

TR-I-0043

言語データベース作成のための
日英対訳対応付け

Corresponding Japanese and English Text
for Linguistic Database

篠崎 直子 小倉 健太郎 森元 暉
Naoko SHINOZAKI Kentaro OGURA Tsuyoshi MORIMOTO

1988.9

概要

現在ATR自動翻訳電話研究所では、各種言語現象の分析を可能とする言語データベースの作成を進めているが翻訳という処理を考えた場合、付加情報として「日英対訳対応情報」を用意することは言語間の変換処理用の基礎データとして重要である。

そこで、まず「日英対訳対応付け」を行う際の問題点を明確にするため、テストデータを用いて予備調査を行ったが、この調査によりデータベース化の見通しを得ることができた。

本稿では、この調査の結果について報告するとともに、実際の言語データベース作成における「日英対訳対応付け」に調査結果をどのように反映したかについて述べる。

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

- 目次 -

1.はじめに	1
2.調査の前提	2
2.1. テストデータ	2
2.2. 対応付けの単位について	3
①単語レベル	4
②文節レベル	5
③格レベル	6
3. 調査結果	9
3.1. 対応がとれたもの	9
3.2. 対応がとれなかったもの	12
3.2.1 レベル別問題点	12
①単語レベル	12
②文節レベル	14
③格レベル	15
3.2.2 問題点の分析	17
①文法にかかわるもの	18
②構文にかかわるもの	20
3.2.3 ランダムレベル	22
4. おわりに	25

1.はじめに

現在我々は、各種言語現象の分析を可能とする大規模言語データベースの構築を進めている。^{[1][2][3]}

この言語データベースには付加情報として「形態素情報(単語情報)」「格、係り受け情報」を与えることとしているが、翻訳という処理を考えた場合、更に「日英対訳対応情報」を用意することは言語間の変換処理用の基礎データとして重要である。

そこで、まず「日英対訳対応付け」を行う際の問題点を明確にするため、テストデータを用いて予備調査を行ったが、この調査によりデータベース化の見通しを得ることができた。

本稿では、まず調査の前提を述べ、次にこの調査の結果について例を挙げて説明する。最後に実際の言語データベース作成における「日英対訳対応付け」をどのような形で行うこととしたかについて述べる。

2.調査の前提

2.1 テストデータ

この調査ではデータの極端な偏りをなくすため、テストデータとして2種類の異なったテキストを用いた。

具体的には以下の通り。

1. 「ニュース報道を目的とした雑誌文」

- a)テキスト種別： 書き言葉
- b)データ量： 1記事,3800語
- c)翻訳方向： 英語 ← 原文
日本語 ← 翻訳文

2. 「国際会議の問い合わせをテーマとしたキーボード会話」

- a)テキスト種別： 通信文
- b)データ量： 9会話,3000語
- c)翻訳方向： 質問者の日本語 ← 原文
質問者の英語 ← 翻訳文
事務局の英語 ← 原文
事務局の日本語 ← 翻訳文

なお語数は日本語の単語数(形態素数)で示している。

2.2 対応付けの単位について

日英の対訳対応付けを行う際に先ず問題となるのは何をもってその単位とするか、及びそれを相手言語のどの単位と対応させるか、である。

単位として考えられるものには
 会話/文章、発話/パラグラフ、文、句、命題、文節、単語
などがある。

このうち会話単位、発話単位は話し言葉の単位であり書き言葉にはない。逆に文章単位、パラグラフ単位は書き言葉の単位であり話し言葉にはない。(通信文は書き言葉と話し言葉の中間に位置付けられる。)また、文節単位は日本語にのみ存在する単位であり、これに当たる英語の単位はない。

