

概念素に基づく音声理解への統計的言語制約の導入

永井 明人、石川 泰

{nagai, yasushi}@media.isl.melco.co.jp

三菱電機株式会社 情報技術総合研究所

〒 247-0056 神奈川県鎌倉市大船 5-1-1

我々は自由発話理解のために概念素に基づく意味理解方式を提案し、実用化検討を進めている。本方式は、言語的頑健さを得るために、統語規則の適用は文節内文法のみとし、統語的段階より上位の理解過程では、時間的に連続する文節を概念素と呼ぶ意味単位にまとめ、概念素の組として文意を理解する枠組である。しかし、意味理解の段階で統語的な制約を行っていないために、理解結果として表層的に不自然なものを受理するという問題があった。そこで、表層的な言語特徴の統計的性質を利用する統語制約を検討した。語彙 1000 語規模の音声対話実験システム上で収集した自由発話音声で評価した結果、第一位理解誤り率は 23.5% から 19.1%へ減少し、また、累積の理解誤り率では半減以下となる改善効果を確認した。

Concept-Driven Speech Understanding Incorporated with a Statistic Language Model

NAGAI Akito and ISHIKAWA Yasushi

MITSUBISHI Electric Corporation

Information Technology R & D Center

5-1-1 Ofuna, Kamakura, Kanagawa 247-0056, JAPAN

We have proposed a method of concept-driven semantic interpretation based on general semantic knowledge of conceptual dependency. In our approach, a concept is a unit of semantic interpretation and an utterance is regarded as a sequence of concepts that have an intention. Accepted understanding results, however, included a considerable number of meaningful hypotheses which were syntactically unreasonable. This is because no syntactic constraint was used in constructing a whole meaning from concepts but only semantic constraint was used to attain linguistic robustness. Therefore, we introduce a statistic language model which calculates plausibility of a sequence of concepts from the viewpoints of syntactic constraint. Experimental results of speech understanding for 1000-word-vocabulary spontaneous speech show that the proposed method significantly improves the system performance.

1 はじめに

一般の利用者の多様な発話表現を理解するために、自由発話の音声理解技術が必要である。我々は、意味主導的なアプローチで言語的頑健さを得る、概念素に基づく意味理解方式 [1, 2] を既に提案し、言語制約知識を一般化した概念依存関係に基づく音声理解方式 [3, 4] をベースに実用化検討を継続して進めている。本方式は、言語的頑健さを得るために、統語規則の適用は文節内文法のみとし、統語的段階より上位の理解過程では、時間的に連続する文節を概念素と呼ぶ意味単位にまとめ、概念素の組として文意を理解する枠組である (図1 参照)。

文献 [5] では、言語制約知識の一般性を損なわずに、対話システムに対する発話文としての意味的妥当性を評価する制約関係知識を検討し、理解性能向上への有効性を示した。しかし、本方式では意味理解の段階で統語的な制約を行っていないために、理解結果として表層的に極めて不自然なものを受理するという問題があった。

そこで、自由発話文の表層的な統語的特徴に着目し、この統計的性質を利用する言語制約を検討した [6]。以下、この統計的言語制約の検討について述べ、語彙 1000 語規模のタスクでの自由発話音声に対する理解性能を評価した結果を報告する。

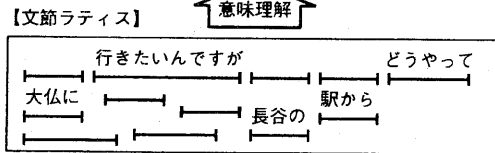
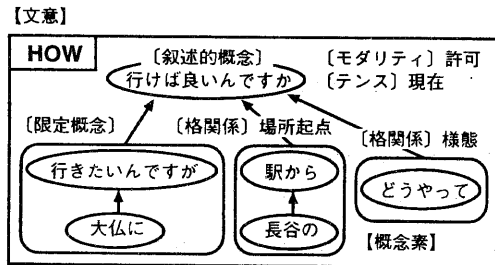


