

助詞「か」に関する計量的調査

吉橋健治 仁科喜久子

東京工業大学

knj4484@ryu.titech.ac.jp knishina@ryu.titech.ac.jp

1 はじめに

自然言語処理の研究の中で並列解析の重要性はたびたび指摘されており[1]、いくつかの解析手法が提案されてきた。従来の解析手法は基本的に文節列の類似性に基づくものである。一方、日下部ら[2]は「…も…も」のような助詞の呼応による並列があれば、それに従って終点を決めるという方法を取っている。並列構造は必ずしも類似性に基づいているとは言えないので日下部らの解析手法が有効であると考え、呼応する並列助詞について調査することにした。今回は並列助詞の中でも特に「か」が用言とともに用いられるものに焦点を当てている。

2 先行研究

本稿の研究対象である助詞「か」に言及している研究として以下のものがある。

首藤ら[3]は文形式の並列の型の出現頻度について約1万の日本語技術文を対象に調査している。その中で「S[止]か(,)Sか」という型が2.5%、「Sか(,)あるいは(,)S」という型が1.0%、「S[止]か(,) S[止]か(,)Sか」が1.0%の出現比率であることを報告している。従って、助詞「か」を含む並列表現全てを含めると4.5%の出現比率であることになる。

国立国語研究所[4]では、助詞について実例を集め意味分類している。そのなかで「用言+か」を含むパターンも多数取り上げられているが、それらに関する分析は行われていない。

3 調査

3.1 データ

構文解析器を利用する際、主な解析対象は新聞データなので、今回の調査では毎日新聞1年分を調査対象とした。

3.2 抽出

まず、JUMAN、KNPを用いて全ての文を解析した。KNPは用言と「か」が含まれている文節に<ID:~か>というタグを付与する。このタグをもとに、「用言+か」を含む文を抽出した。この中には「~するべきか」「~するばかりか」「~するののか」などのように用言と「か」が隣接していないものも含まれる。毎日新聞CD-ROMには重複した記事も含まれているので、同一の文が抽出された場合には片方を排除した。以上の手順により最終的に357文が抽出された。

3.3 分類

抽出された文を、「用言+か」の文節の係り先の文節の種類、および係り先との関係を考慮して分類した。分類作業を進めて「用言+か」には多数のパターンがあることが明らかになったが、構文解析への応用を考えたときに意味のある分類になるように、以下で説明する6タイプに分類し、当てはまらないものを「その他」とした。以下の表記の中でP1、P2は用言を表す。

タイプ1: 選択を意味する並列 [P1 か P2 か]
佐藤が王将奪取に王手をかけるか、羽生が勝負を振り出しに戻すかの一番だ

「P1するか、それともP2するか」と言い換えられるような文で、いくつか挙げた項目の

どれか一つを選択させるような意味の並列構造をタイプ1にした。また、その文自体が「P1か、それとも P2か」の形で現れている場合もタイプ1である。

タイプ2：選択を意味する並列〔P1かP2〕
いろいろな課題をあたえ、またいろいろな状態で被験者の頭の中に浮かぶイメージを語ってもらうか、書いてもらう。

タイプ1と同様の言いかえが可能な並列だが、最後の並列項目には「か」が付かないものをタイプ2とする。

タイプ3：用言の格要素

なぜあんなことが起きたのか、理解できなかった

「～か」の文節が用言の格要素であり、格助詞が省略されているものをタイプ3とする。この例では「～か」が「理解できる」のガ格になっている。

タイプ4：同格名詞節

このままでは10年後の自分たちの食べ物はどうなっているのか不安を覚える

「～か」の統語的な性質は、通常は用言に係ると考えるべきだが、これらの文では単純に用言にかかるとは言いがたい（上記の例で「～どうなっているのか→覚える」という係り受けでは意味を成さない）。むしろ、直後の名詞と内容的に結びついていると考えた方が自然である（「食べ物がどうなっているのか」が「不安」の具体的な内容である）。このように「～か」が、後ろの名詞と同格であるような名詞句と考えられるものをタイプ4とする。

