

# World Wide Web を用いた外国人名の英訳自動獲得

後藤 功雄<sup>†</sup>, 加藤 直人<sup>†</sup>, 田中英輝<sup>†</sup>  
江原 暉将<sup>†</sup>, 浦谷 則好<sup>†</sup>

カタカナで表記された外国人名の英訳語を、関連語をキーワードとする言語横断情報検索と、発音類似性を利用した訳語推定により、World Wide Web (WWW) 文書から獲得する手法を提案する。ニュース記事に出現する人名は新出語であることが多く、対訳辞書に登録されていない場合も多い。提案手法は、カタカナの外国人名が文書中に存在した場合、はじめにその周辺の単語を対訳辞書によって英訳し、これらをキーワードとして英語の WWW 文書検索を行う。次に、検索された WWW 文書中から人名候補となる英単語列を翻字により変換し、発音が類似した英単語列を訳語とする。ニュース記事に出現した外国人名を対象として本手法による実験を行い、有効性を確認した。

## Automatic Acquisition of English Equivalents of Foreign Personal Names Using the World Wide Web

ISAO GOTO,<sup>†</sup> NAOTO KATO,<sup>†</sup> HIDEKI TANAKA,<sup>†</sup>  
TERUMASA EHARA<sup>†</sup> and NORIYOSHI URATANI<sup>†</sup>

This paper proposes a method of acquiring English equivalents of foreign personal names written in katakana characters from the World Wide Web (WWW). In news articles, new foreign personal names appear frequently and are rarely registered in bilingual dictionaries. Our method can automatically obtain the English equivalents of personal names by using two phases: cross-language information retrieval using related words and acquisition of translation based on phonetic similarity. In the first phase, given a katakana foreign personal name appearing in a news article, the method extracts words related to the foreign personal name, translates these words into English using bilingual dictionaries, and retrieves WWW documents in English using the translated words as keywords. In the second phase, our method extracts candidates of English equivalents from the retrieved WWW documents, transliterates the candidates to phonetic expressions, compares them with the phonetic expression of the personal name written in katakana, and obtains the most similar one as the English equivalent. We confirmed the effectiveness of our method with a series of experiments using foreign personal names appearing in news articles.

### 1. はじめに

ニュースは最新の情報を取り扱うため、既存の対訳辞書や対訳記事にない外国人名がしばしば出現する。そのため、ニュースを翻訳する際に翻訳者はこれらの訳語を調査する必要があるが、この調査は時間の

かかる作業となっている。日英翻訳においてこの訳語の調査を効率化するために、カタカナで表記された外国人名の英訳語を、関連語をキーワードとする言語横断情報検索 (Cross-Language Information Retrieval: CLIR) と、発音類似性を利用した訳語推定により、World Wide Web (WWW) 情報から獲得する手法を提案する。

本論文の英訳語取得の問題設定は、検索質問をカタカナで表記された外国の人名とし、その英訳語を WWW 情報から取得するとする。日英対訳辞書や日

<sup>†</sup> NHK 放送技術研究所  
NHK Science and Technical Research Laboratories  
現在, ATR 音声言語コミュニケーション研究所  
Presently with Presently with ATR Spoken Language  
Translation Research Laboratories  
現在, 諏訪東京理科大学システム工学部電子システム工学科  
Presently with Presently with Tokyo University of Science, Suwa

日本語記事を作成する際には外国人名を日本語に翻訳しているので、この情報を保存しておけば改めて英訳を調べる必要はない。しかしながら、実際には、日本語記事を作成することに時間をとられ、このような情報を保存するという作業は行われていない。

英対訳コーパスには検索質問の英訳語は存在しないものとする。これは、異なる言語の情報を取得するので CLIR の手法の 1 つと位置づけられる。CLIR には、過去に多くの研究例がある。これらの研究例と未知訳語の獲得との関係について以下に述べる。

検索質問を検索対象の言語へ翻訳して検索する方式（検索質問翻訳方式<sup>1),2),6),12),13),18</sup>）がある。その翻訳には対訳辞書や機械翻訳、対訳コーパスが利用されているが、検索質問の訳語が未知である場合には対応できない。

逆に検索対象となる文書を検索質問言語へ翻訳して検索する方式（検索対象文書翻訳方式<sup>19)</sup>、検索質問翻訳方式と検索対象文書翻訳方式を組み合わせる方式<sup>17),22)</sup>、検索質問と検索対象文書の両方を単語レベルで中間言語表現に変換して検索する方式<sup>4),8),11)</sup>もある。しかし、いずれの方式も未知訳語を扱うことができない。

翻字により検索質問を検索対象の言語表現に変換して検索する方式<sup>9),14)</sup>がある。また、翻字を利用してカタカナ表現と英語とを対応づける手法<sup>3),16),20)</sup>も提案されており、これらの対応関係も検索質問の変換に用いることができる。翻字を高精度に行うことができれば、キーワード検索だけで訳語を取得できるため、効率的な検索が可能となる。しかし、翻字による変換では正しい訳語が生成されるとは限らない。文献 9) の手法では、自ら収集した文書中に存在する単語にしか変換できない。これは、翻字による変換方法が、被検索文書中に出現した英単語リストからカタカナに対応する英単語を選択する方式であるためである。

発音の類似性を利用して訳語を取得する方式<sup>5),7),15)</sup>が提案されている。この方式は、我々の手法の中でも利用している。翻字により発音表現に変換することで、未知訳語も扱うことができる。しかし、これらの手法は、WWW 文書全体を対象とすると類似度の計算量が多くなってしまふ。特に、1 つの単語ではなく、任意の長さの単語列を比較対象とすると、計算量は多くなる。

また、コンパラブルな報道記事から訳語を獲得する手法<sup>10)</sup>がある。しかし、この手法は単言語の文書から訳語を取得することはできない。

提案手法は、検索質問の周辺に存在する単語を翻訳して作成したキーワードによる検索と、検索質問を翻字により発音表現に変換して行う類似発音表現の探索

という 2 段階の処理を行うことによって、WWW 文書という大きなデータから英訳語を検索する。本手法は、辞書や対訳コーパスに存在しない英訳語を取得することができる。また、外部の WWW 検索エンジンを利用することが可能である。

本論文では、2 章で提案手法について説明する。3 章で実際のニュース文に出現した人名の訳語取得の評価実験と考察について述べ、4 章で本論文についてまとめる。

## 2. 提案する英訳語取得手法

本手法全体の構成を図 1 に示す。

本手法は、次の Phase 1 と Phase 2 の 2 種類の手法を用いて検索を行う。

- Phase 1: 日本語ニュース記事コーパスから検索質問が存在する文書を収集し、その文書中から検索質問の周辺の単語（以下では文脈と呼ぶ）を取得し、その文脈の英訳をキーワードとして WWW 文書を検索する。これによって大量の WWW 文書群の中から検索質問に関連する文書に絞り込む。もちろんキーワードには、検索質問の英訳語は含まれていない。
- Phase 2: 絞り込んだ WWW 文書群から人名の候補となる英単語列を抽出する。抽出した人名候補と検索質問とを翻字により発音表現に変換し、文字単位の類似度計算を行い、英訳語を決定する。ここで、発音表現とは、単語の発音を発音記号により表現したものである。

