

# 文末機能表現シソーラスの編纂に向けて —文末機能表現の網羅的生成—

松木 久幸      佐藤 理史      駒谷 和範

名古屋大学大学院 工学研究科 電子情報システム専攻

{h.matuki, ssato, komatani}@nuee.nagoya-u.ac.jp

## 1 はじめに

日本語には、「文末の用言に接続」し、「機能的な働きをする」表現が多数存在する。本研究では、これらを**文末機能表現**と呼ぶ。文末機能表現は、ヴォイス・テンス・アスペクト・モダリティ・否定などに関わる意味を表す。たとえば、「原稿を書きはじめてのほうがよい」という文において、文末機能表現「～はじめてのほうがよい」は、“ある動作を開始すること(アスペクト)を勧める(モダリティ)”意味を表す。

文末機能表現には、表層的な形式は異なるが、同じような意味を表す表現が存在する。たとえば、「～てくださいか」「～もらえないか」「～していただけませんか」は、いずれも“依頼”を表す文末機能表現である。意見マイニングなどの応用では、これらの同義関係を判定する必要がある。同義関係の判定には、シソーラス(同義語辞書)を用いた方法が一般的である。

機能表現に関するシソーラスとして、日本語機能表現辞書『つつじ』[2]がある。『つつじ』には、16,771の表層形が収録されており、これらに対して199種類の意味IDが付与されている。この意味IDにより、“機能表現間(1対1)”の同義関係の判定が可能である。

ところが、“機能表現の列からなる表現間(m対n)”の同義関係の判定が必要となる場合もある。たとえば、“否定”を表す機能表現「～ない」「～わけがない」の列で構成される「～ないわけがない」は、“必然性”を表す「～に違いない」と同義関係にある。『つつじ』は、言語学の文献に記載されている単位の表現(「～ない」「～わけがない」など)を収録しており、判定できる同義関係が“機能表現間(1対1)”に限定されているため、「～ないわけがない」と「～に違いない」のような同義関係を判定できない。

梶田ら[3]は、上述した“機能表現の列からなる表現間”の同義関係の判定の問題を取り上げ、『つつじ』に収録されている機能表現のN-gramを見出し語として収録したシソーラスの編纂に取り組んだ。しかしな

がら、梶田らの方法では、エントリー数や意味ラベルの異なり数が膨大となってしまうという問題が生じ、実用的なシソーラスを実現するに至っていない。この研究から、機能表現の構成要素として適切な単位設定をすることが重要であることが判明した。

我々は、“機能表現の列からなる表現間”の同義関係の判定を実現するため、「文末機能表現の列全体」を見出し語として収録したシソーラスの編纂を目指す。本稿では、シソーラス編纂の過程うち、「シソーラスの見出し語リストの作成のステップ」、すなわち、「文末機能表現の網羅的生成」について述べる。

文末機能表現の網羅的生成は、以下の手順で行う。まず、文末述語表現の構造モデルを定める(2節)。次に、この構造モデルに基づき、最小構成要素を列挙(3節)し、それらを接続する(4節)。

## 2 文末述語表現の構造モデル

文末の用言とその右側(つまり‘用言+文末機能表現’の部分)を**文末述語表現**と呼ぶ。本節では、この文末述語表現の構造モデルについて述べる。準拠する文法体系には益岡・田窪文法[1]を採用する。その理由は、益岡・田窪文法に従う<sup>1</sup>ことで、文末機能表現を構成する要素数を抑えられるからである。

以下の手順で文末述語表現の構造モデルを定める。

1. 文(実在する文または作例)において、直感に従って、文末の述語表現の範囲を定める。
2. 切り出された文末述語表現内に以下の3種類の境界を認定することで、文末述語表現をいくつかの部分に分割する。

文末となりうる境界 ‘||’

文節末となりうる境界 ‘|’

左が文節として成立しなくなる境界 ‘┐’

<sup>1</sup>益岡・田窪文法は、「書いて」「書いた」などを活用形(テ形・タ形)として扱うので、語として扱う単位が大きくなる。

3. こうして得られた部分を整理・グループ化することにより、文末述語表現のモデルを定める。

たとえば、例文 (1) において、文末述語表現は下線部のように定められる。

- (1) 事実が書かれなくなっているのではないだろうか。

この文末述語表現に上記の 3 種類の境界を挿入し、最小構成要素へと分割すると、次のようになる。

- (2) (事実が) 書か-れなく | なって | いる || のでは | ない || だろう || か。

このように定まった最小構成要素 (以下、ユニットと呼ぶ) を、以下の 5 種類のタイプに分類する。なお、‘U’ はユニットの略である。

- 境界 ‘||’ の直後に現れる表現を、**助動詞 U** または **終助詞 U** とする。
- 境界 ‘|’ の直後に現れる表現を、**準用言 U** とする。
- 境界 ‘-’ の直後に現れる表現を、**接尾辞 U** とする。
- 残った部分は、**用言** とする。