これらの単位の利用を考えると大略以下のようなになる。

☆ 会話単位/文章単位の対応

→ 会話や文章の大きな流れの対応を分析するためのデータとなる。

☆ 発話単位/パラグラフ単位の対応

→ 局所的な流れや意図の対応を分析するためのデータとなる。

☆ 文単位、句単位の対応

→ 表現方法の違いや、構文の対応パターンを分析するデータとなる。

☆ 命題単位の対応

→ 命題の対応を分析すると共に、格要素の変化についても分析するデータとなる。

☆ 文節単位の対応

→ 文末などのひとまとまりの表現を扱うデータとすることができる。
また、単語単位で扱えない複合語相当表現もこの単位で吸収することができる。

☆ 単語の対応

→ 翻訳用変換辞書作成に利用できる。
また、翻訳技術の研究データともなる。

もちろん利用目的はこれらに限定されることはなく、また明確に境界線を引ける類のものでもない。ここにあげたのはその一例である。

今回の調査では、会話単位/文章単位および発話単位/パラグラフ単位の対応は既にとれていることを前提として、まず、文単位の対応を考え、テキストを意味のあるまとまりに分割した。

第 一 回 通訳電話国際会議事務局 だ っ け ？
1 2 3 4 5 6 7

Is it the Secretariat of the First International
5 42.....
of Interpreting Telephony ?

日本語の単語分割は [4] に従って行った。

ただし、「形容名詞+だ、です」と「サ変名詞+する」は1つの単位とみなして対応付けを行った。

先 ず 登 録 が 必 要 だ っ け ？
1 2 3 4 5 6 7
First you need to register .
1 5 3

会 議 に 参 加 し たい の だ っ け ？
1 2 3 4 5 6 7 8 9
I would like to attend the conference.
5 3 1

英語の単語分割は空白によることとした。

「英 → 日」の対応付けは「日 → 英」と逆の考え方で行った。

②文節レベル (日 → 英)

このレベルでは対応付けは
日本語の一文節に対応する英語表現があれば対応させる
という方針で行った。

会 議 に / 参 加 し たい の だ っ け ？
1 2
I would like to attend the conference.
2 1

対応する英語表現が、別の語句をはさんで分離するときもある。

参 加 料 は / 銀 行 振 込 み だ っ け ？
1 2
Do I pay my registration fee to your bank account ?
1 2

文節の構成は基本的には「接頭語*+自立語+接尾語*+付属語*」とする。

(注)ここで(*)は任意個の繰り返しを表わす。

ただし次の2つは例外とする。

1. 複合名詞相当の名詞連続は全体で1つの名詞扱いとする。

/登録 料/
/口座 番号/

2. 補助動詞は付属語扱いとする。

/失礼 致し ます/
/振り込ん で 下さい/

なお、このレベルで「英 → 日」の対応付けがないのは英語には文節に当たる単位がないためである。

③ 格レベル (日 → 英, 英 → 日)

このレベルでは「日 → 英」の対応付けは

まず[5]に従って "述語を中心とした" 日本語の格解析を行って

1. 述語 → Pとアンダーラインで示す。
2. 格要素 → 番号とアンダーラインで示す。
3. 命題全体 → 番号と()で示す。

の3つを単位要素とした上で、これらの単位要素に対してそれぞれ対応する英語表現があれば対応させる。

という方針で行った。

観光 料金は 会議 の 登録 料 に 含ま れ て おり
(1 1 2 P)
ます 。

The tour fee is included with the registration fee.
(1 1 P 2)

私 は 英語 が 全然 分ら ない の です が 、
(1 1 2 P)

同時通訳 の サービス は 有る の でしょう か 。
(2 1 P)

I'm not good at English .
(1 1 P 2)

Do you have simultaneous interpretation ?
(2 P 1)

対応する英語表現が、別の語句をはさんで分離するときもある。

聞いて居るだけでいいのですが、
(1 1 P)
I would just like to attend the conference .
(1 1 P)

また、埋め込み構造は命題番号の後ろに*マークを付して区別した。ただし、次のものを埋め込み構造とした。

1. 連体修飾構造となって後へかかる命題

- a. 被修飾体言が埋め込み構造の格要素となっているときは埋め込み構造の範囲にその体言も含める。

送っていただく宛先は、
(1* P 1)

- b. 被修飾体言が埋め込み構造の格要素となっていないときは埋め込み構造の範囲にその体言は含めない。

もし参加できなくなった場合
(1* P)

2. 格要素の内容自身、別の命題となっているもの

スピーカーも...参加すると思います。
(1* 1 P)
(2 1 P)

