

TR-I-0152

日英対訳対応データの仕様
Specifications of Corresponding Japanese and
English Text
for Linguistic Database

幸山 秀雄 衛藤 純司*
Hideo Kohyama, Jyunji Eto

1990.3

概要

ATR 自動翻訳電話研究所では現在、言語データベースを構築している。言語データベースは形態素データ、格係り受けデータ、日英対訳対応データの3つから作られる。本稿では、日英対訳対応データの対応づけの仕様について述べる。

©ATR Interpreting Telephony Research Laboratories
ATR 自動翻訳電話研究所

*日本アイアール株式会社

1 対応について	1
2 対応のレベル	1
2-1 単語レベル	1
2-1-1 複合名詞の対応	2
2-1-2 助動詞・補助動詞の対応	3
2-1-3 「願う」の対応	7
2-2 文節レベル	7
2-2-1 決定詞の対応	8
2-2-2 述語文節の対応	8
2-2-3 省略語句の対応	9
2-2-4 文末表現の対応	9
2-3 格レベル	10
2-3-1 日本語の格解析	10
2-3-2 英語との対応	11
2-3-3 補文について	11
2-4 ランダムレベル	14
2-5 文レベル	14
2-6 発話レベル	15
3 属性について	16
3-1 意味属性	16
3-2 モダリティ属性	17
3-3 複合名詞属性	17

1. 対応について

ある日本語の表現とある英語の表現が同じ意味を表しているとき、これら2つの表現は対応しているとみなす。ここで、「同じ意味」とは、日本語の表現が指示している物・事と英語の表現が指示している物・事が同一であることである。

観光 ツアー は 登録料金 には 含まれて おりません。
1 2 3 4 5 6 7 8

The sightseeing tour is not included in the registration fee .
1 2 7 8 6 5 3 4

「同一性」の判断はテキストに即して行う。すなわち、ある日本語の表現とある英語の表現が、調査者の日本語および英語の知識に照らしてかけはなれているように見えようとも、当の発話のなかで同じ物・事を指示しているかぎり、対応させる。

両分野とも ディスカッション の 対象 と なっ ています。
1 2 3

Both areas will be included as the subjects of the conference .
3 2 1

2. 対応のレベル

2-1. 単語レベル

日本語の1単語と英語の1単語以上の表現が対応するとき、深さ1の行に記述する。

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 taking place from the 5th to the 8th of August ,
1 7 4 3 6 5 2
isn't it ?

日本語2語以上がひとまとまりになって、英語の1単語以上と対応するとき、深さ2の行に記述する。(ただし、データベースではランダムレベルに記述される)

<u>木曜日</u> か <u>金曜日</u>	<u>Thursday</u> or <u>Friday</u>
<u>言語学</u> の <u>研究</u>	<u>linguistic</u> <u>research</u>
是非 <u>参加</u> したい	I want to <u>participate</u> in it.
一番 <u>簡単</u> な <u>方法</u>	the <u>easiest</u> way
3人 <u>まとめて</u> <u>支払う</u>	pay for the 3 <u>jointly</u>
会議に <u>興味</u> がある	I <u>am interested</u> in it.

N.B. 単語レベルでは、語彙の対応データを集めることが目的であるので普通名詞であれ固有名詞であれ、定冠詞は対応範囲に含めない。

2-1-1. 複合名詞の対応

複合名詞については、深さ2の行に複合名詞全体の対応を、深さ1の行に構成要素となる個々の単語の対応を記述する。

<u>登録</u> <u>申し込み</u> <u>用紙</u>	<u>registration</u> <u>application</u> <u>form</u>
<u>1</u> <u>2</u> <u>3</u>	<u>1</u> <u>2</u> <u>3</u>
<u>4</u>	<u>4</u>
<u>会議</u> <u>論文集</u>	<u>conference</u> <u>proceedings</u>
<u>1</u>	<u>1</u>
<u>2</u>	<u>2</u>

