TR-IT-0242

科学技術文献抄録に対する変換主導型機械翻訳の適用
Applying TDMT to Document Abstructs on Science and Technology

太田　浩子　　隅田英一郎　　飯田仁
Hiroko Ohta　　Eiichiro Sumita　Hitoshi Iida

1997.10.31

概要
旅行会話などの話し言葉の翻訳で有効性が確認されている変換主導型翻訳方式（ＴＤＭＴ）を、典型的な書き言葉の一つである科学技術文献抄録に適用した。本稿では、その適用手順、結果及び課題について述べる。

エイ・ティ・アール音声翻訳通信研究所
ATR Interpreting Telecommunications Research Laboratories

目次

１．はじめに

ＡＴＲでは、話し言葉を対象として用例に基づく枠組みに従って翻訳知識を活用する変換主導型翻訳方式（ＴＤＭＴ）を提案し、多言語のプロトタイプシステムを構築している。この方式の利点は、以下の通りである。

１）文法的に説明が難しい表現を処理できる。
２）意味距離計算により高速な処理が可能である。
３）知識の記述や追加が容易である。

しかし、これらの点は話し言葉を対象に実証されてきてはいるが、書き言葉に対しても同様な利点、有効性が得られるか否かについては確認されていない。

本稿では、典型的な書き言葉の一つとされる科学技術文献抄録を対象とした英日翻訳へのＴＤＭＴの適用を行い、書き言葉に対するＴＤＭＴの有効性を検討する。以下、２章では変換主導型翻訳方式の概要、３章で科学技術文献抄録に対する適用実験について述べ、４章で課題等についての考察する。

2．変換主導型翻訳方式（ＴＤＭＴ）の概要

実際の言語現象から獲得した翻訳知識を活用して翻訳処理を行う、変換主導型機械翻訳（Transfer-Driven Machine Translation、以下、ＴＤＭＴと記す）は［古瀬 94］［Furuse 94］によって提案された。ＴＤＭＴは、形態素解析、変換、生成の３つの主要なモジュールから構成される。いわゆる構文解析処理は変換モジュールに含まれる。変換モジュールにおいて、文から語句までの様々な言語的単位の変換知識を入力文に適用し、翻訳結果を作り出す。

ＴＤＭＴの変換中心の翻訳メカニズムは、文から語句までの様々な言語的単位で、用例に基づく枠組みに従って翻訳知識を最大限に活用することができる。用例に基づく枠組みでは、翻訳知識として蓄積された用例の中から、入力表現とベストマッチする用例を意味距離計算により求め、ベストマッチした用例の対訳情報を使って翻訳結果を作る。

ＴＤＭＴ内処理（英日翻訳の場合）

・英語形態素解析処理
　ＴＤＭＴ英語形態素解析処理には、ｎグラムに基づく尤度が利用されている。
　よって、形態素解析用英語辞書、ｎグラムデータを用いて入力された英文を構成する形態素列を抽出する。

・変換処理

・・原言語（英語）構造の作成
　大きな言語的単位から小さな言語的単位のものへと変換知識を入力文に適用し、変換知識の原言語側を組み合わせた原言語構造を作る。

・・目的言語（日本語）構造への変換
　原言語構造の部分構造ごとに意味距計算によって最尤の部分構造へ変換し、目的言語構造を作る。

・・最尤構造文の決定
　目的言語構造の中から、意味距離計算の総和に基づき最尤の構造を決定する。

・日本語生成処理
　変換処理で作成された形態素列を目的言語文法にそって活用を整える等の処理を行う。

3．科学技術文献抄録に対する翻訳知識の追加

実験データとして、科学技術文献の標題及び抄録の英文を計５００文用意した。これは、「computer」をキーワードとして持つ文献の標題、抄録を抽出している。
５００文中、４５０文を訓練文としてＴＤＭＴへの翻訳知識の追加を行った。
本章では、今回の実験データの特徴、翻訳知識の適用手順、適用結果について述べる。

