

## 音声翻訳のための認識誤りにロバストな類似用例選択手法の評価

廣瀬 良文† 脇田 由実† 小沼 知浩‡ 西崎 誠‡ 遠藤 充‡

†松下電器産業株式会社 先端技術研究所

〒619-0237 京都府相楽郡精華町光台 3-4

‡松下技研株式会社 情報・ネットワーク研究所

〒214-8501 神奈川県川崎市多摩区東三田 3-10-1

E-mail: {hirose,yumi}@crl.mei.co.jp,{tkonuma,mnishi,endoh}@mrit.mei.co.jp

### あらまし

音声翻訳技術を小型携帯端末上で実用化しようとした場合には、音声認識誤りに頑健であることが必要である。われわれはドメインを限定し、キーワード主導型の類似用例選択を行うことによって、音声認識誤りに頑健な翻訳を行うことを試みた。また、キーワード間の依存関係を用いることにより、用例候補を効率よく選択することを試みた。日本語と中国語のコーパスを用いて、誤りを含む音声認識結果のみを入力とした用例選択実験を行った。用例選択率は、日本語で 57.95%、中国語で 60.7%の再現率を得ることができた。主観評価結果においても、類似用例選択を行うことにより、意味理解率が音声認識結果そのものを理解しようとした時と比較して約 2 倍になった。また、キーワード間の依存関係を用いることにより、用例候補数を日本語で 84%、中国語で 26%削減することができた。

**キーワード** キーワード、用例選択、依存関係、多言語

## Evaluation of robust speech translation using example-sentence-driven method

Yoshifumi Hirose† Yumi Wakita† Tomohiro Konuma‡

Makoto Nishizaki‡ Mitsuru Endoh‡

†Matsushita Electric Industrial Co., Ltd. Advanced Technology Research Laboratories  
3-4 Hikaridai, Seika, Souraku, Kyoto, 619-0237, Japan

‡Matsushita Research Institute Tokyo, Inc. Information & Network Research Laboratory  
3-10-1 Higashimita, Tama-ku, Kawasaki, Kanagawa, 214-8501, Japan

E-mail: {hirose,yumi}@crl.mei.co.jp,{tkonuma,mnishi,endoh}@mrit.mei.co.jp

### Abstract

For practical use of speech translation on portable hardware, it is necessary to develop a robust speech translation technique capable of dealing with sentences including recognition errors. We propose a language translation method in which speech recognition results are mapped to example sentences by keywords.

To confirm effectiveness of our method, we conducted example selection method for Japanese and Chinese sentences. We confirmed that proposed method improved speech understanding rate twice, and reduced the number of output sentences by 84% for Japanese, and 26% for Chinese.

**Key words** keyword, example sentence selection, dependency, multi-lingual

## 1はじめに

われわれは、携帯端末上で実現できる音声翻訳技術の実用化を目指し、旅行会話用音声翻訳システムを構築している。音声翻訳の実用化を考えた場合、音声認識の誤りに対して、頑健でなければならない。従来、話し言葉に対応可能な用例選択型の翻訳<sup>[1-4]</sup>はなされていたが、誤りを含んだ音声認識結果への対処は不十分であった。

通常、音声認識誤りを含む入力をそのまま翻訳すると、誤りが翻訳結果に大きく影響を及ぼす。そこで、われわれはキーワード主導型類似用例選択により音声認識誤りに頑健な音声翻訳を行うために音声認識結果からキーワードを抽出し、抽出したキーワードの誤りを推定し、正しいと思われるキーワードにより、類似用例選択を行うキーワード主導型音声翻訳システムを構築している。

また、われわれは旅行などで音声翻訳システムを使うことを想定し、持ち運びに便利な小型携帯端末 (PDA) に実装することを考えている。携帯端末での動作を考えた場合、用例選択型の翻訳において選択する用例数は、できるだけ少なくする必要がある。そこで、キーワード間の依存構造を用いることにより適合率の向上を試みた。

