

説明的独話の談話セグメントに対する題名記述法の検討

竹内和広, 高梨克也, 森本郁代, 井佐原均

独立行政法人通信総合研究所

{kazuh, takanasi, ikuyom, isahara}@crl.go.jp

1. はじめに

本稿の試みは『日本語話し言葉コーパス』(CSJ)における話し言葉の談話構造分析の一環としての試みである[1]。

講演や講義などの説明的独話の談話構造分析を行う上で、談話のそれぞれの部分でどのような内容が語られているかを記述することが重要である。しかし、このような内容を何の制約もなく自然言語で記述した場合には、多様な題名記述の連続になってしまい、例えば、複数の作業で談話の内容記述を談話セグメントの題名の集合により検討する場合などでは、作業結果の妥当性をより緻密に議論する上での問題となる。

本稿では、数文の連鎖から構成される談話セグメントの談話主題とその修辭的説明パターンの類型を談話セグメントの題名記述に反映させることを通じて、より精緻に談話セグメントの内容を手で記述する手続きを検討する。

2. 談話セグメントへの題名付与問題

本プロジェクトが行う談話構造分析の動機となった背景理論に、Grosz と Sidner の談話構造理論 (以下GS)がある[2]。しかし、我々は、CSJの談話の構造解析を行う予備的な試みを様々な作業基準を用いて行う過程の中で、CSJのような10分を越える自然発話に近い説明的独話について意義のある分析をするためには、様々な大きな課題があり、アノテーション作業の焦点を絞ることが重要であることを認識した。この問題の考察と我々の試みの全体像については森本らの発表を参照されたい[3]。

ここでは上記の試みで得られた問題点を要約する。我々はGSに基づく談話分析の先行研究を参考に、談話をセグメントに分け、セグメント間の階層を認定し、さらに各セグメントに対して、作業者が推定した話し手の意図や目的を自然言語で記述した。しかし、我々のこの予備的実験の結果では、この3つの認定作業のどの作業においても、作業間の相互依存的なデータの不一致が見られ、全体として安定的な分析作業を実行できるとは出来るとはいえなかった。なお、GSにおける話し手の意図や目的とは、直感的には、談話目的とは話し手が当該の談話セグメントをなぜ発話したか、その理由の記述である。この意図や目的の記述のことを、本稿では、談話セグメントの「題名」と呼ぶ。

ここで、セグメント例1のような「まとめり」を談話セグメントとして作業者が認定できたと仮定し、本稿の

問題である題名記述の問題を説明する¹。

セグメント例1に対して何も制約をかけずに作業者が自然言語による「題名」を付与すると、例えば、「実験の結果の説明」「プレジジョンとリコールの計算」「実験したことから考えられること」などの様々な「題名」が記述されることとなる。このような題名記述を何らかの形で比較可能かつゆれの少ない形に制約しなければ、作業者が把握した内容に対応する談話セグメントの境界や、談話セグメント間の階層や関係を議論しようがない。そこで我々は、まず、話し言葉で一般的にあらわれる意味的なまとめり (本稿では、そのような談話セグメントをエピソードと呼ぶ) を定義し、その題名を記述する手順を検討することに分析の重点を置くことを選択した。具体的にはセグメント例1の題名記述のゆれが、少なくとも「実験の結果」や「実験の結果の傾向」といった差になるように制約することを考えたい。

説明的独話は、すべての談話部分が相手に何かを説明するというコミュニケーション活動であるため、説明的な談話であるという意味においては書き言葉にも近いと言える。そのため本稿の説明的な独話分析のために題名記述をする手法は、書き言葉の分析にもつながる課題であると考えている。

セグメント例1

- A1) というわけで実験してみました
- A2) プレジジョンとリコールはこのように計算します
- A3) プレジジョンはそんなに良くなかったです
- A4) リコールはこれぐらいでした
- A5) グラフにするとこんな感じになります
- A6) まあ実用に耐えるのではないかと考えています

3. 題名を付与するセグメント単位 : エピソード 3.1 CSJの談話分析における最小単位

従来、書き言葉を対象とする場合には、情報付与の対象となるまとめりの単位としては「文」が用いられてきた。しかしながら、CSJを対象とした場合、文を単位とすることには以下のような問題点がある。