各 Phase には次のような特徴がある。Phase 1 は、キーワード検索であるため、大量の文書を対象とした検索を効率的に行うことができる。Phase 2 は、文字単位で類似した表現を探索しているため、発音表現が完全に一致しない場合でも英訳語を取得することができる。

以下の節では、Phase 1 と Phase 2 の処理の内容と、Phase 2 で用いる探索アルゴリズムについて述べる。

### 2.1 Phase 1: 検索質問に関連する WWW 文書の検索

はじめにカタカナで表記された外国の人名が検索質問として与えられる。本手法は、まず、検索質問が出現する記事を日本語ニュース記事コーパス中から探して、その記事から文脈として周辺の単語を取得する（図 1 の Phase 1.1）。ただし、現在文脈として用いているのは、肩書きや固有名詞（地名、国名、組織名など）などの人名に共起しやすい語である。さらに、肩書きについては検索質問を含む文内から、固有名詞に

日英対訳コーパスに訳語が存在する場合には、その情報で翻訳支援を行うことができる<sup>21)</sup>。

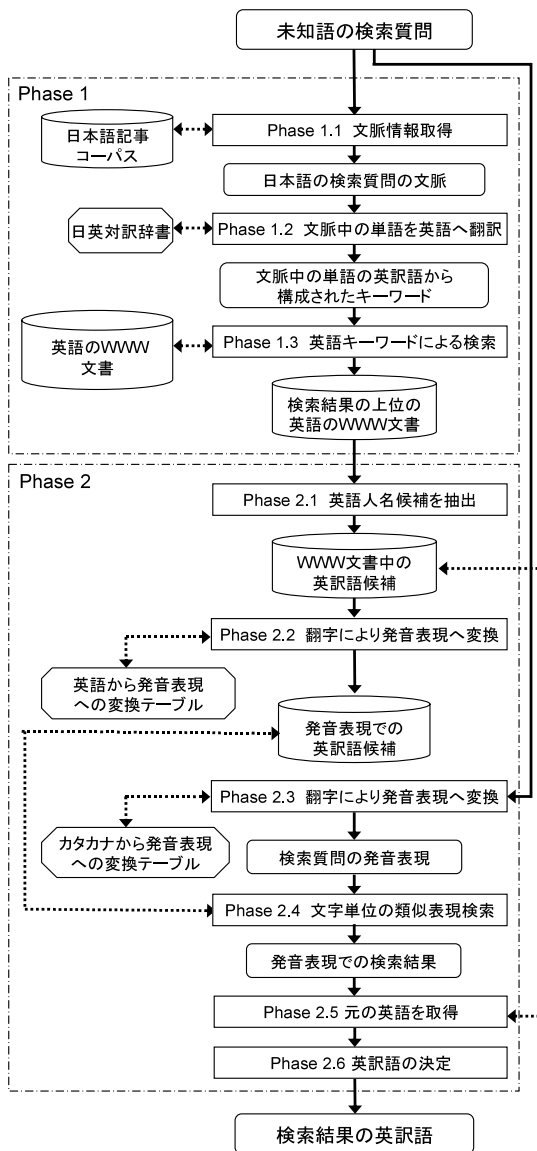


図1 全体の構成

Fig.1 The overall design.

については検索質問を含む記事全体から取得している。次に、取得した文脈を日英対訳辞書により英語に翻訳する(図1のPhase 1.2)。図2に検索質問を「ゴンチグドルジ」とした場合の文脈の例を示す。この例では、同じ文内に出現する肩書き「前国会議長」と、同じ記事内に出現する固有名詞「モンゴル」、「民主党」が取得され、日英対訳辞書を使って翻訳される。ここで、肩書きと固有名詞の取得には、それぞれ肩書きの対訳辞書と固有名詞の対訳辞書を利用する。

このようにして得られた英単語からヒューリスティクスにより検索キーワードのリストを生成し、検索質

問に関連する WWW 文書を検索する(図1のPhase 1.3)。ここで、利用するヒューリスティクスを説明する。キーワードの関連順位を利用して、検索キーワードリストを生成する。はじめに、次の順に優先規則を適用することにより関連順位を決定する。

- 1) 肩書きを固有名詞より優先する。
- 2) 翻訳元の日本語単語が出現する文中に検索質問が存在する場合は、出現しない場合より優先し、さらに、検索質問との距離が短い単語を優先する。距離は文字数により測定する。
- 3) 翻訳元の単語が出現する文中に検索質問が存在しない場合、第1文の単語を優先し、他の文の単語は検索質問を含む文に近い文に含まれる単語を優先する。第1文を優先するのは次の理由による。ニュースの第1文は、記事全体の概要になっており、ここに出現する単語は、記事全体に関する可能性が高く、そのため検索質問との関連も大きい可能性が高いと考えたためである。

次に、生成したすべての英単語の集合から関連順位の低い単語またはフレーズを順番に削除して単語数の異なる複数の集合を生成し、検索キーワードリストとする。このとき、肩書きの訳語が複数ある場合は、肩書きの訳語を分けて取り扱い、検索キーワードは「肩書きの訳語またはフレーズ1つ+固有名詞の訳語1つ以上」となるようにする。この英単語の集合を検索キーワードとする。

例として、図2の文脈から生成した検索キーワードとキーワード検索により取得した WWW 文書の例を図3に示す。図3の例の検索キーワードのリストでは、検索エンジンへの4つのクエリを示している。なお、取得した WWW 文書は、重複があれば1つにまとめる。検索質問を含む記事が複数存在する場合には、記事ごとにPhase 1.3までの処理を行う。この場合にも、すべての文書中において重複があれば1つにまとめる。

## 2.2 Phase 2: 翻訳による発音表現への変換と発音類似性を利用した訳語抽出

### 2.2.1 Phase 2.1: 英語人名候補の抽出

Phase 1で取得した英語の WWW 文書中から、人名の特徴を持つ英単語列を抽出する。ここで、人名の特徴は、大文字で始まり小文字で終わるとした。さらに、人名の特徴を持つ単語に挟まれている単語も人名の候補として抽出した。また、冠詞が付いた単語やストップワード(It, This など)、Phase 1.2で取得した文脈の英訳語は候補から除いた。

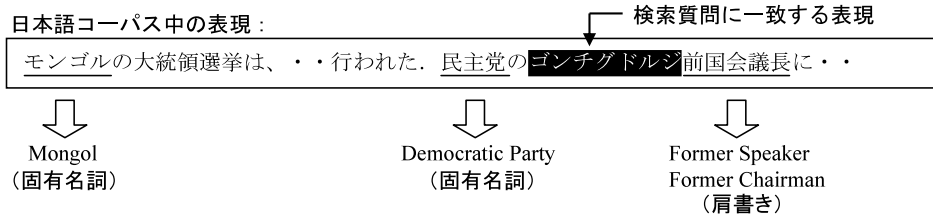


図 2 検索質問の文脈の例  
 Fig. 2 Example of the context of a retrieval query.



