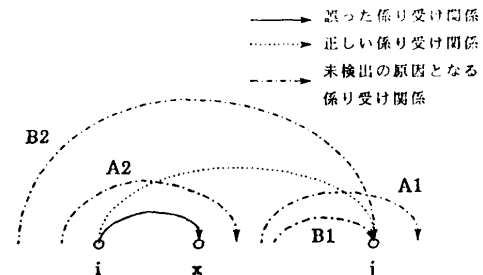


図 9 正しい係り先よりも、前方に係る文節が存在する場合

Fig. 9 Case that a bunsetsu incorrectly depends on a nearer bunsetsu than the right one.

を (i, x) とする。今、文節 i の係り先の別解がすべて検出されなかったとする。文節 i の正しい係り先 (これを j とする) は、文節 i の係り先の別解の一つに当たるから、上記(1)中に示した条件(a), (b)の少なくともいずれか一方により、文節 i から文節 j に係るといふ、正しい係り受け関係 (これを (i, j) とする) の、別解としての検出が抑制されたことになる。したがって、係り受け構造の最優先解において、少なくとも、①正しい係り受け関係 (i, j) と交差する係り受け関係があるか、②文節 i と格が同一の文節が文節 j に係っているかのいずれかであることになる。

①係り受け構造の最優先解において、正しい係り受け関係 (i, j) と交差する係り受け関係がある (これを係り受け関係Aとする) 場合

係り受け関係Aは (i, j) と交差するのであるから、係りの後方性の規則 (2.1節(1)の(a)参照) を考慮することにより、係り受け関係Aは、

- (i) 文節 i, j の間 (文節 i, j は含まない) から文節 j の後方に係るもの (図9中のA1参照)
- (ii) 文節 i の前方の文節から文節 i, j の間 (文節 i, j は含まない) に係るもの (図9中のA2参照)

のうちのいずれかであることになる。正しい係り受け関係は、非交差性の規則 (2.1節(1)の(b)参照) により、 (i, j) と交差しないから、文節 i, j の間にある文節の正しい係り先は、文節 j か文節 j よりも前方の文節でなければならない。また、同じ理由により、文節 i より前方にある文節の正しい係り先は、(イ)文節 i か文節 i よりも前方の文節、あるいは、(ロ)文節 j か文節 j よりも後方の文節でなければならない。したがって、係り受け関係Aは、(i)の場合と、(ii)で(イ)の場合は、「正しい係り先よりも後方の文節に係る係り受け関係」となり、(ii)で(ロ)の場合は、「正しい係り先よりも前方の文節に係る係り受け関係」となる。ところが、仮定より、この場合、「正しい係り先よりも後方の文節に係る係り受け関係」は存在しないので、(ii)で(ロ)の場合しかあり得ない。また、係り受け構造の最優先解は、非交差性の規則 (2.1節(1)の(b)参照) を満たすように決定されるはずであるから、係り受け関係Aは、係り受け関係 (i, x) と交差しない。これらのことにより、係り受け関係 (i, x) とは別の「正しい係り先よりも前方

の文節に係る係り受け関係」が、係り受け関係 (i, x) の外側に存在することになる。

②係り受け構造の最優先解において、文節 i と格が同一の文節が文節 j に係っている (これを係り受け関係Bとする) 場合

係り受け構造の最優先解は、係りの後方性と唯一性 (2.1節(1)の(a)参照) の規則を満たすように決定されるはずであるから、係り受け関係Bは、

- (i) 文節 i, j の間 (文節 i, j は含まない) から文節 j に係るもの (図9中のB1参照)
- (ii) 文節 i の前方から文節 j に係るもの (図9中のB2参照)

のいずれかであることになる。正しい係り受け関係は、非交差性の規則 (2.1節(1)の(b)参照) により、 (i, j) と交差しないし、格の非重複性の規則 (2.1節(1)の(c)参照) により、文節 i と格が同一の文節の正しい係り先は文節 j にはならない。それゆえ、文節 i と格が同一で、かつ、文節 i と j の間にある文節 (文節 i, j は含まない) の正しい係り先は、文節 j よりも前方でなければならない。また、同じ理由により、文節 i と格が同一で、かつ、文節 i よりも前方にある文節の正しい係り先は、(イ)文節 i か文節 i よりも前方の文節、あるいは、(ロ)文節 j よりも後方の文節でなければならない。したがって、係り受け関係Bは、(i)の場合と、(ii)で(イ)の場合は、「正しい係り先よりも後方の文節に係る係り受け関係」となり、(ii)で(ロ)の場合は、「正しい係り先よりも前方の文節に係る係り受け関係」となる。ところが、仮定より、この場合、「正しい係り先よりも後方の文節に係る係り受け関係」は存在しないので、(ii)で(ロ)の場合しかあり得ない。これらのことにより、係り受け関係 (i, x) とは別の「正しい係り先よりも前方の文節に係る係り受け関係」が、係り受け関係 (i, x) の外側に存在することになる。

