

# 話し言葉生成のための発話文表現文型辞書

夏目和子 佐藤理史

名古屋大学大学院工学研究科

natsume @ nuee.nagoya-u.ac.jp

## 概要

多様な言語形式の発話文の自動生成を可能とするために、発話意図毎にその意図を表す表現文型を整理した『発話文表現文型辞書』の編纂を進めている。この辞書の特徴は、それぞれの表現文型に、その文型の「話し方の特徴」をベクトルとして記述する点にある。本稿では、旧版からの改訂のポイントを中心に、その現状を報告する。主な変更点は、(1)発話意図のグルーピング見直し、(2)文型を接続することができる述語の記述方法の変更、(3)例文の整理、(4)表現文型を主要部と末尾部に分割、の4点である。現在の辞書は、60種類の発話意図に対して、8,055文型を収録している。

## 1 はじめに

日本語の特徴のひとつに、文末の述語形式の多様性がある。たとえば、話し手が聞き手に行為の許可を求める「許可要求」を表す表現には、付録Aに示すような多様な形式が存在する。これらの形式のうち、どの形式を使用するかは話し手次第であり、その選択はある種の情報を伝達する。

たとえば、「帰らせてくれ」という表現は男性的で強い要求という特徴を持つため、その表現の話し手は男性であろうと推測される。それに対して、「帰らせてくださいませんか」という表現は女性的で丁寧という特徴を持つため、話し手は大人の女性であろうと推測される。小説の会話では、このような情報伝達を巧みに用い、地の文で明示することなしに会話の話し手が誰であるかを読者に提示する。これは、会話の書き分けと呼ばれ、小説執筆のひとつの技法ともなっている[1]。

我々は、小説の自動生成[2]を進める過程で会話の生成が必要となり、これを実現するための『発話文表現文型辞書』の編纂を進めている[3,4]。この辞書では、発話の目的を「発話意図」として分類し、それぞれの発話意図に対して、その意図を表現する表現文型を収録している。それぞれの表現文型には、

その表現が話し方としてどのような特徴を持つかという情報（「話し方の特徴」）を付与している点に本辞書の独自性がある。

本稿では、旧版[4]からの変更点を中心に、この辞書の現状を報告する。

## 2 辞書の概要

『発話文表現文型辞書』のエントリは、特定の発話意図を表す表現文型であり、主に下記の4種類の情報で構成されている。

1. 発話意図＝発話の目的
2. 表現文型＝発話意図を表現する言語形式
3. 例文
4. 話し方の特徴ベクトル＝6軸20次元

2021年12月現在の本辞書のサイズは、発話意図60種類、エントリ数8,055、例文数23,489である。

## 3 発話意図と文型

本辞書における発話意図とは、話し手の発話時における発話の目的である。その発話がどのように聞き手に理解されるかまでは含意しない。

本辞書の発話意図および表現文型の決定には、荒木[5]、グループ・ジャマシィ[6]、市川[7]などを参考にした。なお、辞書の全体像を把握するために、発話意図を11のグループに分類し、追加・再編成を続けている。表1に、発話意図のグループリストを示す。

発話意図のグルーピングに厳密なルールは決め難いが、宮島[8]の発話・伝達上の文の類型を拝借すると、グループRとOは働きかけの文、JとIとEは意志・願望の表出の文、TとCは現象文・判断文、Qは質問文ということになる。なお、グループAとKの大部分は定型表現である。

グループCには、「だろう」「らしい」「みたいだ」「ようだ」「かもしれない」「ちがいない」「はずだ」などのモダリティを表す文型を収録した。エントリ数は5,700以上であり、今後、適切な発話意図を設定して整理する予定である。

表 1 発話意図のグループリスト

グループ	発話意図	発話意図の数	エントリの数
R	行為要求 命令, 依頼, 勧誘, 忠告, 禁止など	9	378
O	意向伺い 申し出	1	26
J	態度表明 褒める, 貶す, 非難など	5	212
I	意志表示 願望, 後悔, 決心など	8	261
E	感情表出 -	0	0
O	意向伺い 申し出	1	26
T	情報伝達 説明, 伝聞など	6	1,132
C	情報の確からしさ 推量, 断定, 確信など	13	5,714
Q	質問 確認, 念押し, 許可要求など	4	103
A	応答 肯定, 承諾, 否定, 断りなど	10	172
K	感動詞 よびかけ, 別れ, 感謝, 謝罪	4	57
		60	8,055

