

構文解析

2018. 4. 12

国分芳宏

用語があるかないかだけで検索している現行のシステムはもはや限界にきています。文の構造を含めた形で検索をするべきです。構文解析は文の構造を調べるためのプログラムです。翻訳などのための高度な日本語解析でも必須です。

構文解析の基本的な手法

構文解析は係り受け解析ともいいます。どの用語がどの用語に係るかを調べて文の構造を決めます。受けの性格として体言（名詞）と用言とがあります。それに係る係りの性格として連体修飾と連用修飾とどこにも係らない独立とがあります。その関係を調べて係り先を決定して文の構造を調べます。体言は助詞によって、連用修飾格になる場合と、連体修飾格になる場合があります。用言は連用形が連用修飾格になり、連体形が連体修飾格になります。

	連用修飾格	連体修飾格	独立格
単独の品詞	副詞	連体詞	間投詞
名詞	学校で	学校の	学校だ
時の名詞（副詞）	今朝（食べた）	学校に対する	
数量詞	5人（来た）		
副助詞による	学校より（大きい）		
接尾辞による	制度上		
形式名詞による	精度のため		
動詞	連用形 動いて	連体形 動く	動く。
形容詞	連用形 青く	連体形 青い	青い
形容動詞	連用形 元気に 元気で	連体形 元気な 連体形 堂々たる	元気だ。
接続詞	その上	といった	

次の文を例にして説明を進めます。

今年の夏は暑かったので紅葉は遅いと聞いていましたが、先日、老いた母と妻を連れて養老溪谷へ行って来ました。

前の文を係り受けに分解して文節同士の修飾関係を詳しく見てみましょう。

構文解析すると何が何に係っているのかが分かります。

係り (修飾側)		受け (修飾される側)	
今年の	連体修飾格	夏は	体言で始まる格
夏は	連用修飾格	暑かったので、	用言で始まる格
暑かったので	連用修飾格	聞いていましたが	用言で始まる格
紅葉は	連用修飾格	遅いと	用言で始まる格
遅いと	連用修飾格	聞いていましたが	用言で始まる格
聞いていましたが	連用修飾格	行って来ました	用言で始まる格
先日	連用修飾格	行って来ました	用言で始まる格
老いた	連体修飾格	母と	体言で始まる格
母を	連用修飾格	連れて	用言で始まる格
妻を	連用修飾格	連れて	用言で始まる格
連れて	連用修飾格	行って来ました	用言で始まる格
養老溪谷へ	連用修飾格	行って来ました	用言で始まる格

この係り受けの形でデータベースにしておき、検索する文も同じように構文解析して検索すればノイズが大幅に減ることが期待できます。

係り受けの強さ

文、短文、文節の順で強さがあり、より強さが強い文節をまたいで係ることはありません。この性質を利用して文の構造が決められます。

強さ1 文節 格助詞などで終わる

強さ2 単文 読点、接続詞、接続助詞などで終わる

強さ3 文 句点、間投詞、終助詞などで終わる

この例文は次の4つの単文に分かれます。すべてが最後の「行って来ました」に係ります。

今年の夏は暑かったので紅葉は遅いと聞いていましたが、
先日、
老いた母と妻を連れて
養老溪谷へ

行って来ました。

形容詞には数量詞も係ります。

1年―

遅い―

卒業です。―

これは―

遅い―

話です。―

否定に係る副詞があります。

少しも―

ない―

×少しも―

ある―

否定にかかる副詞は相当な数に上ります。いくつか例を上げますがこの他にも否定に係る慣用句もあります。

あきらかに、あたまから、あんのじょう、いうまでもなく、いちども、いたずらに、いっこうに、いっさい、いまや、おちおち、かいもく、義理にも、けして、けっして、金輪際、さっぱり、さっぱり、しょせん、すこしも、する事なす事、ぜったい、ぜったいに、ぜんぜん、だんじて、ちっとも、てんで、どうしても、どうせ、とうてい、どうにも、どだい、どの道、とて、どれもこれも、とんと、二度と、ほんとうに、まさか、まだ、まったく、まるっきり、まるで、もうとう、もはや、やっぱり、やはり、ゆめにも、よもや、ろくすっぽ、ろくに

もう一つグループの例を上げると願望を表す「たい」「ほしい」「てください」などに係る副詞 19 の副詞がありました。

あえて、かならず、しっかり、せいっぱい、ぜがひでも、ぜひ、ぜひとも、ちからいっぱい、できたら、できることなら、できるだけ、できれば、どうしても、とうぜん、なるたけ、なるべく、なるべくなら、なんとしても、

名詞に係助詞、副助詞がついた結果、副詞と同じよう係り先に制限がある文節があります。

程度副詞として形容詞に係る例

学校より―

大きい―

否定に係る例

学校しか―

ない―

慣用句

慣用句はまとめて1つの用語として扱う必要があります。

領主は―
農民の―
手に掛かった。―

×領主は―
農民の― |
手に―
掛かった。―

「手に掛かる」というのは「殺される」という意味の慣用句で慣用句を一まとめにしてあつかわないと文の構造が分かりません。

双主文

主語が2つあるように見える文があります。

象は-T	象は―
鼻が―	鼻が―
長い。―	ある。―

Tは双主文の意味です。

並列

お酒と-P	彼と―
ビールを―	ビールを―
飲みました。―	飲みました。―

Pは並列の意味です。

「お酒」と「ビール」は両方とも「物」なので並列になります。しかし「彼」は「人」で「ビール」は「物」です。並列にすると「彼」を飲んでしまうことになります。

謝辞

mecab の辞書を使わせていただいています。感謝します。

参考文献

国分芳宏 日本語の文の仕組み 電子出版

南不二男 現代日本語の構造 大修館書店

金谷武洋 日本語に主語はいらない 百年の誤謬を正す (講談社選書メチエ)

中島文雄 「日本語の構造 英語との対比」 岩波新書

朝倉日本語講座 5 文法 I 朝倉書店

朝倉日本語講座 6 文法 II 朝倉書店

後書き

日本語処理のシステムを手放してから5年たちました。やはり頭の中では日本語処理についていろいろ考えていて、ときどきホームページに書いていました。mecab の辞書を入手したのがきっかけで新たに構文解析プログラムを開発する気になりました。お陰様で生活に張りが出てきて、毎朝5時には起きてパソコンに向かっていきます。飲み歩くことも少なくなりました。日本語処理でまだやってみたいことがあるのとその有用性を確かめるために今後このプログラムをベースにしていろいろやってみようと考えています。

mecab の辞書のうち慣用句、副詞、連体詞、助詞、助動詞などは私の集めたものに置き換えました。mecab の品詞分類の仕方は私が考えているものと異なるため、ときどき解析ミスをします。システムに実装するためには、辞書の品詞を検査する必要があります。

不勉強で Windows が分からないので c 言語を用いてコマンドプロンプト上で作っています。皆様に使っていただけるようにするにはどうしたら良いのかどなたか教えていただけませんかでしょうか。1000 ステップ程度のプログラムです。ご興味のあるかたはご連絡ください。