文末表現シソーラスの設計と編纂

桝 田 達 也　　佐 藤 理 史

名古屋大学大学院工学研究科

1. はじめに

日本語の文末表現には，多くのバリエーションがあり，それらの中には，意味的に等価な (言い換え可能な) ものが存在する．たとえば，“座る動作” を相手に依頼する場合，以下に示すような表現が可能である．

「座って下さい」,「お座りねがえますか」
「座って下さいませんか」,「座って」
「座ってよ」,「座ってちょうだい」

意味的に等価な表現を一つに集約することは，テキストマイニングや情報検索など，言語表現の意味を扱う応用では必須である．これを実現する方法の一つとして，文末表現を言い換える方法がある．

言い換えとは，ある表現をその意味内容を変えずに別の表現に置き換えることを言う．言い換えを実現するには，同義関係を定義したシソーラス (同義語辞書) を利用する方法が一般的である．

われわれは，昨年，機能表現辞書『つつじ』[1] を用いて機能表現シソーラスを作成することに取り組んだ [2]．具体的には,『つつじ』に収録されている機能表現の N-gram を作成し，その中で，文節機能部 (文節から内容語を切り離した残りの部分) としてコーパスに出現するものを，シソーラスの見出し語として採用した．しかしながら，この方法では，エントリー数や意味ラベルの異なり数が膨大になるという問題が発生し，実用的なシソーラスを実現するには至らなかった．この経験より，われわれは，機能表現のシソーラスの作成において，見出し語として適切な単位を設定することが重要であることを学んだ．

本研究では，対象を文の末尾の文節 (文末述語文節) に限定し，そこに現れる機能表現のシソーラスの実現を目指す．まず，文末述語文節の構造をモデル化し，見出し語として適切な単位を定める．次に，その構造モデルと単位に従って，シソーラスの見出し語を作成する．本稿では，これらの内容について述べる．

2. 文末述語文節のモデル化

本研究では,「基礎日本語文法-改訂版-」[3] に準拠し，文末述語文節の構造を，**基本ユニット**という単位を導入して整理する．

文末述語文節の構造を表 1 に示す．(1) に示すように，文末述語文節は，内容語を中心とした主述部と，それに後続する助動詞ユニット，終助詞ユニットから構成される．主述部は，文末述語文節の内容語の種類によって，動詞部，形容詞ユニット，名詞ユニットの 3 種類に分けられる．動詞部は，(3)〜(5) に示すように，動詞ユニットと補助ユニットから構成される．

基本ユニットの構造を表 2 に示す．この表で定義される**基本ユニットが，文末表現シソーラスの収録単位となる**．この表に示すように基本ユニットは 9 種類あり，内容語を含むものと，それ以外のものの 2 種類に分けられる．表中の “⟨⟩” は，その要素が内容語で置き換えられる要素であることを示す (以下，内容語変数と呼ぶ)．基本ユニットを構成する最小の要素 (テ形補助動詞，接尾辞 A など) を最小ユニットと呼ぶ．表 3 に，内容語変数を除く各最小ユニットに属する表現の例を示す．

2.1 内容語変数を含む基本ユニット

内容語変数を含む基本ユニットには，動詞ユニット，形容詞ユニット，名詞ユニットの 3 種類ある．

動詞ユニットは，内容語変数である動詞本体に，動詞性接尾辞が接続したものである．

形容詞ユニットは，形容詞本体に，接尾辞が接続したものである．形容詞は,「優しい」のようなイ形容詞と,「簡単だ」のようなナ形容詞 (形容動詞) の 2 種類に分けることができる．さらに，ナ形容詞は,「簡単だ」,「簡単である」,「簡単です」のように語尾の形によってダ列，デアル列，デス列の 3 種類に分けられる．これらの種類によって，形容詞ユニットの構造が (7)〜(10) のように異なる．また，ナ形容詞の語幹は,「簡単/かもしれない」(“/” は基本ユニットの境界) のように，直接助動詞ユニットに接続することがある (11)．