このようなキーワード主導型の類似用例選択手法は、言語に依存しないと考えられる。そこで、日本語と中国語において用例選択の性能評価を行った。

## 2. キーワード主導型類似用例選択方式を用いた音声翻訳

### 2.1. 本方式の構成

音声翻訳を行う場合、入力部分の音声認識における認識誤りは避けられない。従って、認識誤りにロバストな翻訳技術が必要となる。われわれは、キーワード主導型用例選択方式を用いた翻訳を試みた。キーワード主導型用例選択方式の利点は、機能語などの誤りに頑健なことである。

図 1 にシステム構成を示す。連続音声認識器が出力した音声認識結果から、キーワードを抽出する。キーワードとは、文の意味を理解するのに最低限必要な単語のことである。対訳コーパスから抽出したキーワード間の依存関係を用いて抽出したキーワードの誤りを推定し、誤りキーワードを除去する。キーワード間の依存関係は、対訳コーパスを構文解析システムによ

り構文解析して、依存関係を抽出し、人手によりチェックしたものを用いた。依存関係により推定された誤りキーワードを除いて、残りのキーワードを用いて用例選択を行う。用例選択は、キーワードの一致度とキーワード間の依存関係の一致度が最も高い用例を用例データベースから選択する。

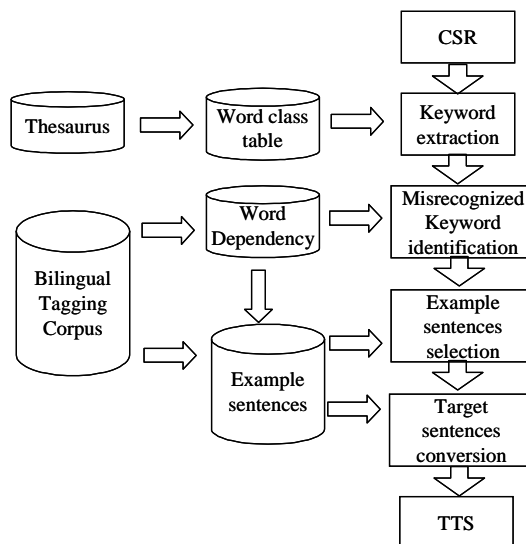


図 1 システム構成図

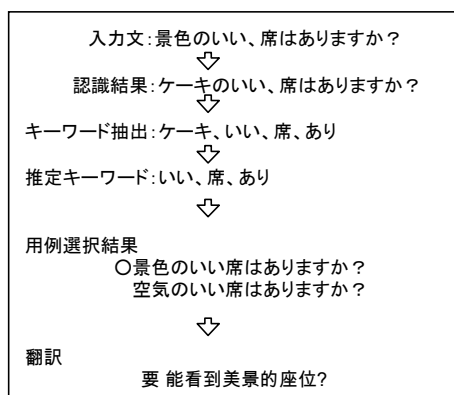


図 2 システムの動作例

選択した用例が複数ある場合には、用例の中から話者は自分の発声に最も近い用例を選択する。選択された用例を用例データベースを用いて翻訳し、合成音声を出力する。図 2 に本システムの動作例を示す。音声認識の段階で、「景色」が「ケーキ」に置換誤りを起こしている。この認識結果に対してキーワードを抽出すると、「ケーキ」、「いい」、「席」、「あり」というキーワードが抽出される。予め、用例から抽出した依存関係から、各キーワード間の関連を調べ、関連の低いキーワードをキーワードから除去

する。例の場合、「ケーキ」が除去される。このようにして、抽出したキーワードから用例検索して翻訳を行う。

## 2.2. キーワードクラスタリング方法<sup>[5][6]</sup>

ドメインを限定した場合、必要な会話の表現をすべて収集することは可能である。しかし、必要なすべての語彙を集めることは、たとえ限られたドメインであったとしても非常に困難である。そこで、キーワードをキーワードクラスに割り当てて、用例検索には、キーワードそのものではなく、キーワードクラスを参照することにより、ドメインに属するすべての用例選択を行うことが可能になる。

