社団法人 情報処理学会 研究報告
IPSJ SIG Technical Reports

2007－ＳＬＰ－69（49）
2007／12／21

音声対話システムにおけるタスク外発話処理の高度化に関する研究

目黒豊美† 鈴木基之† 伊藤彰則† 牧野正三†

† 東北大学大学院工学研究科
〒 980-8579 宮城県仙台市青葉区荒巻字青葉 6-6-5
E-mail: †{meguro-t,moto,aito,makino}@makino.ecei.tohoku.ac.jp

あらまし 音声対話システムにおいて，従来のように記述文法で受理できる文章か受理できない文章かという識別だけでは，柔軟な対応をするためには不十分である．そこで，本研究では、意味的にタスクに沿っている文章かそうでない文章か識別することを目的とする．まず，記述文法を用いた音声認識と大語彙連続音声認識のスコアを用いて，受理可能な文と，受理不可能な文の識別を行ない，正解率 98%という高い確率で識別することを確認した．続いて，受理不可能な文がタスク内の発話かタスク外の発話かを識別するため，受理可能文と比較し，単語の意味的距離を調べ，意味が似ていればタスク内，意味が似ていなければタスク外とする手法を検討した．複数の単語類似度を比較し，平均して 90%程度の正解率を得ることができた．しかし，コーパス等に収録されていない単語については値を与えることができないなど，課題が残った．

キーワード 音声認識，単語類似度，タスク

Examination of judgment method of utterance outside task in voice conversation system

Toyomi MEGURO†, Motoyuki SUZUKI†, Akinori ITO†, and Syozo MAKINO†

† Graduate school of engineering, Tohoku University
6-6-5, Aoba, Aramaki-aza, Aoba-ku, Sendai, Miyagi, 980-8579 Japan
E-mail: †{meguro-t,moto,aito,makino}@makino.ecei.tohoku.ac.jp

Abstract In a small task, to be able to do more flexible processing,the utterance that relates to the task is recognized by the written grammar and the utterance that did not relate to the task is recognized by a large vocabulary speech recognition. Then, the technique for identifying sentences that do not relate to sentences that relate to the task by using semantic distance between words of the noun is examined in this paper.

Key words speech recognition, distance between words, task

1. はじめに

近年，音声認識技術が向上し，音声対話が実現できるようになってきた．現在，音声認識に用いる言語モデルは，N-gram を用いたものと，タスクごとに対応した記述文法を用いるものとの大きく二つに分けることができる．

記述文法は比較的小さなタスクに対しては，高い認識精度を得ることがまた，対話を進めるのに必要な単語を取り出すのも容易であるが，未知語を認識できず，想定外の発話には非常に弱いという特徴がある．

対して，N-gram を用いた大語彙連続音声認識は，言語モデルに含まれる単語をすべて認識できるため，想定外の発話や，語順が入れ替わった場合でも，認識することができる．しかし，認識精度が低いという欠点がある．また，対話を進めるのに必要な単語を取り出すことも難しい．

これらの特徴から音声対話には記述文法が多く用いられている．

しかし，音声対話に記述文法を用いると，前述のように想定外の発話を認識できないため，対話をすすめることができなくなることがある．つまり，ユーザーはシステム側が自分の発話を（本当は認識できるのに）正しく認識できていないだけなのか，それとも記述文法に書かれていないために認識することが不可能なのかを知ることができない．

