TR-IT-0239

質疑応答文に対する協調融合翻訳の適用検討

Preliminary Experiments using Coorperative Integration Translation method on a Q&A task

東海林 里仁　美馬 秀樹　飯田 仁
Rihito Shoji　Hideki Mima　Hitoshi Iida

1997.9.30

概要

協調融合翻訳方式の実験システムである TDMT を使って、質疑応答文を対象とした日英翻訳のための予備的実験を行なった。本稿では、その内容と適用検討の結果について述べる。

ATR 音声翻訳通信研究所
ATR Interpreting Telecommunications Research Laboratories

1. 目的

1.1. 研究目的

質疑応答文に対する協調融合翻訳の適用検討

ATR 音声翻訳通信研究所の協調融合翻訳のプロトタイプシステム（Transfer-Driven Machine Translation, 以下、ＴＤＭＴと呼ぶ）[Furuse94],[古瀬 94],[美馬 96],[美馬 97]を用い、質議応答文のための翻訳知識を構築し、予備的実験により翻訳結果を評価する事を目的とする。

1.2. 研究計画

協調融合翻訳方式の理解

協調融合翻訳方式とはいかなる翻訳方式であるかを理解する。

翻訳知識の記述方法の理解

協調融合翻訳方式のプロトタイプシステムであるＴＤＭＴにおいて、翻訳知識の記述方法を理解する。

翻訳知識の構築

サンプル文を翻訳できるようにすることを目的として、翻訳知識を構築する。

翻訳予備実験

サンプル文に対して、ＴＤＭＴで翻訳を実行する。

2. 協調融合翻訳（ＴＤＭＴ）

ＴＤＭＴは日英翻訳の場合、
- 形態素処理
- 原言語構造の解析
- 文の最尤構造の決定
- 対訳パターンの決定
- 単語レベルでの翻訳
- 生成処理

の順で処理される。

2.1. 形態素処理

単語レベルに文を分解し、各単語に以下の情報を付加する。

発話 ID | 発声 ID | 文節 ID | 単語 ID | 表記形 | 読み | 正規形 | 品詞 | 活用形 | 活用型又は音便 | |

例
「もっとゆっくり言ってくれませんか」
に対して形態素解析を行うと以下のようになる。
0|0010|10|10|もっと|モット|もっと|副詞|||||
0|0010|20|20|ゆっくり|ユックリ|ゆっくり|副詞|||||
0|0010|30|30|言っ|イッ|言う|本動詞|五段ワ|た||
0|0010|40|40|てくれ|テクレ|てくれる|助動詞|一段|連用||
0|0010|50|50|ません|マセン|ません|助動詞|無変化|基本||
0|0010|60|60|か|カ|か|終助詞|||||

また、例のように形態素解析された（TAG 付された)テキストファイルを予め作成することで、形態素解析が終了した状態からの TDMT での翻訳処理も可能である。

2.2. 原言語構造の解析

2.2.1. 単語の合成

形態素解析では別々の単語に分割されたが、一つの単語として処理した方が都合が良い場合、又は一つの単語として認識しないと翻訳できない場合は単語の合成を行う。

例
「グローバル化」
形態素解析を行うと「グローバル」「化」の二つの単語に分解される。
「グローバル化」を globalization と翻訳するために、a-data というディレクトリの下の cn.lisp ファイル(名刺同士の合成なので cn.lisp ファイルを用いる)に以下の書式で追記し、翻訳時に合成が行われるようにする。追記後、保存し、C-c C-c を行えば TDMT に反映される。
また、新しい語合成した場合は後に説明する意味辞書、翻訳辞書にも登録がひつようである。

```
(define-lexical-transformation gousei-gurohbaruka :j-e 2
 (
  (((:word . "グローバル")) ((:word . "化")))
  =>
  (lex ((1 2) (:pos . サ変名詞) (:reg-exp . "グローバル化")))
  )
;;対訳　globalization
)
```

2.2.2. マーカーの付与

ＴＤＭＴでは２つの自立語の関係をパターン化する際、自立語の間に存在する機能語を用いてパターン化する。

例
「私は行く」→ “X は Y”

ところが、自立語間に機能語が存在しない場合はマーカーを挿入する。

例
「もっともっと」→ “X <adv> Y”

