

読点配置の適切性に影響する文の係り受け構造上の要因について¹⁾

伊藤 俊一

情報教育講座

Effect of Sentence Structure on Appropriateness of Punctuation

Toshikazu ITO

Department of Information Sciences, Aichi University of Education, Kariya 448-8542, Japan

1. はじめに

文中のどこに読点を配置すると、より読みやすい文になり得るのか、ということについての明確なルールは未だ得られていないのが現状である。

伊藤・上野^[1]、および、伊藤^[2]では、文中に読点を打つ者、すなわち、読点の配置を決定する者（書き手、あるいは、推敲者）の打点行動を分析することによって、読点配置の適切性に影響する要因を探ることを試みた。

伊藤・上野^[1]では、文節間の係り受け構造において階層的に上位に位置する接点から下方に向かって順に読点を打つ行動を取る傾向が強い者ほど、適切性の高い読点配置を産出しやすいことを示した。さらに、伊藤^[2]では、階層的に上位に位置する接点から下方に向かって順に読点を打つ行動を実験参加者に強制的に促すことによって、より適切性の高い読点配置を産出させることができることを示した。これらの結果は、階層的に上位に位置する接点に対して読点を配置することが、読点配置の適切性を向上させるものであることを示唆する。

岩畑^[3]も、文中のある特定の位置における「読点容認度」をその位置の構造的特性と対応づけ、構造上、上位にある接点ほど「読点容認度」が高いというモデルを提案している。

「階層的に上位に位置する接点に対して読点を配置することが、読点配置の適切性を向上させる」という因果関係をさらに明確なルールとして形式化するためには、読点が配置された接点の、文節間の係り受け構造における位置を算出するための客観的な基準が必要となる。そこで、本研究では、読点配置を数値化して表す8種類の変数について考える。そして、それぞれの変数において算出された数値と読点配置の適切性との関連性の高さについて、実験データを元に比較・検討する。

2. 係り受け関係に基づく読点配置の数値化

2.1. 接点間の階層的関係

本研究では、日本語構文解析システムKNPによる文節間の係り受け関係の解析結果に基づいて、文中に存在する接点間の階層的な上下関係を、次の通り、定める。

接点aと接点cの間に構造木(1)の関係が成り立つとき、接点aは接点cに対して、右方枝分かれ構造における上位の接点とみなすことができる。このことから、接点aを接点cの「上流接点」と呼ぶことにする。

(1)

X ─┐ 接点 a
Y ─┘ 接点 c
Z

また、接点bと接点cの間に構造木(2)の関係が成り立つとき、接点cは接点bに対して、左方枝分かれ構造における上位の接点とみなすことができる。このことから、接点cを接点bの「上流接点」とする。

(2)

W ─┐ 接点 b
Y ─┘ 接点 c
Z

例えば、KNPによる文(3)の係り受け関係の解析結果(4)において、接点の階層的な上下関係は、上流から下流に向かって順に、接点a、接点c、接点bとなる。

(3)

「ウオータースライダーが45度の急傾斜を滑り落ちる。」

(4)

ウオータースライダーが ——— 接点 a
 45度の ㄣ | 接点 b
 急傾斜を 十 接点 c
 滑り落ちる。

2.2. 読点ごとの上流接点数

2.1.で述べた方法で接点間の階層的な上下関係を定めることによって、文中に配置される読点ごとに、その上流に位置する接点の度数「上流接点数」を算出することが可能となる。

例えば、KNPによる文(5)の係り受け関係の解析結果(6)において、読点Bの上流接点数は2個(接点aと接点c)、読点Dの上流接点数は1個(接点c)となる。

(5)

「ウオータースライダーが45度の急傾斜を、一気に滑り落ちると観客は、水しぶきでずぶぬれた。」

(6)

ウオータースライダーが ——— 接点 a
 45度の ㄣ |
 急傾斜を、 十 読点 B
 一気に 十
 滑り落ちると ——— 接点 c
 観客は、 十 読点 D
 水しぶきで 十
 ずぶぬれた。

上流接点数とは、文節間の係り受け関係に基づく文の階層的構造において、当該の読点が、どのくらい下位に位置する接点に配置されているかを示す変数であると言える。

2.3. 読点ごとの残留接点数

2.2.で述べた読点ごとの上流接点数から、読点が配置されている接点の度数を引いた値を、「残留接点数」とする。

例えば、KNPによる文(7)の係り受け関係の解析結果(8)において、読点A、読点Bの残留接点数は、ともに1個(接点c)となる。読点Aが配置されている接点も、接点cと同様に、読点Bの上流に位置するわけだが、しかし、そこに読点Aが配置されていることによって、読点Bの残留接点数には加算されない。

