

TR-I-0151

日英対訳対応データ  
Corresponding Japanese and English Text  
for Linguistic Database II

幸山 秀雄  
Hideo Kohyama

1990.3

概要

収集した日英それぞれの会話に対応づけをし、データベースで検索出来ることは、対訳部分の研究者にとっては非常に有意義なことである。机上で少ない例文を考えるより、幅広く実際の会話から用例を得ることができるからである。

本レポートは日英対訳対応データの基本的な仕様の枠組みと、おおまかな特徴の分析について述べたものである。なお日英対訳対応データの仕様の詳細については、TR-I-0152をご覧頂きたい。

0. はじめに .....	1
1. 目的 .....	1
2. 基本的な使用方法 .....	2
(1)検索方法 .....	2
(2)ソーステキスト .....	4
3.対応づけの枠組み .....	5
4.日英対訳対応データの特徴の分析 .....	9
(1)意識の機能的分類 .....	9
(2)述語出現頻度 .....	13
5.おわりに .....	31

## 0. はじめに

ATRでは大規模な言語データベースの構築を進めている。この言語データベースは次のような工程で作業がおこなわれている。

日英テキスト収集 → 日本語形態素解析 → 日本語係受け解析 →  
日英対訳対応 → データベース

日英対訳対応データは日本語形態素解析結果と格係受け解析結果、英語会話をもちいて作成される。ここでは日英対訳対応データの使われかた、日英対応の方法及び特徴分析について述べる。

## 1. 目的

日本語文から英語文に翻訳するためにはそれらの間に何らかの対応がなければできない。一言に対応と言っても構文的な対応、意味的な対応、推論などを介した対応などいろいろな対応が考えられる。日英対訳対応データでは構文的、意味的な対応をとる。文中の各単位(単語や格など)で、対応が取れない日英の用例文は現在の機械翻訳システムでは用例文と同じ対訳を生成できないといっても過言ではない。このような対応が取れない用例もさらに高度な翻訳を実現するための研究材料として必要なものである。

また、収集した多くの実際に近い会話文から必要な用例を検索することが出来ることはシステムの研究者(特に変換処理部分)には非常に有意義なことである。机上で考える以外の用例を知ることやその頻度的特徴も知ることが出来る。また対訳辞書などをつくる際の基本データともなる。

## 2. 基本的な使用方法

### (1)検索方法

データベース化された日英対訳対応データの基本的な使用方法としては次の3つがある。

①分析用データを得る

表現、対応属性などをキーとして検索し対訳を得る。

②統計情報を得る

表現、対応属性などをキーとして訳語出現頻度を得る。

③変換処理用データを得る

単語や格対応などを検索し、システム用データとする。

日本語表現の対応単位としては次のものがある。

・単語

単語辞書の構築、分析などに使われる。

・文節

英文の省略部分を含めた対応範囲を取り、分析などに使われる。

・格

変換辞書の基礎データ、分析など使われる。

・文

対応文を検索するために使われる。

・発話

文脈をみるために使われる。

また属性としては次の5つが付与されている。

・訳語の文脈依存属性

単語対応で訳語が文脈または言語外知識に依存しているかの属性

・複合名詞属性

複合名詞の対応を識別するための属性

・述語属性

述語を識別するための属性

・補文属性

補文を識別するための属性

・モダリティ属性

モダリティを識別するための属性

このような属性と対応単位から対訳部分を検索するための実際的な基本キーとして次のものがある。

単語

文節

格要素

格構造

文ID

発話ID

文脈依存属性

複合名詞属性

モダリティ属性

述語属性

補文属性

なお日英対訳対応データは日本語会話のある単位と、対応する英語会話の範囲を記述したものである。なお日本語から英語への対応づけであり、英語側から検索するための単位を設定していないため英単語などをキーとして検索しても十分な結果が得られる保証はないことに注意する必要がある。

(2) ソーステキスト

ソーステキストの収集方法としては現在、次の組合せがある。

日本語		発話テキスト		通訳文
	×		×	
英語		キーボード入力テキスト		翻訳文

使用目的によりをソーステキストを選ぶのが賢明である。より自然な会話から分析用データがほしいのなら、発話テキスト-通訳文、発話テキスト-翻訳文で収集したテキストがよいだろうし、システム用のデータがほしいのなら、キーボード入力テキスト-翻訳文で収集したテキストがよいと考えられる。

テキスト収集方法で、比較して日英対応づけが多くつくのはキーボード入力テキスト-翻訳文で収集したテキストである。この理由としては概ね次のことがいえる。

(発話文)

