

意味構造テンプレートを用いた口語文の意味解析手法の提案

菊池久一，樽松理樹，藤田ハミド
岩手県立大学ソフトウェア情報学研究科

g231y005@edu.soft.iwate-pu.ac.jp, {kure, issam}@soft.iwate-pu.ac.jp

概要

本稿では、意味構造テンプレートを用いた口語文の意味解析手法を提案する。本手法では、最初に発話区間に対する音声認識結果を句単位に変換する。次に、その中に現れる動詞句に注目し、文の抽出を行い、動詞(句)を基に選択した意味構造テンプレートと照合することで意味構造候補を構築する。最後に、構築した意味構造候補を評価し、最適なものを意味構造として抽出する。本手法により、通常の構文を逸脱している発話に対し、文単位での処理が行なえることから、音声理解の向上が期待できる。本手法はまだ実装されていないため、実装、実験評価を行ない、その有効性を評価する必要がある。

Semantic Structure Representation Using Template For Natural Spoken Language

Hisakazu KIKUCHI, Masaki KUREMATSU and Hamid FUJITA
Graduate School of Software and Information Science, Iwate Prefectural University
g231y005@edu.soft.iwate-pu.ac.jp, {kure, issam}@soft.iwate-pu.ac.jp

Abstract

In this paper, we propose the semantic-analysis method of the utterance using semantic structure templates. In first, the method changes the speech recognition result of the utterance into phrase structure. Next, the method extracts some sentence candidates paying attention to the verb which appears in speech recognition results and builds some semantic structures from them using semantic structure templates. Finally, the method extractes the optimal semantic structure evaluating slot values in semantic structures. We don't implement this method. It is necessary to implement it and evaluate the validity of it.

1 はじめに

情報化の進展により、ユーザとの親和性の高いことからその重要性が増している音声インタフェースは、大きく音声認識部と音声理解部から構成される。音声認識部を構築する技術は急速に発展しており、ユーザの発話から文への変換が可能となり、PC上においても大語彙連続音声認識ソフトウェアが利用可能となってきた [1]。一方、音声理解部としては、その発展が著しい自然言語処理技術の適用が考えられるが、処理対象の特性の差から、まだ多くの課題がある。そのような課題の一つとして、人間の話す単位と機械が処理する単位との不一致が挙げられる [2]。コ

ンピュータによる意味解析を考えた場合、文を機械の処理単位とするのが適切であり、多くの自然言語処理技術が、文単位の処理を行っている。しかし、発話における文の認識は困難であり、単語を処理単位とするか、ある一定時間以上発話の無い区間(以後、無音区間)に挟まれた部分(以後、発話区間)を処理単位として扱うのが現状である。これらの方法は、意味解析の処理単位としては適しておらず、誤った意味をおこすこととなる。そのため発話内容の意味をより正しく理解するために、発話という言語現象に適した手法を提案する必要がある。

本稿では、このような課題に対し、ユーザの自発音声の音声認識結果に対し、出現する単語から

句、さらに文を推定し、動詞に着目した意味構造テンプレートを利用することにより、文単位での意味構造を抽出する手法を提案する。本手法により発話区間内の文章の文単位での処理が可能となることから、意味解析の精度をあげることが期待できる。また、従来の構文規則を直接適用するわけではないため、構文を逸脱した発話文に対しても、処理が可能となることが期待できる。

以下、第2章では本手法について説明する。さらに第3章で解析例を提示し、第4章において本手法の考察、課題を示す。

2 意味構造テンプレートを用いた解析手法

2.1 概要

本手法の概念図および処理の流れを図1を示す。本手法は、音声認識部の認識結果から文を推定し、後述の意味構造テンプレートを用いて、文単位の意味構造を出力する。本手法の処理サイクルは、図中に示す Step.1 から 6 までの 6 つのステップから構成されており、未処理(未使用)の音声認識結果がある間は、処理サイクルを繰り返す。以下、各ステップについて説明をする。

Step.1 : 音声認識結果の取得 音声認識部の出力は、無音区間を含む仮名漢字混じり文として、バッファに保管されている。本手法は初めに、バッファから音声認識結果を取得する。音声認識結果を新たに獲得した場合、または未使用の音声認識結果がある場合は、次の処理へと進む。無い場合は、再度音声認識結果の取得を行なう。