例文 (2) では、「のでは」「だろう」が助動詞 U、「か」が終助詞 U、「なって」「いる」「ない」が準用言 U、「れなく」が接尾辞 U、「書か」が用言である。

さらに、ユニットよりもう少し大きい単位 (‘部’ と呼ぶ) を導入する。まず、最左の境界 ‘||’ より右側の部分 (‘|| のでは | ない || だろう || か’) に注目し、このうち、終助詞 U (‘|| か’) を取り除く。それ以外の部分 (‘|| のでは | ない || だろう’) をひとまとまりとして **助動詞部** と呼ぶ。次に、最左の境界 ‘||’ より左側の部分 (‘書か-れなく | なって | いる’) のうち、最左の境界 ‘|’ より右側の部分 (‘| なって | いる’) をひとまとまりとして **準用言部** と呼ぶ。このように、‘ユニット’・‘部’ という単位を導入することで、図 1 に示す構造モデルを定めた。

図 1 に示すように、準用言部は準用言 U と接尾辞 U から構成され、助動詞部は助動詞 U と接尾辞 U と準用言部から構成されている。たとえば、先の例文 (2) では、準用言部 ‘| なって | いる’ は準用言 U ‘なる’ ‘いる’ から構成され、助動詞部 ‘|| のでは | ない || だろう’ は助動詞 U ‘のでは’ 準用言部 ‘ない’ 助動詞 U ‘だろう’ から構成されている。

作成するシソーラスの見出し語には、用言を品詞で抽象化した表現 (すなわち、文末機能表現) を採用する。たとえば、先の例文 (2) の場合は、「書か」を ‘動詞’ として抽象化し、「動詞. れなくなっているのではないだろうか」を見出し語として採用する。

$$\text{文末述語表現} = \text{用言} \{ \text{接尾辞 U} \}^? \{ | \text{準用言部} \}^? \{ || \text{助動詞部} \}^? \{ || \text{終助詞 U} \}^?$$

$$\text{用言} = \left\{ \begin{array}{l} \text{動詞} \\ \text{形容詞} \\ \text{名詞} + \text{判定詞} \end{array} \right\}$$

$$\text{準用言部} = \{ | \text{準用言 U} \{ \text{接尾辞 U} \}^? \}^+$$

$$\text{助動詞部} = \{ || \text{助動詞 U} \{ \text{接尾辞 U} \}^? \{ | \text{準用言部} \}^? \}^+$$

図中 ‘?’ は「あってもなくてもよい」、  
‘+’ は「1 回以上繰り返す」ことを表す。

図 1: 文末述語表現の構造モデル

### 3 ユニットとして採用する表現

ユニットとして採用する全 165 種類の表現を、表 1 に示す。これらの表現に対して、「活用型 (どのような活用をするのか)」「左接続条件 (どのような表現に接続するのか)」を記述する。これらの情報は、4 節で述べるユニットの接続を行う際に用いる。

#### 3.1 接尾辞 U

接尾辞 U として採用するのは、基本的に「未然形に接続する」表現 (表 1 の I, II) である。

I のグループには、ヴォイスを表す表現 (‘(ら) れる’ ‘(さ) せる’), 丁寧さを表す表現 (‘ます’), 否定関連の表現 (‘ない’ ‘ぬ’ ‘ざるをえない’) がある。このうち、‘ます’ は連用形接続であるので、原則に従えば準用言 U に配置することになるが、自立性が低いので接尾辞 U として採用する。

II のグループは、I のグループの表現の列で構成される表現である。文末機能表現を構成するユニット数を抑えるため、接尾辞の列を 1 つのユニットとして扱う。

#### 3.2 準用言 U

準用言 U として採用するのは、基本的に「連用形に接続する」表現 (表 1 の III, IV, V) であるが、条件形に接続する表現 (表 1 の VI) も準用言 U に含める。

III のグループは、基本連用形 (‘R’ と表す) に接続する準用言 U である。このグループの中心は、複合動詞の後項 (可能形も含める) である。‘R やすい’ ‘R にくい’ などは、しばしば接尾辞として扱われるが、前項が受身形・使役形をとりうるので、準用言 U として採用する。

IV のグループは、タ系連用テ形 (‘テ’ と表す) に接続する準用言 U である。このグループの中心は、複合動詞の後項と、補助形容詞 (‘テほしい’ ‘テいい’ など) である。‘テいる’ は、‘い’ を省略した形 (‘書いて | いる’ ‘書いて | ない’ など) で出現しうる。これらを文末機能表現として認定するために、‘テる’ ‘テ