N.B. 「論文集」と「proceedings」も対応するが、深さ2の行はすでにふさがっているので、記述しない。ただし、個々の単語に分解することが不可能な複合名詞は、全体の対応のみを深さ2の行に記述する。

<u>人工</u> <u>知能</u>	<u>artificial</u> <u>intelligence</u>
<u>1</u>	<u>1</u>
<u>登録</u> <u>料</u>	<u>registration</u> <u>fee</u>
<u>1</u>	<u>1</u>
	<u>fee</u>
	<u>1</u>

cf. 登録料金

1 2
3

registration fee

1 2
3
fee
2
3

名詞連続句で、複合名詞か否かを判断するのが困難なものは、すべて複合名詞として対応させる。

月刊 コンピューター 雑誌

1 2 3
4

monthly computer magazine

1 2 3
4

白黒 印刷 全面 広告

1 2 3
4

full page ad in black and white

2 3 1
4

cf. 国公立 研究 機関

1
2 3

national and public institute

1
2 3

日本語の複合名詞が英語では意識などされていて対応づけが難しい場合も、意味の対応を考えて可能なかぎり対応させる。

会議の参加者の人数はどれぐらいでしたか。

How many people attended the conference?

会議の開催日は10月27日です。

The conference will be held on October the 27th.

広告担当の方、お願いします。

Can I talk with the person in charge of the advertisement?

N.B. 複合名詞として単語レベルの深さ2の行に記述されたものは、データベースではランダムレベルに記述される。

2-1-2. 助動詞・補助動詞の対応

助動詞・補助動詞はひとまとまりでアスペクトやモダリティや発話意図を表すので、個々には英語との対応が非常にとりにくい語である。しかし、これについても、意味を考慮して極力対応させる。

れる・られる

会議は大阪で開催されます。
The conference will be held in Osaka.

私は自動車を盗まれました。
I had my car stolen.

全部は食べられない。
I cannot eat the whole thing.

せる・させる

子供たちにもっと勉強させる必要がある。
We need to make our children study harder.

自己紹介させてください。
Let me introduce myself.
Allow me introduce myself.
May I introduce myself?

う・よう

映画を見に行こう。
Let's go to the movies.

今年ももっと勉強しよう。
I will study harder this year.

彼女は逃げようとした。
She tried to run away.

らしい・べし

彼女は忙しいらしい。
It seems that she is busy.

あの男は中国人らしい。
That man looks like a Chinese.
That man appears to be a Chinese.

彼は結婚したらしい。
They say that he got married.

きみは彼に謝るべきだ。
You should apologize him.
You have to apologize him.

ぬない

誰が講演するのかまだわかりません。
I don't know which lecturers will be talking.

その計画はうまくいかないと思う。
I don't think the plan will work well.

食糧は何も残っていない。
There is no food left.

たいたがる

彼はあなたに会いたがっている。
He wants to see you.

会議についてお聞きしたいのですが。
I would like to ask you about the Conference.

詳しい資料をお送りしたいと思います。
I will send you the detailed information.

手配したいと思います。
I will certainly try to make the necessary arrangements.

N.B. 次の例は英語表現が変化しすぎなので対応させない。

関係資料を頂きたいと思っております。
I'll appreciate it if you'll let us have the materials.

だです

こちら通訳電話国際会議事務局です。
This is the Secretariat for the ICCS conference.

そうだ

それはおもしろそうだ。
That sounds interesting.

彼は会社をやめたそうだ。
I hear that he quit the company.

ようだ

彼は試験に落ちたようだ。
He seems to have failed the exam.

私たちは急行に間に合うように走った。
We ran in order to catch the express.

言われたように下さい。
Do as you are told.

できる

会議には参加できません。
I can not attend the conference.
I am not able to attend the conference.
It is not possible to attend the conference.

会議には参加できません。
I can't attend the conference.
It is impossible to attend the conference.
The conference isn't open to everyone.