キーボード入力は発話入力と比べて冗長さが少ない。

(翻訳文)

翻訳者の文は通訳者の文と比べて言語構造(おもに格構造)を保存している場合が多い。

### 3. 対応づけの枠組み

データベース化された日英対訳対応データを検索するためには、一貫性のある検索キーが必要である。日英対応づけは対応元である日本語の1つの設定単位に対し、意味的に対応する英語部分の範囲をきめる。

対応づけの仕様は単位の設定とその対応方法からなる。対応元の単位としては、単語、文節、格、文、発話がある。対応の方法は意味的なまとまりを考慮するが、それぞれの単位により対応のさせかたも異なる。それぞれの単位及び属性の対応の概略を次にしめす。

#### (単語)

- ・決定詞は対応範囲に含めない。
- ・訳語の文脈依存の可否を示す属性を付与する。
- ・複合名詞は長単位で取る。

会議は8月5日から8日まで開かれますね。

1        2   3   4   5   6   7   . . .

The Conference is talking place from the 5th to the 8th of August, isn't

1                    7                    4                    3   6                    5                    2

it.

#### (文節)

- ・決定詞も対応範囲に含める。
- ・日本語文で省略され英語文で明示的に表現されているものは修飾関係を考慮して対応範囲とする。

アメリカですと1週間位で着くと思います。

1                    2                    3                    4

I believe that the second announcement of the conference will

4                    3

reach the United States within 1 week.

1                    2

#### (格)

- ・格構造範囲の対応と格要素の対応を取る。

会議のことを新聞で読みました。

< 1                    2                    >

I've heard about the conference in the newspaper.

< 1 2 >

(文)

文を単位にして、ひとまとまりの意味を表す部分に対応させる。

もう一度繰り返しますが、ファックス番号は06、372、6164、もしくは

1

は6127ということですね。

Let me make sure I have those number correctly.

1

063726164 <zero six three seven two six one six four> or 6127 <six

1

one two seven>.

(発話)

1人の発話者の発話に対し対訳に対応させる。

もう一度繰り返しますが、ファックス番号は06、372、6164、もしくは

1

は6127ということですね。

タイアップしたものは来週末までちょっとお送り出来ないんですけれ

1

ども、手続きで十分であろうと思いますので、送らせていただきま

す。

大変、申し訳ありませんでした。

1

Let me make sure I have those number correctly.

1

063726164 <zero six three seven two six one six four> or 6127 <six

1

one two seven>.

I like I say, I'm afraid I won't be able to send a typed copy till the end of

1

next week.

I think my handwritten copy should be adequate, and I'm very sorry

1

about this.

#### ・文脈依存性

単語対応のデータを検索し、システムのデータなどに用いるときに、訳語が文脈などによって選択されているかそうでないかを知っておく必要がある。文脈依存性を調べるためには日本語の語義と英語の語義をそれぞれの辞書を使って比較するのが理想であるが作業コストがかかってしまう。そのため和英辞典を検索し見出し語と訳語が載っていれば文脈依存性がないと判断し、訳語が載っていないければ文脈依存しているとする。見出し語が載っていないければ載っていないという記号を付与する。この結果は人手でチェックされる。

現在は奈良で研究中なのですが

currently doing some work in Nara

#### ・複合名詞の対応

複合名詞は複数の名詞で構成され、名詞連続句も複合名詞として扱う。複合名詞はそれ自体の対応を取るほか、修飾関係を考慮して分解した形態素単位でも取る。ただし固有名詞などの分解しても意味が保証されないものや、分解すると意味が変わるものは対応を取らない。また接頭辞は名詞+接尾辞と分解しないようにする。

○登録/用紙

○月刊/コンピュータ/雑誌

×自動/翻訳/電話/研究/所

×人工/知能(「人工知能」と分解された「人工(の)知能」は意味が違う)