## 4 表現文型と例文

表現文型とは、発話意図を表現する際に使用される言語形式のことである。本辞書の表現文型の大半は文末の述語形式であるため、以下の2つの要素を記述する。

1. 核となる述語（述語核）の活用型とテンス・否定の有無の情報（接続形式）
2. 述語核に後続する助動詞や助詞などの文字列。主要部と末尾部に分けて記述する

発話意図 R05「忠告」の接続形式、主要部、末尾部、例文を表2に示す。

### 4.1 接続形式

旧版では述語核の情報を、日本語教育で用いられている文法関連の記号で記述していた。これを変更し、今回の版からはHaoriBrics3 (HB3) [9]による表層文の生成が容易になるような記述法を採用した。

接続形式は、活用型とテンス・否定の有無を下記の記号を組み合わせで記述する。

表 2 発話意図 R05「忠告」の表現文型と例文

接続形式	主要部	末尾部	例文
Vbnt	方がいい		(食べる/食べない/食べた)方がいい
Vbnt	方がいい	よ	(食べる/食べない/食べた)方がいいよ
Vbnt	方がいい	ね	(食べる/食べない/食べた)方がいいね
Vbnt	方がいい	わ	(食べる/食べない/食べた)方がいいわ
Vbn	ことだ		(食べる/食べない)ことだ
Vbn	ことだ	ね	(食べる/食べない)ことだね
Vbn	こと		(食べる/食べない)こと
Vbn	こと	ね	(食べる/食べない)ことね
Vbn	ことです		(食べる/食べない)ことです
Vb	命令形	よ	食べろよ
Vb	なさい		食べなさい
Vb	お～なさい		お食べなさい
Vn	テは	ね	食べなくてはね
Vn	チャ形	ね	食べなくちゃね

V 動詞型の活用語

A イ形容詞型の活用語

C 判定詞型の活用語

b 無標（テンス・否定なし）

n ない

t タ

k ない+タ

たとえば、表2の発話意図「忠告」では、主要部「方がいい」に接続する述語核は、Vbnt（動詞型：無標、ない、タ）である。これは、この主要部に例文に示すような3種類の形式が接続すること意味する。一方、主要部「ことだ」に接続する述語核は、Vbn（動詞型：無標、ない）である。なお、Vt（動詞型：タ）に接続する「食べたことだ」は非文ではないが、忠告の意図を表さないので接続形式に含まれない。

### 4.2 主要部と末尾部

主要部には、述語核に後続する助動詞（相当句）と接辞などを記述する。これらを含まない文型の場合は、述語核の活用形などを、HB3のルールに従って記述する。具体的には、命令形やチャ形などである。

末尾部には、終助詞と文末記号（?!/…など）を記述する。末尾部を独立して記述するメリットは、第一に過不足なく記述できているか確認しやすいこ

と、第二に、後述する「話し方の特徴」を記述する際に、一貫性を保ちやすくなることである。終助詞は多くの場合、女性的・男性的、大人っぽい・子供っぽい、強弱などの特徴が表されるので、他の発話意図の「話し方の特徴」と比較しながら作業を行うことが容易となる。

### 4.3 例文

例文では原則として、活用型毎に共通の述語を使用した。具体的には、動詞型「食べる」、イ形容詞型「美しい」、判定詞型「綺麗だ」である。

表2で示すように、一つのエントリに複数の接続形式を記述した場合は、記号（/）を用いて、全ての例文をまとめて記述している。

## 5 話し方の特徴ベクトル

話し方の特徴は、表3に示す6軸 20要素のベクトルとして記述する。この記述法を定めるにあたっては、日本語記述文法研究会[10, 11]から、終助詞の機能、談話の文末表現に見られる文体差、日本語の複雑な待遇的意味などを参考にした。

