

主張文・根拠文の配置とレポート評価との関係について

伊藤 俊一

情報教育講座

Report Scoring Based on Handling of Opinion and Fact Statements

Toshikazu ITO

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

1. はじめに

授業の一環としてレポート課題を課した場合、教員はその後、提出されたレポートに対する客観的で妥当な評価を下すことを求められる。評価の客観性と妥当性を確保しつつも、評価者の負担を軽減することが可能となるように、現在も様々な自動評価システムの開発研究が進められている(石岡(2008), など参照)。そのような自動評価システムの精度をより高めるためには、レポートの良し悪しを判定するための手がかりとして有効、かつ、容易に利用できる属性を幅広く見定めていく探究がさらに継続的に必要であろう。

本研究では、授業におけるレポート課題で課されることの多い「自身の意見を論じる」というタイプのレポートを分析の対象とする。そして、それら「自身の意見を論じる」タイプのレポートには当然ながら含まれているであろうことが想定される主張文と根拠文がどのようにレポート内に配置されているかということと、最終的に評価者がそれらのレポートに対して下す評価との間には、いかなる関係が存在するのかについて検討することを目的とする。主張文・根拠文の配置とレポート評価との関係を明らかにすることで、自動評価システムにとって必要な、レポートの良し悪しを左右する新たな手がかりを見出すことができるのではないかと考える。

長谷川・堤(2011)は、「自身の意見を論じる」タイプのレポートにおいて、主張文と根拠文が互いに意見とその論拠という関係性をもって対応付いているはずであるというある種の「ひな形」を評価者は予め持つと考えている。そして、その「ひな形」からの逸脱が生じているレポートにおいては、評価者の読解に要する負担は増大し、その負担が大きければ大きいほど分かりにくいという評価が下されることになる論じている。

その上で、長谷川ら(2011)は、主張文と根拠文の対応付けという観点から、レポートが分かりにくいと評価される要因を次の2点にまとめた。

- a. 主張文のサポートとなる根拠文がない、もしくは、不足している。
- b. 主張文そのものがない、もしくは、不足している。

つまり、長谷川ら(2011)は、主張文と根拠文のいずれが不足しても、それがレポート全体への評価に影響を及ぼし、最終的には分かりにくいという評価が下されることになると考えた。

本研究では、大学で実施された授業においてレポート課題として受講者から提出され、授業を担当する教員によって評価されたレポートを用い、長谷川ら(2011)の指摘する主張文、あるいは、根拠文の不足がレポート評価に及ぼす影響を統計的に検証するとともに、主張文と根拠文の距離といった新たな要因についても、それらがレポート評価に与える影響を検証することを試みる。

2. 方法

レポート・データの収集:

A大学の大学1年生40名を対象とした初年次教養科目「レポートライティング」の初回の授業において受講者は次の課題を課された。

「最近1カ月の新聞記事の中から情報に関連する記事を1つ選び、その内容に対する自分の意見を1,000文字程度で論じなさい。」

その1週間後に提出された第1稿のレポート40編を、授業を担当する教員2名が合議によって5段階(S・A・B・C・D)で総合的に評価した。評価は、それぞれ100点満点における90~100点(S), 80~89点(A),

70～79点 (B), 60～69点 (C), 0～59点 (D) に概ね対応している。これらの評価は、当該授業の単位認定に関わる成績評価を目的として実施されたものであり、最終的には受講生本人に開示されたものである。2名の合議を経た評価結果の内訳は、S評価7編、A評価4編、B評価10編、C評価10編、D評価9編であった。

これらのレポートのうち、S評価とA評価を合わせた11編のレポートを高評価群、D評価9編のレポートを低評価群とし、それら20編のレポートを本研究における分析対象とした。

主張文の抽出：

高評価群のレポート11編、低評価群のレポート9編のそれぞれから主張文を抽出する作業を行なった。一般に、日本語の文章から主張文を抽出するためには、文末表現を手がかりとすることが有効であることが従前から指摘されている(山本・増山・内藤(1995), など)。本研究でも、文末表現に基づいて、それぞれの文が主張文であるか否かの判定を行なった。本研究で主張文を特定するための手がかりとして実際に文章中に出現した文末表現は、Table 1の通りであった。

