

係り受け解析のための動詞連用形用法の曖昧性解消

麻岡 正洋 颯々野 学

{asaoka.masahiro, sassano}@jp.fujitsu.com

(株) 富士通研究所

1 はじめに

「待ち受け」のような日本語動詞の連用形には、通常の動詞的な用法に加えて、名詞的に使われるものが数多く存在している。以下に例を示す。

1. 代表を待ち受け、ねぎらった長嶋監督。(動詞的用法)
2. 子供の写真を待ち受けに設定する。(名詞的用法)

典型的な形態素解析では、形態素解析用の辞書をベースにしており、辞書に記述のある形態素(品詞)を解析の候補としている。そのため、名詞の辞書エントリのない動詞連用形は名詞としては判定できず、係り受け解析で誤りの原因となっている。上記の2.を例にとると、動詞連用形「待ち受け」を動詞と判定して、「写真を」は「待ち受け」に係ると判定する可能性がある。

この問題の解決策の1つとして、名詞的に使用される可能性のある動詞連用形を、すべて名詞として辞書エントリに加える方法が考えられる。しかし、これはすべての名詞的に使用される可能性のある動詞連用形の調査には多大な手間がかかり、現実的ではない。また、辞書の保守の手間も増えるといった欠点もある。

そこで、現在の係り受け解析では、KNP [1] のようにルールベースで形態素解析結果を修正したり、CaboCha [2] のように教師あり学習手法であれば誤りを含んだコーパスから係り受け規則を学習することにより対応している。しかし、修正処理の不完全さやコーパスの網羅性などの理由により、係り受け解析の誤りが生じているのが現状である。

本稿では、動詞連用形用法の調査に基づき、形態素レベルでの動詞連用形の用法を決定付ける要素を明らかにする。そして、用法の曖昧性解消を行うシステムを試作し、その評価結果を示す。さらに、係り受け解析への影響についての調査結果も示す。

2 従来手法：KNP の品詞の修正

従来の係り受け解析の段階での動詞連用形の曖昧性解消の手法として、構文解析システム KNP の品詞の修正について述べる。

KNP では、大きく分けて、2つの修正処理がある。

1. 修正処理

動詞連用形を名詞(普通 or サ変名詞)に修正する処理。「直後が助詞(例外あり)」や「直前が名詞、直後も名詞」など、7つのルールがある。

2. 曖昧性解消処理

JUMAN の解析結果に動詞連用形と名詞の曖昧性がある場合に、どちらかに特定する処理。「直後が名詞なら名詞」や「直後が読点なら動詞」など、4つのルールがある。1.に比べて粗い。

この処理のルールは動詞連用形の曖昧性解消には完全であるとは言えない。また、この処理では、形態素解析結果に曖昧性がない場合には、動詞から名詞への修正しかない。そのため、形態素解析用の辞書に動詞連用形の動詞と名詞のエントリがない場合には、曖昧性の解消を行えない。

3 動詞連用形用法の曖昧性

3.1 動詞連用形の用法

今回対象とする動詞連用形の用法を例とともに以下に示す。

1. 動詞的用法

- a) 連用中止・連用修飾語
例) 首相の退陣を求め、野党のほぼ全議員が辞職した。
- b) 接尾辞・助動詞を伴うために、連用形となったもの
例) 内閣の命運をかけて取り組みたい。
- c) 複合動詞の一部
例) 国民は中選挙区に最も慣れ親しんでいる。

2. 名詞的用法

- d) 判定詞を伴って、述語になるもの(判定詞省略を含む)
例) 死者は数百人に達する見込みだ。
- e) 助詞を伴って、格要素などになるもの
例) 景気の落ち込みをどう見るか。
- f) 「サ変名詞+する」として使われるもの
例) 売却、もしくは間引きする。

g) 複合名詞の一部

例) 懸賞金付き定期預金を二十三日から発売する。

h) 並列した語

例) 政治からの、逃避、しらげ、の両極ばかりが混じり合うこの世相。

3.2 動詞連用形の調査

京大コーパス [3] の 6 日分、6,756 文を対象に、動詞連用形の用法と文脈 (直前や直後) との関係性を調査した。その結果を、直後については表 1 に、直前については表 2 に示す。

表 1 から、用法ごとに直後の語の品詞に大きな偏りがあることが分かる。直後の語が句点であれば、上記 d) の判定詞の省略の可能性が高いことや、動詞性・形容詞性接尾辞であれば、同 b) の連用形変化であろうことや、読点であれば、同 a) の連用中止や、同 h) の並列が疑われることがわかる。

