

オンライン文章要約の機序解明に向けて：段落分けの検討 Toward understanding of on-line text summarization: paragraph detection task

布山 美慕[†], 日高 昇平[†]
Miho Fuyama, Shohei Hidaka

[†] 北陸先端科学技術大学院大学
Japan Advanced Institute of Science and Technology
m-fuyama@jaist.ac.jp, shhidaka@jaist.ac.jp

概要

文章理解の機序解明に向け、著者らは読者が読みながら文章情報のある意味のまとまりごとに圧縮し保持する“オンライン要約”に注目した。このオンライン要約の基礎的な過程の一つは意味のまとまりの範囲の推定である。本研究ではこれを意味的段落分けとして、読者が意味的段落分けする際に利用する情報について調べた。実験の結果、意味的段落分けは当該段落内部の情報のみでも一貫性をもって可能なことが示唆された。この結果は、意味的段落分けが、段落の意図の記述の必要十分性の認識によってなされるという本研究の仮説に整合的である。

キーワード: 文章理解, 段落分け, オンライン文章要約

1. 文章理解におけるオンライン要約

人はどのように文章を理解しているのだろうか。文章の理解とはどのような情報の処理で、その“理解”は読み手においてどのように表現されているのだろうか。著者らはこの文章理解の認知の解明に向けて、読みながら行われる要約(以下オンライン要約)に注目した。本研究では、このオンライン要約の認知機構を明らかにする一歩として、読者がどのように文章の意味的なまとまりを見出していくのか(以降「意味的段落分け」と呼ぶ)を調べる。

多くの読者は、あるまとまった文章を読む際、それまでに読んだ一語一文を全て覚えているわけではない。しかし、文章を全体として理解するには、それまでに読んだ文章を何らかの形で記憶しておく必要があるだろう。従って、人は情報を圧縮・保持する一種の“要約”を行いながら、文章を読み進めていると考えられる。著者らは、こういった“オンライン要約”による意味的な段落分けは、読者の文章の理解がどのように/どのような情報として表現されているかを示唆し、文章理解の認知過程を解明する要点となると考えている。

オンラインの要約は次のような課題だと解釈できる。まず、制約条件として、文章全体に比べて、一度には少ない量の文章しか認識できず(情報の入力之窗が小さい)、記憶容量が文章全体をそのまま記憶するのに必要な容量に比べて非常に少なく(保持できる情報が少ない)、それ以降の文章の内容が読者にとって不明である。オンライン要約は、この条件下において、それ以降の文章理解に長期的に有用な情報を構築し保持する課題である。文章要約については、文章の全体が与えられた際の自動要約の研究が盛んに行われているが(レビュー的な文献として[1, 2]など)、少なくとも先の文章が不明な条件で、それ以降の読解に有用となる情報を構築するという点が新規な課題である(自動要約の場合は、人の認知容量の制約はないため、さらに情報入力のサイズや記憶容量の問題が大きく緩和される)。

2. 意味的まとまりの判断

このオンライン要約の認知過程を明らかにするための一歩として、本研究では文章の意味的なまとまり判断(意味的段落分け)の認知を調べる。オンライン要約を行うためには、要約の単位として、読者は読みながら文章の意味的なまとまりを意識的・無意識的に判断すると考えられる。前述の通り、読者はその認知容量の制約に合わせてある文章範囲ごとに要約を行うと考えられる。このとき、要約は要約内容に合わせた文章の範囲ごとに構築されると考えるのが自然である。この文章の範囲=ある情報で要約される意味的なまとまりの範囲をどう取るかと、要約の内容は相互依存的である。したがって、オンライン要約の観点から考えると、意味的まとまりの判断は読者の文章理解を反映する基礎的な認知過程である。

