

## 文間の接続関係と読み

植田 禎子<sup>†</sup> 齋藤 豪<sup>††</sup> 奥村 学<sup>††</sup><sup>†</sup>東京工業大学大学院 総合理工学研究所 <sup>††</sup>東京工業大学 精密工学研究所

## 1 はじめに

文章は何らかの意味内容を伝えるために構成される。その際、同じ意味内容でも様々な文章の形で表現することができる。しかしその表現の仕方によっては、文章は読みやすくなり、また読みにくくなる。このような読みやすさ・読みにくさは、一般に可読性という用語で表現される。

可読性の高・低を左右する要因として、過去には語彙に関する要因、構文に関する要因、意味・談話に関する要因などが研究されてきた [9][10]。本研究では、文章の可読性を左右する要因として、文章の談話構造および接続表現を取り上げ可読性との関係を探る。

談話構造とは、文章中の文又は文群間の関係による構造のことである。読み手が書き手の意図する談話構造を容易に再構築できるようにするため、その構造を陽に示すための手段として接続詞などの接続表現がある。広瀬 [2] は接続詞について次の文の大まかな姿勢を予告する指標であるとし、このような予告機能が接続詞の重要な機能であると主張している。

本研究では、いち早く読み手に文章の談話構造を理解させるという接続表現のもつ機能に着目し、実際に接続表現が読みを促進するのかを明らかにすることを目標とする。分析には、新聞記事に談話構造と文間の関係が人手で付与されたコーパスを使用する [11]。

## 2 接続表現と読み

接続表現によって可読性が変化する例をあげる。

## 接続表現があった方が理解しやすい例

細菌による感染症は抗生物質のおかげで容易に治療できるようになりました。ところが、50年ほど前から抗生物質に抵抗力を持つ耐性菌が登場しました。新しい抗生物質を開発しても、次々にそれに耐性を持つ細菌が登場し、ほとんどの抗生物質が効かないMRSA (メチシリン耐性黄色ブドウ球菌) という菌まで出てきました。

## 接続表現の有無で理解がそれほど変化しない例

また、同省はダイオキシンが出にくい1日の処理能力が100トン以上の大型炉の導入を地方の自治体に指導している。さらに、焼却炉に比べて安全性の高いガス化溶融炉など

の開発・導入のほか、昨年末にはリサイクル時代に向けた次世代炉が実用化され、従来型の焼却施設の導入は鈍化している。ここ数年は競争激化による単価の落ち込みも大きいという。

1例目の「ところが」は削除すると非常に可読性が下がる例である。ところが2例目の「さらに」は、削除してもそれほど理解が損なわれない。このように、可読性は接続表現の有無に大きく影響をうけるが、すべての文章にただ闇雲に接続表現をつければいいというものではない。どのような場合に接続表現があることで可読性が高まるのか、また不要な時とはどのような場合なのか考えていく。

## 3 談話構造コーパスと接続表現

この節では、分析対象の談話構造コーパスと接続表現について説明する。

## 3.1 談話構造コーパス

先に説明したように、談話構造とは文章中の文間の関係による構造のことである。我々が用いる談話構造コーパス [11] は、文を基本単位とし、談話構造を付与したコーパスで、修辭構造理論 [5] に基づいている。ここでは、文章の談話構造は木構造をしていると考えられており、その木構造は、2文間の係り受け関係で構成されている。係っている方の文を衛星、係られている方の文を核文と呼んでいる。このような木構造がまず段落内で構築される。その後、段落内の木構造の根同士で係り受け関係が構築され、文章全体が1つの木となる。

さらに、核文と衛星がどのような関係で係っているのかも定義されている。文間にラベル付けされる関係には、背景、並列、補足、転換、呼応、因果、例示、対比の8つの関係がある [11]。このうち、テキストの核文には転換の関係がつけられ、並列の関係は多核の関係である。

本研究では、談話構造コーパス中で、隣接している3文を選び、分析の対象とした。このとき段落をまたぐ3文は隣接しているとは考えない。しかし3文が隣接しているだけでは、必ずしも接続しているとは限らない。3文の中に談話上の境界があったり、直前直後の文には係らない文もあるからである。そこで、本研究ではコーパス中で隣接している3文の中から3文中

