

物語テキストにおける会話文の意図の話者属性による特徴

村井 源^{1,a)} 松本 斉子

受付日 2019年4月26日, 採録日 2019年11月7日

概要: BCCWJ 所収の物語テキスト中の発話文に対して, 発話者の意図に基づくタグ付けを行いコーパスを構築した. 構築されたコーパスを発話者の年齢, 性別, 発話者と相手の関係で分析したところ, それぞれ異なった特徴が得られた. また他者の発話への応答のパターンにも話者の属性による違いが表れた. これらの特徴に基づいて属性や状況に合わせた発話の自動生成を行うことができれば, より自然な発話の生成が可能になると考えられる.

キーワード: 物語論, 会話文, 意図, 語用論

Characteristics of Intentions within Conversational Sentences Based on Speakers' Attributes in Story Texts

HAJIME MURAI^{1,a)} NAOKO MATSUMOTO

Received: April 26, 2019, Accepted: November 7, 2019

Abstract: A corpus with speaker's intention tags was developed based on conversational sentences within texts of novels in BCCWJ. By analyzing the corpus from the viewpoints of age, gender, and relationships between the speaker and the hearer, different characteristics have extracted for each category. Moreover, the characteristics in response patterns for other person's utterances were dependent on speaker's attributes. If automatic appropriate utterance generation in each situation and attributes will be enabled based on those characteristics, the more natural utterance would be generated.

Keywords: narratology, conversational sentence, intention, pragmatics

1. はじめに

近年自然言語処理技術や音声認識技術などを用いた機械的な会話の理解や生成を行うシステムの構築が進められてきている. これらのシステムの多くは言語学や社会学における会話に関する種々の研究の結果を応用して用いており, 発話へのキャラクター性の設定 [1] やユーザの情報の自動的な抽出と活用 [2] などにまで発展している.

一方で同じ会話文の自動生成であっても小説などの物語テキストにおけるキャラクター間の発話の自動生成では, テンプレートをを用いた組合せや書き換えなどの初歩的な生成が行われている段階 [3], [4], [5] であり, 物語の場面に適切

な自然な会話を柔軟に生成することは実現できていない. 会話システムで用いられているような発話生成を因果的關係や確率論を用いて連続的に組み合わせ物語を生成する試みも行われているが [6] この場合には, 物語における起承転結やオチなどのマクロ的な構造を導入することが困難であり, 生成されたテキストは物語というよりもキャラクターの単なる行動記録のようなものになってしまう.

物語文における会話を生成するうえでは, 単発の発話としての自然さだけではなく, いくつかの発話を並べた場合の連続的な自然さも必要となる. またいくつかの発話の連続として構成される会話文が, 物語のプロット上における各場面の機能 [7] (たとえば, 冒険物語における最初の「依頼」の機能や, 人間関係の変化を描く物語における「対立」や「和解」などの機能は会話文で構成可能である) を担う必要が生じる. このため, 物語文では会話を構成する各発話の機能や複数の発話によって構成される関係性の構造や

¹ 公立はこだて未来大学
Future University Hakodate, Hakodate, Hokkaido 041-8655,
Japan

^{a)} h_murai@fun.ac.jp

マクロ的な機能を分析する必要があると考えられる。

また、一般的な対話システムと大きく異なり、物語テキストでは質問や依頼に対し説明や承諾を行うだけでなく、時として言い訳や拒絶やはぐらかしなどの応答の生成も必要となる。さらに、非難や罵倒のような機械的対話システムが避けるべき発話の生成も可能でなければならない。

さらに物語テキストにおいては自然なだけではなく、何らかの意味での読者に続きを読ませるためのテクニックも要求される。たとえばキャラクターに対するイメージを惹起させかつ読み手へのインパクトを引き起こすために、実際の発話以上に誇張された「らしさ」を導入した文体の利用なども頻繁に用いられるテクニックである [8]。そのため自然な会話文の場合と異なり、各キャラクターの発話は作者が思い描く登場人物の属性のステレオタイプ的な特徴を含有すると考えられる。ただしそのステレオタイプ的な特徴は、多くの読者にとって理解可能でなければならない。そうでなければ物語が社会的に受容されないためである。出版社での編集と書店での販売、図書館での選別などを経て、社会的に受容された物語が含む会話文は、作者のステレオタイプに基づきつつも、多くの一般的な読者の了解可能性に沿う形になっていると考えられる。

そのため、より自然かつ了解可能なキャラクターらしさをとまなう物語的な会話自動生成を行うには、従来の日常会話文の分析だけでなく物語テキスト中の会話分析と特徴抽出が必要不可欠と考えられる。また言語には語彙・文法や社会・文化的な差異などの様々な要因が影響し、会話文の特徴も日本語と他の言語では様々な観点で異なると予想される。そこで日本語での会話自動生成の基礎データとなる、自然な日本語での物語の会話文の特徴を抽出するために、日本語の物語テキストを対象とした分析を行うこととする。

日本語の物語テキストでの文体的な特徴とキャラクター性に関してはランダムサンプリングされた物語テキストに基づく分析 [9] とその日常会話との比較 [10] がすでに行われており、物語テキストにのみ見られる文体的特徴が明らかになっている。そこで本研究では日本語の物語テキストにおける会話文の主たる機能である語用論的な意図とその連続的な出現のパターンに焦点を絞って計量的な手法での特徴抽出を行う。特徴の抽出においては、多くの先行研究で扱われている性別と年齢による相違と発話における相手との関係による変化がどのように影響するかを分析する。

