

Ore 法太

深沢 朋弘 三浦 孝夫

法政大学工学部情報電気電子工学科

1. 前書き

日本語会話では、“僕”と“俺”の使い分けはスピーチシフト問題の代表として知られ、形式的なスタイルと日常的なスタイルの使い分けを暗黙のうちに行うとされる[4]。

本研究では会話コーパスを定量的に解析し、シフト状況を誘引する状況を特徴付ける。特に「あいづち」がスピーチシフトの特徴的な表象となることを利用し[1]、これらの結びつきについて論じる。本研究では、試作チャットシステム「Ore 法太」を用いて、その効果を検証する。

2. 背景

近年、計算機の高度な知的処理を行うには自然言語処理の利用が必須である。計算機と人間の対話において、両者の話し方の丁寧さは会話にとって非常に重要である。本来、人と人との対話の状態は動的であり、常にお互いがある会話のレベルを一定に保つ働きをすることによって会話の流れを円滑にする。このような動的な関係を一定に保とうとする働きをスピーチレベルと呼ぶ[1]。

このスピーチレベルのシフトは「あいづち」でも表象する。「あいづち」にもカジュアルスタイル(CS)、フォーマルスタイル(FS)がある。「エエ」や「ハイ」は FS の、「ウン」や「ソウナンダ」は CS の例である。この対照は例で示される。

カナリアが好きだ。(CS)

カナリアが好きです。(FS)

スピーチレベルのシフトを決定するには、スタイルが受取人の発話によって後の発話のスピーチレベルが変化することを用いる。スピーチレベルのシフトの例を示す。

法太①：君は鳥が好きなんだね。(CS)

入力：エエ。(FS)

法太②：僕は鳥が好きなんです。(FS)

上述の法太①ように入力の語尾が「ダ体」なので入力文は CS であるが、入力側の発話では、FS の「あいづち」を返している。これによって法太②ではスピーチレベルがシフトしている。

本研究では「あいづち」と同様に、“俺”や“僕”などの 1 人称代名詞を使用したときの特徴(参考の動機付け、情緒、フレームセット)によってこのスピーチレベルがシフトすると考える[4]。この特徴の例を以下に示す。

1. 僕はカナリアが好きなんだ。

2. 僕もカナリア。

3. 僕カナリアなんだ。

1 のような“僕”を発話しないと主語がわからなくなるような文章のことを参考の動機付けと呼ぶ。2 のような感情的な発話を情緒と呼ぶ。3 のような正しい日本語ではないものをフレームセットと呼ぶ。本研究では、“僕”と“俺”の代名詞とその特徴を解析することにより、スピーチレベルのシフトを推定する。以下に例を述べる。

1. 入力：僕はカナリアが好きです。(FS)

法太：そうなんだ。君はカナリアが好きなんだ。

俺は犬が好きなんだ。(CS)

2. 入力：カナリアが好きなんだ、俺。(CS)

法太：そうなんですか。あなたは鳥が好きなん

ですね。僕もなんです。(FS)

ここでは“俺”、“僕”を使った会話でのスピーチレベルシフトは確率的に生じるとして扱う。

対話の受取人として使う「あいづち」にもスピーチレベルのシフトの役割を有する。この状況は次の例で示される。

A: 僕の好きなものはね。(CS)

B: エエエエ。(FS)

A: カナリアなんです。(FS)

本研究では、「あいづち」によってスピーチシフトが生じるとし発生を確率的に制御する。

3. スピーチシフト判定

本実験はコーパスを定量的に解析しその生起確率を基に現象シフトを推定する。本研究では日本語話し言葉コーパスを用い、茶筌により形態素解析したあと、前述スピーチレベルのスタイルを決定する。このために各文章の語尾を抜き出し、「デス・マス体」が生じれば FS とし、「ダ体」が生じれば CS と見る。同様に前章で指摘した代名詞とその特徴を解析して、スピーチレベルのシフトを推定する。

4. 実験

筆者らは自動対話システム法太を提案し[2, 3]、自然言語インターフェイスを介してデータベースにアクセスして自身の知識を獲得する学習機構を有する方式を論じている。法太システムでは対話型に学習し、辞書を拡張しながらその対話方式自体も改善させる。図 1 は法太の辞書である。ここで、「犬が好き」という表現に対し利用者は「犬」の親概念の「動物」も好きと推定し、その別の子概念である猫も好きかと問いかけるようになっている。法太では各語に喜怒哀楽のパラメー

