

『日本語話し言葉コーパス』の談話境界情報について

Version 1.0

竹内和広・森本郁代・高梨克也・井佐原均

独立行政法人 情報通信研究機構

【目次】

1. 談話境界情報の概要.....	2
2. 談話境界情報として付与されるデータ .....	3
2.1 談話境界とその題名の付与 .....	3
2.2 データの仕様.....	3
2.3 作業の流れと対象講演の選択手順.....	5
3. 作業者ラベルの付与.....	6
4. 統合ラベルの付与 1：談話下位目的の認定.....	7
4.1 談話下位目的の認定手順.....	7
4.2 説明タイプの分類.....	9
5. 統合ラベルの付与 2：談話目的の認定.....	11
5.1 作業者ラベルの境界とのすりあわせ .....	11
5.2 談話下位目的にもとづく談話目的の生成 .....	12
6. おわりに.....	13
参考文献 .....	13
付録：対象講演 ID 表.....	14

## 1. 談話境界情報の概要

日本語話し言葉コーパス（以下 CSJ）において、談話境界情報が付与された講演は、コアに分類される説明的モノローグのうち 40 講演である。40 講演の内訳は、模擬講演 25 講演と学会講演 15 講演である。これらの 40 講演を選択した手順は 2.3 節で述べ、対象となった講演の講演 ID は付録として本稿末に添付した。談話境界とは、一つの談話の中でまとまりをなす話題の境界である。このような話題のまとまりは、談話の構成を分析する上で単位になりうる。概念図で示すと以下の図 1 のようになる。

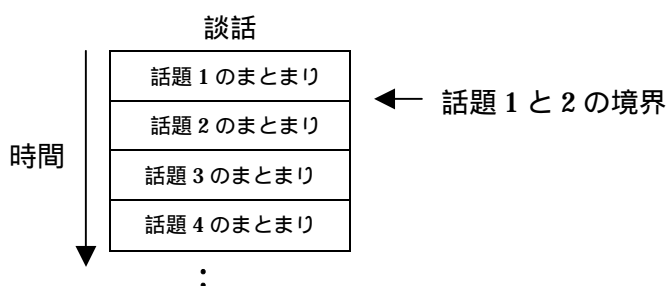


図 1. 談話の構成概念図

談話境界を付与する対象は、原稿読み上げの朗読と比較すれば、自由度の高いモノローグではあるものの、書き言葉に近い性質を持っている部分もあるものと考えられる。書籍や新聞などの書き言葉であれば、章や段落など、書き手側が形式的な手続きに従って読み手に与える「まとまり」の目印が存在する。しかし、話し言葉においては、まとまりを話し手が明示する明確な手続きが存在するわけではなく、分析者が談話境界を分析する手がかりに乏しい。

談話の分析に対しての枠組みは、様々な点において流動的であるため、基本的には試行錯誤を繰り返しながら、得られた知識を積み重ねていくことが重要である。本談話境界情報の付与でも、このような「まとまり」についての理論的定義が研究的課題の域を出ない以上、人間が話題という概念を意識し、作業した結果を尊重することが重要と考える立場をとる。

本談話境界情報で付与されるデータは、大きく以下の 2 種類に分けることができる。

- 作業ラベル
- 統合ラベル

前者は、作業者の分析における直観を尊重したデータとして、複数人の作業者が行う談話境界の認定の作業結果であり、後者は相対的に手順を形式化したデータとして、本稿の著者らがある一定の理論を仮定し、本稿の著者らが手続きに従って認定した談話境界の作業結果である。各ラベルのデータ形式に関わる仕様は 2 章で詳しく述べる。

## 2. 談話境界情報として付与されるデータ

### 2.1 談話境界とその題名の付与

本境界情報付与作業の特徴は2点ある。一つは、1章で述べたように、複数分析者の談話境界結果と一定の手順を用いた談話境界認定を公開することである。他方は、境界情報だけでなく、境界で区切られる談話部分の話題を端的に表すような「題名」も付与していることである。談話境界における題名は、分析者がどのような「話題」を意識して「まとめ」を認定したかを示す証拠となる。実際の作業では、談話境界間で維持されると認定した話題を「題名」として記述する。ここで、本稿が作業的に認定する「まとめ」のことを、「談話セグメント」と呼ぶ。題名は各分析者が談話セグメントの境界として認定した範囲と相互関係にある。