×登録/料

×会議/場

・述語属性

日本語文の述語に属性を付与し対応する英語部分も述語であればそれにも属性を付与する。

・補文属性

日本語側の認定基準としてはつぎのものがある。

・連体修飾節となつてうしろの名詞または準体助詞の「の」にかかる文  
・つぎの語を受ける文

と、て、の、こと、れる、できる、かねる、させる、  
たい、たがる、ほしい、いただく、もらう、やすい、  
にくい、難い

英語側の認定基準としてはつぎのものがある。

・つぎの語が導入される文

that who what when where which how whether if  
~ing for~to

・モダリティ属性

モダリティ表現に対して属性を付与する。

#### 4. 日英対訳対応データの特徴の分析

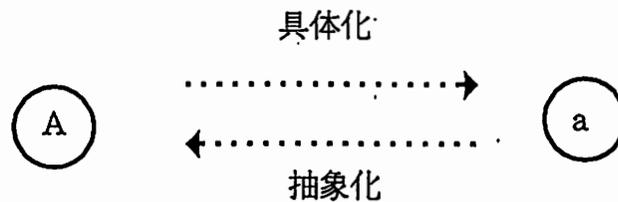
##### (1) 意識の機能的分類

日英で直接対応づかないものをピックアップして分析、分類した。

文がある言語的要素(たとえば名詞、動詞)と関係から構成されると考えると、構文的、意味的に直接対応づけられない文を対応づけるオペレータとして次のものが考えられる。

##### ① 具体化－抽象化

指示内容を具体化(又は抽象化)して翻訳されたもの。

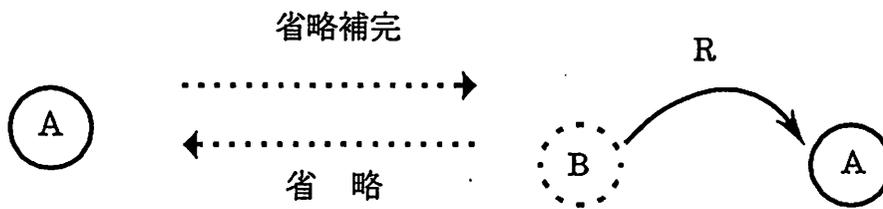


クレジットカードというものは本来外国人のゲストの方、もしくは外国人のスピーカの方にお使いいただければ、支払いがしやすいということで、書いてある項目です。

Well the credit card name and number we hope that it would be used by foreign guests and foreign speakers to the conference.  
It's much easier to pay by credit card.

##### ② 省略－省略補完

日本語文で省略され英語文で補完されている(またはその逆)と解釈できるものの対応。ゼロ代名詞なども含まれる。

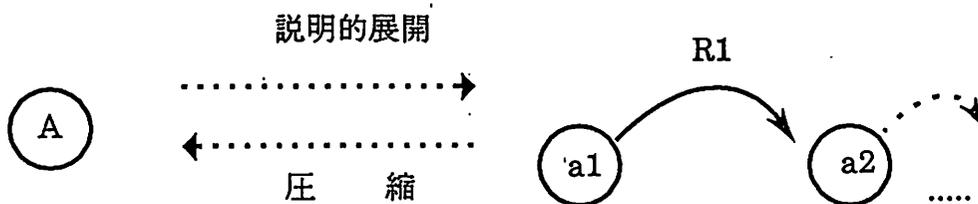


アメリカですと1週間位で着くと思います。

I believe that the second announcement of the conference will reach the United States within 1 week.

③ 説明的展開－圧縮

国語辞書の見出し語とその説明のように対応するもの。

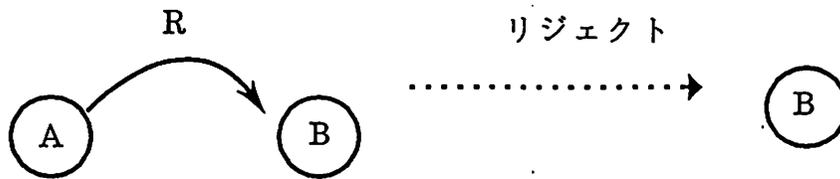


もちろん、専門用語もいくらか出て参りますけれども、....

I'm sure some of the language will be very technical.

④ リジェクト

省略とリジェクトの違いは省略は復元可能であるがリジェクトは不可なことである。



この会議は1988年10月27日から10月30日までの4日間大阪のホテルニューオータニ大阪で開かれます。

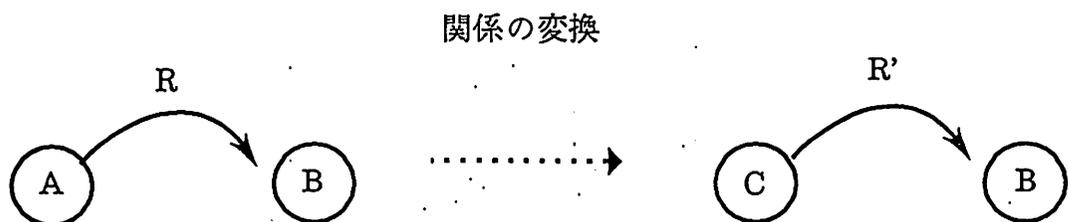
The conference will be held for 4 days from October 27th to the 30th at the Hotel New Otani in Osaka.