なお、このレベルで「英→日」の対応付けがないのは英語には文節に当たる単位がないためである。

命題の内部構造が日本語と英語で違っていて、述語、格要素が対応しないことがある。その場合には対応付けできない。

他に何か御座いますか？
(1 1 2 P)
Will that be all ?
(1)

「英→日」の対応付けは「日→英」と逆の考え方で行った。

ちなみに格解析の際「述語」としてとったものは次のとおり。

1. 日本語

動詞、形容詞、「形容名詞+『だ・です』」、
「サ変名詞+『する』」、「助動詞の『たい』」、
「コブラ文の『だ・です』」

2. 英語

動詞、形容詞、「コブラ文の『be動詞』」

3. 調査結果

調査結果を簡単に述べると、当然の事ながら、対応がとれたものととれなかったものがあり、対応のとれたものもそのとれ方は様々であった。

この調査の過程で日英対訳対応情報をデータベース化する際のポイントが明らかとなった。それは次の2点である。

1. 対応がとれたものについて

対応のとれ方にどのようなものがあるかを整理した上で、その中で区別すべきものはどれか、どの様に区別するか、を検討すること。

2. 対応がとれなかったものについて

対応がとれなかったものにはどのようなものがあるかを整理した上で、何らかの方法で対応付けるべきものはどれか、どの様に対応付けるか、を検討すること。

以下、順に述べる。

3.1 対応がとれたもの

対応がとれたものの「とれ方」を整理すると次のようになる。

(a) 一般の辞書に載っている語の意味として対応をしておき、文脈を離れても対応しているとみなせるもの。

経済 問題
economic problem

登録 用紙
registration form

gray Mao suit
灰色の人民服

attend the conference
会議に参加する

もちろん、これにも程度の違いはあって、翻訳者が多少手を加えているものもある。

severely
厳しい態度で

taking the subway
地下鉄を利用する

(b) 概念の一部を保っているが、文脈を考慮しなければ対応していることがわからないもの。

外国との交流計画
exchange program

総書記に専念する
keep the new post

詳しいことを知らせる
send the information

道路の込み具合による
It depends on the traffic

(c) 同一指示対象をもつために、文脈上でのみ対応の認められるもの。

they unanimously
拡大会議は一致して

用紙を送りましょうか
Shall I send you one?

(d) 品詞が変わるもの

まったく予想外だった
come as a complete surprise

トウ小平の引退
once he retires

締め切らせていただきます
The deadline is ~

(e) パラフレーズされているもの。

自己批判し
delivered a self-criticism

said vaguely that
漠然とした説明が出されていた

observe the conference
聴衆として参加する

これら(a)~(e)のうち、(a)(b)(c)は構造的に対応しているので対応付けは比較的容易である。特に(a)は意味的にも対応しているので対応付けの際にも迷わずにすむ。

しかし、それに比べて(c)は文脈を離れると対応がとれなくなり、厳密には意味の対応がない。そこで(c)は言語データベース化する際にはCマークを付与することとする。(C stands for Context)

また(d)は構造的にずれのある例だが、概念を共有している(意味が対応している)限り、対応付けはそれほど困難ではない。

(e)は構造的に対応していないだけでなく、意味的にも若干のずれがあるので、厳密な対応をとることはできないが、言語分析用資料とするために緩やかな基準で対応付けを行い、言語データベース化する際にはUマークを付与することとする。(U stands for Uncomfortable)

3.2 対応のとれなかったもの

この章では先ず「単語レベル」「文節レベル」「格レベル」とレベル別に対応がとれなかったものにどのような例があるか具体的に見ていく(3.2.1 レベル別問題点)。

次に、対応がとれなかった原因について考察し(3.2.2 問題点の分析)、言語データベース化の際に導入することとした新たなレベルについて述べる(3.2.3 ランダムレベル)。

3.2.1 レベル別問題点

ここでは「単語レベル」「文節レベル」「格レベル」の順にレベル別問題点を列挙する。類似した問題が続くので、少々くどい感じもするが、表層的な問題点の整理の為、敢えて繰り返した。