談話セグメントに題名を付与する上で、Grosz と Sidner の談話構造理論（以下 GS と略す）[GS 86]における、話し手の発話の意図ないしは目的（本稿では談話目的、あるいは単に目的と呼ぶ）の考え方を意識した。GS では、話し手の目的は構造化されており、その目的構造の存在が談話の表層的な言語構造に反映され、談話の一貫性が達成されているものとみなす。ここで、GS における話し手の談話目的という概念を要約すると、

- なぜ（他の行動ではなく）談話という言語行動によって事をなそうとしているのか
- なぜ（他の内容ではなく）この談話の内容を伝達しているのか

に関わるものである。つまり、GS における談話目的とは、直観的には、話し手が当該の談話セグメントを伝達する理由「なぜその談話セグメントを言ったのか？」であり、その談話セグメントは全体の目的の達成にどのように貢献するかに関わる。本談話境界認定では、分析者が認定した談話セグメントに対しての題名を、この談話目的の近似物としてとらえ、以降、「目的」と呼ぶ。

### 2.2 データの仕様

談話境界データとして公開される 40 講演に付与される談話境界情報の XML 属性を表 1 に示す。CSJ の談話は自発性が高く、いわゆる書き言葉における「文」のような発話単位を人手で認定しなくてはならない。そのため、談話境界の最小単位として CSJ で統一的な単位の一つである節単位を利用する(節単位の認定については、マニュアルとして同梱された『日本語話し言葉コーパス』における節単位認定「[clause.pdf](#)」を参照されたい。)

談話境界情報を発話との関係に基づいて表にすると、図 2 のようなイメージになる。図中の講演列は講演の一部分の発話群を抜き出したものである。談話境界は、表 1 に示した作業ラベル、統合ラベルとも、話題の境界となりえる節単位に当該の談話セグメントの目的を記述することにより付与する。この時、境界は、当該の発話の前に存在すると考える。例えば、作業ラベルのうち、談話作業ラベル 3 が図 2 に示した部分談話の中で境界としたのは、節単位「そうやって酷い…」の前と、「僕はですね這いつくばる…」の前である。談話コメント属性のみは、目的の記述を値にとるのではなく、表 2 に示す仕様の属性であ

る。談話コメントの属性値は、本稿の著者らが、談話の分析過程において認定した付加情報である。ほとんど全ての対象講演に付与される談話コメントに談話の開始と終結に係る情報がある。このような情報を付与した理由は、プライバシーの問題から、所属名・発表者名などを談話から除去する関係上、談話境界認定の分析対象から談話の当該部分を除いたことを明示するためである。また、談話の開始や終結、あるいはフロアとの対話部などの談話境界認定の作業対象から外す部分を、あらかじめ認定しておき、作業者が談話を分析する際の参考情報として提供した。

表 1．談話境界情報で使用する属性

作業者ラベル	
英語属性名	日本語属性名
DS_Subject1	談話作業者 1
DS_Subject2	談話作業者 2
DS_Subject3	談話作業者 3
統合ラベル	
英語属性名	日本語属性名
DS_Purpose	談話目的
DS_SubPurpose	談話下位目的
DS_Comment	談話コメント

表 2．談話コメント属性の値

値	意味
%B0	談話開始部
%B1	開始部補足
%E	談話終結部
%C	境界認定除外候補
%I,%O,%S	5 章の認定作業にともなうコメント
自然言語による記述	上記以外で、5 章の認定作業時に付与される当該の談話に特徴的な操作に関する記録的コメント