表 1: ユニットとして採用した表現一覧 (合計:165 種類)

接尾辞 U (19 種類)	I	(ら) れる	(さ) せる	ます	ない	ぬ	ざるをえない
	II	(さ) せ(ら) れる	(ら) れます	(ら) れない	(ら) れぬ	(ら) れざるをえない	
		(さ) せます	(さ) せない	(さ) せざるをえない		(さ) せ(ら) れます	
		(さ) せぬ	(さ) せ(ら) れない		(さ) せ(ら) れぬ	(さ) せ(ら) れざるをえない	
準用言 U (88 種類)	III	R はじめる	R だす	R だせる	R つづける	R おえる	R おわる
		R おわれる	R やむ	R あがる	R あがれる	R あげる	R とおす
		R とおせる	R きる	R ぎれる	R ぬく	R ぬける	R つくす
		R つくせる	R わずれる	R かねる	R そこなう	R そびれる	R そんじる
		R おとす	R おとせる	R しぶる	R なおす	R なおせる	R あう
		R あえる	R かえす	R かける	R つつある	R くださる	R なさる
		R うる	R たい	R よい	R いい	R やすい	R にくい
		R づらい	R がたい	R がちだ	R そうだ	R すざる	
	IV	テいる	テる	テない	テいく	テいける	テある
		テくる	テこれる	テしまう	テしまえる	テいただく	テいただける
		テくれる	テくださる	テもらう	テもらえる	テやる	テやれる
		テあげる	テさしあげる	テおく	テおける	テみる	テみせる
		テほしい	テいい	テよい	テハだめだ	テハならない	
	V	R ある	R なる	R する	R ない		
	VI	バいい	バよい	バいけない	バならない	タラいい	タラいけない
		タラだめだ	タラよい				
助動詞 U (34 種類)	VII	のだ	わけだ	はずだ	ようだ	つもりだ	ことだ
		ものだ	だろう	らしい	みたいだ	です	そうだ
		べきだ	まい	ばかりだ	ところだ		
	VIII	ほうがいい	かもしれない	にちがいない	という	とのことだ	とする
		ことがある	こともある	ことがない	こともない	ことができる	こともできる
		つもりがある	つもりがない	つもりもない	はずがない	はずもない	わけがない
終助詞 U (24 種類)	IX	か	かい	かしら	ぜ	ぞ	つけ
		って	とか	な	なあ	ね	ねえ
		よ	わ				
	X	かしらね	かな	かね	かよ	つけね	つてね
		よね	よねえ	わね	わよ		

‘R’ は基本連用形接続、‘テ’ はタ系連用テ形接続、‘バ’‘タラ’ は条件形接続、‘テハ’ はテ形+助詞「は」の形に接続することを表す。

ない」を準用言 U として採用する。「テハだめだ」「テハならない」は、テ形には直接接続せずに、取り立て助詞「は」が挿入された形(「～ては」)にのみ接続する表現である。

V のグループは、形容詞の連用形に接続する表現である。このうち、「R する」「R なる」は、形容詞を動詞化する働きをもつ表現である。「R ない」は、形容詞・判定詞に接続する否定表現である。接尾辞 U として採用した「ない」は、動詞の未然形に接続する「～ない」であった。これに対して、形容詞・判定詞に接続する「～ない」は、「寒く(φ, は, も)ない」のように取り立て助詞「は, も」の挿入を許す。そのため、これを準用言 U 「R ない」として採用する。

VI のグループは、条件形(「～たら」「～ば」)に接続する表現である。これらは、文節末となりうる境界(I)の右側に出現するので、準用言 U として採用する。

### 3.3 助動詞 U

助動詞 U として採用するのは、益岡・田窪文法が定める助動詞の概念 [1](p. 29) に従って、基本的には

「基本形、タ形、連体形に接続する」表現(表 1 の VI I)とする。例外として「文末形に接続する」表現(表 1 の VIII)も助動詞 U に含める。

VII のグループは、益岡・田窪に助動詞一覧として記載されている表現が大部分である。これらに加えて「ばかりだ」「ところだ」も採用する。

VIII のグループは、助動詞相当の複合辞である。助動詞相当の複合辞として採用する基準は、「文末形に接続する」ことと、「原則として、モダリティに関わる意味をもつ」ことの 2 つである。

### 3.4 終助詞 U

終助詞 U として採用するのは、「文末形に接続する」表現(表 1 の IX, X)である。

IX のグループは、益岡・田窪文法に記載されている終助詞である。

X のグループは、IX のグループの表現の列で構成される表現である。文末機能表現を構成するユニット数を抑えるため、終助詞の列も 1 つのユニットとして扱う。

## 4 文末機能表現の生成

### 4.1 ユニットの接続方法

前節で定めたユニットを接続することで、文末機能表現を生成する。本節では、「個々のユニットをどのように接続するか」と「ユニットをどのような順番で接続するか」について述べる。

