



人文学のためのテキストデータ構造化のチュートリアル

第12章

談話資料の 構造化

中川奈津子

version 1.0

2026.3.21 作成

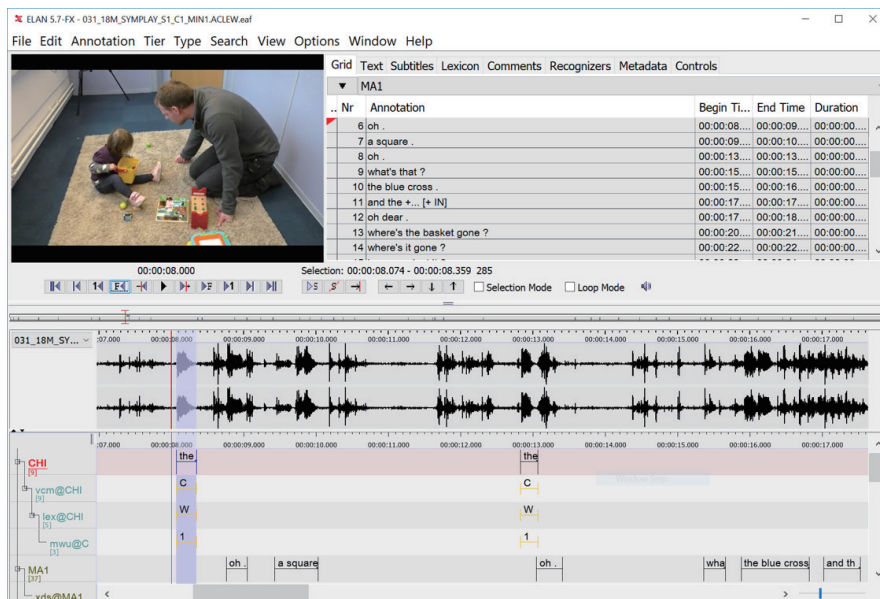
本資料は、文部科学省委託事業「人文学・社会科学のDX化に向けた研究開発推進事業」(JPMXP1624)において、学校法人慶應義塾が、大学共同利用機関法人人間文化研究機構から再委託を受けて作成したものです。本資料の利用にあたっては、出典を必ず記載するなど、「文部科学省ウェブサイト利用規約」を準用（ただし、商用利用は不可とする。）してください。

1. はじめに

本章では、マルチメディアデータ（音声・動画とそれに対応する書き起こしや注釈）の一例として、談話資料の構造とその符号化について述べる。マルチメディアデータとは、テキスト・画像・動画・音声など様々な形態の情報を統合したデータである。画像・動画・音声にテキストの情報が付与されているものなどがこれに該当し、人の会話の録音・録画とその文字起こしを統合したデータはその1例である。ここではこれを談話資料と呼ぶ。

2. 談話資料の作成

音声・動画データを書き起こしたり注釈を入れたりする際は通常、何らかの書き起こし用のソフトウェアを用いる。テキストエディタに音声・動画の時間情報を人手で書き入れながら書き起こすのは手間がかかり、また書き起こされた箇所の音声・動画を再生するのも大変なので、音声・動画の時間情報と書き起こし・注釈を統合してくれるソフトウェアは便利である。例えば筆者の専門である言語学では、ELAN¹⁾ という音声・動画の注釈ソフトがよく使われている。図1のように、ELANは音声波形が表示され、選択した音声を何度も聞き直しながらか該当箇所の書き起こしや注釈を行うことができる。音声と同期した動画も同時に再生することができる。このようなデータを研究に用いることで、文字情報だけでなく音声・動画の情報もすぐに見返すことができる。また、この研究データを公開することで、根拠データとして誰でもすぐに著者の主張を確かめることができる。例えばある話者は無声両唇破裂音 p の氣息が強いとか、ある箇所で母音が長く発音されたというようなことが、文字情報だけでなく、実際に当該箇所を聞いて確かめることができる。さらにこのELANの形式で音声・動画と文字情報をセットで公開すれば、当該の箇所もその周囲の文脈も音声・動画で確認することができる。文字情報を検索することにより、それに紐づいた音声・動画情報を参照することができるのがこのようなソフトウェアの利点である。ELANの他にも、Praatという音声に特化した注釈ソフトもあり、こちらは詳細な音声分析に利用されている。