講演	作業者ラベル			統合ラベル	
	談話作業者 1	談話作業者 2	談話作業者 3	談話目的	談話下位目的 談話コメント
<p>そうやって酷い航海だった訳ですけども/並列節ケドモ/ フェリーにはですねお風呂がありました[文末]</p> <p>で僕はどうせ揺れているんだからお風呂に入ってしまうと船の揺れとですね同期してむしろ揺れを感じなくなるのじゃないかという風にですね浅はかにも考えました[文末] しかし結果は悲惨なものでした[文末] お風呂は浅いですので当然お尻が着きます[文末] そうするとですね僕の視界に飛び込んでくるのはですね湯船をですねたぶたぶんとですね揺れる波でした[文末] やはりそこも波でした[文末] 僕はまたそこでげろを吐きました[文末] 僕はですね這いつくばるようにして風呂から上がりました[文末]</p> <p>でもう吐くものが何もなくなってですねいよいよ風呂に入ってですね脱水症状になってなり掛けたのでやばいと思ってですね僕はボカリスエットを飲みました[文末] そしてシャワーならいいかと思ってシャワーを浴びに行きました[文末] それでですね頭から熱いシャワーをこう浴びているとですね幾分こう救われた気がしたんですけどもね/並列節ケドモ/ 暫くすると一際熱い何かがこう流れて僕の全身をこう伝いました[文末]</p>			<p>入浴が逆効果だったことの説明</p> <p>シャワーも効果がなかったことの説明</p>	お風呂での事件	お風呂での状況

図 2 談話情報の参照例（談話の一部を抜粋）

### 2.3 作業の流れと対象講演の選択手順

談話境界情報の認定作業の流れを図示すると図 3 のようになる。作業者ラベル付与の段階ではそれぞれの講演について 3 名の作業者が同じ手順で独立に行った、3 通りの結果を付与する。また、統合ラベルの付与は 2 段階に分かれ、まず、談話下位目的を認定し、その後、談話目的認定を行うことに注意されたい。

なお、最終的に公開する談話境界情報付与対象の 40 講演は、以下の手順で選択した。ここで、模擬講演のテーマ内容、学会講演の学会種別などの詳細については、「音声収録作業の概要」(recording.pdf) を参照されたい。

1. 模擬講演 35 講演と学会講演 25 講演、計 60 講演を以下の点に留意して選択した。

模擬講演

- (1) 話者が重複しないように講演を選択

- (2) 各テーマが模擬講演全体に占める割合から各テーマ講演を選択
  - (3) 各テーマごとに男女比ができるだけ同じになるように講演を選択  
学会講演
  - (1) 各学会が収録講演に占める割合に比例するように講演を選択
  - (2) 各学会で男女比ができるだけ同じになるように講演を選択
2. 3章で述べる作業を 60 講演について行う
3. 2の結果、3名の作業者が一致して認定した境界数が全境界数に占める割合が低い講演は、話者の個性が談話に強く表れているため、談話分析自体が困難である可能性がある。そのため、4章、5章で述べる詳細な分析を行う上で、作業の信頼性に問題を残すと考え、最終的な談話境界情報付与対象には、講演の種類・男女比を特に考慮しつつ、一致率の高い 40 講演（模擬講演 25 講演、学会講演 15 講演）を選択した。

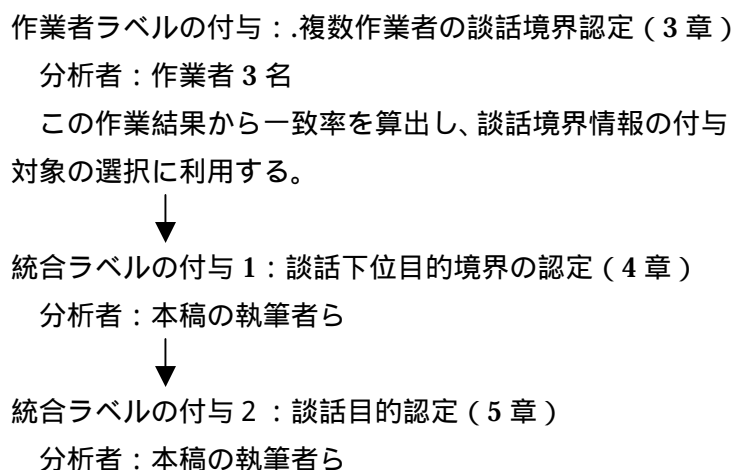


図 3.談話境界認定作業の流れ

### 3. 作業者ラベルの付与

作業者ラベルでは、1つの講演に関して、三人の分析者（以降、統合作業の分析者と区別するため作業者と呼ぶ）が、主に音声を聞きながら、独立に、談話を小説の章のような形で、階層性を意識せずに 5 ~ 15 程度の話題のまとまりに分割する。