図 3 キーワード検索により取得した WWW 文書の例  
 Fig. 3 Example of a retrieved WWW document using keywords.

2.2.2 翻字による発音表現への変換

Phase 2.2, 2.3 では、英語の人名候補とカタカナの検索質問とをそれぞれ翻字により発音表現へ変換する。カタカナの検索質問から発音表現への変換には、訓令式のローマ字表現を用いる。ただし、すべてのカタカナの文字を変換できるように訓令式ローマ字の仕様 (ISO 3602) を一部拡張している。促音は削除して無視した。長音は “\_” で表現した。

一方、英語の人名候補から発音表現への変換には、人手により作成した変換ルールを用いて行う。変換ルールは、変換元の部分文字列、変換先の発音表現文字列、後続制約 (C: 子音, V: 母音, \$: 単語末尾) の 3 つからなる。変換は、英語の人名候補の文字列に対して、部分文字列を照合し、発音表現文字列に変換する。発音表現文字列は複数対応する場合もあるので変換後はラティスで表現される。ここでは、後続制約を使うことにより、訓令式のローマ字表現の仕様に近い発音表現に変換されるようにしている。後続制約は変換元の部分文字列の直後の文字の種類により行う。

変換ルールの例を表 1 に示す。たとえば, Gonchigdorj を発音表現に変換すると図 4 のようになる。

2.2.3 発音類似性を利用した訳語推定

Phase 2.4 から 2.6 では、発音表現に変換された、カ

表 1 英語から発音表現への変換ルールの例

Table 1 Example of conversion rules from English to phonetic expressions.

部分文字列	発音表現文字列	後続制約 (C: 子音, V: 母音, \$: 単語末尾)
g	<i>g, gy, z, zy</i>	V
g	<i>gu</i>	C, \$
gg	<i>g, gy, z, zy</i>	V
gg	<i>gu</i>	C, \$
gh	<i>h, g</i>	V
gh	<i>hu, *</i>	C, \$
ght	<i>t</i>	V
ght	<i>to</i>	C, \$

(\*は発音しないことを示す。大文字小文字の区別は行わない。)

カタカナの検索質問と人名候補の英単語列を比較し、表現間距離を利用して最終的な英訳語を決定する。表現間距離は、発音表現間の編集距離で定義した。発音表現間の編集距離を計算するアルゴリズムについては、2.3 節で述べる。さらに表現間距離には、WWW 文書中における肩書きの出現の有無という情報も考慮した。これは、人名には、周囲に肩書きが出現する可能性が高いという特徴があるからである。そのため、人名の正訳語の周辺には、検索に利用した肩書きの英語が存在する可能性が高いと思われる。そこで同一文中に肩書きが出現していない場合にはペナルティ *P* を編集距離に加算した。*P* は定数であり、後述する実験では、*P* = 0, 0.5, 1.5, 2.5 の 4 通りを比較した。Phase 2.6 では、以上の表現間距離を使って、最終的

ここで用いたルールの部分文字列の異なり数は 103、発音文字列の異なり数は 95 である。

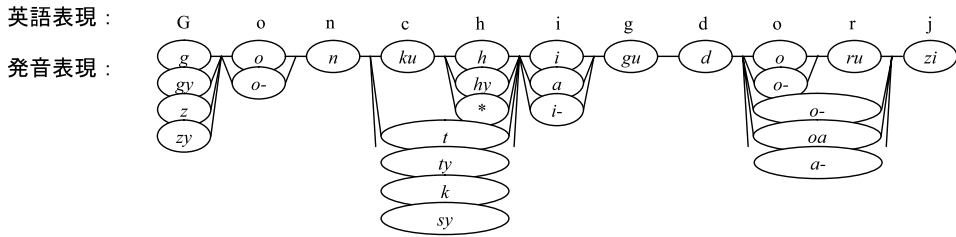


図 4 英語から変換した発音表現のラティスの例

Fig. 4 Example of a phonetic expression lattice converted from an English word.

な訳語を決定する。

### 2.3 発音表現の編集距離が最小となる英単語列の探索アルゴリズム

本節では、発音表現の編集距離を計算するアルゴリズムについて述べる。このアルゴリズムは編集距離が最小になる英単語列を動的計画法により探索するもので、以下の3点に対応できることを特徴とする。1) 比較対象の片方がラティスになっている、2) 編集距離を計算する経路が単語の途中から始まったり終わったりすることがない、3) 訳語候補の単語列内における任意の長さの単語列を対象とする。本アルゴリズムは1次元の表現間の編集距離を計算する一般的な動的計画法による手法を拡張してテーブルを複数段にすることで上記の特徴を実現したものである。

探索アルゴリズム(図5)は、3種類のテーブル( $T_1, T_2, T_3$ )を利用する。 $T_1$ は、英語の単語単位の行を持ち、 $T_2$ は、英語の文字単位の行を持ち、 $T_3$ は、英語から発音表現へ変換する単位の発音表現の行を持つ。列はそれぞれカタカナの発音表現の文字単位となっている。

はじめに、すべての単語の先頭に単語の区切り記号“\_”を付与する。単語列を  $W = w_1 \dots w_n$ 、単語  $w_i$  の文字数を  $l(w_i)$ 、単語中の文字を  $c$  で表し、単語  $w_i$  の文字列を  $w_i = c_1 \dots c_{l(w_i)}$  とし、単語中の文字位置  $p$  から  $q$  までの部分文字列  $c_p \dots c_q$  を  $c_p^q$  で表す。発音表現への変換ルールにより  $c_p^q$  を変換した発音表現の集合を  $S(c_p^q)$  で表し、それらの各発音表現を  $s$  で表す。発音表現  $s$  の文字数を  $l(s)$  とし、発音表現の文字を  $\bar{c}$  で表し、発音表現の文字列を  $s = \bar{c}_1 \dots \bar{c}_{l(s)}$  で表す。単語  $w_i$  の文字位置  $p$  から単語末  $l(w_i)$  までの部分文字列  $c_p^{l(w_i)}$  のうち、文字位置  $p$  から対応する発音表現の集合は  $\cup_{p \leq q \leq l(w_i)} S(c_p^q)$  と表せる。カタ

```

1   $T_1[0, j] \leftarrow j \quad (0 \leq j \leq m)$ 
2   $T_1[i_1, 0] \leftarrow 0 \quad (0 \leq i_1 \leq n)$ 
3  for  $i_1 = 1$  to  $n$  do
4     $T_2[0, j] \leftarrow T_1[i_1 - 1, j] \quad (0 \leq j \leq m)$ 
5    for  $i_2 = 1$  to  $l(w_i)$  do
6      foreach  $s \in \cup_{i_2 \leq q \leq l(w_i)} S(c_{i_2}^q)$  do
7         $T_3[0, j] \leftarrow T_2[i_2 - 1, j] \quad (0 \leq j \leq m)$ 
8         $T_3[i_3, 0] \leftarrow i_3 \quad (0 \leq i_3 \leq l(s))$ 
9        for  $i_3 = 1$  to  $l(s)$  do
10         for  $j = 1$  to  $m$  do
11           if  $\bar{c}_{i_3} = r_j$  then
12              $a \leftarrow 0$ 
13           else
14              $a \leftarrow 1$ 
15           endif
16            $T_3[i_3, j] \leftarrow \min\{T_3[i_3 - 1, j - 1] + a,$ 
17              $T_3[i_3 - 1, j] + 1, T_3[i_3, j - 1] + 1\}$ 
18         end
19       end
20       for  $j = 0$  to  $m$  do
21         if  $T_2[q, j] > T_3[l(s), j]$  then
22            $T_2[q, j] \leftarrow T_3[l(s), j]$ 
23         endif
24       end
25     end
26      $T_1[i_1, j] \leftarrow T_2[l(w_i), j] \quad (1 \leq j \leq m)$ 
27   end
28 return  $\min\{T_1[i_1, m] \mid 1 \leq i_1 \leq n\}$ 

```

図5 発音表現の編集距離が最小となる単語列の探索アルゴリズム  
Fig.5 Search algorithm for a word sequence of the minimum edit distance between phonetic expressions.