名詞ユニットは，名詞本体に，判定詞，接尾辞が接続したものである．名詞は，文末述語文節に現れる場合，基本的には判定詞を伴う．しかし，例外として,「あなた/かもしれない」のように，直接助動詞ユニットに接続する場合がある (15).「名詞＋判定詞」は，ナ形容詞をとみなすことができ，ナ形容詞と同様に，ダ列，デアル列，デス列の 3 種類に分類できる (12)〜(14)．

2.2 それ以外の基本ユニット

内容語変数を含まない基本ユニットには，4 種類の補助ユニットと，助動詞ユニット，終助詞ユニットの合計 6 種類ある．

補助ユニットは，補助動詞，あるいは，補助形容詞に接尾辞が接続したものである．連用形補助動詞や連用形補助形容詞とは，連用形の直後に接続する補助動詞，補

表 1 文末述語文節の構造

文末述語文節＝主述部＋助動詞ユニット*＋終助詞ユニット	(1)
主述部＝動詞部｜形容詞ユニット｜名詞ユニット	(2)
動詞部＝動詞ユニット＋連用形補助動詞ユニット*＋テ形補助動詞ユニット*	(3)
｜動詞ユニット＋連用形補助動詞ユニット*＋テ形補助形容詞ユニット*	(4)
｜動詞ユニット＋連用形補助形容詞ユニット*	(5)

※“*”は，その要素が任意個存在することを示す．

表 2 基本ユニットの構造

動詞ユニット＝〈動詞〉＋動詞性接尾辞*	(6)
形容詞ユニット＝〈イ形容詞〉＋形容詞性接尾辞*	(7)
｜〈ナ形容詞ダ列〉＋形容詞性接尾辞*	(8)
｜〈ナ形容詞デアル列〉＋動詞性接尾辞*	(9)
｜〈ナ形容詞デス列〉	(10)
｜〈ナ形容詞語幹〉	(11)
名詞ユニット＝〈名詞〉＋判定詞ダ列＋形容詞性接尾辞*	(12)
｜〈名詞〉＋判定詞デアル列＋動詞性接尾辞*	(13)
｜〈名詞〉＋判定詞デス列	(14)
｜〈名詞〉	(15)
連用形補助動詞ユニット＝連用形補助動詞＋動詞性接尾辞*	(16)
連用形補助形容詞ユニット＝連用形補助形容詞＋形容詞性接尾辞*	(17)
テ形補助動詞ユニット＝テ形補助動詞＋動詞性接尾辞*	(18)
テ形補助形容詞ユニット＝テ形補助形容詞＋形容詞性接尾辞*	(19)
助動詞ユニット＝助動詞	(20)
終助詞ユニット＝終助詞*	(21)

表 3 最小ユニットの例

連用形補助動詞	つづける，はじめる，たくなる，やすくする
連用形補助形容詞	たい，やすい，がたい，にくい
テ形補助動詞	ている，てほしくなる，てなさすぎる
テ形補助形容詞	てほしい，てよい，てならない
助動詞	はずだ，かもしれない，ことだ，ことになる
判定詞	だ
終助詞	か，かな，ね，わ，ぜ，なあ，よね
接尾辞 A	可能・受身・使役の接尾辞 (「せる」,「れる」など)
接尾辞 B	「うる」,「すぎる」
接尾辞 E	「ウとする」,「なければならない」など
接尾辞 M	「ます」,「ません」,「ませんでした」
接尾辞 N	否定の接尾辞 (「ない」,「ん」など)
動詞性接尾辞＝接尾辞 A*＋接尾辞 B*＋接尾辞 E*＋ (接尾辞 M*｜接尾辞 N*)	
形容詞性接尾辞＝接尾辞 E*＋接尾辞 N*	