次に、一度全体を通して聞いた談話に対して、談話境界と目的を認定してゆく。この際、談話の音声は何度も聞きなおしてもよいとした。目的の記述に関しては、談話セグメントの目的としてどのような対象を仮定しているかについて、2章で説明したものとほぼ同一の説明を作業者に対して行ったが、記述方法については特に制約を設けず自由に記述することとした。その理由は、本公開データ内における作業者ラベルの位置付けが、談話の個別

性に関連する情報を相対的に重視するものであり、制約をかけることによってかえってこの個別性が見えなくなってしまう可能性があるためである。

このようにして、一旦、談話全体を談話セグメントに分割しその目的を記述した後に、その結果を必ず分析者自身が見直すことを義務化した。この際、まとまりの粒度を分析者の主観により確認し、修正することを許可し、音声も何度聞き直してもよいものとした。作業者が談話境界を認定するのは、談話のうち、2章で示した談話開始部、談話終結部を除く部分である。ここで、談話開始部の直後が、作業者2人以上で談話の先頭境界としなかった場合は、開始部補足として、談話コメント属性の値に「%B1」を設定した。

作業者ラベルにおける談話境界は、各作業者の観点を尊重した結果であるため、当然のことながら、作業者ごとの境界認定位置にずれが存在する。予備研究の結果、作業者間の境界のずれの範囲が数発話程度で、かつ、両者の題名記述が類似している場合は、このずれの部分がある一定の性質をもつことを知見として得ている[森本 03]。この知見を前提とし、5章では、作業者ラベルの結果と、4章で行う目的記述の結果を統合し、作業者間で2者以上の一致がみられる境界について、手続き的な目的記述を試みる。

## 4. 統合ラベルの付与 1：談話下位目的の認定

### 4.1 談話下位目的の認定手順

統合ラベルの付与は2段階の認定作業から構成される。この章で述べる談話下位目的認定は、統合ラベルの談話目的を認定するための前処理と位置付けられる。

談話下位目的の認定では、節境界認定作業で得られた節をまとめあげ、内容上一貫性のある節の連鎖パターンを発見し、そのパターンを反映する題名と範囲を決定する。節の連鎖パターンとは、談話主題に相当する、あるキーワードに関する説明を、聞き手が抱くであろう疑問を見越して、話し手がどのように説明を展開するか/伝えるかを表現したものと捉える。形式的に言えば、我々が想定する談話下位目的の認定作業は、ある談話主題 X をどのような説明タイプ Y で説明しているかを示す<X,Y>の組を認定する作業である。ここで、説明タイプ Y は、話し手の X に対する「評価」を、節の一貫性をもつ連鎖として発話した説明方法のタイプ類型である。この方法のさらに詳細な理論的背景については高梨ら[高梨 03]を参照されたい。

統合ラベルとして付与する下位談話目的は、この<X,Y>の組を「X の Y」型、あるいは、X を Y に連体接続することにより名詞句を作り記述する。このように単なる組<X,Y>ではなく、自然言語して目的を記述し、確認する理由は、分析者の言語直観を利用して、題名付けの妥当性を検討するためである。

例を挙げる。以下の図3に示す談話セグメントでは、談話主題に関わるキーワード X は「壁も非常に薄い」であり、Y は「問題点」を列挙している。さらに、組<X, Y>の認識が言語直観に鑑みて適切であるかどうかを判断するために、X の「壁も非常に薄い」と Y の問題点を結合させて言葉として適切な題名として「壁が薄いことの問題点」形に言い換え

る。この例の言い換えでは「こと」を補ったが、他にも照応をたどって復元できるような語を補ったり、あるいは適切な助詞を補う / 言い換えることを許す。ここで注意されたいのは、文章における談話主題と異なり、談話セグメント上の談話主題は、書き言葉のように特定の言語表現を伴って導入されるとは限らないことである。また、言語表層に現われる表現がそのまま談話主題ととらえることに適当な表現とは限らない。