①単語レベル

(a) 2語以上でひとまとまりとなり、全体として相手言語と対応するものは、それぞれの語自体の対応がとれなくなる。

厳肅な 同志的 批判
serious and comradely criticism

特別 臨時 ニュース で
in a special bulletin

誤りを 自己 批判 し
delivered a self-criticism

拡大 会議 は 一致 して
they unanimously

会議 の 後 の 御昼 ご飯 の とき に
at lunch after the morning session

スピーカー の 方々 と 意見 交換 出来る 機会
any chance to discuss with the speakers

the People's Republic of China
中華人民共和國

By the end of the week
週末 を 迎える 頃 には

ここまでは相手言語の一語と対応しているので、逆方向の対応はとれる例である。しかし、相手言語の二語以上に対応しているときはそれができないし、極端な例として、翻訳に際して構文が全面的に組み替えられ、単語単位の対応を考えることがほとんどできない場合もある。

彼の書く暴露記事には熱心な読者が多く
His exposes are avidly read ~

無難な道を選ぶ工場長が増えそうだ
many managers play it safe

今回の運動に対する胡耀邦のやり方は行きすぎだ
Hu had taken the campaign too far

胡耀邦がそろそろ危ない
Hu's days had to be numbered

(b) 助詞、助動詞、補助動詞を含む表現は、単語単位では対応が対応がとりにくい。特に、アスペクトや時制、態、法などを表す表現はほとんど対応がとれない。

灰色の人民服を着ていた
Sporting a gray Mao suit

人民服を脱ぎ捨てて背広を着た
he shed his Mao suits and wore Western clothes

厳しい態度で臨まなければならない
we must deal severely with ~

詳しいことをお知らせ致しましょうか
Shall I send you the information?

登録料を払って戴かなくてはなりません
And you will need to pay the registration fee

また、ある種の否定表現は日本語と英語で否定の位置や表現がずれることがある。

そう思わない連中は夢を見ているのだ
anyone who thought otherwise was dreaming

こうした筋書きが実現しそうには見えない
that scenario seemed unlikely

党のスポークスマンが～肯定も否定もできないと答えた
a party spokesman said he could neither confirm nor deny ~

登録料はドルでなければ御受取りできません
we only accept the fee in US dollars

ホテル代と食費は含まれておりません
The hotel and food expenses would be extra.

(c) 文脈に依存してはじめて対応する表現がある。

いずれは首相を辞任して総書記に専念する
to relinquish his old job and keep the new post

支払いの締切は6月30日です。
The deadline for the registration fee is June 30

特に、代名詞や関係代名詞は文脈によってさまざまに訳し分けられる。

After offering Hu ~, the party leaders accepted his resignation
党の指導部は胡に対して～その辞任を認めた

they unanimously elected Prime Minister Zhao Siyang,
拡大会議は一致して趙紫陽を総書記代行に任命した。

Do you have one?
カードは御持ちですか

You also need to pay the registration fee, which is 100 dollars (US)
登録料は100(アメリカ)ドルです

②文節レベル

(a) 2文節以上でひとまとまりの意味を持ち、全体として相手言語と対応するものは、それぞれの文節自体の対応がとれなくなる。

上等な/仕立ての/背広
well-tailored suits

声高な/主張を/掲げる/強硬派
assertive hard-line faction

ただ/死者を/出さぬ/よう/最善を/尽くして/もらいたい
Just try as much as possible not to kill anyone

こうした/説明は/~/ますます/信憑性が/薄く/なっていた
those excuses were looking thinner and thinner

申し訳/ありませんが、/ドルで/お願いします
I'm sorry but we would prefer for you to pay in \$ (US) only.

話は/変わりますが、/観光ツアーも/行われる/ようですが
By the way, I've heard there is a sightseeing tour.