3.1　科学技術文献抄録の特徴

本実験で用意した科学技術文献抄録の特徴は、以下の通りである。

［1］長文が多い。（平均２０～３０ word。旅行会話文は平均１０ word）
(例) "Basic construction of land record data bases using the personal computer" is one subject in the total subjects, and the purpose are the improvements of records as the bases of land record data bases using the personal computers, the attempt of land record data base popularization in the regions, and assistance of the database enrichment in the region.
「パソコンを用いた地図データベースの基礎構築」はその中の一課題で，パソコンを用いて地図データベースの基礎となる地図データを整備し，地域における地図データベース普及を計り，地域におけるデータベース拡充の助けにすることを目的としたものである。

［2］数値、記号類が多用される。
(例) The WIDOM model gave *-1062.0.MU.kj/mol* and Shing and Gubbin model gave *-880.0.MU.kj/mol* as compared to the measured value of *-893.1.MU.kj/mol*.
実測値がー８９３．１μｋｊ／ｍｏｌであるのに対してＷＩＤＯＭモデルではー１０６２．０μｋｊ／ｍｏｌ，またＳｈｉｎｇ　ａｎｄ　Ｇｕｂｂｉｎモデルではー８８０．０μｋｊ／ｍｏｌであった。

［3］複合語が多用される。
(例) Research report on international cooperation relating to the image information technology.
映像情報技術に係る国際協力に関する調査報告書。

ＪＳＴ[1]の科学技術文献用の辞書（本実験では活用できなかった）では、Research report、 international cooperation relating to the image、information technology は複合語として登録されている。一般に複合語の認定は難しく、予め辞書に登録しておくことが望ましい。

---

[1] ＪＳＴ＝Japan Science and Technology Corporation

［4］並列構造が多用される。

（例）This report summerizes a *systematic description method of operations aiming at standardization of word-processor operations* and a *method to derive draft standards of word-processor operations from the described results of operations.*
ワードプロセッサ操作の標準化を目的とした操作の系統的な記述法と，操作を記述した結果からワードプロセッサ操作の標準案を導出する方法についてまとめた。

Patella behavior in the largest extension of the knee joint was photographed by X-ray, and three-dimensional shapes of the pattela and thighbone was obtained from computed tompgraph images of the knee joint.
膝関節最大伸展時における膝蓋骨の挙動をＸ線で撮影し，また膝関節のＣＴ画像より膝蓋骨および大腿骨の三次元形状を求めた。

［5］関係節が多用される。

（例）The main memory of card 386 employed 2 substrate sheet structure *in which* the memory module mounted on the surface of 2 layer substrate was connected with the main subsate using flexible tape.
カード３８６ではメインメモリは２層基板に表面実装したメモリモジュール基板をフレキシブルテープでメイン基板に接続した２枚基板構造とした。

［6］文頭に論文特有の表現が多用される。

（例）This paper explains～
This paper describes～
This paper presents～
This paper presents～

3.2 適用手順

［1］科学技術文献抄録英日対訳コーパス５００文を用意。

(例) 1件につき1標題と平均3文が含まれる。

Commodity Trend and Technological Trend of "X" Window Terminal.
Xウィンドウターミナルの商品動向と技術動向。

The products with higher performance of over 200,000 Xstone value are shipped and their performance will be improved by advanced CPU and graphics accelerators in future.
性能面ではＸｓｔｏｎｅ値２０万をこえるものが出荷されており，今後はＣＰＵの高性能化やグラフィックアクセラレータによりさらに性能アップがはかられる。

X11R6 is the main function and some multi-media functions are incorporated.
機能面ではＸ１１Ｒ６が主流となっており，マルチメディアの機能が若干盛り込まれた。

The multi-media functions will be enhanced in future.
今後はマルチメディアの機能が強化される。

［2］５００文を対象としてタギングデータ(TDMT コーパス) 作成
形態素解析の誤りを除き、正確に変換処理部分の評価を行うために、形態素解析処理後のデータを人手で作成し、変換入力データは誤りがない状態とする。