表 1: 動詞連用形の直後の品詞

品詞	動詞的用法	名詞的用法
句点	0	63 (5.3%)
読点	1214 (59.9%)	21 (1.8%)
助詞	89 (4.4%)	825 (68.8%)
名詞	131 (6.5%)	145 (12.1%)
動詞性接尾辞	138 (6.8%)	0
形容詞性接尾辞	287 (14.2%)	0
動詞	89 (4.4%)	25 (2.1%)
判定詞	0	54 (4.5%)
括弧始	34 (1.7%)	0
名詞性接尾辞	12 (0.6%)	23 (1.9%)
その他	33 (1.6%)	44 (3.7%)
合計	2027 (100.0%)	1200 (100.0%)

次に、表 2 から、直前の語にも偏りが見られる。特に助詞が顕著で、接続助詞は圧倒的に名詞的用法、格助詞は主に動詞的用法ということがわかる。

表 2: 動詞連用形の直前の品詞 (語)

品詞	動詞的用法	名詞的用法
格助詞	1642 (81.0%)	188 (15.7%)
「を」	690 (34.0%)	20 (1.7%)
「に」	445 (22.0%)	46 (3.8%)
「が」	226 (11.1%)	35 (2.9%)
「と」	164 (8.1%)	31 (2.6%)
「で」	71 (3.5%)	16 (1.3%)
「から」	20 (1.0%)	6 (0.5%)
副助詞	135 (6.7%)	66 (5.5%)
「も」	79 (3.9%)	11 (0.9%)
「は」	55 (2.7%)	53 (4.4%)
接続助詞	7 (0.3%)	371 (30.9%)
「の」	1 (0.0%)	351 (29.3%)
「や」	0	15 (1.3%)
動詞	42 (2.1%)	153 (12.8%)
形容詞	43 (2.1%)	72 (6.0%)
名詞	16 (0.8%)	16 (1.3%)
読点	31 (1.5%)	140 (11.7%)
<文頭>	7 (0.3%)	108 (9.0%)
合計	2027 (100.0%)	1200 (100.0%)

最後に、動詞連用形の用法別の出現数を表 3 に示す。この表から、動詞によっては用法がほぼ固定されているということがわかる。

曖昧性の解消の決め手となるものはどのようなもの

表 3: 動詞連用形の用法別出現数

単語	動詞的	名詞的	単語	動詞的	名詞的
し	496	0	続き	25	2
あり	81	1	述べ	23	0
なり	78	1	思い	10	22
調べ	3	59	違い	4	21
対し	49	0	戦い	6	20
動き	1	33	より	19	1
見通し	0	33	はじめ	17	7
考え	8	31	見直し	2	17
さきがけ	0	28	見込み	0	17
疑い	0	26	受け	15	5

があり、それはどのくらいをカバーできるのかをまとめた結果を表 4 に示す。「その他」とは主に形態素レベルでは曖昧性の解消のできないものである。

表 4: 曖昧性解消の決め手

決め手	頻度 (%)
語自身	660 (20.4%)
直後の語	1364 (42.3%)
直前の語	39 (1.2%)
直前・直後の組み合わせ	734 (22.7%)
その他	429 (13.3%)
合計	3227 (100.0%)

この表から、単語自身とその前後 1 語を参照するだけでも、ほとんどの用法を判別することができるということがわかる。

4 曖昧性解消システム

3 節の調査結果に基づき、動詞連用形用法の曖昧性の解消を行うシステムを試作した。

曖昧性解消システムは次の 3 ステップからなり、入力形態素解析結果で、出力は修正された形態素解析結果である。

ステップ 1 対象の特定

曖昧性解消の対象である、動詞連用形やその同形の名詞を特定する。

ステップ 2 曖昧性の解消

特定した対象の用法の曖昧性を解消する。

ステップ 3 後処理

曖昧性を解消した結果、修正が必要な形態素を修正する。

4.1 ~ 4.3 節にて、各ステップの処理内容について詳しく述べる。

4.1 対象の特定

曖昧性解消の対象は、形態素解析結果で動詞連用形のものやその同形の名詞である。

そこで、形態素解析結果が動詞のものは、活用形が連用形のもを対象とする。形態素解析結果が名詞のものは、形態素解析辞書から動詞エントリを抜き出し、それを連用形に活用させたものと表記が同じものを対象とする。ただし、形式名詞、副詞的名詞、代名詞は対象外とした。

4.2 曖昧性の解消

判別ルールによって、曖昧性の解消を行う。判別ルールとしては、次のものを用意した。なお、調査したコーパスからの影響を抑えるために、単語自身からの判定ルールを極力無くすようにした。