カナの発音表現を  $R = r_1 \dots r_m$  とする。 $r$  は、カタカナの発音表現の文字を示す。

ここで、例を示す。 $w_i$  が “George” の場合、 $c_1 = \_$ 、 $c_2 = G$ 、 $c_3 = e$ 、 $c_4 = o$ 、 $c_5 = r$ 、 $c_6 = g$ 、 $c_7 = e$  となり、 $c_4^5 = or$  となる。 $c_4^5 = or$  が、3つの発音表現  $o-$ 、 $oa$ 、 $a-$  に変換される場合は、 $S(c_4^5) = \{o-, oa, a-\}$  となる。発音表現  $s$  が  $oa$  の場合は、 $\bar{c}_1 = o$ 、 $\bar{c}_2 = a$  となる。

以下、図5を用いてアルゴリズムを説明する。英語の単語列とカタカナの発音表現を対応させたテーブル  $T_1$  を作成する。テーブルには、編集距離の初期値

単語間の区切り(カタカナは“\_”，英語はスペース)に、無音を示す発音記号“\_”を割り当てる。英語の単語の区切りは必ずしもカタカナでは“\_”として表現されず省略される場合も多いため、英語の発音記号の“\_”については、削除操作を行っても編集回数を加算しないこととする。

の設定を保存するための行と列を先頭行と先頭列に追加する。この追加した行をダミー行、追加した列をダミー列と呼ぶ。図5では、テーブルの  $[0, \cdot]$  がダミー行を表し、 $[\cdot, 0]$  がダミー列を表している。T1では、ダミー行は、左端の値を0とし、残りは対応するカタカナの発音表現を追加する編集距離とする(図5の1行目)。ダミー列は、すべての値を0とする(図5の2行目)。次に、各英単語とカタカナの発音表現とを対応させたテーブル T2 を作成する。T2 のダミーの行は、1つ前の英語の T1 の行の値を保持する(図5の4行目)。さらに、T2 の英語の文字列から変換ルールにより生成される発音表現のラティスの各ノードとカタカナの発音表現とを対応させたテーブル T3 を作成する。T3 のダミー行は、1つ前の英語の文字の T2 の行の値を保持する(図5の7行目)。距離計算を行う際には、T1 の行から対応する T2 を作成し、さらに、T2 の行に対応する T3 を作成して上の行から下の行へ計算を進める。T3 では、動的計画法により、末尾行までの編集距離の計算を行う(図5の9行目から18行目)。T3 の末尾行の値を対応する T2 の行に入力する。ただし、T2 の行にすでに値が入力されている場合、T3 の値が小さい場合のみ値を入力して更新する(図5の19行目から23行目)。T2 の末尾行まで計算が終了した時点で、T2 の末尾行の値を対応する T1 の行に入力する(図5の26行目)。T1 の末尾行まで値が入力された後で、T1 の末尾列で最も小さい値を最小の編集距離として出力する(図5の28行目)。最小の値となる行に対応する単語が最小の編集距離となる単語列の末尾の単語である。最小の編集距離となる単語列の開始の単語は、経路を保存しておくことで、調べることができる。ただし、図5では、簡略化のために、経路を保持するためのコードと英語の発音記号“\_”の削除操作の例外処理を省略している。

例として英単語列 “Bill Clinton George Bush John Kerry” とカタカナ表現「ジョージ・ブッシュ」との発音表現での編集距離を計算する例を図6に示す。図6では、ダミー行または列に対応する表現として記号“#”を用いている。テーブル間を結ぶ矢印は行単位で値を保存することを示す。T2 は英単語の数だけ生成されるが、例として “\_George” に対応するテーブルを示している。T3 は、発音表現のラティスのノードの数だけ生成されるが、そのうちの “g” に対応する “g”, “gy”, “z”, “zy” の4つのテーブルを示している。この英単語列中の任意の長さの英単語列において、最小の編集距離は1で、最小の編集距離となる英単語列は、経路をたどることで “George Bush” であるこ

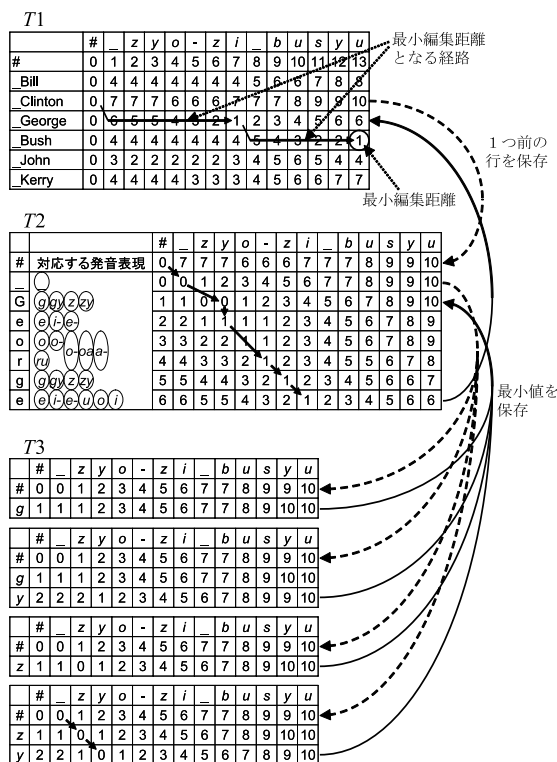


図6 探索アルゴリズムの適用例  
Fig. 6 Example of applying the search algorithm.