- 書き言葉では書き手自身が句点によって区切りを確定するのに対して、話し言葉にはこうした情報がない。

¹ 本稿で提示する例は作例で、各発話単位を短く簡略化してある。発話単位については3.1節で触れる。CSJに含まれる談話は、本稿で示した例よりもさらに多様でかつ長く、相当数の先行文脈を提示する必要があるため、作例により問題点を明示することに重きを置いた。

- 文法的に明確な文末形式が用いられるとは限らない。
- 自発的な話し言葉では、言い直し、言い換え、言いやめなどの要因により文の範囲が確定しにくい場合や語や文の断片だけで発話が構成される場合がある。

そこで、CSJでは、統語的・意味的に利用しやすい単位を認定し、それを基本単位として採用することにした。その単位を便宜上「節」と呼ぶ[4]。

3.2 文間関係を用いたまとまり認定の問題

談話には多様な「まとまり」の単位を定義しようが、談話内にまとまりを感じさせる要因を、Hallidayは、照応や省略のような談話の結束性(cohesion)という概念を言語的な機能に基づいて提案している[5]。これに対し、省略、照応、繰り返しといった言語的デバイスだけでは捉えられない種類の談話内容の「まとまり」を指すものとして、一貫性(coherence)という概念が用いられてきた。しかし、一貫性の内実については様々な理論が提唱されてきているもの、決定的なモデルが存在するとは言いがたいのが現状である。

ここで、我々が「まとまり」すなわちエピソードとして想定するのは、特定の結束性を手がかりに決定的に決まる「まとまり」ではなく、一貫性を持つ、書き言葉の段落に相当する「まとまり」である。談話の一貫性に関する先行研究として有力なものに修辞構造理論(RST)がある[6]。RSTでは談話内の節間に見られる二十数種類の関係が規定されており、各関係を認定する際の基準が明示されている点に特徴がある。これらの関係は他の単位との間に再帰的に適用でき、最小単位である文間関係の判定から始め、より上位の単位間関係の判定へと進むことにより談話を階層構造で記述することが想定されている。

しかし、最小単位をまとめあげて、一貫性を保持する単位にボトムアップに分析していくことには、問題がある。以下に、話し言葉における2つの単位の例を示し、それを用いて説明する。

B1) 監督は一年だけで辞めてしまいました

B2) 後援会からも批判が出ていたみたいです

この2つの節は、中心的な節を前者とも後者とも取ることが可能である。例えば、「監督は一年だけで辞めてしまった」ことが話者の中心的な批判であり、話者の批判を強調するために(話者だけではなく)「後援会からも批判が出ていた」と発話したと捉えることも、「監督は一年だけで辞めてしまった」ことが背景にあり、その結果「講演会からも批判が出ている」ことが重大な問題なのだと捉える可能性もある。これは、書き言葉であれば、複文の形式的制約から節間の関係のある程度限定できる可能性もあり、文や節の形式が相対的にあいまいな話し言葉特有の問題であるようにも思われる。しかし、この問題

は、言語形式上の問題から2単位間の関係付けが多様になることだけが問題の本質なのではなく、RSTなどの文間関係に基づく分析を設定する上での本質的な問題である[7,8]。

このような決定的には決まらない基本発話間の関係を、あいまいなまま残しておき、その部分を包含する範囲の意味内容、あるいは文脈を適切に表現してやることにより、発話間の関係や遷移のあり方を制約することが考えられる。例えば、範囲とその中での中心文(あるいは重要文)といった観点を与えてやれば、その範囲の中での発話間関係のあいまいさを軽減できる。

3.3 話者が提供する「まとまり」へのヒント

3.2節で説明したように、文の連鎖をまとめていくためには、どのような観点で文がまとまるかを制約する要件が必要となる。

非常に一般的な言い方ではあるが、我々が一連の発話が「まとまる」ための観点として想定するのは「話題」である。問題は、話題をどのように捉え、扱うかである。

例えば、3.2節のB1,2がある種の「まとまり」を感じることができる以下のような環境下にあるときを考える。このようなまとまりを形成する話題を作るキッカケを形成しているのはB1,B2に先行するC1,C2,C3である²。