Table 1. 主張文の文末表現

いけない	言える
いられない	考えられる
ざるを得ない	思う
ではないか	重要である
ではないだろうか	正しい
ならない	足りないだろう
はずである	大切である
べきである	当たり前である
ほしい	難しい
過言ではない	必要である
感じた	必要になってくるだろう
期待したい	不可欠である
気がする	厄介である
求められる	要注意だ
言いようがない	(立場に)立つ
言うまでもない	

根拠文の同定：

抽出された主張文のそれぞれに対する根拠となる内容を含む文を根拠文として同定する作業を行なった。

レポートの評価を行なった授業担当者とは異なる2名が、それぞれの主張文に対して根拠となる内容を含むと考えられる文すべてを選んだ。この作業は2名がそれぞれ単独で行なった。その後、2名のうち1名のみが選んだ文については、2名の合議によって根拠文として同定するか否かを決定した。両者がともに選んだ文については、そのまま根拠文として同定した。

3. 結果

主張文の出現頻度：

レポート1編あたりの主張文の出現頻度は、高評価群において平均6.00文(標準偏差1.47)、低評価群において平均4.11文(標準偏差2.02)であった(Fig.1参照)。分散分析の結果、主張文の出現頻度には群間で有意な差が認められた($F(1,18)=5.22, p<.05$)。

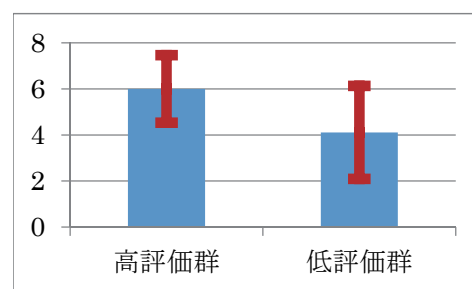


Fig. 1 主張文の出現頻度 (単位：文数)

主張文あたりの根拠文数：

それぞれの主張文に対して根拠となる内容を含む文(根拠文)の数は、高評価群において平均2.61文(標準偏差0.75)、低評価群において平均3.17文(標準偏差2.37)であった(Fig.2参照)。分散分析の結果、主張文あたりの根拠文数には群間で有意な差が認められなかった($F(1,18)=0.50, p>.10$)。

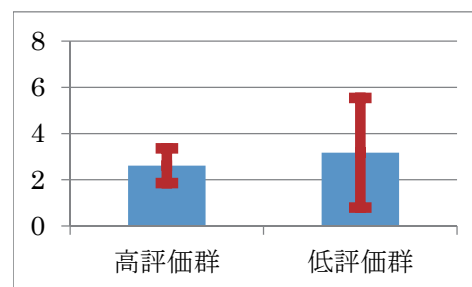


Fig. 2 主張文あたりの根拠文数 (単位：文数)

群間において標準偏差の大きさがかなり異なったことから、両群の違いは分散の差に表れていると推察し、分散の有意差検定を行なった。その結果、主張文あたりの根拠文数の分散には群間で有意な差が認められた($F(8,10)=10.28, p<.01$)。このことは、高評価群の主張文あたりの根拠文数は比較的一定であるのに対して、低評価群の主張文あたりの根拠文数はレポートごとにまちまちであることを示唆している。

両群を合わせた20編のレポートの主張文あたりの根拠文の数は平均2.86文(標準偏差1.71)であった。「平均±標準偏差」の範囲に収まらなかった4編のレポートはいずれも低評価群のレポートであり、うち2編は平均より少なく、2編は平均より多かった。

主張文と根拠文の距離：

それぞれの主張文とそれに対する根拠文との間に介在する文の数を主張文と根拠文の距離と見做し、以下の検定を行なった。

主張文と根拠文の距離は、高評価群において平均 3.72 (標準偏差 1.50)、低評価群において平均 3.42 (標準偏差 3.33) であった (Fig.3 参照)。分散分析の結果、主張文と根拠文の距離には群間で有意な差が認められなかった ($F(1,18)=0.06, p>.10$)。