ください

登録用紙をお送りください。
Please send me the registration form.
Could you send me the registration form?

致す

詳しい資料を早急にお送り致します。
I will immediately send detailed documents.
I am able to send you detailed documents immediately.

頂く

資料を送って頂けないでしょうか。
Can you send me such information?
Will you please send me such information?

アブストラクトをタイプして頂きます。
We would like you to type out the abstract.

ネームカードを胸につけていただきます。
Everyone will be required to put a name tag.

N.B. 次のように日本語と英語の意味が大きく変化していて対応させるのが困難な場合は対応させない。

事前に情報を知っておいて頂きたい。
It is important to let you know details in advance.

2-1-3. 「願う」の対応

「～お願いします」で終わる文では主動詞が省略されており、英語ではしばしばその主動詞に対応する特定の動詞が用いられている。このような場合、復元された部分も含めて対応させる。

登録の係の方(につなぐこと)をお願いします。
Please put me through to someone in charge of registration.

御名前、御住所(を教えること)をお願いします。
Will you give us your name and address.

2-2. 文節レベル

自立語と付属語からなる日本語の1文節に対応する英語の表現があるとき、深さ1の行に記述する。

こちら 北京から 電話して る んですが。
1 2 3
I'm calling from Beijing.
1 3 2

単語レベルで深さ2に記述するものが、文節レベルでは分割されてしまうことがある。そのような場合は、文節レベルでは対応させない。

会議への登録についてお知らせください。
Please tell me about the registration to the Conference.

会議への登録についてお知らせください。

Please tell me about the registration to the Conference.

×

2-2-1. 決定詞の対応

英語の決定詞は対応範囲に含める。

会社の特徴等を含む内容を記した物を

1 2 3 4

a document describing the features of these companies

4 3 2 1

会場のそばにホテルありますか。

1 2 3

Are there any hotels near the conference hall?

3 2 1

N.B. 決定詞には次のようなものがある。(研究社・新英語学辞典)

a[an], the, some, any, no, every, either, neither,
this [these], that [those], which (ever), what (ever),
my, your, his, her, its, etc.

ただし、決定詞でも、日本語でその意味が明示されている場合は、そちらに対応させる。

そちらの御住所をお聞かせください。

1 2 3

Would you please give me your address.

3 1 2

2-2-2. 述語文節の対応

述語文節は助動詞・補助動詞も含めて、全体の意味を表す英語表現を対応させる。

この論文集はいつ頂けるんでしょうか。

When I wonder I can get this abstract?

特別の資格があるんでしょうか。

Are you going to require any qualifications?

そちら登録用紙をお持ちでしょうか？

Do you have a registration form?

会議の参加料は割引されますでしょうか？

Does this entitle me to a discount rate for the conference?

2-2-3. 省略語句の対応

日本語で省略され(主語、目的語、修飾語、etc.)、英語で明示された部分の対応は、明示された部分が係る述語、もしくは明示された部分が修飾する語句とともに対応させる。

当日行ってみなければわからないんですね。

1 2

I can't know anything about this until I physically go there

2 1
on the conference day.

登録用紙をお送りすればよろしいのですね。

I only have to send the registration form to your office?

アメリカですと1週間で着くと思います。

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

2-2-4. 文末表現の対応

モダリティや発話意図を表す文末表現を、最長一致で深さ2の行にまとめて対応させる。

会議に参加したいのですが。

I would like to attend the conference.

料金は払い戻しして戴けるのでしょうか。

Would you please pay back the fee to me?

ファクシミリをお使いいただいで結構です。

You could also fax it to us.

登録用紙をお送りしなければなりませんか

Do I have to send the registration form?