図1 音声とそれに対応する書き起こし²⁾

ELAN や Praat で作成した注釈ファイル (eaf, TextGrid 形式) はテキストエディタで開くことができるので汎用性が高いと言えるが、より一般的には動画の字幕用に SubRip (srt) 形式が広く用いられている。SubRip 形式は、(1) ID、(2) 字幕の表示開始時刻 --> 表示終了時刻、(3) 1 行以上のテキスト、(4) 字幕の終了を示す空白行からなる。SubRip 形式は動画・音声の階層的な注釈 (具体例は後述) を行うには不十分である。ELAN や Praat の注釈は階層構造に対応しているが、ブラウザ上で音声・動画と連動して文字情報を HTML 形式で表示させたり、IIIF 動画ビューワに読み込ませたりするには何らかの変換が必要である。そこで、ELAN や Praat で作成されたテキストを、テキスト構造化の標準である TEI 形式で保存しておけば他の便利な形式への変換可能性も高まるので、TEI 符号化する価値がある。以下では、談話資料の構造を解説し、その符号化の例を提示する。

3. 談話資料の構造

談話資料の構造化の課題は、資料がマルチメディアデータであることの特徴に由来するもの、複数人による話しことばであることの特徴に由来するものと、データを階層化するという要請に由来するものがある。

3-1. マルチメディアデータの特性

音声・動画とその書き起こし・注釈などのテキストを統合するために、テキストデータ中に音声・動画に対応する時間情報、必要に応じて動画の位置情報を記録する要請が生まれる。例えば図1のように、ある時点の発話の内容をテキストで記録すると、いちいち聞き直さずにその箇所の発言内容を確認することができる。あるいはその箇所の発音の特性などを注釈として書き、特徴別に集計することもできる。これはテキストと音声・動画の時間情報が関連付けられていることにより可能になっている。動画の場合は時間情報とテキストの結びつきに加えて、動画の座標情報とテキストも関連させることができる。図2では地方選挙の様子を映した動画の投票用紙に注釈が施されている。



図2 動画へのアノテーション³⁾

このように、テキストとその他のメディアが統合されたマルチメディアデータには、テキストの2次元的な広がりだけでなく、時間的・空間的広がりが存在する。3Dデータの場合にはさらに扱う次元が増える。

話しことばの研究などにおいてテキストと音声・動画の結びつきが重視されるのは、テキストよりもむしろ音声・動画が一次資料であるとみなされるからである。ある表現がどのように発音されているのか、会話の参加者がどのように発話を重複させたり言いよんだり身振りを使いながら会話を進行させているのかを調べるには、抽象化された文字情報では不十分で、実際に見たり聞いたりできる音声・録音データのほうが調べたいことをより精密に表現している。しかしテキストデータは検索したり、一覧してざっと見るには便利であるため、テキストデータも必須の要素である。

3-2. 会話の特性

複数人が発話する会話においては、書きことばとは異なる構造が出現する。まず、会話においては話者交替が起こるため、どの発話が誰によるものなのかを発話ごとに明示する必要がある。書きことばとは異なり、話しことばはあらかじめ発話内容を計画して産出されるとは限らないため、(A) 言い淀み、(B) 言いかけて途中で計画を変更してからの言い直し、(C) 沈黙などが発生する。(D) 複数の話者が同時に発言する、発話の重複も起こる。(E) 笑い声や咳などの非言語的音声も産出される。(F) 同時発話、不明瞭な発話、言い間違いなどによって、書き起こしが困難な箇所も出てくる。さらに、(G) 会話中に電話がかかってくる、客が来る、救急車や飛行機が通過するなど突発的な事態が起こり、録音・録画が中断されたりもする。(H) 話者間で当然のこととして共有されている知識や背景状況などは言及されることがなく、文脈を補足するための注釈も重要である。