(7)

「ウオータースライダーが、45度の急傾斜を、一気に滑り落ちると観客は水しぶきでずぶぬれた。」

(8)

ウオータースライダーが、 ——— 読点 A
 45度の ㄣ |
 急傾斜を、 十 読点 B
 一気に 十
 滑り落ちると ——— 接点 c
 観客は ———
 水しぶきで 十
 ずぶぬれた。

2.4. 遠隔接点数・隣接接点数

村田・大野・松原^[4]は、社説記事を対象とした調査を行ない、「係り受け関係にある隣接文節間192,540箇所に対して、読点が挿入されたのは5,866箇所、挿入率は3.04%に過ぎなかった。一方、係り受け関係のない隣接文節間への挿入率は36.99%であった。」と報告している。

本研究では、係り受け関係のない隣接文節間の接点を「遠隔接点」と呼び、係り受け関係にある隣接文節間の接点を「隣接接点」と呼ぶことにする。

村田らの報告は、読点が配置される可能性は「遠隔接点」において高く、「隣接接点」においては極めて低いことを示している。このことは、上流接点あるいは残留接点が、遠隔接点に該当するか、あるいは、隣接接点に該当するかによって、それらの度数が読点配置の適切性に及ぼす影響に違いが生じる可能性を示唆する。

そこで、本研究では、上流接点あるいは残留接点を、さらに遠隔接点と隣接接点とに分け、「遠隔上流接点数」・「隣接上流接点数」・「遠隔残留接点数」・「隣接残留接点数」をそれぞれ異なる変数として扱うことにする。

例えば、KNPによる文(9)の係り受け関係の解析結果(10)において、接点cは隣接接点である。このとき、読点Bの遠隔上流接点数は2個(読点Aが配置された接点と接点d)、隣接上流接点数は1個(接点c)、遠隔残留接点は1個(接点d)、隣接残留接点数は1個(接点c)となる。

(9)

「国土交通省が、熊本県の球磨川水系に、建設を進めている川辺川ダムに漁民がノーを出している。」

(10)

国土交通省が、
熊本の
球磨川水系に、
建設を
進めている
川辺川ダムに
漁民が
ノを
出している。

読点A
読点B
接点c
接点d

2.5. 延べ接点数・異なり接点数

複数の接点に読点が配置されている文においては、文全体としての上流接点数・残留接点数を算出する際に、読点ごとの接点数の「延べ数」として算出する方法と、「異なり数」として算出する方法の、2通りの方法を考えることができる。

例えば、先に示した構造木(8)において、読点Aの上流接点は接点cであり、読点Bの上流接点は読点Aが配置された接点と接点cである。これらのうち、接点cは、読点Aと読点Bに共有された上流接点である。また、構造木(8)において、読点Aの残留接点、読点Bの残留接点は、ともに接点cであり、接点cは、読点Aと読点Bに共有された残留接点である。このとき、構造木(8)におけるそれぞれの値は、次の通り、算出される。

構造木(8)において：

- (11) 読点Aと読点Bの延べ上流接点数：
読点Aの上流接点数(1) + 読点Bの上流接点数(2) = 3
- (12) 読点Aと読点Bの延べ残留接点数：
読点Aの残留接点数(1) + 読点Bの残留接点数(1) = 2
- (13) 読点Aと読点Bの異なり上流接点数：
(読点Aの上流接点数(1) + 読点Bの上流接点数(2))
- 読点A・B共有の上流接点数(1) = 2
- (14) 読点Aと読点Bの異なり残留接点数：
(読点Aの残留接点数(1) + 読点Bの残留接点数(1))
- 読点A・B共有の残留接点数(1) = 1

構造木(8)においては、読点が2箇所の接点に配置されているので、読点あたりの延べ上流接点数・延べ残留接点数・異なり上流接点数・異なり残留接点数は、式(11)、(12)、(13)、(14)の値をそれぞれ2で割った値、すなわち、1.5、1.0、1.0、0.5となる。なお、構造木(8)においては、読点Aおよび読点Bの上流に隣接接点が存在しないので、これらの値は、すべて遠隔

