

文テンプレートによる発話文認識

望主 雅子 室井 哲也

{masako,muroi}@ic.rdc.ricoh.co.jp

(株) リコー 情報通信研究所

スポッティングを制御するための記述方式として文テンプレートという枠組を提案する。文テンプレートは発話文中からスポッティングする語の位置関係を記述したもので、話しことば特有の省略や語順の変化にも対応できる。また、話しことばにはどのような制約が有効であるかを調節しながらさぐるツールの意味を持つ。文テンプレートを使うことで、スポッティングに頻出するわきだしのエラーを抑え、認識率をあげることができ、また、対話処理や意味処理との結合や特定タスクの記述がしやすくなる。本稿では、この文テンプレートの記述方式と実際のタスクでの記述、実験の結果を報告する。

Speech Recognition by Using Sentence Templates

MOTINUSI Masako, MUROI Tetsuya

Information & Communication R&D Center, RICOH Co., Ltd.

To control word spotting, we have developed a method that uses "sentence templates." The sentence templates describe locations of target words, which may appear in varying positions in a spoken sentence. The templates are also useful to examine and adjust the constraint and optimum unit to identify spoken words. By using the templates, we reduced the number of false alarms and raised the overall recognition rate. Because of their flexible syntax, the templates are also capable of incorporating information that is necessary to process dialogues, analyze semantic features, and describe different recognition tasks. In this paper, we discuss the syntax of the templates and the results of a word spotting task we performed using the templates.

1 はじめに

音声認識の認識制御や認識結果の絞り込みに言語的な制約を使うことが認識率を向上させる有効な手段となっている。言語的な制約にも主に二通りある。一つはいわゆる文法というものを制約とするものである。従来は特に計算機にのせやすい、書き換え規則の形で記述するものが多かった。これらの文法を認識制御に使った場合の問題として、以下がある。

- 非文の生成

文法の記述が十分でない場合に、過剰な生成がおこり、文として誤っている表現を生成し、制約として働かない。

- いい淀み、不要語の扱い

「えー」「あー」などの不要語¹やいい淀みも、文法の一タームとして記述しておく必要がある。これらの現象は今のところどの箇所にも出現するなどが不明確であるので、規則化しにくい。

- 意味処理との整合性

文法はそもそも構文的な性質を記述するものである。このため、意味的な制約は、文脈自由文法では規則の一分岐として吸収して記述するなど、意味と構文を混在させた規則で扱うことがあり、その記述が繁雑になることがあった。

書きことばをベースに記述したものが多い。書き言葉では「文」という単位が存在し、その単位内での整合性、規則化をはかったものが文法である。この点からして書きことばベースの文法は話しことばに向かないのではないのか。

もう一つの方法として意味主導でスポッティングをする方式がある。これはタスク上必要なキーワードだけを発話中から抽出するものであり、以下のような特徴、問題点がある。

- 不要語、語順の変化に対応できる

いい淀み、不要語を含んだ発話を、それらを見捨てるという意味で、扱うことができる。また、語順が倒置など、書き言葉の文法から逸脱している場合も、語順などの制約を用いないので、扱うことができる。

¹ これらの語が不要であるかどうかは、別として、従来不要語と呼ばれてきたので、ここではこのように呼ぶ

- 湧きだしエラーが多い

発話中から純粋に音として似ているものを抽出するので、湧きだしのエラーが多くなる。

認識候補を抽出後に時間的な重なり、文法などでふるい落とすという方法がある。

我々は後者の意味主導、スポッティングベースで話文の認識を行なっている。本稿では、スポッティングを制御する一手法として文テンプレートという記述方式を提案し、記述の枠組、認識制御の実験と結果について報告する。

2 文テンプレート

2.1 文テンプレートの狙い

スポッティングで問題になるのは湧きだしのエラーである。キーワードだけをスポッティングするので、

例えば、

会議がしたいんですけど 抽出キーワード：「会議」「明日」

のように、発話中の助詞、助動詞部分から「明日」というキーワードを抽出することがあった。

また、

3時からします 抽出キーワード：「3時半」

のように、助動詞部分が別の語として認識されることがあった。

こういった湧きだしのエラーを減らすには、そこには別の語が話されているということがわかるということ、つまり、発話中からできるだけわからないところを少なくするといったことが必要だと考える。

現段階では、発話の現象で「わかるところ」つまり、規則として書けるものがすべてわかっているわけではない。「わからないところを少なくする」ということは、現象のすべてを説明できなくても、あるタスク、対話で少なくとも有効であるものを制約としてとりいれ、発話現象全体から少しずつわからないところをなくしていくということの意味する。文テンプレートは、発話を扱う上で制約として有効であるものを徐々に少しずつとりいれ、記述し、実験して確認していくためのツールの意味を持つ。現在、制約として想定しているものには、以下のものがある。