セグメント例2

C1) そんな厳しい監督だったわけですけど

C2) 監督となぜ離れてしまったかというのは

C3) 監督はまだまだ若いんですけどね

B1) 監督は一年だけで辞めてしまいました

B2) 後援会からも批判が出ていたみたいです

D) いろいろ苦労があったようです

ここで我々が注目したのはC2のような発話である。C2は、伝統的な言い方では、このセグメントの主題あるいは話題を提供し、このセグメントをまとめる手がかりを提供していると考えられる。

機能文法の伝統では、書き言葉のパラグラフに相当するレベルで観察される共通の話題を談話主題としてとらえる試みがある。C2はこのような談話主題を導入する存在と考えることもできる。しかし、自発性の高いCSJの談話主題の導入は非常に多様であり、必ずしもC2のような「Xというのは」や「Xとしては」という言語形式により談話主題を導入しているわけではない。また、仮に談話主題を導入しようとする言語形式を談話マーカースとして列挙したとしても、同一表現であるからといって常に談話主題を導入する機能を果たすとは限らない。談話主題の粒度も問題である。C1からC3まではすべてが「監督」について述べられているため、談話主題が「監督について」ならば、これらの発話すべてが談話主題の導入

² Dは「まとめ」の性質をもち「まとまり」を示す大きな要因となる。しかし、4節で述べるように、その生起の傾向をつかむことは難しい。

となるのか、初出の監督のみが談話主題の確立に寄与するのか疑問が残る。

我々は、上の C2 にみられる機能は単に談話主題を導入するのではなく、C2 機能の働きの一環において、談話主題が導入されたと考える。また、その機能が、話し言葉における「談話主題」の導入形式を多様化している原因だと考えている。ここで、談話主題の導入と区別するため、我々が対象とする談話セグメントであるエピソードの話題を導入する部分を以降「話題導入」と呼ぶ。

我々は C2 のような話題導入は、話し手が話題（談話主題）X に関しての「X の焦点が何か」「X の感想はどうだったか」といった、伝えようとしている内容を「予告」する意思表示、あるいは、話し手自身が今から何を述べるかを確認する自身のための題名付けの痕跡ととらえる。つまり、このような予告的存在を手がかりに、話者がどのような戦略で談話のまとまりを計画しているかを分析しようとしたことが我々の試みの始まりである。

まず、予告を対話の相互行為の観点から「X の Z は W か」という形式に一般化する。この形式化の意義と X,Z,W それぞれ詳細については、4 節で詳しく述べるが、おおまかには、

- 話題の X
- X についての話し手の焦点を示す抽象名詞 Z
- 「何か」、「なぜか」等の疑問詞に相当する W

である。ここで、実際の談話では、X,Z,W の項すべてが満たされる、ある種抽象的な「予告」は少ないこと、あるいは予告的な要素そのものが存在しないこともあることに注意されたい。

我々がエピソードと呼ぶ「まとまり」は、この<X,Z,W>の組を観点にまとまる談話セグメントである。

4. エピソードの認定と題名記述

4.1 エピソード認定作業

エピソードの認定は2つの相補的な側面から成る。3.2 節で述べた「予告」は抽象的な存在で、実際には X,Z,W のすべてが談話表層上に表出するわけではなく、そもそも予告に相当する部分すらない場合もある。そのため、<X,Z,W>のうちのどれかを導入する断片的な予告が話題導入の典型であると考えている。すなわち、エピソード認定には「予告」の機能を部分的に果たす話題導入を認定する側面がある。

他方、明らかに話題導入があっても、その「予告」に見合った発話内容を話し手が後続させるか、あるいは後続させたとしても、それが作業（聞き手）が認定できるかは別の問題である。このように、ある話題導入について、それがどのような形で充足され、エピソードを形成するかを認定する側面がある。