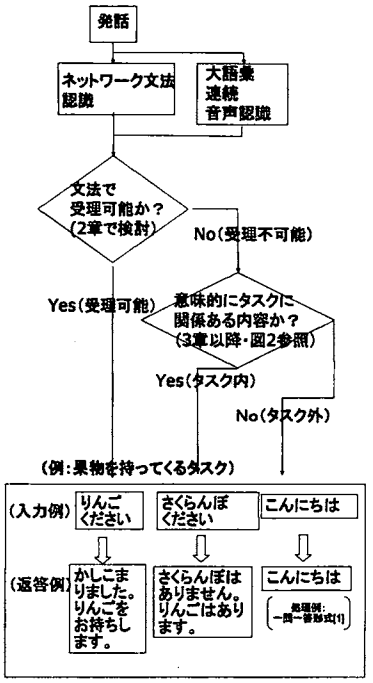

図 1　想定するシステムのフローチャートと処理例

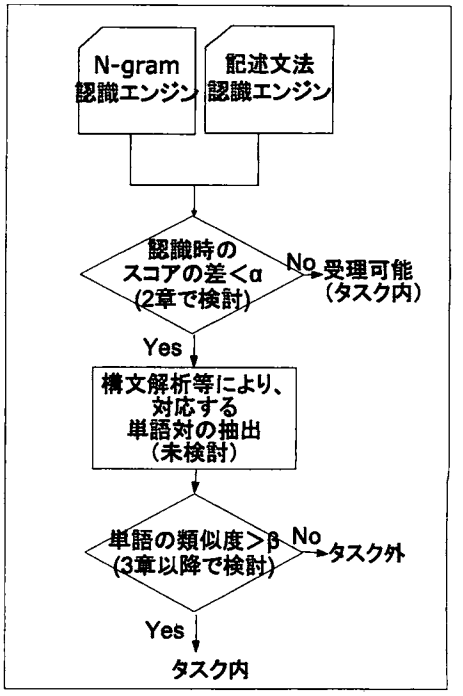

図 2　想定するシステムの具体的な処理の流れ

この欠点を補い，音声対話をスムーズに進めるため，以下のようなシステムを提案する．

（1） 発話が受理可能かどうかを識別する (2. 章)
受理可能な場合は，記述文法の認識結果をシステムの認識結果とする．

（2） 次に受理不可能な場合は，タスク内かタスク外かを識別する (4. 章)
この手法により，入力音声が「記述文法で受理できる」「文法で受理できないが，意味的にはタスクに沿っているもの」(以後，この 2 種をタスク内と呼ぶ) 「完全にタスク外」の 3 つのクラスに分類できる．（図 1 参照）

また，発話が未知語を含む場合，ユーザに発話のどの単語が未知語であるかを図 1 の例ように「さくらんぼはありませんが，りんごはあります」とフィードバックするためどこが未知語であるかまでを同定したい．そのために，記述文法による認識と N-gram による大語彙連続音声認識を同時に行い，その結果を比較する方法を検討する．

具体的には，図 2 のように 2 つの認識器から構文解析等を行ない対応する名詞対を抽出し，単語類似度を測るという手法を検討する．

2. 受理不可能文の識別

2.1 識別方法

今回提案する方法は，N-gram を用いた認識時の音響スコアと言語スコアをあわせたスコアと，記述文法を用いた認識時の音響スコアを比較するというものである．N-gram を用いた際のスコアは，式 (1) を用いた．

$$\mathrm{score}_1 = \log p(X|W) + \alpha * \log p(W) + \beta|\omega| \qquad (1)$$

α = 8.0 ：言語モデル重み
β = -2.0 ：単語挿入ペナルティ
ω ：単語数

手書き文法を用いた認識時の音響スコアは，式 (2) を用いた．

$$\mathrm{score}_2 = \log p(X|W) \qquad (2)$$

$$\mathrm{score}_1 - \mathrm{score}_2 \geqq \theta \qquad (3)$$

式 3 の θ が，閾値以上高ければ受理可能文，以下であれば受理不可能文とし，判別を行なった．

2.2 実験方法，結果

タスクとして，今回は，図 3 に示すような，果物を注文するタスクを用いた．図 3 の" くだもの" と" 数" はそれぞれ下のように単語の集合を表している．

くだもの ： みかん　りんご　ぶどう
数　　： 0-9

この文法で受理可能文と，その他に受理できないがこのタスクに意味的に似ている文章や，まったく関係ない

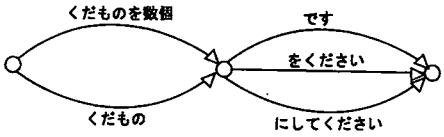

図 3　実験に用いた認識文法

表 1　受理可能，不可能と判別された文章数

文章数 (二名平均)		識別結果		計
		受理可能	受理不可能	
入力文	受理可能文	20	0	20
	受理不可能文	2	80	82

文章など計 102 文を，認識，スコア算出し，受理できるかできないかを判別した．認識には，大語彙連続音声認識システム Julius および Julian [3] を用いた．被験者は男性 1 名女性 1 名．音響モデルは，日本語の不特定話者モデル，言語モデルは，連続音声認識コンソーシアムによる語彙数 6 万の Web モデルを用いた．