- キーワードと共起する語の認識

キーワード以外の語も積極的に認識することで、発話部分からわからないところを減らす。助詞や動詞や助動詞など、キーワードとの共起が考えられるものを認識する。

- 語順の制約

発話中にこういった意味の語がこういった語順で現れるかを記述するようにし、湧きだしエラーや誤認識を減らすことを狙う。

- 認識単位の調節

発話中からキーワードやキーワード以外の語を認識するとき、どのような単位で認識するかということが重要である。「文節」「単語」「連文節」などの構文的な単位が合致するとは限らないが、ここでは、記述の単位を単語、文節とし、認識時にはそれらを接続可能にして認識のしやすさに応じて調節できるようにする。

- 省略の有無

発話現象では書き言葉では省略されない語が省略される。どのような語が省略される傾向にあるかを調べ、省略されるかどうかを記述できるようにする。

- 発話の局面と語

発話の局面では少なくとも何か語を一つ発すると考える²。この局面が一つの意図を伝える単位となり、少なくとも一つ、意図と対応づけられる語が発せられると考える。この意図と対応する語を発話の局面での中心となる語とし、認識制御に用いる。一局面で扱う範囲は人間が一発話で話せる、単文程度の長さとする。

- 文法と文テンプレート

発話中の単語内、文節内などの狭い範囲は比較的书きことばに近いきつい制約が有効で、単語間や文節間はその制約がゆるいと考える。単語内、文節内など狭い範囲は書き換え規則の形で記述し、単語間、文節間を文テンプレートで記述する。

2.2 文テンプレートの記述

文テンプレートはある発話の局面で、認識したい語とその時間的な順番を記述したものである。以下が記述形式である。

2.2.1 記述形式

文テンプレート = <対話状態タグ><テンプレート>

² 無言というもの、対話では相手への非同意、待ちなどが考えられるが、現在は扱っていない

対話状態タグ = 登録 | 応答 | …

テンプレート = <記述項> …

記述項 = (&)&<指定項>(:[品詞])(:[頻度])

記述項 = (+)(^)(&)(*)(%)<指定項>(:[品詞])(:[頻度])

指定項 = *|\$[意味名]| [語表記]

注)「()」はテンプレート上で省略可能、「|」は「または」の意味、「<>」は以下に定義あり、「[]」は一般にいわれている意味を表す。

2.2.2 フラグの意味

表 1に記述のフラグの意味を示す。

表 1: フラグと意味

@	発話の局面でキーとなる語
&	共起制約の中心になる語
^	「&」の語と共起する語
*	ワイルドカード
:	制約(品詞や頻度)
+	前の項とつなげて一単位で認識する
%	省略可能な語
*	繰り返し可能
#	タスク上のキーワードとして抽出する

例えば以下の表現を記述できる。

- 共起表現の記述

格要素などは構文的には名詞句という同じものが1発話中に連続し、意味的にはある句と共起し、タスクのスロットをうめていくような性質をもつ。これらを、名詞句で、ある語に共起するものという記述ができるようにした。以下の共起スロットは、「会議」を意味する語を共起の中心とし、これと日付、時間、場所を意味する語が共起するという意味である。この場合の格要素の助詞は省略可能として扱う。

記述	*:*:名詞 &\$会議
共起スロット	\$会議(\$日付,\$時間,\$場所)
文例	明日会議 3時から明日打ち合わせをします

- 一単位で認識する表現の記述

一単位で認識した方が認識しやすい表現を「+」

でつないで、一単位で認識する。「+」で接続された語について、接続検定を行ない、接続可能なものを生成する。

記述	&\$日付 +%の +%時間 +%から
文例	明日の3時20分 来週の水曜日、2時から

● 制約の記述

発話中に現れる語について、主に意味を第1の制約とし、構文的な制約やタスク上の制約（頻度など）を各語に付加する形にし、意味と構文的な制約を切り分ける。

記述	\$会議:名詞:hindo>1 +\$断定
文例	会議です ミーティングです

● 局面でキーとなる語

発話の局面で中心となり、省略されることが少ない語や、他の語と識別され、認識しやすい語にマークし、最初に認識する。これを共起スロットと組み合わせることで、キーとなる語と関係する語が次々と認識されるようにする。例えば「〇〇会議室」という長めで認識しやすい語を中心にした記述を以下にします。

記述	*\$:名詞 &\$会議室 *\$:名詞
スロット	\$会議室(\$日付,\$時間,\$会議)
文例	明日、C2会議室で会議です A2会議室にて会議があります

