

# 実対話コーパスに対する意味単位・構造情報のタグ付け

牧本慎平, 柏岡秀紀, ニック・キャンベル  
奈良先端科学技術大学院大学 情報科学研究科  
{shimpei-m,kashioka,nick}@is.naist.jp

## 1 はじめに

対話は人間にとって最も一般的な情報伝達方法の一つであり、大量の情報のやり取りが対話によって行われている。今後音声認識技術の発達とともに、より高精度な対話のディクテーションが行えるようになることが予想され、書き起こされた大量の対話情報を扱うことが容易になると考えられる。このような対話情報を対象に機械翻訳や情報抽出などのアプリケーションの研究・開発を行うことは有用であるが、対話の構造は書き言葉と異なり文法的に厳密な構成になっておらず、計算機による処理が困難だという問題がある。このため、対話に対して自然言語処理のアプリケーションを適用する前段階として、対話の抽象化・構造化を行い、対話がどのような流れで行われているのかを理解することが必要である。

発話単位タグ [1] などをはじめとして、対話・談話の構造を表現するという試みは行われてきているが、その大半が何らかの目的を指向したものであり、雑談のような日常の大部分を占める目的・指向のない会話に関して適応できない場合が多い。また、それらを自動的に獲得する研究 (例えば [6] など) も同様に多数行われているが、高い精度を得ているとは言いがたい。

本研究の目的は実対話、主に日常会話などに見られる強い目的を持たない自由対話に対してその構造を明確に表現し、また自動獲得が可能なタグ付けを行うことにより、対話情報を容易に扱うことを可能にすることである。そのために我々は現在、実対話コーパスに対して構造情報のタグ付けを行っている。

我々が対象としているコーパスは電話を介して行われた内容指定のない自由対話であり、フィラーや言い誤りなど話し言葉特有の現象が含まれ、発話プランの中止や変更、割り込みなどが行われており、明確な構造になっていない、言わば断片化された状態にあり、意味のある情報が時間的に離れた状態で発せられることがある。

本稿では我々が実対話に対し行っている意味単位・構造情報のタグ付けについて述べ、そこから得られた知見を述べる。

## 2 自由対話コーパス

我々は自由対話コーパス ESP\_C [3][4] に対し、対話の意味単位・構造情報の付与を行っている。

ESP\_C は JST/CREST による「表現豊かな発話音声のコンピュータ処理システム」プロジェクト\*1によって構築された JST/ATR ESP Corpus のサブセットである。このコーパスは 2 名の話者による実対話の録音と書き起こしのセットである。話者は合計 10 名 (男女ともに 5 名ずつで日本語を母語としない者 4 名を含む) で、互いに面識のない状態で、電話による 1 セッション 30 分の対話を行った。コーパスには各組み合わせ週 1 回でおおよそ 10 回 (日本語の母語としない者を含む組み合わせでは 5 回) のセッションが収録されている。

本研究の対象となっているのは収録されていることを意識しない自由発話で構成される対話である。大規模な対話コーパスには Call Home や Switchboard Corpus [5] などがあるが、これらは話す内容について何らかのディレクションがあり、完全に自由な対話とは言えない。話題について指示がされていない対話は、発話プランが明確に立てられないため、一般に発話が断片化されやすいと考えられる。そのため、実際の雑談に近い状態で収録された ESP\_C に対して構造化を試みている。

## 3 意味単位・構造情報のタグ付け

### 3.1 タグ付けの方針

人間の対話には、表面に出現しない様々な世界知識や認知的な要素などが関係して生じる現象が多分に含まれる。しかしながら、計算機によってそれらを解析するのは困難である。従って、発話の音声や書き起こし文を参照することによって構築できる意味・構造をタグ付けを行う。つまり、音声や書き起こし文といった比較的容易に獲得可能な表層的現象から、その対話の構造や流れを理解することを目的として、対話コーパスに対してタグ付けを行う。また、これらのタグ付けは既存の意図情報などのタグと共存できるようなものとする。

\*1 <http://feast.atr.jp/esp/>

タグ付けはパタンマッチなどによるヒューリスティックな手法を施した後、人手による修正を加えるという半自動の手法で行った。

### 3.2 タグの仕様

以下に我々が採用しているタグの仕様について述べる

#### 3.2.1 時間発話単位

対話コーパスから最も客観的につけられる単位として、時間発話単位(タグ名は *utter*) を付す。これは、各話者ごとに 300ms 以上の発話されていない時間があれば、機械的にそこを発話の切れ目と処理するものである。