2.3. 文の最尤構造の決定

各語は意味辞書に意味的な分類を示すコードが登録されており、その値を基に意味的に最も近い対訳用例を決定する。

例
「これは何ですか」
jma-sem-code.txet ファイルに
("これ" 代名詞 "158" "101")
("何" 普通名詞 "101")
という行が存在する。
この値を基に p-data ディレクトリの
1．sen-de.lisp ファイル中の
(?X ですか)パターン中の翻訳構造を(it is !X)とする X が「何」、
2．kakarijo-ha.lisp ファイル中の
(?X は?Y)パターン中の翻訳構造を(!X be !Y)とする X が「ここ」Y が「どこ」、
という用例に最も近いと決定する。

2.4. 単語レベルで翻訳

単語変換辞書に登録されている訳語に従って翻訳する。

例
「これは何ですか」
/dic/japanese-to-english.lisp ファイル中の
((代名詞 "これ") (PRON "this"))
((普通名詞 "何") (WH "what"))
が含まれる行を参照し、変換する。

2.5. 生成処理

2．3で決定された翻訳構造で活用が必要な場合、generation ディレクトリにある活用辞書を参照し正しく活用させる。

3. ＴＤＭＴの質疑応答文への適用

質問文２０、肯定文５、「の」の表現を含む句３７の計６２文のサンプルについて翻訳を実行した。

入力文と模範正解出力

1	もっとゆっくり言ってくれませんか	Could you speak a little slower?
2	これは何の意味ですか	What does this mean?
3	これは何ですか	What is this?
4	この単語はどこにありますか	Where is this word?
5	著作はだれですか	Who is the author?
6	この雑誌はどこにありますか	Where is this magazine?
7	これはいつ発明されたのですか	When was this invented?
8	なぜ情報は増大するのですか	Why is information increasing?
9	いつこの計算は終了するのですか	When will this calculation terminate?
10	この式はどういうふうに導かれるのですか	How will this formula be introduced?
11	計算量はいくらになるか	How much calculation is required?
12	計算時間はどのくらいになるか	How long will the calculation take?
13	誤りは何個あったか	How many errors were there?
14	どちらが正しいか	Which one is correct?
15	どうすればとけますか	How is the problem solved?
16	この値は間違いですか	Is this value incorrect?
17	私の結論は奇妙ですか	Does my conclusion sound strange?
18	数字の順序は逆ではないか	Isn't the order of numbers reversed?
19	質問してもいいですか	May I ask you a question?
20	参考書を紹介してくれませんか	Can you recommend a good reference book?
21	私の名前は山本です	My name is Yamamoto
22	おっしゃることがわかりません	I don't understand what you mean
23	わかりました	I understand
24	プリントをください	Printed material please
25	ヒントを教えてください	Please give me a hint
26	コンピュータの発展	Development of computers

27	触覚などの感覚	One of the human senses such as the sense of touch
28	複数の情報	A variety of information
29	デジタルの技術	Digital technology
30	別の媒体	Other media
31	移動の向上	Increase of movement
32	情報処理の技術	Fundamentals of information processing
33	八つの章	Eight chapters
34	誤りの検出	Error detection
35	現在の通信	Present communications
36	社会のグローバル化	Globalization of societies
37	汎用のコンピュータ	General purpose computers
38	多数のコンピュータ	Many computers
39	情報の交換	Exchange of information
40	外部のコンピュータ	Outside network computers
41	異なる業種間のビジネス	Interdisciplinary business
42	世界の最新情報	Latest world information
43	ネットワークの間を	Between networks
44	自分専用のコンピュータ	Private computers
45	同軸ケーブルのイーサネット	Ethernet composed of coaxial cables
46	百メガビット／秒のイーサネット	High-speed 100Mbit/s communications
47	広域のネットワーク	Wide area networks
48	映像のメディア	Media involving images
49	性能の向上	Performance improvement
50	教育の分野	Area of education
51	携帯型の端末	Portable terminal
52	ユーザ好みの形	User preference
53	サービスの機能	Service function
54	ネットワークアクセスのモデル	Network access model
55	帯域の制限	Bandwidth limitation
56	個別の通信網	Individual communications network
57	均一の形	Uniform
58	端末の接続	Terminal connection
59	音声の伝送	Speech transmission
60	呼制御用のDチャンネル	D channel for call control
61	一定の間隔	Constant interval
62	データの一部	Part of the data

