

文章作成における読点推敲支援

愛知教育大学大学院教育学研究科 発達教育科学専攻 情報教育領域
森部 大悟 伊藤 俊一

1. はじめに

文を読みやすくする上において、読点はとても重要なものである。読点は文を読みやすくしたり、理解しやすくしたりするものである。読点一つで文の意味が変わってくることもある。読点の用法は多くの研究者が論じているが、定まった用法は確立されていない。

伊藤・上野[1]の研究では、人が文に読点を打つ際の行動を分析することで、その行動の特徴と、その行動が読点の適切性にどれだけ影響しているかを調べている。その結果、階層的に上にある接点から順に読点を打った被験者の方が、読点の配置の適切性の高い評定値となることが明らかとなった。構造上高い接点から読点を打つと評定値が高いことを利用し、森部[2]では、被験者が読点を打つ際に用いる方略を制限することによる読点配置の適切性への影響を調べた。その結果、読点を打った箇所から枝分かれしていく、階層的な順序で、次に読点を打てる箇所を制限していくことにより、実験で用いた条件の中でもっとも適切な読点配置が得られることがわかった。

そこで、本研究では、文を推敲する際に文の構造上高い接点に読点が打たれるような支援方法を考えていく。また、その支援方法を用いたソフトを作成し、支援方法の効果とその支援方法の改善すべき点を特定することを目的とする。

2. 不適切な読点の指摘基準の作成

不足している読点箇所と不必要な読点箇所を指摘する際の基準を、村田ら[3]・伊藤[4]の研究に基づき作成した。

2.1. 村田ら (2010) による分析結果

村田らは係り受け関係を隣接文節の性質で2タイプに分けて分析を行っている。

1) 隣同士が係り受け関係にある文節間

2) 隣同士が係り受け関係にない文節間

図1では、灰色で囲まれた接点 (BC 間・CD 間・FG 間) の3つが隣同士が係り受け関係にある文節間であり、AB 間・DE 間・EF 間の3つが隣同士が係り受け関係にない文節間である。

隣同士が係り受け関係にある文節間への読点の「挿

入率は3.04%」であるという結果が得られている。このことから、隣同士が係り受け関係にある文節間に読点が打たれることはほとんどないことが明らかとなった。

2.2. 伊藤 (2011) による分析結果

伊藤は村田らの1)と2)のそれぞれを以下のように表している。

1) 遠隔接点 …村田らによる1)と同じ

2) 隣接接点 …村田らによる2)と同じ

さらに、1) 2)それぞれにおいて、残留接点の考え方を導入し、以下のようにまとめている。

I) 遠隔残留接点…ある打たれた読点から見て、文構造で上流にある文節の内、読点の打たれていない遠隔接点。1) の内で残留接点である文節。

II) 隣接残留接点…ある読点から見て、文構造で上流にある文節の内、読点の打たれていない隣接接点。2) の内で残留接点である文節。

図1では、AB 間から見ると、上流にある①DE 間は読点の打たれていない遠隔接点であるため、遠隔残留接点である。また、②CD 間が隣接残留接点である。

伊藤の研究によって、遠隔残留接点が多いほど読点配置の適切性は損なわれるという結果が得られた。つまり、ある打たれた読点から見て、文構造で上位にある文節の内、読点の打たれていない遠隔接点の数が多ければ、読点配置の適切性が低くなることが明らかとなった。

2.3. 不適切な読点の指摘基準

2.1節より、(1)隣同士が係り受け関係にある文節間に読点の打たれている場合は、読点が必要のない可能性があることを書き手に指摘する。

2.2節より、遠隔残留接点を生み出している読点と遠隔残留接点自体の文節間を指摘箇所とする。遠隔残留接点を生み出している読点は削除することによって遠隔残留接点が減る。また、遠隔残留接点自体に読点を打つことによっても遠隔残留接点は減る。そのため、(2)遠隔残留接点を生み出している読点が必要であるなら、遠隔残留接点にも読点が必要である可能性が高いことを書き手に指摘する。