で<接続詞>それから天井が<P>簡単に落ちるぐらいで<P><並列節デ>壁も非常に薄い<P>です<P>[文末]  
んで<接続詞>そもそも柱と壁の間に隙間があって<P><テ節>夜<P>電気を消すと<条件節ト>隣の明かりがちょっと見えるぐらいの<P>あれなので<理由節ノデ>声が物凄くよく聞こえるんですね<P>[文末]  
声とか音とかがよく聞こえて  
廊下を挟んで<P><テ節>五つずつかな<P>部屋があるんだけど<P>/<並列節ケド/>  
廊下を挟んだ向かいの<P>三つ先ぐらいの部屋の目覚まし時計で目が覚めてしまうと<引用節>いう<P>[文末]  
で<接続詞>そのうちこれはどこ<P>誰々の<P>部屋の目覚まし時計だから<理由節カラ>俺は起きなくて<テ節>いいんだとか<トカ節>そういう風になってきて<P><テ節>-  
非常に音がよく聞こえるのですよね<P>[文末]

図 3. 目的「壁が薄いことの問題点」を付与した談話セグメント例

僕は小学校一年生になる時に(昔は僕は足立区に住んでいたんですが/並列節ガ/)+今住んでる文京区に引っ越しまして/テ節/  
小学校に上がると同時に引っ越すってことは要するに友達とかそういうのが誰一人としていないという状態な訳ですね[文末候補]  
それで慣れない土地で一応でも小学校には行ったんですけども/並列節ケレドモ/  
その小学校がケー小学校っていうところで-  
僕が住んでいる新しく引っ越した文京区ってところは{人口自体が少なくて/テ節/:しかも殆ど子供がいない}ということなんでドーナツ化現象のあの穴の開いてる部分に大体当たるもんですから-  
その小学校が{全校の児童数が入学した時に三百五十人いるかいないかぐらいで/並列節デ/:各学年二クラスずつ二十五人から三十人ぐらいのクラスが二つずつ}っていう小ぢんまりとした構成でした[文末]

図 4. 目的「小学校一年生になった時の状況」を付与した談話セグメント例

図 3 以外の具体的な例として図 4 のような例も示しておく。図 4 の例は、図 3 で例示した目的「壁が薄いことの問題点」が主観的評価の度合いの強い談話セグメントであることに対し、相対的に主観的評価の度合いが低いと考えた例である。図 4 の談話セグメントの X は「小学校一年生になる時」で Y は「状況」と認定し、談話下位目的は「小学校一年生になった時の状況」とした。この認定では、この談話セグメントの中心的評価が、小学校の属性である「こじんまり」とした程度を中心的に述べているのではなく、話者が小学校に入った時の状況的属性を順に列挙したと考えた。



## 4.2 説明タイプの分類

本談話境界認定において<X,Y>を認定する道具立てとして用意した説明タイプは、話し手が X に対してどのような評価を下しているのか、そして、その評価をどのような戦略で聞き手に伝えようとしているかを主眼に Y を認定する候補として設定したものである。

表 3 に Y の基本候補として整理しリスト化したものを整理する。Y の候補となる語は「 」の後に列挙した。これらの語は、予備的な談話分析の試行錯誤の過程で、談話目的を認定する上で典型的かつ基本的なものと考えたもので、対象談話 40 講演の談話下位目的記述すべてがこの基本候補のみにより記述されているわけではないことに注意されたい。

説明タイプを表 3 に整理する上で基軸に据えたのは、 に分類した図 3 の例のような評価に対しての主観性の強い説明が中心となるタイプと、 に分類した図 4 の例のような属性列挙型などの主観性の度合いが低い説明タイプとの対立である。この対立は連続的であり境界的な例が多いことも確かだが、この判断を分析者が行うことが、説明タイプを分類する上で最も基本的な対立と考えた。候補の優先順位としては、表 3 の分類のうち、 と を最も基本的な説明タイプであり区別と考え、まず、説明に主観が介在する度合いを検討する。検討した結果、分類 I の「感想」、分類 II の「内容」といった一般的な説明タイプになる場合、さらに Y を詳細化することにより適切な目的を表現できないかを検討する。なお、表 3 の分類は、説明タイプを整理して説明する上での上位分類の一例を示したに過ぎない。Y の候補となる説明タイプは、様々な観点から分類可能である。