図 1: 概念素に基づく意味理解方式の原理

2 統計的言語制約の対象

発話表現の中には、可能な言い回しではあるが通常用いられないと思われる表現が存在し、妥当な意味解釈が可能でも表層表現として頻度が低い以下のような例がある。

- 「そのお寺って、拝観料が、いくらですか」(高頻度)
- ×「拝観料が、そのお寺って、いくらですか」(低頻度)

このような自由発話文の構文的な言語特徴の性質を、本理解方式での統語的制約として用い、理解性能の向上を図る。

この制約の実現にあたっては、知識源としては自由発話の多様な表現を受理でき、かつ、発話の統語的な妥当さを表わす重み付けを与えることが必要であり、統計的文法を利用することが望ましい。さらに、自由発話の表層的特徴に対するタスク非依存の一般的制約を実現するために、制約の対象は、学習テキストに依存する単語ではなく、表層的な言語特徴を用いるべきだと考える。

そこで、本理解方式における統計的言語制約の対象として、概念素の表層的な属性に着目する。概念素の属性とは、(1) 格性(主格、話題格、対象格、など)、(2) 連用修飾性(連用節、副詞)、(3) 連体修飾性(連体形の述語、名詞接続の助詞)、である。文意の構成に直接関係を持つのは、(1)と(2)であり、これら属性の生起順序の統計的性質が、自由発話文の統語的特徴を表現していると考えられる。

属性は、付属語、品詞、活用などの形態素情報により決まる。しかし、属性のうちの格性については曖昧性があり、格の認定は一般に困難である。そこで、この属性を、概念素の最後尾の表層的言語特徴である付属語カテゴリ (図2) により近似する。すなわち、付属語カテゴリの生起順序が属性の生起順序を近似的に表現しているとする。属性と付属語カテゴリとの具体的な対応の例を、表1に示す。

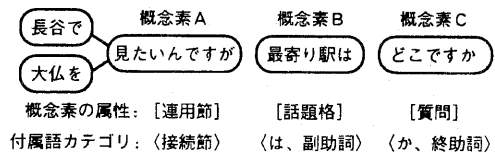


図 2: 概念素の付属語カテゴリ列

表 1: 属性と付属語カテゴリとの対応例 (一部)

話題格	助詞「は、って」
主格	助詞「が」
対象格	助詞「を、も」
起点格	助詞「から」
終点格	助詞「まで」
条件節	助詞「ば」、助動詞「なら、たら」
テ形連用節	接続助詞「て、で」
接続連用節	接続助詞「が、けど」

3 統計的言語モデル

本章では、具体的に用いる統計的言語モデルと、その定式化について述べる。

3.1 概念素単位の付属語 trigram

付属語の生起順序を制約する統計的言語モデルとして、 N -gram モデルを用いる。ここでは、付属語連鎖の特徴は bigram では表現しにくいと考え、trigram を用いる。先行研究では、文節単位で付属語の N -gram (bigram) を利用する方法 [7] が検討されている。ここで検討する概念素単位の付属語連鎖は、文節単位の付属語連鎖よりも大局的な構文的特点を表現すると考えられる。

また、付属語の連鎖確率の学習では、自由発話での付属語連鎖に対する一般的な妥当さを抽出することが我々の目的である。従って、付属語連鎖の確率がタスクに依存しないこと、及び自由発話文で言語的に妥当であると判断される付属語連鎖は同一に扱いたい、という観点から、付属語連鎖の確率がある値を越えたときに、その連鎖は言語的に妥当であるとみなし、同一の評価値を与えた。ここでは、この処理を付属語連鎖の規則化と呼ぶ。