表 1 は，閾値を 50 とし識別を行った結果である．受理可能な文かどうか 99%の確率で識別できている．これは，認識文法をかなり小さいサイズのものを使ったので，スコアの差がはっきり出てきたためである．

3.　認識結果文の比較による認識率の改善

4. 章で単語類似度を用いるためには，大語彙連続音声認識の認識結果が信頼できるものでなければならない．そのため，高い認識率が求められる．2. 章の実験で，大語彙連続音声認識の認識率は 102 文合計では 66%であった．しかし，受理不可能かつタスクに関連ある文章 (表 2 内の番号（以下，番号）2~ 5 の文章）は 59% , タスク外の文章（番号 6,7）は 85%と，タスク内の文章の認識率がかなり低かった．これは，大語彙音声認識に用いた言語モデルがタスク内に特化したものではないので，認識しづらいものになっていたからである．よって，これらの認識率をあげることを考える．

2. 章の実験において，大語彙音声認識と記述文法による認識のそれぞれについて，認識スコアと認識結果を比較した．その結果，以下のことを確認できた．

- 認識文法で受理できる文と発話文がまったく違う場合
  - 記述文法での認識時のスコアと大語彙音声認識時のスコアに大きく差がでる．
  - 記述文法での認識結果文と大語彙音声認識結果文の Levenshtein 距離が大きくなる
- 類似文 (番号 2,3,5,6)
  - スコアの差が小さくなる
  - 距離も小さくなる

表 2　実験に用いた文章の種類

文章の種類	タスク内外	例
1 受理可能文	内	蜜柑をください
2 受理不可能な言い回し	内	蜜柑食べたい
3 受理不可能なくだもの	内	メロンください
4 受理不可能な言い回し+くだもの	内	メロン食べたい
5 タスク内発話に似た文	外	蜜柑おいしい
6 受理可能文を誤認識した文	外	日本の子です
7 タスク内発話に関係ない文	外	こんにちは

類似文では，発話文の一部は記述文法で受理できるので，記述文法の認識結果文と大語彙音声認識結果文の距離が近いものがより上位にくるようにすることによって，単語正解精度が上がると考えられる．

よってまず，Levenshtein 距離が 0 より大きい，または スコア差が 50 より小さいという条件の元，類似文 (番号 2,3,5,6) を抽出する．この抽出法の正解率は 95%であった．

この類似文に対して，大語彙音声認識結果 1000best の中で，記述文法を用いた認識結果文との Levenshtein 距離がもっとも小さい文を選択する．これを認識結果とすると，1best の文章と比較して，単語正解精度が 65%から 73%に上昇することがわかった．

4.　単語類似度によるタスク外発話の識別

4.1　はじめに

音声対話システムが柔軟な対応をするためには，記述文法で受理できない文章が，タスクに関係ある文かタスクに関係ない文であるかを判別する必要がある．そこで，単語類似度を用いた識別を提案する．

4.2　識別手法

認識文法で受理できる文と発話文の一部が同じ場合 (類似文と呼ぶ．)，タスク内の文である確率が非常に高いと言える．

例えば，2. 章の記述文法と大語彙音声認識器を用いて，「さくらんぼください」という発話を正しく認識できた場合 (さくらんぼは受理できない)，

記述文法の認識結果　「りんご　　ください」
大語彙音声認識結果　「さくらんぼ　ください」

と，認識結果が返って来たとする．「さくらんぼ」と「りんご」の意味が近いことがわかれば，この発話がタスク内であると識別することができる．この例のように正しく認識し，かつ 2 つの認識結果から単語対を抽出できたという仮定する．

このように，記述文法と大語彙音声認識の認識結果の対応をとった結果，「りんご」と「さくらんぼ」のように名詞の部分だけ異なる場合，単語類似度を用いて 2 つの名詞の意味的距離をはかる．発話内の単語と文法に含まれる単語の類似度が大きいときに，その発話をタスク内で

あるとする.