こういった読者の意味的まとまり判断に関しては、先行研究でも、文章の意味的な分割が読者間で一貫性をもってなされるとして報告されてきた(3節でより

詳しく述べる)。たとえば, Zacks ら (2009) [3] は, event segmentation の一種として, 読者間で一貫性のある文章分割がなされることを確かめ, これが“出来事”のような意味的単位と見なせると論じた。また, この Zacks ら [3] も参照しているように, 文章理解の研究でも, 読者は状況モデルと呼ばれる現実世界に類似する物語世界のモデルを構築しており, このモデルが大きく更新される際に読解時間が延びるとされる [4, 5]。これはモデル更新までの範囲を意味的なまとまりと見なす仮説と解釈できるだろう。以上の知見は, 読者がなんらかの意味的なまとまりを文章理解の際に見出している可能性を示唆する。ただし次節で議論するように, これらの先行研究の“意味的なまとまり”が, 本研究で仮定する“要約”が対応する意味的単位と同じ対象を指すかは明らかでない。

3. 意味的段落分けに関する本研究の仮説

本研究では, 読者の文章の解釈に依存する最小の意味的なまとまりとして「意味的な段落」を考える。単語や文のまとまりの単位が比較的明確なのに対し (少なくとも現代の日本語や英語では文は終端記号を持つ), いくつかの文からなる意味的な“段落”はそのまとまりが必ずしも文中に明示されず, 前述の通り読者の文章解釈に依存して可変である。本研究の仮説では, この意味的な段落を最小の要約に対応する意味単位と考え, よって, 要約に相当するなんらかの文章の“意図 (言いたいこと)”を必要十分記述している単位が意味的段落であると考え。

著者らはこの意図の記述の必要十分性は, より大きな単位では, (良い) 小説全体を読んだときに, きちんと話が“終わった”感じを受けることと対応していると考えている。小説を読了してすっきりと終わった感じがしない場合, 伏線が回収されないなど, 作品の意図に対し必要十分な記述がなされていないと読者が感じる事が要因と考えられる。一方で, 作品の意図が必要十分記述された小説では, 最後まで読むと物語が“終わった”感じを感じることができであろう。これと類似で, 本研究の仮説では, 段落の単位でもこういった意図が十分述べられた～“終わった”感じの判断がなされ, これが意味的段落分けの判断と考える。

一方, 文章分割に関する先行研究の多くは, 段落内外の意味的類似性の差 (段落内では意味的類似性が比較的高く, 段落間では比較的低い) を手がかりとした意味的段落分けを主張してきた。たとえば, Zacks ら [3] は, 文章理解の一つのモデルである Event-Indexing Model [5] を参照し, Cause, Character, Goal, Object,

Space, Time の6つの要素が大きく変化する地点で読者は文章分割を行うと予測し, これを実験から部分的に確かめた。これは, 段落内外でこれらの要素の差が大きいことを手掛かりに意味的段落分けをすとの仮説だと見なせる。また, 自然言語処理の分野でも, Hearst (1997) [6] をはじめとして, 段落の境界で文章の類似度が変化する (Hearst (1997) では境界前後の特定の数の単語集合間の類似度が大きく変化した地点を段落の境界とみなす) と考え, 文章の自動分割を行ってきた。

本研究の仮説は, 読者は各段落の意図が必要十分書かれたか否かを基準に段落分け可能と予測する。よって, 類似度の変化の情報がなくとも段落分け可能と考える (本研究の仮説では類似度の変化は段落ごとの作品の意図の変化に伴う副次的な変化だと考えられる)。本研究では, 本仮説検証の第一歩として, 類似度変化の情報がなくとも読者はもとの作品の段落分けを認識できるか, 読者間で共通した段落分けが行われるかを調べた。