Step.2 : 形態素解析 Step.1 で取得した仮名漢字混じり文に対し、形態素解析を行なう。この時、未使用の仮名漢字混じり文があった場合は、新規に取得したものの前に追加し、追加した文に対して形態素解析を行なう。また、無音区間に対しては形態素解析を行わず、そのまま無音区間として保持する。

Step.3 : 句候補列リスト構築 Step.2 で得た形態素列を、句構築ルールに基づいて句候補列リストに変換する。ここで句候補とは、一つ以上の隣接する形態素列を統合したものを意味する。この処理により形態素列は、句候補列に変換される

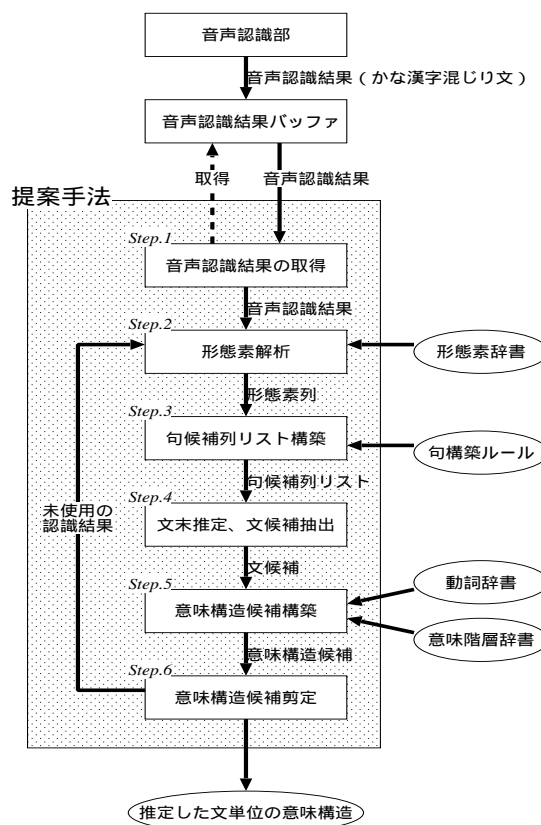


図 1: 概念図・処理の流れ

こととなる。また、未知語が存在する場合は未知語句に、無音区間は無音句に変換する。さらに形態素列に対し、考えられるすべてのルール適用を行い、複数の句候補列(句候補列リスト)を作成する。

Step.4 : 文末推定、文候補作成 次に、生成された各句候補列中の動詞句に注目し、文末推定を行い、現在の仮名漢字混じり文(音声認識結果)の最初の一文を抽出する。この際に、文末候補が複数考えられるが、それらすべての場合の文候補を抽出する。

Step.5 : 意味構造構築 抽出した文候補に対して、文に含まれる動詞をキーとして、動詞辞書から意味構造テンプレートを取り出し、そのテンプレートに値を埋めることで、意味構造を構築する。この処理では、構築しうるすべてのパターンを抽出する。

Step.6 : 意味構造候補剪定 Step.5 で作成したすべての意味構造候補に対し、評価値を計算し、絞り込みを行なう。絞り込んだ意味構造を構築する

部分を文として確定する。この結果、確定された文に利用されていない音声認識結果の部分は、未使用の音声認識結果として、次の処理サイクルで扱う。

2.2 各ステップの詳細

本章では、各ステップについて説明を加える。

2.2.1 音声認識結果の取得

例えば IBM 社製の ViaVoice[3] のように、現在市販されている音声認識エンジンの出力では、ユーザの発話を仮名漢字混じり文に変換するものが多くある。本手法においても、発話を仮名漢字混じり文に変換したものを音声認識結果と想定し、処理対象とする。また、発話区間全体の音声認識処理後に音声理解処理を行うと、待機時間が発生し、処理が遅くなることから、音声認識部と本手法とは同期せず、バッファを介して情報のやり取りを行う。すなわち、音声認識部は、音声認識結果をバッファに出力し、本手法は、バッファ内にある音声認識結果を所得し、以後の処理を行う。

2.2.2 形態素解析

音声認識結果である仮名漢字混じり文に対し、形態素解析を行う。形態素解析については、既存の形態素解析ツールを用い、一つの解析結果を得る。解析結果は、(形態素(音声認識結果内の表層形), 形態素(基本形), 品詞) という3つ組のリストで示される。なお、誤認識と思われるものや未知語については、削除などを行わず、そのまま次の処理へと渡す。