(b) 「まるで～ない」のような一種の係り受け表現の対応がとれない。

胡の/リベラルな/考え方は/軍首脳と/まるで/そりが/合わない/もの
だった
Hu's relatively liberal views never sat well with the Chinese military establishment

どんなに/高い/地位のもの/だろうと、/いかに/高い/名声の/持ち主で/
あろうと
no matter how high his position and no matter how big his reputation

(c) 構文が全面的に組み替えられているとき、文節単位ではほとんど対応づけることができない。

わたしが/困った/立場に/なるかもしれない
It may get back to me

国内の/知識人に/対する/締め付けが/今まで/以上に/強化される
the country's intellectuals would be kept on a tighter rein from now on

今後の/中国情勢を/見る/上で/興味深い/ところだ
offers some intriguing possibilities about the direction in which China might be heading

趙紫陽を/起用すれば/政治的な/危険が/小さく/なると/トウ小平は/
考えたのかもしれない
for Deng, that probably translated into fewer political risks

もし/、二日目から/参加できなくなった/場合/
If I can come only for the first day of the conference,

地下鉄と/バスで/行くよりも/、タクシーの/方が/速い/ですか？
Will it take longer if I ride the subway and a bus instead of a taxi?

③格レベル

(a) 日本語では「名詞+述語」で表されるものが、英語ではただ一つの述語で表されることがある。

そう思わない連中は夢を見ているのだ
anyone who thought otherwise was dreaming

今日でも彼は、上等な仕立ての背広に～を身につけている
Even today he sports well-tailored suits, ~

指導体制をめぐる新たな抗争の引金となる可能性もある
it could also trigger a new round of factional fighting within the leadership

その影響には深甚なものがあるだろう
the repercussions could be far-reaching

こうした説明はますます信憑性が薄くなっていた
those excuses were looking thinner and thinner

また、日本語では名詞句で表現されているものが、英語では文単位に展開されることがある。

申し込みの方法を
How do I register?

会議参加の方法を教えてください
What do I have to do to attend the conference?

- (b) 日本語のある種の文末表現は述語を含んでいるが、それを格構造に分割していくといたずらに繁雑になり、しかも英語との対応がとれない。

先週の発表はまったく予想外だったわけではない
Last week's announcement did not come as a complete surprise

このような事態はトウ小平としても面白かるうはずがない
That can scarcely give Deng any pleasure

趙紫陽の後継首相に誰がなるかは今後の中国情勢を見る上で興味深いところだ
the question of Zhao's successor offers some intriguing possibilities about the direction in which China might be heading

それはトウ小平の引退に伴う政権引き継ぎ構想をさらに混乱させかねないものだ
that could further disrupt Deng's transition plans for China once he retires

- (c) 「～によって、～による」のような複合助詞相当語句は動詞を含んでいるが、英語では前置詞に対応することがある。

趙紫陽は75年に、文化大革命による混乱收拾のため四川省に派遣されている
In 1975 Zhao was sent to Sichuan Province to clean up the damage left by the Cultural Revolution.

命令を無視する者に対しては厳しい態度で臨まなければならない
We must deal severely with those who defy orders

同市の建設プロジェクトや住宅計画、新しい道路システムに関する李市長の功績
his work on construction projects housing and a new road system for the city

胡耀邦を辞任させるという決定はトウ小平にとって容易ならざるものだったはずだ
the decision to fire Hu could not had been easy for Deng

トウ小平の近代化路線全体がより保守的な線に沿って手直しされる可能性もある
Deng's entire modernization campaign might be recast along more orthodox lines

国営北京中央テレビネットワークを通じて、このアナウンサーは特別臨時ニュースで次のように伝えた
he said in a special bulletin on the Chinese state network

- (d) 英語の副詞や形容詞に対応する動詞がある。

拡大会議は一致して趙紫陽首相を総書記代行に選出した
they unanimously elected Prime Minister Zhao Ziyang

中国は依然として多くの経済問題に悩まされている
The country is still plagued with a host of economic problems

トウ小平 自身 最近 三度にわたり、面と向かって 胡耀邦 の 行きすぎた
自由化政策を非難した

Deng had berated Hu three times in recent face to face meeting for his overly liberal
policies

中国 首脳 が これほど 思い切った 行動に出たのは初めてだった
the first time anyone so high-ranking had been so bold

胡耀邦 ほど 陽気な人間ではないが、とにかく 仕事のできる
男だ

He is not as good-humored as Hu, but he is competent

- (e) 命題の内部構造が全面的に組みかえられている場合、格関係の対応はとれない。

トウ小平 が 最高 実力者 の 地位 に とどまる のも そう 長く は ない
だろう

Deng probably does not have much time left at the center of power.