とが分かる。

図6の T2 で英語の文字  $c_4^5 = \text{“or”}$  の発音表現  $S(c_4^5) = \{o-, oa, a-\}$  に対応する T3 のテーブルは、直前の英語の文字  $c_3 = \text{“e”}$  に対応する T2 の行がダミー行の値となる。また、これらの T3 の末尾行の値と  $S(c_5) = \{ru\}$  に対応する T3 の末尾行の値の中で、最小値を T2 の  $c_5 = \text{“r”}$  の行に保存する。このように T2 において、T3 の初期値を取得する行と T3 から最小値を保存する行を発音表現のラティスにあわせることによって、本節のはじめに述べたアルゴリズムの特徴1)の「ラティスへの対応」を可能としている。また、T2, T3 の初期値の設定と T1 の末尾列から最小値を取得することにより、アルゴリズムの特徴2)の「経路が単語の途中から始まったり終わったりすることがない」を可能としている。T1 の初期値の設定と T1 の末尾列中の最小値を取得することにより、アルゴリズムの特徴3)の「任意の長さの単語列を対象とする」を可能としている。

### 3. 評価実験と考察

#### 3.1 評価実験の方法

NHK の放送で用いられたニュース記事を利用して

実験を行った。このニュース記事から検索質問と日本語ニュース記事コーパスを、次のようにして選択した。検索質問として 60 人のカタカナで表現された外国人名を用いた。これは、2002 年 9 月 15 日から 2002 年 10 月 14 日までの 1 カ月間の「国際」の区分<sup>1</sup>の日本語ニュース記事から、カタカナ表記の外国人名で、2002 年 9 月 14 日以前の 1 年間に日本語原稿に出現しなかった 60 個である。一方、日本語ニュース記事コーパスとして、2002 年 9 月 15 日から 2002 年 10 月 14 日までの記事を用いた。実際の翻訳者の利用状況に合わせて、はじめに人名が翻訳された記事より後に出稿された記事は含めないこととし、文脈は 1 記事のみから取得した。日英対訳辞書は、規模が大きいものを用意できなかったため、評価実験用に対訳辞書を作成した。すなわち、実験で出現した肩書きと固有名詞に、人手で訳語を与えた。実際には肩書きとして 48 個、固有名詞として 175 個を登録した。ストップワードは EDR 英語単語辞書<sup>2</sup>から作成した。

Phase 1 のキーワード検索では、Google<sup>3</sup>を利用し、検索は 2002 年 11 月 6 日から 7 日に行った。Phase 2 でのペナルティの値は、 $P = 0, 0.5, 1.5, 2.5$  の 4 種類を個別に用いて結果を比較した。また、Phase 1.3 では、検索エンジンに対して指定する WWW 文書取得数の上限を、各検索質問について 500 以上となるように設定した<sup>4</sup>。各検索質問に対する検索結果の上位として実際に取得した WWW 文書の平均数は重複を除いて 275 であった。

### 3.2 評価実験の結果

表 2 に英訳語取得成功率を示す。英訳語取得成功率は、すべての検索質問のうち、正解の訳語を取得できた割合である。取得した単語が訳語の正解例と一致した場合に正解と判定した。ただし、英語圏以外（たとえばロシア語）の人名については、英語表記に揺れがあるので、英訳語の正解例と綴りが完全に一致しない場合でも、同一人物であると判断できた場合には正解とした。 $P = 1.5$  の場合の検索結果の正解率が一番高

表 2 英訳語取得成功率

Table 2 Success rate of acquiring English equivalents.

順位	累積成功率 (%)			
	$P = 0$	$P = 0.5$	$P = 1.5$	$P = 2.5$
1 位	78.3	78.3	85.0	83.3
2 位以上	83.3	85.0	88.3	88.3
3 位以上	85.0	86.7	90.0	88.3
5 位以上	86.7	88.3	90.0	88.3
10 位以上	88.3	88.3	90.0	90.0

く、1 位の正解率は 85%、3 位以内に正解を含む率は 90%であった。また、付録 1 の表 3 に検索質問の番号と用いた人名、その英訳語の正解例、本手法 ( $P = 1.5$ ) により取得した英語中での正解英訳語の順位、英訳語の正解例をキーワードとして Google で検索した際のヒット件数を示す。付録 2 の表 4 に、検索質問の番号と英語キーワード生成に利用した検索質問の文脈を示す。表 4 の文脈は Phase 1.3 の関連順位順になっている。なお、検索を行った時点で、Google は一度に入力できるキーワード数を 10 単語までに制限していたので、英語キーワードで 11 番目以降となった語は、検索に利用していない。

### 3.3 考察

まず、誤りの原因について考察する。原因は大きく分けて 2 種類に分かれる。1 つは Phase 1 で取得した WWW 文書に正解となる英訳語が含まれていない場合である。もう 1 つは、Phase 2 で正しい英訳語が 1 位にならなかった場合である。

Phase 1 では、取得した WWW 文書の中に正解の英訳語が含まれていた率は 90%であった。特に WWW 文書全体の中で訳語を含む文書の数が少ない場合は、訳語を含む記事の取得に失敗している傾向があることが付録 1 の表 3 から分かる。取得に失敗した中には、「という人物」と表記されていて肩書きが存在しないものが 1 件あった。

Phase 2 での  $P = 1.5$  の場合の検索結果について述べる。Phase 1 で取得した WWW 文書の中に正解が含まれている場合では、Phase 2 で 1 位が正解ではなかったものは 3 件であった。これらは、綴りが類似した別人の名前がたまたま文書中に存在したために、それらが 1 位となったものである。しかし、これらの場合でも、正解は 2 位または 3 位という上位にランクされていた。実際の利用を考えると、上位結果とそれらを含む WWW 文書へのリンクを提示することで、翻訳者は、英訳候補とともにリンク先の WWW 文書を参照することができるので、訳語が上位解に含まれれば、翻訳支援には有効であると考えられる。また、 $P = 0$  の実験と比較すると、6 件の検索質問 (オジャ

<sup>1</sup> NHK では、「政治」「経済」「社会」「国際」「地域」「スポーツ」に分類している。

<sup>2</sup> 次の品詞の単語を用いた。基数詞、序数詞、人称代名詞、疑問代名詞、指示代名詞、不定代名詞、関係代名詞、be 動詞、形容詞、関係副詞、疑問副詞、副詞小辞、接続副詞、指示限定詞、不定限定詞、冠詞、助動詞。

<sup>3</sup> <http://www.google.com> 検索エンジンへのアクセスには、Google APIs service を利用した。

<sup>4</sup> 同じ検索質問についての各検索キーワードで取得する文書数の上限は同じとした。検索エンジンからの取得単位が 10 であったので、この上限の数は 10 の倍数で検索キーワードの数 × 上限の数が 500 以上となる最小の値を用いた。

ラン, キム・ケグアン, クアシ, サレハ, ドウドウ, バク・ミョンチョル) について, 正解の順位が上がり, 1 件の検索質問 (イ・ヨンテク) について順位が下がった. 下がった原因は, 日英対訳辞書の登録不足によるものであった. これにより, 検索された英語の文書中の肩書きの有無も利用することが有効であることが分かる. また, Phase2.1 で英訳候補から除いた文脈中の訳語が人名の英訳語と偶然一致すると, 英訳語取得失敗の原因となるが, 今回の実験では, 偶然一致した例はなかった.