キーワードをキーワードクラスに割り当てる方法としては、シソーラスを用いることが考えられる。しかし、シソーラスを用いると、キーワードクラスを定義することは可能であるが、以下に挙げるような問題点がある。

- (1) ある単語は複数の意味クラスに属するため、1つのキーワードクラスを定義することが困難である。
- (2) 同じ意味クラスに属していても、用法が異なるものが存在する。

そこで、依存構造解析を用いて、クラス内の各単語間の類似度を定義し、図3に示すように、対訳コーパスに適合するようなクラスタリングを行う。

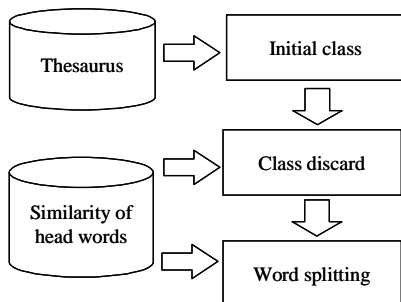


図3 クラスタリング方法

初期クラスは、シソーラスによるクラスを用いる。各クラスについて、類似度を算出し、以下の処理を行うことによりクラスタリングを行う。

- A) クラス内の各単語間の類似度の平均が閾値以下の場合には、そのクラスを除去する。
- B) クラス内の他の単語との類似度がすべて閾値以下である場合、そのクラスから

単語を除外する。

結果として、シソーラスのクラスから、ドメインに最適なクラスが選択される。また、他の単語と意味が異なる単語をオリジナルのシソーラスクラスから除去することができる。

なお、クラスAに属する単語 $W_j$ と $W_i$ の間の類似度は、以下の式により計算する。

$$\text{Sim}(W_j^A, W_i^A) = \sum_{k=1}^K R(W_j^A, Z_k) \times \sum_{k=1}^K R(W_i^A, Z_k)$$

$$R(W_j^A, Z_k) = \frac{\text{FreqPair}(W_j^A, Z_k)}{\text{Freq}(W_j^A)}$$

where

if  $R(W_j^A, Z_k) = 0$ , then  $R(W_i^A, Z_k) = 0$

if  $R(W_i^A, Z_k) = 0$ , then  $R(W_j^A, Z_k) = 0$

$\text{Sim}(W_j^A, W_i^A)$ :  $W_j$ と $W_i$ の類似度

$Z_k$ : k番目の修飾語句

$\text{FreqPair}(W_j^A, Z_k)$ :  $Z_k$ に $W_j^A$ が修飾された頻度

$\text{Freq}(W_j^A)$ :  $W_j^A$ の出現頻度

$K$ : 修飾語句の数

## 3. 実験

### 3.1. 実験条件

本翻訳の性能は類似用例が正しく選択できるかによる。そこで本翻訳手法の有効性を確かめるために、音声認識誤りを含む入力のみに対して、類似用例選択を行う実験を行った。表1に本実験の実験条件を示す。用例は、旅行会話文であり、日本語が1055文型、中国語が924文型である。評価文はネイティブ話者により用例と同じclosedな文章を発声した。音声認識は、N-gramによる連続音声認識を用いた。言語は日本語と中国語を用いた。音声認識結果のうち、認識誤りを含んだ日本語302文、中国語1284文のみを対象とした。

表1 実験条件

	日本語	中国語
用例種類数	1051 文型	924 文型
話者数	30 名	18 名
異なり単語数	1426	1588
認識誤りを含む文	302 文	1284 文