接点数である。

2.6. 読点配置の数値化

上流接点数と残留接点数の別(2.2および2.3参照)、遠隔接点数と隣接接点数の別(2.4参照)、延べ接点数と異なり接点数の別(2.5参照)を組み合わせることによって、読点配置を数値化した次の8通りの変数を考えることができる。

| | |
|-----------|------------|
| 延べ遠隔上流接点数 | 異なり遠隔上流接点数 |
| 延べ隣接上流接点数 | 異なり隣接上流接点数 |
| 延べ遠隔残留接点数 | 異なり遠隔残留接点数 |
| 延べ隣接残留接点数 | 異なり隣接残留接点数 |

例えば、先に示した構造木(10)におけるそれぞれの変数の値は、次の通り、算出される。

構造木(10)において：

| | |
|---------------|----------------|
| 延べ遠隔上流接点数=1.5 | 異なり遠隔上流接点数=1.0 |
| 延べ隣接上流接点数=1.0 | 異なり隣接上流接点数=0.5 |
| 延べ遠隔残留接点数=1.0 | 異なり遠隔残留接点数=0.5 |
| 延べ隣接残留接点数=1.0 | 異なり隣接残留接点数=0.5 |

本研究では、これらの変数が読点配置の適切性に及ぼす影響の大きさについて、実験データを元に比較・検討する。

3. 方法

【実験参加者】

大学生28名であった。

【材料】

朝日新聞記事データベース(CD-HIASK)2001年版を用いて、次の条件を満たす文を50種類、抽出した。

- ・日付の異なる社説記事から抽出された文であること
- ・日本語構文解析システムKNPによる解析結果において計4箇所の遠隔接点を含むこと
- ・原文中の1箇所ないしは2箇所の接点に読点が配置されていること

以上の条件によって抽出された50種類の文に対して、文中に含まれる4箇所の遠隔接点から選ばれた任意の2箇所に読点を配置することとした。4箇所の遠隔接点をそれぞれa, b, c, dとした場合、読点の配置には(a, b)(a, c)(a, d)(b, c)(b, d)(c, d)の6通りのパターンが存在することになる。これらすべての読点配置パターンを、本実験における評定課題の対象とする。すなわち、50種類の文×6通りの読点配置パターン=300文が対象となる。

【質問紙】

被験者ごとに、25ページからなる冊子を作成した。各ページには、用意した50種類の文のうちの1文を、「基本文」として1パターン、「比較文」として6パターン、計7パターンの読点配置とともに印刷した。基本文には読点を1箇所も配置しないパターン、比較文には4箇所の遠隔接点のうち任意の2箇所に読点を配置した計6パターンを用いた。

それぞれの比較文の右端には、その比較文が基本文と比べて、どの程度、読みやすいか、あるいは、読みにくいかを実験参加者が評定するための5段階スケールを設けた。5段階の内訳は、「5：かなり読みやすい」「4：やや読みやすい」「3：同程度」「2：やや読みにくい」「1：かなり読みにくい」であった。

実験参加者に対する文（ページ）の割り当て、文（ページ）の提示順序、および、各ページにおける比較文の並び順は、実験参加者間でカウンターバランスされた。

【手続き】

実験参加者は、ページごとに、一番上に印刷されている基本文を読んで、内容をよく理解する。その基本文を読むときに感じた「読みやすさ」、あるいは、「読みにくさ」の程度を、比較文に対する評定課題が完了するまで、しばらくの間、頭の中で覚えておくように指示される。

続いて、比較文を上から順に読み、それぞれの比較文を読むごとに、読点が増えられたことによって、基本文よりも、どの程度、読みやすくなったか、あるいは、読みにくくなったかを5段階スケール上で評定する。

4. 結果**【相関係数】**

比較文として評定課題の対象となった300文（50種類の文×6通りの読点配置パターン）のそれぞれについて、「読みやすさ」評定値（基本文と比べて、どの程度、読みやすいか）の平均を求めた。

また、300文のそれぞれについて、2.6.で述べた読点配置を示す8通りの変数の値を求めた。なお、本研究で評定課題の対象とした比較文においては、すべての読点が遠隔接点に配置されたため、これらの比較文における隣接上流接点数と隣接残留接点数は同じ値を取ることになる。