趙 と 西欧 的 な 政治 思想 と の つながり は それ ほど 密接 で は ない
Zhao has not been all that closely linked with Western political ideas:

それらの解決には、中央統制の緩和や資本主義国に
対する門戸解放政策の継続が必要

~that will be hard to solve without some easing of central control and
maintaining an Open Door to capitalist countries

討論に参加するのではなく
But I don't read paper

登録料を払って戴ければ誰でも参加できます。

If you want to attend, you have to pay the registration fee just like everybody else.

3.2.2 問題点の分析

「3.2.1 レベル別問題点」では、いわば表層的な問題点整理を行ったのであるが、この節ではそれを内容別に整理する。それには大きく分けて2つのタイプがある。

ひとつは主として文法に関わるもので、名詞や動詞などの実質語の周辺にあってそれらの関係を支える語——具体的にいうと、助詞、助動詞、前置詞、接続詞、ある種の形式的な語句など——に関するものである。

もうひとつは構文構造に関わるもので、日本語と英語の発想の違いによって構文が大きく組み替えられたり著しく意識されたりするタイプのものである。

この2つについて順に述べる。

①文法にかかわるもの

以下のような表現は、日本語と英語の文法形式の違いによるものなので、個々の語に着いて対応をとろうとするのが無理なのかもしれない。

(a) 日本語の助詞は、英語では構文的な位置関係で表されることがある。

こうした筋書きが実現しそうにはみえない。
that scenario seemed unlikely

詳しいことをお知らせ致しませうか。
Shall I send you the information?

(b) 時制、アスペクト、態などを表す助動詞、補助動詞は、英語では動詞の変化で表されることが多い。しかも、そのあいだに、規則的な対応を見つけることが難しい。

党の指導部はその辞任を認めた。
the party leaders accepted his resignation.

警察当局は学生に圧力を加え続けている。
police were putting the heat on the campuses.

市内観光が予定されていますか。
Are sightseeing tours planned ~ ?

(c) モダリティや待遇を表す語句は規則的な対応を見つけることが難しい。

円で御支払いしてもよろしいでしょうか。
Can I pay by YEN?

まだ研究発表を募集して居られるのですか。
Do you still accept submission of papers?

(d) 日本語の文末を構成する形式的な語句には対応する英語がない。

彼らはそう言いたかったわけだ。
They wanted to show that ~

北京駐在の外交官の注目するところとなっていた
Diplomats in Peking noted with interest that ~

会議に参加したいのですが
I would like to attend the conference.

(e) 英語の冠詞や限定詞は、日本語では訳されないことが多い。

a grey Mao suit
灰色の人民服

I would like to attend the conference.
会議に出席したいのですが。

I don't have any credit card.

クレジットカードを持っていないのですが。

- (f) 前置詞は日本語に訳出されることもあれば、されないこともある。

at the conference

会議場で

experts on China

中国専門家

- (g) 文脈から明らかな英語の代名詞や主題語句は、日本語には訳出されないことが多い。

I can't speak to you.

あなたと話すことはできない。

We can afford to shed some blood.

多少の血は流れてもかまわない。

The Registration fee is \$100 (US).

アメリカドルで100ドルです。

- (h) 形式的な "it" や "there" は訳出されない。

it was plain that ~

~ことは明らかである。

There was certainly plenty of evidence

証拠はいくらもあった。

It would be more convenient to pay by the credit card.

クレジットカードの方がいいのですが。

- (i) 制限用法の関係代名詞は訳出されない。

those who defy orders

命令を無視する者

anyone who thought otherwise

そう思わない連中

- (j) 並列の接続詞や補文標識の "that" は訳出されないことがある。

serious and comradely criticism

厳粛な同志的批判

the possibility that Deng's entire modernization ~

トウ小平の近代化路線 ~ 手直しされる可能性

②構文にかかわるもの

構文が組み替えられたり意識されたりすると、単語のレベルから格のレベルまで、さまざまなずれが生じて対応がとれなくなるが、その種の組み替えには、次のようなタイプがある。

- (a) 日本語では動詞構文であったものが、英語では名詞構文になるもの、あるいはその逆。これにともなって、主語や目的語が名詞にかかる所有格になったり、副詞が形容詞になったりといった、規則的な変換が生じる。

声高な主張が一気に噴き出した
The bursts of hard-line rhetoric

彼は経済改革と、西側への接近を熱心に提唱してきた。
He has been an ardent advocate of economic reforms and the overtures to the West.