助形容詞のことであり，たとえば,「〜続ける」,「〜始める」,「〜たい」,「〜やすい」などがそれに該当する．テ形補助動詞やテ形補助形容詞とは，タ系連用テ形の直後に接続する補助動詞，補助形容詞のことであり,「ている」,「てしまう」,「てほしい」,「てよい」などのことである．補助形容詞は,「なる」,「する」,「すぎる」を伴い，動詞化することがある．これらは，補助動詞として定義する．

助動詞ユニットは，助動詞単体と定義する．動動詞は，基本的に述語の基本形，タ形に接続し，複雑な述語を作る語である．

終助詞ユニットは，終助詞の列と定義する．終助詞は，一般的に終助詞と呼ばれているものである．

3. 文末表現シソーラスの概要

3.1 収 録 情 報

シソーラスのエントリーの例を表 4 に示す． この表に示すように，エントリーは 6 つの構成要素からなる．このうち，意味コードに関しては次節で述べる．

見出し語は，エントリーの ID である．

表記は，その基本ユニットの表層形 (複数可) を表す．表記において，内容語変数は，以下のように記述する．

〈英字〉-〈活用形〉

英字の部分は，V が動詞，A が形容詞，N が名詞を示す．

形態素情報とは，そのエントリーがどのような形態素とみなすことができるかを示す情報であり，以下のように記述する．

〈活用型の種類〉-〈活用形〉

エントリーが動詞，あるいは，形容詞単体の場合，その動詞や形容詞の活用型の種類によって，形態素情報が異なる．そのため，活用型の種類には，Mv，Ma という変数を記述する．

左接続情報とは，そのエントリーの直前に接続可能な

表 4 文末表現シソーラスのエントリーの例

見出し語	表記	形態素情報	左接続情報	右接続情報	意味コード
V-意志形■	V-意志形	Mv-意志形		■	意志$_1$:N:n
つづける-基本形	つづける	母音動詞-基本形	子音動詞-連用形; 母音動詞-連用形;…	all	継続$_2$:N:n
つづける-意志形■	つづけよう	母音動詞-意志形	子音動詞-連用形; 母音動詞-連用形;…	■	継続$_2$.意志$_1$:N:n
つづける. られる-基本形	つづけられる	母音動詞-基本形	子音動詞-連用形; 母音動詞-連用形;…	all	継続$_2$.可能$_1$:N:n
A-連用形. ない-基本形	A-連用形. ない	助動詞ナイ-基本形		all	否定$_1$:N:n
テいる. ない-推量形■	いなかろう	助動詞ナイ-推量形	子音動詞-タ系連用テ形;…	■	継続$_1$.否定$_1$.推量$_3$:N:n
テいる-タ系連用テ形	いて	母音動詞-タ系連用テ形	子音動詞-タ系連用テ形;…	other	継続$_1$:N:n
テいる-タ系連用テ形■	いて	母音動詞-タ系連用テ形	子音動詞-タ系連用テ形;…	■	継続$_1$.依頼$_1$:N:n

表 5 大まかな意味を示すノードのリスト

ムード・モダリティ	意志，依頼，受身，内-受益，内-授与，回想，確認，可能，完遂，感嘆，願望，疑問，許可，極端，経験，限定，命令，困難，使役，自然発生，自問，推量，勧め，相互動作，伝聞，当為，同意，判断，比況，頻度，不可能，不許可，不遂行，付帯，不必要，他-授与，容易，理由，改善，返報
アスペクト	事前，開始，継続，最中，完了，事後，発継続，着継続
否定	否定

表 6 意味情報が “推量” である基本ユニットの例

推量$_1$	V-連用形. そうだ
推量$_2$	かもしれない かもわからない のかもしれません
推量$_3$	にちがいない にきまっている にそういない らしい だろう みたいだ ようだ V-推量形