#### 3.2.2 フィラー、笑い、言い淀み

話し言葉、特に目的が明確でない場合の対話において、フィラーなどの発話と発話の間に生じる情報を持たない発話現象は多く見られる。我々のタグ付けでは、ESP\_C 書き起こし文に記述されているフィラー、笑い、言い淀みについて、それぞれ *filler*, *laugh*, *disfluency* というタグを付した。これら 3 つは発話間を埋める現象という意味合いで捉えれば広義には同一の現象と考えることができる。

話し言葉のフィラーなどを自動抽出するという研究はなされているが [2]、今回は単純なパタンマッチの結果を人手で修復するという方法でタグ付けを行った。

#### 3.2.3 断片

断片 (*fragment*) は我々のタグ付けにおける発話の最小の単位である。発話において、1 つの意味は 1 つ以上の断片から構成されると考えることができる。断片タグについては以下の規則を基に付した。

- 時間発話単位をまたがない。(300ms 以上の間を空けない)
- *filler*, *laugh*, *disfluency* が生じたらそこまでが 1 つの断片となる。
- 助動詞「です」或いは終助詞があった場合、そこまでが 1 つの断片となる。ただし、対話の場合、その後前発話を修飾する語が頻出するので、1 つの内容語で完結する修飾に限り、前の断片に加える。

#### 3.2.4 関係属性

*fragment* に対して他の *fragment* に対する関係性を示す属性を付す。これらは一方の発話から対象となる発話へのポイントの役割を果たす。関係は以下の 5 種類と定義される。

結合 (*combine*) 相手の割り込みやフィラーなどによって 1 つの意味の発話が複数の断片に分かれてしまったとき、それらを接合する。

働きかけ (*approach*) 質問や確認など発話の相手に何らかの反応を求める発話について、対応する反応の断片を値とする。

参照 (*refer*) 特に働きかけを行ってない発話断片を参照した発話についてつける。

継承 (*succeed*) 自身の発話を受け、連続して発話する場合につける。

換言 (*paraphrase*) 自身の発話の途中で発話の一部を言い換えて新たに発話する場合、言い換えもとの発話についてつける。

これらの関係は比較的表層的な情報のみで判別できると考えられる。例えば、「……でしょう」という表現があった場合、質問と自身の意見の主張とどちらでもありえる。しかし、これらを判別するには文脈の理解が必要となり、計算機では困難である。そのため、一律に働きかけという関係を付した。今回はコーパスに対して手作業にて関係性をつけたが、今後自動での関係獲得を目指す予定である。

### 3.3 タグ付けの結果

前節までに述べたタグ付けを ESP\_C コーパスから選択した 2 人の男性話者による 30 分間の対話 1 セッションに付した。表 1 にセッションに含まれる各タグの数を、表 2 にその中の断片タグに伏せられた関係属性の数を示す。

表 1 1 発話内の話者ごとのタグの数

	utter	frag.	filler	laugh	disfl.
話者 A	414	606	80	114	37
話者 B	439	564	227	168	44
合計	853	1169	357	283	81

表 2 1 発話内の話者ごとの関係属性の数

	comb.	appr.	refer	succeed	paraph.
話者 A	52	47	88	202	42
話者 B	67	100	105	85	26
合計	119	147	193	287	68

また、以上のようなタグ付けを行うことによって、対話は発話断片をノード、関係性をエッジとした図 1 のような有向グラフの形状 (対話グラフ) になる。

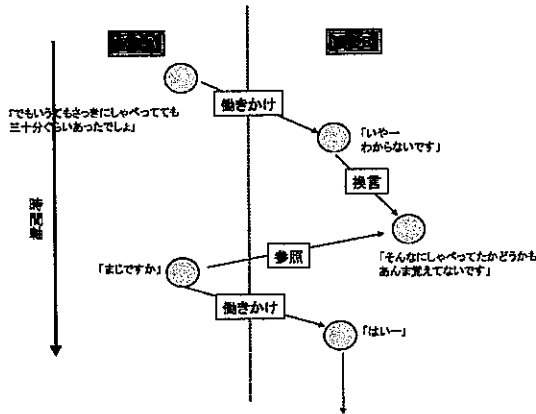


図1 対話グラフの例

#### 4 考察

タグ付けした対話データから得られる知見を述べる。このセッションでは話者間の発話数及び発話断片の数がほぼ同数である。これはそれぞれの話者の発話する機会がほぼ同じであったことを示している。しかしながら、関係属性の数を見ると話者Aは継承の関係が多く、話者Bは働きかけの関係が多いことが分かる。これらにより、このセッションでは話者Aが話者Bの問いかけに答えるという形式で行われたことが予想される。このように、関係属性によってその対話の様式を推定することが可能である。

