

呼応関係を産み出す構文手がかり

木田敦子^{1,2} 山本英子² 神崎享子² 井佐原均^{1,2}

{kida, eiko, kanzaki, isahara}@crl.go.jp

1 通信・放送機構 2 通信総合研究所

1 はじめに

共起関係が二つの語が同一文内に出現する関係を指すのに対し、呼応関係は呼要素 α が出現したら応要素 β が出現するというある種の拘束関係を指す。述語が文末に現われる日本語の場合、文の終末近くまで文内容が確定しない。これに対し、古語には係助詞と文末の活用形とが形態的な呼応関係を持つ係り結びの用法があり、係助詞が後に出現する要素の予告を可能にしていた。係り結びが消滅した現代語においても、呼応関係は存在する。直観でもわかる「しか～ない」「決して～ない」などが挙げられる。

このように後ろに出現する要素を予告する性質は、構文解析の曖昧性解消や係り受け関係の決定などに有効と考えられる。そこで我々は、大規模な電子化コーパスから自動的に呼応関係を抽出して呼応関係テーブルを作成することを試みている[木田・山本・井佐原 02]。本稿では[木田・山本・井佐原 02]を承けて、呼応関係抽出のために行った分析について報告する。

2 先行研究と本研究の位置付け

2.1 先行研究

[大野 93]は、係り結びが消滅した現代語においても、「時間的に線状的に発展し連続していく言語表現の早い部分で、一文の行く手、肯定か否定か疑問かなどを予告しておこうとする」表現として、古語の係助詞に代わる「ある種の副詞」が存在することを示唆している。

[益岡 91]は[大野 93]とは別の角度からアプローチを行い、文要素の呼応関係について論じている。研究史的に陳述論の流れを汲む益岡氏のモダリティ論では、文を階層構造と呼応関係を持つものと捉え(表 1)のような呼応関係を挙げている。

[益岡 91]では、呼応関係にある呼要素と応要素の語を同一

カテゴリ内にまとめて入れており、個々の語がそれぞれどの語と呼応関係にあるのかは記述されていない。また、挙げられている要素の数が少ない、挙げられている呼要素と応要素の妥当性が検証されていない、などの弱点がある。

2.2 本研究の目標

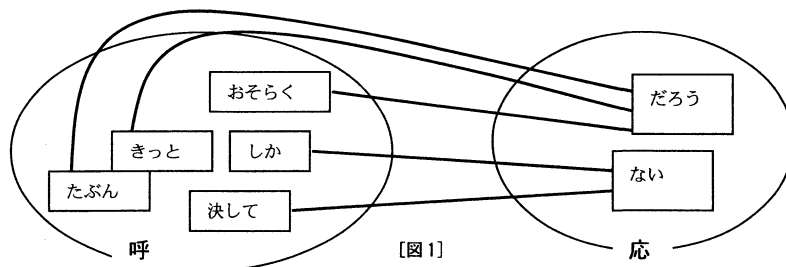
本研究において我々は、大規模な電子化コーパスから自動的に呼応関係を抽出することで、1) 直観によらず客観的かつ実用にたえる規模の呼応関係データを作成する、2) どの呼要素がどの応要素と呼応関係にあるかの対応関係を明らかにする、3) 抽出したそれぞれの呼要素と応要素について関係の強さを明らかにする、4) 抽出した呼要素と応要素の妥当性の検証することを目標としている。

3 主な論点と方法

3.1 本稿の論点と方法

[山本・梅村 02]では、出現パターンの包含関係に強い補完類似度を用い、コーパスから事象間の一対多関係の抽出実験を行っている。本研究で扱っている呼応関係も、応要素が呼要素を包含する一対多関係と見ることができる(図 2)。[木田・山本・井佐原 02]では、いわゆる係助詞・副助詞、陳述副詞を調査対象語とし、補完類似度を用いて調査対象語の「応」となる要素を探す手続きについて報告した。

本稿は[木田・山本・井佐原 02]を承けて、補完類似度を用いた呼応関係の抽出のための分析を行う。本稿の主な論点は、1) 抽出方法において[木田・山本・井佐原 02]を改善した点の報告、2) 呼応関係か否かの判定の手がかりとなる構文の分析結果の報告とする。



3.2 調査対象

本稿では、『基礎日本語文法』で「提題助詞」「取り立て助詞」「陳述の副詞」に分類されている以下の語を調査対象語とし、これを呼応関係の呼要素と仮定して分析を進める。調査データには、毎日新聞記事データ 10 年分(1991 年～2000 年)、日経新聞記事データ 10 年分(1991 年～2000 年)、読売新聞記事データ 10 年分(1991 年～2000 年)を使用した。

