

国立国語研究所学術情報リポジトリ

Politeness and consideration expression Corpus making and analysis : Towards the application of a dialogue system that considers interpersonal relationships

メタデータ	言語: jpn 出版者: 公開日: 2020-02-06 キーワード (Ja): キーワード (En): 作成者: 太田, 博三, Ota, Hiromitsu メールアドレス: 所属:
URL	https://doi.org/10.15084/00002570

ポライトネス及び配慮表現コーパス作成と分析手法の一考察 ～対人関係を考慮した対話システムの適用に向けて～

太田 博三 (放送大学) †

Politeness and consideration expression Corpus making and analysis ～ Towards the application of a dialogue system that considers interpersonal relationships ～

Hiromitsu Ota (The Open University of Japan)

要旨

昨今、ディープラーニングを中心とした機械学習の進展が見受けられ、従来の画像処理・音声認識・自然言語処理の3分野での進展、とりわけ、機械翻訳での取り組みにおいて、言語学や社会学からポライトネスや配慮表現が取り込まれることで、技術面での質的向上が図られようとしている。従来のQ&Aのような対話応答から、対人関係を配慮した対話応答の取り組みは、必要不可欠である。ここで、機械学習への配慮表現の適用は、教師ありデータとして準備する必要がある、PJ上、時間を要するものである。そこで、本稿では、ポライトネスもしくは、配慮行動や配慮表現を主とした機械翻訳や対話文生成の基となる小規模なデータベース（またはコーパス）を構築し、先行研究の多様な定義や議論を踏まえて場面別の発話ストラテジーの傾向を簡易なベイズ論的アプローチで試みたものである。既存のコーパスと対話システム・機械翻訳との懸け橋になればと考えている。

1. はじめに

国内外でのディープラーニングの進展は、スポーツ分野でのデータ解析への適用やコスモロジーなどの宇宙分野や天文学への専門分野への応用が試みられている段階にあり、応用段階にあると思われる。

一方、自然言語処理における機械翻訳の分野では、語用論の中の一つのポライトネスが必要との認識の下で、機械学習時の学習データ（教師データ）の準備の手間暇がネックとなっている。途中から、新しいタイプのデータを反映させるため、制御の効かないディープラーニングの分野では、小規模のテストが望ましいと考えられる。しかし、ポライトネスや敬語を中心とした待遇表現、または配慮表現の分野では、文化庁の国語審議会等を見る限り、ここ数十年での進展はあまり見受けられない。よって、技術面への適用も、例えば、ポライトネスを機械翻訳に反映させるべきなどの段階で停滞している。

本稿の位置付けは、ポライトネスや待遇表現、配慮表現を技術面に適用することが目的である。厳密な議論はその次の段階にある。

ポライトネスや待遇表現、または配慮表現に関するコーパスを収集し、ユーザの運用を想定した、有効な分析手法を検討したものである。

2. 語用論の登場と技術面でのブレークスルー

2.1 語用論の登場

語用論では、ポライトネスが日本に紹介され、その位置づけや厳密な定義や捉え方において、様々な議論がなされてきた。結果的に次の2つに分かれて収束された。1つはグラ

† 9924658973@campus.ouj.ac.jp

イスの発話論やゴフマンのフェイスの概念，そしてブラウン&レヴィンソン（以下，B&Lと記す）のポライトネス理論に乗っ取った捉え方であり，もう一つは，日本語の敬語と主体としたものである．ポライトネス理論や待遇表現，配慮表現は重複するものであるとの認識は否定的ではない段階でまとまっているように見受けられる．本研究では，生田（1997）の用いた配慮表現を主体的に用いるが，これに限らない．

2.2 語用論の技術面への適用可能性

ポライトネスや待遇表現，配慮表現の捉え方が多岐に議論された中で，機械翻訳や対話自動生成への技術面への適用に限っては，アプリケーションやWebサービスでの提供になり，ユーザーからの評価に左右されるため，比較的，自由度が広いと考えられる．

これまでも，学問的な厳密性でとん挫した場合は，技術面への応用で，一気にブレークスルーすることがある．音声認識や自然言語処理など多岐にわたる分野がマツコロイドなどの対話ロボットに適用され，ユーザーの評価が高まることが正解となる．