次に, 日英対訳辞書による影響について考察する. Phase 1 では, 日英対訳辞書の登録語によって, 検索される WWW 文書が異なってくることもある. たとえば, イラクの「共和国防衛軍」の訳語 “Republican Guard” が登録されている場合と, その訳語が未登録で「防衛軍」の訳語 “defense forces”, “security forces” が登録されている場合とでは, 検索される文書は異なる. 後者の場合は, 検索質問の文脈と検索された WWW 文書の内容の一致度合いが低くなり, 英訳語を取得できない可能性が高くなると思われる. しかし, このような場合でも, ほかに「イラク」などの語もあれば, これらを利用することで, 訳語を取得できる可能性がある. また, 今回は実験に合わせて日英対訳辞書を作成したが, 一般的には大規模な対訳辞書を作成しておく必要がある. たとえば, EDR 日英対訳辞書 V1.5 を利用することが考えられる. しかしながら, 本実験の日英対訳辞書と EDR 日英対訳辞書を比較すると, 肩書きで 48 語のうち 22 語しか, 固有名詞で 175 語のうち 91 語しか共通に存在したにすぎず, EDR 日英対訳辞書でも十分とはいえない. そこで, たとえば, 過去の対訳記事から辞書を自動的に構築することを考えている.

さらに, WWW の利点についても考察する. WWW のデータは規模が大きい, それに比べて通信社や新聞社の英語記事は規模が小さい. そのため, WWW を用いると, 通信社や新聞社の英語記事データに含まれていない英訳語を取得できる可能性がある. 参考までに, 英語の新聞記事から Phase 2 の手法を用いて訳語を取得する実験も行った. 利用したデータは新聞の Daily Yomiuri で, 期間は 2002 年の 9 月と 10 月の 2 カ月間, 記事数は 1,498 である. 3.1 節の実験と同じ 4 通りの  $P$  の値を用いた. これらの記事に訳語が含まれていたのは 7 件 (12%) であった. この 7 件の訳語は  $P$  の値にかかわらず, Phase 2 の手法ですべて 1 位となった. この結果と比較すると, WWW の利用が有用であるといえる.

### 3.4 おわりに

本論文では, カタカナで表記された外国人名の英訳語を WWW 文書から取得する手法を提案した. 本手法は, 文脈から生成した英語キーワードによる言語横断検索と, 翻字により変換した発音表現での発音類似性を利用した訳語推定を組み合わせている. 本手法による実験を行い, 有効性を確認した.

本手法は, 文脈を翻訳する際に日英対訳辞書を利用するので, 実際に翻訳支援に活用していくためには辞書の整備を進めていく必要がある. さらに, 本手法は, 検索対象を人名としたが, 会社名にも本手法を応用することができる. この場合は「株式会社」のようなキーワードが利用できる. また, 検索質問を翻字により変換することでも単言語のデータを検索することが可能であるので, このような方法と本提案手法との両方の訳語取得結果を提示することで, より検索カバレッジが広い英訳語取得を実現できると思われる. これらの課題への対処や検索質問を翻字により変換する方式との併用による翻訳支援への効果については, 稿を改めて報告したい.

### 参 考 文 献

- 1) Ballesteros, L. and Croft, W.B.: Phrasal translation and query expansion techniques for cross-language information retrieval, *Proc. 20th SIGIR Conf.*, Philadelphia, PA, USA, pp.84-91 (July 1997).
- 2) Ballesteros, L. and Croft, W.B.: Resolving ambiguity for cross-language retrieval, *Proc. ACM SIGIR Conf.*, Melbourne, Australia, pp.64-71 (Aug. 1998).
- 3) Brill, E., Kacmarcik, G. and Brockett, C.: Automatically Harvesting Katakana-English Term Pairs from Search Engine Query Logs, *Proc. 6th Natural Language Processing Pacific Rim Symposium*, Tokyo, Japan, p.393-399 (Nov. 2001).
- 4) Carbonell, J.G., Yang, Y., Frederking, R.E., Brown, R.D., Geng, Y. and Lee, D.: Translingual information retrieval: A comparative evaluation, *Proc. IJCAI*, Nagoya, Japan (Aug. 1997).
- 5) Chen, H.H., Huang, S.-J., Ding, Y.-W. and Tsai, S.C.: Proper Name Translation in Cross-Language Information Retrieval, *Proc. 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics*, Quebec, Canada, pp.232-236 (Aug. 1998).
- 6) Chen, H.H., Bian, G.-W. and Lin, W.C.: Re-



- solving translation ambiguity and target polysemy in cross-language information retrieval, *Proc. 37th Annual Meeting of the Association for Computational Linguistics*, Maryland, USA, pp.215–222 (June 1999).
- 7) Collier, N.H. and Hirakawa, H.: Acquisition of English-Japanese proper nouns from noisy-parallel newswire articles using Katakana matching, *Proc. 4th Natural Language Pacific Rim Symposium*, Phuket, Thailand, pp.309–314 (Dec. 1997).
  - 8) Dumais, S.T., Letsche, T.A., Littenman, M.L. and Landauer, T.K.: Automatic cross-language retrieval using latent semantic indexing, *Proc. AAAI Spring Symposium on Cross-Language Text and Speech Retrieval*, Stanford, USA (Mar. 1997).
  - 9) Fujii, A. and Ishikawa, T.: Japanese/English cross-language information retrieval: Exploration of query translation and transliteration, *Computers and the Humanities*, Vol.35, No.4, pp.389–420 (Nov. 2001).
  - 10) Fung, P. and Yee, L.Y.: An IR Approach for Translating News Words from Nonparallel, Comparable Texts, *Proc. 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics*, Quebec, Canada, pp.414–420 (Aug. 1998).
  - 11) Gilarranz, J., Gonzalo, J. and Verdejo, F.: An approach to conceptual text retrieval using the EuroWordNet multilingual semantic database, *Proc. AAAI Spring Symposium on Cross-Language Text Retrieval*, Stanford, CA, USA (Mar. 1997).
  - 12) Hull, D.A.: Using structured queries for disambiguation in cross-language information retrieval, *Proc. AAAI Spring Symposium on Cross-Language Text and Speech Retrieval*, Stanford, CA, USA (Mar. 1997).
  - 13) Jang, M.G., Myaeng, S.H. and Park, S.Y.: Using mutual information to resolve query translation ambiguities and query term weighting, *Proc. 37th Annual Meeting of the Association for Computational Linguistics*, Maryland, USA, pp.223–229 (June 1999).
  - 14) Jeong, K.S., Myaeng, S.H., Lee, J.S. and Choi, K.S.: Automatic identification and back-transliteration of foreign words for information retrieval, *Information Processing and Management*, Vol.35, issue 4, pp.523–540 (July 1999).
  - 15) Kang, B.J. and Choi, K.S.: Two Approaches for The Resolution of Word Mismatch Problem Caused by English Words and Foreign Words in Korean Information Retrieval, *International journal of computer processing of oriental language*, Vol.14, No.2, pp.109–131 (June 2001).
  - 16) Knight, K. and Graehl, J.: Machine transliteration, *Computational Linguistics*, Vol.24, No.4, pp.599–612 (Dec. 1998).
  - 17) McCarley, J.S.: Should we translate the documents or the queries in cross-language information retrieval?, *Proc. 37th Annual Meeting of the Association for Computational Linguistics*, Maryland, USA, pp.208–214 (June 1999).
  - 18) Nie, J.Y., Simard, M., Isabelle, P. and Durand, R.: Cross-language information retrieval based on parallel texts and automatic mining of parallel texts from the web, *Proc. 22nd ACM SIGIR Conf.*, Berkeley, CA, USA, pp.74–81 (Aug. 1999).
  - 19) Oard, D.W.: A comparative study of query and document translation for cross-language information retrieval, *Proc. 3rd Conf. of the Association for Machine Translation in the America*, Langhorne, Pennsylvania, USA (Oct.1998).
  - 20) Qu, Y., Grefenstette, G. and Evans, D.A.: Automatic Transliteration for Japanese-to-English Text Retrieval, *Proc. 26th Annual International ACM SIGIR Conf.*, Toronto, Canada (July 2003).
  - 21) 熊野 正, 後藤功雄, 田中英輝, 浦谷則好, 江原暉将: 翻訳用例提示システムの設計・開発・運用, 信学論 (D-II), Vol.J84, No.6, pp.1175–1184 (2001).
  - 22) 藤井 敦, 石川徹也: 質問翻訳と文書翻訳を統合した日英言語横断情報検索, 信学論 (D-II), Vol.J84, No.2, pp.362–369 (2001).