一応工学部でそういった関係のことをやっておりますので興味がありました。

I'm in the engineering department and I'm interested in this field.

⑤ 論理の変換

論理を変換させなくては対応が取れないもの。(p(not(a)) → not(p(a))など)



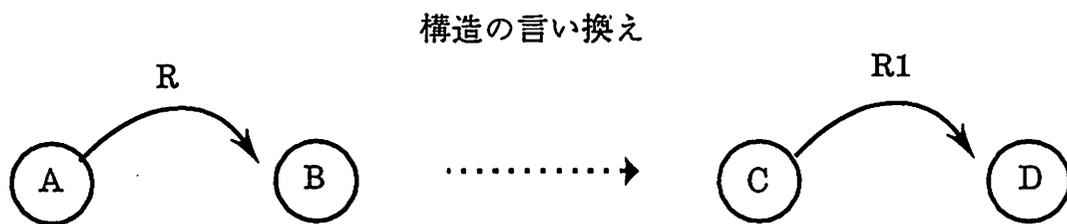
多分そんなことはないと思うんですけれども、万が一ね...

Well I don't think it will happen but if for example...

会議期間中は特にバスを運行しております、土曜ダイヤでバスが運行するようになっておりますから、その際には、ダイヤの方を正確にご覧いただき、まちがった時間表をご覧にならないようご注意ください。

And we are running special buses during the conference from that station and these buses will operate on the Saturday diagram on the timetable. make sure to read the right and correct bus timetable.

### ⑥ 構造レベルの言い換え



これは誰でも参加できるのでしょうか。

the attendance is open to the general public ?

で、返答用紙の方はスピーカとしてご講演なさる方のみ必要ということですので...

The reply form is only necessary if you choose to be a speaker...

## (2)述語出現頻度

現在、国際会議の問い合わせと旅行の問い合わせという2つの領域で会話を収集している。

言語データベースを用いる場合、領域に依存したテキストのもつ特徴を大まかにでも知っておくことは重要である。ここではテキストがどのような特徴をもっているかについて述語の出現頻度によって述べる。

### (方法)

作業手順を以下に示す。

- ①日英対訳対応データ(-注1)に付けられている述語識別子により述語を切出す。
- ②切出した述語の出現頻度をプログラムによりカウントする。
- ③カウントした出現頻度が3以下のものを捨てる。(-注2)
- ④上述①~③の作業をファイル単位を任意に増分しながら行なう。

### (結果)

#### a.電話会話の頻度

それぞれ任意の形態素数(47,105形態素、91,133形態素、125,079形態素)の述語頻度をカウントした。以下にそれぞれの上位20までの結果を示す。また125,079形態素中の上位18述語の対訳および出現頻度も示す。

---

注1)データは正規化されていない検索用データを用いた。(日英対訳対応データにはデータベース用と検索用の2種類のデータフォーマットがある。)

注2)本特徴分析ではおおざっぱな特徴を得るのが目的であり、データ分析作業を軽減するため出現頻度3未満のデータは無視した。

表1. 電話会話の述語頻度

47,105形態素		91,133形態素		125,079形態素	
述語名	出現頻度	述語名	出現頻度	述語名	出現頻度
だ・です	96	だ・です	188	だ・です	287
思う	64	いただく	143	いただく	189
いただく	62	送る	124	送る	164
ある	43	思う	109	思う	147
なる	42	ある	96	ある	129
送る	40	参加(する)	64	なる	110
申す	36	ございます	62	する	84
する	33	できる	62	できる	82
参加(する)	30	する	55	ございます	80
ございます	27	結構(です)	53	参加(する)	73
できる	25	申す	53	よろしい	69
いい	21	よろしい	47	結構(です)	62
わかる	21	わかる	46	願う	58
行く	20	願う	43	行く	52
結構(です)	19	行く	39	いい	50
待つ	18	いい	37	わかる	49
願う	17	出す	32	出る	37
よろしい	16	用意	29	聞く	36

だ

be (3) is (3) It's (1) it's (1) take (1) that's (1) you're (1)

で

am (2) applies (1) are (7) attending (1) be (3) been (1)  
decided (1) had (2) having (1) hold (1) I'm (4) include (1)  
Is (5) is (21) is held (1) It's (2) it's (3) like (1) send (1)  
take place (1) that's (1) you're (1)