3.2 定式化

以上述べた概念素単位の付属語 trigram を次のように定式化する。まず、付属語 trigram の連鎖確率の学習では、概念素単位の付属語カテゴリ連鎖を、 $C_1, C_2, \dots, C_n, \dots, C_N$ (ただし、 N は概念素数、 $3 \leq n \leq N$) として、 C_{n-2}, C_{n-1} の付属語カテゴリ連鎖の条件下で C_n が出現する生起確率 $p(C_n | C_{n-2}, C_{n-1})$ は、

$$p(C_n | C_{n-2}, C_{n-1}) = \frac{C_{n-2}, C_{n-1}, C_n \text{ の出現頻度}}{C_{n-2}, C_{n-1} \text{ の出現頻度}} \quad (1)$$

により、学習コーパス中の出現頻度から算出できる。ただし、付属語連鎖の規則化を考慮すれば、連鎖確率の規則化のためのいき値を $P_{threshold}$ ($0 < P_{threshold} < 1$)、規則としての同一の評価値を $Const.$ として、次式により算出する。

$$P(C_n | C_{n-2}, C_{n-1}) = \begin{cases} \frac{C_{n-2}, C_{n-1}, C_n \text{ の出現頻度}}{C_{n-2}, C_{n-1} \text{ の出現頻度}} & (< P_{threshold}) \\ Const. & (\geq P_{threshold}) \end{cases} \quad (2)$$

次に、発話文全体の付属語連鎖に対する生起確率 $P(S)$ は、発話文の開始と終了を意味するカテゴリをそれぞれ C_0, C_{N+1} として、次式で算出する。

$$P(S) = p(C_1 | C_0, C_0) \prod_{n=2}^{n=N+1} P(C_n) \quad (3)$$

上記の付属語カテゴリの連鎖確率を、理解処理の中で実際に利用する際には、以下のように、対数尤度として言語尤度 S_{ngram} を計算する。

$$S_{ngram} = -\log(P(S)) = -\log(p(C_1 | C_0, C_0)) - \sum_{n=2}^{n=N+1} \log(P(C_n)) \quad (4)$$

4 付属語 trigram の学習

本章では、付属語 trigram について、学習の方法と、学習結果に対する評価を述べる。

4.1 付属語カテゴリの設定

設定した付属語カテゴリは、格性を代表する 11 種類の助詞、連用修飾性を代表する 3 種類 (連用節「円覚寺に行きたいんですが、」、条件節「円覚寺に行ったら、」、テ形節「円覚寺に行って、」) のカテゴリ、それら以外を洗わず記号 <else>、及び文の開始終了の共通記号 - の、16 カテゴリである。実際に用いた全ての付属語カテゴリを、表 2 に示す。

表 2: 設定した付属語カテゴリ

<は, 副助詞>, <って, 副助詞>, <が, 格助詞>, <を, 格助詞>, <も, 副助詞>, <に, 格助詞>, <で, 格助詞>, <から, 格助詞>, <まで, 副助詞>, <へ, 格助詞>, <か, 終助詞>, <連用節>, <条件節>, <テ形節>, <else>, - (開始終了)

4.2 学習用テキストへのタギング

学習テキストは鎌倉観光案内に関する対話文であり、現在まで収集した音声対話データの書き起こし文と、被験者が対話を想定して任意の言い回しで作成した文とを合計した 1091 文を用いた。このテキストに前述の付属語カテゴリのタグを付与する。タギングは以下の手順で行なった。

1. 各テキストに対し、概念素境界の区切り記号 (/) を付与する。これは、概念素単位での付属語を抽出するためであり、この作業のみ人手で行なう。
2. 区切り記号を含んだテキストを形態素解析する。形態素解析には、奈良先端大学提供のフリーソフトウエア「茶筌 version 1.5」を用いた。
3. 概念素境界の区切り直前の助詞、助動詞、または用言活用形などの形態素情報を利用して、付属語カテゴリを決定し、タグ付けする。この際、「茶筌」で定義している品詞体系と、概念素理解方式で用いている品詞体系とが一部異なるために、簡単な品詞翻訳規則を用いて付属語カテゴリを決定した。