発話 ID | 文 ID | 形態素 ID | 出現形 | 正規形 | 意味活用 | 一致活用 | 品詞 | コメント

(例)
```
;;; Research report on international cooperation relating
to the image information technology.
|10|10|research|research|||CN||
|10|20|report|report|||CN||
|10|30|on|on|||PREP||
|10|40|international|international|||ADJ||
|10|50|cooperation|cooperation|||CN|?UNDEF?|
|10|60|relating|relate|ING||V|?UNDEF?|
|10|70|to|to|||PREP||
|10|80|the|the|||DET||
|10|90|image|image|||CN||
|10|100|information technology|information technology|||CN||
|10|105|.|.|||SYMBOL||
```

[3] ５００文を対象として変換辞書作成　２４１８語

(例) ((CN "airtightness") (普通名詞 "気密化"))
((CN "algographics, co., ltd.") (普通名詞 "(株) アルゴグラフィックス"))
((CN "base") (普通名詞 "基礎"))
((CN "communication market") (普通名詞 "通信市場"))
((CN "computational fluid dynamics") (普通名詞 "計算流体力学"))
((V "secure") (形容名詞 "安全")(格助詞 "に")(本動詞 "する"))
((V "run") (本動詞 "実行する"))
((V "satisfy") (本動詞 "満足する") (助動詞 "せる"))

[4] ５００文を対象として合成語作成 (a-data)　５５８語

(例) (define-lexical-transformation oa_equipment :e-j 2
(
(((:word . "oa")) ((:word . "equipment")))
=>
(lex ((1 2) (:all-copy . 2) (:word . "oa equipment")
(:reg-exp . "oa equipment")))
)
)

[5] ５００文を対象として英語意味辞書作成　２４１８語

(例) ("airtightness" CN "236")　##236：開閉
("algographics, co., ltd." CN "713")　##713：団体
("intermediate distributing frame" CN "990")　##990：機械
("office computer" CN "992")　##992：電気機具
("working machine" CN "990")

[6] 変換パターン作成

用例追加 (斜体用例は科学技術文献抄録用に追加した用例)

(例) (define-pattern as-n+n :e-j n+n
(?x "as" ?y) (:head 1)
=>
(((!y "として" "の" !x) (格助詞 2) (連体助詞 3))
(
(("introduction") ("ews of ibm"))
(("letter") ("proof"))
(("personal computer") ("instrument"))
(("plan") ("communication system"))
(("role") ("system integrator"))
(("use") ("teaching material"))
))
:
)

新規パターン追加

```
(例) (define-pattern in_accordance_with-n+n :e-j n+n
       (?x "in accordance with" ?y) (:head 1) (:x group-n) (:y group-n)
        =>
       (((!y "に対応する" !x) (連体助詞 2))
         (
          (("sensor") ("purpose"))
          (("system") ("environment"))
          )
       ))
```

boundary-marker パターン作成

```
(例) (define-local-transformation cn-bev :e-j 5
       (
        (((:pos . cn)) ((:pos . bev)))
        =>
        (1 <cn-bev> 2)
        )
       (
        (((:pos . cn)) ((:reg-exp . ")")) ((:pos . past)) ((:pos . bev)))
        =>
        (1 2 <cn-bev> 3 4)
       )
      (
       (((:pos . cn)) ((:pos . SYMBOL)) ((:pos . bev)))
        =>
        (1 2 <cn-bev> 3)
       )
      )
```

［7］５０文オープンテスト

旅行会話２８００文と科学技術文献抄録４５０文から得られた翻訳知識（パターン、用例、辞書等）を用いて、科学技術文献抄録残り５０文のオープンテストを行った。

3.3 パターン追加

パターン作成の段階評価を行うため、４００文までは用例のみを追加した時点、新規にパターンを追加した時点の各段階でパターンを固定した。４００～４５０文までの５０文は用例もパターンも同時に作成した。