3-3. データの階層性

言語学では、理解できる人がごく少数しかいない言語を記録することが重要な課題の1つであるため、言語表現1つ1つの意味や機能を、理解できる人が多くいる共通語（日本語、英語など）で記述することがある。このとき、典型的に各発話は図3の構造を持つ（例は国立国語研究所(1980, p.171)から）。この例は1976年に研究者が京都府綾部市を訪れて録音した会話の一部で、会話は基本的にこの地域の言語で行われている。1行目に当該の言語の書き起こしがあり、適切な単位（ここでは文節）に分割される。2行目に、その単位ごとの逐語訳（グロス）が、共通語（ここでは日本語）で書かれる。3行目に、この発話全体でどのような意味になるか、共通語で書かれる（が、京都府綾部市の言語と日本語は、語順がほとんど同じでグロスだけでも意味が通るため、国立国語研究所(1980)には書かれていない）。

アンタ	、	オチャカ°	サメル	。	ヨバレテー。
あんた	、	お茶が	さめるわ	。	召し上がって [ちょうだい]
あんた、お茶が冷めるわ。召し上がってちょうだい。					

図3 グロス付き例文⁴⁾

この3層を基本構造とし、発音記号の層、形態素が分析された層、追加の言語（英語など）への翻訳層、注釈層などを追加することがある。

4. 符号化の例

ここまでの議論を踏まえたうえで、談話資料の TEI 符号化例を提示してみよう。符号化は中

川ほか (2023) に基づいているが、一部改変した。図 3 の 3 層の基本構造は、図 4 のように符号化できる。type="書き起こし" の div は図 3 の 1 行目にあたる。type="グロス" は 2 行目、type="共通語訳" は 3 行目にあたる。図 3 の例では文節ごとに分割されていたため <phr> (phrase) タグを用いたが、語ごとに分割されている場合は <w> (word) タグを用いても良い。

```
<div type="書き起こし">
  <u xml:id="発話 ID" who="発話者の ID" start="#時間 1" end="#時間 2">
    <phr>文節 1</phr> <phr>文節 2</phr> <phr>文節 3</phr> <pc>。</pc>
  </u>
</div>

<div type="グロス">
  <u corresp="#発話 ID">
    <phr>グロス 1</phr> <phr>グロス 2</phr> <phr>グロス 3</phr> <pc>。</pc>
  </u>
</div>

<div type="共通語訳">
  <u corresp="#発話 ID">
    共通誤訳
  </u>
</div>
```

図 4 談話資料の TEI 符号化の基本構造

この基本構造を図 3 の例に適用した例を図 5 に示す。

```
<div type="書き起こし">
  <u xml:id="u0417110" who="#B" start="#時間 1" end="#時間 2">
    <phr>アンタ</phr><pc>、</pc><phr>オチカ°</phr>
    <phr>サメル</phr><pc>。</pc><phr>ヨバレテー</phr><pc>。</pc>
  </u>
  <u xml:id="u0417111" who="#C">
    <phr>エー</phr><pc>。</pc><phr>ヨバレマス</phr><phr>ワ</phr><pc>。</pc>
  </u>
</div>
```

```
<div type=" グロス ">
  <u corresp="#u0417110">
    <phr> あんた </phr><pc>、 </pc><phr> お茶が </phr>
    <phr> さめるわ </phr><pc>。 </pc><phr> 召しあがって
    <add> ちょうだい </add></phr><pc>。 </pc>
  </u>
  <u corresp="#u0417111">
    <phr> はい </phr><pc>。 </pc>
    <phr> いただきます </phr><phr> わ </phr><pc>。 </pc>
  </u>
</div>

<div type=" 共通語訳 ">
  <u corresp="#u0417110">
    あんたお茶がさめるわ。召し上がってちょうだい。
  </u>
  <u corresp="#u0417111">
    <phr> はい </phr><pc>。 </pc><phr> いただきます </phr><phr> わ </phr><pc>。 </pc>
  </u>
</div>

<timeline xml:id=" タイムライン 1" unit="s">
  <when xml:id=" 時間 1" absolute=" 秒数 " />
  <when xml:id=" 時間 2" absolute=" 秒数 " />
</timeline>
```