このような手順でタギングした例を以下に示す。

中華街の近くで/どこか/安いホテルは/ありますか
<で, 格助詞> <else> <は, 副助詞> <か, 終助詞>
そこは/駅から/歩いて行けますか
<は, 副助詞> <から, 格助詞> <か, 終助詞>
湘南海岸の近くで/きれいなプールのあるホテルを/
教えて下さい
<で, 格助詞> <を, 格助詞> <else>
八月四日から/江ノ島周辺で/二泊したいのですが
<から, 格助詞> <で, 格助詞> <else>
部屋からの眺めの良くて/食事の美味しい宿は/あり
ますか
<テ形節> <は, 副助詞> <か, 終助詞>

4.3 付属語 trigram の学習結果

付属語 trigram の学習では、妥当な付属語連鎖規則の認定基準として、生起確率が 0.05 を越える trigram に対して一律に確率 1.0 を与えた。付属語 trigram の学習結果を、表 3 に示す。また、学習された付属語 trigram に関して、高頻度で出現した三つ組、及び未出現の三つ組の例を表 4 に示す。

表 3: 付属語 trigram の学習結果

出現した付属語 trigram	375 種類
付属語 trigram の総サンプル数	3955
可能な trigram の総数	4096 (16 ³)

表 4: 学習された付属語連鎖の例。() 内は頻度。

<-/-/は> (160), <は/か/-> (125), <-/-/で> (119), <else/か/-> (115), <を/else/-> (98), <-/-/else> (92), <-/-/連用節> (82), <else/else/-> (79), <-/-/って> (55), <-/-/を> (54), <が/か/-> (44), <-/-/は/か> (42), <で/を/else> (36), <は/に/か> (22)
未出現:<は/まで/も>, <が/が/か>, <を/まで/は>, <に/から/で>

4.4 テキストによる評価

学習された付属語 trigram を用いて、学習テキスト、及び理解誤り例に対する尤度計算を行なった。尤度は (4) 式に従った。解析の結果、学習テキストに対しては 0.1 ~ 2.0、理解誤りのテキストに対しては、以下に示すように、4.5 ~ 13.1 程度の尤度を与えられ、理解誤りの抑制への効果を確認した。

- そこも/カレー/駅から/銭洗い弁天に/宿は/五月/ワインですか : 13.1
- 門からきたら/銭洗い弁天に/宿を/五月/行けば良いんですか : 8.3
- 鎌倉駅から/銭洗い弁天に/宿は/行けば良いんですか : 5.2
- リラ/玄って/教えて下さい : 4.5

5 音声理解実験

概念素を単位とする付属語 trigram による言語制約力を、音声理解実験により評価する。

5.1 実験条件

評価音声データは、鎌倉観光案内実験システムを用いてオフィス環境で収録した、話者6名による自由発話音声 [5] である。付属語 trigram の性能評価のために、これらの 260 評価文の中から、単一の概念素のみからなる評価文(「はい」「明日です」などの宣言的な発話)を除いた 162 文を本実験に用いた。

音声認識部の文節スポッティングは、文献 [8] の未知語処理なしの方式を用いた。文節内ネットワークは不特定話者の音節 HMM で構成され、ネットワークの規模は語彙 1005 単語(自立語 848、付属語 134、不要語 23)である。意味理解部では、概念素を 102、概念依存関係を 13 種類用いた。

理解結果である文意候補の総合尤度は、(1) 音響尤度、(2) 制約関係知識を用いた言語的妥当さの評価値であるペナルティスコア [5]、(3) 付属語 trigram 尤度、の三つの評価値を用いて算出する。具体的には、下記の算出式に従って、係数 W_1, W_2 による重み付きの線形和とした。

$$S_{total} = S_{acoustic} + W_1 * S_{penalty} + W_2 * S_{ngram} \quad (5)$$

ただし、 S_{total} : 総合尤度、 $S_{acoustic}$: 音響尤度、 $S_{penalty}$: ペナルティスコア、 S_{ngram} : 付属語 trigram 尤度、 W_1 : ペナルティスコアの重み、 W_2 : 付属語 trigram 尤度の重み、である。