読点配置を示す8通りの変数と、実験で得られた「読みやすさ」評定値の平均との相関係数をFig. 1に示す。

遠隔接点数と「読みやすさ」評定値との相関が総じて高い一方で、隣接接点数と「読みやすさ」評定値との相関は低い。また、遠隔接点の中でも、残留接点数と「読みやすさ」評定値との相関が、上流接点数と「読み

やすさ」評定値との相関に比べて高い。残留接点、上流接点とも、延べ数として算出した値のほうが、異なり数として算出した値よりも、「読みやすさ」評定値との相関が高い。

【分散分析】

読点配置を示す8通りの変数の値ごとに、「読みやすさ」評定値の平均を求めた。それらをFig. 2～Fig. 5に示す。以下に、変数ごとの分散分析の結果を挙げる。

延べ遠隔残留接点数の効果は有意であった（ $F(4, 295) = 63.87, p < .01$ ）。Tukey-kramer法を用いた多重比較によれば、「読みやすさ」評定値は、0.0個 > 0.5個 > 1.0個 > 1.5個 = 2.0個の順に高かった。（Fig. 2.参照）

異なり遠隔残留接点の効果は有意であった（ $F(2, 297) = 110.43, p < .01$ ）。「読みやすさ」評定値は、0.0個 > 0.5個 > 1.0個の順に高かった。（Fig. 3.参照）

延べ遠隔上流接点数の効果は有意であった（ $F(4, 295) = 52.88, p < .01$ ）。「読みやすさ」評定値は、0.5個 > 1.0個 > 1.5個 > 2.0個 = 2.5個の順に高かった。（Fig. 4.参照）

異なり遠隔上流接点数の効果は有意であった（ $F(2, 297) = 47.64, p < .01$ ）。「読みやすさ」評定値は、0.5個 > 1.0個 = 1.5個の順に高かった。（Fig. 5.参照）

延べ隣接残留接点数および延べ隣接上流接点数の効果は有意であった（ $F(5, 294) = 3.15, p < .05$ ）。Tukey-kramer法を用いた多重比較による対間の差は、いずれも有意ではなかった。（Fig. 2.およびFig. 4.参照）

異なり隣接残留接点数および異なり隣接上流接点数の効果は有意であった（ $F(3, 296) = 4.53, p < .01$ ）。「読みやすさ」評定値は、0.0個 > 0.5個 = 1.0個 = 1.5個の順に高かった。（Fig. 3.およびFig. 5.参照）

5. 考察

本研究で検証した読点配置を示す8通りの変数のうち、「読点あたりの延べ遠隔残留接点数」が、読点配置の適切性と最も関連性が高く、延べ遠隔残留接点数が多いほど読点配置の適切性は損なわれることが示された。この結果は、「階層的に上位に位置する接点に対して読点を配置することが、読点配置の適切性を向上させる」という先行研究の知見を裏付けると同時に、さらに詳細に、以下のことを明らかにしたものと言える。

(A)：配置された読点の上流に接点が存在していたとしても、すなわち、配置された読点自身はそれらの接点の下流に位置していたとしても、上流の接点にも同様に読点が配置されていた場合には、それら上流の接点の存在は、読点配置の適切性を損なわせる主な原因にはならない。

