

# 類似する用例を用いた日本語係り受け解析とその評価

2L-3

廣田 啓一 椎野 努 河合 敦夫 田添 文博

三重大学工学研究科

鈴鹿高等専門学校

## 1. はじめに

日本語における係り受けの曖昧性を解決する手法の一つとして、当該文書や同分野の文書から確実あるいは確実性の高い係り受け用例を学習し、曖昧な係り受けの解析に利用する試みが幾つかなされてきた。しかし、これまでの方法では、獲得できる用例の数がそれほど多くない事、獲得した用例を利用できる機会が乏しい事などから、効果的な結果を得るに至っていない。

本稿では、学習する係り受けの範囲を拡大し確実性に依拠して評点を与えるとともに、適用に際しては学習用例との意味的距離を考慮して適用する事で、学習量、適用量ともに拡張する方法を提案する。

## 2. 係り受けの確実性と評点

係り受け用例を抽出するにあたっては、抽出規則が構文的な情報に依存する事、また用例適用時に妥当性を判断する基準が必要となる事から、その抽出規則の確実性を論ずる必要がある。これまでの研究では主に確実・準確実・不確実などの三段階の基準を設けるものが多かったが、規則自体が大まかなものであり、評価基準も必ずしも明確ではなかった。そこで機能語や単語間の距離など構文的に評価できる情報から抽出規則を細分化し、文章中における出現数や適合数などの統計的調査の結果に基づいて、その確実性を示す評点を与えた。

主な抽出規則とその評点は次の通りである。表中「後に複数の動詞」がある場合とは、同一の助詞あるいは句読点の出現までに別の動詞が存在する場合を言い、表現が入れ子構造になっ

ていて係り受けが曖昧な場合が多いので、前出の各規則の評点から-3する事を示している。

パターン	抽出規則	評点
名詞+を+動詞	直後に動詞	10
	修飾語の後、動詞	9
	補語の後、動詞	7
名詞+に+動詞	後に複数の動詞	-3
	直後に動詞	10
	修飾語の後、動詞	9
名詞+が+動詞	目的語の後、動詞	8
	直後に動詞	9
	後に複数の動詞	-3
名詞+の+名詞	間に補語・目的語	7
	直後に動詞	9
	後に複数の動詞	-3
名詞+の+動詞	後ろが普通名詞	10
	後ろがサ変名詞	7
	三つ以上の連結	5
名詞+の+動詞	直後に動詞	9

以上のような抽出規則に基づいて、構文的に用例が抽出される。例えば、下の例文では、規則「名詞+の+名詞」から「新任+の+教授」という用例が、また規則「名詞+を+動詞」から「教授+を+案内する」という用例が、それぞれ抽出される。

**新任の教授を案内する**  
 名詞の名詞      名詞を動詞

## 3. 単語の類似度と意味的距離

抽出した用例を適用する際に、単純に単語の一致を調べるだけの方式では適用可能な用例が非常に少なく、高い効果を望む事はできない。したがって、ここでは、抽出用例との間に意味の類似性をもつ対象単語にも用例を適用可能として解析可能な範囲を広げ、その効果を調べた。

類似性を調べるための辞書として、一般に広く用いられている国立国語研究所編纂の分類語彙表を利用した。分類語彙表では、品詞および意味分類に即した5桁のコードが各単語に振ら

れており、意味距離を計算する手段として主に一致する桁数を調べる方法がよく知られている。すなわち認識：1.3062と意味：1.3070であれば、1.30の部分が一一致しており、後ろ2桁が一致しない事から意味距離を2として考える。

類似コード	単語	意味距離
1.3062	認識	—
1.3060	知識	1
1.3070	意味	2
2.3062	認識する	0

#### 4. 学習用例の適用

用例適用の対象となるのは、構文解析で特定できない係り受け関係、およびその深層格構造が決定できない係り受け関係である。すなわち先の確実性に基づく抽出パターンに該当しないもの、または該当してもその確実性が低いもの、および構文的には確実であっても意味的には複数の解釈が存在するものを対象に、用例の適用可能性を調べるのである。

**研究室の教授も案内する**  
教授を案内する

たとえば上の例文の係り受けは構文的には確実だが、名詞が主格または目的格となりうるので意味的に曖昧である。そこで抽出用例の中に目的格と判断できる用例があれば、適用によりその関係は目的格と考えられるので、「教授+を+案内する」として確定する。

用例の適用にあたり、対象となる助詞の持ちうる格の種類によって、適用できる用例の制限がある。例えば、先の「教授も案内する」という文に対しては、助詞「も」に主格または目的格が考えられるので、「が・を」用例が適用可能であるが、「に・の」用例は適用する事が出来ない。今回用いた適用制限を下の表に示す。

助詞	が	は	を	に	の	で	も	と	や
が用例	○	○	×	×	○	×	○	○	○
を用例	○	○	○	×	○	×	○	○	○
に用例	×	×	×	○	×	○	×	○	○
の用例	×	×	×	×	○	×	×	×	○

また用例適用の際に、パターンの異なる複数の用例が適用可能である場合、また適用可能な用例の確実性が極端に低く解析の効果が薄い場合が考えられる。したがって、抽出用例の確実性と適用単語との類似性の二つの評点をもとに、係り受け用例の適用が可能かどうかを判断する基準値を設定する事にした。これにより不適格な用例の適用を除去できるため、全体の解析精度を高める事が出来るものと考えられる。

#### 5. 実験結果と考察

今回は文献中の一節（24文、388文節）を対象に抽出・適用処理を行い、評価を行った。

まず、類似性無しで単語の一致のみで適用した場合と意味距離0~2で適用した場合のそれぞれについて調べ、表にまとめた。左の数字は適用が可能であった用例数、右はその内適用が妥当であった用例の割合である。

適用	類似無し	距離=0	距離=1	距離=2
が用例	1	100%	2	100%
を用例	2	100%	7	86%
に用例	0	—	2	50%
の用例	0	—	1	100%
全体	3	100%	12	83%

次に用例適用の際に確実性評点と意味距離に基づいた適用可能基準値を設定し、評価を行った。左の数字は実際に適用された用例の数、右はその内適用が妥当であった用例の割合である。

用例	7点以上	4点以上	制限なし
が用例	1/1	100%	2/4
を用例	21/29	72%	24/35
に用例	1/1	100%	3/8
の用例	8/8	100%	8/9
全体	31/39	79%	37/56

表から明らかなように、抽出規則に対し確実性に基づく評点を付与する事で、より詳細な係り受け解析を行う事が出来、効果がある事が分かる。また類似用例の利用により解析数を大幅に増加出来、かつ確実性評点との組み合わせにより解析精度の低下はわずかにとどめる事が出来る事も分かる。評点の付与および意味距離の利用は非常に有効な手法であるといえる。