これら(1)・(2)を不適切な読点箇所の指摘基準とする。

3. 読点推敲支援ソフトの作成について

推敲時に、読点を打つことの支援を行うソフトを作成する。

3.1. ソフトの支援基準

2.3節より、(1)においては該当する読点を赤色にし、

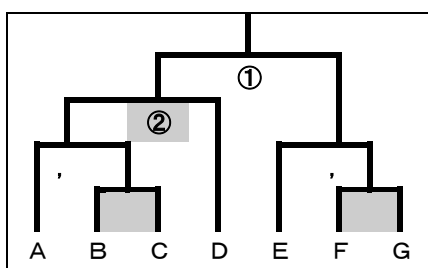


図1. 係り受け関係によって木構造化された文の例

不必要である可能性があることを指摘する。(2)においては、遠隔残留接点を生み出している読点を黄色、残留接点である文節間を緑色にし、黄色の文節に打たれた読点が必要であるなら、緑色の文節にも読点が必要である可能性が高いことを指摘する。以上をまとめて、ソフトの支援基準となる表1を作成した。

接点の状態				支援方法
読点	係り受けの性質	自身の残留接点性	自身に起因する遠隔残留接点	
有	隣接	残留接点	有	×
有	隣接	残留接点	無	×
有	隣接	非残留接点	有	赤色
有	隣接	非残留接点	無	赤色
有	遠隔	残留接点	有	×
有	遠隔	残留接点	無	×
有	遠隔	非残留接点	有	黄色
有	遠隔	非残留接点	無	指摘なし
無	隣接	残留接点	有	×
無	隣接	残留接点	無	指摘なし
無	隣接	非残留接点	有	×
無	隣接	非残留接点	無	指摘なし
無	遠隔	残留接点	有	×
無	遠隔	残留接点	無	緑色
無	遠隔	非残留接点	有	×
無	遠隔	非残留接点	無	指摘なし

表1. ソフトの支援基準

3.2. 接点の状態判断のための動作と入出力

文の解析と係り受け関係の判別に関しては、既存の文構造解析ソフト『Cabocha』[5]を使用する。『Cabocha』の係り受け解析レイヤの出力フォーマットによる出力結果から、接点の状態を判断し、表1の16通りのどこに当てはまるのかを分類していく。

本読点推敲支援ソフトでは、文を1文入力すると、内部で文節間の接点の状態の判断を行い、表1に示す支援方法で色付けされた文が表示される。

4. 書き手の違いによる読点推敲支援の必要性の調査

書き手の違いによって、本読点推敲支援ソフトの二通りの指摘基準(2.3節の(1)・(2))のどちらをどれくらい必要とするのかを調査する。

4.1. 方法

4.1.1. 材料

- ・小学生(60人分)卒業文集
- ・中学生(60人分)卒業文集
- ・大学生(60人分)愛知教育大学1年前期開講『レポートライティング』の提出課題の内、添削される前の文章(初回と2回目)
- ・社説(60記事分)…日経新聞(96)

4.1.2. 手続き

作成した読点推敲支援ソフトを使用し、収集した文の解析を行う。それぞれの文において、文・文節の数、読点推敲支援ソフトの表1における支援方法が赤色・黄色・緑色となる文節の数を調べる。

4.2. 結果

①問題指摘箇所数の比較

文節を5つに分類し、分析を行った。『指摘なし読点有』・『指摘なし読点無』とは、読点推敲支援ソフトで問題の指摘がされなかった箇所である。それらの文節の内、読点が打たれた文節が『指摘なし読点有』であり、読点が打たれていない文節が『指摘なし読点有』である。『赤』・『黄』・『緑』とは、読点推敲支援ソフトによって問題が指摘された箇所である。

書き手×問題指摘箇所数で χ^2 検定を行った結果、小学生・中学生の文章は『赤』が多く、大学生は『黄』と『緑』が多く、社説は『赤』『黄』『緑』の全てにおいて少ないことが明らかとなった。 $(\chi^2(12)=678.271, p<.01)$

②1文あたりの文節間の数の比較

1文あたりの文節間の数が異なるのかどうかを調べた。1文あたりの文節間の数は『小学生=中学生<社説<大学生』の順で多いことが明らかとなった。 $(F(3,6702)=355.40, p<.01)$

③文節に打たれた読点数の比較

読点が打たれた文節間の数で χ^2 検定を行った結果、『社説=大学生<中学生=小学生』の順で読点の打たれた文節数が多いことが明らかとなった。 $(\chi^2(3)=422.839, p<.01)$

4.3. 考察

書き手によって、支援の必要箇所が異なることが明らかとなった。

小学生・中学生の文は全体的な文節数が少なく、係り受けに複雑性が生まれにくい。また、読点が打たれている文節が多い。このことから、隣接接点に読点が打たれることが多くなったと考えられる。小学生・中学生に対しては、隣接接点に打たれる読点における支援が必要であることが考えられる。