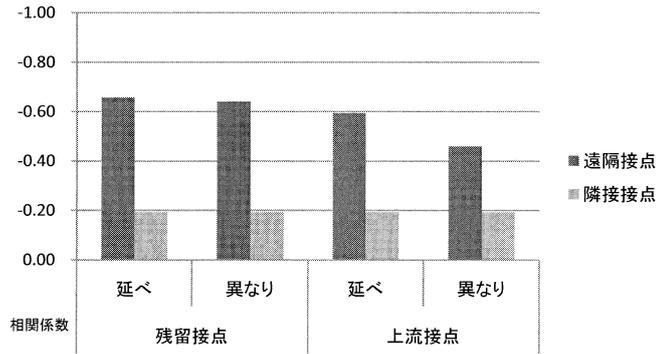


Fig. 1. 読点配置と「読みやすさ」評定値の相関

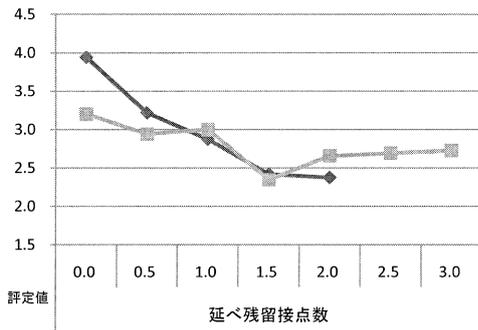


Fig. 2. 延べ残留接点数毎の「読みやすさ」評定値

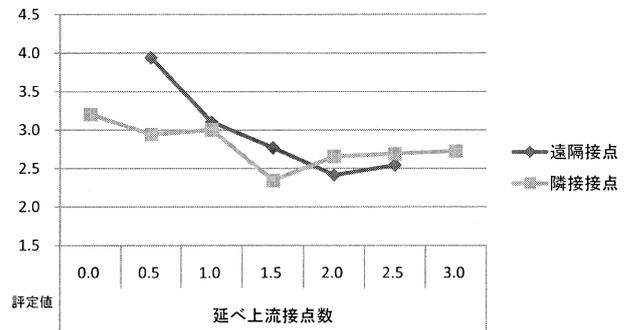


Fig. 4. 延べ上流接点数毎の「読みやすさ」評定値

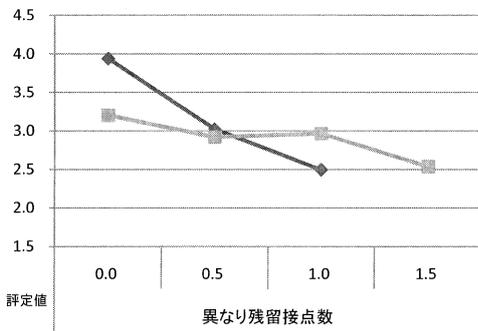


Fig. 3. 異なり残留接点数毎の「読みやすさ」評定値

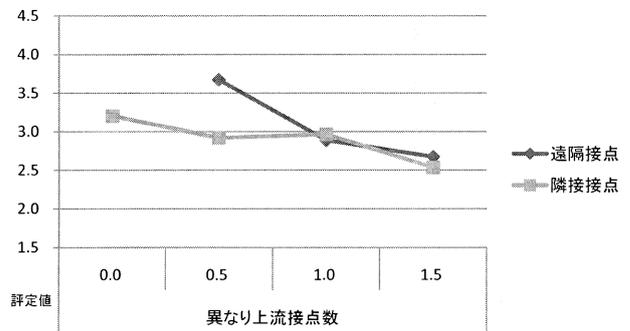


Fig. 5. 異なり上流接点数毎の「読みやすさ」評定値

(B)：配置された読点の上流に接点が存在していたとしても、すなわち、配置された読点自身はそれらの接点の下流に位置していたとしても、上流の接点が係り受け関係にある隣接文節間の接点（隣接接点）である場合には、それら上流の接点の存在は、読点配置の適切性を損なわせる主な原因にはならない。

(A)と(B)から、読点の適切性は、文節間の係り受け構造における階層的な深さによって単純に規定されるのではなく、むしろ、その読点の上流に新たな読点が配置され得る可能性を残した接点が存在するか否かによって規定されていると言える。既に読点が配置されている上流の接点、および、上流に存在する隣接

接点は、新たな読点が配置される可能性を残した接点からは除外されることとなるわけである。（新たな読点が配置される可能性を残した接点から隣接接点が除外される理由については、2.4.を参照のこと。）

要するに、上流に読点を配置することが可能な接点が残っているにもかかわらず、それらを“差し置いて”下流の接点に読点を配置することが、読点配置の適切性を大きく損なわせる原因となり得ることを、本研究の結果は示していると言える。

今後は、本研究におけるこれらの知見に基づいた読点配置の客観的な評価方法を確立することとともに、その推敲支援への応用が望まれる。

引用文献

- [1] 伊藤俊一・上野慎之介 (2008) 文推敲者による読点打ち行動の分析 言語処理学会第14回年次大会発表論文集, 1101-1104.
- [2] 伊藤俊一 (2010) 打点方略が読点配置の適切性に及ぼす影響 言語処理学会第16回年次大会発表論文集, 407-410.
- [3] 岩畑貴弘 (2004) 読点の使用とその決定要素について 神奈川大学「人文研究」, 154, 51-81.
- [4] 村田匡輝・大野誠寛・松原茂樹 (2010) 日本語テキストにおける読点位置の検出 言語処理学会第16回年次大会発表論文集, 812-815.

注

- 1) 本研究の一部は, 言語処理学会第17回年次大会にて発表された。

(2011年8月31日受理)