図5 談話資料のTEI符号化の基本構造の例

3.2節で会話の特性としてあげた、(A) 言いよどみ、(B) 言いかけ、言い直し、(C) 沈黙、(D) 発話の重複、(E) 笑い、咳などの非言語要素、(F) 聞き取り困難箇所、(G) 録音・録画の中断、(H) 補足情報の符号化案を表1にあげる。

符号化対象の現象	TEI エlement
(A) 言いよどみ	<unclear> </unclear>
(B) 言いかけ、言い直し	<del type="言い直し">
(C) 沈黙	<pause dur="PT 秒数 S" />
(D) 発話の重複	<anchor /> による ID 指定
(E) 笑い、咳	<vocal> </vocal>
(F) 聞き取り困難	<unclear> </unclear>
(G) 突然の訪問客などにより、収録を中断した箇所	<incident> </incident><incident />
(H) 補足情報	<add> 訳の補足 </add> <note> 文脈の提供 </note>

表 1 符号化対象の現象と TEI の Element の対応

以下に (A) - (H) の実例をあげる。例はすべて国立国語研究所 (1980) の京都府綾部市の言語であり、中川ほか (2023) の成果に基づいている。なお、本来 <u> の属性として書かれているべき発話開始、発話終了時間は煩雑なので省略した。図 4 の基本構造に埋め込まれた一部だけを表示していることに注意されたい。

4-1. (A), (F) 言いよどみ、聞き取り困難箇所

言いよどみや聞き取り困難な箇所は <unclear> タグで囲んだ。正確には発話計画中であるなどの理由で話者が言いよどんだ (その結果聞き取りが困難になった) 箇所と、書き起こし手の能力や録音機の限界で聞き取り困難な箇所は意味的に異なり、元データである国立国語研究所 (1980) では区別されているが、ここでは一括して同じタグを付与した。図 6 にその例を示す。

```
<u xml:id="u0418207" who="#C">
  <phr>ヨイ</phr><phr>ウチニ</phr>
  <unclear>イヤハン</unclear>
  <phr>ノ</phr><pc>。</pc><phr>ソレヤツタラ</phr><pc>。</pc>
</u>
```

図 6 言いよどみ・聞き取り困難箇所の符号化例 (04IIKyoto02.xml)

4-2. (B) 言いかけ、言い直し

話者は発話途中に計画を変更し、発話を途中で中断して別のことを言い直すことがある。その中断された発話には図 7 のように タグを使用した。

```
<u xml:id="u0416602" who="#B">
  <phr>セヤケドー</phr><phr>アノ</phr><phr>オヨキ°ワ</phr><phr>トモカク</phr>
  <phr>アノー</phr><phr>クロタニット</phr><phr>ユー</phr><phr>トコワ</phr>
```

```

<del type=" 言い直し "> エ </del>
<phr> アノ </phr><phr> チカイ </phr><phr> シンルイカ° </phr>
<phr> ナー </phr><pc>。 </pc>
( 後略 )
</u>

```

図7 言いかけ・言い直しの構造化例 (04IIKyoto01.xml)

4-3. (C) 沈黙

話者は発話計画中や、会話の前提に齟齬があるなどの問題が発生したときに沈黙することがある。その際には <pause> タグを使用した。このタグは開始・終了を持たない空白タグである。図8では、属性に沈黙の長さを示す @dur を指定している。"PT5S" は ISO 8601 に従った時間の長さの書き方で、ここでは5秒間の沈黙があったことを示している。

```

<u xml:id="u0419605" who="E">
  <phr> アー </phr><pc>。 </pc>
  <phr> ソーデス </phr><phr> カ </phr><pc>。 </pc>
  <pause dur="PT5S"/>
  <phr> アノー </phr><phr> ウンドーカインアンテ </phr><phr> ユーノワ </phr>
  <phr> ドーユー </phr><phr> コト </phr><phr> ヤットリマシタ </phr>
  <phr> カ </phr><pc>。 </pc>
</u>

