

読点を伴う連体修飾の係り先

4F-8

荻野紫穂¹⁾ 成原克恵²⁾

1) 日本アイ・ビー・エム(株) 東京基礎研究所

2) 東京女子大学 文理学部 日本文学科

1. はじめに

我々は、機械翻訳システムの日本語解析において、文節間係り受けの曖昧さを減らすために、読点を伴う連体修飾の係り先を調査した。読点を伴う連体修飾は、直後の体言列に係りにくいと考えられがちだが、実際に直後の文節に係るものの数は、例外と考える範囲を越えている。また、文の途中に現れる、動詞型や形容詞型の活用語の終止・連体形は、全てが連体形の機能だけを持つと考えられがちだが、読点が後接している場合には、終止形として直後の動詞等に接続するものや他の文節中の述語等と並列になるものが、多く見られる。

本稿では、調査結果に基づき、連体形と終止形双方の可能性のある活用語に読点が後接している時、それが終止形として働いているならどのような文節に係り先の候補に残せばいいのか、連体修飾に読点が後接しているなら直後の文節に係るのはどのような場合かについて考察する。

データとしては、1987年1月の新聞記事1ヶ月分から抽出された、連体修飾機能を持つものに読点が後接している例(以下、RT+〈、〉と略記;終止形に読点が後接する場合は、SH+〈、〉と略記)640例を使用する。ただし、この中には助詞ノに読点が後接する場合を含まない。助詞ノには、連体修飾と主格表示の二つの働きがあるので、その二つの識別問題が、読点を伴う連体修飾の調査に絡むのを防ぐためである。

2. 終止形として働いている場合

2.1. 調査結果

データ640例中、終止形として働いているものは211例ある。

211例のうち、助詞その他助詞相当語に直接接続しているものが110例ある。これらに対しては、1)読点とその直後の助詞類を合わせて一つの助詞扱いにする、2)読点直後の助詞まで続けて1文節になるよう接続表を作る、3)SH+〈、〉は直後の助詞に係る規則にする、など、様々な解決策が考えられるが、どれを採用にしても、それほど難しくはない。例えば、文節は必ず自立語1語を含むとし、且つ、

読点の運ぶ情報を生かそうとするならば、接続表による処理が無難だろう。接続先の助詞類の内訳を、表1に示す。

終止形の残り101例は、後続する他の文節内の述語相当語等と並列のものである。このうち、係り先の文節に、並列または引用を示す助詞相当語が含まれているものが63例、係り先の文節が文末にあるものが31例、箇条書きの一部のものが2例、その他が5例である。並列または引用を示す助詞相当語の内訳を、表2に示す。

箇条書きの文は、一般の文とは違った処理をした方がよい(箇条書きであることが判ったら、その文中に現れたSH+〈、〉は箇条書きの最後の部分に係る可能性を残しておけばよい)。他の例について考えると、並列や引用を表す助詞相当語を含む文節と文末の文節とを、係り先の候補に残しておけば、終止形の働きをしている211例の約96%である204例について、正しい係り先が候補に残ることになる。「SH+〈、〉は述語相当語と並列になる」という理由から述語相当語を含む文節全てを係り先の候補に残すよりも、この方法の方が曖昧さが少ない。

2.2. 問題点

問題点として挙げられることは、まず、文末の文節に係るもののうち、2例は倒置の例だった。これらは、表層の係り先決定は並列文と同様にできても、その後の処理では、並列文と同様の解析をできない点で、問題を含んでいる。係り先決定以後の段階に何らかの情報を運ぶことが、必要となろう。

次に、本来ならば修飾を受けない副詞「そう」が係り先になるものがある。#1にそれを示す。

助詞類	数
など	100
なんて	4
って	1
は	1
ぐらい	1
だけ ^{*)}	2
で	1

*) だけじゃ 1
だけでは 1

助詞類	数
と ¹⁾	50
など ²⁾	13
って	

1) と 24
と 26

2) など 8
など 5

表1.
読点に後接する助詞の
種類とその頻度

表2.
SH+〈、〉を受ける引用・
並列を示す助詞

Modifier of Adnominal with "Touten"

Shiho OGINO¹⁾ Katsue NARIHARA²⁾

1) Tokyo Research Laboratory, IBM Research, IBM Japan Ltd.