趙と西欧的な政治思想とのつながりは、それほど密接ではない。
Zhao has not been all that closely linked with Western political idea

- (b) 視点をどこに置くかによって、主語に選ばれる語が変わり、それによって、構文が組み替えられることがある。

国内の知識人に対する締めつけが強化されるだろう。
the country's intellectuals would be kept on a tighter rein

それだけで彼が政治的苦境に陥ったわけではない。
Hu's political problem did not stem solely from ~

趙のテクノクラートとしての手腕が、トウ小平の気に入っていた。
he was also impressed with Zhao's technocratic skills.

視点の違いは、次の「(c) 能動・受動の違い」にも影響を与える。

- (c) 日本語では能動態で表現されたものが、英語では受動態で表現されることがある。また、その逆のケースもある。

作家の王若望を公に非難、党から除名した。
~, writer Wang Ruowang, who was expelled from the Communist party.

攻撃の調子には、文化大革命当時のレトリックが反映されていた。
The tone of the attacks echoed some rhetoric of the Cultural Revolution.

文法的な能動態・受動態ではないが、意味的にはそれに似た表現もある。

党は事態を完全に掌握している。
the party was in control.

(d) 英語の分詞構文や不定詞を伴う構文は、日本語ではさまざまに言い替えられる。

テレビのアナウンサーは、灰色の人民服を着ていた。
Sporting a gray Mao suit, the television newscaster ~

国務院も、先週この提案を承認した。
~a move endorsed just last week by the State Council

趙紫陽に代わる首相の後任がいつ決まるか
when a successor would be chosen to replace to Zhao as prime minister

(e) 数量的な表現を伴うとき、日本語と英語で構文が変わることがある。

無難な道を選ぶ工場長が増えそうだ
many managers would play it safe

ニュースは、王には異端の言説が多かったと非難した。
Newscasts accused Wang of numerous heresies

トウ小平が最高実力者の地位にとどまるのも、そう長くはないだろう。
Deng probably does not have much time left at the center of power

これは、より一般的には、簡潔に表現されたものが翻訳に際してパラフレーズされることがあるということである。

国営新華社通信も、名前を特定せずに、一部の党員が～と非難した。
The saterun Xinhua News Agency charged that some unnamed party members had shown ~

追放の対象となった知識人の中で、最も重要と思われる人物
Potentially the most significant target of the purge of intellectuals

(f) 日本語と英語で肯定と否定が逆転することがある。

更迭劇はまだ続く可能性がある。
the purge may not be over yet.

それらの解決には、中央統制の緩和や～が必要と考えられる
that will be hard to solve without some easing of central control and ~

(g) 慣用表現によるもの

わたしが困った立場になるかもしれない
It may get back to me.

胡耀邦がそろそろ危ない
Hu's days had to be numberd.

趙紫陽は従来の伝統的な中国指導者とは似ても似つかない。
Zhao hardly cuts the figure of a traditional Chinese leader from the past.

実際には、以上のタイプのいくつかが同時に働いていることが多い。また、こうした構文変換は規則というよりも、翻訳の技術とみなした方がよいので、必ず適用されるというものではない。文脈により、視点の置き方により、また翻訳者により、さまざまに変化すると思われる。

3.2.3 ランダムレベル

「3.2.1 レベル別問題点」「3.2.2 問題点の分析」によって、単語レベル、文節レベル、格レベルの3つのレベルで対応付けられないものを洗い出すことができた。

それらのうち、対応が実際に存在しないもの、例えば翻訳時に省略してしまっただけなどは対応のつけようがないが、レベルを固定さえしなければ何らかの形で対応をとれるものは言語データベース化にあたっては「ランダムレベル」として対応付けることとした。