表3 話し方の特徴ベクトルの軸と要素

	軸	要素	値
1	ジェンダー	男性的	0,1,2,3
2		女性的	0,1,2,3
3	年齢	大人っぽい	0,1,2,3
4		子どもっぽい	0,1,2,3
5	長さ	長い	0,1,2,3
6		短い	0,1,2,3
7	強さ	強い	0,1,2,3
8		弱い	0,1,2,3
9	a 敬意の有無	上扱い	0,1,2,3
10		下扱い	0,1,2,3
11	b 距離	遠ざけ	0,1,2,3
12		親しみ	0,1,2,3
13	c 場面差	あらたまり	0,1,2,3
14		くだけ	0,1,2,3
15	d 品格	丁寧	0,1,2,3
16		ぞんざい	0,1,2,3
17	印象	明るい	0,1
18		暗い	0,1
19		可愛い	0,1
20		古風	0,1

各要素の値域は、0から3（印象軸は0か1）の整数値である。対立する要素の値は、かならず一方は0であり、両方に1以上の値を付与することはない。

以下では、ベクトルの値をどのように付与しているか、その概略を示す。

### 5.1 ジェンダーと年齢の軸

ジェンダーと年齢の軸は、話者の属性に関係がありそうだが、あくまで話し方の特徴、口調についての要素とする。値は、おおよそ文型を構成する要素に基づいている。具体例を以下に示す。

1. 終助詞さ、ぞ 男性的-1
2. 終助詞ぜ 男性的-2
3. 終助詞わ U 女性的-1
4. 終助詞わよ、わね 女性的-2
5. からだ、からだよ 男性的-1
6. からよ、からね 女性-1
7. からじゃ 大人っぽい-3
8. のだ、のだよ 男性的-1
9. の、のよ 女性的-1
10. のね 女性的-2
11. です/ます+ぞ/な 男性的-2, 大人っぽい-2
12. です/ます+の/のよ 女性的-2, 大人っぽい-2
13. だろう、だろうな 男性的-1
14. でしょうね 女性的-1
15. たまえ 男性的-2, 大人っぽい-1
16. やがれ 男性的-2
17. じゃん 子どもっぽい-1

なお、終助詞「わ」は、必要に応じてイントネーションをU（上昇調）とD（下降調）の2種類に区別している。No.3のように上昇調の「わ」には女性的-1を付与するが、下降調の「わ」には付与しない。

3段階のうちどの値を付与するか原則の一部を以下に示す。

- 男性的-2, 女性的-2は、反対のジェンダーの話者がめったに使わない表現文型に付与する
- 大人っぽい-2は中年、-3は老人が使う表現文型に付与する
- 子どもっぽい-1は若者言葉、-3は幼児語に付与する

### 5.2 長さ強さの軸

長さ軸は、表現文型の見目の長短に基づいて付与する。たとえば、「助動詞+終助詞」を基準とす

ると、助動詞のみの文型は、**短い-1**、「助動詞+接辞+終助詞」は、**長い-1**を付与する。

強さ軸は、表現文型に、口調や意味を強める文字列(だぞ、とも、よね)や記号(!)が含まれていれば、**強い-1**を、省略やぼかし表現(とか、けど)や記号(…)が含まれていれば、**弱い-1**を付与する。強調要素、ぼかし要素が複数含まれていれば、**強い-2**や**弱い-2**などを付与する。

### 5.3 待遇表現の軸

待遇表現には、上向き vs. 下向きの対立がある。日本語の待遇的意味は複雑であるため、表3に示すとおり、待遇表現の軸の下にaからdの4軸を設け、それぞれに(上向き vs 下向き)を表す要素を設定した。a 敬意の有無(上扱い vs 下扱い)、b 距離(遠ざけ vs 親しみ)、c 場面差(あらたまり vs くだけ)、d 品格(丁寧 vs ぞんざい)である。これによって、ある軸では上向きであると同時に、別の軸では下向きである表現の特徴を数値化できるようになった。

たとえば、大人が子どもに対して、命令する表現で、「お～なさい」や「ごらんなさい」を含む文型は、a 敬意の有無軸が下向きで、d 品格軸が上向きである。本辞書では以下の値を付与している。

- R01「お食べなさい」  
下扱い-1, あらたまり-1, 丁寧-1
- R01「食べてごらんなさい」  
下扱い-1, 親しみ-1, あらたまり-1, 丁寧-1

また、近年よく耳にする「です」を「ッス」に置き換える表現について、中村[12]は、体育会男子学生が先輩に「そっすね」「おもしろいっす」などと言うのは、親しさと丁寧さを同時に表すためであると分析している。本辞書では「ッス」を含む表現文型は、敬意と親しみを同時に表すものとして、以下の値を付与している。