4.3 web 上の共起頻度を用いた名詞の識別 [4]

まず web 上で同じページにある単語を共起している (つまり似ている意味である) とし,YahooAPI の検索ヒット数を用いて計算した. a_i と a_j を距離を調べたい 2 つの名詞であるとする. これらの単語類似度は, (4) 式のように表される.

$$sim(a_i, a_j) = \frac{2f_{ij}}{f_i + f_j} \tag{4}$$

f_i は a_i を検索クエリとしたときのヒット数

ある閾値より値が大きければタスク内, 小さければタスク外となる.

4.4 シソーラスを用いた名詞の識別

シソーラスを用いた手法では荻野 [5] の作成したシソーラスと, 日本語語彙体系 [6] 上の距離を意味的距離とした. これらシソーラスは木構造になっており, 枝 1 本を距離 1 とし, 意味的距離を求めた. この場合, 親は距離 1, 兄弟は距離 2 となる.

ある閾値より値が大きければタスク外, 小さければタスク内となる.

4.5 LSA を用いたタスク外名詞の識別

今回は, 毎日新聞 94 年~ 03 年 10 年分で LSA を作成し, それぞれの単語ベクトルをコサイン類似度で評価した. LSA 作成には, Infomap-NLP [7] を用い, 特異値の数は 100 とした.

ある閾値より値が大きければタスク内, 小さければタスク外となる.

4.6 実験に使用したタスク

表 3 のような飲み物注文タスク, 洋服選択タスクと果物注文タスクを用意し, さらに 3 つのタスクとは関係ない名詞をタスク外単語として 50 個用意した.

それぞれの単語のタスク内との距離は, 式 5 のように計算した. つまり, タスク内の単語は他のタスク内単語との類似度を求め平均し (CrossValidation) , タスク外の単語はタスク内の単語との類似度を求め平均した.

$$d(x) = \frac{\sum_{w \in W} sim(x, w)}{|W|} \begin{cases} \text{タスク内}: & W = V - \{x\} \\ \text{タスク外}: & W = V \end{cases} \tag{5}$$

V:タスク内の単語の集合

w:単語

5. 実験結果

実験結果は, 図 4~ 図 7 のようになった. 未収録の単語は無視している. 縦軸は正解率, 横軸はそれぞれの閾値を表している.

表 3 タスクに含まれる名詞とタスク外の名詞例と個数

タスク名	例	総数
飲み物	お茶 紅茶 緑茶 etc.	30 個
洋服	シャツ ズボン ジャケット etc.	30 個
果物	りんご みかん ぶどう etc.	30 個
タスク外	宇宙 薬 医師 etc.	50 個