でし

am (1) are (2) is (2) possible (1)

でしょ

are (2) be (1) calling (1) cost (2) Is (7) is (28) mean (2)  
need (1) send (1) take place (1)

です

am (1) are (4) are listed (1) be (4) be held (2) call (1)  
come (2) expecting (1) finishes (1) given (1) go on (1) had (1)  
has (2) have (2) I'm (7) informed (1) Is (7) is (87) is held (1)  
It's (1) it's (4) link (1) look at (1) make arrangements (1)  
may (1) mention (1) opens (1) pay (1) pays (1) post-congress  
tour (1) representing (1) send (1) specified (1) take (1) take  
place (1) takes (3) takes place (1) that's (3) there be (1) think (2)  
use (1) used (1) was (2) was programmed (1)

な

accept (1) am (5) are (1) attending (1) be (3) been (1)  
come (1) given (1) going (1) happens (1) I'm (6) is (11)  
it's (1) take (1) that's (1)

ください

can (1) could (2) had (1) have (2) have to (1) need to (1)  
Please (1) please (2) received (2) sent (1)  
we would like you to (1)

ください

can (1) could (1) get (1) have to (2) like (1) need (2)  
please (2) please feel free (1) received (2) should (2) will (1)  
would (2)

ください

accepted (1) accommodate (1) have (2) have to (3) is going to (1)  
need to (1) pay (1) please (1) would (1)

いただけ

be able to (1) can (2) Could (8) could (5) have (1) may (1)  
need (1) Please (1) please (2) receive (1) Would (1) would (1)  
Would you mind (1)

いただける

able to receive (1) are going to (2) are you going (1) be able to (2)  
can (1) could (2) have time (1) Is going to (1) is going to (2)  
issuing (1) receive (1)

いただけれ

could (2) give (1) please (2) send (1) will (1) would (2)

送っ

get (2) include (2) mail (4) make the presentation as speakers (1)  
need (1) obtain (1) received (1) receiving (1) send (15)  
sending (3) sent (7) submit (1)

送ら

belong (1) mail (2) make the payment (1) provide (1) send (11)  
send out (1) sending (2) sent (2) sent back (1)

送り

afford (1) appear (1) bring (1) by (1) distributed (1)  
forward (1) get (2) get off (1) get out (1) give (1) have (1)  
include (1) is (2) issued (1) mail (2) make a decision (1)  
make the application (1) make the payment (1)  
make the payment (1) notify (1) paid (1) pay (1) poster (1)  
Send (1) send (50) send back (3) sending (8) sent (6)  
submission (1) submit (2) transfer (1) was (1)

送る

send (2) sent (1)

送れ

address (1)

送れる

can send (1)

思い

able (1) afraid (5) believe (6) can (1) consider (1)  
distributing (1) expect (2) fear (1) greetings (1) guess (6)  
have (1) hope (4) hopefully (1) hoping (1) is (1) know (1)  
may be (1) presume (1) send (1) should (1) sounds (6) sure (5)  
think (52) think interesting (1) thought (1) wondering (1)

思う

appreciated (1) assume (1) believe (1) decided (1) guess (2)  
hoping (1) I'm afraid (1) imagine (2) like (1) remember (1)  
seems (1) sounds (2) suppose (2) sure (1) think (10)  
understand (2) wish (1) wondering (1)

思っ

assume (1) hope (1) like (1) planning (1) sure (1) think (3)  
thinking (1) wondering (2)

あっ

crop up (1) happens (1) have (3) is held (1) necessary (1)

あり

Are (7) be (3) find (1) had (1) have (25) implies (1)  
involves (1) Is (1) is (9) it's placed (1) last (1) located (2)  
need (1) presents (1) there are (1) there is (2)

ある

are (2) arrange (1) be (5) be held (4) cause (1) enclosed (1)  
give (1) going on (1) had (1) have (22) having (1) holding (1)  
is (10) it's (1) organize (1) run (1) take place (2) there be (3)  
there is (1) there was (1) there're (2) there's (2)

あれ

have (1) there's (1)

なっ

are (1) be (4) being (1) getting back (1) go (1) have (2)  
is (13) It's (2) it's (2) mentioned (1) paid (1) read (1) said (1)  
set up (1) takes (1) were (1)

なら

are (1) allow for (1) be (9) happen (1) happens (1) is (5)  
It's (1) it's (2) means (1) set up (1)

