

テキストにおける首尾一貫性(coherence)と 文脈表現構造

4D-2

内田ユリ子 石崎 俊 井佐原 均

(電子技術総合研究所)

1. はじめに

テキストは、その構成部品がある原理に従って配列されたものであるが、その配列原理の如何について広く受け入れられるような解答は存在していない。[1]

テキストを構成する部品の単位には、語、句、節、文、更にパラグラフや、章、巻なども考えられる。ここでは、これら全てを総称して、セグメントとよぶことにする。

テキストにはそれを構成するセグメントの間の統語的關係(結束性、cohesive)と共に、内容的な連関性(首尾一貫性、coherence)が存在する。

結束性が統語的な言語上の手段—例えば、代名詞による指示、接続詞や接続副詞による結び付け、類義語や類義表現による言い替えなど—によるのに対して首尾一貫性は、主として、書き手と読み手の言語外の知識や手段—例えば、現実世界に関する知識、書き手・読み手の仮定や推論など—による談話やテキストの内容のまとまりを指す。[2]

複数パラグラフよりなるテキストの場合は、1つのパラグラフ内部にはもとより、パラグラフ同士の間にも内容的な連関性が存在する。首尾一貫性は、テキスト全体に及ぶものであるが、結束性についても局所的なものばかりでなく、パラグラフを超えて働く統語制約(例えば、再参照による冠詞theの生成など)も考えられる。

従来の研究では、テキストの構成要素の間の内容的連関性に関しては、主として文や節を対象として議論されてきたが、本稿では、複数パラグラフよりなるテキストのパラグラフ間の構造的連関性にも適用できる理論として、テキストの意味表現構造と首尾一貫性の関係、および読み手の持つ情報の新旧と首尾一貫性との関係に付いて述べる。また、首尾一貫性のあるテキストの実例を示し、そのパラグラフ構成とテキストの背後にある情報構造との関係について検討し、テキスト生成との関連も述べる。

2. テキストの首尾一貫性と意味構造

テキストの首尾一貫性は、それらを構成する幾つかのセグメントの間の構造的関係の表われである。その関係はテキストの意味構造に基いている。すなわち、セグメント間の構造的関係とはテキストの意味表現の構造関係の表出にほかならない。書き手の思考内容、すなわちテキストの意味表現構造が一次的にマッピングされたテキストから、その構造関係に関する言語上の手段がテキストの中に存在しなくても、読み手が元の構造を再現できるとき、そのテキストは内容的なまとまりがある(coherent)といえる。

首尾一貫性は、テキストの中のあるセグメントとそれの1つ前のセグメントとの間の関係に付いてだけを指すものではなく、テキスト全体におけるセグメントの間の連関性を意味する。(Hobbsは、テキストの中で隣合ったセグメント同士の連関性を連結性(relevance)と呼び、首尾一貫性(coherence)と区別した。[3])

連結性は言わば局所的なものであるが、首尾一貫性は大局的なものである。テキスト生成のばあい、特に複数パラグラフよりなるテキストを生成する場合は、パラグラフ順序決定の根拠として、パラグラフ間に首尾一貫性をもたらすパラグラフ順序決定規則が必要である。連結性に依存するだけでは、パラグラフを繋いでいくことが出来ない場合もある。実際に、新聞記事などには連結性の無いパラグラフが隣合っている例がある。それらは(テキストの中の位置とは無関係に)首尾一貫性に依存して他のパラグラフと繋がっているのである。

3. テキストの首尾一貫性と意味情報の新旧

多次元的な構造的関係を持ったセグメントを、一次的にテキストの中に並べていくプロセスで、連結のための言語上の手段を必要とするか否かは、そのセグメントを生成する基礎となった意味表現の構造的関係が、読み手にとって古い知識か、新しい知識かに依存する。

テキストの中に幾つかのセグメントを生成するとき、それらの間の構造的関係が古い知識として書き手と読み手に共有されているならば、表層言語上の連結手段は一般的に不要である。一方、その関係が読み手に取って未知の知識であれば、その関係を知らせる言語表現が必要である。例えばここで、幾つかの事象にたいしてそれぞれセグメントを生成してテキストを構成する場合を考えると、事象間の構造的関係に関する知識の新旧が、首尾一貫性の有無を決めることになる。すなわち、構造関係が既知の事象同士はそれらに対応するセグメント間に首尾一貫性が存在し、事象関係を示す言語表現を必要としない。構造関係が既知であるとは、既出のテキストで既に関係が示されている場合もあるが、書き手と読み手に共有されている言外の知識による場合もある。