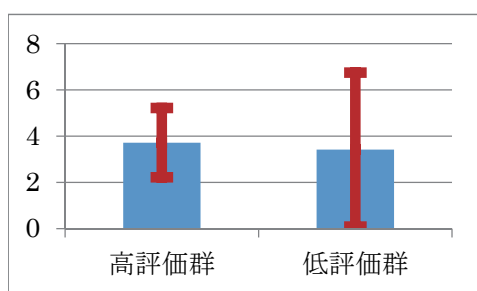


Fig. 3 主張文と根拠文の距離 (単位：文数)

群間において標準偏差の大きさがかなり異なったことから、両群の違いは分散の差に表れていると推察し、分散の有意差検定を行なった、その結果、主張文と根拠文の距離の分散には群間で有意な差が認められた ($F(8,10)=4.99, p<.05$)。このことは、高評価群における主張文と根拠文の距離は比較的一定であるのに対して、低評価群の主張文と根拠文の距離はレポートごとにまちまちであることが示唆している。

両群を合わせた 20 編のレポートの主張文と根拠文の距離は平均 3.59 (標準偏差 2.50) であった。「平均±標準偏差」の範囲に収まらなかった 5 編のレポートはいずれも低評価群のレポートであり、うち 3 編は平均より小さく、2 編は平均より大きかった。

4. 考察

本研究で得られた結果をもとに、主張文と根拠文の配置に関して高評価群のレポートと低評価群のレポートのそれぞれが持つ特徴を要約すると、次のようになる。

高評価群のレポートの特徴：

- ・主張文の出現頻度が高い。
平均 6.00 文 (標準偏差 1.47)
- ・主張文あたりの根拠文数が一定である。
平均 2.61 文 (標準偏差 0.75)
- ・主張文と根拠文の距離が一定である。
平均 3.72 (標準偏差 1.50)

低評価群のレポートの特徴：

- ・主張文の出現頻度が低い。
平均 4.11 文 (標準偏差 2.02)
- ・主張文あたりの根拠文数の分散が大きい。
平均 3.17 文 (標準偏差 2.37)
根拠文が過度に少ない(「平均-標準偏差」よりも少ない) レポートと、逆に、根拠文が過度に多い(「平均+標準偏差」よりも多い) レポートは、すべて低評価群に属するレポートであった。
- ・主張文と根拠文の距離の分散が大きい。
平均 3.42 (標準偏差 3.33)
主張文と根拠文の距離が過度に小さい(「平均-標準偏差」よりも少ない) レポートと、逆に、主張文と根拠文の距離が過度に大きい(「平均+標準偏差」よりも大きい) レポートは、すべて低評価群に属するレポートであった。

以下では、特に低評価群のレポートの特徴に着目し、レポートの評価を低下させる原因について考察していく。

まず、主張文が不足することによってレポートの評価が低下する傾向にあることは長谷川ら (2011) によっても指摘されていたことである。本研究で分析の対象としたレポートがそうであったように、授業の課題として書かれるレポートの多くが「自分の意見を論じる」ことを目的としたものである以上、主張の不足が低い評価を招くのは当然のこととも言えよう。

主張文あたりの根拠文が不足することによってレポートの評価が低下する傾向にあることも長谷川ら (2011) によって指摘されていたことと一致する。根拠文の不足は、主張を裏付ける論拠となる事実が十分には示されない事態を招き、その結果、主張自体の説得力が損なわれ、ひいては、レポート全体の評価を低下させる原因となると考えられる。

一方、本研究では、主張文あたりの根拠文が不足する場合と同様に、主張文あたりの根拠文が過剰に存在することによってもレポートの評価が低下する傾向にあることが示された。主張文あたりの根拠文が過剰に存在するレポートにおいては、最終段落に集中的に主張文が配置される重点後置型 (木下 (1994) 参照) の構成を取る傾向が強かった。そのような構成においては、最終段落以外の段落では、主張文に対する根拠として最終的には用いられることになる事実が列挙されることになる。しかし、それらの事実がどの主張文と結びつくかは最終段落までは明示されないため、読み手にとっては、書き手の意図が不明確な状態のまま多くの事実を理解し記憶しながら文章を読み進めなければならない。このような根拠文の過剰な羅列が読み手の負担を増大させ、そのことがレポートの評価を低下させる原因となったと考えられる。