実験1では, 検証の前段階として, 元の作品の段落分け情報を除いた際に, 読者がどの程度元の作品の段落分けを推定可能か調べた。実験1では読者には作品の全文を提示し, 段落分けをしてもらうため, 読者は段落分けにあたり内容の類似度変化の情報も利用できる。実験2では, 被験者に一文ずつ作品の本文を提示し, その文で段落分けされると思う程度を5件法で回答してもらった。この実験では, 読者は当該段落内部の情報のみで意味的段落分けを行うことになる。このとき, 読者は次の段落の文を参照せずに段落の終端を予測する必要があり, 類似度変化の情報を利用できない。なお, 本研究は進行中の研究であり, 実験1・2ともに予備的な実験である。そのため, 被験者はそれぞれ7名と9名と少数である。

4. 実験1

4.1 実験方法

改行情報 (作品の元の段落分け情報) を削除した10作品 (小説7作品, エッセイ・論述3作品。表1に作品名を記載) の各作品冒頭900~2200文字を7名の被験者に読んでもらい, 意味的な段落の切れ目を報告してもらった。被験者は日本語を母語とする大学院生および研究者 (男性6名, 女性1名) であった。10作品分を印刷した紙を渡し, 読み返しや先読みも自由でできる状態で, 自由な速度で読み, 時間制限なく意味的な段落の切れ目を本文中に書き入れてもらった。

表 1 実験結果 1: 検出能力を表す d-prime (表中は d' 表記) の大きい順 (降順) に結果を並べた. 開業の頻度が少ないものほど d-prime は高くなる傾向がある. また, 作品 No.1,2,5 は随筆, それ以外は小説で, 随筆の d-prime が高い傾向がみられる.

作品 No	作品名	P(正検出)	P(誤検出)	d'	元作品段落分け頻度
1	回顧と展望	0.61	0.08	2.43	0.14
2	趣味の遺伝	0.40	0.04	2.01	0.13
3	桐壺	0.64	0.15	1.92	0.07
4	細雪	0.32	0.01	1.85	0.24
5	家の話	0.57	0.12	1.81	0.14
6	冗談に殺す	0.28	0.05	1.03	0.36
7	銀河鉄道の夜	0.29	0.05	1.03	0.35
8	城	0.23	0.04	1.02	0.23
9	ラプンツェル	0.22	0.05	0.84	0.31
10	レ・ミゼラブル	0.26	0.08	0.83	0.19

4.2 実験結果・議論

結果を表 1 にまとめた. 被験者の意味的段落分け報告を集計し, 被験者 7 名分の結果から, 元の作品で段落分けがある箇所では被験者が段落分けありと判断した確率 (P (正検出)), 元の作品で段落分けがない箇所では被験者が段落分けありと判断した確率 (P (誤検出)) を算出した. そして, これらの値を元に信号検出理論で用いられる信号 (本実験では元の段落分け) の検出能力を表す d-prime の値を算出した. d-prime は被験者にとっての信号+ノイズの分布の平均値からノイズの分布の平均値を引き, これをノイズの分布の標準偏差で割ったもので, ノイズの標準偏差を単位として信号+ノイズとノイズの分布がどの程度離れているかを示す. よって, たとえば, 最も d-prime が大きかった作品 No.1 の $d-prime = 2.43$ は被験者が平均的に信号+ノイズとノイズをノイズの標準偏差の 2.4 倍程度離れた分布として識別可能であることを示唆する.

以上の分析の結果, 表 1 中の d-prime が算出された. d-prime が 0.83~2.43 であることから, 信号とノイズは被験者にとって十分識別可能であると推測される. また, 作品の元の段落分け頻度と d-prime の間には相関係数-0.72 程度の負の相関があり, 段落の数が多ほど段落分けの予測が難しくなっていることを示唆する. これは, 段落分けが少ない場合, 大きな意味的变化を反映した段落分けであるため, 被験者が推測しやすかった可能性を示唆する. 加えて, 作品 No.1, 2, 5 は随筆・論述文であり, 小説に比べて文章の展開が予測しやすいと考えられる. その結果, d-prime の

値が比較的高くなっている可能性が考えられる.

まとめると, 作品の種類や改行の多さに応じて差はあるものの, 読者は総じて文章から元の作品の段落分けを推定可能であることが示唆された.