タイプ5：二つの疑問文の並列

学習にとっては、いかにして発見するか、いかに楽しくやるかということが、ひじょうに大切なことなのです。

この文では「いかにして発見するか」と「いかに楽しくやるか」という二つの部分が並列になっている。しかし、タイプ1やタイプ2のように、どちらかを選択させる意味を表し

ているわけではなく、二つの疑問文をいわゆる「並べの並列」にしたものである。このような並列構造の場合をタイプ5とする。

タイプ6：不確実な理由を表す従属節

親に言いつけられたのか、同級生は近寄らずいつも独りぼっちだった

「～か」の部分が従属節を作り、不確実な理由を表している。例では「親に言いつけられた」と「一人ぼっちだった」が因果関係を持つが、断定を避けた表現である。

その他

上記の6つのタイプに分類できなかったものを例を挙げていくつか説明する。

バーチャルでも学生が授業についてこられるようにするにはどうしたらいいか、教材づくりを考えてきた

この文は、一見すると「～か」は「考えてきた」に係ると考えられる。しかし、その場合は「～か」が「考える」のヲ格であると同時に「教材づくりを」が「考える」のヲ格として存在することになってしまい、矛盾が生じる。従って、この文は「～か」の係り先が存在しない非文法的な文である。

グラビアのような感じと云えばいいでしょうか、動きのある写真ばかりです

この文も「～か」の係り先は存在しない。本来は、「～か」の後ろの読点を句点にするべき文である。

ただ、優等生的というか常識的な答えが多かったとも感じた

この文のように「PかP(か)」ではなく「PかN」の形になるものは、今回の調査ではその他にした。しかし、「というか」のような特別な場合には複合助詞を介した名詞の並列と考えた方が良いと思われるものもある。

4 分類結果と分析

分類結果の内訳を表1に示す。タイプ3が圧倒的に多く45%を占める。また、並列構造のもの(タイプ1、2、5)の合計は全体の17%になる。

表1 各タイプの出現数

タイプ	出現数
1 選択を意味する並列[P1かP2か]	34
2 選択を意味する並列[P1かP2]	16
3 用言の格要素	159
4 同格表現	36
5 二つの疑問文の並列	14
6 不確実な理由を表す従属節	22

各タイプの詳細を調べてみると以下のようなことが分かった。

【選択を意味する並列で「か」が省略される条件】

タイプ2は構造的にはタイプ1と同じで最後の並列文節から「か」を省略したものだが、どのような場合がタイプ2として出現したか例を挙げて説明する（数字は出現数）。

連体修飾節(5)

企業会計のように資産と負債を記すバランスシート（貸借対照表）を作成したか、検討している市町村が昨年8月現在、2883団体と、全体の89%に達したことが19日、総務省のまとめで分かった

「作成したか、～検討しているかする市町村」のように連体修飾する文節が「か」を含むパターンも理論上は可能だが、今回の調査では一例もなかった。

条件節(3)

森内が藤井猛九段（31）に勝つか、佐藤が森下卓八段（35）に敗れると、森内の挑戦が決定

「～勝つか、～敗れるかすると」というように条件節の文節に「か」が含まれるパターンも理論上は可能だが、今回の調査では一例もなかった。

【「～か」を格要素に取る用言】

タイプ3の用言について、表2のように類義表現ごとにまとめ出現回数を数えた（出現回数少ないものは省略してある）。まとめた系統ごとの出現回数を円グラフにして図1に示す（系統の中で代表的な語を凡例にした）。こ

表2 タイプ3の用言の分類とその出現数

語	出現数
分かる,わかる,おわかり	62
考える,考えさせる	9
注目する,注目される,注目だ	8
検討する	7
調べる,調査	6
明らかだ,不明,不透明,はっきりする	6
見る,見守る,見定める	5
確かめる,確認する	5

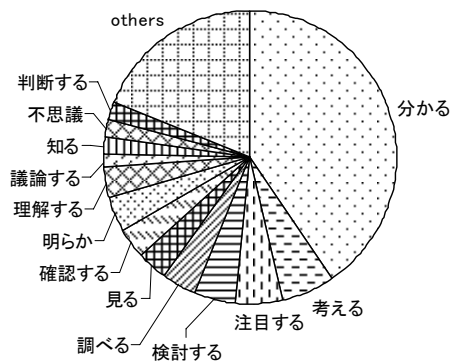


図1 タイプ3の用言の系統ごとの頻度分布

これらの図表から、「～か」を格要素としてとることができる用言には様々なものがあることが分かる。また、「分かる」「考える」「注目する」の上位3項目で約50%を占めており、その分布にはある傾向があることが分かる。

