

TR-IT-0191

日独変換主導翻訳の中間時評価

Report on Mid-term performance evaluation of Transfer-Driven

Machine Translation between Japanese and German

Michael Paul

Yasuharu Den

Osamu Furuse

伝 康晴

古瀬 蔵

1996 年 9 月

要旨

本稿では変換主導翻訳 (TDMT) の日独翻訳モジュールの評価結果 (1996 年 7 月 30 日) について報告する。まず、評価の概要とおのこの評価者の評価結果について述べた後、翻訳過程で生じる問題点を議論し、システムの性能を改善するための提案について述べる。

エイ・ティ・アール音声翻訳通信研究所

ATR Interpreting Telecommunications Research Laboratories

©(株) エイ・ティ・アール音声翻訳通信研究所 1996

©1996 ATR Interpreting Telecommunications Research Laboratories

もくじ

1	はじめに	1
2	翻訳結果の評価	1
2.1	概要	1
2.2	評価の質	2
2.3	評価の結果	4
2.4	評価判断の理由	7
3	翻訳の問題点: パターン + 語彙選択 + 文脈	8
3.1	句の訳し間違い	8
3.2	句の訳し間違い(文脈依存)	12
3.3	語彙選択の誤り(文脈依存)	14
3.4	逐語訳が不可能な場合	16
3.5	SUB/OBJ/IND-OBJ の省略	17
3.6	重複	18
3.7	P-DATA + loss of input parts	18
3.8	P-DATA + default rules	19
3.9	P-DATA (その他)	21
4	生成の問題点	22
4.1	埋め込みによる情報の喪失	22
4.2	等位 / 従属節における主語の一致	23
4.3	冠詞の挿入	23
4.4	その他	24
5	おわりに	24
	参考文献	24

図一覧

1	実行時間	6
---	----------------	---

表一覧

1	評価のズレ	3
2	TT の分布	5
3	TS の分布	5
4	評価値 A, B, C	5
5	構造正解率	6
6	判断の理由	7

1 はじめに

変換主導翻訳 (TDMT) プロジェクトでは、未知データに対するシステムの性能を調べるために評価試験 (オープンテスト) を行なっている。本稿では、日独翻訳モジュールの評価結果について報告する。多言語 TDMT システムの他の言語ペアのモジュールの評価については、別稿 (日英間: [2], 日韓間: [3]) を参照されたい。

まず、評価試験の概要について述べた後、日独翻訳モジュールの評価を行なったおのおのの評価者の評価結果を比較し、次に、その評価結果をもとにしてシステムの性能を検討する。さらに、システムの性能を改善するために、翻訳過程のおのおのの部分 (形態素解析, 構造解析, 変換処理, 生成) について調べる。

形態素解析の誤りによって単語の分類に間違いが生じ、変換辞書へのアクセスや変換規則の適用が失敗するなどして、翻訳の精度が落ちる。しかし、翻訳の問題点のほとんどの部分は、むしろ、構造解析 (変換規則のマッチング処理における誤ったパターンの選択) や生成 (語順や活用の誤り) において生じる。これらの問題点について、具体例をあげるとともに、改善のための提案を行なう。

2 翻訳結果の評価

2.1 概要

システムデータのサイズ

学習文	のべ	931
	異なり	855
形態素合成	語彙規則	421
基本マーカ挿入	ローカル規則	86
変換	規則	541
	用例	2941
単語変換辞書項目数		4937
評価文	のべ	1247
	異なり	1021

入力データ

対象: 日英バイリンガル 69 会話の日本語話者発話部分 (比搔検討のため日英, 日韓, 日独)

フォーマット: 形態素タギングデータ (形態素解析の問題を分離)

評価者	(1) TT
	(2) TS
翻訳出力の評価値	A no problem
	B fair
	C acceptable
	D nonsense, wrong sense
	NIL no output

2.2 評価の質

評価値 NIL 以外の文 (のべ: 876) について 2 人の評価のズレを比較した (表 1 参照). 表で 'diff = X' は 2 人の評価値に距離 X のズレがあることを示す (例えば BD は距離 2 のズレ). また, 'positive' はどちらの評価者がどれくらいの回数甘い評価をしたかを表す. 例えば, 'TT: 296 (86.1%)' はズレの距離が 1 であった 344 文のうち, 評価者 TT のほうが甘く評価したものが 296 文あったことを示す.

90.7% の文については 2 人のズレが距離 1 以内に納まっているので, 評価の質としては十分高いといえる. しかし, 2 人の評価の下し方には一般的な傾向があり, TS は TT よりも厳しい評価を下している. 第一著者自身が評価文を調べたところ, TT の評価が甘過ぎると思われるものが数文あったが, TS の厳し過ぎる評価は多くの場合単に間違いであった (TS の評価には同意できない部分が多い). 以下に例をあげる.

AT120072(28)	
source:	“番号は八三八一四零零です” (morph: 10)
target:	Die Nummer ist acht drei acht eins vier null null. (5.0e-6)
rankTT:	A
rankTS:	C → style info determiner
著者の評価:	A

表 1: 評価のズレ

diff	計 (%)	評価値 (%)	positive
0	450 (51.4%)	AA (26.4%)	-
		BB (11.3%)	
		CC (16.7%)	
		DD (45.6%)	
1	344 (39.3%)	AB (33.2%)	TT: 296 (86.1%)
		BC (27.9%)	
		CD (25.0%)	
		BA (2.9%)	TS: 48 (13.9%)
		CB (2.0%)	
		DC (9.0%)	
2	78 (8.9%)	AC (80.7%)	TT: 74 (92.8%)
		BD (14.1%)	
		CA (0.0%)	TS: 4 (5.2%)
		DB (5.2%)	
3	4 (0.4%)	AD (50.0%)	TT: 2 (50.0%)
		DA (50.0%)	TS: 2 (50.0%)

AT130051(17)

source: “それではお客様のお名前をお願いいたします” (morph: 10)
 target: Also, Ihren Namen bitte. (7.4999994e-6)
 rankTT: A
 rankTS: C → selection style info
 著者の評価: A

CT220042(11)

source: “お値段のほうはどうなっていますか” (morph: 10)
 target: Wie wird der Preis? (0.2666792)
 rankTT: B → selection style
 rankTS: D → “Wrong word selection, sentence does not make sense.”
 著者の評価: B
 comment: “Wie hoch ist der Preis”