5. 実験 2

5.1 実験方法

夏目漱石『こころ』の冒頭 172 文を, 一文ずつ改行し, 被験者に一文ずつ提示しながら各文で段落が終わると考える程度を各文ごとに 5 件法 (5 がその文で段落が終わると思う程度が最も強い) で評価してもらった. 該当文の評価時に前の文を読み返すことは許可したが, 該当文の評価を次の文以降を読み直して修正することは許可しなかった. 読む速度は被験者の自由で, 課題全体の時間制限も行わなかった. 被験者は 9 名で, 日本語を母語とする大学院生および研究者 (男性 7 名, 女性 2 名) であった. うち, 1 名の被験者が『こころ』を読んだことがあったが, 冒頭部分の展開をはじめ作品についてあまり記憶しておらず, 実験への影響は少ないと判断した¹.

5.2 実験結果・議論

表 2 に被験者ごとの正検出/誤検出の確率および d-prime の値を示した. 表では d-prime の値が降順になるように結果を示した. この結果から, 被験者によって差はあるものの, 上位 3 名はノイズの 1 標準偏差程度

¹実際, 実験で特に目立って精度良く作品の元の段落を予測することはなかった.

でノイズ+信号とノイズを弁別している。実験1の結果と比べると、全体的に実験1よりも検出力は落ちるが、上位4名程度は実験1の下位程度以上の検出力をもって段落分けを判定した。また、表2の一番下の行に9名分の平均値を記した。作品が異なるので単純比較はできないが、d-primeの平均値は、実験1の最低値に近い値となった。

また、被験者間の意味的段落分け評価値の相関係数を各被験者間で求めた。d-primeが作品の元の段落分けを正解データとした検出力を示すのに対し、被験者間の意味的段落分け評価値の相関は、被験者間の意味的段落分け評価の一貫性を示す。この分析結果として、全ペアの相関係数をまとめ、図1に相関係数のヒートマップとして示した。相関係数は $-0.06 \sim 0.42$ 、相関のp値が0.05以下のペアは36ペア中27ペアと、75%のペアにおいて5%有意水準で有意な相関を示した。

この結果は、d-primeが低い被験者でも他の被験者と高い相関係数を持つことから、実際の作品の段落分けとは異なる作品の意味的段落を被験者が共通して推定した可能性を示唆する。これは、読んでいる文が含まれる段落に関して、被験者がなんらかの共通の手がかりを用いて、作者の意図とは異なる意図を推定し、意味的段落分けを行った結果とも解釈可能である。

表2 実験結果2:d-prime (表中はd'表記)の大きい順(降順)に結果を並べた。

被験者	P(正検出)	P(誤検出)	d'
1	0.54	0.15	1.47
2	0.17	0.01	1.23
3	0.46	0.16	1.12
4	0.21	0.05	0.84
5	0.33	0.16	0.63
6	0.33	0.17	0.60
7	0.38	0.21	0.56
8	0.08	0.03	0.44
9	0.25	0.15	0.39
average	0.31	0.12	0.81

6. 総合議論

本研究では、オンライン要約の認知過程解明に向けて、文章の意味的段落分けが最小の要約相当の文章の“意図”が必要十分書かれる単位でなされるという仮説の部分的な検証を行った。具体的には、意味的段落内部の情報のみで、意味的段落分けが可能であるかを2つの実験から検証した。実験1では、読者が段落情