エントリーを示すもので，形態素情報の集合として記述する．ただし，そのエントリーが内容語変数を含む場合，左接続情報は記述しない．

右接続情報は，そのエントリーの直後に接続可能な基本ユニットの種類を表す情報である．以下に示す 3 種類のいずれかで記述する．

(1) ■：終助詞ユニット，あるいは文末
(2) other：終助詞ユニットと文末以外の基本ユニット
(3) all：すべての基本ユニット

表 4 の「テいて」のように，右接続情報によって意味コードが異なる場合は，同一表記の基本ユニットを複数のエントリーに分けて定義する．

3.2 意味コード

意味コードは，以下のような 3 つ組の情報からなる．

⟨ 意味ラベルの列 ⟩:⟨ テンスラベルの列 ⟩:⟨ 文体ラベル ⟩

意味ラベルは，ムード，モダリティ，アスペクト，否定を表し，以下のような二層構造となっている．

レベル 1 大まかな意味を示す (“推量”，“継続” など)
レベル 2 言い換え可能であることを示す (添字)

レベル 1 の意味ラベル (49 種類) を表 5 に示す． 意味ラベルは，2 階層すべて合わせると，68 種類ある．意味ラベルが “推量” である基本ユニットの例を表 6 に示す．

テンスラベルは，その文末表現の時制が過去か，あるいはそれ以外かを表す情報である．テンスラベルは以下の 3 種類で示される．

- P：タ形
- N：タ形以外
- 0：テンスを持ちえない表現

文体ラベルは，その文末表現が敬語表現であるかどうかを示す情報で，以下のように 3 種類存在する．

- p：敬語表現である
- n：敬語表現でない
- 0：敬語表現を持ちえない

意味コード中で，意味ラベルとテンスラベルは列を構成する場合がある．その理由は，4.2 節で紹介する．

3.3 基本ユニットの接続条件

文末述語文節は，基本ユニットの列として解析される．このとき，基本ユニットは，表 1 に示す順序で接続する．ただし，2 つの基本ユニットが接続可能かどうかの条件は，この順序以外に，左接続情報と右接続情報の 2 種類の情報として記述されている．例として，以下に示す 2 種類の基本ユニットの接続を考える．

- 「動いて」=形態素情報：子音動詞-タ系連用テ形，右接続情報：other
- 「テいる」=左接続情報：子音動詞-タ系連用テ形;…，基本ユニットの種類：テ形補助動詞ユニット

「動いて」の右接続情報は other であり，終助詞ユニット以外の基本ユニットは接続可能である．また,「テいる」の左接続情報の要素には,「動いて」の形態素情報である “子音動詞-タ系連用テ形” が含まれている．これにより,「動いて」と「テいる」は接続可能である．

4. 文末表現シソーラスの作成方法

以下の手順に従って文末表現シソーラスを作成した．

(1) 最小ユニットをリストアップする
(2) 最小ユニットから基本ユニットを合成する

このように文末表現シソーラスを作成した後に，基本ユニットの列が非合成的意味を持つ場合，その列を新たなエントリーとしてシソーラスに追加した (4.3 節)．

4.1 最小ユニットのリストアップ

最小ユニットに対して，以下の 2 種類のリストを作成した．

(1) 辞書形のみが収録されているリスト (辞書形リスト)

表 7 テンスラベルの合成ルール

左側	右側	合成結果
N	N	N
N	P	P
0	X(X は任意)	X
X	0	X
P	N	P.N
P	P	P.P

表 8 文体ラベルの合成ルール

左側	右側	合成結果
n	n	n
p	n	p
n	p	p
p	p	p
0	x(x は任意)	x
x	0	x

(2) 活用形が展開されているリスト (活用形展開リスト)

辞書形リストは，実際に新聞に出現する文末表現をもとに作成した．また，機能表現辞書『つつじ』に収録されている機能表現も一部参考にした．辞書形リストのエントリー数は 494 件である．

活用形展開リストは，辞書形リストを，形態素解析用辞書 JUMAN ☆ の活用分類辞書を利用し，機械的に展開させたものである．活用形展開リストのエントリー数は 6,343 件である．