2.2.3 句候補列リスト構築

このステップでは、句構築ルールに基づいて、形態素列を、(句の種類, 句の内容, 代表形態素) から構成される句候補の列に変換する。ここで、代表形態素とは、句に含まれ句を端的に表わす形態素である。句構築ルールの適用は、適用できるルールがある限り続ける。また、複数のルールが適用できる場合についても、すべてのパターンの生成を行う。

句構築ルールは、基本的に句構造文法の考え方

に基づいている。ただし、文にまでまとめず、隣接する形態素を句にまとめることを試みる。具体的には、1) 隣接する動詞と助動詞を動詞句にまとめる、2) 隣接する名詞をまとめて複合名詞とする、3) 隣接する形容詞をまとめて形容詞句とする、4) 隣接する副詞をまとめて副詞句とする、5) 隣接する形容詞句と名詞句をまとめて名詞句とする、6) 助詞により結合される名詞や名詞句をまとめて名詞句とする、7) 名詞句に隣接する助詞をまとめて名詞句とする、8) 隣接する無音区間を無音句とする、などの規則を記述する。

また句構築ルールは、句候補を構成する形態素(や句候補)の中から句候補を端的に表わす形態素を代表形態素として選択する操作も含み、句構築ルールを適用する際に、代表形態素の抽出も行う。また無音区間の句である無音句の代表形態素としては、無音区間の時間を用いる。

この処理によって形態素列は、1つ以上の句候補列(句候補列リスト)に変換される。

2.2.4 文末推定、文候補抽出

このステップでは、各句候補列に対し、次に示すアルゴリズムにそって文末を推定し、仮名漢字混じり文を構成する最初の文候補を抽出する。この処理では、処理対象内の最初の文のみの抽出を行ない、仮名漢字混じり文全体を文に分割するわけではない。なお、文末推定は、一文には一つの動詞句が含まれているという仮定に基づいている。また無音区間の判断に使われる閾値は、ユーザが設定する。

- 1.1 i ; 文候補リスト;
2. if (i 番目の句候補が、閾値以上の無音区間) then { 処理 4 へ } else if (i 番目の句候補が動詞句) then { 処理 5 へ } else { $i+1$ i }
3. if ($i > n$) then { 処理 9 へ } else { 処理 2 へ }
ここで n は句候補列の句候補数
4. 無音区間の直前までの句候補列 (1 番目から $i-1$ 番目までの句候補列) を文候補リストに追加し、処理 8 へ
5. i x ; $i+1$ i
6. i 番目の句候補から順にチェックを再開する。
if (i 番目の句候補が閾値以上の無音区間) then { 処理 4 へ } else if (i 番目の句候補が動詞句) then {
 x から i 番目の間に文末があるとし、それらの組み合わせ、すなわち、1 番目の句候補から x 番目の句候補、1 番目の句候補から $x+1$ 番目の句候補

補, ..., 1 番目の句候補から i-1 番目の句候補までのすべての組み合わせを文候補リストに追加し、処理 8 へ

```
} else if (i 番目の句候補が接続詞句) then {  
  i-1 番目 (接続詞の直前) に文末があると推定し、1 番目の句候補から i-1 番目の句候補までを文候補リストに追加し、処理 8 へ
```

```
} else {i+1 i}
```

7. if (i>n) then { 処理 9 へ } else { 処理 6 へ }

8. 文候補リストのうち、動詞句を含まない文候補を削除し、残りのものを文候補として出力し、処理を終了する。

9. 文候補は無しとして、処理を終了する。

2.2.5 意味構造候補構築

このステップでは、各文候補に対し、意味構造テンプレートを用いて、意味構造を構築する。ここでは意味構造を構築する際に利用する意味構造テンプレートを説明したあと、意味構造構築手順を説明する。

意味構造テンプレート 意味構造テンプレートは Fillmore が提唱した格文法 [4] を基本としている。本テンプレートは、動詞名スロットと基本スロット群から構成される。基本スロット群は、動作主格スロット、経験者格スロット、道具格スロット、対象格スロット、源泉格スロット、目標格スロット、時間格スロットの 8 つのスロットから構成される。各スロットは、スロット名と値を持つ。動詞名スロットの値には、動詞の基本形が与えられ、検索時に利用される。基本スロットのスロット名は、動詞に対する意味的役割を示しており、スロット値は、それに当てはまる概念が 0 個以上与えられている。スロット値が 0 個の場合は、そのスロットに当てはまる概念が無いことを意味する。

意味構造候補構築手順 各文候補から、前述の意味構造テンプレートを用い、次の処理手順にしたがって、意味構造に変換する。ここで作成された意味構造を、次ステップへ渡す。