なる

allow (1) be (4) can (1) cost (3) have (1) held (1)  
including (1) is (4) means (1) play (1)

さ

acting (1) be (3) far away (1) give (2) He's (1) involving (1)  
made (1) make (9) making (1) performing (1) would (1)

し

am (1) arranged (1) be (1) bringing (1) do (3) finish (1)  
get (1) give (2) have (4) I'm (2) involved (1) is (1)  
is engaged (1) know (1) Let (1) made (9) paid (1) play (1)  
put (1) set up (1) substitute (1) take care (1) write (1)

する

cost (1) give (1) have (4) is (1) make (1) organize (1)  
planning (1) presenting (1) send (1) take (1)

すれ

do (3) follow (1) get (1) make (1)

せ

do (1) let (1) looking forward to (1) not (1)

でき

able (8) allowed (1) allows (1) available (1) Can (1) can (10)  
compiled (1) confirmed (1) Could (3) could (6) get a chance (1)  
good at (1) have (1) is built (1) May (1) may (1) over (1)  
possible (10) sending (1)

できる

able (2) accept (1) agreeable (1) be able to (1) Can (2) can (15)  
could (2) expressed (1) give (1) have (1) have sufficient time (1)  
it's easy (1) possible (1) willing to undertake (1)

ございまし

are (1) find (1) had (1) have (10) make (1) many (1)  
send (1) there is (1)

ございます

Are (1) are (5) contains (1) find (1) has (1) have (13) Is (8)  
is (6) is located (1) Is there (1) made (1) preparing (1)  
provide (1) running (1) There's (1) there are (1) want (1)

ございませ

be (2) been (1) good (1) have (5) Is (2) is (4) It's (1)

There's (1) there's (1)

参加

attend (6) attendance (1) attending (7) in (1) join (1) make (2)  
participate (5) participation (1) suddenly (1) take part (3)  
the [ uh ] attendance (1)

参加さ

attend (4) attending (3) be (2) coming (1) going (1) make (1)  
participants (1) participate (1) preparing (1) register (1)  
take part (7) take part in (2)

参加し

attend (1) attending (1) come (1) involved (1) participate (5)  
participating (1) represent (1) take part (2)

参加する

attend (1) attending (1) participate (1) participating (2)  
participation (1)

よろしい

accept (2) advisable (2) am I right (1) be able to (1) best (1)  
better (4) can (5) convenient (1) correct (1) could (3) give (1)  
good (1) have (1) have to (6) Is correct (1) is there [ uh uh ] any  
time (1) need (3) O.K. (1) possible (4) right (1) shall (1)  
Should (1) should (15)

よろしかっ

O.K. (1)

よろしけれ

wanna (1)

よろしゅう

accept (2) all right (1) can (1) should (1)

結構

can (5) do (1) fine (5) good (2) great (2) mail (1) no problem  
(1)

結構だ

can (1)

結構で

acceptable (1) be able to (1) fine (1) fine (1)

結構です

accept (1) acceptable (1) all right (1) can (15) could (1)  
doesn't matter (1) fine (5) may (4) no necessary (1) O.K. (4)  
possible (3) welcome (1)

結構な

can (2) convenient (1)

願

ask (1) asked (1) Can (1) can (2) Could (1) count on (1)  
do (1) famous (1) have to (3) hold (1) hope (1) I'm (1)  
like (1) make contact (1) May (2) may (2) must (1) need (1)  
need to (1) needs to (1) Please (2) please (13) please feel free (1)  
representing (1) request (2) required (1) should (1)  
thank you (1) wanna (1) Would (2) would (1)  
would you please (1)

願え

can (1) Could (3) is (1) please (1)

行か

come around (1) go (1) take (1)

行き

come (1) fly (1) go (4) participate in (1)

行く

approach (1) attend (3) be there (1) bound (1) coming (1) get  
(1) go (7) is (1) reach (1) take part (1) transportation (1)  
visit (1)

行け

attend (2) get (2) go (6) participate (1) see (1)

行ける

be able to accompany (1) can get (1) can take part (1) down (1)  
fine (1) get (1) reach (1) take (3)

行っ

been (1) being (1) fastest (1) go (4) informed (2) travel (1)

いい

accept (1) alright (1) are (1) beautiful (1) best (2) better (1)  
can (5) do (1) excellent (1) fine (1) get (1) good (1) great (2)  
have to (1) is (6) it's (2) need (1) nice (1) perfect (1) pick up  
(1) possible (1) right (1) should (13) shown (1) sound more  
like my style (1) the New Otani (1)