の1文を核文とし、核文以外の2文の係り先が3文内にある3文に限定する。言い換えると以下のようになる。

- 同じ段落の連続している3文、かつ
- 段落の核文を含む、かつ
- 核文以外の2文の係り先が当該の3文の中にある3文。

### 3.2 接続表現

接続表現は、文頭に出現し先行する文との接続関係を示す表現で、主に接続詞として出現するが、接続詞以外でも文頭で文の接続関係を示す表現は存在する。そこで接続詞以外で、接続表現と考えられるものについて説明する。

まず、コーパスを目視し、接続詞以外で文を接続していると考えられる表現を調査した。図1に、コーパスに出現した接続詞以外の文頭の接続表現の品詞カテゴリを示す。

- 指示、参照表現を含むもの
  - 連体詞
  - 代名詞
  - 副詞
- 副詞

図1: 文頭の接続表現

指示・参照表現を含むものは、いわゆる「コ・ソ・ア・ド」が語頭につく語と、そのほかに、「同」(e.g. 同日)、「両」(e.g. 両日)などを含む表現がある。しかし、接続詞と違い指示表現のすべてが文章の接続を示しているわけではない。単に、先行する単語を参照しているだけのものもある。

指示詞はその修飾形態によって分類することができる。「これ」「ここ」「こちら」の名詞形態や、「この」「こんな」「このような」「こういう」の名詞修飾形態、「こう」「こんなに」「このように」「こんなふうに」の述語修飾形態がある[6][1]。しかし修飾形態による分類では、全く意味の違う表現も同じに分類されてしまう。

例えば、「ここで・そこで・あそこで・どこで」は、表面上は似ている語だが、「そこで」だけ意味合いが違い、接続詞として用いられる語である。また「この日」、「この学校では」、「このため」が同じカテゴリに分類されてしまうが、「このため」は実際には接続詞に近い用いられ方をしている。このように、あるカテ

ゴリ内でも、接続詞とほぼ同様の用いられ方をしている表現とそうでない表現がある。

そこで、修飾形態による分類ではなく、その語の用いられ方に注目して、どのような文脈でも接続を表現していると考えられる指示詞を接続表現として本研究では採用した。採用した表現、しなかった表現を図2、図3に示す。

- 「この+名詞」のうち
  - このうち
  - このため
  - このところ
  - そのうち
  - そのため
  - その上で
  - その後
- これに対し
- こうしたことから
- そこで
- それでも
- それゆえ
- それなら
- それだけに
- そうでないと

図2: 採用した表現

- ここ  
(e.g. ここ数年、ここまで)
- これ  
(e.g. これは、これを、これが)
- この+普通名詞
- 指示表現が含まれる文中で、その指示表現が主語になっているもの  
(e.g. こうした実情は、その後は)

図3: 採用しなかった表現

上の指示表現中、「そこで、それでも」は形態素解析器である茶筌[8]では、接続詞と分類される。また「それゆえ」は同じく形態素解析器であるJUMAN[7]では、接続詞と分類され、「それだけに」は茶筌とJUMANの両方で接続詞と分類される。

### 4 文の接続関係と接続表現

読みにおける接続表現の働きは、今までの文に対して、次の文がどのような意味を持つのかを示すことである。読み手は、これまでの文と今読んでいる文がどのような関係にあるのか探りながら読み進めていく。そのとき、今読んでいる文がこれまでの文章に対しど

のような位置づけにあるのかを読み手が早くつかめると、文章として可読性が高まると考えた。

そこでまず、どのような関係の時にどのような接続表現が出現しやすいのかコーパス中で調べた。コーパスにラベル付けされている関係ごとに、どのような接続表現が用いられるのか下に示す。

- 背景  
接続表現は出現しない。
- 並列  
「また」や「さらに」の表現が出現する。
- 補足  
接続表現は全く出現しない。
- 因果  
係り先に「このため」などの表現が出現することがあるが少ない。
- 例示  
「例えば」は入ることがあるが少ない。
- 呼応  
接続表現は出現しない。
- 対比  
係り先に「しかし」「ところが」「ただ」などの表現が出現する。