TAS13015(33)	
source:	“宿泊は十月二十五日から五泊” (morph: 9)
target:	Die Übernachtung für fünf Nächte vom fünfundzwanzigsten Oktober an. (0.0)
rankTT:	A
rankTS:	C → selection style
著者の評価:	A
TAS32001(35)	
source:	“じゃ朝食の方も予約しておいてください” (morph: 9)
target:	Also, bitte reservieren Sie auch ein Frühstück. (1.0e-5)
rankTT:	A
rankTS:	C → selection style
著者の評価:	A
TCS32012(8)	
source:	“じゃあ料金の方はおいくらぐらいでしょう” (morph: 8)
target:	Wie teuer ist die Gebühr? (1.00001)
rankTT:	A
rankTS:	C → selection style info
著者の評価:	A

さらに、TSは「同等な」文に異なる評価値を与えている場合もあった。例えば、以下の例は「です」のあるなしの違いしかないので、翻訳結果は同等であり、それゆえ、評価値も同じになるべきである。

AT120072(18)	
source:	“番号は五二七九三九二零” (morph: 10)
target:	Die Nummer ist fünf zwei sieben neun drei neun zwei null. (0.0)
rankTS:	B → style determiner
AT120072(28)	
source:	“番号は八三八一四零零です” (morph: 10)
target:	Die Nummer ist acht drei acht eins vier null null. (5.0e-6)
rankTS:	C → style info determiner

評価者がどのような理由でその評価を下したかをより良く理解するためには、評価者自身による(正解)翻訳文を付記してもらうことが望ましい。これは次回のオープンテストの評価の際に導入してよいだろう。

2.3 評価の結果

まず、おのおの評価値の分布は表2、表3の通りである。

TDMTの翻訳結果として使いものになるもの(評価値A, B, C)についてまとめると、表4のようになる。日英(60.2%)や日韓(70.5%)に比べてかなり劣っている。

表 2: TT の分布

評価値	のべ	(%)	異なり	(%)
A	298	23.897355	165	16.160635
B	168	13.472335	96	9.4025465
C	168	13.472335	156	15.279145
D	242	19.406585	234	22.918715
NIL	371	29.751450	370	36.238985
計	1247	-	1021	-

表 3: TS の分布

評価値	のべ	(%)	異なり	(%)
A	131	10.505215	77	7.5416260
B	176	14.113875	85	8.3251715
C	265	21.251500	192	18.805095
D	304	24.378515	297	29.089135
NIL	371	29.751450	370	36.238985
計	1247	-	1021	-

表 4: 評価値 A, B, C

評価者	のべ (1247)	異なり (1021)	評価者	のべ (876)	異なり (651)
TT	50.8%	40.8%	TT	72.4%	64.0%
TS	45.9%	34.7%	TS	65.3%	54.4%
平均	48.4%	37.8%	平均	68.9%	59.2%

(a) すべての文に対する割合

(b) NIL 以外の文に対する割合

表 5: 構造正解率

評価者	のべ (1247)	(%)	異なり (1021)	(%)
正解	678	54.4%	457	44.7%
誤り	208	16.7%	204	20.0%
NIL	361	28.9%	360	35.3%

一方、原言語構造の正解率は表 5 の通りである。

次に、実行時間は入力文の形態素数が増えるほど長くなる (図 1 参照)。形態素数が大きな部分 (morph>16) における変動は、評価データのサンプル数が少ないことによる。

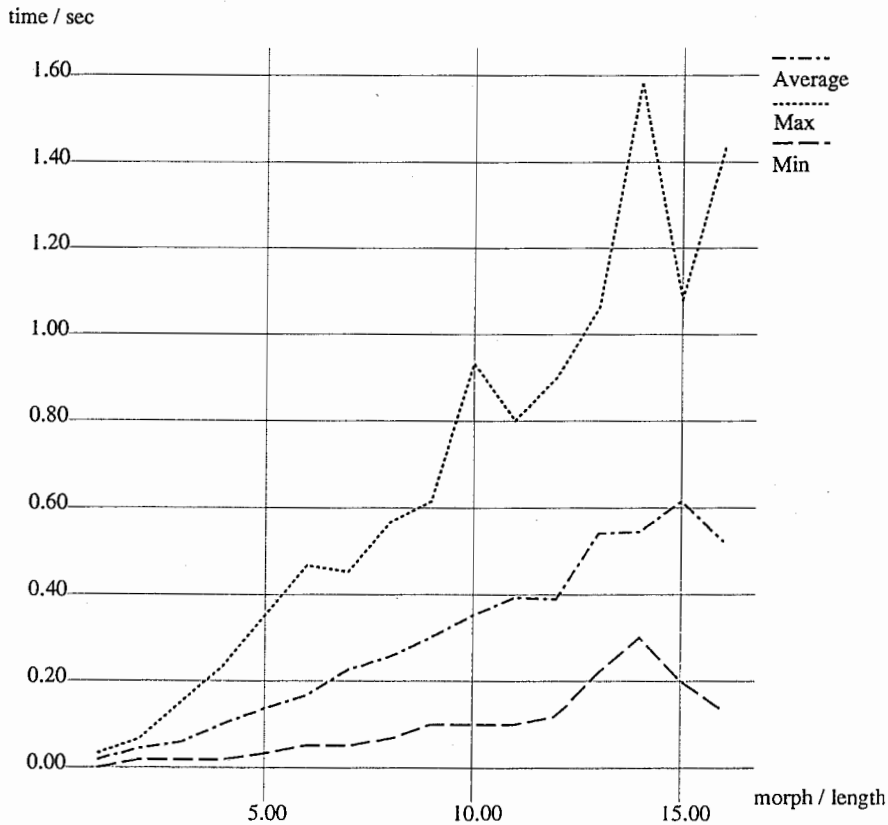


図 1: 実行時間

2.4 評価判断の理由

評価者は自分が下した判断の理由を選べるようになっている。現在のところ、以下の7つの誤りのカテゴリから理由を選ぶようになっている。

1 - selection 3 - order 5 - info 7 - others
2 - conjugation 4 - style 6 - determiner

おのおののカテゴリの分布は表6の通り。

表6: 判断の理由

評価者	評価値	1	2	3	4	5	6	計
TT	B	55	6	17	101	13	5	197
	C	105	34	98	40	18	8	303
	D	176	56	141	24	92	5	494
	計	336	96	256	165	123	18	994
TS	B	116	0	4	163	2	11	296
	C	250	17	58	256	20	70	671
	D	21	5	11	21	3	12	73
	計	387	22	73	440	25	93	1040
計		723	118	329	605	148	111	2034