したがって、①, ②のどちらの場合でも、係り受け関係 (i, x) とは別の「正しい係り先よりも、前方の文節に係る係り受け関係」が係り受け関係 (i, x) の外側に存在することになる。

以上のことより、もし、係り受け構造の最優先解において、「正しい係り先よりも、前方の文節に係る係り受け関係」が存在し、かつ、「正しい係り先よりも後方の文節に係る係り受け関係」が存在しない状態で、「正しい係り先よりも、前方の文節に係る係り受け関係」のうちのどの一つについても、係る側の文節

の係り先の別解が検出されないならば、任意の一つの「正しい係り先よりも、前方の文節に係る係り受け関係」の外側に、それとは別の「正しい係り先よりも前方の文節に係る係り受け関係」が存在し、さらに、その外側に、それらとは別の「正しい係り先よりも、後方の文節に係る係り受け関係」が次々と存在することになり、無限個の係り受け関係が、係り受け構造の最優先解の中に存在することになってしまう。ところが、係り受け構造の最優先解の中に、係り受け関係は有限個しか存在しないので、仮定が誤っていることになる。ゆえに、係り受け構造の最優先解において、「正しい係り先よりも、前方の文節に係る係り受け関係」が存在し、かつ、「正しい係り先よりも、後方の文節に係る係り受け関係」が存在しなければ、上記(a), (b)の条件により、別解の検出を抑制しても、「正しい係り先よりも、前方の文節に係る係り受け関係」のうちの一つについては、必ず、係る側の文節の係り先の別解が検出されることになる。

(1), (2)を総合することにより、「係り受け構造の最優先解との非交差性、格の非重複性を満たす範囲に別解の検出を抑制しても、係り受け構造の最優先解において、誤った係り受け関係があれば、そのうちの一つについては、必ず、係る側の文節の別解が検出される」が成り立つことになる。

付 録 II

全解方式の適合率

全解方式においては、

検出した文節の総数

= 構文的に係り先の曖昧な文節の総数

ヒットした検出の数

= 係り先の最優先解が誤っている文節の総数

である。よって、

適合率

$$= \frac{\text{係り先の最優先解が誤っている文節の総数}}{\text{構文的に係り先の曖昧な文節の総数}} \times 100 (\%)$$

となる。これに対して、

構文的に係り先の曖昧な文節の総数

= 係り先の最優先解が誤っている文節の総数

+ 係り先の最優先解が正しい文節の総数

という関係を適用すると、

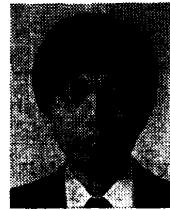
適合率=100%

$$\frac{\text{係り先の最優先解が正しい文節の総数}}{\text{構文的に係り先の曖昧な文節の総数}} \times 100 (\%)$$

∴ 適合率=100%-正解率

(昭和 63 年 9 月 8 日受付)

(平成 2 年 7 月 10 日採録)



平井 章博 (正会員)

昭和 32 年生。昭和 54 年早稲田大学電子通信学科卒業。昭和 56 年同大学院修士課程修了。同年(株)日立製作所入社。現在、同社システム開発研究所研究員。昭和 63 年～平成 1 年カーネギー・メロン大学客員研究員。自然言語処理、ニューラル・ネットワークの研究に従事。電子情報通信学会、ACM 各会員。



梶 博行 (正会員)

昭和 26 年生。昭和 48 年京都大学工学部電気工学第二学科卒業。昭和 50 年同大学院修士課程修了。同年(株)日立製作所入社。現在、同社システム開発研究所主任研究員。自然言語処理、機械翻訳システムの研究開発に従事。電子情報通信学会、ACM、ACL 各会員。



芦沢 実 (正会員)

昭和 36 年生。昭和 58 年早稲田大学理工学部電子通信学科卒業。昭和 60 年同大学院修士課程修了。同年、(株)日立製作所システム開発研究所入社。以来、日英機械翻訳、自然言語処理の研究に従事。電子情報通信学会会員。