## 付 録

表 3 検索質問の人名

Table 3 Personal names of the query.

No	検索質問	英訳語の正解例	英訳語の正解 順位 ( $P = 1.5$ )	正解例の ヒット件数
1	アミル・サーディ	Amir al-Saadi	1	533
2	アン・サンヨン	Ahn Sang Young	1	149
3	アンドリュウ・ファストウ	Andrew Fastow	1	21800
4	イ・ヨンテク	Lee Yun Taek	3	134
5	ウダイ	Uday	1	37300
6	ウッドロー・ウィルソン	Woodrow Wilson	1	239000
7	エブダネ	Ebdane	-	2470
8	オジャラン	Ocalan	1	27400
9	カルロス・フエンテス	Carlos Fuentes	1	16900
10	キム・イルチョル	Kim Il Chol	1	626
11	キム・ケグァン	Kim Gye Gwan	2	1000
12	キム・ソンジェ	Kim Song Chae	1	2
13	キム・ソンチョル	Kim Song Chol	-	89
14	キム・ヨンイル	Kim Yong Il	2	414
15	クアシ	Kouassi	1	3070
16	クシュナハン	Cushnahan	1	2030
17	クルト・ビュートリッヒ	Kurt Wuethrich	1	610
18	ゲオルギエフスキ	Georgievski	1	10900
19	コシュトニツァ	Kostunica	1	35900
20	コルシノ	Corsino	1	1890
21	サレハ	Saleh	1	87800
22	シドニー・ブレナー	Sydney Brenner	1	4950
23	ジミー・カーター	Jimmy Carter	1	423000
24	シャタロフ	Shatalov	1	3270
25	ジャビル	Jaber	1	36800
26	ジュリンダ	Dzurinda	1	7010
27	ジョゼ・セーラ	Jose Serra	1	9240
28	ジョン・サルストン	John Sulston	1	3550
29	ジョン・フェン	John Fenn	1	1600
30	スダルマン	Sudarman	1	263
31	スティーブン・パーティニ	Steven Berinti	-	6
32	ダイ・バクティアル	Da'i Bachtiar	1	989
33	ダニエル・カーネマン	Daniel Kahneman	1	4110
34	チェ・スホン	Choe Su Hon	1	276
35	チャールズ・ロバート・ジェンキンス	Charles Robert Jenkins	1	511
36	ドヴィルパン	de Villepin	1	5400
37	ドウドゥ	Doudou	1	9410
38	ドラノエ	Delanoe	1	3830
39	バーノン・スミス	Vernon Smith	1	6140
40	パク・ソンスク	Park Sun-sook	1	504
41	パク・ミョンチョル	Pak Myong Chol	1	84
42	バラシンハム	Balasingham	1	9000
43	ファデーエフ	Fadeev	1	4510
44	フェドートフ	Fedotov	1	9570
45	フワイシュ	Huweish	1	77
46	マッカーサー	MacArthur	1	329000
47	マトリ	Matori	1	1940
48	マヘンドラ	Mahandra	1	222
49	ムバリス	Mubarez	-	101
50	メチアル	Meciar	1	16700
51	ヤファイ	Yafai	1	2050
52	ヤン・ビン	Yang Bin	1	8610
53	ラプス	Labus	1	7710
54	リ・ソンチョル	Ri Song Chol	-	9
55	リ・ミョンウォン	Ri Myong Won	-	5
56	リップポネン	Lipponen	1	8400
57	ルーラ・ダシルバ	Lula da Silva	1	15800
58	レイモンド・デイビス・ジュニア	Raymond Davis Jr.	1	915
59	レザ・ハタミ	Reza Khatami	1	966
60	ロバート・ホルビッツ	Robert Horvitz	1	15500

表 4 利用した検索質問の文脈

Table 4 Used context of the query.