“order” と “info” 以外はより詳細な理由が付記されているほうが良いだろう。例えば、以下のような具合である。

- selection
- どの語の翻訳がまずいのか。
 - どの文脈?
 - 正しい翻訳の例
- conjugation
- 時制, 数, 人称などのより詳しい情報
- style
- どのようなスタイルに関して (formal/informal/colleague/...)
 - どの文脈?
 - 正しい翻訳の例
- determiner
- definite/indefinite/none

しかし、特定の理由が頻繁に選ばれていることから、いくつかのヒントは与えられる(この話題は将来もっと詳細に検討すべきだが)。上記の結果によると、もっとも重大な問題は語彙選択(selection)(36%)のようだ。語彙選択は文脈に依存する(例: “行く” – “gehen” (human), “fahren” (train), “fliegen” (airplane))。現在のアプローチでは、翻訳辞書は入力側の単語を出力側のもっとも一般的な単語に1対1に対応づけるようになっている。これは、(少なくとも)理解可能な翻訳文を出力することを目指しているからである。しかし、文脈に依存する単語翻訳を正確に行なうためには、意味辞書のようなものが必要になる(表 [1])。

もう一つの問題はスタイル(style)(30%)の扱いである。2.2 節で述べたようにTSの評価はうのみにはできないが、日本語のさまざまな敬語表現を目標言語の適切な表現に翻訳する必要があるということはいえる。

語順(order)(16%)に関する問題は、翻訳結果中で言語的情報を過度に指定し過ぎていることによる。この点は4.1 節でより詳しく述べる。

対話の文脈に関する詳細な情報が与えられるようになると、翻訳の情報内容(info)(7%)や冠詞の選択(determiner)(6%)に関する誤りも減らすことができるだろう。

活用(conjugation)(6%)の問題は主に、時制の体系が日本語(現在, 過去)とドイツ語(現在, 過去, 現在完了, 過去完了, 未来)で異なることによる。

3 翻訳の問題点: パターン + 語彙選択 + 文脈

以上の結果をふまえて、翻訳レベルでの問題点について考察する。問題のパターンと関連する例(オープンテストの出力例)を以下にあげる。

3.1 句の訳し間違い

“お客様”

これは単に丁寧さを表すだけの表現であり、翻訳に必要な統語的・語彙的な情報は含んでいない。

⇒ 単に無視すれば良い(少なくとも以下の例では)。(DUMMY “Gast”)のような翻訳辞書項目を使って空文字列を生成する。

AT130082(49)	
source:	“お客様のお越しをお待ちしております” (morph: 9)
target:	Wir freuen uns auf Ihren Besuch <i>Gast</i> . (0.5)
⇒	“Wir freuen uns auf Ihren Besuch”

OT330021(11)	source:	“お客様金沢はいかがでしょう” (morph: 5)
	target:	<u>Lieber Gast</u> , wie ist Kanazawa? (1.1428572)
	⇒	“Wie ist Kanazawa”/“Wie gefällt Ihnen Kanazawa?”
TAS13015(26)	source:	“それではお客様のお名前とお電話番号をお願いします” (morph: 13)
	target:	Also, Ihren Namen und die Telefonnummer bitte. (1.2499998e-5)
	⇒	“Gast” dropped using transfer rule ‘お客様の Y → Y’

姓と名

姓と名を含む名前は1つのまとまりとして扱う必要があり、また、敬称の性 (“Frau/Herr”) は姓ではなく名に依存する。

⇒ P-DATA rules 中で名前は名詞句として扱う必要があり、その際、主要部の情報は名のほうから取り出す。

CT120051(3)	source:	“五〇八号室の鈴木和夫です” (morph: 9)
	target:	<u>Suzuki</u> vom Zimmer fünf null acht <u>Kazuo</u> . (2.4999996e-5)
	⇒	“Kazuo Suzuki vom Zimmer fünf null acht.”
AT130082(24)	source:	“メアリーフィリップス様” (morph: 3)
	target:	<u>Mary Herr</u> Phillips. (1.4999999e-5)
	⇒	“Frau Mary Phillips”
KT330011(42)	source:	“それではジョンフィリップス様あしたの朝八時にお待ちしております” (morph: 15)
	target:	Also, <u>John Herr</u> Phillips. Wir warten morgen um acht Uhr morgen. (0.50003207)
	⇒	“Also, Herr John Phillips, wir warten morgen um acht Uhr morgen.”

名 + “様”

ドイツ語では、名だけの場合は敬称 (“Herr/Frau”) をつけない。

TAS33002(31)	source:	“ファーストネームがジョン様でラストネームがフィリップス様でございますね”
	target:	Der Vorname ist <u>Herr John</u> ? Und der Familienname ist Herr Phillips, nicht wahr. (3.9999998e-5)
	⇒	Der Vorname ist John? Und der Familienname ist Phillips, nicht wahr?

“N 名様”

“大人 / お子様 / ...” と組み合わせ使った場合、“N Personen Erwachsene/Kinder” のような冗長な翻訳 (“Personen” も “Erwachsene” も人を表す) になってしまう。

⇒ “N Erwachsene/Kinder” となるべきで, “Person”(“名様”)は無視しなければならぬ. しかし, そのような P-DATA rule を書くと, “loss of input parts” の問題が生じる (3.7 節参照). 別の解決策は “Erwachsene/Kinder N Personen” と訳すことである (あまり自然ではないが).

TAC23013(37)	
source:	“それからご宿泊人数のほうは大人の方四名様でよろしかったでしょうか” (morph: 18)
target:	Und die Übernachtung ist in Ordnung <u>vier Personen Erwachsene</u> gewesen, richtig? (5.0116487)
⇒	“Und die Übernachtung ist für vier Erwachsene gewesen, richtig?”

TAC23013(40)	
source:	“では大人の方二名様お子様二名様ですね” (morph: 13)
target:	Also, <u>zwei Personen Erwachsene, zwei Personen Kind.</u> (2.888899)
⇒	“Also zwei Erwachsene und zwei Kinder.”

日付・時間表現

日付・時間表現で問題になるパターンは以下の通り.