1. 語自身による判別 (1つ)

- 語自身が「し(する)」「なり(なる)」「あり(ある)」なら動詞的用法

2. 直後の語による判別 (6つ)

- 直後が動詞性、形容詞性述語、名詞性述語接尾辞なら動詞的用法
- 直後が句点、判定詞なら名詞的用法
- 直後が助詞「つつ」「ながら」「さえ」なら動詞的用法
- 直後が上記3つ以外の助詞なら名詞的用法
- 直後が「する」「できる」などのサ変名詞接続語なら動詞的用法
- 直後が動詞なら動詞的用法

3. 直前の語による判別 (5つ)

- 直前が接続助詞(ただしその前が動詞や用言系の接尾辞ではない)なら、名詞的用法
- 直前が格助詞「の」なら、名詞的用法
- 直前が名詞接頭辞なら、名詞的用法
- 直前が指示詞「あの」「その」「この」「どの」なら、名詞的用法
- 直前がサ変名詞で、語自身が「できる」「いたす」などのサ変名詞接続語なら名詞的用法

4. 直前・直後の組み合わせによる判別 (4つ)

- 直後が読点、括弧始で、
 - 直前が格助詞、副助詞、副詞、接続助詞「に」なら、動詞的用法
 - 形容詞の連用形、格(副)助詞+動詞の連用形なら、動詞的用法
 - 直前が形容詞の基本形・連体形なら名詞的用法
- 直後が読点、直前も読点で、その前が動詞や用言系の接尾辞でないなら、名詞的用法

これらのルールに従って、動詞連用形の用法の曖昧性を解消する。なお、ルールに当てはまらないものは、主に形態素レベルでは曖昧性を解消できないものであり、修正する必要がないものとして扱う。

4.3 後処理

動詞を名詞に修正する場合には、『直後が「する」「できる」などのサ変名詞接続語』のルールを適用したものはサ変名詞に、それ以外のは普通名詞に修正する。名詞を動詞に修正する場合には、読みを参考にして活用型を決定し、活用型を参照して原形を決定し、それぞれ修正し、活用形は連用形に修正する。

5 実験と検討

4節で述べた機械処理について、1) 曖昧性解消と2) 係り受け解析への影響、の2種類の実験を行った。実験対象とする文は、京大コーパスから抽出した、1,264文。形態素解析には JUMAN (Ver. 3.61)[4] を、係り受け解析には KNP (ver. 2.0b6) を使用した。

5.1 曖昧性解消の実験

文を JUMAN で形態素解析をし、対象を特定した上で曖昧性解消を行い、正解と比較した。

実験結果を表5に示す。ここで、「ベースライン」とは形態素解析結果そのままであり、「ルールのみ」とは判別ルールによる曖昧性解消結果、「ルール+そのまま」とはルールに当てはまらないものはそのまま出力するもので、今回のシステムの出力である。

判別ルールによる曖昧性解消により、変化が顕著に見られた名詞的用法で、14.8 point の精度向上が見られ、全体でも 5.6 point 精度向上し、98.2% となった。また、「ルールのみ」の場合には、失敗が1例しかなく、判別ルールの判定はほとんど誤らないことが確認できた。

表 5: 曖昧性解消実験

	動詞的用法			名詞的用法		
	正解	失敗	合計	正解	失敗	合計
ベースライン	385	3	388	180	42	222
	99.2%	0.8%		81.1%	18.9%	
ルールのみ	344	1	345	196	0	196
	99.7%	0.3%		100%	0.0%	
ルール+そのまま	386	2	388	213	9	222
	99.5%	0.5%		95.9%	4.1%	

	全体		
	正解	失敗	合計
ベースライン	565	45	610
	92.6%	7.4%	
ルールのみ	540	1	541
	99.8%	0.2%	
ルール+そのまま	599	11	610
	98.2%	1.8%	

次の文は、失敗した例である。

- 結局、ベビー・ブーマー・ジェネレーションは経済発展に尽くしただけで祈りと冥想、つまり、神と語ることに自らと語ることを忘れた。

これは、動詞連用形が副詞的に使われたものであり、このような動詞連用形の副詞(や接続詞)への曖昧性は今回は考慮していない。しかし、今後は対処が必要であると考えている。

5.2 係り受け解析への影響

次に、形態素解析結果を修正したことで、係り受け解析にどのような影響が出るかを調べるための実験を行った。実験として、通常形態素解析結果と、5.1 節で述べた実験の結果（形態素解析結果の修正）を入力として、KNP で係り受け解析を行い、京大コーパスの正解との比較を行った。