「ランダムレベル」として対応付けられるものは以下の通りである。

(a) 複合語や固有名詞のように、概念としては1語のものが2語以上に分割されているもの。

Mao suit
人民服

food expenses
食費

登録者
registered conference goers

これには、構成要素の全部または一部が単語レベルで対応しているが全体でひとまとまりの表現とみなした方がよいものも含まれている。

コールフォーペーパー
call for paper

また、接頭辞/接尾辞を伴うものも多い。

厳粛な同志的批判
serious and comradely criticism

共産党主義的 きまじめさ
communist rectitude

特別臨時ニュース
a special bulletin

ただし、同じ様な例でも、文節の中の自立語として文節レベルで対応させることの出来るものもある。

灰色の / 人民服 を / 着ていた
Sporting a gray Mao suit

(b) 2文節以上でひとまとまりの意味を持ち、全体として相手言語と対応するもの。

こちらからも / 連絡 が / 取れる / 訳です。
we may contact you

厳しい / 態度 で / 臨む
deal severely

声高な / 主張 が / 一気に / 噴き出した / こと
the bursts of hard-line rhetoric

これの極端な例として、構文が全面的に組替えられていることがあり、その場合は文全体を対応させた。

御用件は何でしょうか？
May I assist you?

党は / 事態 を / 完全に / 掌握 している
the party was in control

(c) イディオム

By the way,
話は変わりますが

for the moment
差し当りは

in favor of political liberalization
政治的自由化を支持している

(d) イディオムではないが、日本語との対応という点で、ひとまとまりの表現とみなした方がよいもの。

One last item : could you give us your name
それから、御名前をお聞きしたい

the same distance to Kyoto International Conference Hall
距離はどれも変わりません。

a leading astrophysicist
天体物理学の第一人者

have a greater say
発言力を強めている

(e) 法や待遇を表す表現

I would like to attend the conference.
会議に参加したいのですが。

詳しいことをお知らせ致しましょうか。
Shall I send you the information?

将来もそうあり続けるだろう
and would stay that way

厳しい態度で臨まなければならない
we must deal severely with

(f) 単語または文節で対応付けられているが、前後の語を加えてもなお対応の取れるもの。

外国から(徐々に)入り込んできた
crept in from abroad

それでも(なお)
all the same

言論の自由要求(の声)
calls for free speech

(g) その述語が相手言語にはなく、しかも全体としては対応付けることのできるもの。

重大な政治問題で犯した誤り
his mistakes on major political issues

今月行われた学生デモ
the student demonstration this month

これら(a)~(g)のうち、イディオムは単位として特殊であり、後でイディオム辞書の作成に利用できるように、Iマークを付与することとした。(I stands for Idiom)

実際の言語データベース化にあたっては、「日→英」方向と「英→日」方向の両方の対応付けを行うことは、重複が多いのでコスト上問題があるが、このランダムレベルを導入することにより「英→日」方向でしかとれない対応は吸収でき、基本的に「日→英」方向の対応付けのみを行えば良いということが明らかとなった。

4.おわりに

本報告では言語データベースに「日英対訳対応付け」を行うためのレベル設定について説明し、問題点およびその対応方法について述べた。

今後は今回の結果を利用し、言語データベース化を進めるとともに対訳対応付けのノウハウを更に細かな作業規則として蓄積していく予定である。

<謝辞>

本研究の機会を与え、ご指導下さった樽松明社長に感謝します。また、熱心に討論して下さいったデータ処理研究室および日本アイアール株式会社の諸氏に感謝致します。

<参考文献>

- [1] 森元・小倉・飯田：
「自動翻訳電話研究用言語データベースの収集について」、
情報処理学会第36回全国大会4U-5、1988
- [2] 小倉・篠崎・森元：
「言語データベース収集支援システム」、
情報処理学会第36回全国大会4U-4、1988
- [3] 篠崎・小倉・森元：
「言語データベースの品質管理」、
情報処理学会第36回全国大会4U-3、1988
- [4] 吉本：
「日本語品詞の分類」、TR-I-0008
- [5] 井ノ上、小倉、森元：
「言語データベース用格、係り受け意味体系」、TR-I-0029