3.1. ＴＤＭＴに手を加えない状態での日英翻訳の実行

ＴＤＭＴにサンプル文を入力したところ、全６２文中３５文で形態素解析が行えず、NIL 表示となった。これは、形態素辞書に含まれない単語がサンプルに多数存在したため、解析が行えなかったことが最も大きな原因と考えられる。

初期状態での出力結果

1	もっとゆっくり言ってくれませんか	Slowlier _ _ will you tell?
2	これは何の意味ですか	What is this meaning?
3	これは何ですか	What is this?
4	この単語はどこにありますか	NIL
5	著作はだれですか	NIL
6	この雑誌はどこにありますか	Where is there this magazine?
7	これはいつ発明されたのですか	NIL
8	なぜ情報は増大するのですか	NIL
9	いつこの計算は終了するのですか	When and this calculation end?
10	この式はどういうふうに導かれるのですか	NIL
11	計算量はいくらになるか	How much the quantity is in the calculation _
12	計算時間はどのくらいになるか	How the time is in the calculation _
13	誤りは何個あったか	NIL
14	どちらが正しいか	Which is righter?
15	どうすればとけますか	NIL
16	この値は間違いですか	NIL
17	私の結論は奇妙ですか	NIL
18	数字の順序は逆ではないか	NIL
19	質問してもいいですか	Could I ask?
20	参考書を紹介してくれませんか	NIL
21	私の名前は山本です	My name is Yamamoto
22	おっしゃることがわかりません	I don't know to say
23	わかりました	I see
24	プリントをください	Gives me the print
25	ヒントを教えてください	NIL
26	コンピュータの発展	NIL
27	触覚などの感覚	NIL
28	複数の情報	A plural information
29	デジタルの技術	NIL
30	別の媒体	NIL
31	移動の向上	NIL

32	情報処理の技術	NIL
33	八つの章	NIL
34	誤りの検出	NIL
35	現在の通信	The communication in now
36	社会のグローバル化	NIL
37	汎用のコンピュータ	NIL
38	多数のコンピュータ	Many _ the number computer
39	情報の交換	Information exchange
40	外部のコンピュータ	The computer the outside
41	異なる業種間のビジネス	NIL
42	世界の最新情報	Latest of the world information
43	ネットワークの間を	NIL
44	自分専用のコンピュータ	An exclusive computer one
45	同軸ケーブルのイーサネット	NIL
46	百メガビット／秒のイーサネット	NIL
47	広域のネットワーク	NIL
48	映像のメディア	NIL
49	性能の向上	NIL
50	教育の分野	The education field
51	携帯型の端末	The type end the end of the carrying
52	ユーザ好みの形	NIL
53	サービスの機能	The function of service
54	ネットワークアクセスのモデル	NIL
55	帯域の制限	NIL
56	個別の通信網	NIL
57	均一の形	The shape during uniform
58	端末の接続	The end end the connection
59	音声の伝送	NIL
60	呼制御用のDチャンネル	NIL
61	一定の間隔	The interval certain
62	データの一部	A data the part

3.2. tag 付きサンプル文での日英翻訳の実行

3．1ではそのままのサンプル文では形態素解析が行えなかった。翻訳の変換知識の向上に研究の重点をおくため、あらかじめ形態素解析をし、tag を付与したサンプル文に対して TDMT 翻訳を実行した。

その結果、NIL 表示は消えたが全６２文中４４文が意味不明な翻訳結果を示した。TDMT は旅行会話の翻訳を想定して作られているため、情報処理分野の単語の辞書が存在せず、翻訳できなかったと考えられる。