- 男性的-1, 子どもっぽい-1, 上扱い-1, 親しみ-1

付録に示したQ06「許可要求」のNo.7とNo.19がそれに該当する。

### 5.4 印象の軸

印象軸は、他の5軸の補助的な位置付けで、現状では対立する要素を設けず、フラグとして運用している。

たとえば、I05後悔「食べちゃった」には、**可愛い-1**を付与している。

また、R02依頼「食べておくれよ」は時代劇の子どもの発話の印象があるので、**古風-1**を付与している。

## 6 本辞書の利用

本辞書は、発話文生成システムの開発とともに改訂を重ねてきた。刀山[13]では初版を、木村[14]では2版を、米田[15]では本稿の版を利用して発話文生成を実現している。

発話文生成における本辞書の役割は、与えられた発話意図に対して、その発話意図を表す表現文型を提供することである。可能な表現文型の中から実際に提供する文型を決定する際に、表現文型に付与されたベクトルを利用する。

初版から2版への主な変更は、「性別・年齢を中心とした8次元の話者属性ベクトル」から「20次元の話し方の特徴ベクトル」への変更である。刀山[13]のシステムは、入力された話者属性ベクトルからその話者らしい発話を生成するのに対し、木村[14]のシステムは、入力された話し方の特徴ベクトルに基づいてその特徴を有する発話を生成する。この変更は、同じ人物でも、状況や感情や話し相手に応じて様々な表現文型を用いることに対応するためである。

米田[15]のシステムは、どのような性格の人(キャラクター)がどのような話し方の特徴(口調)を用いるかを収集したデータベースを使って、性格から口調を推定する。それにより、入力された性格に対して、その性格を反映した口調の発話を生成する。

## 謝辞

本研究はJSPS科研費JP18H03285、JP21H03497の助成を受けたものです。

## 参考文献

1. 大沢在昌. 小説講座売れる作家の全技術. 角川書店, 2012.
2. 佐藤理史. コンピュータが小説を書く日. 日本経済新聞出版社, 2016.
3. 夏目和子, 刀山将大, 佐藤理史. 発話文自動生成のための日本語表現文型辞書の作成. 言語資源活用ワークショップ2016発表論文集, pp. 126-136, 2017.
4. 夏目和子, 佐藤理史. 発話文表現文型辞書の設計と編纂. 言語資源活用ワークショップ2019発表論文集, pp. 295-312, 2019.

- 
5. 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.
  6. グループブジャマシィ. 教師と学習者のための日本語文型辞典. くろしお出版, 2009.
  7. 市川保子. 日本語類義表現と使い方のポイント: 表現意図から考える. スリーエーネットワーク, 2019.
  8. 宮島達夫, 仁田義雄編. 日本語類義表現の文法, 上:単文編. くろしお出版, 1995.
  9. 佐藤理史. HaoriBricks3: 日本語文を合成するためのドメイン特化言語.自然言語処理, Vol. 27, No. 2, pp. 411-444, 2020.
  10. 日本語記述文法研究会. 現代日本語文法 4. くろしお出版, 2003.
  11. 日本語記述文法研究会. 現代日本語文法 7. くろしお出版, 2009.
  12. 中村桃子. 新敬語「マジヤバイっす」: 社会言語学の視点から. 白澤社, 2020.
  13. 刀山将大, 夏目和子, 佐藤理史, 松崎拓也. 話者の特徴を反映した発話文生成器の作成. 言語処理学会第 23 回年次大会発表論文集, B1-3, 2017.
  14. 木村遼, 夏目和子, 佐藤理史, 松崎拓也. 発話表現文型辞書を利用した多様な発話文生成機構. 2018 年度人工知能学会全国大会 (第 32 回), 2E2-02, 2018.
  15. 米田智美, 佐藤理史. 話者の性格を反映した発話文の生成. 言語処理学会第 28 回年次大会発表論文集, 2022 (発表予定).