物語テキストにおける会話文の特徴が明らかになれば、それを基礎データとしてより自然な会話文の自動生成が可能になると期待される。またその結果は物語テキストの自動生成だけでなく、キャラクター性を付与した対話システムの実現や、複数対話システム間での自然な会話文の生成などにも活用可能であり、種々の情報処理研究の基盤的なデータとして活用可能であると考えられる。

2. 会話の特徴と分析

会話文は日常的な言語利用における基本的な形式の1つであり、言語学や社会学、哲学などの様々な領域にまたがって古くから会話を対象とした研究が行われてきている。会話分析 [11] と呼ばれる分野では対話の連鎖における話者の交代の問題や [12]、会話中で対をなす2つの発話 [13] の問題などが熱心に研究されてきた。また語用論の中で言語行為論 [14] と呼ばれる領域では、会話文における表面的な文字列としての言葉の意味とそれが間接的にもたらす機能(遂行動詞などで表される発話内行為)の齟齬に着目し、機能の分類や整理 [15] などが行われている。これらの直接・間接的な会話文の機能は発話者の意図とも呼ばれる。

一方で、前述のとおり近年情報処理技術の進展により、これらの会話文に関する研究結果などに基づいて、機械的な対話の理解や生成を行う対話システムの構築が進められてきている。これらのシステムが依拠するのは多くの場合実際の日常的な会話文を電子データ化したコーパスであるが、単に会話文を電子化するだけではなく種々の意味的なタグを付与して特徴を分析する試みも行われてきている。たとえば、語用論的な機能に相当するタグを付与してコミュニケーションにおける発話の機能を調査した研究 [16], [17] や、感性的な情報を付与した研究 [18] などがある。これらのタグ付きコーパスの多くは人手によるが、まだ精度は十分に高いとはいえないながらも自動化の試みも行われている [19]。このような各種のタグ付きコーパスを用いることで、雑談での話題展開の構造的パターン [20] や、年齢性別による話題展開の差違 [21] などの抽出も可能である。

3. 分析対象と手法

3.1 分析に用いるコーパス

一般的な物語テキストにおける会話文の特徴を分析するため、『現代日本語書き言葉均衡コーパス』 [22] (The Balanced Corpus of Contemporary Written Japanese, 以下 BCCWJ) をデータとして用いた。BCCWJ には販売された書籍の ISBN に基づくサンプリングで作られたコーパスと、図書館に所蔵された書籍のリストに基づくサンプリングで作られたコーパスが含まれているが、実際に流通している物語テキストにおける会話文の実態をとらえることを目的とするため、図書館の蔵書に基づくコーパスを用いた。また、日本語文学テキストを対象とするため十進分類で日本語の物語・小説自体にあたる 913 に分類された可変長のサンプルのテキスト(文章構成的なまとまりで区切られた部分を抽出したテキスト)のみを選択した。結果として条件に該当する 2189 サンプルが抽出されたが、そこからランダムに 100 サンプルを抽出し人手でのタグ付けの対象とした。

表 1 作成した意図のカテゴリ
Table 1 Created categories of intentions.

導入	慣習的な挨拶, 相手への呼びかけ, 人の紹介, 会話の導入的な発話など. 例:「今晚は」
陳述	普遍的事実や現在の状況・過去の出来事の単なる叙述, 相手への説明, 報告や伝聞的な情報伝達など. 例:「コンニャクは胃ではどうにもならないというわけで, 早く送り出されてしまうのです」
思考	現在や過去の状況に対する推測や自問自答, 仮定や仮説を用いた思考の言語化や, 理由や言い訳の提示, 相手に返答を求めない疑問や疑いの表明, 自説の主張などの思考に関する発話を含める, ただし将来の状況への推測は「将来」に分類. 例:「ただ, 正造だけは犯行当日まで, 何も知らされてなかったんじゃないか」
意思	将来行う予定の自身の行為をあらかじめ伝達する発話や, 自身の決意や義務の表明, 相手との約束などを含む. 例:「彼と相談してみるわ」
願望	発話者の願望や期待に関する発話, ただし何らかの行動を行う意思の表明を含む場合には「意思」に分類する, 神仏への祈りや魔法等での呪文も「願望」に含める. 例:「連合艦隊の勇戦に期待しておりますぞ」
依頼	発話者が相手に何かを頼む場合, 丁寧な依頼や懇願だけでなく, 高圧的な命令や脅し, 許可や禁止, 相手が果たすべき義務に関する示唆もこの中に含める. 例:「助けてくれえ!」
提案	発話者と相手がかともに行動することを提案する場合や, 相手のために相手の行動を促す発話, 発話者からの相手のための行動を申し出る場合などを含める, 発話者が自身の将来の行動を決定している場合には「意思」に分類する. 例:「正月はパリで過そう」
質問	相手に対する問いかけや質問, 詰問, 物事に対する疑問など. 例:「そんなに凄い匂い?」
応答	相手の発話に対する相槌や, 肯定否定の明確ではない応答, 相手の発話に対する聞き直しや言い直しなどの確認的な行為を含める. 例:「へえ」
了解	相手の依頼や提案に対する了解や受諾, 請け合いなどの肯定的な反応, 相手の発話に対する同意や納得の表明, 相手への屈服や服従の表明などを含める. 例:「仰せのとおり」
拒否	相手の依頼や提案に対する拒否や拒絶, 相手の発話内容に対する否定などの否定的な応答を含める. 例:「だめだってば」
将来	発話者の行動の意思を明示的には伴わない将来に対する予定や予測に関する発話を分類する, 発話者の行動の意思を明示的に含む将来の予定は「意思」に分類する. 例:「これ, 絶対, 売れるわよ」
評価	自己や他者, 事物や状況に対する評価を行う発話を肯定的・否定的両方とも分類する, ただし「自責」「他責」に分類可能な評価は「自責」「他責」に優先的に分類する, また相手に配慮するためのお世辞や謙遜と解釈可能な場合は「配慮」に分類する. 例:「これ, いいわね」
自責	自己の非を認めることや, 謝罪や反省, 犯罪の自供などに関する発話を分類する. 例:「…情けねー…」
他責	他者の非を責める非難・批判的な発話や, 相手の欠点や問題点を指摘する発話, 相手への皮肉や侮辱や悪口, 相手を挑発したりからかったり茶化したりするような発話を分類する. 例:「こら, おまえのせいだぞ」
感謝	相手への感謝や謝意を表す発話を分類する. 例:「助かるわ」
感情	発話者自身の感情状態や身体的感覚等に関する表明, 笑い声や驚きの感嘆詞など発話者の内的状態の表明に関する発話を分類する, ただし「感謝」「自責」「他責」に分類可能な場合はそちらを優先する. 例:「アハハハハハ…アツハツハツハツ」
配慮	相手の感情状態を改善させるための労わりや慰め, かばい, なだめ, 激励, はげまし, もてなし, お世辞や謙遜などの機能を持つ発話を分類する. 例:「風間さん, そう力を落とさずに」
冗談	冗談やとぼけ, おどけ, はぐらかし, などを分類する, ただし相手を茶化したりからかったりする冗談は「他責」に分類する. 例:「火星人も見たいし…」