実際の分析作業では、表 3 のタイプ からタイプ までで分類した Y の基本候補を利用して、目的記述をした上で、分析者の言語直観と合わないものについては、以下の優先順位で検討を行った。

1. X の妥当性の確認
2. 新規説明タイプ導入の検討

基本候補に対して、実際の談話下位目的認定の過程で、新規説明タイプの導入を行う場合は、既存の説明タイプとどのような関係にあるかを明示し、作業管理データベース上に登録を行った。

表 3.説明タイプの分類

I. 中心的評価が相対的に明確な説明タイプ
感情の好悪や、利益、不利益など話し手の主観的评价が色濃く表れる説明。
利点・長所・欠点・問題点 X についての、話し手の肯定的 / 否定的判断である Y という特徴を中心に説明。
程度・良さ・ひどさ・うれしさ、など X が「どの程度」Y であるかにより特徴を説明。
特徴・特色 X の特徴・特色が「どのようなもの」であるかを、他者の比較や特殊性を論理的に説明することなどにより、話し手の肯定的 / 否定的な判断を相対的にあまり加えず説明。
解釈・意義

<p>Xの特徴を単に述べるのではなく、特徴に関連したXへの評価・解釈・意義を論述的に説明。</p> <p>感想・印象・思い</p> <p>Xへの主観評価を中心とする談話セグメントとして一般的な度合いが強い説明タイプ。</p>
<p>II. 中心的評価が相対的に弱い説明タイプ</p> <p>分類 と比較して、属性叙述を列挙して説明するなど、相対的に主観の度合いを低く説明するタイプ。</p>
<p>内容・状況・様子</p> <p>Xの内容・状況・様子などが「どうであるか」といった下位属性を列挙して説明。一般的な度合いが強い説明タイプ。</p> <p>種類・機能・形状・所属・効果、など</p> <p>Xの特定の属性が「何」であるかを説明。</p>
<p>III. 出来事と時間軸の関係が特徴的なタイプ</p> <p>分類 あるいは でも一般的な説明になってしまう出来事に関する説明に対して、動的な観点、あるいは時系列の観点を導入することによって適切な分類ができる説明タイプ。</p>
<p>経緯</p> <p>Xに関連する出来事の始点から終点までを述べるとともに、帰結に重点を置いて説明。</p> <p>帰結・結論・結果</p> <p>Xの出来事の終点・終点以降に対する評価を主眼に説明。</p> <p>変化</p> <p>Xに関連する出来事の始点と終点の変化を主眼に説明。</p> <p>思い出・事件・経験</p> <p>出来事そのものを説明する一般的な説明タイプ。</p>
<p>IV. 評価の中心が関係付けである説明のタイプ</p> <p>Xに関する話し手の評価の中心が、因果関係などの関係付けである説明。</p>
<p>理由</p> <p>Xに対しての理由が「何か」を、話し手の評価としての因果性を中心に説明。</p> <p>きっかけ</p> <p>時系列において出来事や行動に至るきっかけが「何か」を評価し、説明する。</p>
<p>V. 宣言的な談話セグメント</p> <p>説明行為に関連した特徴的な談話セグメントと考えることが適当なもの。</p>
<p>目標・目的</p> <p>談話中、あるいは談話外の対象について、話し手が何らかの宣言を行う。</p> <p>まとめ</p> <p>主に学会講演の終わりなどで、談話において言及してきた内容を談話セグメントの粒度で形式的に再評価を行う。</p> <p>分類</p> <p>Xの属性としての種類を説明するのではなく、Xに関しての分類を行う。</p>
<p>VI. 学会講演に特徴的な説明タイプ</p> <p>I から V は模擬講演を基本に学会講演にも適用できる一般的な説明タイプとして採用した。それに対し、ここに列挙する説明タイプは必ずしも学会講演のみに表れるわけではないが、学会講演の説明タイプを適切に記述することを想定して用意したものである。</p>
<p>定義・構成・対象・基準</p> <p>学会講演では、専門用語、概念やデータの構成などを説明する以前に文脈に導入しておく必要がある。このようなものを説明するタイプ。</p> <p>図示・例示</p> <p>学会講演は、OHP や予稿などを用いて説明をする点で、言語のみを用いる説明行為とは言いがたい点があり、言語外の参照表現を多用する場合があり、他の説明タイプに分類しがたい場合も多い。そのような場合に用いる説明タイプ。</p>