### 3.2. 単語認識率とキーワード認識率

表 2 に認識誤りを含む文のみに対する音声認識率を示す。表において単語認識率は入力文の全単語に対する認識率であり、キーワード認識率は、入力文中に含まれるキーワードのみの認識率である。なお、認識率は正解精度 (accuracy) により評価し、以下の式で算出した。

$$accuracy = \frac{N - Sub. - Del. - Ins.}{N} \times 100 [\%]$$

日本語においては、単語正解精度率は、57.17% であり、キーワード正解精度は 58.43% とほぼ同等であった。

一方、中国語に関しては、単語正解精度は 56.50% であり、キーワード正解精度は 61.39% とキーワード正解率が約 5% 高かった。

表 2 誤りを含む文のみに対する音声認識率[%]

	日本語	中国語
単語認識率	57.17	56.50
キーワード認識率	58.43	61.39

### 3.3. 用例選択率による評価

表 3 に用例選択率を示す。評価は再現率 (recall) と、適合率 (precision) により行い、以下の式により算出した。ここで、用例選択に成功したとは、用例選択部が出力した用例に、入力した用例が含まれていることを示す。したがって、入力した用例を正しく翻訳できる上限値を示す。

$$recall = \frac{\text{用例選択に成功した文数}}{\text{用例数}} \times 100 [\%]$$

$$precision = \frac{\text{用例選択に成功した文数}}{\text{出力文数}} \times 100 [\%]$$

再現率では、日本語 57.95%、中国語 60.70% と、約 6 割の入力に対して、正しい用例を選択できていることがわかる。表 4 に日本語において、キーワードが 3 個以上の文で、キーワード誤り率が 50% 以下の入力に対する用例選択率を示す。キーワードが 3 個以上である文数は全体の 81% であり、大部分の用例に対して依存関係を用いることが可能であることがわかる。表 4 に示すようにキーワード誤りが半数以下の場合には、再現率は 83.23% であった。このことから誤りを含んだ入力に対しても、キーワードが半数程度認識できている場合には、正しい用例を選択できることがわかる。

一方、適合率では、日本語の場合には、

39.77% と、中国語と比較すると不要に選択される用例数が多いことがわかる。これは、日本語のクラス数が 63 と、中国語の 37 と比較して多いため、クラス化された用例を多く選択するためである。また、日本語用例中に指示代名詞クラス(「この」、「あの」等)が多く含まれているため、キーワード認識率が高くない場合には、用例を誤って選択することがあるためである。

表 3 用例選択率[%]

	再現率	適合率
日本語	57.95	39.77
中国語	60.70	49.31

表 4 キーワード誤り率 50% 以下の時の用例選択率(日本語)[%]

キーワード数	誤り単語数	文数	再現率	適合率	
3	0	20	100.00	86.96	
	1	24	70.83	45.95	
4	0	16	93.75	93.75	
	1	20	90.00	85.71	
5	2	21	66.67	53.85	
	0	10	80.00	80.00	
6 以上	1	9	100.00	60.00	
	2	16	93.75	83.33	
	0	5	80.00	80.00	
	1	11	81.82	69.23	
	2	6	83.33	83.33	
	3	6	66.67	50.00	
	4	2	50.00	50.00	
Sum			167	83.23	71.25

### 3.4. 主観評価

表 5 に用例選択結果に対するの主観評価結果を示す。評価は、ネイティブ(日本語 2 人、中国語 5 人)により 5 段階で行った。評価基準を以下に示す。

1. 違う意味である。
2. 部分的には同じ意味であるが、入力文を予想することはできない。
3. 部分的に同じ意味であり、入力文を予想することができる。
4. ほぼ同じ意味である。
5. 完全に同じ意味である。

但し、入力文は対話文であるため、あらかじめ会話の場面を想定しないと、意味が理解できない場合がある。そこで、入力文が話された場面を提示して、回答して貰った。

スコア3～5が「意味がわかる」と判断されたと考えると、日本語の場合、音声認識の段階では、スコア3～5が28%であるのに対し、用例選択結果では、51%とほぼ倍増していることがわかる。また、「完全に同じ意味である」と回答した割合は約25%であり、音声認識が誤った場合においても4分の1は、完全に回復していることがわかる。