〔調査対象語〕

「こそ」「しか」「さえ」「は」「も」「ばかり」「のみ」「すら」「なら」「くらい(ぐらい)」「だけ」「なんて」「決して」「おそらく(恐らく)」「たぶん(多分)」「ぜひ(是非)」「まるで」「もし」「きっと」

4 補完類似度を用いた呼応関係の抽出

処理の流れを(図 2)に示す。

- ① 茶筌で形態素解析した新聞記事データに対し、補完類似度で類似度計算を行う
- ② 類似度計算の結果から調査対象語を含むペアを抽出する
→ 共起ペア
- ③ 共起ペアが呼要素、応要素の順で出現しているものを抽出する
→ 呼応候補ペア 1
- ④ 信頼度(呼応候補ペア 1 の数/呼要素の数)が 0.05 以上の呼応候補を抽出
→ 呼応候補ペア 2
- ⑤ 構文手がかりによる判定
→ 呼応関係データ

(図 2)の処理の流れのうち、①から③の過程は[木田・山本・井佐原 02]で報告した。これに対し、④信頼度による判定、⑤構文手がかりによる判定、の過程は今回新たに導入したものである。

5 信頼度による判定

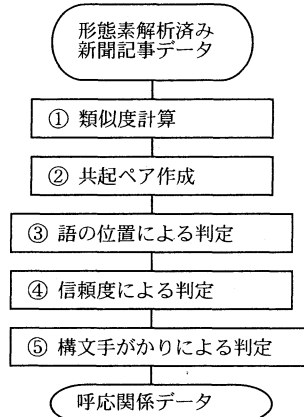
③の出現順序に基づく判定の後で得られる「呼応候補ペア 1」は膨大な数になる。「決して」を呼要素とするペアは「決して～ない(助動詞/特殊・ナイ)」「決して～だ(助動詞/特殊・夕)」など 4039 組、「さえ」を呼要素とするペアは「さえ～ば(助詞-接続助詞)」「さえ～ない(助動詞/特殊・ナイ)」「さえ～ある(動詞-自立/五段・ラ行)」など 3712 組、「は」を呼要素とするペアは 18590 組にもなる。

これらのペアの中には「きっと～出好き(名詞-形容動詞語幹)」「きっと～わらわせる(動詞-自立/一段)」「きっと～ながれつく(動詞-自立/五段・カ行イ音便)」のような除外すべきものが多く含まれる。そこで、信頼度(呼応候補ペア 1 の数/呼要素の数)が 0.04 以上になるペアを抽出した。

信頼度の基準を 0.04 に基準に定めた理由は以下の通りである。信頼度を 0.05 以上と定めると、「まるで～みたい(名詞-非自立-形容動詞語幹)」「ぜひ～みる(動詞-非自立/一段)」「ぜひ～欲しい(形容詞-非自立/形容詞・イ段)」「ぜひ～下さる(動詞-非自立/五段・ラ行特殊)」「きっと～から(助詞-接続助詞)」「きっと～ね(助詞-終助詞)」「きっと～で(助詞-接続助詞)」「おそらく～ようだ(助動詞/ナ形容詞)」

[表 1: [益岡 91]の呼応関係]

「呼」要素	「応」要素
ねえ、おい	ね、よ
ぜひ、なんて	て下さい、なあ
たぶん、どうも、	だろう、らしい、
いったい	ようだ、か
むかし、かつて、もうすぐ	た
決して、必ずしも、	ない



[図 2: 処理の流れ]

「こそ～たい(助動詞/特殊・タイ)」などの呼応関係として着目する可能性を残すべきペアをはねてしまう。一方で、0.04 以上まで値を下げると、「おそらく～初めて(副詞)」「は～者(名詞-接尾-一般)」「は～的(名詞-接尾-形容動詞語幹)」などの除外すべきものも拾ってしまう。だが、ここでは着目の可能性を残すべきペアをはねるよりは、不要なものを拾って次の段階ではねることを考えるのが妥当と判断し、0.04 以上に値を定めた。

信頼度 0.04 以上を基準に呼応候補ペアを抽出すると、4039 組あった「決して」を呼要素とするペアは 14 組に、3712 組あった「さえ」のペアは 16 組に、18590 組あった「は」のペアは 23 組になった。この方法で調査語ごとのペアは概ね 20 組前後に落ち着いた。

6 呼応関係判定のための構文手がかり

6.1 前接、後接要素から

同じ応要素でも、調査語によって呼応と判断できるものとできないものがある。以下の例は、応要素が「ない」の例である。「さえ」「しか」は「ない」と呼応していると判断できるが、「だけ」「くらい」は「ない」と呼応しているとは判断できない。