手法・手順・方法

学会講演では、手法や手順の再現性も説明対象となるため、以上のものを説明タイプとして採用した。

傾向・分布

「結果」や「特徴」としたのでは一般的すぎて適切でない場合に採用した。実験や提案手法の結果の傾向や分布の実例といった提示方法に基づいて詳細化する。

## 5.統合ラベルの付与 2：談話目的の認定

### 5.1 作業者ラベルの境界とのすりあわせ

3章で述べた作業者ラベルは、作業者の観点を尊重してつけたラベルであるため、当然のことながら、境界の認定箇所にはずれが存在する。しかし、2人以上の作業者が同一箇所、あるいは近隣する箇所に境界を認定した場合、それらの箇所は相対的に信頼性高く、話題境界を認定できた部分であると考えられる。本談話境界ラベル付与作業では、4章で述べた談話下位目的認定の結果を参考に、作業者ラベルの結果を統合することを試みる。

統合ラベルにおいて、「談話目的」に対応する談話セグメント（以下、談話目的セグメントと呼ぶ）は、作業者ラベルと談話下位目的ラベルの付与作業の結果に基づいて、まず境界を決め、その後、談話下位目的に基づいて、当該の談話目的の記述を決定する。具体的には、談話目的セグメントの境界は、作業者ラベルの結果で二者以上（作業者ラベルは三人の作業者で行っているため、多数決の結果となる）が同じ位置を境界と認定した箇所とする。

ここで、多数決で境界が認定できない境界においても、以下の要件を合議することにより、統合境界として含めることを検討する。対象となる境界候補は、以下の二条件である。

- 作業者ラベルにおいて2名の作業者が、1, 2の節単位をはさんで境界認定を認定している箇所
- 境界が談話下位目的で認定した談話セグメント（以下、談話下位目的セグメントと呼ぶ）と交差しないこと

上記の条件下でゆれが存在する箇所について、「はい」「そうですね」など、フィラー的な表現がある時は、無条件で、作業者2者一致に準じる談話目的境界とした。その他のゆれが存在する箇所についても、談話下位目的で認定した目的と、作業者ラベルにおける当該箇所の2者の談話目的がそれぞれ整合しているものとみなされる場合は、作業者2者一致に準じる談話目的境界とみなした。その場合の境界位置は、本稿の筆者らが、音声を再確認しつつ合議によって決定した。このようなみなし談話目的境界は、談話コメント属性に「%S」を値に持つ。

なお、談話下位目的セグメントは、先にも述べたように、作業者ラベルにおける作業者が認定した談話セグメントの粒度よりも小さい傾向を持つことを予備実験により確認している。作業者ラベルにおける2者一致以上の境界一致により定義される統合ラベルの談話目的セグメントは、談話下位目的セグメントと交差がなければ、1つ以上の談話下位目的

セグメントを包含する。

談話目的セグメントと談話下位目的セグメントの間の交差はほとんど起こらないが、ごく一部の談話下位目的の境界が、作業ラベルの境界認定の結果に対して交差するなど整合的でない場合は、合議で対処を検討し、当該の対処方法を談話コメント属性の値に自然言語の注釈として記述した。

## 5.2 談話下位目的にもとづく談話目的の生成

5.1 で説明した手続きにより認定した談話目的セグメントに対して、談話下位目的の認定結果を利用して、目的記述を行う。

この際、談話目的セグメントが1つの談話下位目的セグメントしか持たなかった場合は、基本的に談話下位目的と談話目的は同一とする。

それ以外の、談話目的セグメントが2つ以上の談話下位目的セグメントを持つ場合は、基本的には以下の2つ方法を用いて談話目的が生成できないか検討する。

- 支配する談話下位目的のどれかを用いて代表させる
- 例えば、支配する談話下位目的セグメントが「X1のY1」及び「X2のY2」の2つである場合、「X1とX2」を包含する上位表現 X'、及び、「Y1とY2」を包含する上位表現 Y'を検討し、「X'のY'」形式で記述することを基本に、支配下の談話下位目的セグメントの題名を包含する記述を形式的に生成することを意識しつつ検討する。この場合、X1とX2、あるいはY1とY2がそもそも、同一であるか類似する場合も多い。