“N 年 N 月”/“N 年の N 月”	“N 日か N 日”	“N 日に”
“N 日から”	“N 日と N 日”	“N 日 <CN-CN>N 日”
“N 日から N 日まで”	“N 日ですと”	“X 時と Y 時”
“N 日から N 日までの”	“N 日の”	“午後 X 時と Y 時”
“N 日から N 日の”	“N 日の方”	

⇒ 上記のパターンについて新たに規則を作る. “N 年 N 月” と “X 時と Y 時” を除くパターンについては, 「序数 + 品詞」の情報を使って, 入力から序数表現を抜き出す. 例えば, “十六日から” は “from the 26th” のように訳す (“26 days” ではない). “N 年 N 月” は特別な扱いが必要: “1996 年” は “nineteen ninety-six” のように訳す (“one nine nine six” ではない). “午後 X 時と Y 時” では “午後” は “X 時” と “Y 時” の両方にかかる.

AT120072(20)	
source:	“有効期限は九十六年四月です” (morph: 8)
target:	Gültiger Termin <u>sechsendneunzig und nach dem Jahr.</u> (1.6889038)
⇒	“Der gültige Termin ist April neunzehnhundert sechsendneunzig.”

CT130032(14)	source:	“五百八号のフィリップス様に二十六日から三十日までジャパントイムスをお届けいたします” (morph: 18)
	target:	Gibt die Japan Times an fünfhundertacht Herrn Phillips von <u>sechszwanzig Tagen</u> bis zu <u>dreißig Tagen</u> ab. (2.6944711)
	⇒	“Die Japan Times an fünfhundertacht Herrn Phillips vom sechszwanzigsten bis zum dreißigsten.”
OT230111(17)	source:	“大阪城発の運行時間ですが午後六時と七時がございますので” (morph: 16)
	target:	Es gibt die Fahrt ums Schloß Osaka <u>sieben Uhr</u> . <u>Achtzehn Uhr</u> . (3.6313224)
	⇒	“Es gibt die Fahrt um das Schloß Osaka um achtzehn und neunzehn Uhr.”
OT320031(18)	source:	“そうですかでは三日からお願いします” (morph: 12)
	target:	Ach so. Also, ab <u>drei Tagen</u> bitte. (0.9444495)
	⇒	“Ach so. Also, ab dem dritten bitte.”
OT330071(9)	source:	“祇園祭のハイライトは十七日の山鉾巡行です” (morph: 10)
	target:	Im Höhepunkt vom Gion Fest ist Umzug in die Yamahoko <u>siebzehn Tage</u> . (1.8750134)
	⇒	“Der Höhepunkt vom Gion Fest ist der Yamahoko Umzug am siebzehnten.”
TAC23024(11)	source:	“十四日十五日ともツインのほうご用意できます” (morph: 12)
	target:	Ich kann ein Zweibettzimmer für beide für <u>vierzehn Tage</u> , <u>fünfzehn Tage</u> , arrangieren. (3.305555)
	⇒	“Ich kann ein Zweibettzimmer am vierzehnten und fünfzehnten arrangieren.”
TAS13008(9)	source:	“十月二十七日と二十八日ですね” (morph: 9)
	target:	Siebenundzwanzigster Oktober und <u>achtundzwanzig Tage</u> . (0.66667503)
	⇒	“Siebenundzwanzigster und achtundzwanzigster Oktober.”
TAS13008(39)	source:	“では二十七日にお待ちしております” (morph: 9)
	target:	Also, wir warten <u>auf siebenundzwanzig Tage</u> . (0.5833358)
	⇒	“Also, wir freuen uns auf den siebenundzwanzigsten.”

数量表現 “X が二つ”/“X を二つ”

名詞句 “2 X” に翻訳されるべき。現状では、数量と X の間に関係がつかない。

⇒ 新しい規則: ‘n+n’ → (NP !Y !X)/(SUB ...)/(OBJ ...)

BT320011(34)	source:	“そのダブルベッドが二つある部屋ということですね” (morph: 10)
	target:	Das, daß Zimmer, das <u>ein Doppelbett zwei</u> hat. (1.1666842)
	⇒	“... zwei Doppelbetten ...”
TAC23013(36)	source:	“九月十一日一泊でツインのお部屋を二つ” (morph: 13)
	target:	<u>Zwei</u> für eine Nacht am elften September das <u>Zweibettzimmer</u> . (0.9722298)
	⇒	“... zwei Zweibettzimmer ...”

3.2 句の訳し間違い (文脈依存)

以下のパターンは翻訳が曖昧である。

⇒ 文脈情報が必要

“お願い” (頻度: 98)

⇒ “bitte/ich bitte darum/vielen Dank/auf wiedersehen/...”

BT320011(11)	source: “ぜひお願いいたします” (morph: 5)
	target: Ich bitte. (0.8333333)
BT320011(37)	source: “お願いします” (morph: 4)
	target: Bitte. (0)
OT320031(28)	source: “はいそれでお願ひします” (morph: 7)
	target: Ja, das bitte. (0.0)
OT320031(32)	source: “はいでは七月三日に伺いますのでどうぞよろしくお願ひします” (morph: 16)
	target: Ja, also, ich frage am dritten Juli. Bitte, vielen Dank. (1.9416666)
OT330021(48)	source: “どうぞよろしくお願ひします” (morph: 6)
	target: Bitte, vielen Dank. (1.0)

“そうです” (頻度: 85)

⇒ “na ja/also/ach so/gut/dann ist es ja gut/inOrdnung/...”

BT320011(15)	source: “そうですね” (morph: 3)
	target: Na ja. (0)
TAS13015(44)	source: “そうですか” (morph: 3)
	target: Ach so. (0)

“かしこまりました” (頻度: 50)

⇒ “in Ordnung/es ist in Ordnung/ja/ja natürlich/gern geschehen ...”

AT130051(30)	source: “かしこまりました” (morph: 1)
	target: In Ordnung. (0)

“よろしい” (頻度: 28)

⇒ “ist es in Ordnung?/ist es Ihnen recht?/stimmt es/stimmt das/...”

BT320011(48)	source: “よろしいですか” (morph: 3)	target: Ist es in Ordnung? (5.0e-6)
TAS13008(36)	source: “二名様でデラックスツイン二万一千円のお部屋をご予約ということでよろしいでしょうか” (morph: 19)	target: Es ist zwei Personen das Luxusdoppelzimmer, das Zimmer zu einundzwanzigtausend Yen in Ordnung bitte, richtig? (2.555563)
TAS13015(35)	source: “これでよろしいでしょうか” (morph: 5)	target: Das ist in Ordnung, richtig? (0.11111945)

“分かりました” (頻度: 26)

⇒ “In Ordnung/Okay/Ich habe verstanden/Ach so/...”