ペナルティスコアの重みに関しては、 W_1 を定数として実験的に定めた。付属語 trigram 尤度の重みに関しては、発話文が長いほど尤度が悪化するために、正規化が必要である。付属語 trigram は概念素単位であることから、以下のように、概念素数による正規化を行なった。

$$W_2 = C/(n+1) \quad (6)$$

ただし、 W_2 : 付属語 trigram 尤度の重み、 C : 重み定数、 n : 文意候補内の概念素数 (+1 は、発話文前後の開始終了記号を含むことを示す)。

5.2 実験結果

付属語 trigram の理解性能向上への寄与を図 3 に示す。これより、音響尤度とペナルティスコアを併用した場合と比較して、付属語 trigram の導入により、第一位での誤り率は、23.5%から 19.1%へと減少した。さらに、第三位と第十位までの累積誤り率では、誤り率が半減以下となる改善効果を確認した。

また、ペナルティスコアの重みを $W_1 = 0.2$ に固定して、付属語 trigram 尤度の重みを、(1) $W_2 =$ 定数、とした場合(正規化なし)、(2) $W_2 = C/(n+1)$ とした場合(正規化あり)、の結果を、それぞれ表 5 と表 6 に示す。これらの結果より、以下がわかった。

- 付属語 trigram の導入による理解率の向上が、正規化なし/ありの両者の場合とも確認された。
- 正規化なしの場合は、重み W_2 の増加に対して理解率が低下した。これに対し、正規化ありの場合は、このような理解率低下はなかった。
- 最良の理解率は、正規化ありの場合が、正規化なしの場合を上回った。

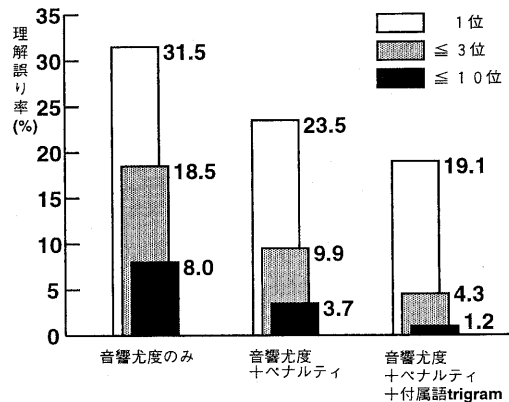


図 3: 付属語 trigram の理解性能向上への寄与: 理解誤り率 (%)

表 5: 文意理解率 (%): 正規化なし、話者 6 名の平均。

W_2	0.02	0.04	0.08	0.1	0.2	0.4
1 位	78.4	78.4	79.6	79.6	80.2	79.6
≤ 3 位	93.2	93.2	93.8	93.8	94.4	93.2
≤ 10 位	98.1	98.1	98.1	98.1	98.1	98.1

表 6: 文意理解率 (%): 正規化あり、話者 6 名の平均。

C	0.2	0.3	0.4	0.5	0.6	0.7
1 位	79.6	80.2	80.9	80.2	80.9	80.9
≤ 3 位	94.4	94.4	95.1	95.1	95.7	95.7
≤ 10 位	98.8	98.8	98.8	98.8	98.8	98.8

6 考察

本検討で導入した統計的言語制約方式に関する課題について考察する。

制約対象の言語単位：本検討では、統計的な言語制約の対象として、文意の大局的な発話構造を特徴付ける格や連用節(概念素の属性)に着目し、これらの近似表現として概念素単位の付属語カテゴリを用いた。まず、付属語カテゴリの種類としては、ここで設定した16の付属語カテゴリに加えて、さらにカテゴリ数を増やすことで、理解性能向上の可能性があると考える。例えば、「～で」は、場所を表す助詞と、条件を表す助動詞の場合があり、本検討ではこれらの両者を一つの付属語カテゴリとして扱っている。また、16付属語カテゴリにおける助動詞は、ここでは条件節を判別する「なら、たら」のみ用いており、他の助動詞を利用する余地が残されている。次に、付属語ではない他の表層的言語カテゴリの利用が考えられる。これは、自立語の一般的な品詞、述語の終止形や連用形などの活用形、単語の表層的な分類カテゴリ(意味クラス)などが考えられる。