本研究では物語中に会話として明示的に表れる部分のみに限定した分析を行う. タグ付けの対象はテキスト中の時間軸で現在発話されている「」や『』でくくられた会話箇所とした. 回想, 置手紙, 語り手や登場人物による想像上の会話などその場で行われていない発話的な描写は除外するが, 電話の会話やファンタジーにおけるテレパシーで

の会話は分析対象に含めた. なお, 物語と日常いずれでも会話ではボディランゲージ (たとえば「うなづく」「首を横に振る」など) や沈黙などで相手と非言語的なコミュニケーションを行うことがある. また物語文には地の文として記述された会話的要素 (たとえば, 「A は承知の返事をした。」など) も出現する. しかしそれらの非会話的な

コミュニケーション要素を含める場合には、どこまでをコミュニケーションと見なすか解釈の幅があり、客観的な基準の設定が困難である。また、本研究の主目的は物語らしい会話の自動生成に向けた基礎データの取得であるため、会話文の形で記述されていないこれらの箇所は分析の対象外とした。

結果として、「」や『』などの区切り記号を単位とすると日本語の物語文 100 サンプル中の 5632 発話が抽出された。ただし、小説においては発話の途中で説明や情景描写の地の文を挟むなどの描写法が多用されるため「」の数が発話の数とは限らない。なお、BCCWJ の各テキストは小説の一場面を切り出したものであるため、場面転換による発話の切断は基本的に含んでいない。そこで、連続する同じ人物による「」単位での発話は 1 つの発話であると見なすこととして分析を行った。この場合全発話数は 4,548 となる。

これらの発話に対し、発話者とその性別、年齢（若年層、成年層、老年層の 3 種類に人手で分類）に関しては国立国語研究所で作成された BCCWJ Speaker Information Corpus [23] を用いてタグを付与し、それに加え発話者から見た発話の相手の関係性（友人、家族、上司など）もタグとして付与した [24]。なお BCCWJ Speaker Information Corpus の話者属性付与は、会話文内の情報に合わせて地の文や、サンプル外のテキスト中の情報も含めて行われている。また BCCWJ Speaker Information Corpus には対話を構成しない発話（内言など）も含まれているが本研究では対話構造の分析を目的とするため、内言などは省略し、より単純な構成としている。

3.2 分析手法

登場人物の発話の文体的傾向は先行研究で調査が行われており [9], [10], 本研究は発話の持つ語用論的な機能である意図に焦点を絞り、先行研究 [16], [17] を参考に 19 のカテゴリを作成し人手での付与を行った。意図付与の単位は句点で区切られる文とした。意図数（文数）は全体で 9277 となった。作成したカテゴリとその定義を表 1 に示す。

本研究での分類の妥当性を確認するため第 2 分析者による分類の一致度の検定を行った。一致度の検定においてはランダムに発話中の 200 文を抽出し、第 2 分析者にカテゴリの第 1 候補と第 2 候補を選択させ第 1 分析者との一致度を κ 係数によって計算した [25]。 κ 係数は本研究のような対象の分類分析においてその結果の科学的な再現性を数値的に評価するための手法として心理学や社会学などで広く用いられている指標である。 κ 係数は下記の式で計算する。 $P(A)$ は第 1 分析者と第 2 分析者の一致率、 $P(E)$ は期待される一致率（ランダムな場合でも一致しうる確率）である。

$$\text{Kappa} = \frac{P(A) - P(E)}{1 - P(E)}$$

計算された一致率、期待される一致率、 κ 係数を表 2 に

表 2 カテゴリの妥当性検証結果

Table 2 Results of validation experiment for categories.

	第一候補のみ	第二候補も含む
一致率	0.70	0.89
期待される一致率	0.16	0.17
κ 係数	0.64	0.86

表 3 発話者の性別・年齢の属性と発話数・意図数

Table 3 Number of utterances/intentions according to gender/age of speakers.

		人数	発話数	意図数	発話当たり平均意図数
男性	若年層	29	304	516	1.7
	成年層	258	2426	4978	2.1
	老年層	4	16	56	3.5
女性	若年層	29	283	536	1.9
	成年層	96	976	2234	2.3
	老年層	2	2	7	3.5
不明・その他		124	541	950	1.8
合計		542	4548	9277	2.0

表 4 発話者からみた相手の関係と発話数・意図数

Table 4 Number of utterances/intentions according to relationships between speakers and listeners.