AT120072(40)	source: “分かりました” (morph: 3)	target: Okay. (5.0e-6)
CT220042(21)	source: “はい分かりました” (morph: 4)	target: Ja, okay. (5.0e-6)

“と言います / と申します” (頻度: 16)

⇒ “heißen/mein Name ist/sein...”

CT220051(10)	source: “わたしは前田由美と言います” (morph: 7)	target: Ich heiße <i>mich</i> sein Maeda Yumi. (0.16668418)
OT320031(24)	source: “はいわたしの名前は鈴木和夫と申します” (morph: 10)	target: Ich heiße <i>Ihren Namen sein</i> Kazuo Suzuki. (0.50002253)
OT320031(25)	source: “名字が鈴木で名前が和夫です” (morph: 8)	target: Der Familienname ist Suzuki und Ihr Name ist Kazuo. (2.9999998e-5)
OT330141(15)	source: “ええ清水焼は清水寺の近くにゆかりがありますので清水焼と言います” (morph: 16)	target: Ja, es hat eine Verbindung die Shimizu Keramik in der Nähe vom Kiyomizudera Tempel. <i>Ich heiße</i> Shimizu Keramik. (1.4333634)

“失礼いたします” (頻度: 7)

⇒ “Auf Wiedersehen/Auf Wiederhören/bis bald/tschüß/...”

CT130032(17)	
source:	“失礼いたします” (morph: 1)
target:	Auf Wiedersehen. (0)

“教えてください” (頻度: 7)

⇒ “informieren ueber/unterrichten/empfehlen/...”

TCS12023(38)	
source:	“はい教えてください” (morph: 3)
target:	Bitte sagen Sie mir etwas. (5.0e-6)

“ありがとうございます” (頻度: 6)

⇒ “Vielen Dank/Ja bitte/...”

AT130082(4)	
source:	“ありがとうございます” (morph: 1)
target:	Vielen Dank. (0)

TAC23013(1)	
source:	“ありがとうございます京都観光ホテルでございます” (morph: 4)
target:	Vielen dank, Kyoto Kanko Hotel. (1.0e-5)

3.3 語彙選択の誤り (文脈依存)

数量接辞

翻訳は計量されているものの種類による。文中に計量の対象が明示的に出てくる (下例の“水割り”, “ベッド”) なら, 接辞はいらない。

CT320062(6)	
source:	“寝る前に一杯水割りが飲みたいんですが” (morph: 12)
target:	Vor eins und ich möchte gerne nach der <i>Schale</i> trinken und schläft. (2.522244)
⇒	“Bevor ich schlafen gehe, möchte ich noch ein Glas Whiskey mit Wasser trinken.”

TAS13015(15)	
source:	“いいえベッドは二台です” (morph: 6)
target:	<i>Ständer</i> nein, zwei ein Bett. (0.33335337)
⇒	“Nein, es gibt zwei Betten.”

“ございます” (頻度: 128)

⇒ “wir haben/es gibt/...”

OT330021(6)	
source:	“はいございますよ” (morph: 3)
target:	Ja, wir haben. (5.0e-6)

“確認いたします” (頻度: 21)

⇒ “確認” → “nachschauen/wiederholen/bestätigen/...”

AT130082(23)	
source:	“確認いたします” (morph: 2) <--- number of morph wrong !!
target:	Ich schaue nach. (0.0)

TAS13008(27)	
source:	“確認いたします” (morph: 3)
target:	Ich schaue nach. (0.0)

TAS13015(31)	
source:	“はいそれでは確認させていただきます” (morph: 7)
target:	Ja, also, ich wiederhole. (0.6666667)

“結構” (頻度: 13)

⇒ “nicht nötig sein/in Ordnung sein/genug sein/ausreichen/...”

DT120032(17)	
source:	“いえ今は結構です” (morph: 5)
target:	Nein, jetzt ist nicht nötig. (1.500005)

DT120032(22)	
source:	“はい今は結構です” (morph: 5)
target:	Ja, jetzt ist nicht nötig. (0.500005)

KT330011(34)	
source:	“予約金の方はもうそのままお預かりいたしますのでキャンセル料金などはもう結構でございます” (morph: 19)
target:	Die Anzahlung bewahrt noch auf. Die Gebühr der Annullierung und so weiter ist in Ordnung. (4.108)

OT320031(19)	
source:	“ツインで結構です” (morph: 4)
target:	Es reicht Zweibettzimmer aus. (0.2916742)

“コース” (頻度: 10)

⇒ バス: “Tour” / 料理: “Angebote” / その他: “Kurs”

DT120032(7)	
source:	“コースについて教えてください” (morph: 4)
target:	Bitte sagen Sie mir etwas über die Tour. (0.9166667)

“入れる” (頻度: 3)

⇒ “hineintun”/ 本を: “eintragen”/ 予約を: “machen”/ “eintragen”

OT320031(22)

source: “それじゃあそれで予約を入れてください” (morph: 7)

target: Also, bitte tun Sie die Reservierung die hinein. (1.5555606)

“キロ” (頻度: 1)

⇒ “Kilometer/Kilogramm”

TCS32005(10)

source: “西に一キロ行って右手右手ですね” (morph: 10)

target: West eins und ich gehe ein Kilogramm und nach der rechten Hand. (1.833346)

“お電話お待ち”

⇒ “Telefonanruf erwarten/...”

TCS33035(33)

source: “それではお電話お待ちしております” (morph: 8)

target: Also, wir warten ein Telefon. (0.87501)

3.4 逐語訳が不可能な場合

“申し遅れました”

DT320012(33)

source: “申し遅れました” (morph: 3)

target: Hat verspätet informiert. (5.0e-6)

“いいえどういたしまして”

⇒ “いいえ”は無視

OT330122(19)

source: “いいえどういたしまして” (morph: 2)

target: Nein, gern geschehen. (5.0e-6)

“いいいえまだなんです”

⇒ “Nein, noch nicht”

TCS32012(25)	
source:	“いいえまだなんです” (morph: 5)
target:	Nein. Noch. (1.4999999e-5)

“心配”

⇒ “kein Sorge/es ist in Ordnung/...”