わかっ familiar (1) find out (1) heard (1) know (2) understand (1)

わから calling (1) clarify (1) familiar (1) had any idea (1) know (10) sure (1)

わかり asked (1) be decided (1) decided (1) find (1) have (1) have any idea (2) have information (1) know (6) notify (1) read (1) see (1) sure (3) understand (2) understood (1)

わかる find (1) get (1) know (3)

わかん understand (1)

出 attending (1) away (1) come up with (1) go (1) involved (1) is (2) is run (1) leaving (1) left (1) make (1) receive (1) speak and make the presentation (1) start (1) there is (1) was given (1)

出さ be (1) dispatch (1) mail (1) submit (1) submitted (1)

出し include (1) issue (1) provide (1) putting (1) send in (1) sent (1) submit (2)

出す issuance (1) payed (1) present (1) send (1) submit (1)

出せる present (1)

出る attending (1) serve (1) take part (1)

聞い familiar (1) got (1) guess (1) have (1) hear (3) heard (7) information (1) know (1) known (1) mentioned (1) solicit information (1) sponsor (1) talk (1) told (1) understand (2)

聞かせ ask (1)

聞き

ask (2)   contacted (1)   familiar (1)   have (1)   hear (1)   heard (1)  
know (2)   questions (1)   said (1)   starts (1)

聞ける

hear (1)

(特徴)

・出現率

結果をみると会話の中に出現率(頻度数/形態素数)の高い述語と低い述語が存在することは明白である。これは総形態素数が増えるにしたがって出現率の高い述語と低い述語の頻度差がひらくことを意味する。

・領域に依存する述語の頻度

国際会議という領域に依存する述語で12万形態素中、上位20内に入っているのは次の6つである。

送る    参加(する)    行く    用意(する)    出る    聞く

「送る」という述語が多いのは何かを参加者達に送ることは国際会議事務局の基本的な仕事であり、そのため出現率が高いと考えられる。

「願う」「できる」「よろしい」「結構(です)」などのように願望、可能、了解を表わす述語が多いのは問い合わせ文の特徴といえる。

b. キーボード会話の頻度

キーボード会話28,674形態素について述語頻度をカウントした。キーボード会話の述語頻度表とその対訳および出現頻度をつぎに示す。

表2. キーボード会話の述語頻度

述語名	出現頻度
です	170
送る	78
参加(する)	47
いただく	44
できる	39
ありがとう	33
ある	33
なる	30
願う	30
教える	30
する	28
申す	25
下さい	25
思う	23
聞く	21
良い	20
用意	18
待つ	17

だ

are (1)

で

am (1) be (1) comes (1) I'm (1) Is (1) is (16) make (1)  
range (1) were (1)

でし

was (1)

でしょ

are (1) be (3) is (2)

です

am (7) are (4) available (1) be (13) be held (4) cost (2) I'm  
(2) Is (7) is (91) it's (1) last (1) start (1) take (1) takes (1)  
That's (1) was (3)

な

am (2) are (1) I'm (1) is (6) was (1)

おくっ

send (1) send back (1)

おくら

send (1)

おくり

let you have (1) send (3) sending (1)

送くり

informed (1)

送っ

forward (1) get off (2) get off (1) include (1) invoice (1)  
mailed (1) Send (1) send (4) sending (1)

送ら

get (1) get off (1) provide (1) send (2) sending (1) sent (1)

送り

forward (2) get (1) get out (1) know (1) let you have (1) post  
(1) return (1) returned (1) send (40) sending (3) sent (5)  
sent back (1) tell (1)

送る

send (1) submit (1)

送れる

able to send (1) put in the mail (1)

参加

act (1) attend (12) attendance (1) attended (1) attending (6)  
come (1) make (1) participate (6) participation (1) travelling (1)

参加さ

attend (1)

参加し

apply (1) attend (6) go (1) join (2) participate (1)

参加する

attend (1) hold (2) join (2)

いただき

are requested (1) Could (1) get (1) received (1)

いただく

are requested (1) paid (1) Would (1)

いただけ

can (1) could (1) have (1) Please (1) please (2) possible (1)  
would (1)

いただける

can (1) could (1)

いただけれ

could (1)

頂い

received (1)

頂か

should (1)

頂き

have (2) please (2)

頂く

could (1) have to (1)

頂け

Can (3) Could (5) could (4) get hold of (1) Please (1)

頂ける

able(1) arrive(1) can(1) could(1) Will(1)

頂けれ

can(2) please(2) send(1) would(1)