	人数	発話数	意図数	発話当たり平均意図数
友人	40	544	1017	1.9
同僚	25	240	513	2.1
先輩・上司	21	216	487	2.3
家族	28	224	477	2.1
恋人	12	190	408	2.1
後輩・部下	11	214	391	1.8
不明・その他	405	2920	5984	2.0
合計	542	4548	9277	2.0

示す。 κ 係数が 0.61-0.80 の場合は本質的 (substantial) な妥当性であるといわれており、また κ 係数が 0.80-1.0 の場合にはほぼ完全な一致 (almost perfect agreement) であるといわれている [25]。このため、本研究のカテゴリ分類には十分な妥当性があると考えられる。なお、全体的には高い一致度であったが、「他責」「拒否」に関しては分析者間の解釈の相違が複数みられた。ネガティブな意図では隠蔽や婉曲表現が多用されることが原因と考えられる。本研究では第 1 分析者の分類結果に基づいて以降の分析を行った。

コーパスに付与された発話単位の発話者本人の性別と年齢層ごとに発話者の人数と、意図数（文数）、発話あたりの平均意図数を表 3 に示す。また発話を発話者からみた相手の関係によって同様に分類した場合の結果を表 4 に示す。

表 5 発話者の属性に基づく意図の集計と χ^2 乗検定結果
 Table 5 Results of chi square test and number of intentions according to attributes of speakers.

	男性		女性	
	若年層	成年層	若年層	成年層
陳述	112 ▽▽	1583 ▲▲	132 ▽▽	637
質問	98 ▲▲	636	83 ▲	244 ▽▽
思考	55	531	35 ▽▽	290 ▲▲
依頼	40	360	43	153
応答	25	260	32	128
評価	23	225	20	124 ▲
他責	25	227	35 ▲	87
了解	14 ▽	270 ▲▲	27	65 ▽▽
導入	21	208	37 ▲▲	64 ▽▽
意思	16	154	11	80
提案	18	125 ▽▽	6 ▽▽	96 ▲▲
拒否	18	117 ▽▽	20	69
感情	21 ▲▲	81 ▽▽	26 ▲▲	60
配慮	7	44	7	32
自責	7	34	6	30
将来	4	49	8	15
願望	6	34	4	21
感謝	4	23	4	29
冗談	2	17	0	10

4. 分析結果

4.1 意図の頻度と属性による差異

発話者の属性である性別と年齢層によってカテゴリに分類した発話の意図数を集計した結果を表 5 に示す。表 5 では該当発話数の多い男女の若年層・成年層のみを示している。また、頻度の少ないカテゴリ（「配慮」、「自責」、「将来」、「願望」、「感謝」、「冗談」）を除いて χ^2 乗検定の残差分析を行い、5%有意水準で多い箇所には▲を、1%有意水準で多い箇所には▲▲を、5%有意水準で少ない箇所には▽を、1%有意水準で多い箇所には▽▽をそれぞれ右側に記している。なお、除外されたカテゴリ以外は χ^2 乗検定における各セルの期待値は 10 以上である。 χ^2 乗値は 193.6、有意水準は 1%、Cramer's V は 0.091 となった。

表 5 では、男性と女性の若年層の共通傾向として、「陳述」が少なく「感情」が多いという結果がみられる。一方男性と女性の成年層には共通の傾向がみられず、かえって「了解」や「提案」では正反対の傾向を示している。また男性の若年層と成年層（「陳述」、「感情」）、女性の若年層と成年層（「思考」、「導入」、「提案」）も反対の傾向は示すが共通の傾向は含まれない。これより、物語テキストでの登場人物の発話意図の傾向は性別や年齢単体に依拠しているわけではないと分かる。これらの結果より若者は感情的、成年男性は非感情的で、成年女性は熟考して提案するという

ような、小説の世界におけるステレオタイプ的なイメージが表れていると考えられる。これらは個々の作者のステレオタイプの平均的なものと考えられるが、以降の結果から得られる考察に関しても同様である。

同様に、各発話における発話者から見た相手の関係性のタグに基づいて意図を集計した結果を表 6 に示す。表 6 においても、 χ^2 乗検定における各セルの期待値は 10 以上となっている「陳述」から「感情」までのカテゴリに対して χ^2 乗検定の残差分析を行い、5%と 1%水準での多寡を同様の記号で示している。 χ^2 乗値は 361.8、有意水準は 1%、Cramer's V は 0.151 となった。

発話の相手が友人、家族、恋人であった場合に共通する傾向としては、「陳述」が少なくなるということがある。親しい相手との会話では事実関係の説明のような発話は減ることを示していると考えられる。また、友人への発話では「評価」と「感情」が多い傾向がみられる。いずれも発話者の心的な状態の開示であり、相互にそのような開示を行うことのできる関係が友人的であるという、作者たちの友人関係に対する平均的な認識を示していると考えられる。恋人関係の相手に対しては「依頼」「提案」「拒否」が多いことが表から見て取れる。これらはいずれも場合によっては人間関係を壊す可能性を多分に含んだ機能であり、そのようなリスクをも包含可能なごく親しい間柄であることを作者が示唆していると考えられる。

予想どおりではあるが、先輩・上司と後輩・部下に対する発話は正反対の傾向にある。後輩・部下に対しては、「質問」をして問い質し、自身の「思考」を伝え、また場合によっては非難や叱責の言葉を送る（「他責」）。逆に先輩・上司に対しては事実関係を説明する「陳述」が多く、返答としては「了解」が増える。逆に叱責の言葉や問い質す言葉は出現しにくい。同じ立場の同僚に対しては、発話者の「意思」を伝達する状況が描かれることが多いと分かる。