# A 付録

## Q06 「許可要求」

連番	発話意図	接続形式	主要部	末尾部	例文	イントネーション	男性的	女性的	大人っぽい	子どもっぽい	長い	短い	強い	弱い	上扱い	下扱い	遠ざけ	親しみ	あらたまり	くだけ	丁寧	ぞんざい	明るい	暗い	可愛い	古風
1	Q06	Vb	おいしい	?	食べていい?			1	1		1				1	1	1	1	1							
2	Q06	Vb	おいしい	か	食べていいか		2								1	1	1	1								
3	Q06	Vb	おいしい	かい	食べていいかい		2								1	1	1	1								
4	Q06	Vb	おいしい	かな	食べていいかな			1						1	1	1	1									
5	Q06	Vb	おいしい	かしら	食べていいかしら			2						1	1	1										
6	Q06	Vb	おいしいです	か	食べていいですか						1				1							1				
7	Q06	Vb	おいしいっす	か	食べていいっすか		1		1	1					1		1	1	1	1	1		1			
8	Q06	Vb	おいいでしょ	か	食べていいでしょか						1			1	1							1				
9	Q06	Vb	てよるしい	か	食べてよるしいか				1					1								1				
10	Q06	Vb	てよるしいです	か	食べてよるしいですか				1	1				1								1	1			
11	Q06	Vb	てよるしいでしょ	か	食べてよるしいでしょか				1	1				1	1							1	2			
12	Q06	Vb	てもいい	?	食べてもいい?			1	1							1	1	1	1							
13	Q06	Vb	てもいい	か	食べてもいいか		2				1					1	1									
14	Q06	Vb	てもいい	かい	食べてもいいかい		2				1					1	1					1				
15	Q06	Vb	てもいい	かな	食べてもいいかな			1			1			1	1	1	1									
16	Q06	Vb	てもいい	かね	食べてもいいかね		1		2		1					1						1				
17	Q06	Vb	てもいい	かしら	食べてもいいかしら				2		1			1	1	1										
18	Q06	Vb	てもいいです	か	食べてもいいですか						2			1	1							1				
19	Q06	Vb	てもいいっす	か	食べてもいいっすか		1		1	2				1	1			1	1	1	1	1	1			
20	Q06	Vb	てもいいでしょ	か	食べてもいいでしょか						2			1	1							1				
21	Q06	Vb	てもよるしい	か	食べてもよるしいか		1		1	1						1	1					1				
22	Q06	Vb	てもよるしいです	か	食べてもよるしいですか				1	2					1		1	1	1	1						
23	Q06	Vb	てもよるしいでしょ	か	食べてもよるしいでしょか				1	2				1	1		1	1	1	2						
24	Q06	Vb	させて		掃らせて			1				1	1			1	1	1								
25	Q06	Vb	させて	よ	掃らせてよ			1						1		1	1	1								
26	Q06	Vb	させてくれ		掃らせてくれ		2							1		1	1	1								
27	Q06	Vb	させてくれ	よ	掃らせてくれよ		2			1	1				1	1	1									
28	Q06	Vb	させてくれない	?	掃らせてくれない?			1		1						1	1									
29	Q06	Vb	させてくれない	か	掃らせてくれないか			1			2					1	1	1								
30	Q06	Vb	させてくれませんか	か	掃らせてくれませんか						2					1						1				
31	Q06	Vb	させてください		掃らせてください					1	1	1		1									1			
32	Q06	Vb	させてくださらない	?	掃らせてくださらない?			2	1	2					1		1					1				
33	Q06	Vb	させてくださいませんか	か	掃らせてくださいませんか			1	1	3					2	1						2				
34	Q06	Vb	させていただけない	?	掃らせていただけない?			2	1	2												1		1		
35	Q06	Vb	させていただけませんか	か	掃らせていただけませんか				1	3					2	1							2			
36	Q06	Vb	させてもらえない	?	掃らせてもらえない?			1	1	1						1	1	1								
37	Q06	Vb	させてもらえん	か	掃らせてもらえんか		1		2	2						1	1									
38	Q06	Vb	させてもらえませんか	か	掃らせてもらえませんか				1	3							1						1			
39	Q06	Vb	させてほしい		掃らせてほしい							1	1			1						1				
40	Q06	Vb	させてほしいの	の	掃らせてほしいの			2						1		1	1	1								
41	Q06	Vb	させてほしいんだ	が	掃らせてほしいんだが		2		2	1						1	1									
42	Q06	Vb	させてほしいんだけど		掃らせてほしいんだけど					1	1					1	1					1				