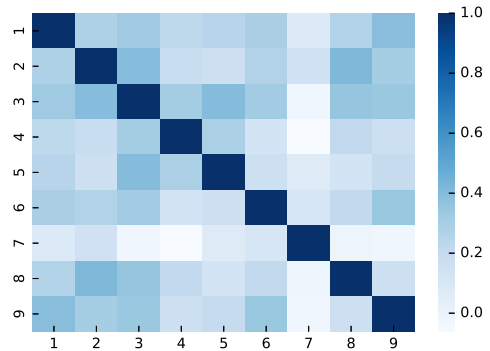


図1 被験者9名の1～5段階の段落分け評価値の相関係数のヒートマップ。被験者1～9の順でd-primeが高くなるよう並べた。

報を落とした文章から元の段落分けをどの程度推定可能か調べた。その結果、信号検出理論のd-prime=0.83～2.43程度の十分な精度で作品の元の段落を推定可能であった。実験2では、読者が段落内部の情報のみで、意味的段落分け可能か調べた。その結果、実験1よりも精度は落ちるものの作品の元の段落分けが推定可能であること、被験者間で有意に相関した意味的段落分け評価がなされることが示唆された。

以上の結果は、読者は段落内部の情報のみでも意味的段落分けが可能であることを示唆する。可能という意味は、元の作品の段落分けを推定できること、およびなんらかの一貫した文章の手がかりを用いて他の読者と共通した段落分け評価をなせることを指す。実験1よりも実験2の方がd-primeの値が低いことから、通常、読者は該当段落よりもあとの文章の情報も利用して意味的段落分けを行なっていることを示唆する。

これらを総合すると、先行研究が主張するように、なんらかの段落内外の差も意味的段落分けに利用するが、段落内部の情報のみでも十分一貫性のある意味的段落分けが可能であると考えられる。この結果は、本研究の、読者がなんらかの文章の意図の推定から、意味的段落分けを行っているとの仮説に整合的である。

著者らは、オンライン要約の認知過程解明に向けて、本仮説検証のための今後の課題として以下を考えている。まず、意味的段落分けに利用可能な段落内部の情報として、本研究の仮説が主張するような段落全体の“意図”情報のみならず、多くの情報が考えられる。たとえば、局所的な段落最後の文の文末や接続詞など特徴的な一語・二語によって、読者が段落の終端を予測している可能性がある。よって、意味的段落分けに際

し、こういった段落内部の局所的な情報にどの程度依拠しているか、段落全体が示唆する大局的な“意図”がどの程度利用されているかを分けて調べる必要がある。このためには、文をシャッフルしたり（大局的な情報を弱める）、接続詞や文末の表現を変える（局所的な情報を弱める）などして読者の意味的段落分けがどの程度変化するか調べる等の方法が考えられる。また、テキストの局所的な情報から段落分けを推定するモデルを、コーパス等から学習し、局所的な情報のみで（原理的には）どの程度段落分けが推定可能か調べることも可能であろう。加えて、より本質的には、段落全体の“意図”がどのように文章から推定され、どのように認知過程に表現されるのかモデルを構築することが必要であると考えられる。これらの知見を元に、オンライン要約の基礎となる意味的段落分けの認知過程を明らかにし、文章理解過程の解明を目指したい。

謝辞

本研究は科学研究費補助金若手研究 A 16H05860 の補助を受けた。

文献

- [1] 奥村学 & 難波 英嗣, (2005), “テキスト自動要約”, オーム社.
- [2] 高村大地, (2017), “文章要約のための数理的手法”, オペレーションズ・リサーチ, Vol.62, No.11, pp.711-716.
- [3] Zacks, J., Speer, N., & Reynolds, J. (2009), “Segmentation in reading and film comprehension.”, *Journal of Experimental Psychology: General*, Vol.138, No.2, pp.307-327.
- [4] Gernsbacher, M. (1997), “Two decades of structure building.”, *Discourse Processes*, Vol.23, No.3, pp.265-304.
- [5] Zwaan, R., Langston, M., & Graesser, A. (1995), “The construction of situation models in narrative comprehension: An event-indexing model.”, *Psychological Science*, Vol.6, No.5, PP.292-297.
- [6] Hearst, M. (1997), “TextTiling: Segmenting text multi-paragraph subtopic passages.”, *Computational Linguistics*, Vol.23, No.1, pp.33-64.