Baseである旅行会話用のパターン数、本実験作業の１００文毎の追加用例数、新規パターン数、総合計を以下に示す。

	用例数	パターン数
ＴＤＭＴ旅行会話用翻訳知識（Base）	10300	1375
１～１００文まで用例のみ追加	970	0
１０１～２００文まで用例のみ追加	905	0
２０１～３００文まで用例のみ追加	560	0
３０１～４００文まで用例のみ追加	706	0
１～１００文まで用例＋新規パターン追加	132	47
１０１～２００文まで用例＋新規パターン追加	124	37
２０１～３００文まで用例＋新規パターン追加	204	35
３０１～４００文まで用例＋新規パターン追加	105	16
４００～４５０文まで用例＋新規パターン追加	326	16
科学技術文献抄録４５０文の追加翻訳知識合計	4032	151
総合計	14332	1526

以上より

（1）１文につき、追加すべき用例数は平均９．６用例であり、旅行会話文の平均３．７用例に比べて多い。

（2）４００文までの用例のみ追加が終了した時点で、用例のみで完成した文が２０５文／４００文であった。しかし、異なる新規パターンが約２００パターン必要なのではなく、同じパターンを必要とする文章が多く見られたため追加パターン数が減ってきているのがわかる。最終的に１５１パターンで済んでいる。パターン追加が飽和状態となるまで拡充が必要である。

【 新規追加パターンの特徴 】

[1] **副詞句 ＋”，”＋文**

文例：As the basic preparation for computerizing promotions, the environmental improvement is necessary for the utilization capable of personal computers (Pasocon) with one set by one staff.

パターン：("by" "ing+" ?x "," ?y) ("as" ?x "," ?y)

上記形のパターンは、旅行会話パターンの中にも存在したが、","ではなく、解析マーカ「〈〉」が入ったパターン (("by" "ing+" ?x 〈〉 ?y)、("as" ?x 〈〉?y)) であり、これは話し言葉パターンの特徴と言える。

[2] **Symbol (「"」 「 () 」) を含む形**

文例：The report is a summary of research and investigation for 5 years by the "operation-description research subcommittee. "

パターン：("¥"" ?x "¥"") ("(" ?x ")")

書き言葉特有の「"」「 () 」等が、主部、述部問わずあらゆる場所に現れるためその出現位置毎にパターン作成が必要であった。

また、マーカの挿入が必要な場合もあるが、マーカを挿入すると、そのための副作用も多く、マーカ挿入には充分な検討が不可欠である。

[3] **関係節**

文例：A system in which the calculation and plotter output of the performance test are carried out with the personal computer has also been developed.

パターン：(?x "in" "which" <> ?y) (?x "," "which" <> ?y)

[4] **名詞句 ＋”，” ＋名詞句 ＋ ”，” ・・**

文例：This paper describes a derivation method of an NS equation, a discretization method, the stability of the numerical calculation, the accuracy, etc.

パターン：(?a "," ?b "," ?c "," ?d "," ?e "," "etc.") (?a "," ?b "," ?c "," "and" ?d)

書き言葉特有の、","で区切られた句並列、多変数構造のパターンが多く出現し、その都度パターン作成が必要になる。

3.4　翻訳結果（オープンテスト）

書き言葉用の知識の追加は小規模であったので翻訳結果を評価するには尚早ではあるが、以下に、表層構造が正しい文、Separate が入る文、失敗の原因について例示、考察を行う。

● 表層構造が正しい文：１８／５０文

下記出力結果は、
① 英文
② 翻訳結果
③ ＪＳＴ抄録文
の順である。

[1] ("A master-slave manipulator, based on a decentralized control, was developed by use of a cost-effective microcomputer-based memory-to-memory communication method.")
("分散する制御に基づくマスター・スレーブ・マニピュレータは、費用対効果の高いマイコンベースメモリー間通信方法の利用によって開発されました。")
マイコンベースの安価なメモリ間通信法を用いて，分散制御によるマスター・スレーブ・マニピュレータを開発した。