表 4 コーパスと共起条件

	コーパス	共起条件
web	Yahoo 内全データ (YahooAPI を利用)	同ページ内

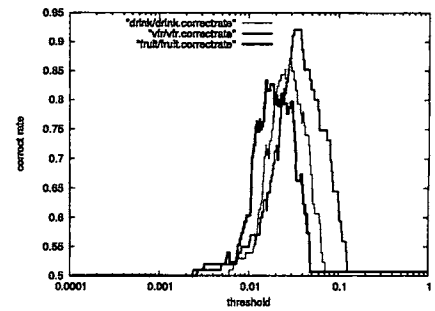

図 4 web の単語類似度の閾値と正解率

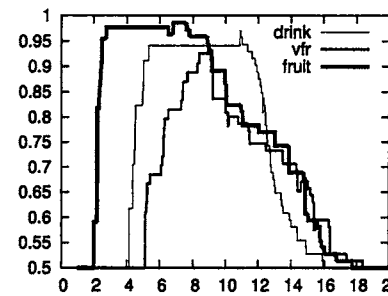

図 5 日本語語彙大系での単語間距離閾値と正解率

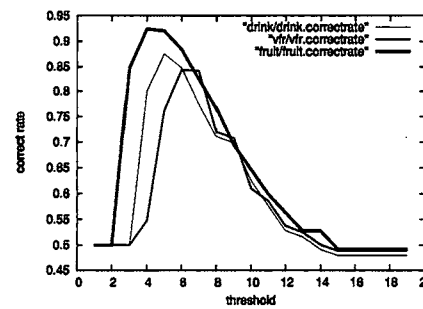

図 6 分類体系での単語間距離閾値と正解率

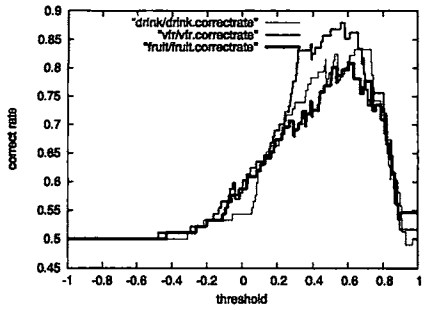

図 7 LSA のコサイン類似度閾値と正解率

5.1 結　　果

最適な閾値と，正解率は表 5 と表 6 のようになった．

正解率を見ると，日本語語彙大系を用いた場合が最も高くなっている．シソーラスは統計的なものではなく，人手で作っているもののため，単語の意味を正確に表している．しかし，シソーラスには必ず未知語が含まれているため，すべての単語に対して値を与えることができない．(理想的には，言語モデルに含まれるすべての単語を表したシソーラスがあるといいのだが，シソーラスは人手で作るため言語モデルを作るたびに作り直さなければならず実現は難しい) また，シソーラス内の木構造が，システム設計者の意図した単語の意味の木構造が同じであるとは限らない．そこで，ひとつのシソーラスを使うより，複数のシソーラスを用いることで，それらの欠点をカバーできる可能性がある．

対して，Web 上の共起頻度は正解率がシソーラスに比べ劣るものの，未知語はないと言っていい．

そこで，これら複数の方法を組み合わせることで，すべての単語に対して，よりよい正解率を得られる方法を今後検討していく．

表 5 正解率が最も高くなる閾値

最適な閾値	飲み物	果物	洋服
分類体系	5-6	4-5	5-6
日本語語彙大系	10.88-10.94	6.81-7.59	8.38-9.14
新聞記事 LSA	0.6377-0.7310	0.6027-0.6317	0.5582-0.5826
Web 共起頻度	0.0298-0.0299	0.0154-0.0158	0.0319-0.0387

表 6 最適な閾値のときの正解率

正解率 (%)	飲み物	果物	洋服
分類体系	87.56	92.4	84.37
日本語語彙大系	97.05	98.64	92.60
新聞記事 LSA	83.3	80.8	88.0
Web 共起頻度	86.6	83.3	92.1

表 7 未知語数

正解率 (%)	飲み物	果物	洋服	タスク外
分類体系	15	5	8	9
日本語語彙大系	13	8	2	13
新聞記事 LSA	11	14	11	4

6. ま と め

音声対話システムの発話が，受理可能文か受理不可能文であるかを，認識時のスコアを用いて識別する方法を検討した．小さなタスクにおいては，高い確率で識別が行なえることがわかった．次に，受理不可能文が意味的にタスクに近いものであるか，または関係ないものであるかを，名詞間の類似度を用いて識別する方法を検討した．

未知語に対しては識別できない手法（シソーラス）では高い正解率を得ることができ，逆にすべての単語の識別ができる手法（共起頻度）ではシソーラスに比べて低い正解率だった．

今後はこれらの複数の手法の組み合わせ方を検討していく．

文　　献

[1] 音声対話エージェントによる生駒市コミュニティセンターの案内システム
西村竜一，西原洋平，鶴身玲典，李晃伸，猿渡洋，鹿野清宏
情報処理学会第 65 回全国大会
[2] 鶴身玲典．“タスク文法による N-gram 確率の部分強化を用いた音声認識アルゴリズム”(修士論文) 奈良先端科学技術大学院大学 2003
[3] 大語彙連続音声認識システム Julius: http://julius.sourceforge.jp/
[4] 新納浩幸 佐々木稔 “検索エンジンを利用した未登録単語に関する単語間距離の測定”，言語処理学会第 12 回年次大会，D2-1, pp.380-383 (2006)
[5] “現代日本語名詞シソーラスによる意味の分類体系” 荻野綱男 東京都立大学 人文学部
[6] ”日本語語彙体系”，岩波書店
[7] Infomap-NLP Software, http://infomap-nlp.sourceforge.net/