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手段の 1 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．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可能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，対話では言語情報は様々な要因によって断片化しており，計算機による処理が困難だという問題がある．このため，対話に対して様々な応用を行う前処理として，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解析することが必要である．

発話単位タグ [1] 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いるが，対象とする対話・談話が何らかの目的を指向したものであり，雑談など明確な目的のない会話に関して適応できない場合が多い．また，自動的にタグ付けを行う研究 (例えば [6] など) も同様に多数行われているが，会話の流れや発話者の意図などがタグ付けに必要なため，高い精度を得ているとは言えない．

本研究では日常会話などに見られる明確な目的のない自由対話に対してその構造の表現方法を考案する．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自由対話であり，フィラーや言い淀み，発話プランの中止や変更，割り込みなどのため，発話の意味情報が断片化している．本稿では，発話構造の自動獲得のモデル構築のための資源として，実対話コーパスに対して付するタグの仕様を提案する．

2 自由対話コーパス

我々は電話対話コーパス ESP_C[3][4] を対象とした，対話における意味単位・構造情報のタグの仕様を提案する．

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 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目的が存在する．明確な目的のない対話は，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．従って，我々は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．

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 名の話者間の日本語による実対話の録音と書き起こしのセットであり，CALLHOME[7] などと類似したコンセプトである．話者は収録以前には面識はない合計 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) である．それぞれの 2 名ずつの組み合わせで 1 セッション 30 分の電話による対話を行った．各組み合わせ週 1 回の頻度で合計 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た．

本研究では日本語を母語とする男性話者 2 名による対話を対象にタグを付した．

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，発話の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要因となっ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深層的な現象を解析するのは困難である．従って，対話コーパスに対して，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・構造のアノテーションの仕様を提案する．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．

今回，関係属性の付与については手作業にて行い，その他の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によって処理を施した後，人手による修正を加えるという半自動の手法にて行った．

[*1] http://feast.atr.jp/esp/

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 ポーズによる発話単位

対話コーパスから最も客観的につけられる単位として，**ポーズによる発話単位**(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言語的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，対話コーパス内のフィラー，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．

- **ポーズによる発話単位**を跨がない．つまり，300ms 以上の間があれば分割する．
- *filler*, *laugh*, *disfluency* が発話単位内にあれば，そこで断片は途切れる．
- 助動詞「です」，終助詞，接続助詞があった場合，そこまでが 1 つの断片となる．ただし，対話の場合多く見られる，前の断片を修飾する句はその断片に加える．例えば，**「負けずぎらいっていうのはよくいわれーましたね中学んときとか」**という発話があったとき，**「中学んときとか」**は終助詞**「ね」**の後にあるが，この発話は 1 つの断片とする．

3.2.4 関係属性

断片に他の断片に対する関係性を示す属性を付す．これらは一方の断片から対象となる断片へのポインタの役割を果たす．

対話内の発話の関係性は複雑であり，厳密にすべての属性を論じることは困難である．本研究では以下の 5 つの関係属性を定義した．これらの関係属性は文脈によらずにテキストの表層的な情報から知ることができると考えられる．現段階での関係属性の定義では対話内のすべての事象を言及できるとは言えず，細分化や新たな属性を導入する必要がある可能性もあるが，より詳細な関係属性定義は今後の課題とする．

結合 (combine)　相手への働きかけなどによって 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新たな表現に換言されるときつける．

以上のようなタグ付けを行うことによって，対話は断片をノード，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．

