

コーパスに基づく雑談を目的とした発話役割同定

平尾 卓也[†] 松本 和幸[‡] 北 研二[‡] 任 福継[‡]

[†] 徳島大学大学院 先端技術科学教育部

[‡] 徳島大学大学院 ソシオテクノサイエンス研究部

[†]hirao-takuya@iss.tokushima-u.ac.jp

[‡]{matumoto, kita, ren}@is.tokushima-u.ac.jp

1 はじめに

従来の対話研究では，学内案内システムや車内案内などのように対話を通して何らかのタスクを遂行することを目的としてきた．しかし，近年の対話研究では 1 人暮らしの高齢者の話相手のように対話すること自体を目的とした研究も盛んに行われるようになってきた．

対話とは「二人の人がことばを交じわすこと」とされている [1]．対話ではことばを交じわす事が必要であり，ことばを交じわさず話す場合は対話とはいえない．また，発話に対して無関係な回答をされると不快に感じ，話を続けることは困難であると考えられる．円滑な対話をするためには，たとえば [依頼] された場合は [依頼] に対する断りや受け入れなどの返答をすることが必要である．つまり，雑談を行っていく中で，対話者の発話した内容の発話役割を同定することは円滑な対話を進めるうえで必要不可欠である．

我々は，コーパスに基づく雑談を目的とした発話役割の同定の研究を行っている [2]．西原らは発話役割の一部は助詞・助動詞の組み合わせで表されるとし，助詞・助動詞の意味を用いたルールによって発話役割の同定を行った [3]．そこで本稿では品詞情報を用いた雑談の発話役割に有効な重みを模索し，コーパスに基づく雑談を目的とした発話役割の同定手法を提案する．

2 発話役割の定義

人間が発話した文には機能がある事がほとんどである．例えば「今日は晴れです。」という文は [事実] という機能があり，「返事お願いします。」という文は [依頼] という機能がある．このように話者視点での文の持つ機能のことを本研究では発話役割と定義する．

3 提案手法

3.1 発話役割コーパス

本研究で使用するコーパスを発話役割コーパスとし，下記のように構築した．まず，Web 掲示板からテキス

トの収集を行う．本研究の目標は雑談における発話役割の同定であるため，収集するテキストは雑談である必要がある．そのため，ブログなどのように一名が書いているものは対象とせず，二名以上が対話のようなやりとりをしている掲示板を対象とした．このように収集したテキストの 1 文ごとに人手で発話役割タグの付与を行う．また，文の表す発話役割が複数あると判断出来る場合は適当と思われる発話役割タグのすべてを付与する．本研究では [依頼]，[事実]，[提案]，[心情]，[教える]，[確認] の 6 種類の発話役割を使用する．

3.2 発話役割の同定手法

図 1 の手順で発話役割の同定を行う．

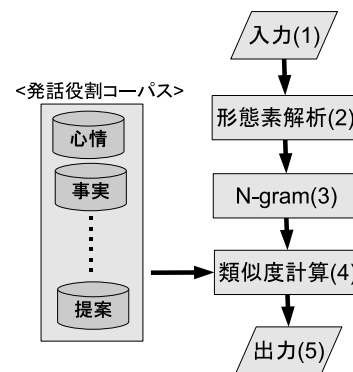


図 1: 発話役割同定の流れ

- (1) 入力
任意の 1 文を入力とする．
- (2) 形態素解析
入力された文を MeCab[4] を利用して形態素解析し，形態素表層形と品詞情報の抽出を行う．
- (3) N-gram
形態素表層形を利用して形態素 N-gram に分割する．

(4) 類似度計算

入力文 D とコーパス内のある文 E の類似度を算出する．入力文 D の形態素 N-gram の頻度ベクトルを $\vec{D} = (wx_1, wx_2, \dots, wx_n)$ で表す (w : 重み)．文 E の形態素 N-gram の頻度ベクトルを同様に表したとき，式 (1) を用いてコサイン類似度の計算を行う．

$$\text{sim}(\vec{D}, \vec{E}) = \frac{\vec{D} \cdot \vec{E}}{|\vec{D}| |\vec{E}|} \quad (1)$$

重み w は形態素の品詞情報により変化させる．

(5) 出力

入力文とコーパス内のすべての文に対して類似度計算を行う．その結果から発話役割タグごとに最も類似度の高い数値を取り出す．その数値を比較し最も高い数値を持つ発話役割を入力文の発話役割と同定する．

4 評価実験

重みを品詞ごとに決定し発話役割を同定する手法が有効であるか評価するために，品詞を考慮せず重みを一定にした場合との比較を行った．

助詞・助動詞は話者の心情などの一定の意味を添える場合や，判断を表現するために使われる語のため発話役割にとって重要である場合がある．名詞は人名や場所など事物の名を表すことで多く使われるため発話役割にとって重要でない場合が多い．そこで助詞・助動詞の重みを重くし，名詞の重みを軽くする．

4.1 実験条件

本稿では形態素 2-gram を用いて実験を行った．発話役割コーパス内の文に付与された発話役割タグの数は各 600 個に統一した．またコーパス内にテストデータは含まない．人手で付与した発話役割タグを正解とし正否の判断を行った．

4.2 実験結果

品詞によって重みを変化させて発話役割の同定を行った結果を表 1 に示す．

表 1: 結果

	(i)	(ii)	(iii)	(iv)
正解率	41.66(%)	40.78(%)	42.60(%)	42.03(%)

(i) 重みを一定にした場合

(ii) 名詞を含む形態素 2-gram の重みを $\frac{1}{2}$ 倍にした場合

(iii) 助詞または助動詞を含む形態素 2-gram の重みを 2 倍にした場合

(iv) 条件 (ii) と条件 (iii) を合わせた場合

4.3 考察

条件 (i)，条件 (ii)，条件 (iii)，条件 (iv) で正解率はほとんど変わらない結果となった．またすべての条件において，発話役割ごとの正解率は [心情] がもっとも悪く，30%程度となった．悪くなった原因としては，[心情] は文で使用される形態素にばらつきがあるため同定が困難であると考えられる．条件 (i) では 854 文を [心情] と誤って同定したが，条件 (iii) では誤りは 691 文に減っている．

5 おわりに

本稿ではコーパスを用いた発話役割の同定手法について提案した．形態素 N-gram に含まれる助詞・助動詞の重みを 2 倍，名詞の重みを $\frac{1}{2}$ 倍と変化させて実験を行った結果，正解率は 42.03%となった．今後の課題として重みの作成方法や，重み付けをする品詞を増やすことで精度向上を目指したい．

謝辞

本研究の一部は，科学研究費補助金・挑戦的萌芽研究（課題番号：21650030）により行われた．

参考文献

- [1] 広辞苑（第五版），1998，株式会社 岩波書店，新村出編
- [2] 平尾卓也，任福継，“自由発話できる雑談を目的とした発話役割同定について”，電気学会電子・情報・システム部門大会論文集，pp.1365-1366，2009．
- [3] 西原陽子，砂山渡，谷内田正彦，“発話テキストからの人間の仲の良さと上下関係の推定”，電子情報通信学会論文誌 D Vol.J91-D，pp.78-88，2008．
- [4] MeCab:Yet Another Part-of-Speech and Morphological Analyzer, <http://mecab.sourceforge.net/>，mecab of 0.98