この中で、並列の関係に含まれている「また」、「さらに」や例示の関係に含まれている「例えば」は出現頻度が低い。つまり、並列や例示の関係であっても必ず接続詞が出現するわけではない。それとは対照的に、対比に出現する接続詞「しかし」は、関係中での出現頻度が高い。「しかし」以外の接続表現として「ところが」や「ただ」なども出現する。「ただ」は、前の文に対して制限を加える時に出現する接続詞であるが、それまでの文章とは、相反する情報を加える文につけられるという意味で、逆接の接続詞と併せて反予想型の接続詞とまとめられることもある [4]。

接続表現が出現する頻度の高い関係は、その関係の文がそれまでの文脈から予測することが難しい関係であると考えられる。ここで接続表現が頻出しているのは、対比の関係である。背景、並列、補足、因果、例示の関係の文がそれまでの文の意味内容に沿った情報を加えるのに対し、対比の関係の文は前の文とは反する内容を持つ文である。読み手は、これまでの文を受け入れた上で次の文を理解しようとするため、逆の意味を持つ対比の関係の文では、文頭に何らかの表現をたして接続を示さないと理解が難しくなってしまうと考

えられる。そのため、対比の関係の文では、接続表現の「しかし」などの表現が頻出し、その文が前の文と逆のことを説明している文だということを読み手に知らせていると考えることができる。結果、前の文と同じような意味内容を付け加える並列のような関係に比べて、接続表現が付加されることが多いと考えられる。

しかし、対比関係でも「しかし」などの接続表現が出現しないこともある。そのような場合でも、「しかし」などの接続表現を付加した方が読みやすくなるのだろうか。もともと逆接の接続詞を含む文では、その接続詞の必要度が高いのは確認されているが [3]、もともと接続詞を含まない対比関係に関しては接続詞の必要度が不明である。コーパス中で対比の関係で接続表現が出現しない文章には、複文の場合が多く、従属節に接続助詞の「が」を伴っている。この場合、「が」の手前までは前の文と順接の関係であり、「が」の後ろが逆接の関係となる。そのため、文頭には「しかし」を伴うことができないのではないかと考えた。

そこで対比の関係における接続表現の要・不要を調べるため、対比の関係をもつ文章において、接続詞「しかし」を削除し、文章の可読性に与える影響を調べた。また、もともと接続詞を含まない文章に対しても「しかし」を加えることで、可読性が変化するかどうか調べた。

「ところが」「ただ」を含む文も同じような役割を果たすと考えられたので、「しかし」と同様に実験を行った。これらはすべて、前の文と対比的な意味を持つ文に付け加えられる接続表現だが、対比の関係には「その後」という接続表現も出現していた。「その後」という接続表現が対比の関係においてどのような役割を果たしているのか調べるため、「その後」を含む文章も実験の対象とした。具体的な手法は次に述べる。

#### 4.1 実験

対比の関係において接続表現のある・なしで可読性が変化するか調べるため、両方のパターンを被験者に呈示して、どちらが読みやすいと感じるか二者択一で回答してもらった。またその接続表現の必要度を5段階スケールで回答してもらった。もともと接続表現を含む文章に対しては、原文章とその接続表現を削除した文章の2パターンを作成した。接続表現をもともと含まない文章に対しては、原文章と対比の関係の係り先に接続詞「しかし」を挿入した2パターンを作成した。

呈示した文章は対比の関係を含む文章から、接続詞「しかし」、「ところが」、「ただ」を含む文と接続詞を含まない文を選択して用いた。内訳は、接続詞「しかし」を含む文を9例、接続詞「ところが」を含む文を

2例、接続詞「ただ」を含む文を2例、「その後」を含む文を1例選び、実験の素材とした。また、接続詞を含まない文は18例であった。そのうち、接続助詞「が」を含む文が9例、含まない文が9例だった。計32の例文を用意した。

もともと接続詞を含んでいる文章とそうでない文章はランダムな順で呈示した。また接続詞のあるパターンとないパターンについてもバランスした。被験者は大学院生6人、大学4年生1人の計7人で、全員日本語を母国語とする。実験には約30分ほどかかった。

## 4.2 結果

原文に接続詞「しかし」「ところが」「ただ」「その後」がついていた14例は、13例で7人中5人以上が接続表現が必要であると回答した。1例で7人中3人しか必要であると回答しなかったが、対比関係が薄く、順接と解釈できなくもなかったのが原因と考えられる。原文に接続表現を含んだその他の文はすべて、接続表現が必要だという結果となった。