以上の統合方法で談話目的セグメントの意味内容を適切に示すような目的を生成できないときは、次のような方法により目的の生成を検討した。

- 談話目的セグメント内の表現から談話セグメントの範囲の意味内容を適切に示す目的を生成する

この方法で統合的な目的を生成する場合、談話セグメントの先頭で話者が「X1のY1」の談話下位目的セグメントだけでなくそれ以降の内容をも予告するような発話をしている場合がある。このような発話は、談話下位目的の認定時に、談話下位目的の先頭に存在するXの導入に関する補足、背景情報の提示、あるいは前置きと解釈できる部分である場合が多い。このような場合に限り、当該の部分を超えて上位の話題を導入している部分と考え、談話コメント属性の値として反映した。具体的には、上位の話題を導入する効果が及ぶ範囲が、談話目的セグメント内で、談話下位目的を統合しているとみなされる場合は「%I」を、当該の談話目的セグメント以降の複数の談話目的セグメントに関して影響をもつとみなされる場合は「%O」を値に設定した。

なお、以上の方法により作業ラベルと統合ラベルの談話目的、談話下位目的の認定との整合性を総合的にチェックし、整合性の原因が談話下位目的の不備によるものと考えられる場合は、談話下位目的の認定をやり直した。

## 6. おわりに

本稿で説明した、作業者ラベル付与における談話境界認定作業では、特定の言語表現やポーズをマニュアルに基準化して、談話境界認定を行う方法を、あえて採用しなかった。その理由は、談話セグメントを決定する上で決定的な観点があるわけではなく、複数の作業者が言語情報・パラ言語情報・意味内容から総合的に認定した結果を尊重すべきだという立場をとったからである。この結果は、分析者の観点や境界が必ずしも一致しない理由をモデル化する一助になるものと信じる。

他方、統合ラベルの付与では、談話セグメントの話題についての手続き的な分析方法を談話下位目的の認定に導入し、談話下位目的の認定結果と作業者ラベルとの整合性を検討した結果を統合的な境界情報として付与した。この統合ラベルにより、40 講演内での認定作業の整合性を意識した目的記述と、それぞれの談話の特徴をより個別的・具体的に記述したであろう複数の作業者の目的記述との差を検討する試料を提供できるものとする。

本稿でのこの実験的な試みは、『日本語話し言葉コーパス』作成のプロジェクトにおいて、対話処理、談話処理の研究者が共同して分析を行う機会を得たことにより可能となったと考えている。今後、この試みが、独話をより高い視点から分析する方法論として、精緻化・発展し、新しい談話分析の方法論を構築する礎になれば幸いである。

最後ではあるが、このようなプロジェクトの企画・遂行に携わった数多くの方に敬意を表し、このような分析を行う機会を与えていただいたことに深く感謝する。

## 参考文献

[GS 86] Grosz, B.J. & Sidner, C.L. Attention, intention, and the structure of discourse. *Computational Linguistics*, 12 (3), 175-204.

[森本 03] 森本郁代, 竹内和広, 高梨克也, 井佐原均. 「『日本語話し言葉コーパス』への談話構造タグ付与の問題点」. 人工知能学会研究会資料 SIG-SLUD-A302-09, 49-54.

[高梨 03] 高梨克也, 竹内和広, 森本郁代, 仲本康一郎, 井佐原均. 「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 6 回大会 Programs & Abstracts, 76-79.

付録：対象講演 ID 表

模擬講演(25 講演)				
S00F0209	S00F0210	S00M0065	S00M0071	S00M0117
S00M0213	S01F0157	S01F0166	S01F0183	S01M0051
S01M0227	S02F0100	S02F0189	S02M0011	S02M0161
S03F0119	S03F0214	S03M0089	S03M0098	S03M0194
S04F0013	S05F1600	S05M0412	S05M0613	S06F1034
学会講演(15 講演)				
A01F0145	A01M0025	A01M0056	A01M0070	A01M0137
A01M0157	A02F0116	A03F0072	A03M0005	A03M0059
A03M0138	A05F0043	A05M0031	A06F0028	A06F0073