4. 実例におけるテキストの首尾一貫性

Hobbsは談話を構成するセグメント間に見られる様々な内容的連関性を整理し、それらの関係を図式化した。[3]

Hobbsの扱ったセグメントは、文や節の様な小さな単位であるが、それらの間に首尾一貫性を作り出す関係の中には、パラグラフの様により大きなセグメント間の関係にも適用出来るものがある。例えば、Hobbsが文間に内容的連関性をもたらす関係の中に挙げた、詳細化(ellaboration)、時間順(temporal)、背景(background)などは、パラグラフとパラグラフの間の連関にも適用されることが実際の新聞記事などに観察される。

図1の例では、第1パラグラフはテキスト全体の要約を与えており、第2パラグラフから第7パラグラフまでのパラグラフは、それぞれ第1パラグラフの詳細化や背景説明という関係に依存してつながっている。また第7パラグラフは、第2パラグラフに内容的に連関している。

表1に、図1のテキストの各々のパラグラフで述べられた事象を抽出し、それらの間の内容的連関性を示した。

図1の例では、第1パラグラフはこの位置に有ることが必要であるが、第2から第6パラグラフまではこの順序で並ぶ必要は無く、順番を入れ替えてもテキストの意味内容は変わらない。(ただし、その場合は第2パラグラフで人

Coherence in a text and Context Representation
Structure

Yuriko UCHIDA, Shun ISHIZAKI, Hitoshi ISAHARA
Electrotechnical Laboratory

名や年齢などを述べるように、若干の手直しがいる。) また、第7パラグラフについても、元の第2パラグラフの後方であればどれとでも入れ替えられる。すなわち、図1の例では、パラグラフ間の連結性(relevance)は存在しない。

図1の例では、第1パラグラフで、誘拐というトピック全体のストーリーが要約的に示され、読み手はトピックの全体的な情報構造を得て、後続のパラグラフの内容をその構造と結び付けて理解する。

ところで、図1のテキストで第1パラグラフが欠如していたと仮定する。それでも大方の読み手はこのストーリーを誤り無く読み取ることが出来るはずである。解放(第2パラグラフ)の後に身代金要求(第5パラグラフ)があった、とか、犯人逮捕(第6パラグラフ)の後にまた新たな誘拐があった、とかいう誤解はしないであろう。それは、読み手が誘拐という概念に付随するストーリーについて枠組的知識を持っており、それに各々の事象をあてはめていくからである。

[誘拐トピック]の主な事象項目は、[被害者の略奪、(身代金などの)交渉、事件の集結(例えば、被害者の救助、解放、死体発見など)]であり、それらの主事象を構成する副次的な事象項目(たとえば、被害者の略奪という事象では、[犯人による捕獲作業、被害者の運搬、監禁])もある。これらは、誘拐というトピックの基幹になる事象である。この様に、誘拐というトピックを構成する構造は、どの読み手も知っているので説明する必要が無い。

典型的でだれでも連想できる枠組的知識は、テキスト自体としては一見ばらばらに見える連結性の無いセグメント同士に首尾一貫性をもたらすものである。

図2に誘拐トピックを構成する事象の枠組的構造を示す。

5. テキスト生成との関係

枠組的知識は、首尾一貫性を持ったテキストの理解においても生成においても不可欠のものである。筆者等はこの枠組的知識にたいする知識表現形式を文脈表現構造と呼び、これの構造自体の持つ情報を首尾一貫性のあるテキストの解析と生成に生かしている。[4][5][6]

例えばテキスト生成の場合、文脈表現構造の中で先験的に連結されている節点同士の間には旧知識としての構造的関係を前提とできるので、それらに対して生成したセグメントは、テキストの中で隣接する必要は無いし、また接続詞のような表層テキストにおける接続装置も必要としない。しかし、新しく因果関係などで連結された事象に対して生成したセグメントは、言語上の連結手段が必要である。