2.3 コーパスとコーパスの技術面への適用

国立国語研究所では，国家プロジェクトとして，『梵天』を除き，中立性を担保しつつ，大規模なコーパスが開発されている．一方，コーパスを機械翻訳や対話応答の自動生成に活かすには，正解としての教師データが必要不可欠であるため，csv出力などデータとしての保存方法を統一し，ラベリングすることが簡易にできることが求められている．最終的には，ユーザーの評価，つまり，他人の使い勝手に帰属するため，機械学習にも適用可能なデータのあり方も分析手法への適用から有用なアウトプットも考慮されることになる．

3. ビジネス場面での配慮表現のストラテジーの分析とその妥当性について

コールセンターなどのビジネス面での会話につながるクレーム処理を取り上げる．服部（2009）は図3-1のように，大きく3つの段階に分けて，発話ストラテジーを捉えている．まずは，このように大雑把に分類し，次に，山岡（2017）の2種類の分類を追加することで，ラベルとして，教師データに近づけることが考えられる（図3-2-1と図3-2-2を参照のこと）．

一方，発話ストラテジーの捉え方が適切かについては，ベイズ統計学のベイズ更新を用いて，判断する．4節で，その詳細を検討する．

	内部構造	構成要素	例（実例と筆者作例による）	
終結部	【前段終結】	終結へのメタメッセージ終結同意	「じゃ，よろしくお願ひします」 「はい」「分かりました」	
		【人間関係の再確認】	謝礼及びお互いの関係に言及するもの	謝罪
	感謝			「お電話ありがとうございました」
	責任への言及			「責任は当方にございます」
	友好関係への言及		「これまで，貴社とはよいお付き合いをさせていただいております」	
	今後の行動や対処について言及するもの	再接触の約束	「折り返しすぐお電話します」	
		努力を伝える	「今度も努力します」	
		行為の保証	「必ず解決いたします」	
		行動を表明する	「担当者へすぐ連絡します」	
		儀礼的決まり文句	「よろしくお願ひします」	
【別れの挨拶】	一般的な別れのことば	「失礼します」		
	別れのことば以外	「はい」		

図3-1 終結部の内部構造と構成要素

形式分類
① 副詞 ちょっと、ぜひ、ただ、まったく、たしかに、そこそこ、ぜんぜん、なるほど、そろそろ、なんか
② 副詞句 おかげさまで、悪いけど、すみませんが、恐縮ですが、失礼ですが、よろしかったら、ご多忙のところ、つまらないものですが、僭越ながら、どちらかというと
③ 形容詞・形容詞句 大丈夫、さすがですね、お見事、大変ですね、よかったですね
④ 接尾語・補助動詞 のほう、的な、的には、とか、なんか、～っぽい、～てくれる
⑤ 文末表現 ～でしょう、～かもしれない、～と言えなくもない、～のように思う、～ので(言いさし)、～ていただけるとありがたい、～てくれてもいいのに、～させていただきます
⑥ 慣用語 ご笑納ください、ご高名は伺っております、おつかれさまです

図 3-2-1 形式分類 山岡(2017)

I 利益表現
① 自利大 おかげさまで、ぜひ、～れば幸いです
② 他利小 つまらないものですが、何もありませんが、ご笑納ください
II 負担表現
① 他負大 お忙しいところ、わざわざ、お足元の悪いところ、お手数ですが、ご面倒ですが、可能なら、よろしかったら
② 自負小 ぜんぜん、大丈夫、喜んで、ついでに
III 緩和表現
① 侵害抑制 かもしれない、どちらかと言えば、言えなくもない
② 不一致回避 かもしれない、のほう、的には、とか
IV 賞賛表現
さすが、すごい、お見事、恰幅がいい、健康的、恐れ多くも
V 謙遜表現
① 自賛抑制 まだまだ、そこそこ、一応、自慢じゃないけど
② 自己非難 僭越ながら、若輩者、不束者、出来の悪い、高い所から
VI 賛同表現
なるほど、たしかに、まったく、もの、ごもつとも
VII 共感表現
大変ですね、よかったですね、おつかれさま、それな、ですよー

図 3-2-2 機能分類 山岡(2017)