個々のユニット(左側ユニットと右側ユニット)は、以下の手順で接続する。

1. 左側ユニットを、その活用型に従ってすべての活用形を展開する。
2. これらの左側ユニットのうち、右側ユニットの左接続条件を満たす表現のみを接続する。
3. 接続した表現のうち、コーパス<sup>2</sup>に存在する表現のみを採用する。

なお、接続後のユニットの左接続条件には“左側ユニットの左接続条件”を、活用型には“右側ユニットの活用型”を採用する。

例として左側ユニット「Rなる」と右側ユニット「テいる」の接続を考える。まず、「Rなる」の活用型に従い「なる、なら、なり、なつて、…」などすべての活用形を展開する。次に、右側ユニット「テいる」の左接続条件に従って、タ系連用テ形の「Rなつて」と「テいる」の組のみを接続する。最後に、接続した表現「Rなつて|いる」をコーパスと照合し、コーパスに存在した場合のみ採用する。

上述の接続方法を用い、構造モデル(図1)に従って、以下の順に接続することで文末機能表現を生成する。

1. 図1の準用言部の構造式に従い、準用言部を生成。
2. 図1の助動詞部の構造式に従い、助動詞部を生成。
3. 「手順1で生成した準用言部」と、「手順2で生成した助動詞部」を接続。
4. 「接尾辞U」と、「手順3で生成した表現(準用言部/助動詞部)」を接続。
5. 「手順4で生成した表現(接尾辞U/準用言部/助動詞部)」と、「終助詞U」を接続。

たとえば、2節で取り上げた例文(2)の「～れなくなっているのではないだろうか」は、次のような過程を経て生成される。手順1では、「Rなる」「テいる」を接続することで、準用言部「| なつて | いる」が生成される。手順2では、「のだ」「Rない」「だろう」を接続することで、助動詞部「|| のでは | ない || だろう」が生成される。手順3で「手順1で生成した表現」と

<sup>2</sup>コーパスには、毎日新聞15年分(1991-2005年版)を使用した。

表2: 生成された文末機能表現の数

接続した要素	表現の数
準用言部	2,658
助動詞部	2,234
準用言部+助動詞部	13,568
接尾辞U+準用言部+助動詞部	21,207
接尾辞U+準用言部+助動詞部+終助詞U	41,061

「手順2で生成した表現」を接続し、手順4で左側に「-れなく」を、手順5で右側に「|| か」を接続することで、「-れなく| なつて | いる || のでは | ない || だろう || か」が生成される。

### 4.2 生成結果

表1のユニットとして採用した表現(全165種類)を4.1節の方法で接続し、文末機能表現を生成した。生成された文末機能表現の数を表2に示す。

ユニット数が最多な表現は、「-れる || ように | なつて | きた || のでは | ない || だろう || か」などの8つのユニットからなる表現であった。最終的に生成された文末機能表現は、41,061種類となった。

## 5 おわりに

文末述語表現の構造モデルを定め、このモデルに基づいて最小構成要素であるユニットを列挙(165種類)し、それらを接続することで41,061種類の文末機能表現を生成した。

残された課題として、生成した文末機能表現の網羅性の評価を行うことが挙げられる。単純な評価の方法は、文末機能表現を網羅的に列挙したリストと照合して、その網羅性を評価する方法である。ところが、そもそも文末機能表現を網羅的に列挙したリストは存在しない。つまり、何らかの形で間接的に評価を行う必要がある。たとえば、本稿とは別の方法で文末機能表現の一覧を作成し、その一覧と本研究の結果を照合する方法が考えられる。

今後は、シソーラスの編纂の次のステップとして、それぞれの見出し語に対して意味の記述を行う予定である。

**謝辞** 本研究は、科学研究費補助基盤研究(B)「辞書自動編纂のためのテクノロジー」課題番号21300094の支援を受けている。本研究では、毎日新聞1991-2005年版を使用した。

## 参考文献

- [1] 益岡隆志, 田窪行則. 基礎日本語文法-改定版-. くろしお出版, 1992.
- [2] 松吉俊, 佐藤理史, 宇津呂武仁. 日本語機能表現辞書の編纂. 自然言語処理, Vol.14, No.5, pp. 123-146, 2007.
- [3] 榊田達也, 佐藤理史, 藤田篤. 言い換えのための機能表現シソーラスの作成. 言語処理学会第15回年次大会論文集, pp. 88-91, 2009.