4.2 話題導入の認定

エピソードを認定する上で、まず、話題導入の側面を

考えるためには、独話に「予告」が存在する理由を考えなくてはならない。

独話は基本的に聞き手からの具体的な問い返しがないことが特徴であるが、その代わりに、話し手が聞き手の理解を見越すことによって生み出されたと見なしうる現象がある。その現象が「予告」と捉えることができる。これは、コミュニケーションにおいて「話者が、聞き手の理解を考慮しながら自分自身の談話を意識する」ことを考慮にいれ談話を分析することを意味する。高梨はこの点に関して、文間関係の認定を通じて談話の一貫性を説明するためには進行中の談話に対して聞き手が抱くであろう「疑問」という観点の必要性を主張した [9]。つまり「疑問」という観点から独話を分析すれば、話題導入は話し手が聞き手の理解状態、特に話し手自身の発話に対して聞き手が抱くかもしれない「疑問」を見越す、あるいは配慮して発話される要素と、捉えることができる。

具体的に話題導入を探す作業例で考えてみよう。例えば、セグメント例 2 の話題導入を C2 と仮定すれば、<監督と離れたこと、φ、なぜ>という空項を含む組ができる。これが、エピソードの話題導入と認定されるのは、この Z の位置の φ が話題導入以降で充足された時とのみである。

また、Z も W も話題導入にない場合がある。セグメント例の 1 話題導入を A1 と仮定すれば、<実験、φ、φ>がエピソード認定の候補になりうる。つまり、話題導入の候補が必ずしも言語表現から認定できないのであれば、すべての発話のすべての要素が X の候補に成りえ、エピソード候補の数は膨大になる。実際の作業では、候補の数を制限するため、談話の音を聞きながら作業するなどの形で作業者の言語直感を取り入れている。

実際の談話では、セグメント例 1 のように<X,Z,W>のうち X のみが明示的に導入されることが多い。この場合、X の導入のされ方により、話者が抱く疑問、例えば「なぜか、どのように」という形で、一通りでないにせよ有限の疑問 W を持ちうる。実際のエピソード認定でも、これらの疑問の想起とその解消を必須の観点に分析を行う。

なお、X は談話上の位置を特定できるようにするため、できるだけ話題導入部分から拾うことを優先したが、対象が発話言葉であるため、省略や言い間違いが少なくなく照応に関連する語の補完など、条件付きの X の修正は認めた。

4.3 話題についての「応要素」の認定

話題導入が聞き手側の疑問の励起とするならば、その疑問に解答を与える要素は直感的には「まとめ」である。しかし、実際の談話において、明確なまとめは必ず生起するわけではなく、生起しても複数の談話セグメント内容に関わる内容をまとめていたりするため、話題導入と 1 対 1 の関係には必ずしもならない。また、まとめは、

話題に関して、いくつかの発話の連鎖が行われた後に言い換えや解釈の性質をもって生起するため、発話の連鎖をもって話題を充足させると考えた方が自然である。

談話における「評価」の重要性を強調した研究としては Labov のナラティブ分析がある [10]。彼によれば、物語は語るに値する（あるいは聞くに値する）ものでなければならない、物語において「評価」が重要なのは、「評価」こそが語り手が語りの存在意義であり、当該内容を話す理由を示す動機だからである。つまり、話題導入が聞き手側の聞く動機を高める話し手の戦略の一つと考えるならば、話す動機に強く関連するのが評価の概念である。

評価は、例えば、言及している対象が、発話者によって好ましいことを伝えたい、ということが談話目的となることに関連する。しかし、単に感情的な好悪のみを限定して評価と定義していたのでは、説明的な談話を部分的にしか説明することができない。そこで、評価を言及世界と話し手の主観を交差させる手段と考え、例えば、「ある事件についての原因を話し手が発見し、それを伝えたい」というものも、事件に対しての話し手の評価であると解釈する。

認定の作業では、道具立てとして、このような説明的な評価を、話題 X をどのように総括したかと捉え、抽象名詞を用いて表記し、Z とする。この Z にとりうる抽象名詞はリストとして整理した。本作業では、当初は「結果、理由、利点、問題点、きっかけ、経緯、特徴、程度、状況、方法、分類、構成、前提」といった相対的に広い概念をもつ抽象名詞の Z 候補リストを作成し、4.4 節で示すチェック作業と実験を繰り返す中で、詳細化・グループ化を行い、Z 候補リストのそれぞれの名詞について特徴付けを精緻化していった³。また、Z として適切な抽象名詞を決めるために、それぞれの抽象名詞に対応する X として適切な対象を、事物や出来事、行為という形で条件付けする。さらに、Z と W はほぼ対に決定できるように Z の粒度を設定した。例えば、Z が「理由」を選ぶ時には、X は行為であり、話題導入で生じた疑問詞 W は「なぜ」でなければならないと条件付けされる。