4.2 発話内構造と対話における応答の構造

発話の連続的なパターンを可視化するため、前後で発話者が変化しない場合の 2 つの連続的な意図から構成される遷移のパターンと、前後で発話者が変化する場合の 2 つの連続的な意図から構成される遷移のパターンを抽出する。まず発話者が変化しない場合の遷移のパターンであるが、Simpson 係数を用いて主要な遷移に絞ってネットワーク状に可視化することを考える。Simpson 係数は A と B をそれぞれ異なる 2 つの意図の出現会話文の数としたとき、A と B が A→B の順で 2 つ連続して含まれる会話文の出現数を分子、A と B いずれかを含む 2 連続の会話文数のうちでより出現数の小さい方を分母として下記の式で表される。

$$\frac{|A \cap B|}{|\min(A \rightarrow B)|}$$

Simpson 係数の最大値は 1 であり、この値が大きいほど

表 6 発話者からみた相手との関係性に基づく意図の集計と χ^2 乗検定結果
 Table 6 Results of chi square test and number of intentions according to relationships between speakers and listeners.

	友人	家族	恋人	先輩・上司	同僚	後輩・部下
陳述	261 ▽	109 ▽▽	94 ▽	230 ▲▲	159 ▲	73 ▽▽
質問	148	74	56	31 ▽▽	62	88 ▲▲
思考	136 ▲	60	28 ▽▽	37 ▽▽	48	68 ▲▲
依頼	49	28	36 ▲▲	16 ▽	25	25
応答	34 ▽▽	34	20	23	25	56 ▲▲
評価	82 ▲▲	17 ▽	33	32	26	10 ▽▽
他責	52	28	14	8 ▽▽	18	28 ▲
了解	30	25	11	28 ▲	23	6 ▽
導入	42	16	9	18	17	13
意思	25	18	14	11	27 ▲▲	7
提案	44	26	30 ▲▲	15	22	6 ▽▽
拒否	34	13	22 ▲▲	17	16	2 ▽▽
感情	38 ▲▲	8	16	7	12	4 ▽
配慮	13	4	5	6	11	0
自責	4	0	5	3	10	2
将来	7	6	1	1	4	0
願望	8	5	6	2	2	2
感謝	1	4	6	1	4	1
冗談	9	2	2	1	2	0

2つの意図の連続的な出現傾向が強いことを示す。類似の指標としてはDice係数やJaccard係数が存在するが[26], テキスト中の要素などの出現頻度の大きく異なる対象間の関係性を計算する場合, Simpson係数を用いると片方が極端に頻出する場合にもその影響を受けにくく, 出現頻度の低い要素からみた場合の関係性の強さを反映しやすい。本研究で扱う対象も上位と下位で極端な頻度の差があるためSimpson係数を用いることとした。また言語においては出現順の違いは意味を持つ場合が多いため, 2つの意図がABの順で出現した場合の頻度とBAの順で出現した場合の頻度とSimpson係数をそれぞれ別に計算した。

全19の意図をノード, Simpson係数上位50位までの意図の連続パターンをエッジとして有効グラフの形で可視化したものが図1である。ただしSimpson係数は頻度が低い場合に大きな値を示しやすい傾向があるため意図が連続するパターンの出現数が3回未満の場合はネットワーク中から除いている。ネットワーク中のノードのフォントサイズは各ノードが示す意図の出現数のルートに対応している(フォントサイズの二乗が出現数に比例)。また図中の数値は各エッジのSimpson係数を示している。このとき, 図1はある発話者が何かの意図を示す発話を行った後に連続的にどのような意図の発話をしやすいかを示している。

図1には「質問」する前に「導入」的な発話を行いやすい, 自分の「意思」を示してから「依頼」を行いやすい, 「拒否」した後は自身の「思考」について説明するなどい

くつかの特徴的遷移がみられるが, 大部分は同じ意図の繰返ししか, 「陳述」へのあるいは「陳述」からの遷移である。1人の発話者の連続的な発話における意図の平均はただだか2程度であるため(表3), 事実の陳述と何らかの意図の組合せ, あるいは同じ意図の繰返しのような単純なパターンが大部分を占めたと考えられる。

次に, 同様の手法で発話者が前後で異なる場合の2つの意図の連続するパターンをネットワーク化した場合を図2に示す。図2においても「陳述」と他の意図の組合せによる遷移が多数みられるが, それ以外にも「質問」に対する反応としての「拒否」, 「了解」や「応答」, また逆に「提案」, 「思考」, 「評価」, 「依頼」などに対する反応としての「質問」のパターンが存在することが分かる。

そこでこのような異なる話者間での発話における反応のパターンの話者属性による差異を検討するため, コーパス中で比較的頻度が高く意図の2連続においてもパターンの出現が期待できる属性の組合せである成人男性と成人女性を選択する(表3)。成人男性と成人女性がそれぞれ直前の他者の発話に反応する箇所のみを抽出して, 同様のネットワークを構築した結果が図3(成人男性が他者の発話に反応する場合のパターン)と図4(成人女性が他者の発話に反応する場合のパターン)である。

図3の成人男性の応答のパターンでは, 「質問」がなされた場合に「拒否」, 「了解」, 「思考」などが後に続いている。また相手の発話の意図が「拒否」, 「思考」, 「提案」や「評