実験結果を表 6 に示す。この結果、動詞連用形用法の曖昧性解消を実施することにより、0.12 point の精度向上が見られた。

表 6: 係り受け解析実験

	JUMAN+KNP	+曖昧性解消
係り受け正解	9559 (84.60%)	9571 (84.72%)
係り受け失敗	1740 (15.40%)	1726 (15.28%)

曖昧性解消を行ったことで、係り受け関係に変化が生じたものを述べる。ここで、「～ヶ所」というのは、係り受け関係に変化のあった動詞連用形の数であり、それに伴う係り受けの数ではないため、表 6 とは数が異なる。

- 正しくなったもの (5ヶ所)
複合名詞化した動詞連用形や、「米国へ買い出しに行く」など「～に行く」、連用中止などがある。図 1 は動詞連用形の曖昧性解消により、対象部分以外の係り受けが改善された例である。この図では、左が本手法適用前、右が適用後である。
- 誤ったもの (3ヶ所)
曖昧性解消で誤った「つまり」による誤り、並列の係り先の違いなどがある。
- いずれも誤り (3ヶ所)
JUMAN の解析誤りなどがある。

アンヘラさんも 昨年まで 週末には、 米国へ 買い出しに 行く 長蛇の 行列に 並んでいた。	アンヘラさんも 昨年まで 週末には、 米国へ 買い出しに 行く 長蛇の 行列に 並んでいた。
--	--

図 1: 曖昧性解消が係り受け解析で有効に働いた例

6 考察・課題

形態素情報を手がかりとした、動詞連用形用法の曖昧性解消は、本稿により多くの部分は実現できたと考えている。しかし、10% 強残った、曖昧性の解消できなかった部分（表 4 の「その他」）の検証は十分にできていない。ここの曖昧性が、本稿であまり扱わなかった語自身の制約により解消できるのか、直前・直後の

語以上を参考にして解消できるのか、もしくはやはり形態素レベルでは解消できないのかを十分に検証する必要がある。

曖昧性解消システムでは、ステップ 1、対象の特定手法において、形態素解析辞書の動詞エントリからの同形名詞探索手法には不足を感じている。解決策として、名詞の語尾や周辺の語の情報を利用したり、機械学習による特定手法を採用するなどを考えている。

次に、ステップ 2、曖昧性の解消においては、今回考慮していなかった、動詞連用形と副詞・接続詞との曖昧性解消を考慮する必要がある。

また、今回の知見をもとにした機械学習による曖昧性解消手法との比較も試してみたい。

最後に、本手法の有効性の検証としては、JUMAN + KNP 以外の形態素解析・係り受け解析への影響等を調べる必要があると考えている。

7 まとめ

本稿では、動詞連用形の用法について、京大コーパスの 6,756 文、3,327 ヶ所の調査を行い、形態素レベルでの用法を決定付ける要素とその比率を明らかにした。

その調査に基づき、ルールによる曖昧性解消システムを作成した。その方法の特徴は、2 つある。1 つは、形態素解析用の辞書のエントリの有無に関わらず、動詞連用形とその同形の名詞の曖昧性の解消を行えることである。2 つ目は、形態素解析用の辞書を充実させるような、大きな手間がかからないことである。

そして、実験を行い、曖昧性解消では 5.6 point の精度向上が見られ、有用性を示した。また、係り受け解析の実験も行い、0.12 point の精度向上への影響が確認でき、KNP より高い性能で曖昧性解消できることがわかった。

この動詞連用形用法の曖昧性解消は、係り受け解析だけでなく、キーワード抽出、複合語抽出、専門用語抽出などへの効果も大きいと思われる。

今後は今回対処できなかった問題について取り組んでいく必要がある。

参考文献

- [1] 黒橋禎夫. 日本語構文解析システム KNP version 2.0 b6 使用説明書. 京都大学大学院 情報学研究科 (1998)
- [2] 工藤拓, 松本裕治. チャンキングの段階適用による日本語係り受け解析, 情報処理学会論文誌, Vol.43, No.6, pp. 1834-1842 (2002)
- [3] 黒橋禎夫, 長尾真. 京都大学テキストコーパス・プロジェクト. 言語処理学会 第 3 回年次大会, pp. 115-118 (1997)
- [4] 黒橋禎夫, 長尾真. 日本語形態素解析システム JUMAN version 3.61 使用説明書. 京都大学大学院 情報学研究科 (1998)