【「～か」による同格表現をとる名詞】

タイプ4に出現した名詞は、タイプ3との関連性から二つに大別できる（表3）。

表3 タイプ4に出現した名詞

a	サ変名詞 形容動词语幹	不安 調査 検討 考慮 確認 説明 結論 下見 解釈 議論 メモ 予想
b	(a)以外	予断 自信 話題 理念 方程式 答え 真実 真相 声

a の名詞群のほとんどはタイプ3の用言と対応している。例えば、タイプ3の「～か不安だ」に対して、タイプ4に「～か不安を覚える」「～か不安が残る」があり、「～か調査す

る」に対して「～か調査を始める」がある。このように対比してみると、タイプ3の表現を基本として、その用言を名詞化して、別の用言をつけたしたものがタイプ4であると考えられることもできる。一方、bの名詞群は対応するものがタイプ3にはない。

5 考察

今回の調査結果の並列構造解析への応用について考察する。

1) 「用言+か」が複数ある文
これは、タイプ1またはタイプ5に当たるが、どちらも構造的には並列構造である。次の例文
彼は、忙しかったのか、子供がちゃんと学校に行っているか確かめなかった。
のように、「用言+か」が二つ現れ並列構造でないものも文法的には成立するはずであるが、今回の調査を通して、その例は存在しなかった。したがって、日下部らの解析手法のように助詞「か」が二つ現れたときには、無条件にその二つを並列構造にしてしまうヒューリスティックは有効であると考えられる。

2) 助詞「か」が一つだけの文

助詞「か」が一つしか現れない文(タイプ2、3、4、5)の解析方法は以下のようなものが考えられる。「用言+か」の後ろに「分かる」「調べる」などのタイプ3の形をとる用言が現れた場合は、並列になる可能性は低いので、通常に係り受け関係としてその用言に係るものとする事ができる。ただし、この処理を実現するためには、タイプ3になる動詞のリストをあらかじめ用意しておく必要がある。タイプ4も同様に名詞のリストを用意しておくことで識別できる。

残るパターン2、5のなかで、今回の調査では「用言+か」が「～のか」「～してか」の形のものパターン5の大部分を占めたが、タイプ2には存在しなかった。逆に、単純な「～か」はパターン2に支配的に現れている。従って、「～のか」「～してか」の形で現れる場合はタイプ5で、「～か」の場合はタイプ2であるとするのがよいであろう。た

だし、タイプ5の場合、係り先を特定することは難しい。

タイプ2は並列構造であるが、これを正しく解析するための手がかりは、前節で述べたような後ろ側の文節の特徴が有効であろう。また、今回の調査対象中には「～するか」と「～しかない」が現れる文がタイプ2として現れたが、この二つの文節が並列構造になる可能性は十分に高い。

6 今後の課題

今回の調査は毎日新聞1年分だけなので、調査結果に毎日新聞に固有な文体や語彙などの特徴が現れている可能性がある。また、出現数の少ないタイプについての分析は信憑性が低い。したがって、対象データを他社の新聞コーパスや新聞以外のデータまで広げて調査する必要がある。また、今回の調査は助詞「か」についての調査だが、「か」と同様に用言の並列を作る助詞「たり」についても調査する。そして、それらの調査結果をもとに、並列解析の手法を実装し、解析精度が上がることを検証することとする。

謝辞

本研究は21世紀COEプログラム「大規模知識資源の体系化と活用基盤構築」の支援を受けている。

参考文献

- [1] 黒橋,長尾: 長い日本語文における並列構造の推定, 情報処理学会論文誌, Vol.33, No.8, pp.1022-1031 (1992)
- [2] 日下部,田中,徳永: 自然言語の並列構造解析へのスキッピングパーザの応用. 情報処理学会自然言語処理研究会. Vol.113. No.11. pp.67-74. (1996)
- [3] 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)
- [4] 国立国語研究所: 『国立国語研究所報告 3 現代語の助詞・助動詞 一用法と実例一』, 秀英出版. (1951)