2-3. 格レベル

格構造の対応を調べるのが、このレベルの目的である。すなわち、日本語を格解析し、命題および命題内部の格関係の対応をとる。

2-3-1. 日本語の格解析

日本語の格解析はMCWに従う。すなわち、MCWによる「係り受けファイル」から、01、03のコードを与えられており、かつ述語となる語が、助動詞・補助動詞以外のもの(「だ・です」は例外とする)を抽出し、命題の範囲と述語・格要素を確定する。

命題の範囲は、01のコードを与えられたものについては、述語のHEADに係る最初のHEADを含む名詞句の初めの語から、当該の述語のHEADまでとする。ただし、サ変名詞が述語になっている場合は補助動詞の「する」まで、形容名詞が述語になっている場合は助動詞の「だ、です」までを命題範囲に含める。副詞等も、述語のHEADに対する係り受け関係が記述されていれば、命題範囲に含める。

会議のことを新聞で読みました。
< >
来年会議を開催されるそうですが。
< >
なかなか出来上がるのが遅くなっております。
< >

「思う」「聞く」など、内容格をとる述語をもつ命題は、述語に係る係り受けの全体を命題範囲に含める。(MCW 05による)

先生のおっしゃる方でしたら、どなたもきっと素晴らしい方だと
<
思います。
>

03のコードを与えられたものは、述語のHEADに後ろから係るHEADまでを命題範囲に含める。

父がくれた本
< >

述語は、MCWが指定した1語に限定される。ただし、「サ変名詞+する」「形容名詞+だ[です]」は、「する」「だ[です]」も述語に含める。

会議のことを新聞で読みました。

< _____ >

来年 開催 される会議。

< _____ >

MCWが、省略されていると思われる述語を復元している場合でも、日英対応調査では復元は行わない。

格助詞、係助詞、および格助詞相当語句にがぎって、格要素の中に含める。

2-3-2. 英語との対応

MCWによって01、03のコードを与えられたものについて、日本語の命題に対応する英語の表現を確定し、さらに、格要素のそれぞれに対応する英語の表現を抽出する。

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 >

会議に参加したいのですが。

<1 >

I would like to attend the conference.

<1 >

N.B. "I" を命題に含めたのは、"attend" の意味上の主語だからである。

父がくれた本

<1 2 >

the book that my father gave me

<2 1 >

2-3-3. 補文について

日本語の補文に"*"を付加し、対応する英語の命題が補文であれば、同じく"*"を付加する。ただし、日本語と英語の補文をそれぞれ以下のものとする。

日本語:連体修飾構造となって後ろの名詞または準体助詞の「の」にかかる文。

a: 日本語を話す人

<* >
詳しい日にち
<* >

b: 8日に終る予定です

<* >

N.B. (a)は関係節で、修飾された名詞が補文の格要素となっている。そこで、この場合は被修飾名詞も命題の中に含める。(b)は同格節で、修飾された名詞は補文。

補文標識の「と、って」「の」「こと」で受ける文。

通訳電話国際会議があると聞いた
<* >

スピーチをするのは構いません
<* >

会議に参加することが出来ます
<* >

受身、可能、自発の「れる、られる」、可能の「できる」で受ける文、および、不可能の「かねる」で受ける文。

開会式は開かれるんですか
<* >

登録料は引き下げられません
<* >

プロシーディングスは入手できますか
<* >

登録料は払い戻ししかねます
<* >

使役の「せる、させる」で受ける文。

登録申請書を送らせていただく
<* >

希望の「たい、たがる、ほしい」で受ける文。

車で行きたいんですが
<* >

登録手続きについて説明してほしい
<* >

受給表現の「いただく、いただける、もらう、」で受ける文。

キャッシュでお支払いいただいで
<* >

登録用紙を送ってもらえますか
<* >

難易を表す「やすい、にくい、難しい」で受ける文。

支払いがしやすい
<* >

それについてはお答えしにくい
<* >

N.B. 命題と、以上のような補文を受ける語との間に他の語がある場合も、その命題を補文とする。

料金を払わなければいけないわけ
<* >

英語: 補文標識“that”によって導入される文

I think that the Gion is more expensive.
<* >

I guess I have the reply form with me.
<* >

補文標識“who, what, when, where, which, how, whether,if”によって導入される文

We haven't decided who are going to be there
<* >

Do you know if I can stay there?
<* >

補文標識“poss-ing”によって導入される文

I prefer attending the session.
< * >

補文標識“for-to”によって導入される文

I ask you to participate in this conference.
< * >
Is it possible for me to obtain the summary?
< * >