また、フィラーの生じた数についても発話者それぞれの特徴を捉えることができる。例えば、話者Aと比較して話者Bのフィラーの数は2倍以上である。フィラーなどの発話現象は話者固有の性質を持っていることが考えられる。

結合、継承、換言については、自身の別の発話断片について言及している関係である。従って、これらを1つのチャンクと捉えることが可能であり、そのチャンクは1つの意味を持ちうる(意味単位)。また、意味単位を働きかけによって相手の発話と繋げることによって、一種の対話単位を得ることができる。

これらの構造の例を図2及び図3で見ることが出来る。ここで、円は発話断片で、実線の四角形で囲まれた部分は意味単位(断片1つで意味単位を形成しているものについては四角を省略)、破線の四角形で囲まれた部分是对話単位である。また、実線の矢印は結合、継承、換言、働きかけのいずれかであり、破線の矢印は参照を意味する。

図2では両話者が互いに働きかけを行っている状態

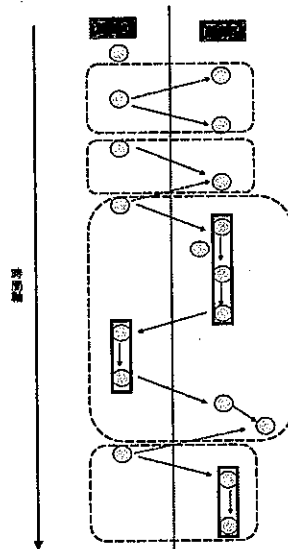


図2 対話構造の例(伝達層に重点が置かれた状態)

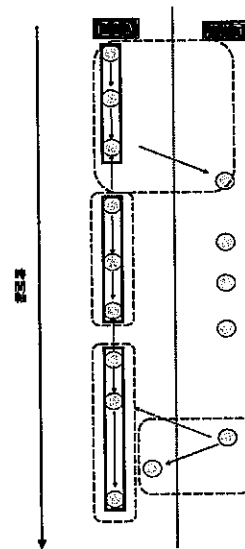


図3 対話構造の例(内容層に重点が置かれた状態)

である。一般に意味単位も対話単位も短く、前の話題を参照しつつ会話を進行させている状態である。一方、図3では話者Aが中心となって話を進めている。また、話者Aの話の合間に両話者間で働きかけなどが行われている。これらから、対話には伝達層と内容層が存在していると解釈することができる。一般に、伝達層では話者間のコミュニケーションを円滑にすることと知識の共有が行われており、内容層では一方の話者

による知識の提供が行われていると考えられる。

## 5 まとめと今後の課題

本稿では、対話書き起こし文の表層的な情報のみで獲得できる実対話コーパスに対する意味単位・構造情報の新たなタグ付けの手法とその特徴について述べた。これによって、明確な構造化がされていない実対話において、その会話の流れや情報として処理できる部位を発見することができる。しかしながら、現段階では人手で行わなければならない作業が多いため、対話の解析に大きなコストがかかってしまう。今後は、より大規模な対話データにタグ付けを行うと同時に、それをもとに高精度な自動タグ付けの手法を検討していきたい。

## 参考文献

- [1] 荒木雅弘, 伊藤敏彦, 熊谷智子, 石崎雅人. 発話単位タグ標準化案の作成. 人工知能学会誌, Vol. 14, No. 2, pp. 251-260, 1999.
- [2] Masayuki Asahara and Yuji Matsumoto. Filler and disfluency identification based on morphological analysis and chunking. In *Proc. ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition*, pp. 163-166, 2003.
- [3] Nick Campbell. On the use of non-verbal sounds in Japanese conversational speech. 第8回音声言語シンポジウム・第64回音声言語情報処理研究会, 2006.
- [4] Nick Campbell. Selecting speech fragments for affect display in concatenative expressive speech synthesis. 日本音響学会 2007 年春季研究発表会, 2007.
- [5] J. J. Godfrey, E. C. Holliman, and J. McDaniel. SWITCHBOARD: Telephone speech corpus for research and development. In *Proc. IEEE-ICASSP*, Vol. 1, pp. 517-520, 1992.
- [6] Ken Samuel, Sandra Carberry, and K. Vijay-Shanker. Dialogue act tagging with transformation-based learning. In *Proc. COLING-ACL*, pp. 1150-1156, 1998.