統計的言語モデル：本検討では、言語モデルとして trigram を用いて、付属語の頻出連鎖を規則として学習テキストから抽出する方式を用いた。ここで設定した、規則化のための確率値の閾値は実験的に定めており、統計的言語制約モデルとしてどのような確率分布のモデルが良いかは、検討課題である。

総合尤度の算出方法：音響尤度と、言語制約のための知識源に基づく評価値とを組み合わせる場合、尤度を統合する重みを制御することにより、音声理解系の理解性能の特性をユーザに適応させることが可能である。例えば、通常は不特定話者用に一般的な統計的言語モデルに重みをおいて理解し、音響尤度が悪く認識しにくい話者の場合は、ヒューリスティックな理解戦略に基づく評価値の重みを上げることで、理解特性の話者適応ができる。このような異種の知識源の組み合わせ方法も今後重要と考える。

理解誤り：本理解方式は基本的に二つの概念間の依存関係を規定する知識源に基づいて理解するために、複数の概念依存関係で結合された語の共起関係が適当でない文意候補を受理するという問題点があり、今後の課題である。

7 おわりに

自由発話理解のための意味主導的な理解方式である概念素に基づく意味理解方式に、表層的な統語特徴に対する統計的言語制約を新たに導入した。音声理解実験の結果、理解性能向上の効果を確認した。今後は、複数の依存関係での語の共起制約を検討し、具体タスクでの自由発話理解の音声対話応用システムを開発目標として、実用性能を得るための改良検討を進める。

謝辞：研究の機会を与えて下さった、三菱電機株式会社 情報技術総合研究所 音声・言語インタフェース技術部 部長 中島邦男氏に感謝致します。

参考文献

- [1] 永井明人, 石川泰, 中島邦男 “概念素に基づく意味理解における音声認識への統合手法,” 情報処理学会 音声言語情報処理研究会, SIG-SLP 7-4, pp. 23-28 (Jul. 1995).
- [2] A. Nagai, Y. Ishikawa and K. Nakajima, “Integration of Concept-Driven Semantic Interpretation with Speech Recognition,” Proc. ICASSP'96, Atlanta (U.S.A.), pp. 431-434, (May 1996).
- [3] 永井明人, 石川泰, “概念依存関係による概念素統合に基づいた音声理解,” 情報処理学会 音声言語情報処理研究会, SIG-SLP 15-22, pp. 125-130 (Feb. 1997).
- [4] Akito Nagai and Yasushi Ishikawa, “Speech Understanding Based on Integrating Concepts by Conceptual Dependency,” Proc. EuroSpeech'97, Rhodes (GREECE), pp. 2747-2750, (Sep. 1997).
- [5] 永井明人, 石川泰, “対話システムのための概念素理解方式による対話音声理解,” 情報処理学会 音声言語情報処理研究会, SIG-SLP 18-5, pp. 25-30 (Oct. 1997).
- [6] 永井明人, 石川泰, “概念素理解方式による音声理解への統計的言語制約の導入,” 日本音響学会 春季研究発表会 講演論文集, pp. 15-16 (Mar. 1998).
- [7] 磯谷亮輔, 嵯峨山茂樹, “自立語と付属語の連鎖統計モデルを用いた音声認識のための候補選択,” 電子情報通信学会 信学技報 SP93-33, pp. 73-78 (Jun. 1993).
- [8] 花沢利行, 阿部芳春, 中島邦男, “文節スポッティングにおける未知語検出方式の改良,” 日本音響学会 秋季研究発表会 講演論文集, pp. 179-180 (Sep. 1996).