[2] ("Introduction of a porogy net room in Tomonoura.")
("ともの浦の鯛網部屋の導入です。")
ともの浦鯛網部屋の紹介。

[3] ("This paper describes the effort of Fukuyama Human Resources Development Junior college's Production Technology Department to the robot sumo wrestling.")
("福山職業能力開発短大のロボット相撲の生産技術の努力について述べます。")
福山職業能力開発短大・生産技術科のロボット相撲への取り組みについて述べた。

[4] ("This paper describes development of a robot in the title by Industrial Equipment Department of Oyama Human Resources Development Junior College.")
("小山職業能力開発短大の産業機器の部門によるタイトルにおけるロボットの発展を述べます。")
小山職業能力開発短大・産業機械科の標記ロボット開発について述べた。

[5] ("Considering the number of competitions held, the robot sumo wrestling is useful for the regional development.")
("持たれる競合数を考えて、ロボット相撲は地域の開発の有用です。")
開催大会数の多さからみてロボット相撲は地域振興に役立つ。

[6] ("Performance of the calculation by MD using large computer.")
("大きいコンピュータを使用したMDによる計算の性能です。")
大型コンピュータを用いたMDによる計算の実演。

[7] ("A specific example is shown on how the simulation by MD method is actually carried out.")
("明確な例はどのようにMD方法によるシミュレーションが実際には実施するか上で示されています。")
MD法によるシミュレーションが実際にどのように進められるかを具体例を通して解説した。

[8] ("Computer environment is assumed in a way that a personal computer is connected with a supercomputer through LAN.")
("コンピュータ環境はパソコンがＬＡＮを通してスーパーコンピュータと接続されるという方法の中で仮定されています。")
コンピュータ環境はパソコンからＬＡＮを介しスーパーコンピュータに接続することを想定した。

[9] ("Utilization of on-the-market or published programs are also referred to.")
("市販の利用こと、または公開されたプログラムことは同様に言及されています")
市販・公開プログラムの利用にも触れた。

[10] ("He fills up the newspaper contents cooperating with photographer companions in the same job.")
("同じ業務で写真家仲間と協力して新聞内容について満たします。")
同類の写真家仲間との情報交換を広げながら新聞の内容を盛り上げて行く。

[11] ("The specification, performance and system of the camera are introduced.")
("カメラの仕様、性能、システムは紹介されます。")
本機の仕様，性能，システムなどを紹介した。

[12] ("This paper describes outline and future image of the software of the title under development as a dynamic characteristic simulation program of hydraulic control system by personal computer.")
("パソコンによる油圧制御システムの動的な特徴シミュレーションプログラムとしての開発中のタイトルのソフトウェアの概要と今後の画像を述べます。")
パソコンによる油圧制御系の動特性シミュレーションプログラムとして開発中の標記ソフトウエアの概要と将来像を述べた。

[13] ("The table shows hardware and OS needed to run VisSim.")
("ＶｉｓＳｉｍを実行しなければならないれたハードウエアとＯＳを示します。")
ＶｉｓＳｉｍを実行するのに必要なハードウエアとＯＳを表に示した。

[14] ("This paper presents simulation of bond graph as an example of the expansibility.")
("拡張性の例としてのボンドグラフのシミュレーションを示します。")
拡張性を表す例としてボンドグラフのシミュレーションを示した。

[15] ("EX.TDs for Windows and Macintosh are being sold.")
("ウィンドウズとＭａｃｉｎｔｏｓｈ用のＥＸ．ＴＤは販売されています。")
ＥＸ．ＴＤは現在ウィンドウズ用とマッキントッシュ用が販売されている。