1. 文候補中の動詞句を取出す。
2. 動詞辞書と動詞句を照合し、動詞名スロット値と代表形態素が一致する意味構造テンプレートを取出す。
3. 動詞句の助詞や活用に注目し、時制スロット、および状態スロットを付加する。状態スロ

ットの値には、伝聞、可能、といった情報が与えられる。

また未知語を保存するための未知スロットを追加する。

4. テンプレートの持つ各スロットの値と、文候補を構成する各句候補の代表形態素とを比較する。代表形態素が、意味階層辞書において、スロット値のどれか一つと同じ概念または下位概念である場合、その句候補をそのスロットに割当てて。この時、各スロットに割当てられる句候補は、複数許し、また、一つの句候補を複数のスロットに割当てることができる。
5. 基本スロットのいずれにも割当てられない句候補については、未知スロットに割り振る。

2.2.6 意味構造候補剪定

最後に、作成された意味構造候補を以下の基準によって剪定する。

1. 各意味構造に対し、次の数式に従い、評価値を計算する。
評価値 = 基本スロットの評価値の合計 - 未知スロットの評価値
各スロットの評価値 = 含まれる句の評価値
句の評価値 = 基本スロットの種類数 / その句が含まれるスロットの数
2. 1 で計算した評価値が最も高い意味構造を抽出する。
3. 抽出された意味構造が複数ある場合、文を構成する形態素数が最も少ないものを意味構造として抽出する。

利用する形態素数が少ないほど、次の音声認識結果への処理時への利用が期待でき、ユーザの与えた情報の欠落を防ぐことができると考える。また以上の方法で絞りこんだとしても、意味構造としては複数のものが選択される可能性はある。これらについては文脈レベルでのあいまい性を含むものとし、すべて取出す。

最後に後処理として、剪定された意味構造(文)を構成しない音声認識結果を、未使用の音声認識結果として保存する。これは次の処理サイクルの際に利用する。

表層形	基本形	品詞
明日	明日	時相名詞
無音区間(XXms)		
3	3	数詞
時	時	名詞性名詞助数
から	から	格助詞
会議	会議	サ変名詞
が	が	格助詞
ある	ある	動詞
無音区間(YYms)		
場所	場所	普通名詞
は	は	副助詞
無音区間(XXms)		
会議	会議	サ変名詞
室	室	普通名詞
無音区間		
駅ビル	駅ビル	普通名詞
の	の	格助詞
無音区間(YYms)		

図 2: 形態素解析の結果

3 解析手法の処理例

3.1 処理例

本章では、2章に示した枠組みにそった処理例を示す。この例は、実際に実行した例ではなく、想定例である。

3.1.1 音声認識結果の取得

音声認識結果の取得を行なった結果、次に示す仮名漢字混じり文を得たものとする。これが、処理を行う対象となる。

「明日、3時から会議がある。場所は、会議室、駅ビルの。」

ここで読点は、ユーザが与えた閾値より短い無音区間、句点は閾値より長い無音区間を示す。

3.1.2 形態素解析

音声認識結果に対する形態素解析の結果を図2に示す。この結果は、京都大学の長尾らが開発したJUMAN[5]による解析結果に処理を加えたものである。図中のXX、YYは無音区間の時間を示しており、XXは閾値よりも短く、YYは閾値よりも長い。

{1,[(名詞句, 明日, 明日), (無音句, 無音区間, XXms), (名詞句, 3時から, 時), (名詞句, 会議が, 会議), (動詞句, ある, ある), (無音句, 無音区間, YYms), (名詞句, 場所は, 場所), (無音句, 無音区間, XXms), (名詞句, 会議室, 室), (名詞句, 駅ビルの, 駅ビル), (無音句, 無音区間, YYms)]}

図 3: 構築した句候補列リスト

スロット名	スロット値
動詞名	(動詞句, ある, ある)
時制	現在
動作主	(名詞句, 会議が, 会議)
対象	(名詞句, 会議が, 会議)
時間	(名詞句, 明日, 明日) (名詞句, 3時から, 時)

図 4: 構築した意味構造候補

3.1.3 句候補列リスト構築

句構築ルールを用い、形態素列を句候補列へと変換した結果を、図3に示す。図中において各句候補は、(句の種類, 句の内容, 代表形態素)の形で記述している。

3.1.4 文末推定、文候補作成

求めた句候補列リストから、2.2.4章で提示したアルゴリズムに従って文末を推定する。結果、動詞句「ある」と2つめの無音句との間に文末を推定し、「明日」「3時から」「会議が」「ある」を、文候補として抽出する。この時点で、「場所は」以降の部分は、現在のサイクルの処理対象外となる。