「さえ」「しか」

(1) 職場を早く離れたいという気持ちから、退社時、置き忘れた傘を取りに戻ることにさえできない。

(2) 債権額のわずかに五%、平均五千円程度しかもどってこないことが判明。

「だけ」「くらい」

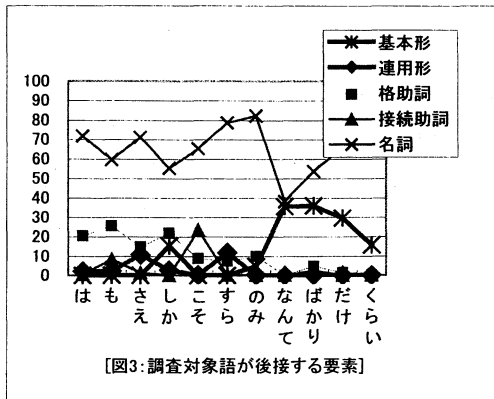
- (3) 野球の投手の動きを観察すると、手だけで投げているのではない。
- (4) 物価上昇を一・五%程度として実質三・五%くらいでないと予算が組みづらいという事情だ。

原因は助詞の種類の違いにあると考えられる。いわゆる取り立て助詞は、前後両項を結びつける「係助詞」と体言を形成する「副助詞」に分けることもできる[木田 98]。そして、呼応に関係するのは「係助詞」である。

係助詞の働きの一つに「行きはしない」「見もしない」のように述語に割って入る二分結合がある。この場合、係助詞は連用形に後接する。

ノヤナを伴って連体修飾するのは体言の性質である[星野・丸山 93]。また、体言の構文上の働きの一つに助動詞ダを伴って文の述語になることが挙げられる[橋本・桑畑 96]。このことから、「解散するくらいの英断」「眠るだけのスペース」「前進あるのみのアメフト」「待つばかりの状態」、「国民の不満は高まるばかりだ」「やさしく思えるくらいだ」それぞれの下線部は体言相当語句と判断できる。そして、この場合、茶筌の形態素解析結果では、「解散する」「眠る」などは基本形となる。

そこで係助詞か副助詞かを判断する基準として、連用形に後接する傾向、基本形に後接する傾向を観察した。(図 3)は、調査対象語が後接する要素の種類と該当数の割合である。「は」「も」「さえ」「しか」「すら」は連用形に後接する傾向が見られる。「しか」「のみ」「なんて」「ばかり」「だけ」「くらい」は基本形に後接する傾向が見られるが、「しか」だけは基本形に後接していても「春を待つしか出口がない」「実績で証明していくしか なさそうだ」のように「基本形+しか」で体言を形成しておらず他の語とは質を異にする。



【図3: 調査対象語が後接する要素】

調査対象語が後接する要素の傾向から、「のみ」「なんて」「ばかり」「だけ」「くらい」を副助詞、「は」「も」「さえ」「しか」「こそ」「すら」を呼応に関係する係助詞とみなす。

1 「こそ」は非常に強い力で体言を形成する性質を持つがそ

6.2 呼応ペアの詳細検討 — 「きっと」を例に
一次に構文手がかりによる判定のための分析を行った。本節では、呼要素「きっと」に対する応要素の候補を類似度順に並べたものの上位 20 件に対して行った検討の手順を示す。

6.2.1 応要素候補の分類

応要素の候補は、(表 2)のように分類できる。

- 単独で応要素候補が応要素になるもの
- 複数の応要素候補の組み合わせが応要素になるもの
- 応要素にならない応要素候補

呼要素「きっと」に対する応要素上位 20 件には、「応要素にならない応要素候補」は見られなかった。

6.2.1.1 単独で応要素となるもの 実例とその分類

「はず(名詞-非自立一般)」「よ(助詞-終助詞)」は単独で応要素になることができる。

- (5) 広大な海辺で見る作品はきっと魅力的に映るはずだ。
- (6) きっとウオークに夢中になりますよ。
- (7) 自信を持ってやれば、きっと自分の最高の演技が出来るよ。

6.2.1.2 複数候補の組み合わせの応要素 (1) 実例とその分類

複合 1	複合 2
きっとーてくれる	きっとーだろう
きっとーでしょう	きっとーと思う
きっとーに違いない	きっとーのではないかと
	きっとーのだろう
	きっとーからだ
	きっとーからです
	きっとーんだ
	きっとーことだろう
	きっとーことでしょう

・「複合 1」のグループ

- (8) 後輩たちがきっと僕たちの分も勝つてくれます。
- (9) きっと緊張するでしょうね。
- (10) かみしめれば、きっとサケの味がするに違いない。

・「複合 2」のグループ

- (11) きっと大観衆の前でプレーするプロ野球やサッカー選手のような快感を味わっているのだろう。
- (12) みんなもきっとあっちゃんのことが好きだと思います。
- (13) きっと別の道をたどったのではないかと。
- (14) きっと照れ臭くて言わなかったのだろう。
- (15) きっと成功すると確信したからだ。
- (16) きっと将来、役に立つと思うからです。
- (17) きっと育たないから、神様が悲しい思いをさせないようにしてくれたんだ。
- (18) きっと猛反発することだろう。