TCS13005(14)	
source:	“その心配はございません” (morph: 4)
target:	Es gibt die Sorge nicht. (0.4583417)

TCS13019(21)	
source:	“ご心配ありません” (morph: 4)
target:	Man braucht nicht Sorge. (0.8333433)

“申し上げます”

TCS33050(8)	
source:	“はい申し上げます” (morph: 3)
target:	Ich sage Ihnen. (1.0e-5)

3.5 SUB/OBJ/IND-OBJ の省略

日本語の1つの特徴は主語や目的語の省略である。目的言語の生成には、省略されたゼロ代名詞の指示対象を以前に言及された事物(通常名詞句)から見つけておかなければならない。そのためには、何らかの短期あるいは長期記憶(例えば、ゼンタリング機構 [3])を使って、対話中で言及された事物を後続の発話の処理のために憶えておく必要がある。現在はこのような文脈処理機構がないので、省略された主語や目的語を適切に補完することはできない。

⇒ しかし、動詞の自動性/他動性の情報は翻訳辞書に簡単に追加できる。すなわち、デフォルトの目的語を動詞の辞書に記述しておく。これによって、本来あるべき目的語が翻訳結果中に存在しない場合は、デフォルトの目的語を与えることで文法的に正しい文が生成できる。ただし、2つの例外があり、以下の場合は挿入されたデフォルトの目的語を落さなければならない: 受動態およびwh-疑問文。

AT130082(15)	
source:	“これは一名様のご宿泊料金でございます” (morph: 11)
target:	<u>Das</u> ist Gebühr der Übernachtung für eine Person. (2.0833385)
⇒	“der/die/das”

AT130082(35)	source: “烏丸口を出て市バス二百六番にお乗りください” (morph: 11)
target: <u>Man</u> geht den Karasuma Eingang hinaus. Und bitte nehmen <u>Sie</u> einen städtischen Bus Nummer zweihundertsechs. (1.2000117)	
⇒	different default subjects (“man”/“Sie”) in both sentence.
CT130051(30)	source: “二千円でございます” (morph: 4)
target: <u>Das</u> kostet zweitausend Yen. (0.0)	
⇒	“er/sie/es”
TCS32005(28)	source: “それ探してたんですよ” (morph: 8)
target: <u>Ich</u> habe das gesucht, nicht wahr? (0.750015)	
⇒	wrong SUB “ich” → “Sie”
TCS33035(21)	source: “それでしたら御用意できると思います” (morph: 7)
target: Also, <u>ich</u> glaube, <u>Ø</u> kann es arrangieren. (1.7904092)	
⇒	(SUB “ich”) in second part !

3.6 重複

同じような意味を持つ句が組み合わさって使われている場合に、冗長な出力が生成されることがある。

⇒ 1つの解決策は、入力中の同じ意味の句の一方を無視すること。しかし、これは危な過ぎる (cf. “loss of input parts”).

AT130051(31)	source: “それではお越しをお待ち申し上げます” (morph: 8)
target: Also, ich freue mich <u>auf Ihren Besuch auf Ihren Besuch</u> . (1.4999999e-5)	
TCC23012(18)	source: “シングルルームご用意できますのでどうぞお越しくださいませ” (morph: 10)
target: Ich kann ein Einzelzimmer arrangieren. <u>Bitte, bitte</u> kommen Sie. (1.8214388)	
TAC23013(7)	source: “はいツインのお部屋二部屋ほど空いております” (morph: 11)
target: Es ist frei etwa <u>zwei Zimmer des Zweibettzimmers</u> . (2.0238194)	
⇒	“... zwei Zweibettzimmer ...”
TAS33002(27)	source: “それではツインルームのシングルユース一万四千円のお部屋をお取りいたします” (morph: 14)
target: Also, ich nehme <u>ein Zimmer</u> zu vierzehntausend Yen mit dem <u>Doppelzimmer</u> für eine Person. (0.8333384)	
⇒	“... ein Doppelzimmer zu vierzehntausend Yen ...”

3.7 P-DATA + loss of input parts

入力の一部を無視するような規則で頻繁に問題が生じる。すなわち、入力パターン “X [pattern] Y” を単に X(または Y) に変換するような場合である。これらの規則

は、学習文では正しく動くが、オープン試験文では重要な情報(動詞, 主語など)を無くしてしまうことがある。

(KAKUJO-DE-NP) NP :(?X で ?Y) → (!X in ordnung sein) ... ((席)(結構))
(KAKUJO-DE-NP) NP :(?X で ?Y) → (!X in ordnung sein) ... ((それ)(良い))
(KAKUJO-DE-NP) NP :(?X で ?Y) → (!X in ordnung sein) ... ((部屋)(十分))
(KAKARIJO-HA-NP) NP :(?X は ?Y) → (!Y) ... ((時間)(N時))
(KAKARIJO-HA-NP) NP :(?X は ?Y) → (!Y) ... ((料金)(N円になる))
(RENTAIJO-NO-N+N) N+N :(?X の ?Y) → (!Y) ... ((お客様)(電話番号))
(MARKER-CN-CN-IS) IS :(?X <CN-CN> ?Y) → (jeweils !Y) ... ((それぞれ)(とる))
(MARKER-CN-CN-N+N) N+N :(?X <CN-CN> ?Y) → (!Y) ... ((どこか)(ところ))
(MARKER-SN-CN-N) N :(?X <SN-CN> ?Y) → (!X) ... ((観光)(コース))
(MARKER-SN-CN-N2) N :(?X <SN-CN> ?Y) → (!X) ... ((宿泊)(分))
(MARKER-PRON-CN-N) N :(?X <PRON-CN> ?Y) → (!X) ... ((いつ)(頃))

(RENTAIJO-NO-N+N) N+N :(?X の ?Y) → (!X) ... (() (者))
(MARKER-CN-CN-N) N :(?X <CN-CN> ?Y) <i>rightarrow</i> (nach !X) ... (() (行き))
(MARKER-/KIHON-FN-S+N) S+N :(?X </KIHON-FN> ?Y) → (wo man !X) ... (() (ところ))
(IMP-KAKUJO-WO-NP) NP :(?X を ?Y) → (!X bitte) ... (() (おねがいします))
(KAKARIJO-DEHA-NP) NP :(?X では ?Y) → (damit !Y) ... ((それ)())
(CONJ-KEREDOMO-SE) SE :(?X けれども) → (!X) ... (())

3.8 P-DATA + default rules

目標言語の具体性の高い表現にデフォルト規則を用いると多くの問題を生じる。以下の例は少なくとも1つデフォルトの用例を持っている。そのため、距離計算の際に距離が非常に小さくなり、信頼できる値を出せない。

⇒ “の → in” のような具体性の高い表現にデフォルト規則を使うのは避け、“です → be” のような一般的な表現にのみデフォルト規則を使う。

(RENTAIJO-NO-N+N) N+N :(?X の ?Y) → (in !Y von !X) ... (() (近く))