タをつけ、会話時における会話感情をこの値の合計から決定する。下記の図2は“楽”という表情であり、図3は“喜”を表す。

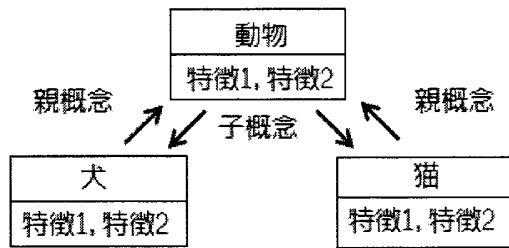


図1. 法太の辞書システム

図2はベースラインの法太の利用状況を示し、図3は本研究で提案するアプローチより拡張された結果を示している。

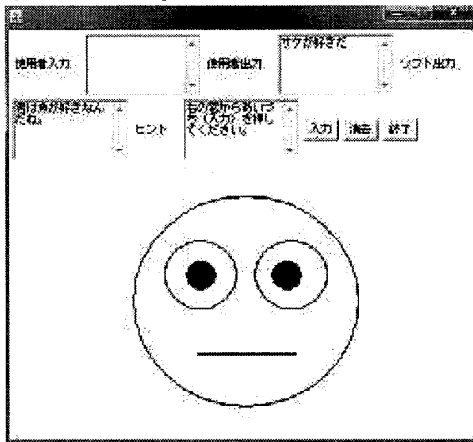


図2. ベースライン法太

5. 考察

表1は、日本語コーパスを形態素解析し、参考の動機付けであるCSの“俺”を用いた発話をすることによってCSに遷移した状況を示す。本研究では、「デス・マス体」、「ダ体」だけでなく、“僕”と“俺”や代名詞の特徴を同時に解析する。

単位(%)	デス体	マス体	ダ体
CS	71.0	61.1	77.8
FS	28.8	37.8	22.2

表1: ダ・デスマス体のスピーチシフト

表1は入力文の語尾が「デスマス体」、「ダ体」のときに“法太”が遷移する状態の実験結果である。これ表よりマス体以外はCSに遷移する割合が高いことを述べていることがわかる。表2はあいづちによるスピーチレベルのシフトを示し、スピーチレベルが「あいづち」では遷移可能性を示す。例えば、FSの“法太”の発話に対して「あーそうなんだ」という「あいづち」では、次の“法太”の発話ではCSモードになる。表3では“僕”、“俺”を用いたレベルシフトの遷移の確率を表す。代名詞の特徴であるフレームセットの“僕”がFSに使われたときにCSに遷移すると言ったような結果を得た。これより代名詞の特徴によって遷移しな

いスタイルも存在する。

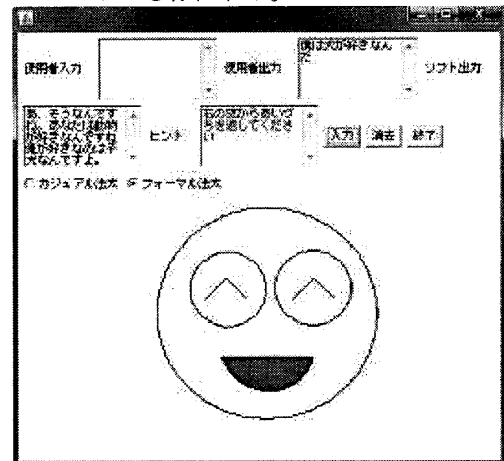


図3. 本実験の法太

単位(%)	法太(CS)		法太(FS)	
	CS	FS	CS	FS
遷移後				
うーん	34.1	65.9	67.3	32.7
ええ	29.0	71.0	64.3	35.7
あーそうなんだ	30.0	70.0	100.0	0.0

表2: あいづちによるスピーチシフト

	参考の動機付け			
	俺(CS)	僕(CS)	俺(FS)	僕(FS)
CS(%)	57.9	66.7	50.0	50.0
FS(%)	42.1	33.3	50.0	50.0
情緒				
	俺(CS)	僕(CS)	俺(FS)	僕(FS)
CS(%)	0.0	50.0	0.0	0.0
FS(%)	100.0	50.0	0.0	0.0
フレームセット				
	俺(CS)	僕(CS)	俺(FS)	僕(FS)
CS(%)	100.0	100.0	0.0	0.0
FS(%)	0.0	0.0	100.0	100.0

表3. “僕”、“俺”を用いたレベルシフトの遷移

6. 結論

本研究では、対話において「ダ体」・「デスマス体」によってスタイルを決定し、「あいづち」によってスタイルシフトする方式の特性を示し、これを試作システム法太上を実現し、スピーチレベルが安定して動作することを確認した。

[参考文献]

- [1] 内藤真理子: あいづちのスピーチレベルとそのシフトについて—日本語母語話者と韓国人学習者の相違—, 世界の日本語教育13, 2003年
- [2] 小林柱輔, 三浦孝夫: 質疑応答からの状況獲得, 情報処理学会第68回全国大会 5N-1, 2006
- [3] 平伸也, 三浦孝夫, 塩谷勇: 抽象化を用いた会話機能, 情報処理学会第66回全国大会 3T-2, 2004
- [4] Ono, T. and Thompson, S.: Japanese (w)atashi/ore/boku: They're not just pronouns, Cognitive Linguistics 14:4 (2003), 321-347