[16] ("The system LSI technical committee conducted research in four fields to promote research and development of microcomputer technology and development of associated industries.")
("システムＬＳＩ技術委員会は研究とマイクロコンピュータの発展技術と関連させる産業の発展を促進する四分野の研究を実践しました。")
システムＬＳＩ技術委員会は，マイクロコンピュータ技術の研究開発促進及び関連産業の発展を促進するため，４分野に分けて調査を行った。

[17] ("With recent advancement in international standardization of environmental control matters, Japan is compelled to positively tackle environmental problems.")
("環境制御事の国際的な標準化の最近の進行では、日本は積極的環境問題を取り組むことが強います。")
環境管理についての国際標準規格化が進む中で，日本も積極的に対処する必要に迫られている。

[18] ("To meet the purpose, it is essential to build new life cycles and social systems.")
("目的を満たすため、新しいライフサイクルと社会システムを造ることが重要です。")
そのためには新しいライフサイクルや社会システムの構築が不可欠となっている。

● Separate が入る文：１１／５０文
(Separate は入るが解釈可能文　６／１１文)

(例) ("The titled digital camera is a third product as Kodak compact camera with lower price below fifty thousand yen.")
("標記のディジタルカメラはより低い価格のＫＯＤＡＫの小型カメラとしての三番目の製品です‥‥五万円です。")
標記ディジタルカメラはコダックコンパクト機としての第３弾で，５万円以下の低価格を設定した。

("The environmental problem is now the most impending task in the industrial world.")
("環境問題は今‥‥‥最も‥産業世界における切迫した課題です。")
環境問題は産業界において最も切迫した課題となってきた。

● Separate が入った文の原因

［1］ パターンがない　→　新規パターン追加

［2］ 既存パターンに２ head 用例がない
(例) ("Communication control was done between five microcomputers.")
("通信制御はされました‥‥五台のマイクロコンピュータです。")
５台のマイコン間で通信制御した。

上記例は、下記のパターン、用例が存在したにもかかわらず、パターンが適用されず Separate が入ってしまった。これは、変数xに１head用例しか存在しなかったためと思われる。極端に言えば、("difficult") を ("be" "difficult") と書き換えると上記文章は完成するため、2headに関する処理が必要である。(安藤のＴＲ参照)

```
(define-pattern vp=between=np-pn :e-j pn
 (?x "between" ?y) (:head 1) (:x ps pn p v bev) (:y group-n)
 =>
 (((!y "の" "間" "で" !x) (連体助詞 2) (普通名詞 3) (格助詞 4))
    ((("difficult") ("machine"))
 ))
)
```

● 構造が正しくなかった文の原因

［1］ 多並列、多変数構造の解釈

(例1)("The simulation is effective in reduction of experiment frequency and short-term correction of any defect.")
("シミュレーションは欠点の実験の頻度と短期改善の削減上で効果的です。"
"シミュレーションは欠点の実験の頻度と短期の改善の削減上で効果的です。"
"シミュレーションは実験の頻度と欠点の短期改善の削減上で効果的です。"
"シミュレーションは実験の頻度と欠点の短期の改善の削減上で効果的です。")
シミュレーションは実験回数の低減と不具合の短期解決に有効である。

基本並列「A of B and C of D」→ ((A of B) and (C of D))
と訳したいが、曖昧性解消は難しい。

(例2) ("In *the future*, *multi-direction detection*, *strengthening of defense*, *equipment of attack weapon*,／diversification of the program will be the design principle for improvement.")
("今後、防御の強化する多方向検知、攻撃武器の機器では、プログラムの多様化は改良の設計方針です。")
今後は，多方向検出化，守備の強化，攻撃武器の装備，プログラムの多様化を設計方針として改良する。

上記「in the future」で区切られるべき副詞句が、オープンテストでは「in ～, 」と４つ目のカンマまでを副詞句と判断されている。

［2］ベストターゲットパターンにマッチしない
→ **用例の不足、意味コード体系見直しが必要**

本実験では、５００文中に現れる単語、合成語について、角川新類語辞典を体系とした意味辞書を作成し、意味距離計算を行っている。しかし、科学技術分野のテキストについてこの体系を適用させると、ほとんどの単語に「機械」コードが付与される。
よって、訳し分けのキーポイントとなる意味概念が狭く、正しい訳し分けができない。テキスト分野を設定する場合は、その分野中の、より詳細なコード体系が必要と思われる。