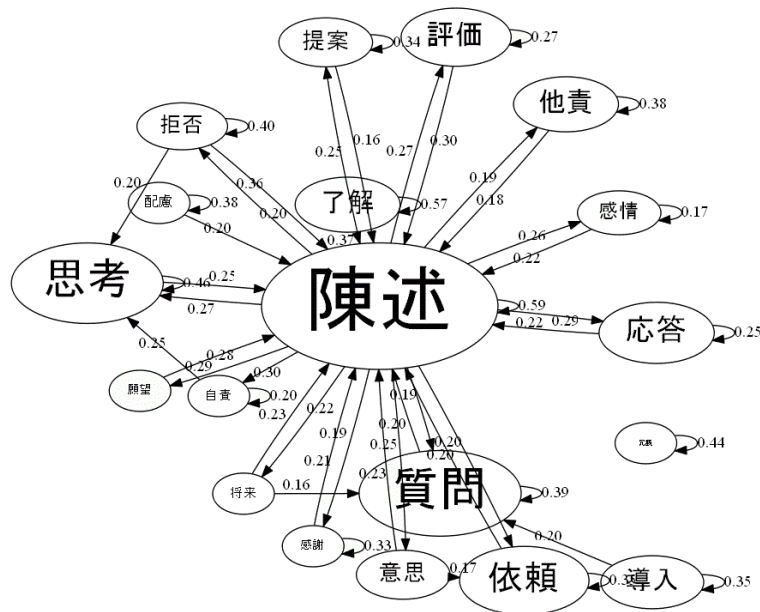


図 1 同じ発話者による連続的な意図の遷移傾向

Fig. 1 Transition pattern of serial utterances by same speakers.

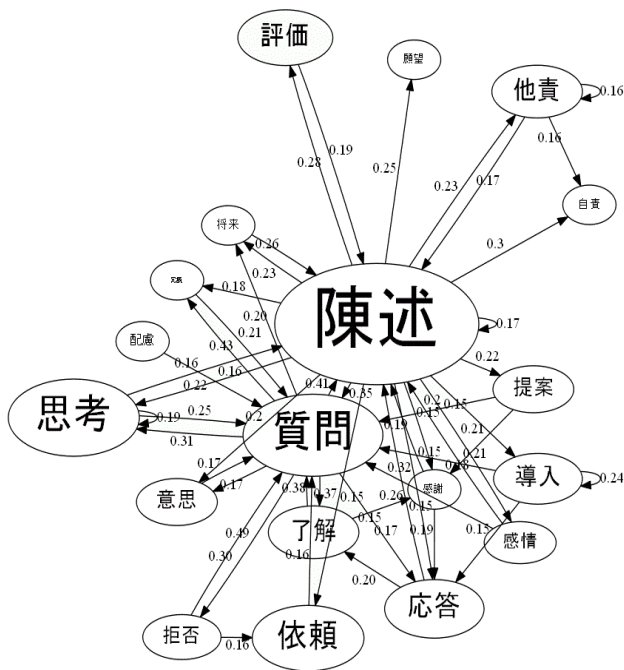


図 2 異なる発話者間の反応の遷移傾向

Fig. 2 Transition pattern of serial utterances between different speakers.

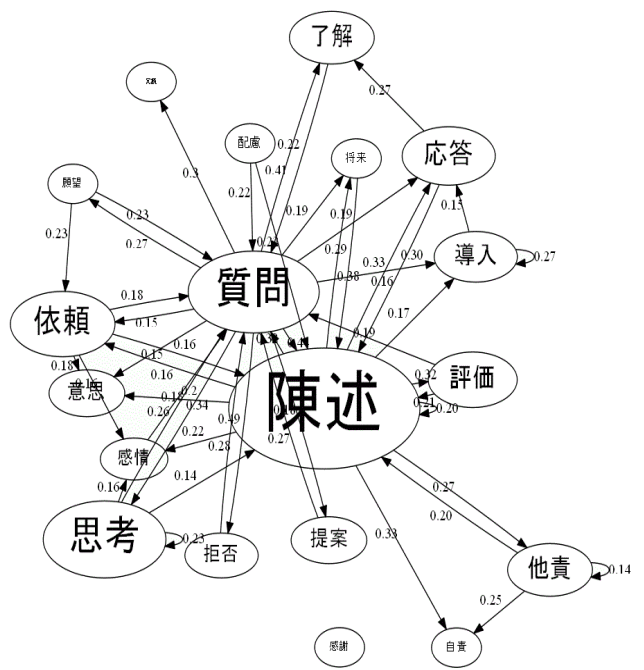


図 3 成人男性話者の異なる発話者への反応の遷移傾向

Fig. 3 Transition pattern of adult male responses for different speakers.

備」であった場合にも「質問」で返すパターンがみられる。
 図 4 の成人女性の応答パターンにおいても、「質問」がなされた場合に「拒否」, 「了解」, 「思考」などが後に続いており、この点に関しては属性による差異はみられない。しかし、相手の発話の意図が「拒否」の場合には「依頼」や「思考」で返すパターンが顕著であり、「拒否」されると「質問」で返す成人男性の反応パターンとは異なっている。また「評価」には「評価」や「感情」の意図で応答し、「提

案」には「感情」や「感謝」の意図で応答していることも読み取れる。この点でも成人男性と成人女性の相手の発話意図への応答のパターンは大きく異なっているといえよう。
 比較すると成人男性が相手に対して総じて「質問」で応答するような状況において、成人女性は「質問」以外の様々な意図を用いて多様な会話のパターンを構成しているといえよう。ただし、相手の質問に対する応答の仕方に関しては成人男性と成人女性では大きな差はみられない。