大学生は1文あたりの文節数が多い。このことから、複雑性が生まれやすく、遠隔接点において必要となる箇所に読点が打たれていないか、余分な読点が打たれてしまっていることが多くなったと考えられる。大学生に対しては、遠隔残留接点と遠隔残留

接点を生み出す読点における支援が必要であることが考えられる。

社説は、全体的に問題の生じている箇所が少なく、支援の必要性が低いと考えられる。

5. 読点推敲支援による推敲行動の分析

作成した読点推敲支援ソフトを大学生に使用してもらい、どのような推敲が行われたかのデータを収集する。被験者は与えられたテーマに関して、本読点推敲支援ソフトを用いて作文を行う。

5.1. 方法

5.1.1. 被験者 大学生 20人

5.1.2. 手続き

実験用コンピュータを用いて実験を行う。被験者は、本実験の説明と読点推敲支援ソフトの使い方の説明を読み、読点推敲支援ソフトの操作練習を行なった後、1つのテーマに沿って30文以上の作文を行う。作文の終了時には、読点推敲支援ソフトの指摘についてのアンケートを行う。

5.2. 結果

5.2.1. 推敲行動の分析

本読点推敲支援ソフトでは、ユーザの推敲行動がログとして自動的に保存されるようになっている。実験から得られたログを用いて、どのような推敲が行われたかの分析を行う。

実験で得たデータの内、最初に入力した文（初稿の文）と、推敲をした後に最終的に完成した文（最終稿の文）を比較し、読点以外の文字が変化している文を、分析対象から除外した。

①文節あたりの初稿と最終稿の変化の分析

初稿と最終稿で、読点にどのような変化があったかの分類を、4.2節の①で考えた文節間の問題指摘箇所の分類に従い、行った。その結果、表2のようになった。

		最終稿					合計
		指摘なし 読点無	指摘なし 読点有	赤	黄	緑	
初稿	指摘なし 読点無	3327	1	0	0	0	3328
	指摘なし 読点有	0	512		0	1	513
	赤	32		28			60
	黄	30	38		47	1	116
	緑	21	51		1	46	119
合計		3410	3410	602	28	48	4136

表2. 文節あたりの初稿と最終稿の変化

初稿と最終稿を比較した結果、読点推敲支援ソフトを用いることにより、赤の指摘では53.33%、黄

の指摘では58.62%、緑の指摘では60.50%、全ての色を合わせると58.3%で色が消された。色の指摘を行うことによって、半数以上の問題指摘箇所の色が解消されていることが明らかとなった。

5.2.2. 各色の支援の評価

読点推敲支援ソフトを使用してもらった際に、各色の支援が参考になったかどうかを5段階で評価してもらった。その結果、表3のようになった。

	赤評価	黄評価	緑評価
平均	3.80	3.50	3.45

表3. 各色の支援の評価

5.3. 考察

本読点推敲支援ソフトを用いることによって、最初に指摘していた箇所があったものの内、色が消えたものは、文節としては58.3%であり、半分以上の問題指摘箇所をなくすことができた。また、使用した被験者の評価からも、全ての色の評価値が3.45以上であり、全ての色の指摘が参考になったという結果を得た。このことから、本読点推敲支援ソフトの支援方法が受け入れられ、推敲の参考となったと考えられる。

6. 推敲支援によって得られた文の評価

5章の実験で得た推敲後の文が最初の文と比べ、読みやすい文となったかどうかの評価を第三者にしてもらう。

6.1. 方法

6.1.1. 被験者 愛知教育大学の学生 40人

6.1.2. 材料

5.2節で分析の対象となった文の内、本読点推敲支援ソフトの支援の結果、読点の位置が変更された文をすべて抽出する。

6.1.3. 手続き

被験者は、アンケート用紙の各行に印刷されている文ごとに、読点配置の適切性を5段階で評価する。アンケートは一人45文（3ページ分）行う。

6.2. 結果

指摘箇所のない文節（『指摘なし読点無・指摘なし読点有』）と指摘箇所のある文節（『赤・黄・緑』）の有無により、『指摘なし』の文・『赤』のみを含む文・『緑黄』を含む文・『赤緑黄』を含む文の4つに、初稿と最終稿の文を分類した。その内、初稿が『赤』で最終稿『指摘なし』の文と、初稿『緑黄』で最終稿『指摘なし』の文において分析を行った。