原文に接続表現がなかった18例では、接続表現が7人中5人以上が必要であると回答したのは、6例であった。これらの例では、原文章になかった接続表現が合った方が良かったことになる。

一方で、「しかし」が必要であると回答したのが5人未満であった例文も12例あった。そのうちの9例は複文で、従属節に接続助詞の「が」を伴っていた場合であった。予想通り、従属節の「が」は、文頭の接続表現「しかし」とは相容れないことがわかる。

ところが、接続助詞の「が」を伴っていないにも関わらず、接続表現が必要ないという結果となった文章が3例あった。これらはすべて、文頭に時を表す表現を伴っていた。時を表す表現は3節では接続表現とは見なさなかったが、実際には接続表現と同じような動きをしていることがわかる。これは、「しかし」「ところが」「ただ」といった接続表現とは一線を画すが、実験により必要性が高くなった「その後」が時を表す接続表現であることも関連するのかもしれない。

## 5 まとめ

本研究では、接続表現が読みに与える影響を調べた。ある文ともう一方の文が対比関係にあるとき、それを明示する「しかし」などの接続表現を加えることで、可読性が向上するか調べた。具体的には対比の関係で接続している文章で、接続表現のある・なしで可読性が変化するかどうか被験者実験を行った。

接続詞「しかし」を含む文章に対し、原文章と「しかし」を削除した2パターン、もともと接続表現を含まない文章に対しては、原文章と対比の関係の係り先の文に「しかし」を付加した文章の2パターンを作成

し、被験者に接続表現があった方がよいか、ない方がよいか二者択一で回答してもらった。結果、もともと接続表現があった文章では14文中13文で、あった方が読みやすいという回答が得られた。もともと接続表現がなかった18例中6例で接続表現が必要であるという結果が得られた。必要なかった例のうち9例は接続助詞の「が」をともなっていた。それ以外の3例は時を表す表現が文頭に出現していた。時を表す表現も接続表現となりうるということがわかった。

対比の関係以外の関係では必ず出現する接続表現はなかったが、必ず出現しなくても出現しているからにはなんらかの意味があるはずである。それらの接続表現と読みの関係についても調べる必要がある。また、時を表す表現は文章の接続関係を強く示していることがわかったが、すべての表現が接続表現となりうるのか明らかではない。今後詳しく調べる必要がある。

## 参考文献

- [1] 橋本 さち恵, 日本語文生成における照応表現の選択に関する研究, 東京工業大学大学院修士論文, 2001.
- [2] 広瀬 正宣, 文章における接続形式について, 井上和子(編), 昭和58年度文部省科学研究費補助金(特定研究1)報告書「明確で論理的な日本語の表現(中間報告)」, 1984.
- [3] 伊藤俊一, 阿部純一, 文章理解における接続詞の働き, 心理学研究, 59-4, 1988.
- [4] 伊藤俊一, 阿部純一, 接続詞の機能と必要性, 心理学研究, 62-5, 1991.
- [5] Daniel Marcu, The theory and practice of discourse parsing and summarization, The MIT Press, 2000.
- [6] 益岡隆志, 田窪行則, 基礎日本語文法-改訂版-, くろしお出版, 1992.
- [7] 松本裕治, 黒橋慎夫, 宇津呂武仁, 妙木 裕, 長尾 真, 日本語形態素解析システム JUMAN 使用説明書 version 2.0, NAIST Technical Report, 1994.
- [8] 松本裕治, 北内啓, 山下達雄, 平野善隆, 日本語形態素解析システム『茶筌』version 2.0 使用説明書, NAIST Technical Report, 1999.
- [9] 松岡正男, 文構造に着目した日本語文の理解しやすさ・しにくさの指標について, 京都大学大学院修士論文, 1996.
- [10] 山本聡美, 乾健太郎, 乾裕子, 聾者向け読解支援のための文可読性基準のモデル化, 情報処理学会自然言語処理研究会, NL-141-21, 2001.
- [11] 横山憲司, Support Vector Machine を用いた談話構造解析, 東京工業大学大学院修士論文, 2003.