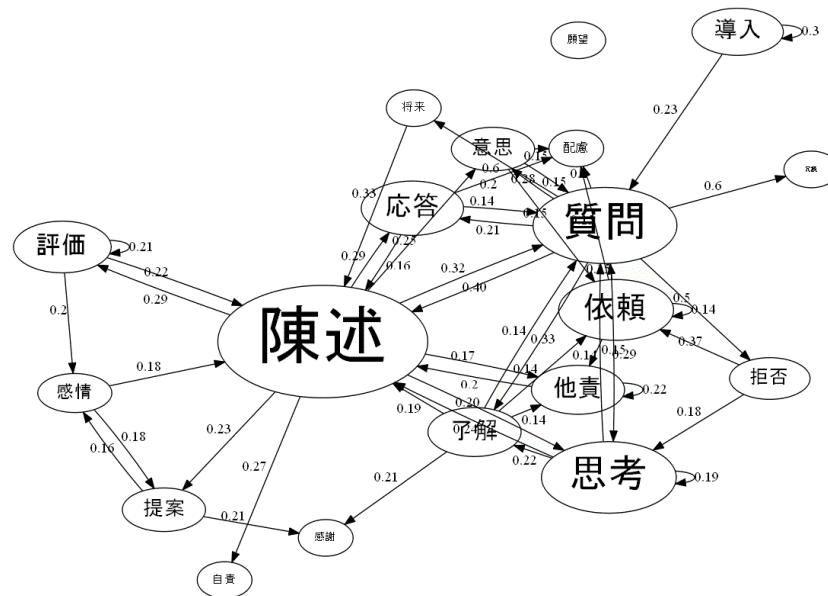


図 4 成人女性話者の異なる発話者への反応の遷移傾向

Fig. 4 Transition pattern of adult female responses for different speakers.

「質問」以外の異なりとしては他にも、成人女性は「配慮」や「感謝」で応答するパターンがあるが、成人男性にはこのような応答パターンがみられない点もあげられよう。

ただしこれらの傾向はあくまで上位 50 位までの相手の意図に対する反応のパターンを抽出した場合の相違であり、より下位の項目までみれば成人男性にも「配慮」や「感謝」を用いて反応するパターンはみられる。男性と女性で会話相手の属性に偏りがある可能性もあるが、本研究の目的は作中のキャラクター「らしい」会話のパターン抽出であるため、特定の相手に特定の意図で話しかけることが物語中で多くみられることをパターンとして扱っている。

5. 結論と今後の課題

本研究では、BCCWJ に収録された物語テキストにおける会話文のランダムサンプリングを対象として、語用論的な機能である意図を 19 種類に分類してタグ付けし、それらの意図の出現傾向が発話者の年齢や性別、発話者と発話の相手の関係性によって異なることを明らかにした。また、他者の発話意図に対して反応する場合の発話意図の傾向のパターンをネットワークを用いて可視化した。さらに頻度の高い成人男性と成人女性の発話において、他者の発話意図に対する反応としての発話意図のパターンを可視化し比較を行った。結果として、物語テキスト中の会話では成人男性は相手への反応として「質問」を利用する形で描かれることが顕著であるのに対し、成人女性は「依頼」「感情」「感謝」など「質問」以外にも多様なパターンを用いる形で描かれる傾向が強いことが明らかになった。特に「感謝」や「配慮」で相手に反応するパターンは成人女性に顕著であった。本研究の成果を自然言語処理技術などに適用することで、キャラクターの「らしさ」を担保した形での会話の

自動生成などが可能になると考えられる。

本研究で得られた分析結果は物語テキストにおける仮定の登場人物間の会話であり、小説の作者が思い描く平均的なキャラクターのステレオタイプの発話傾向といえよう。そのため物語の登場人物的なキャラクターの特徴をとらえられていると考えられるが、より一般的な発話の傾向を理解するためには、今後これらの特徴がどのように実際の日常会話と相違しているか比較検討を行うことも必要であると考えられる。また、意図の分類に関してはネガティブな意図で多用される隠蔽や婉曲表現においても解釈を一定化できるような工夫も必要であるといえよう。

また発話と意図のデータ数を増加することで、発話者と相手との関係性による場合の応答のパターンや発話の展開のパターンも抽出可能になると考えられる。発話のデータが増えれば、属性や関係性の分類の細分化が可能となる。たとえば家族のより詳細な単位への分割も実現できであろう。家族が細分化できれば、上下関係を意識しやすい日本文化において同じ家族内でも親子や兄弟の会話がどのように異なるのかなどのより踏み込んだ相違も情報システムで利用可能データとして抽出できると期待される。

謝辞 本研究は科学研究費基盤研究 C 「物語文と日常会話文における発話文体と話者属性の分析」、国立国語研究所「大規模日常会話コーパスに基づく話し言葉の多角的研究」、日本語日常会話コーパスの構築」の支援を受けた。

参考文献

- [1] 宮崎高明, 平野 徹, 東中竜一郎, 牧野俊朗, 松尾義博, 佐藤理史: 文節機能部の確率的書き換えによる言語表現のキャラクター性変換, 人工知能学会論文誌, Vol.31, No.1, pp.DSF-E-1-9 (2016).
- [2] 小林峻也, 萩原将文: ユーザの嗜好や人間関係を考慮す