3.1.5 意味構造候補構築

求めた文候補に対し、意味構造テンプレートを利用して、図4に示す意味構造を構築する(なおスロット値が無いスロットは省略している)。

3.1.6 意味構造候補剪定

現在のサイクルでは、一つしか意味構造候補がないため、これが意味構造となり、本手法の出力となる。この意味構造を構築する部分が音声認識の結果から削除される。

参考までに、この意味構造の評価値を示す。

評価値 = 動作主格の評価値 + 対象格の評価値 + 時間格の評価値 = $(8/2) + (8/2) + (8+8) = 24$
よって、評価値は 24 となる。

以上で 1 サイクルが終了したため、再び、音声認識結果の取得を行なう。この時点では、「場所は、会議室、駅ビルの。」が未使用の音声認識結果として、次のサイクルの処理対象となる。

4 考察

本稿で提案した手法は、現時点においていくつかの問題点が存在する。本章では、代表的な問題点と今後の課題を述べる。

4.1 問題点

処理対象となる発話の制限： 現在の手法は、文には必ず動詞句を一つ含むことを前提としている。しかし、人の発話の中には、動詞句を含まないものも存在する。現在の手法では、そのような発話を扱えない。また、複数の動詞句を含む複文に対しても、同様に対応が困難である。これらの点は、ユーザの自発音声の認識に対する本手法の弱点となる。

計算量の問題： 現在の手法は、多義が発生する場合についてすべて計算し、候補として保持している。また、同じ発話区間内に対し、同じ処理を複数回試みる。結果、本手法は非効率であり、処理時間がかかると予想される。

手法の評価： 現在、手法に対しては提案を行なったのみであり、実装、評価実験を行なっておらず、どの程度有用かが十分に評価されていない。言わば、静的解析を行ったのみである。

4.2 課題

以上の問題点を踏まえ、以下の課題に取り組む。

実装および評価実験、実験結果のフィードバック：

本手法は、上記のような問題点があるが、第一に実装、評価実験を行なう。実装を進めることにより、本手法の静的解析が進み、新たな問題点の洗い出し、現在の問題点のより詳しい分析ができると考えられる。それらの解析結果を基に手法の改善策を検討する。

また実装時には、できるだけ既存のリソースを利用する。例えば、形態素解析部については、JUMAN などの形態素解析ツールを、意味階層辞書としては、日本語語彙体系 [6] などの利用を検討する。

実験においては、PC 上で利用可能な音声認識システムを利用し、数名の被験者による音声理解実験を試みる。この際、被験者の意図をシステムが抽出できたかを被験者に評価してもらい、その結果を解析し、フィードバックすることで、本手法の精度の向上を図る。

他の手法との比較評価： 実装・評価実験とともに、本手法の有用性の評価するため、Dialogos[7] などのフレームを利用する他の類似手法との比較を行う。

5 終わりに

本稿では、音声理解システムを構築する上で必要となる、音声理解の方法として、発話区間内に現れる句や無音区間から文末を推定し、動詞に注目した意味構造テンプレートを利用し意味構造を抽出する手法を提案した。本稿では、本手法について説明し、想定例にてその処理の流れをしめし、現状で判明している問題点・課題を提示した。今後、手法の検証を進め、問題点を解決するとともに、実装、評価実験を行ない、本手法の改善、有効性の検証を行う。

参考文献

- [1] 河原達也: "道しるべ: ここまできた音声認識技術", 情報処理学会論文誌, Vol.41, No.4, (2000)
- [2] 竹澤寿幸: "道しるべ: いまこそ話し言葉処理技術の研究を", 情報処理学会論文誌, Vol.42 No.2, (2001)
- [3] 日本 IBM: "ボイスランド", <http://www-6.ibm.com/jp/voiceland/index.html>
- [4] C.J.Fillmore, 田中春美, 船城道雄訳: "格文法の原理", 三省堂, (1980)
- [5] "JUMAN", <http://www-nagao.kuee.kyoto-u.ac.jp/nl-resource/>
- [6] 池田悟, 宮崎正弘他編: "日本語語彙体系", 岩波書店, (1997)
- [7] R.Billi, G.Castagneri and M.Danieli: "Field trial evaluation of two different information inquiring systems", Speech Commun. Vol.23, (1997)