れはやや特殊なものであり、他の要因から「こそ」は係助詞であると判断できる[木田 98]。

2) 文末表現の階層構造から見た応要素の連続

文末表現が階層構造を成すという性質[木田・乾・落谷・西野 00]から、「複合2」のグループの応要素は「複合1」のグループの応要素の後に続くことができる。

- ・複合1「てくれる」+複合2「と思う」
- [19] きっと、いろんな人が応援してくれると思う。
- ・複合1「てくれる」+複合2「だろう」
- [20] 作曲者の思い描いた古都を、きっと陰影深く再現してくれるだろう。

6.2.2 呼要素と応要素候補以外の要因

呼要素と応要素候補以外の要因の影響があると考えられるものもいくつか見られる。

ここでは、「きっとーます(助動詞/特殊・マス)」に着目して考察を進める。この文では、「きっとーます」よりも前に、条件節、仮定節、理由節が見られる文が多く存在する。

- ・条件節+「きっとーます」
- [21] がんばっていたら、きっといいことがあります。
- [22] 日本でもNIE運動を進めればきっと将来役立ちます。
- ・仮定節+「きっとーます」
- [23] 人間も動物のようにも少しシンプルになれば、きっと優しい眠になるかも知れないと思います。
- ・理由節+「きっとーます」
- [24] 最後まできちんとやり遂げる子なので、きっと病気に打ち勝つてくれると信じています。

本稿では、呼応関係を呼要素αが出現したら応要素βが出現するというある種の拘束関係とみなし、応要素を導く要因を呼要素のみと仮定しているが、呼要素以外にも応要素を導く構文的要因があることも考えられる。また、呼要素と応要素を含む呼応関係そのものを導くような構文的要因があるのかもしれない。この解明は今後の課題としていきたい。

7 おわりに

本稿では、1) 補完類似度を用いた呼応関係抽出の手順主な論点は、2) 呼応関係が否かの判定に使う手がかりを得るために行った分析結果について報告した。

今後は、分析によって得た構文手がかりを処理にのせることに取り組み、呼応関係データの自動作成を目指す。

謝辞 本稿を纏めるにあたり、毎日新聞社、読売新聞社、日経新聞社の新聞記事データを使用させて頂きました。また、データ加工やツール作成などの面でご助力、ご指導を下さいました通信総合研究所の内山将夫氏に心より感謝致します。

参考文献

大野晋：係り結びの研究。岩波書店 (1993)。
 木田敦子，山本英子，井佐原均：後続要素を予告する表現の分析，情処研報，NL152-20，pp. 137-143 (2002)。
 木田敦子：係助詞と副助詞，東京女子大学日本文学 vol. 89，pp. 69-84 (1998)。
 木田敦子，乾裕子，落谷亮，西野文人：情報抽出のための文末表現分析，言語処理学会第6年次大会発表論文集，pp. 304-307 (2000)。
 橋本三奈子・桑畑和佳子：第一章 名詞辞書の概要，計算機用日本語名詞辞書 IPAL (Basic Nouns) 一解説編一，情報処理振興事業協会 (1996)。
 星野和子，丸山直子編：日本語の表現，圭文社 (1993)。
 益岡隆志：モダリティの文法，くろしお出版 (1991)。
 益岡隆志，田窪行則：基礎日本語文法一改定版一，くろしお出版 (1992)。
 松本裕治，北内啓，山下達雄，平野善隆，松田寛，高岡一馬，浅原正幸：形態素解析システム『茶筌』version 2.2.9 使用説明書，奈良先端科学技術大学院大学 松本研究室 (2002)。
 山本英子，梅村恭司：コーパスの中の一対多関係を推定する問題における類似尺度，自然言語処理 Vol. 9 No. 2，pp. 45-75 (2002)。

[表2：応要素の候補]

	単体	応要素の組み合わせ											複合1	複合2			
		う	だ	と	て	思	で	な	ま	の	は	く			い		
う(助動詞)	×	-															
だ(助動詞)	×	○	-														だろう
と(引用)	×		-	○													と思う
て(接続助詞)	×			-						○							てくれる
思う(動詞)	×				-												
です(助動詞)	×	○			-												でしょう
ない(助動詞)	×					-											
ます(助動詞)	△						-										
の(名詞)	×		①			①		-						①			①のではないか ②のだろう
はず(名詞)	○		②	②					-								
くれる(動詞)	×									-							
いる(動詞)	×										-						
よ(終助詞)	○											-					
なる(動詞)	△												-				
から(接続助詞)	×		①														①からだ ②からです
ん(名詞)	×		○														んだ
こと(名詞)	×		①	①													①ことだろう ②ことでしょう
違い(名詞)	×																
か(終助詞)	×																
に(格助詞)	×								○					○		-	に違いない