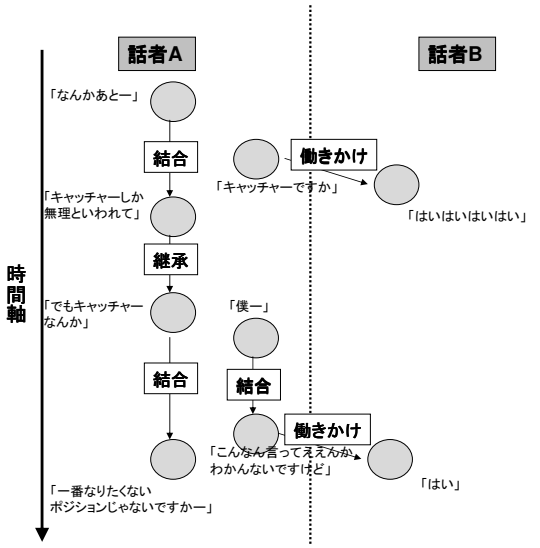

図 1　対話グラフの例

3.3 対話の層構造と意味単位

対話構造において，関係性が発話者自身の中で閉じているもの (**話者内関係**) と 2 話者間で行われているもの (**話者間関係**) が区別されると考えられる．話者内関係は，話者自身が組み立てた内容に起因し，一方，話者間関係は相手話者とのインタラクションの中で成立する．

我々の定義した 5 つの関係属性では**接合**，**参照**，**継承**，**換言**が話者内関係になり得る．また，**参照**，**働きかけ**が話者間関係になり得る (**参照**はどちらにもなり

得る).

また，対話には**伝達層**と**内容層**(更には，フィラーなどの含まれる**非言語情報層**) が存在していると解釈することができる．伝達層では，発話は話者間関係によって相手話者の発話に連結され，話者間のコミュニケーションの円滑化が主に行われている．内容層では，話者内関係によって構成された一方の話者による知識の提供が主に行われている．従って，内容層に注目すれば，発話の持つ 1 つの言語的意味のある情報 (**意味単位**) を獲得することができる．

例えば，図 2 は図 1 の同じ例であるが，断片**#1, #4, #5, #9** の流れは内容層に含まれ，**#2**(**「キャッチャーですか」**) や**#6** と**#7** の組み合わせ (**「僕ー/こんなん言ってええかわかんないすけど」**)，それらへの相手のリアクションである**#3**, **#8** などは伝達層に含まれている．ここで，断片**#1, #4, #5, #9** の一連の流れ (**「なんかあとー/キャッチャーしか無理といわれて/でもキャッチャーなんか/一番なりたくないポジションじゃないですかー」**) は 1 つの意味単位であると考えることができる．

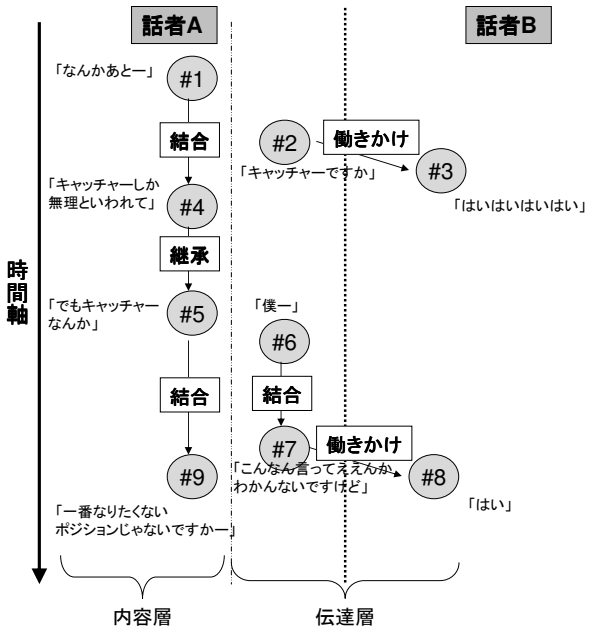

図 2　対話の層構造

4　タグ付けの実施と考察

我々は提案するタグを 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 セッションに対して付した．表 1 にセッションに含まれる各タグの数を，表 2 にセッション内の断片タグに付された各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発話属性の数

	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

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，実際このセッションでは話者 A の話題が中心となる傾向があった．このように関係属性の出現数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が話者固有の性質を持っているかは現段階では判断できないが，多くの発話に対しタグ付けを行った後の課題とする．

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の仕様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．また，今回定義した関係属性だけでは記述できない関係性も存在するため，新たな関係属性を提案する必要がある．今後は，

より厳密に関係性を記述できる属性を考案し，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.

[7] Barbara Wheatley, Masayo Kaneko, and Megumi Kobayashi. *CALLHOME Japanese Transcripts*. Linguistic Data Consortium, Philadelphia, 1996.