最後に、文脈表現構造から目的言語によるテキストを生成するためには、題材と言語に依存知識が必要であることを指摘しておく。言語に依存しない中間表現から、首尾一貫性を持った表層テキストへの変換する道筋は幾通りも考えられる。その中から「自然な」表現を選びだすには、題材および言語の両方に依存した知識が必要である。すなわち、構造的には同じ文脈表現構造から新聞記事を生成する場合と、小説を生成する場合では異なるセグメント配列を取るべきである。また、新聞記事の場合、同じトピックに対する記事でも、英語と日本語では異なるセグメント配列で首尾一貫性を表現している。この様な、題材と言語に依存した首尾一貫性に関する知識は現実の多様なテキストの分析から得られる。

6. おわりに

従来はあまり明確に論じられてこなかったテキストの首尾一貫性と連結性、および結束性の違いを本稿では検討した。また、表層テキストの上で論じられてきた首尾一貫性が、テキストの背後にある大局的な情報構造に基くことを

指摘した。そしてそれが文節間だけでなくパラグラフ間の内容的連関性にも適用出来ることを実例で示した。

7. 参考文献

[1] Mann W.C.:Text Generation The Problem of Text Structure,in Natural Language Generation Systems,1988
 [2] 田中春美 他:現代言語学辞典、1988
 [3] Hobbs,J.R. :WHY DISCOURSE IS COHERENT ?, SRI Technical Note 176,1978
 [4] 石崎 他:機械翻訳システムCONTRASTにおける文脈情報の利用、第33回情報処理学会全国大会、1986
 [5] 内田 他:機械翻訳システムCONTRASTにおける英語テキスト生成、第34回情報処理学会全国大会、1987
 [6] 内田 他:機械翻訳システムCONTRASTにおける英語テキスト生成、情報処理学会自然言語処理研究会、1987

WASHINGTON (AP)—The wife of a former Salvadoran ambassador, held six days by kidnapers demanding \$1.5 million in ransom, said Friday that her rescue from a Washington telephone booth was "just like watching TV."

Clelia Eleanor Quinonez, 53, was kidnapped as she returned to her Miami, Florida, home July 8. She was freed late Thursday by FBI agents as her captors were having her husband, Roberto Quinonez Meza, to say she was safe. Quinonez, 56, a wealthy businessman, served as the Salvadoran ambassador to Washington under Gen. Carlos Humbert Romero from

1977 - 79. A military coup overthrew the Romero Government in 1979.

Officials appeared uncertain about the motive for the kidnapping.

One agent in Washington said the ransom money was intended for leftist guerrillas battling the Salvadoran Government. But an agent in Miami said he thought the kidnapping was more criminal than terrorist.

Seven people were arrested. Quinonez told reporters at a midday news conference at the FBI's Washington field office that she was abducted in the driveway of her suburban Miami home.

(Asahi Evening News 1983年7月16日)

図1 連結性の無いパラグラフ間の首尾一貫性を示す実例

表1. 図1のパラグラフで述べられている事象とパラグラフ間の内容的連関性

パラグラフ番号	他のパラグラフとの内容的連関性	記述されている事象
1	要約	誘拐事件
2	パラグラフ1の詳細化	略奪、救出(終結)
3	パラグラフ1の背景	(人物記述)
4	パラグラフ1の背景	(誘拐の動機)
5	パラグラフ1の背景	(身代金の目的)
6	パラグラフ1の詳細化	逮捕(終結)
7	パラグラフ2の詳細化	捕獲

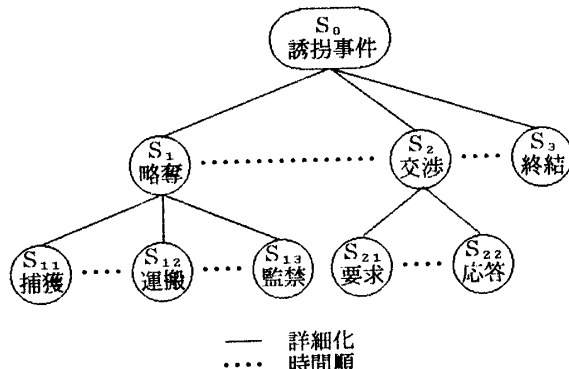


図2 誘拐トピックを構成する事象の枠組的構造