No	肩書き	固有名称
1	大統領 首相	フセイン イラク バグダッド イギリス プレア 国連 国際原子力機関
2	市長 団長	プサン 韓国 北朝鮮 朝鮮民主主義人民共和国 マンギョンボン号 ウォンサン
3	元 C F O	エンロン C F O アメリカ F B I 連邦捜査局 司法省 アンダーセン ワールドコム タイ
4	会長 体育指導委員会委員長	オリンピック 韓国 北朝鮮 プサン 朝鮮民主主義人民共和国 ソウル ビョンヤン 中国
5	大統領	フセイン イラク 国連 アメリカ サプリ オマーン
6	大統領	ノーベル平和賞 アメリカ プッシュ イラク オスロ ノルウェー カーター・センター アトランタ エジプト イスラエル
7	長官	フィリピン イスラム アブ・サヤフ ミンダナオ サンボアンガ アメリカ オサマ・ビンラディン
8	党首 兵士	クルド トルク EU ヨーロッパ連合
9	作家	アラブ イタリア シチリア パレルモ
10	人民武力相 国防相 最高人民会議常任委員長 大統領 総書記	韓国 ビョンヤン 朝鮮民主主義人民共和国 キム・ジョンイル 北朝鮮 キム・ヨンナム キム・デジュン EU ヨーロッパ連合 スウェーデン ペーション インドネシア
11	外務次官 国務次官補 総書記 報道官	クレー 朝鮮民主主義人民共和国 北朝鮮 アメリカ プッシュ パウチャー 韓国 ソウル ビョンヤン
12	体育指導委員会委員長	韓国 北朝鮮 プサン 朝鮮民主主義人民共和国 韓国 ビョンヤン ソウル
13	副部長	韓国 北朝鮮 チャン・ボンリム
14	外務次官 国防相 人民武力相 最高人民会議常任委員長	ビョンヤン キム・ヨンナム 北朝鮮 朝鮮民主主義人民共和国
15	国防相	コートジボワール アフリカ アビジャン ブアケ ケイ
16	選挙監視団長 大統領	EU イスラマバード ムシャラフ ヨーロッパ連合 パキスタン
17	首相 党首	スイス アメリカ ノーベル化学賞 スイス連邦工科大学 バージニア・コモソウェルズ大学 島津製作所 スウェーデン王立科学アカデミー
18	大統領 副首相	マケドニア アルバニア ユーゴスラビア NATO 北大西洋条約機構
19	所長	セルビア ユーゴスラビア
20	大統領	北朝鮮 W F P 北京 国連 世界食糧計画 朝鮮民主主義人民共和国 ビョンヤン
21	大統領	イェメン シラク フランス アルジャー ランブール
22	元大統領 大統領	カリフォルニア アメリカ イギリス ケンブリッジ サンガー マサチューセッツ工科大学 ノーベル医学・生理学賞 エイス スウェーデン カロリンスカ ストックホルム
23	首相 国防相	ノーベル平和賞 アメリカ ジョージア ニカラグア ベル レーガン イラク プッシュ キューバ イスラエル エジプト キャン・デービット 北朝鮮 朝鮮民主主義人民共和国 キム・イルソン
24	首相	北オセチア共和国 イタリアス通信 ロシア 非常事態省 ウラジカフカス プーチン
25	国防相	クウェート イラン イラク アメリカ テヘラン シヤムハニ 国連
26	首相	スロバキア NATO 北大西洋条約機構 EU ヨーロッパ連合
27	候補 名誉総裁 大統領	社会民主党 カルドーゾ アルゼンチン ブラジル
28		サンガー ケンブリッジ カリフォルニア アメリカ イギリス マサチューセッツ工科大学 ノーベル医学・生理学賞 エイス スウェーデン カロリンスカ ストックホルム
29	臨時代理大使 大使	アメリカ ノーベル化学賞 スイス バージニア・コモソウェルズ大学 スイス連邦工科大学 島津製作所 スウェーデン王立科学アカデミー
30		インドネシア バリ
31	国家警察長官	ケネス LA アメリカ ロサンゼルス ロングビーチ
32	教授	インドネシア バリ デンパサル ダウナー オーストラリア アメリカ名誉領事館 クアビーチ
33	外務次官	プリンストン大学 アメリカ ノーベル経済学賞 ジョージメソン大学 スウェーデン ストックホルム
34	兵士	北朝鮮 ニューヨーク 国連 キム・ジョンイル ロシア 朝鮮民主主義人民共和国 アメリカ イラク
35	外相	アメリカ 北朝鮮 国防総省 ニューヨーク
36	内相	フランス イラク 安保理 国連 安全保障理事会 アメリカ プッシュ イウノフ 中国 ロシア サプリ アラブ エジプト マーヘル ニューヨーク
37	市長	アビジャン コートジボワール アフリカ ブアケ ケイ
38	教授	パリ フランス シラク
39	報道官 大統領	ジョージメソン大学 プリンストン大学 アメリカ ノーベル経済学賞 スウェーデン ストックホルム
40	体育指導委員会委員長 文化観光相	韓国 北朝鮮 朝鮮民主主義人民共和国 ビョンヤン 朝鮮半島
41	代表	北朝鮮 韓国 プサン 朝鮮民主主義人民共和国 ビョンヤン ソウル
42	鉄道相	タイ タミル スリランカ シンハラ ベイリス
43	外務次官	ロシア ビョンヤン 北朝鮮 韓国 ウラジオストク シベリア鉄道 鉄道省 ソビエト 朝鮮半島 ハサン
44	副首相	国連 ロシア インタファクス通信 イラク 外務省 フランス 安保理 アメリカ イギリス
45	司令官 最高司令官	イラク 国連 アメリカ 国防総省
46	国防相	連合国軍総司令部 GHQ 外務省
47	元国王 国王	インドネシア バリ アルカイダ
48	情報文化副大臣	キャネンドラ カトマンズ ネパール
49	首相	アフガン アフガニスタン アメリカ ニルファ タイ カルザイ タリバン NGO ホナバード クンドゥズ カブール
50	交通海運相	スロバキア NATO 北大西洋条約機構 EU ヨーロッパ連合
51	行政長官	イェメン フランス 外務省 イラン サヌア アッシュ・シール ランブール
52	候補 副首相 大統領	オランダ 中国 シニジュ ビョンヤン 北朝鮮 朝鮮民主主義人民共和国 アメリカ
53	総書記	セルビア ユーゴスラビア
54	団長	チャン・ボンリム 韓国 北朝鮮 朝鮮民主主義人民共和国 キム・ジョンイル
55	首相	北朝鮮 韓国 プサン 朝鮮民主主義人民共和国 マンギョンボン号 ウォンサン
56	候補 名誉総裁 大統領	フィンランド ヘルシンキ
57	名誉教授	カルドーゾ 社会民主党 アルゼンチン ブラジル
58	国会副議長 大統領 副議長	ペンシルベニア アメリカ ノーベル物理学賞 ニュートリノ 超新星
59		イラン アメリカ イラク 国連
60		マサチューセッツ工科大学 サンガー ケンブリッジ カリフォルニア アメリカ イギリス ノーベル医学・生理学賞 エイス スウェーデン カロリンスカ ストックホルム

(平成 17 年 4 月 18 日受付)

(平成 17 年 12 月 2 日採録)



後藤 功雄 (正会員)

1995 年早稲田大学理工学部電気工学科卒業。1997 年同大学院修士課程修了。同年日本放送協会に入局。1999 年より同放送技術研究所に勤務。2004 年より ATR 音声言語コミュニケーション研究所に出向、現在に至る。自然言語処理、情報検索、機械翻訳の研究に従事。電子情報通信学会、言語処理学会各会員。



加藤 直人 (正会員)

1986 年早稲田大学理工学部電気工学科卒業。1988 年同大学院修士課程修了。同年日本放送協会に入局、同放送技術研究所に勤務。1994 年より 3 年間 ATR 音声翻訳通信研究所に出向。2003 年より ATR 音声言語コミュニケーション研究所に再び出向、現在に至る。博士 (情報科学)。機械翻訳、対話処理、音声言語処理、自動要約の研究に従事。電子情報通信学会、言語処理学会各会員。



田中 英輝 (正会員)

1982年九州大学工学部電子工学科卒業。1984年、同大学院修士課程修了。博士(工学)。同年NHK入局。1987年より放送技術研究所勤務。ATR音声言語コミュニケーション研究所等への出向を経て、現在、NHK放送技術研究所(人間情報科学)、主任研究員。機械翻訳、自動要約等の研究に従事。言語処理学会、映像情報メディア学会、ACL各会員。



浦谷 則好 (正会員)

1975年東京大学大学院修士課程(電気工学)修了。同年NHKに入局。富山放送局を経て、1979年から放送技術研究所にて、情報検索、ニュース文の機械翻訳、テレビ受信機インターフェースの研究に従事。この間、計4年間ATRに出向し、音声翻訳の研究に従事。現在、放送技術研究所(人間情報科学)主任研究員。工学博士。



江原 暉将 (正会員)

1967年早稲田大学理工学部電気通信工学科卒業。同年NHKに入局。主として放送技術研究所に勤務。2003年より諏訪東京理科大学教授。自然言語処理、音声情報処理、Eラーニング等の研究に従事。電子情報通信学会、映像情報メディア学会、言語処理学会、機械翻訳協会各会員。工学博士。