● キーワードとして抽出する語

タスク達成上取り出したい語にマークする。テンプレートではタスクに直接必要な語と誤認識を防ぐための語の両方を認識する。

記述	#\$日付 +\$です
文例	明日です
キーワード	\$日付:明日

2.2.3 文テンプレートの記述例

会議室の予約の対話での第1発話でのテンプレートの例と認識可能な文の例を以下に示す。

登録 *%\$:名詞 @\$会議:名詞\$サ変 +%\$を
*%\$:名詞 %\$する +%\$意志

えー、明日の3時から会議をしたいんですけど
ミーティング、10時に、第3会議室で
打ち合せ、明日します

3 認識システム

音声認識システムの構成を図1に示す。特定話者方式である。

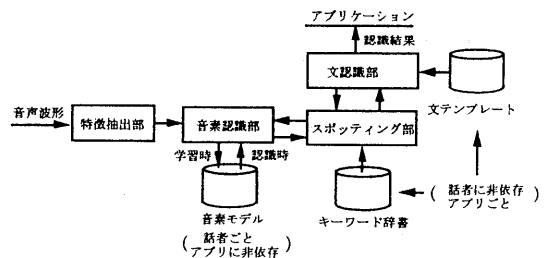


図1: 音声認識システムの構成図

3.1 認識制御

実際の制御は対話やタスクの状態から対応する発話状態タグをもつ文テンプレートに基づいて行なう。各テンプレートの「@」が付与された項を認識し、スコアの最も高いものを持つテンプレートについて、それと共起する語、前後に位置する語の順に認識し、すべての項が認識されれば、そのテンプレートが発話と合致したとする。これをそのタグをもつテンプレートすべてについて行ない、もっともスコアの高いものを認識結果とする。「@」が付与されていない場合は、先頭の項から処理を行なう。

3.2 規則（キーワード辞書）の自動生成

テンプレートの各項が認識時のスポッティング部への命令となる。テンプレートの項の間はテンプレートで制御するが、項の中は一単位として、あらかじめ書き換え規則の形に展開しておき、スポッティング部に渡しておく。単語、文節間は人間がタスクにあわせて記述し、文節内程度の構文的な制約はシステム側で面倒をみるという方針をとっている。この書き換え規則を一部自動生成している。

生成はテンプレートの記述にもとづいて単語辞書、共起スロットデータを用いて行なう。「+」の記号がある場合には「+」で結ばれた項どうしについて接続チェックを行ない、文法を生成する。名詞や助詞、名詞どうしなど活用しない語については接続表を用いて接続検定をする。活用する語については接続情報を前

の語への接続要求として付与・活用形化し、接続していく。具体的な処理は次節で述べる。

3.2.1 接続制限のある語の規則生成

接続制限のある語を前の語への接続要求ととらえなおし、要求する形を辞書に記述しておく。規則生成時に接続制限を辞書から検索し、前接する単語を接続しうる形に変形する。変形の際、要求した形を要求された側、要求した側の両方に残した記号を中間の記号として生成する。テンプレートの記述が「\$話す+\$受身」の場合を説明する。

- 最上位の項を生成する \$話す+\$受身
- 辞書から品詞と接続制約を検索する

意味	品詞	接続制約
\$受身	助動れる	未
	助動られる	1用
意味	表記	品詞 (品詞の意味)
\$話す	話す	動詞 5 さ (さ行 5 段活用)
	伝える	動詞 1 (一段動詞)

- 語尾テーブルから活用形を調べ、接続要求の形があるか調べる。

品詞	活用形	語尾
動詞 5 さ	未	さ
動詞 1	1用	(なし)

- 接続要求の形のある語に要求活用形を付加した記号を生成し、最上位の記号と結び付ける。「:」は接続の制約を受けている意味、「<」は接続の制約を与えている意味である。

左辺	右辺
\$話す+\$受身	\$話す:未 \$受身<未
	\$話す:1用 \$受身<1用

- 語表記とよみを生成し、先の記号と結び付ける。

左辺	右辺
\$話す:未	話さ
\$受身<未	れる
\$話す:1用	伝え
\$受身<1用	られる
話さ	はなさ
伝え	つたえ

テンプレートが「\$話す+%%\$受身+%%\$意志」の場合の展開例を表 2 にあげる (表記レベルまで)。

表 2: テンプレート「\$話す+%%\$受身+%%\$意志」の例

左辺	右辺列
\$話す+%%\$受身+%%\$意志	\$話す+\$受身+\$意志
	\$話す+\$受身
	\$話す+\$意志
	\$話す
\$話す+\$受身+\$意志	\$話す:未 \$受身<未:用 \$意志<用
	\$話す:1用 \$受身<1用:用 \$意志<用
\$話す+\$受身	\$話す:未 \$受身<未
	\$話す:1用 \$受身<1用
\$話す+\$意志	\$話す:用 \$意志<用