- る非タスク指向型対話システム, 人工知能学会論文誌, Vol.31, No.1, pp.DSF-A-1-10 (2016).
- [3] 鈴木諒輔, 佐々木奨之, 袴田 翔, 田中瑞穂, 三浦隆太郎, 城田晃希, 高橋翔太, 南部太雅, 山田康貴, 吉田拓海, 松浦史佳, 松原千里, 寺島啓悟, 津沢慎吾, 渡邊広基, 村井源, 迎山和司, 田柳恵美子, 平田圭二, 角 薫, 松原 仁: 物語と情景描写を自動生成する統合的システムの検討と開発, 情報処理学会研究報告, Vol.2018-EC-50, No.28, pp.1-8, 2018 (2018).
- [4] Sato, S.: A Challenge to the Third Hoshi Shinichi Award, *Proc. INLG 2016 Workshop on Computational Creativity in Natural Language Generation*, pp.31-35 (2016).
- [5] 松山諒平, 佐藤理史, 松崎拓也: 人狼ログからの小説の自動生成, 言語処理学会第 23 回年次大会発表論文集, pp.32-35 (2017).
- [6] The What-If Machine Project: "The What-If Machine," available from http://ccg.doc.gold.ac.uk/research/whim/resources/poster_UC.pdf (accessed 2019-3-22).
- [7] Propp, V.: *Morphology of the Folk Tale*, University of Texas Press (1968).
- [8] 金水 敏: ヴァーチャル日本語役割語の謎, 岩波書店 (2003).
- [9] Murai, H.: Factor Analysis of Utterances in Japanese Fiction-writing Based on BCCWJ Speaker Information Corpus, *Advances in Human-Computer Interaction*, Vol.2018, Article ID 5056268, pp.1-9 (2018).
- [10] Murai, H.: Factor Analysis of Japanese Daily Utterance Styles, *Proc. LREC 2018 Joint Workshop LB-ILR2018 and MMC2018*, pp.26-29 (2018).
- [11] Schegloff, E.A.: Confirming Allusions: Toward an Empirical Account of Action, *American Journal of Sociology*, Vol.102, No.1, pp.161-216 (1996). 西阪 仰 (訳), 会話分析の方法—行為と連鎖の組織, 世界思想社 (2018).
- [12] Sacks, H., Schegloff, E.A. and Jefferson, G.: A Simplest Systematics for Organization of Turn-taking for Conversation, *Language*, Vol.50, pp.696-735 (1974). 西阪 仰 (訳), 会話分析基本論集—順番交替と修復の組織, 世界思想社 (2010).
- [13] Schegloff, E.A.: *Sequence Organization in Interaction: A Primer in Conversation Analysis*, Cambridge University Press (2006).
- [14] Austin, J.L.: *How to Do Things with Words*. Oxford University Press (1962). 坂本百大 (訳), 言語と行為, 大修館書店 (1978).
- [15] 久保 進, 阿部桂子, 越智希美子, 鈴木光代, 向井留実子: 発語内行為の意味ネットワーク—言語行為論からの辞書的対話事例分析, 晃洋書房 (2002).
- [16] 松本斉子, 村井 源, 往住彰文: 看護師対話ログにおける行為遂行の分析, 日本認知科学会第 25 回大会, pp.18-21 (2008).
- [17] 徳久良子, 寺嶋立太: 非課題遂行対話における発話の特徴とその分析, 人工知能学会論文誌, Vol.22, No.4, pp.425-435 (2007).
- [18] 徳久雅人, 前田浩佑, 村上仁一, 池原 悟: 心的状態を表す対話行為タグ付きテキスト対話コーパスの構築, 電子情報通信学会技術研究報告, 思考と言語, Vol.107, No.387, pp.25-30 (2007).
- [19] 大山鉄郎, 稲垣陽一, 関 洋平: 発話と発語内行為命名動詞を用いた言語行為コーパスの自動構築と分析, 情報アクセスシンポジウム, pp.15-22 (2012).
- [20] 河内彩香: 日本語の雑談の談話における話題展開機能と型, 早稲田大学日本語教育研究, Vol.3, pp.41-55 (2003).
- [21] 宇佐美まゆみ, 野口英美, 木林理恵: 初対面二者間会話における主題導入と展開のプラクティス—対話相手との年齢差・性差に着目して, 言語・音声理解と対話処理研究

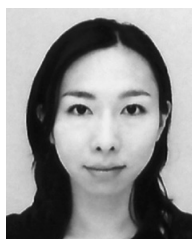
会, Vol.71, pp.23-28 (2014).

- [22] 国立国語研究所: 現代日本語書き言葉均衡コーパス, 入手先 (https://pj.ninjal.ac.jp/corpus_center/bccwj/) (参照 2019-04-05).
- [23] Yamazaki, M., Miyazaki, Y. and Kashino, W.: Annotation and Quantitative Analysis of Speaker Information in Novel Conversation Sentences in Japanese, *Proc. 11th International Conference on Language Resources and Evaluation (LREC 2018)*, pp.1078-1081 (2018).
- [24] 村井 源: 日本語小説の会話タグ付コーパスの開発に向けて, 第 31 回人工知能学会全国大会論文集, 1D2-OS-29a-2 (PDF) (2017).
- [25] Landis, J.R. and Gary G.K.: The Measurement of Observer Agreement for Categorical Data, *Biometrics*, Vol.33, pp.159-174 (1977).
- [26] 松尾 豊, 友部博教, 橋田浩一, 中島秀之, 石塚 満: Web上の情報から人間関係ネットワークの抽出, 人工知能学会誌, Vol.20, No.1, pp.46-56 (2005).



村井 源 (正会員)

東京大学工学部卒業。東京工業大学大学院社会理工学研究科博士課程修了。博士 (工学)。東京工業大学社会理工学研究科助教を経て、現在、公立はこだて未来大学複雑系知能学科准教授。専門はデジタルヒューマニティーズ, 計量文献学, 物語論, テキスト分析, 修辞構造分析等。



松本 斉子

聖心女子大学卒業。東京工業大学大学院社会理工学研究科博士課程修了, 博士 (学術)。東京工業大学大学院研究員, 九州大学大学院システム情報科学研究院研究員を経て退職。専門は語用論, ヒューマンインタフェース。