(例) ("Considering the number of competitions held, the robot sumo wrestling is useful for the regional development.")
("持たれる競合数を考えて、ロボット相撲は地域の開発の有用です。")
開催大会数の多さからみてロボット相撲は地域振興に役立つ。

オープンテストでは「 is useful for the regional development」の部分に
(?x "for" ?y)ソースパターンが適用された。ターゲットパターンに「!y "に対し" !x」、「!y "で" !x 」等が存在するにもかかわらず、「!y "の" !x」が適用されている。
この例の場合は、「!y "に対し" !x」に旅行会話用の用例しか登録されていなかったため、用例不足の問題と思われる。（下記パターン参照）

```
(define-pattern for-n+n :e-j n+n
 (?x "for" ?y) (:head 1)
   =>
   (((!y "の" !x) (連体助詞 2))
                :
   (((!y "に対し" !x) (格助詞 2))
                :
   (((!y "で" !x) (格助詞 2))
                :
```

4．考察

本実験において、変換主導型翻訳方式ＴＤＭＴは、話し言葉同様、科学技術文献抄録に対してパターン作成、用例追加によって適用可能であることが確認された。実験で使用した科学技術文献抄録は、ある程度、旅行会話用パターンに包含されていた。書き言葉特有のシンボル等を含んだパターン、関係代名詞節等の修飾節に関わるパターンを新たに追加することが多かったが、逆に旅行会話で多く使われる疑問文、口語文に関するパターンは、科学技術文献抄録では出現頻度が少なかった。

英日翻訳では、文法的に説明の難しい表現はあまり見受けられなかったが、そのような表現に対しても、ＴＤＭＴのパターンの記述が容易である。従って、多種多様な表現の文章に柔軟に対応できる点はＴＤＭＴの大きな利点と思われる。

しかし、非訓練文の訳質は充分ではなかった。これは、書き言葉、科学技術分野における訓練が足りないことが大きな要因と考えられるため、更なる訓練が必要とされる。
また、現在のＴＤＭＴのいくつか課題点、特定分野に限ったゆえの問題点も現出しており、これを列挙する。

［1］変換辞書デフォルト訳の選定と訳し分け

変換辞書は英語、日本語１対１構造であるため、多義語であってもデフォルト訳を付与しなければならない。訳し分けの問題は、明確なローカル辞書作成規準、辞書構成変更、分野別辞書作成等で解決するような何らかの対策が求められる。

参考までに、ＪＳＴの既存日英辞書を英日に逆転させた場合、一英単語につき多数の訳語が対応する。科学技術全般という広い分野で活用するためにはこれらの訳し分けを行う必要がでてくるであろう。

（例）

operation|演技:手術操作:手術作動:稼動:か動:か働:業務活動:作働:動作:術:操業:運転操作:運航面:手術:操作途中:運用:切開手:オペレーション:運行:操業法:運転:操作:施術:操縦法:運航:演算:施行

scale|鱗:尺度:ドレミファ:目盛り:目盛:スケール:板状鱗屑:鱗粉:規模:鱗屑:りんせつ:りん片:鱗片:鱗片:体重計:測定尺度:ヘルスメーター:ウロコ:うろこ:鱗:脚鱗:階段標準:湯あか:湯アカ:計量装置

［2］意味辞書体系

特定分野の入力文を扱う場合は、訳し分けのキーポイントとなる用例のcoverageと意味距離計算が重要であり、その意味距離計算の基礎となる意味コード体系に広範囲、深階層が求められる。本実験で使用した角川新類語辞典による体系では意味コードの偏りが見られたため、下記のような別体系を検討する必要があると思われる。