```

図8 沈黙の構造化例 (04IIKyoto03.xml)

4-4. (D) 発話の重複

複数人で行われる会話では、複数の話者が同時に発話することがよくある。これにより、話者の相槌のタイミング、他者の発話が終了することの予想などを読み取ることができ、会話を詳細に分析する研究者にとっては有用な情報となりうる。図10が発話の重複の構造化例であるが、込み入っているので、まず構造化する対象の言語現象を図式化したものを図9に示す。

C:	ヨネン	ソツキ° ヨーシタラ
	ホシューカエ	イツ [テー]
A:		[エー]
D:		[フン]

図9 発話の重複 (04IIKyoto03.xml)

この例では、話者Cが「ヨネンソツキ° ヨーシタラ ホシューカエイッター」(四年 [を] 卒業したら 補修科へ行って) と発話した「テー」の部分に、話者AとDがそれぞれ「エー」「フ

ン」と相槌を重複させている。これを表現するために、Cの「テー」の重複(overlap)開始・終了箇所に<anchor /> タグを配置し、それぞれ@type属性を"overlapStart"(重複開始)、“overlapEnd”(重複終了)と指定し、@xml:id属性の値を指定する。次に、AとDの「エー」「フン」という発話に、「テー」と同時発話していることを示すために、属性@sychによって、Cの重複開始・終了位置の@xml:idの値を<anchor /> タグの属性として指定する。これで図10の符号化が完成する。

```
<u xml:id="u041950" who="#C">
  <phr> ヨネン</phr><phr> ソツキ° ヨーシタラ</phr><phr> ホシューカエ</phr>
  <phr> イツ
    <anchor type="overlapStart" xml:id="u041950os"/>
    テー
    <anchor type="overlapEnd" xml:id="u041950oe"/>
  </phr>
  <phr> ノーカ° ツコーエ</phr><phr> イツタンヤデ</phr><phr> ナー</phr><pc>。</
pc>
</u>

<u xml:id="u041950-1" who="A">
  <phr>
    <anchor type="overlapStart" sych="#u041950os"/>
    エー
    <anchor type="overlapEnd" sych="#u041950oe"/>
  </phr><pc>。</pc>
</u>

<u xml:id="u041950-2" who="D">
  <phr>
    <anchor type="overlapStart" sych="#u041950os"/>
    フン
    <anchor type="overlapEnd" sych="#u041950oe"/>
  </phr><pc>。</pc>
</u>
```

図 10 発話重複の構造化例 (041IKyoto03.xml)

4-5. (E) 笑い、咳などの非言語要素

笑い、咳などの非言語要素は <vocal> によって符号化された。この要素は、音声ではあるものの必ずしも語彙項目として書き起こせない発話のためのタグで、他にも鼻すすり、舌打ち、ため息などがこれによって符号化されうる。図 11 の例では、笑いを符号化している。

```
<u xml:id="u0421611" who="#E">
  <phr> ナカナカ </phr><phr> オクサン </phr><phr> オモシロイ </phr><pc>。 </pc>
  <vocal><desc>laugh</desc></vocal>
  <phr> チャント </phr><phr></phr><phr> カキノ </phr><phr> キカ° </phr>
  <phr> アリマシタ </phr><pc>、 </pc>
  <vocal><desc>laugh</desc></vocal>
</u>
```

図 11 非言語要素の構造化例 (04IIKyoto04.xml)

4-6. (G) 収録中断箇所

日常の風景を収録していると、予期しない事態が起こって収録を中断せざるを得ないときがある。突然の来客、電話、飛行機や選挙カーの通過など、公開する音声・動画に入れるべきでない判断された箇所は、<incident> タグにより、中断したことを示すことができる。図 12 の例では中断の理由が示されていないなかったため、空白タグ <incident /> が用いられているが、開始タグと終了タグを用いて偶発的出来事の内容を記述しても良い。

```
<u xml:id="u0420505" who="#D">
  <phr> エー </phr><pc>。 </pc><phr> オンナノ </phr><phr> コ </phr><pc>。 </pc>
  <incident />
</u>
```