4. ベイズ論的アプローチによる会話ストラテジーの検討

配慮表現に該当する語句や品詞の組み合わせの出現確率を計算し、語句とストラテジーの流れをモデル化し、その傾向性を検討したものである。手順は以下のとおりである。

まず、水準を次の3つに定めて、モデル化する。

- 1) ポジティブ, 2) ニュートラル (ゼロ), 3) ネガティブ

次に、尤度を算出し事前確率を設定する。ここでは、理由不十分の原理から事前確率を等しく設定する。

「ちょっと」とく申し訳ございません>とをベイズで推定				
(1)「ちょっと」とく申し訳ございません>のモデル化				
		ポジティブ(H ₁)	ゼロ(H ₂)	ネガティブ(H ₃)
憎因子の数		3	3	3
愛因子の数		4	2	3
(2)尤度の算出				
データ(D)		P(D H ₁)	P(D H ₂)	P(D H ₃)
悪印象(憎因子選択)		0.429	0.600	0.500
良印象(愛因子選択)		0.571	0.400	0.500
(3)事前確率の設定				
		ポジティブ(H ₁)	ゼロ(H ₂)	ネガティブ(H ₃)
最初の事前確率		0.3	0.4	0.3
(4)データ入力と事後確率の算出				
デート回数	データ(D)	ポジティブ(H ₁)	ゼロ(H ₂)	ネガティブ(H ₃)
		0.300	0.400	0.300
1回	良	0.414	0.276	0.310
2回	悪	0.414	0.276	0.310
3回	良	0.527	0.176	0.297
4回	悪	0.527	0.176	0.297
5回	良	0.630	0.105	0.266
6回				
7回				
8回				
9回				
10回				

図 4-1-1 ベイズ更新の結果

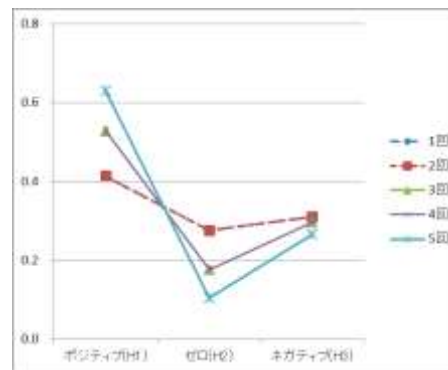


図 4-1-2 ベイズ更新の推移

表 4-2 クレーム (JB01) スクリプト

表 4. 1 JB01	
01 R:	えと私がお持ちできるからちょっと他の担当者(あの:になるか)が分からないんですが(。)あの:まず商品の在庫のほうだけ私のほうから御連絡入れさせていただきますので: =
02 C:	=はい
03 R:	え(。)え(。)であの:また別にちょっと担当者のほうからはC様へお電話入れさせていただくようにはしますので=大変申し訳ございません。ご迷惑おかけいたしました=
04 C:	=いえ。あ(。)あので(。)できるだけ早くお願いします=
05 R:	=じゃああの:在庫のほうの確認だけして先にお電話させていただきます。よろしいですか。私JB01と申しますので
06 C:	[あ(。)はい。はいR様で
07 R:	何かありましたらまた御連絡くださいませ
08 C:	あ(。)はい。よろしくお願ひいたします。
09 R:	[申し訳ございません。よろしくお願ひいたします。はい[失礼いたします
10 C:	[はい=では失礼いたします。

※ 記号Rは受け手(JB)、Cはかけ手

5. まとめ

配慮表現における会話ストラテジーにおいて計量的分析に基づき、その傾向を捉え、再度、機械学習用のコーパスに入れなおすなどで、教師データを整備してゆきたい。

文献

- 生田少子(1997). 「ポライトネスの理論」大修館書店, 『月刊言語』, pp.66-71.
- 彭 飛(2004). 『日本語の「配慮表現」に関する研究—中国語との比較研究における諸問題』 和泉書院
- 山岡政紀(2017). 「日本語配慮表現の分類と語彙リストについて」『日本語コミュニケーション研究論集』第7号, 日本語コミュニケーション研究会, 3-11
- 服部明子(2009)「電話会話における日本人ビジネス関係者のクレームへの対応」『言葉と文化』(10), 77-93, 2009-03
- 久野暲 (1978)『談話の文法』大修館書店
- 堂下・白井・溝口・新美・田中(1998)『音声による人間と機械の対話』「対話過程の基本的特徴と対話における省略の処理(第2章)」オーム社
- ベネロピ・ブラウン・スティーヴン・C・レヴィンソン(1987)「ポライトネス 言語使用における、ある普遍現象」“Politeness: Some Universals in Language Usage”
- 瀧田・西島編著 (2019). 「機械翻訳と未来社会—言葉の壁はなくなるのか」社会評論社, pp.53-104.
- 松原望(2008). 「入門ベイズ統計—意思決定の理論と発展」東京書籍

関連 URL

『国語研日本語ウェブコーパス』検索系『梵天』 <http://bonten.ninjal.ac.jp/>