・ＪＳＴシソーラス
・ＪＳＴ日英機械翻訳意味コード
・ＮＴＴ日本語語彙体系

［3］並列構造、修飾係り受けの解釈

日本語の並列句、曖昧性解消、長文解析についてはいくつかの有効な手法［黒橋 92］が存在するが、英語の曖昧性解消問題は、以前、困難な課題と考えられている。

［4］複数同点文の選択
学習させた文においても複数の同点文が現れる。現在のＴＤＭＴは、複数同点文の中から一文を選択する処理を行っていない。
同点文を減らす処理工夫もしくは、出力文からよりよい結果を選択する処理が必要と思われる。

［5］主格「が／は」の訳し分け
節を単位に分割し、節の種類により「が／は」の訳し分けを行う処理があるが、現在、機能していない。

［6］数値、記号の扱い
科学技術文献抄録に多く見られる記号、数値、単位類は、現段階では単語として認識されタギングされている。しかし、数値、記号は組み合わせが複雑でありその都度のパターン作成では、処理時間もかかり、曖昧性が増加する。
よって、数値、記号類は入力の時点で表層形のまま変換処理に渡される方が望ましい。

(例) ("The WIDOM model gave **minus one thousand sixty two point zero .MU.kj/mol** and Shing and Gubbin model gave minus eight hundred eighty point zero . MU.kj/mol as compared to the measured value of **minus eight hundred ninety three point one .MU.kj/mol.**")
("実測値**マイナス八百九十三点一μｋｊ／ｍｏｌ**に比較して、ＷＩＤＯＭモデルは**マイナス千六十二点零μｋｊ／ｍｏｌ**を示しました、Ｓｈｉｎｇ　ａｎｄ　Ｇｕｂｂｉｎモデルは**マイナス八百八十点零μｋｊ／ｍｏｌ**を示しました。")

("The company developed an image compression elongation one chip LSI suitable for bidirection video communication by original algorithm and high-speed architecture, and in addition, it developed three kinds of basic systems (add-in board for a personal computer, image storage transmission system, radio moving image transmission system) equipped with this LSI, and started supply **in nineteen ninety six.**")
("独自のアルゴリズムと高速アーキテクチャによる双方向画像通信に適している画像圧縮伸長１チップＬＳＩを開発しました、更に、このＬＳＩを装備した三種類の基本システム(パソコン用の拡張ボード、画像蓄積伝送装置、無線動画伝送システム)を開発しましたて、**十九、九十六年**に供給を開始しました。")

5．終わりに

話し言葉と書き言葉では、使用頻度が多少異なるが、基本的パターンは両者に共通であり、本実験においても旅行会話用のパターンを生かし、書き言葉用のパターンを追加することによって変換主導型翻訳方式の科学技術文献抄録への適用も可能かつ有効であることが確認された。しかし、訳質を評価するにはまだ書き言葉（科学技術文献抄録）固有のパターン、用例が不足しているので、旅行会話に準じて、追加拡充する必要がある。また、翻訳質向上においてＴＤＭＴ自体の課題はあるものの、従来方式に比べて、非常に柔軟に言語表現に対応できるので質向上に期待が持てる。

参考文献

[古瀬 94] 古瀬蔵、隅田英一郎、飯田仁. 経験的知識を活用する変換主導型翻訳. 情報処理学術論文、Vol.35, No3, PP414-425, 1994.

[Furuse 94] Furuse,O. and Iida,H. Constituent Boundary Parsing for Example-Based Machine Translation, Proc. of Coling '94, pp.105-111, 1994.

[黒橋 92] 黒橋禎夫、長尾真. 長い日本語文における並列構造の推定. 情報処理学会論文誌、Vol.33, No8, PP.1022-1031, 92.

[安藤 96] 安藤真一、隅田英一郎. ＪＥＩＤＡ機械翻訳システム評価基準を用いたＴＤＭＴの評価. TR-IT-0188