図 12 収録中断箇所の構造化例 (04IIKyoto03.xml)

4-7. (H) 補足情報

グロスや共通語訳において、当該言語の表現の理解を助けるために、短く共通語が補われることがある。このような補足情報は図 13 のように <add> タグで表現した。

```
<div type="書き起こし">
  <u xml:id="u0417007" who="#E">
```

```

    <phr> ココじゃ </phr><phr> ナイ </phr>
    <phr> ワケ </phr><pc>。 </pc>
</u>
</div>

<div type=" グロス ">
<u corresp="#u0417007">
    <phr> ここじゃ </phr><phr> ない </phr>
    <phr> わけ <add> ですか </add></phr><pc>。 </pc>
    </u>
</div>

```

図 13 訳の補足の構造化例 (04IIKyoto01.xml)

注釈を行う者が、さらに詳しく、用いられている表現の文法事項や詳細な発音、理解の助けとなる文脈情報や、発話者の意図などを書いた場合には、<note> タグも用いることができる。

```

<u xml:id="u0425405" who="#C">
    <phr> アー </phr><pc>、 </pc><phr> ソー </phr> カ <pc>。 </pc>
    <phr> フン </phr><pc>。 </pc>
    <phr> ホラ </phr><phr> アカン </phr><phr> ナー </phr>
    <note n="1"> [1] 「話題にならない」の意。 </note>
    <pc>。 </pc>
</u>

```

図 14 文脈・話者の発話意図の補足の構造化例 (04IIKyoto07.xml)

5. メタデータ

メタデータ一般に関する説明は第 6 章に譲り、談話資料においては、音声・動画の時間情報とテキストが結びついているので、ヘッダでは本文 <body> で音声・動画に記述された時間情報がどのファイルを参照しているのかを指定する。図 15 では、<recordingStmnt> の中の <recording> 要素に音声と動画の情報が記述されている。1 つ目の <media> は音声ファイルのためのもので、@xml:id 属性により ID を指定している。そしてこの ID を図 4 の <timeline> の中の属性 @corresp で参照し、どのファイルの時間を指定しているのかを明示している。@url により音声ファイルの場所を指定、@mimeType によりこのファイルが wav 形式の音声ファ

イルであることを示している。2つ目の <media> は動画ファイルのために、動画ファイルの同様の情報を記述している。

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title> タイトル </title>
    </titleStmt>

    <publicationStmt>
      <publisher> 出版者 </publisher>
    </publicationStmt>

    <sourceDesc>
      <recordingStmt>
        <recording xml:id="recording">
          <media xml:id="audio"
            url="audio.wav"
            mimeType="audio/wav"/>
          <media xml:id="video"
            url="video.mp4"
            mimeType="video/mp4"/>
        </recording>
      </recordingStmt>
    </sourceDesc>
  </fileDesc>

  [...]

</teiHeader>
```

図 16 音声・動画ファイルをふくむ teiHeader の例

注

- 1 <https://archive.mpi.nl/tla/elan>
- 2 Image credit: The Language 0-5 project / The Language Archive. Licensed under CC BY 4.0.

3 「iiif-prezi3 を使って、動画にアノテーションを付与する」(中村覚氏による記事)

<https://zenn.dev/nakamura196/articles/b7f52827f1ff80> (2026/1/1 最終アクセス)

4 [] は逐語訳では不十分な表現を訳者が補った箇所である。

参考文献

・国立国語研究所, 国立国語研究所資料集 10 方言談話資料 (4) —福井・京都・島根—

<https://doi.org/10.15084/00002273> (1980) .

・中川奈津子ほか (2023) 日本語方言談話資料の TEI による構造化の試み. 『じんもんこん 2023 論文集』, 83–90.

謝辞

TEI 協会東アジア／日本語分科会主催の TEI 研究会で符号化した例を使用しています。研究会参加者の皆さまに感謝します。