2-4. ランダムレベル

日英対応入力検索システムでは、ランダムレベルに何も対応させないが、データベースでは単語レベルの深さ2の行、文節レベルの深さ2の行で記述されたものが記述される。

2-5. 文レベル

日本語のテキストと英語のテキストには、それぞれ文IDが与えられている。この文を単位にして、ひとまとまりの意味を表す部分に対応させる。日本語と英語が1対1、1対多、多対1に対応してよい。

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
6127ということですね。

タイプアップしたものは来週末までちょっとお送りできないんですけれ
2
ども、手書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。

Let me make sure I have those number correctly.
1
063726164 <zero six three seven two six one six four> or 6127 <six one
two seven>.

I like I say, I'm afraid I won't be able to send a typed copy till the
2
end of next week.

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

2-6. 発話レベル

日本語のテキストと英語のテキストには、それぞれ1人の発話者の1回の発話に対して発話IDが与えられている。発話レベルでは、日本語の1人の発話者の発話に対して英語を対応させる。

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
6127ということですね。

タイプアップしたものは来週末までちょっとお送りできないんですけれ
ども、手書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。

Let me make sure I have those number correctly.

1

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

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

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

1人の発話者が発話しているときに他の発話者があいづちを打っているときは、あいづちと、あいづちでさえぎられたそれぞれの発話に発話IDが与えられている。この場合、あいづちをギャップとして、1人の発話者の発話の全体に英語を対応させる。

事務局:会議は8月の5日から8日までで、

1

質問者:はい。

事務局:1987年ですけれども、5日は土曜日にあたります。

It will take place from August 5th through 8th.

1

It begins on a Saturday 1987.

3. 属性について

3-1. 意味属性

単語レベルにおいて対応関係のある単語について、プログレッシブ和英中辞典(第1版)を参照し、日本語の見出し語、英語の訳語が共に記載されているものに対して、属性“T”を与える。

国際コンピューター会議の 記事

T

an article about the International Conference on Computer Science

T

きじ【記事】 an article; an news story [item]

日本語の見出し語はあるが、それに対応する英語の訳語が記載されていないものに対して、属性“C”を与える。

会議場付近のホテルの 案内

C

some informations about the hotel accommodations

C

あんない【案内】 1〔人を導いていくこと〕 guidance

プログレッシブに日本語の見出し語そのものが記載されていないものに対して、属性“N”を与える。

まず、コールフォーペーパーズをお送りします。

N

We will send you our call-for-paper as soon as possible.

N

固有名詞については、日本語・英語ともに固有名詞のものには属性T、どちらかが固有名詞でないものには属性Cを与える。

通訳電話国際会議

T

International Conference on Interpreting Telephony

T

通訳電話国際会議事務局

C

office

C

ただし、以下のものについては意味属性を付与しない。

日本語・英語どちらかが代名詞・指示詞であるもの。

日本語・英語どちらかが数詞であるもの。

日本語で「願う」を省略範囲を含めて対応付けているもの。

日本語が助詞・助動詞・補助動詞・接続詞、英語が前置詞・助動詞・接続詞であるもの。

3-2. モダリティ属性

文節レベルの深さ2の行に対応づけたモダリティ表現に対して、属性“M”を与える。

会議に参加したいんですけれども。

M

I'd like to attend the conference.

M

3-2. 複合名詞属性

単語レベルの深さ2の行に対応づけた複合名詞に対して、属性“A”を与える。

登録申し込み用紙

A

registration application form

A

(謝辞)

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