左辺	右辺列	左辺	右辺列
\$話す:未	話さ	\$受身<未:用	れ
\$話す:1用	伝え	\$受身<1用:用	られ
\$話す:用	話し	\$受身<未	れる
	伝え	\$受身<1用	られる
\$意志<用	たい	\$話す	話す
	ます		伝える

4 実験

4.1 作例による実験と結果

会議室予約の対話の第 1 発話を想定し、格要素とその語順、ポーズや不要語の有無を変えた作例による小規模な実験を行なった。テンプレートを使ったものと使わない単純なスポッティングだけのものとを比較した。いい淀み、いい誤りを含まない、比較的に明瞭な発話である。認識対象語は 68 文中のキーワード 168 語で、予約に関する 10 単語と日付 (40 通り)、時間 (84 通り) である。認識はテンプレートの記述を認識命令とし、認識結果のスコア第 1 位のものだけをとっている。現象の内訳を表 3 に示す。実験の結果を表 4、

表 3: 認識対象文の言語現象の内訳

キーワード数	1	2	3	4
数	8	32	17	11
格要素	0	1	2	3
数	7	29	17	15
現象	倒置	省略	不要語	
数	5	11	11	

表 5 に示す³。テンプレートで誤認識を減らすことの

³ 検出率は認識対象語 (B) 中で抽出できたキーワード (A) の割合、適合率は抽出されたキーワード (C) 中の正解 (A) の割合を意味する

表 4: キーワードの検出率と適合率 (括弧内は語数)

テンプレート	検出率 (A/B)	適合率 (A/C)
使用	97.0(163/168)	97.0(163/168)
未使用	94.0(158/168)	85.8(158/184)

表 5: 誤認識の種類と数

テンプレート	湧きだし	もれ	誤り
使用	3	3	2
未使用	20	4	6

できたものを表 6 に示す。テンプレートを使用して

表 6: テンプレートで低減された誤認識

現象	現象数
動詞、形式動詞、助詞の認識	9
助詞付きの認識	7
テンプレートのスコアの総計	2

も誤認識になったものは、不要語部分の湧きだし (3 例: 「えー会議室」が「A 2 会議室」) や数詞部分の誤り (5 例: 「C 1 会議室」が「1 時」) だった。

4.2 キーワード表示による発話での実験

発話時にキーワードだけを表示し、キーワードを含む文をその場で考えて発話したものを認識対象とした。発話数は 44 である。先の比較的明瞭な発話に比べ、いい淀み、不明瞭な発声、速度のばらつきがある。認識結果を表 7 に示す。明瞭な発声に比べて検出率も

表 7: キーワード表示による発話の検出率と適合率

テンプレート	検出率 (A/B)	適合率 (A/C)
使用	77.6(83/107)	86.5(83/96)
未使用	72.0(77/107)	82.8(77/93)

適合率もかなり落ちる。

考えながらゆっくり話した場合には、認識が可能であったが、逆に速く話した場合、キーワード数の多い発話の先頭の語が認識できなかった。また、いい淀みがある場合、発話全体が不明瞭で、特に速度が速い場合は音素自体が脱落して認識できなかった。

テンプレートを用いることで、似た語のある箇所での湧きだしや誤認識を減らせたが、語を時間的に離れたところや短い長さで抽出する誤りは減らせなかつ

た。また、中心となる語が認識できずに全体が認識できない場合があった。

4.3 考察

文テンプレートについて以下の有効性を確認した。

- 動詞、助動詞部分の認識が湧きだしエラーをおさえた。長さの長い述部にも効果的である。これらは他の語の位置の予測の手がかりにもなる。
- 付属語をつけた認識が日付や時間などの数詞の短い語の抽出に有効だった。他の語と識別できることや長くなることで認識しやすくなる。
- 共起の中心となる語句を認識することが有効だった。誤認識や省略の多い数詞表現 (日付、時間) より先に、認識しやすい語を共起の中心にすえて、認識することで効果的があった。

しかし、以下のような問題点がある。

- 構文的な制御の限界
記述の単位が構文単位なので、時間など音響的性質による調節ができない (特に後者の実験で)。普通に発声したときとの時間のずれや発話単位と概念の関係 (例: 20 日の 3 時から) を考慮する必要がある。
- 構文的制限の不備
省略や共起のしやすさが語や共起状況によって異なり、すべてを一律に記述する方法に問題がある。より自然な発話では共起の効く範囲は狭いと予想され、その範囲を調べる必要がある。

5 終わりに

文テンプレートという記述方式を提案し、特定タスクの発話文で湧きだしエラーに効果のあることを確認した。より自然な発話に対してはまだ制約としては不足であることもわかった。実際に機械を使ったときの人間の発話について調査し、記述可能な制約の検討を進めていく。

参考文献

- [1] 望主、室井「文テンプレートによる文認識方法」情報処理学会第 47 回全国大会, 5V-2(1993).