2) Department of Japanese Literature, Tokyo Women's Christian University

#1 これで帰れる、そう思った

「～帰れると、」なら、「思った」が係り先だとすれば特殊な処理はいらない。#1の場合、「思った」に「帰れる、」が係ると並列文と紛れやすくなるが、「そう」に係るなら「そう」が「帰れる、」以前の内容を受け継いでいることが分かりやすく、並列文との紛れもない。このため、「そう」を係り先にした。しかし、ここには、「帰れると、」の場合と差があってよいか、「そんな」などの連体詞との兼ね合いはどうか、といった問題が残る。今回の調査では、「そんな」には#2の例があったが、これは(下線部「気持ち」を修飾する)連体修飾とも考えられるため、連体修飾の部類に入れた。SH+<,>以前の内容を受ける副詞や連体詞の例は#1、#2の2例しか見つからなかったため、今回は#1だけを問題点とするが、追跡調査の余地がある部分である。

#2 ～観光滞在を楽しむ、そんな気軽な気持ちで『居住者ビザ』を取ったら～

更に、並列だが連体修飾との区別がつけにくいものがあった。その例を#3に挙げる。

#3 わが国は専守防衛、近隣諸国に脅威を与えない、平和外交の3点を貫く

3. 連体修飾として働いている場合

データ640例中、329例が連体修飾として働いているが、その85%に当たる281例が、直後でない体言列に係っている。そのうち4例は、一般には離れた位置にある連体修飾を受けにくい「ため」「まま」の形式名詞に係っている。

329例の約15%に当たる48例が、直後の体言列に係っている。このうち8例は、係り先の体言よりも後に体言がないため、RT+<,>の係り得る先は直後の文節しかない。更に、係り受けは交差ししないという基本規則を採用するなら、もう1例、直後の文節にしか係り得ないものがある。これら9例を除

体言類	数
「こと」	9
代名詞 ^{※)}	8
固有名詞	12
準固有名詞	5
その他	7

※) これ 2
それ 6

表3.
直後の文節でRT+<,>を受ける体言

く39例について、係り先の体言の種類を、表3に示す。

まず、「こと」だが、RT+<,>の直後に「こと」が現れている例も全部で9例だった。つまり、RT+<,>の後に「こと」が続いている場合、そのRT+<,>は全て直後の「こと」に係っていた。代名詞にも同様のことが言える。RT+<,>の直後に「これ」「それ」が現れる例は合わせて9例あったが、そのRT+<,>が直後の代名詞に係らないのは、代名詞が「これまた」の形で現れた場合1例だけだった。これらから、RT+<,>の直後が「こと」「これ」「それ」だった場合、そのRT+<,>の係り先は、ほぼ直後の文節に決定できる、という予想が立てられる。ただし、確定するにはデータ数を増やす必要がある。

固有名詞とそれに準ずるものについては、RT+<,>の直後に現れる固有名詞が必ずそのRT+<,>を受けるとは限らない。直後でない文節に係っている281例のうち、RT+<,>の直後が固有名詞のものが12例ある。よって、正しい候補を残すことはできても、正しい係り先に一意に決定することは難しい。

その他の5例については、正しい候補を残すことも難しい。

以上を考慮した処理をすると、連体形の働きをしている329例中、324例について正しい係り先を候補に残すことができる。そのうちの、直後でない文節に係っている281例中、RT+<,>の直後に固有名詞が現れていない269例については、係り先の曖昧さを減らすこともできる。

4. おわりに

本稿では、読点を伴う連体修飾の係り先として、

- (1) 引用や並列を表す助詞を含む文節と文末の文節を、読点が後接した終止形で終わる文節の係り先の候補に残す
- (2) 読点が後接した連体形で終わる文節は、直後の文節に係らないとする。ただし、文節頭に固有名詞類、「これ」「それ」が来ているものは係り先として残しておく

この2点を考慮に入れることにより、640例の約98%に当たる630例について、正しい係り先候補を残したまま、文節の係り先の曖昧さを減らせることを述べた。今後は、本稿で問題点となった引用文の内容を受ける副詞や連体詞の研究も含め、構文の曖昧さを減らす研究を続ける予定である。

参考文献

- [1] 丸山, 渡辺(1988)「制約伝搬アルゴリズムを用いた日本語文の解析」情報処理学会第37回全国大会予稿集pp.1065-1066