本研究では、また、主張文と根拠文の距離が過度に大きい場合にレポートの評価が低下する傾向にあることも示された。主張文と根拠文の距離が過度に大きいレポートと、主張文あたりの根拠文が過剰に存在するレポートには、大きな重なりが認められた。すなわち、主張文と根拠文の距離が「平均+標準偏差」以上に大きかったレポートは、主張文あたりの根拠文数が「平均+標準偏差」以上に多かったレポートと同一のものであった。先に指摘したように、それらのレポートにおいては、最終段落以外の段落では、主張文に対する根拠として最終的には用いられることになる事実が列挙される構成が取られていた。それらの事実がどの主張文と結びつくかは最終段落までは明示されないため、読み手にとっては、書き手の意図が不明確な状態のまま多くの事実を理解し記憶しながら文章を読み進めなければならない。さらに、そのような文章構成においては、最終段落に集中的に配置される主張文と、それ以外の段落に分散する根拠文の距離が大きくなりがちであり、読み手にとっては主張文と根拠文の適切なペアリングを行なうことにも支障を来すことになると考えられる。

また、本研究では、主張文と根拠文の距離が過度に大きい場合と同様、過度に小さい場合にもレポートの評価が低下する傾向にあることが示された。主張文と根拠文の距離が過度に小さいレポートにおいては、「根拠文→主張文」というペアリングが局所的に成立していて、なおかつ、「根拠文→主張文」のペアの繰り返しが多数列挙される、という構成を取る傾向が強かった。例えば「根拠文 a→主張文 a→根拠文 b→主張文 b→根拠文 c→主張文 c…」というような構成が多く認められた。このような構成においては、個々の主張文に対する根拠となり得る事実が、その都度、直前に示され、それによって一つ一つの主張に対する論拠は担保されるものの、それぞれの主張は断片的な事実（根拠文）から条件反射的に導き出されたものになりがちであり、レポート全体を通しての主張の一貫性は乏しくなる傾向があった。例えば「根拠文 a→主張文 a→根拠文 b→主張文 b→根拠文 c→主張文 c…」という構成における主張文 a, b, c の間には強い関連性が認められないような例も見受けられた。そのような構成では、レポート全体の結論を明確にすることが難しく、そのことがレポート全体の評価を低下させる原因となり得ると考えられる。

5. 総合的考察

本研究では、レポートの評価を低下させる原因として、「主張文の不足」、「主張文あたりの根拠文の不足」、「主張文あたりの根拠文の過剰」、「主張文と根拠文の

過度に大きな距離」、「主張文と根拠文の過度に小さな距離」が示唆された。

これらの要因だけでレポートの評価が一意的に決まるわけでは勿論ない。しかし、評価者がレポートを精読する前に、これらの要因に基づいてある程度の仕分けをしておくことで、評価者の負担が軽減される可能性があると考えられる。例えば、主張文と根拠文の距離が過度に小さいレポート、あるいは、過度に大きいレポートは、その主張の意味的な内容に関わらず、問題のある可能性が高いレポートとして、さらなる精査を行なうための優先度を上げておく、というような作業工程を組むことが可能であろう。

さらに進んで、これらの要因に基づく評価を自動的に行なうことを想定するときには今後の課題となるのは、主張文と根拠文との対応付けの自動化の問題であろう。主張文、あるいは、事実文をそれぞれ単独で同定することは、文末表現等の表層的な手がかりを用いて、ある程度、可能である（山本・増山・内藤（1995）、など）。しかし、主張文とその論拠となる事実文を対応付けるには、表層的な手がかりに加えて、意味的な手がかり、例えば、因果関係の存在を示す手がかり、等がさらに必要となるであろう。

引用文献

- 長谷川哲子・堤良一（2011） アカデミックライティングにおける「分かりにくさ」の要因は何か？—意見文の分析を通じた一考察— 大阪産業大学論集（人文・社会科学編），11，21-34.
- 石岡恒憲（2008） 小論文およびエッセイの自動評価採点における研究動向 人工知能学会誌，23(1)，17-24.
- 木下是雄（1994） レポートの組み立て方 ちくま学芸文庫
- 山本和英・増山繁・内藤 昭三（1995） 文章内構造を複合的に利用した論説文要約システム GREEN 自然言語処理，2(1)，39-55.

（2018年9月14日受理）