

文節の並べ替えが可読性に与える影響について

情報科学科 服部 祐己

指導教員：山村 毅

1 はじめに

世の中には言っていることは同じなのに、読みやすい文章と読みにくい文章が存在する。例えば、『私は小林が中村が鈴木が死んだ現場にいたと証言したのかと思った。』と『私は鈴木が死んだ現場に中村がいたと小林が証言したのかと思った。』があるとする。この二つは同じことを言っているのだが、情報の伝え方が違うため前者の方が読みやすい人もいれば、後者が読みやすい人もいる。では、より多くの人が読みやすいと思える文の特徴は何だろうか。

本研究では、“文節の順序”に焦点を当てることにした。新聞記事コーパスを用いて、文における文節の順番を調べ、どのような順番にすると人は読みやすいと考えているのかを調査した。また、それによって作られたリストを元に文章評価を行うシステムを作り、人手・システムによって分けられた読みやすい文章と読みにくい文章の一致率を確かめてみた。

2 関連研究

原 [1] は、“文節の順序”に焦点を当て、ルート節（文の最後の節）にかかった文節の並べ替えを行い、可読性の向上を図った。まず、人手によって読みやすい文を決め、次に、それらの文のルート節とそれに係る各文節の距離を測り、各文節の理想の距離を求めた。そして、全ての文のルート節とそれに係る文節の距離を求め、理想の距離の誤差を以下の式で計算した。ここで D は各文の文節の距離、 d は理想の文節の距離である（ f は誤差にあたる）。

$$f = (D_1 - d_1)^2 + (D_2 - d_2)^2 + \dots + (D_n - d_n)^2$$

この f がどれほどであれば、読みやすい文と判定するのか、その閾値を決め、文章を評価したところ、閾値 0.39 のとき、正解率 90.3% であった。

3 語順評価リストの作成

本研究では、原 [1] と同じく、“文節の順序”に焦点を当てた。まず、2005 年、2006 年の毎日新聞から約 100 万文の文章を、日本語依存構造解析システム CaboCha で解析を行った。解析で得られた結果から、同じ文節に係る複数の文節を対象に、それら文節の順番から二組のペアを全て作り、その出現数をカウントする。ただし、複数の品詞から構成されている文節は最後の品詞を、単一の品詞から構成されている場合はその品詞をキーとしてグループ化した。

カウントされた出現数から、ペアとその反ペア（ペアの順番を逆にしたもの）とを比較した時の出現割合を計算し、語順評価リストを作成する。表 1 に作成したリストの一部を示す。

4 語順評価リストによる文章評価

3 で作成した語順評価リストを用い、1995 年の毎日新聞から 1000 文を無作為に選んで、読みやすい文章と読みにくい文章の仕分けを以下の順番で行った。

1. CaboCha を使って依存構造解析を行う。

表 1 語順評価リストの一部

ペア	出現数	出現割合
(など 副助詞, の 連体化助詞)	2579	99.88
(という 格助詞, な 助動詞)	852	99.88
(と 接続助詞, と 格助詞)	578	99.82
(も 係助詞, な 助動詞)	436	99.77
(接続詞節, と 格助詞)	362	99.72

2. 解析した文のルート節に係っている文節を網羅的に入れ替えて新たな文を作成する。
3. 作成した文に対し、読みやすさの評価値 F を計算し、それが閾値を超えていれば、読みやすいと判定する。 F は以下のように定式化する。ここで、 f_n は n 番目に出てくる語順の出現割合、 i は語順の個数とする。

$$F = \frac{1}{i} \left\{ \sum_{n=1}^i (-\log_2 f_n) \right\}$$

これにより、計算した結果について閾値 0.8 の時、表 2 の結果を得た。

表 2 再現率、適合率について

手法	読みやすい文	読みにくい文
人手	2211 文	18433 文
システム	2228 文	18416 文
一致率	935 文	17140 文
可読性	再現率	適合率
読みやすい	42.24 %	41.92 %
読みにくい	92.99 %	93.07 %

5 まとめ

語順評価リストについて、読みやすい語順が決定できそうなものが多かった。また、このリストを使った文章評価について、読みやすい文に関して再現率・適合率は必ずしも高くない。しかし、無作為に行うよりは精度が高く、本研究で用いたリストによる方法は、有効であることがわかる。今後の課題としては『私は、～に』の時、本研究では（は、に）として行っていたが（は、に）として読点を考慮するとまた違う結果が得られるかもしれない。また、本研究では人手による評価を自分のみで行ったが、複数人で行うことでより理想的な読みやすい文を作成することができると思われる。

参考文献

- [1] 原 源記：“文節の順序による可読性の向上手法の提案”，愛知県立大学情報科学部卒業論文，2016