一方、中国語に関しても、スコア3～5が音声認識の段階では、19%であるのに対し、用例選択結果では45%と、2.3倍であることがわかる。また、「完全に同じ意味である」と回答した割合は約25%であり、日本語の場合と同様の結果であった。

表 5 主観評価結果[%]

スコア		1	2	3	4	5
日本語	認識結果	46.69	24.67	13.58	9.44	5.63
	用例選択	33.94	14.74	17.88	8.11	25.33
中国語	認識結果	25.86	55.48	12.34	4.64	1.68
	用例選択	25.79	29.63	16.67	2.48	25.44

#### 4. 考察

##### 4.1. 依存関係利用が用例選択性能に及ぼす効果

小型携帯端末(PDA)に実装を考えた場合には、用例選択候補を効率よく選択する必要がある。本方式では、キーワード間の依存関係を用いることにより、用例選択において出力される候補文を効率よく絞り込むことを試みた。依存関係を用いた場合の効果を検証するために、依存関係の有無による用例候補出力数と用例選択率について調べた。結果を表6に示す。

表6では、依存関係を用いることができた用例のみについて調べた。結果は、日本語においては、依存関係を用いなかったときには、用例候補が多数出力されることが多かった。一方、依存関係を用いることにより、用例候補数は84%削減できた。結果として、再現率では約8%依存関係を用いない時の方が高い結果であるが、適合率に関しては5.6倍の結果を得た。

一方、中国語に関しては日本語の場合ほど顕著には表れないものの、依存関係を用いること

により、用例候補数を24%削減することができ、適合率に関して1.2倍の結果を得た。

日本語と中国語における適合率の差は、3.3節で述べたようにクラス数の差と、日本語用例に指示代名詞が多かったためであるが、依存関係を用いない場合、指示代名詞を含む用例が過剰に選択されたため、影響が顕著に現れた。

また、中国語における出力文数の累積頻度について調べた結果、依存関係を用いた場合、出力文数が1文である割合が88%であるのに対し、依存関係を用いなかった場合では、72%と低い値になっている。また、出力文数が3文以内である割合は、依存関係を用いた場合では、99%であるのに対して、依存関係を用いない場合では、91%であった。したがって、依存関係を用いることにより、類似用例選択を行う際の用例候補数を効率よく絞り込むことが可能であるといえる。

表 6 依存関係を用いることによる用例選択率の変化[%]

	該当文数	依存関係	再現率	適合率
日本語	223	あり	74.89	49.26
		なし	82.96	8.72
中国語	936	あり	81.2	69.03
		なし	85.89	55.76

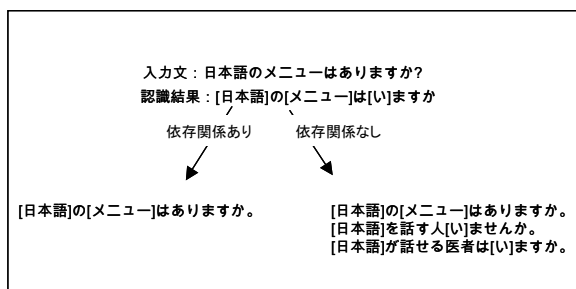


図 4 依存関係を用いることにより適合率が上昇する理由

図4に依存関係を用いることにより、適合率が良くなる例を示す。認識結果に含まれるキーワードにおいて、「日本語」と「メニュー」に依存関係があるが、「日本語」と「い」(いる)には、依存関係が無いため、依存関係を用いた場合には、正解文のみが選択される。しかし、依存関係を用いない場合には、認識文中のキーワードを含む用例をすべて選択してしまうため、結果として適合率が低下する。