1) 初稿『赤』で最終稿『指摘なし』の文の分析

初稿『赤』で最終稿『指摘なし』の文27文で、最終稿の文の評価が下がった文に共通する特徴を調べた。その結果、共通する特徴をまとめ、以下のように考え

られる。

[I] 誤検出された接点（隣接接点ではない接点）が存在する文。[II] 動詞句が並列する隣接接点が存在する文。[III] 同一文字種の連続する隣接接点が存在する文。

[I]～[III]・それら以外の文の評価値は表4となる。

	評価値		合計
	3より大きい	3以下	
[I]誤検出	4	9	13
[II]動詞句並列	1	3	4
[III]同一文字種の連続	1	3	4
上記以外	4	2	6
合計	10	17	27

表4. 初稿『赤』で最終稿『指摘なし』の文の特徴分類とその評価値（文の数）

2) 初稿『緑黄』で最終稿『指摘なし』の文の分析

初稿『緑黄』で最終稿『指摘なし』の文 55 文、最終稿の文の評価が下がった文に共通する特徴を調べた。その結果、共通する特徴をまとめ、以下のように考えられる。

[IV] 誤検出された文。（緑色の接点が黄色の接点の上流にない文。）[V] 複文における主題の係り先が複数通り考えられる文。[VI] 名詞節・引用節を含んでいる文。

[IV]～[VI]・それら以外の文の評価値は表5となる。

	評価値		合計
	3より大きい	3以下	
[IV]誤検出	1	27	28
[V]複文における主題の係り先	1	9	10
[VI]名詞・引用節を含む	0	3	3
上記以外	8	6	14
合計	10	45	55

表5. 初稿『緑黄』で最終稿『指摘なし』の文の特徴分類とその評価値（文の数）

6.3. 考察

1)と2)で、最終稿で評価値の下がった文の特徴をまとめた結果、それぞれを3つに分類することができた。1)の[I]と2)の[IV]は、係り受け関係の誤った解析を行なっている。本読点推敲支援ソフトでは、読点を手がかりとしないように、構造解析ソフト『Cabocha』で解析する際に、読点を削除して解析を行なっている。そのため、『Cabocha』ではうまく判断できなかったことが考えられる。表4より、[II] 動詞句が並列する隣接接点が存在する文と、[III] 同一文字種の連続する隣接接点が存在する文において、最終稿で評価値の下がった文が多くなっている。これらの箇

所においては、赤の指摘をしないようにすることが必要であることが考えられる。表5より、[V] 複文における主題の係り先が複数通り考えられる文と、[VI] 名詞節・引用節を含んでいる文において、最終稿で評価値の下がった文が多くなっている。[V]複文における主題の係り先においては、主題の後に読点が必要か否かはユーザに判断を任せ、それ以上の指摘を行わないことが必要であると考えられる。[VI] 名詞節・引用節が存在する際は、それらの節内を一つの文として判断する必要があり、名詞節・引用節以外の主節とは切り分けて扱う必要があると考えられる。名詞節・引用節内とそれ以外の主節、それぞれを独立した文とし、推敲支援を行う。1)で[I]～[III]の特徴がなかったもの、2)で[IV]～[VI]の特徴がなかったものは、最終稿で評価値が上がったものが多くなった。1)と2)の[I]～[VI]を表1に示した指摘対象のルールに追加することで、適切な読点推敲支援に近づけると考える。

7. まとめ

6章より、係り受けの観点から支援を行う際に、以下の基準を追加することで、効果の高い支援になることが考えられる。

1) 隣接接点（赤）の指摘に関して

①動詞句が並列する隣接接点で指摘を行わない。

②同一文字種の連続する隣接接点で指摘を行わない。

2) 遠隔残留接点（緑）とそれを生み出す接点（黄）の指摘に関して

①複文における主題の係り先が複数通り考えられる文では指摘を行わない。

②名詞節・引用節が存在する場合、名詞節・引用節内とそれ以外の主節を、独立した文として、推敲支援を行う。

8. 参考文献・参考資料

[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] 森部大悟(2009)「打点方略が読点配置の適切性に及ぼす影響」愛知教育大学情報教育課程 卒業論文

[3] 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回次大会発表論文集

[4] 伊藤俊一(2011)「読点配置の適切性を規定する文構造上の要因について」言語処理学会第17回年次大会発表論文集, 695-698.

[5] 工藤拓, 松本裕治(2002)「チャンキングの段階適用による係り受け解析 Cabocha」情報処理学会論文誌, Vol 43 No. 6, pp.1834-1842