でき

able(5) allowed(1) available(1) be(1) Can(1) can(7)  
Could(2) completed(1) finished(1) May(1) may(1) possible  
(6) there is time(1)

できる

can(6) could(1) may(1) possible(2)

ありがとう

Thank you(27) Thanks(6)

有難う

appreciate(1) Thank you(5) Thanks(1)

あり

be held(1) find(1) has(1) have(12) Is(3) is(3) need(1)

ある

are(1) be(2) give(1) grants(1) have(3) Is(1) is(2)

あれ

have(2)

なっ

be(2) be included(1) charged(1) is(3)

なら

be(1)

なり

amount(1) be(1) is(4)

なる

be(2) is located(1)

願い

ask(1) asked(1) asking(1) Could(2) could(1) delegating(1)  
hope(1) leave in your hands(1) May(4) may(1) please(5)  
prefer(1) recommend(1)

願え

Can(2) Could(2) could(2) like(1) please(2)

教え

advise(1) ask(1) attach(1) furnish(1) give(7) give a  
suggestion(1) have(2) inform(1) know(2) let me know(1)  
Let us know(2) suggest(1) tell(8) told(1)

さ

arrange(1) done(1) have(1) make(2) provide(1)

し

delivered(1) do(1) done(1) give(2) giving(1) have(2)  
help(1) inform(1) make(2) making(1)

する

arrange(1) is run(1) making(2) read(1)

すれ

do(1) take(1)

せ

have(1) kept(1) Let(1) let(1)

せる

have(1)

もうし

am(1) is(5)

申し

is(18)

下さい

Could(6) could(1) hope(1) May(1) may(2) need(1)  
Please(1) please(10) Please feel free(1) Would(1)

思い

assume(1) believe(2) certain(1) know(1) recognize(1) see  
(1) seems(2) suppose(1) sure(3) think(5) thinking(1)  
understand(1)

思う

know(1)

思っ

believe(1) I'm interested in(1) think(1)

思わ

foresee(1) think(1)

聞い

give (1) hear (2) heard (1) speak (1)

聞か

tell (2)

聞かせ

give (1) let me have (1) listen to (1)

聞き

ask (2) give (1) have (1) heard (2) informed (1) know (3)  
make an enquiry (1) mentioned (1) speak (1)

よい

all right (1) enough (1) may (1) need (1) should (3)

よろしい

better (1) can (1) Could (1) May (2)

良い

acceptable (1) convenient (1) could (1) good (3) interesting (1)  
May (1) nice (1) OK (1) shall (1) should (2)

良かつ

Excellent (1)

用い

using (1) utilize (1)

用意

arrange (1) available (1) booked (1) get (1) have (1) having  
(1) offering (1) prepare (3) prepared (2) preparing (1) ready  
(1)

用意し

have (1) prepare (1) prepared (1) provide (1)

持ち

get (1) have (1)

持つ

bring (1) have (3) was (1)

待ち

expecting (1) hoping (1) look forward to (3) looking forward to (8)  
looking forwards to (1) waiting (3)

待つ

expecting (1)

待つ

look forward to (1) wait (1)

(特徴)

電話会話に比べ、形態素数2万8千と総数が少ない割には頻度が高い。「です」などは電話会話の8万形態素相当に値する。キーボード会話は電話会話に比べ、だいたい2倍くらいの述語の数があるといえる。

出現する述語は電話会話と似ている。しかし出現率の高低を述べるには5、6位以下は形態素量が少ないため差がついていないのでなんとも言えない。

5. おわりに

本レポートでは日英対訳対応データの基本的な使用法、対応の枠組み、特徴分析結果について述べた。特徴分析については、もっといろいろな特徴を今後調べてみる必要がある。また頻度的特徴を分析するために言語データベースのデータ量をもっと増やすことも急務である。

(謝辞)

前向きに作業し問題提起してくれた日本アイアール株式会社の衛藤氏に感謝します。また日頃、御指導をいただいている森本室長、熱心に討論してくださったデータ処理研究室と言語処理研究室の諸氏に感謝します。

参考文献

- [1]篠崎ほか 言語データベース作成のための日英対訳対応付け TR-I-0043
- [2]篠崎ほか 形態素情報利用解説書 TR-I-0077
- [3]井ノ上 言語データベース用格・係り受け意味体系 TR-I-0029
- [4]幸山ほか 日英対訳対応データの仕様 TR-I-0152
- [5]橋本 自動翻訳電話研究用言語データベースの構成 TR-I-0150