図 5 に依存関係を用いることにより、再現率が低下する例を示す。認識結果に含まれるキーワードは「電池」のみであり、依存関係を用いることができない。したがって、正しい用例を選択することができなくなる。一方、依存関係を用いない場合には、認識結果中のキーワード「電池」を含む用例すべてを選択するため、結果として、再現率が上昇する。

以上の理由で依存関係を用いることにより、過剰な用例選択を防ぎ、適合率を改善することができた。

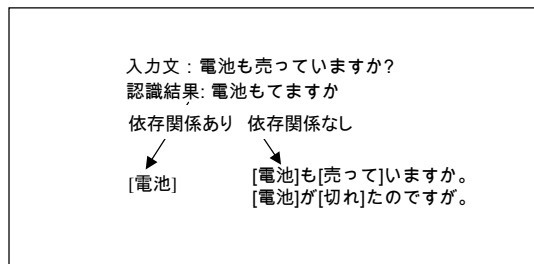


図 5 依存関係を用いることにより、再現率が下がる理由

#### 4.2. 言語依存性

本方式では、キーワードと、キーワード間の依存関係のみを用いて、類似用例選択を行っている。したがって、文法などの言語に依存する知識を用いた場合と比較すると、多国語に展開し易い。

本実験においては日本語と中国語において、キーワード主導型の類似用例選択を行ったが、誤りを含む認識結果のみに対して、音声認識率に多少異なりはあるが日本語で 57.95%、中国語で 60.7%の再現率を得ることができた。一方、適合率に関しては、日本語 39.77%に対して、中国語 49.31%と差が見られた。しかし、これは中国語のキーワードクラスが用例中にあまり存在しなかったため、過度の用例選択が起きなかったためであり、本方式において言語依存性があるからではない。

以上のことからキーワード主導型類似用例選択手法は、言語に依存することなく利用することが可能であると思われる。

#### 5. まとめ

日本語と中国語において、キーワード主導型の類似用例選択手法の評価を行った。音声認識誤りを含む文のみを入力とした場合、日本語で

は、キーワード認識率が 61.51%で、再現率 57.95%達成できることを示した。また、本方式によりキーワード誤りが半分以下であるときには、日本語において 83.23%の再現率を得ることも示した。中国語においても同様の結果であった。

一方、適合率に関しては、キーワード間の依存関係を用いることにより、過剰な用例選択を防止し、日本語においては、用例候補数を 84%、中国語において 26%削減でき、適合率を改善できること示した。しかし、依存関係を用いることにより、再現率が低下するので、今後は再現率を低下させない方法を検討する必要がある。

今回は、評価文として **closed** な文章を用いて、音声認識誤りに対する頑健性を示したが、**open** な文章に対する実験を行い、本手法による類似用例選択が機能するかを検証したい。また、日本語と中国語で検証をおこなったが他の言語に関しても検証を行いたい。

#### 参考文献

- [1]佐藤理史,MBT1:実例に基づく訳語選択、人工知能学会誌,Vol.6 No.4,pp592-600,1991
- [2]佐藤理史,MBT2:実例に基づく翻訳における複数翻訳例の組み合わせ利用、人工知能学会誌,Vol.6,No.6,pp861-871,1991
- [3]古瀬蔵,隅田栄一郎,飯田仁,経験的知識を活用する変換主導型機械翻訳、情報処理学会論文誌,Vol.35(3),pp414-425,1994
- [4]隅田栄一郎,古瀬蔵,飯田仁,英語前置詞句係り先の用例主導あいまい性解消、電子通信学会論文誌, D-II, Vol.J77-D-II, pp557-565, 1994
- [5]Y.Wakita,K.Matsui,Y.Sagisaka,Robust speech translation using fine keyword clustering, Workshop on Multi-Lingual Speech,2000
- [6] Y.Wakita,K.Matsui,Y.Sagisaka, Fine keyword clustering using a thesaurus and example sentences for speech translation, ICSLP2000