4.2 基本ユニットリストの作成

最小ユニットの活用形展開リストから基本ユニットのリストを合成した．この基本ユニットのリストが文末表現シソーラスの本体である．合成手順は以下に示すとおりである．

(1) 文末述語文節の構造に従って，最小ユニットを接続させ，基本ユニットリストを作成する．

(2) 毎日新聞 15 年分 (1991-2005 年版) に出現するもののみに限定する (動詞ユニット，補助ユニットに対してのみ，この処理を行う)

手順 1 を行う際，意味コードの合成処理を行う．合成方法は，意味コードの要素である意味ラベル，テンスラベル，文体ラベルによって異なる．

意味ラベルは，原則として連結する．たとえば，意味ラベル “継続$_2$” と “可能$_1$” は，“継続$_2$.可能$_1$” のように連結する．連結されたラベル列は，それらの合成的意味を表すものとする．合成的意味を持たない場合は，意味ラベル列の書き換えを行う．たとえば,「テしまう “完了$_2$”」と「バよい “勧め$_1$”」が接続すると,「テしまえばよい “完了$_2$.勧め$_1$”」となるが,「テしまえばよい」は「バよい “勧め$_1$”」と同義であるため，“完了$_2$.勧め$_1$” を “勧め$_1$” に書き換える．

テンスラベルの合成ルールを表 7 に示す． この表に示したように，基本的には，テンスラベルは縮退させる．ただし，表の下部に示すように，連結する場合もある．

文体ラベルの合成ルールを表 8 に示す． 文体ラベルは，必ず縮退させる．

☆ http://nlp.kuee.kyoto-u.ac.jp/nl-resource/juman.html

表 9 活用形と意味コードの対応

活用形	例	意味コード
意志形	行こう	意志$_1$:N:n
推量形	優しかろう	推量$_3$:N:n
タ系推量形	行ったろう	推量$_3$:P.N:n
命令形	行け	命令$_1$:N:n
タ系連用テ形	行って	依頼$_1$:N:n
条件形	行けば，行ったら	勧め$_2$:N:n

作成された基本ユニットが “意志形”，“推量形” などのムード・モダリティを持つ活用形であった場合，基本ユニット自体の意味コードと，表 9 に示す各活用形に対する意味コードを，上記の合成方法を用いて合成する． たとえば，基本ユニット「〜つづけよう」の意味コードは，基本形「〜つづける」の意味コード “継続$_2$:N:n” と，意志形の意味コード “意志$_1$:N:n” を合成した “継続$_2$.意志$_1$:N:n” となる．

4.3 基本ユニット列の登録

シソーラスのエントリーは，原則として，基本ユニットである．ただし，基本ユニットの列が非合成的意味を持つ場合に限って，その列をエントリーとして登録することを許す．たとえば，基本ユニットの列である「テくれません/か」は，“依頼$_1$” という非合成的意味を持つので，シソーラスのエントリーとして登録する．

この作業は，終助詞ユニットを中心に調査し，人手で行った．また，以前われわれが作成した機能表現シソーラス [2] の意味ラベル書き換え規則も参考にした．

5. ま と め

われわれは，文末述語文節の構造を基本ユニットという単位を用いてモデル化し，その単位を収録単位とした文末表現シソーラスを作成した．作成したシソーラスのエントリー数は 17,169 件であり，意味コードの種類は 2,549 である．すでに，MeCab 用の juman 辞書にこのシソーラスのエントリーを追加した辞書を作成済みであり，これを用いて，与えられた文末述語文節の自動解析 (エントリーの列に分解) が可能である．

謝辞 本研究では，毎日新聞 1991-2005 年版を使用した．本研究の一部は，NTT サイバースペース研究所との共同研究として実施した．

参 考 文 献

1) 松吉 俊・佐藤 理史・宇津呂 武仁2007 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
2) 桝田 達也・佐藤 理史・藤田 篤2009 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
3) 益岡 隆志・田窪 行則2002 基礎日本語文法-改訂版-, くろしお出版