AT130082(28)
source: “はいこちらは一泊二食付きのお値段でございます” (morph: 13)
target: Diese ist um <u>in</u> Preis <u>von</u> für eine Nacht mit zwei Mahlzeiten. (0.6666842)

CT320062(8)
source: “冷蔵庫の中ですか” (morph: 5)
target: <u>In</u> drinnen <u>von</u> Kühlschrank? (0.16667919)

TAC23013(22)
source: “カードの種類はどのようなものでしょうか” (morph: 8)
target: Was für ein etwas ist <u>in</u> der Art <u>von</u> der Karte? (1.6333435)

TCS13019(27)
source: “八六一の一六四五番です” (morph: 10)
target: <u>In</u> der Nummer eins sechs vier fünf <u>von</u> acht sechs eins. (0.20001753)

(RENTAIJO-NO-N+N) N+N :(?X の ?Y) → (!Y in !X) (()) (席)	
CT220042(4)	
source:	“セントラルパーク付近のレンタルバイクのお店を教えてください” (morph: 10)
target:	Bitte sagen Sie mir eine Umgebung ein Geschäft <u>im</u> Leihmotorrad Central Park. (1.2222297)
TCS33050(23)	
source:	“はい窓際のお席は承れます” (morph: 8)
target:	Ist einen Sitzplatz <u>in</u> am Fenster akzeptierbar. (0.33334586)
(RENTAIJO-NO-N+N) N+N :(?X の ?Y) → (!X) (()) (者)	
⇒	! loss of input !
CT120051(19)	
source:	“希望のものが貸し出し中の場合予約はできますか” (morph: 13)
target:	Leiht man einen Wunsch aus und kann im Fall von drinnen reservieren? (2.063903)
TCS33009(12)	
source:	“日本語と英語の通訳ですね” (morph: 7)
target:	Japanisches und Englisch. (1.0416741)
(RENTAIJO-NO-N+N) N+N :(?X の ?Y) → (!Y mit !X) (()) (景色)	
CT130051(20)	
source:	“エーは京の四季の風物をかたどった干菓子セットでございます” (morph: 14)
target:	In bezug auf die trocknen Süßigkeiten Set, auf dem A eine Landschaft <u>mit</u> vier Jahreszeiten mit Kyoto ist bitte. (2.344471)
(IMP-KAKUJO-WO-NP) NP :(?X を ?Y) → (!X bitte) (()) (おねがいます)	
⇒	! loss of input !
TAC23013(33)	
source:	“ではご予約の内容を確認させていただきます” (morph: 11)
target:	Also, ich einem Inhalt einer Reservierung <u>bitte</u> . (1.42064)
TCS12023(18)	
source:	“デザートはフルーツタルトの方をケーキは生クリームの方をお願いします” (morph: 16)
target:	Der Nachttisch ist eine Früchtetorte <u>bitte</u> . (2.3333435)
TCS12023(49)	
source:	“はいだいたい七時ぐらいを考えております” (morph: 9)
target:	Ja. Es gegen sieben Uhr ungefähr <u>bitte</u> . (0.5833509)
(KAKARIJO-DEHA-NP) NP :(?X では ?Y) → (damit !Y) .. ((それ)())	
⇒	! loss of input !
TAS32001(9)	
source:	“予定では八月の十日から二泊お願いします” (morph: 15)
target:	<u>Damit</u> für zwei Nächte vom zehnten August an bitte. (0.5000025)

(MARKER-CN-CN-N) N :(?X <CN-CN> ?Y) → (nach !X) (()(行き))

⇒ ! loss of input !

AT120072(27)

source: “はいケーキアンドケーキ運輸 ニューヨーク支店 になります” (morph: 6)

target: Es ist nach der K And K Transportation. (0.944467)

BT320011(51)

source: “はい東京商事 ニューヨーク出張所 の鈴木和夫と申します” (morph: 9)

target: Ich heiße Kazuo Suzuki des nach Tokyo Shoji. (0.8333559)

KT330011(1)

source: “はいもしもしジャパンレンタカー 大阪営業所 でございます” (morph: 6)

target: Hallo, nach Japan Rent-A-Car. (0.5000175)

TAS13016(17)

source: “まず玄関からキッチンバス トイレ と続いております” (morph: 10)

target: Es wird fortgesetzt, wenn man von der Haustür aus die Küche, nach dem Bus zunächst. (3.493063)

TCS33009(16)

source: “三日間とも 商談 でいらっしゃいますか” (morph: 9)

target: Man nach für drei Tagen? (2.166674)

(MARKER-/KIHON-FN-S+N) S+N :(?X </KIHON-FN> ?Y) → (wo man !X) (()(ところ))

⇒ loss of input !

DT320012(9)

source: “それでその説明会を来月そちらのホテルで開きたいなと今思っているところなんです
が” (morph: 21)

target: Und es ist das, wo man eine Erklärung bitte. (3.8149085)

(MARKER-CN-CN-IS) IS :(?X <CN-CN> ?Y) → (jeweils !Y) ((それぞれ) (となる))

AT130051(4)

source: “十月二十八日ですと一泊二食付きで二万三千円となりますけれども” (morph: 18)

target: Es kostet jeweils mit zwei Mahlzeiten zu dreiundzwanzigtausend Yen.
(1.6666842)

3.9 P-DATA (その他)

以下のパターンの翻訳規則を新たに追加する.

“X から Y へ” ⇒ “von X nach Y”

OT330141(4)

source: “お客様大阪から京都へ交通機関はJRを御利用でしょうか” (morph: 13)

target: Lieber Gast, verwendet das Verkehrsmittel nach Kyoto JR von Osaka aus?
(2.5833533)

様 <PERSONALPNOM-V> Y

TAS13025(30)

source: “ジョンフィリップス様お待ちしております” (morph: 8)

target: John Herr Phillips. Wir warten. (1.4999999e-5)

“X の <name>” ⇒ “<name> von X”

TCS33009(2)
source: “フロントの鈴木でございます” (morph: 5)
target: Suzuki der Rezeption. (0.33333835)

?X <CN-SN> ?Y

⇒ (!x “sein” !Y)(SUB 1)(*HILFSVERB 2)

CT130051(15)
source: “シーは漬物セットです” (morph: 5)
target: C ist in bezug auf das gesalzene Gemüse Set. (0.9666742)
⇒ “C ist ein gesalzene Gemüse Set.”