作業の例としては、4.2 節でのエピソード候補となった例では、例えば以下のような<X,Z,W>が入りうる。

セグメント例 1 : <実験,結果,どのような>

セグメント例 2 : <監督と離れた,理由,なぜ>

となる。すなわち、典型的な X が話題導入として導入される場合であれば、話題導入以降の発話とその連鎖パターンに最も適切な Z を見つけ<X,Z,W>を決定することによりエピソードが認定される。

4.4 エピソードに対しての題名化と整合性チェック

<X,Z,W>からの自然言語表現化した題名の生成は、前

節で述べたように、W と Z はほぼ対応付けが可能なように Z を設定した関係から、意味的な受容性が高い「X の Z」接続、あるいは、X が行動や動作などの場合は X を Z に連体接続することで行った。

例えば、4.3 節のエピソード認定例であれば、セグメント例 1 の題名は「実験の結果」となり、セグメント例 2 の題名は「監督と離れた理由」である。

これらの題名記述に関する、最終的なチェックは、話題導入以降の範囲を代表「題名」として適切かどうかを言語直感によりチェックした。そのチェックにおいて、不整合を感じた場合は、まず X の認定が間違っていないかをチェックし、X 側に問題がない場合は Z のニュアンスが適切でないかをチェックした。Z のニュアンスが適切でない場合、必要に応じてそのエピソードに適切な抽象名詞を検討し、Z 候補リストに追加した。この際、新しく追加する名詞が、Z 候補リストに既に存在する抽象名詞とどのような点が異なり、どの抽象名詞の下位分類となりえるかを意識し、その抽象名詞の特徴付けを行った。

5.まとめ

本稿はポスター発表の予稿であるため、あえて我々が行っている試みの理念的背景を中心に述べた。さらに詳細なエピソードの語用論的な理論背景については高梨ら [11]を参照されたい。CSJ のデータは多様性に富むため、必ずしも本稿で述べた理論と整合性が高い例ばかりではない。ポスター会場では公開直前の最終データと内部作業データとともに、分析の難しかった個別的な事例を紹介する予定である。本稿の内容を実践した結果について議論させていただければ幸いである。

参考文献

- [1] 井佐原均 2003 『日本語話し言葉コーパス』への情報付与。平成 15 年度国立国語研究所公開研究発表会予稿集。
- [2] Grosz,B.J.& Sidner,C.L. 1986. Attention, intention, and the structure of discourse. *Computational Linguistics*, 12(3): 175-204.
- [3] 森本、竹内、高梨、井佐原 2004 『日本語話し言葉コーパス』における二方向の談話構造分析。In this volume.
- [4] 高梨、内元、丸山、井佐原 2003. 『日本語話し言葉コーパス』における節境界認定。平成 15 年度国立国語研究所公開研究発表会予稿集。
- [5] Halliday,M.A.K. & Hasan,R. 1976 Cohesion in English. Longman Group Ltd.
- [6] Mann,W.C. & Thompson,S.A. 1988 Rhetorical structure theory: Toward a functional theory of text organization. *Text*, 8(3): 243-281.
- [7] Moore,J.D. & Pollack,M.E. 1992 A problem for RST: The need for multi-level discourse analysis. *CL*, 18(4): 537-544.
- [8] Moser,M. & Moore,J.D. 1996 Toward a Synthesis of Two Accounts of Discourse Structure. *CL*, 22(3): 409-419.
- [9] 高梨克也 1998 一貫性の多様な次元-日常会話と思考の「自然な流れ」の解明に向けて-. デュナミス-ことばと文化-. (京都大学大学院人間・環境学研究科文化環境言語基礎論講座) 3: 73-104.
- [10] Labov,W. 1972. The transformation of experience in narrative syntax. *Language in the Inner City*. University of Pennsylvania Press. 354-396.
- [11] 高梨、竹内、森本、仲本、井佐原 2003 談話を語る/聞く動機とエピソード構造、日本語用論学会第六回大会予稿集。

³ この原稿を執筆する時点では、公開データのための最終調整をしている。