4 生成の問題点

4.1 埋め込みによる情報の喪失

変換結果は、翻訳知識の適用時に得られた言語的情報を指定するリストからなる。このリストは文レベルから単語レベルまで再帰的な構造になっている。例えば、(SUB (NP “<word1 word2 ...>”)) では、トップレベルの表現 SUB は埋め込まれた名詞句 (NP) が主語であることを指定している。しかし、文中での語順に関して、トップレベルの要素 (例では SUB) だけが処理の対象になる。この結果、言語的情報の損失の問題が生じる。すなわち、重要な句範疇 (例えば、述部、直接目的語など) がサブレベルに埋め込まれてしまう (例えば、(SUB (VERB <verb>) ...) では VERB はトップレベルにない) と、その情報にアクセスできなくなる。この例では、述部は SUB の値として埋め込まれているので、正しく処理されず、<verb> は述部 (VERB) ではなく主語 (SUB) として認定されてしまう。これにより、語順の違い、述部の欠如、誤ったデフォルト動詞の補完などの問題が生じる。

⇒ アドホックな規則を使って、これらの句カテゴリをトップレベルにあげてしまう。

TOS33005
source: “お泊まりのお客様はフィリップス様でございますね” (morph: 10)
gen-input: (((SUB (VERB “übernachten”)) (*HILFSVERB “sein”) (NP {OHNE} “herr” (EIGENNAME+ “phillips” :GENDER NTR)) (FIX-END “, nicht wahr?”)) . 0.50002503)
target: (“übernachtet ist Herr Phillips, nicht wahr?” . 0.50002503)

4.2 等位 / 従属節における主語の一致

異なる主語を持つ等位節

以下の例の出力文において、最初の節と2番めの節の主語が異なってしまっている。理由は、それぞれの節に対して独立にデフォルトの主語を補っているからである。

AT130082(35)	
source:	“烏丸口を出て市バス二百六番にお乗りください” (morph: 11)
target:	<u>Man</u> geht den Karasuma Eingang hinaus. Und bitte nehmen <u>Sie</u> einen städtischen Bus Nummer zweihundertsechs. (1.2000117)
⇒	“Sie gehen den Karasuma Eingang hinaus. Und bitte nehmen Sie den städtischen Bus Nummer zweihundertsechs.”

TCS12023(3)	
source:	“千二百七号室に泊まっている鈴木直子です” (morph: 9)
target:	<u>NaokoSuzuki</u> und <u>es</u> übernachtet im Zimmer eintausendzweihundertsieben. (0.16668001)
⇒	“Naoko Suzuki und ich übernachtete im Zimmer eintausendzweihundertsieben.”

主語が1つしか出ていない等位節

⇒ 最初の節の主語を(主語が欠けている)2番目の節でも使う。

B4JP-101-150(9)	
source:	“ソウル駅から地下鉄に乗って東大入口駅で下りて五百メートルぐらい歩きます” (morph: 14)
target:	A und B <u>ist</u> Süßigkeiten. (0.8333409)
⇒	“A und B sind Süßigkeiten.”

4.3 冠詞の挿入

現在、名詞の定・不定の判断はかなりアドホックである。そこでは、優先順位に基づいて、名詞の前に冠詞を挿入しない / 定冠詞を挿入する / 不定冠詞を挿入するの判断を行なっている。優先順位は以下の通りである。

LocalDic → TransDic → GenMarker → part-of-sentence default

変換モジュールにおいて、ローカル辞書の項目を指定する可能性があり、それによって、メインの変換辞書の定義が上書きされる。したがって、変換辞書で明示的に指定された冠詞生成のマーカ(例: “Adresse” (address) に対する {BESTIMMT})は、ローカル辞書項目にしたがって変更され得る。変換処理の過程において、変換規則の目標言語側のボディで明示的に指定された冠詞が追加されたり、あるいは、生成モジュールへの入力に対して冠詞生成マーカ({OHNE}, {BESTIMMT},

は、生成モジュールへの入力に対して冠詞生成マーカ ({OHNE}, {BESTIMMT}, {UNBESTIMMT}) が追加されたりする。このような冠詞に関する情報が利用できない場合には、冠詞付与の対象となる名詞が属する句の種類に応じてデフォルトの冠詞を決める。主語 (SUB), 前置詞 (PP), 間接目的語 (IND-OBJ) に対しては定冠詞を挿入し、直接目的語 (OBJ) には不定冠詞を挿入する。その他の句範疇の場合は冠詞の挿入を行なわない。

より満足のいく扱いのためには、冠詞の決定に談話知識と焦点情報が必要になる。例えば、定冠詞は談話中ですでに言及された対象に対して付与されなければならない。

冠詞に関するその他の問題としては、1つの句の中に複数の冠詞生成マーカが付与されてしまった場合の扱いや並立句において各構成素に同一の冠詞を付与することなどがある。

IT320042(8)	
source:	“市電かバスで行きたいんです” (morph: 8)
target:	Ich möchte gerne mit der Straßenbahn oder Bus fahren. (1.0e-5)
⇒	“Ich möchte gerne mit der Straßenbahn oder <u>dem</u> Bus fahren.”

4.4 その他

年の表現 (“N 年 N 月”)

⇒ 序数の代わりに “N 00 + N” の形式

例: 1996 → “Neunzehnhundert neunundsechzig” (19-00 + 96)

NOT: Eintausendneunhundertsechsunneunzig (1996)

5 おわりに

学習データの量が少ないため、日独翻訳モジュールの性能は、日英間や日韓間に比べて、まだかなり劣っている。しかし、本稿で報告した評価試験から、語順の任意性や複雑な活用体系というドイツ語固有の難しさにも関わらず、日独 TDMT システム開発のよい見通しが得られた。評価試験の結果に基づく改善案を導入するとともに、学習データの量を増やすことによって、日英間や日韓間モジュールとの差はすぐに埋まるであろう。

参考文献

- [1] Ogura, K., Yokoo, A., Shirai, S. and Ikehara, S.: Japanese to English machine translation and dictionaries. Proceedings of the 44th Congress of the International Astronautical Federation, Graz, Austria, 1993.
- [2] 美馬、古瀬: 日英間変換主導翻訳の中間時評価, ATR Technical Report TR-IT-0189. エイ・ティ・アール音声翻訳通信研究所, 1996.
- [3] Seligman, M.: CNTR: Basic Functions for Centering Experiments with ASURA. ATR Technical Report TR-IT-0085. ATR Interpreting Telecommunications Laboratories, 1996.
- [4] 山本、古瀬: 日韓間変換主導翻訳の中間時評価, ATR Technical Report TR-IT-0190. エイ・ティ・アール音声翻訳通信研究所, 1996.