TAG付き入力での出力結果

1	もっとゆっくり言ってくれませんか	Slowlier _ _ will you tell?
2	これは何の意味ですか	What is this meaning?
3	これは何ですか	What is this?
4	この単語はどこにありますか	Where is there this?
5	著作はだれですか	_ _ who is it?
6	この雑誌はどこにありますか	Where is there this magazine?
7	これはいつ発明されたのですか	When has this been ?
8	なぜ情報は増大するのですか	Why is information?
9	いつこの計算は終了するのですか	When and this calculation end?
10	この式はどういうふうに導かれるのですか	What sort of air this ceremony is _ _ _ _
11	計算量はいくらになるか	_ _ how much _
12	計算時間はどのくらいになるか	How the time is in the calculation _
13	誤りは何個あったか	_ _ how many was there?
14	どちらが正しいか	Which one is righter?
15	どうすればとけますか	If what does _ _ _
16	この値は間違いですか	_ _ _ you a mistake?
17	私の結論は奇妙ですか	I _ _ _ _ is it strange?
18	数字の順序は逆ではないか	Isn't the number opposite?
19	質問してもいいですか	Could I ask?
20	参考書を紹介してくれませんか	_ _ will you recommend?
21	私の名前は山本です	My name is Yamamoto
22	おっしゃることがわかりません	I don't know to say
23	わかりました	I see
24	プリントをください	Gives me the print
25	ヒントを教えてください	_ _ please tell me
26	コンピュータの発展	Of the computer
27	触覚などの感覚	_ _ _
28	複数の情報	A plural information
29	デジタルの技術	_ _ technology
30	別の媒体	Not included _ _
31	移動の向上	The movement
32	情報処理の技術	An information disposition _ _
33	八つの章	Eight _ _
34	誤りの検出	_ _

35	現在の通信	The communication in now
36	社会のグローバル化	_ _ _
37	汎用のコンピュータ	_ _ the computer
38	多数のコンピュータ	A computer
39	情報の交換	Information exchange
40	外部のコンピュータ	The computer the outside
41	異なる業種間のビジネス	Is different _ _ _ _ the business the period
42	世界の最新情報	Latest of the world information
43	ネットワークの間を	_ _ the period _
44	自分専用のコンピュータ	An exclusive computer one
45	同軸ケーブルのイーサネット	_ _
46	百メガビット／秒のイーサネット	One hundred _ _ _ _ the communication for the second , the freeway
47	広域のネットワーク	_ _
48	映像のメディア	_ _
49	性能の向上	_ _
50	教育の分野	The education field
51	携帯型の端末	_ _
52	ユーザ好みの形	_ _ the shape of the taste
53	サービスの機能	The function of service
54	ネットワークアクセスのモデル	_ _ the model
55	帯域の制限	_ _ the restriction
56	個別の通信網	Separately _ _
57	均一の形	The shape during uniform
58	端末の接続	_ _ the connection
59	音声の伝送	The voice _ _
60	呼制御用のDチャンネル	_ _ the channel
61	一定の間隔	The interval certain
62	データの一部	A data the part

3.3. 日英単語辞書登録

日本語の単語から英語の単語に翻訳できるように翻訳辞書に登録を行った。その結果、翻訳文中に下線が入り、文がとぎれとぎれになる[1]ものの、無理矢理理解しようと思えば理解できないことはない文が増えた。翻訳文中に下線が含まれるのは、新たに翻訳辞書に登録した単語を

[1] パターンもしくは用例の不備により、原言語構造の解析が続行できない場合、ＴＤＭＴは入力文をいくつかの部分構造に分割することにより翻訳処理を進行させる。その際、分割部分に対して下線を挿入している。

含めて意味辞書に登録されていない単語が多いため、文の構造を決定できなかったことが主たる原因と考えられる。

翻訳辞書登録後の出力結果

1	もっとゆっくり言ってくれませんか	Slowlier _ _ will you tell?
2	これは何の意味ですか	What is this meaning?
3	これは何ですか	What is this?
4	この単語はどこにありますか	Where is there this word?
5	著作はだれですか	The author _ _ who is it?
6	この雑誌はどこにありますか	Where is there this magazine?
7	これはいつ発明されたのですか	When has this been invented?
8	なぜ情報は増大するのですか	Why is information?
9	いつこの計算は終了するのですか	When and this calculation end?
10	この式はどういうふうに導かれるのですか	What sort of air this formula is _ _ _ _
11	計算量はいくらになるか	The calculation _ _ how much _
12	計算時間はどのくらいになるか	How the time is in the calculation _
13	誤りは何個あったか	The error_ _ how many was there?
14	どちらが正しいか	Which one is righter?
15	どうすればとけますか	If what does _ _ _
16	この値は間違いですか	_ the value _ _ you a mistake?
17	私の結論は奇妙ですか	I _ _ the conclusion _ _ is it strange?
18	数字の順序は逆ではないか	Isn't the number opposite?
19	質問してもいいですか	Could I ask?
20	参考書を紹介してくれませんか	The reference book _ _ will you recommend?
21	私の名前は山本です	My name is Yamamoto
22	おっしゃることがわかりません	I don't know to say
23	わかりました	I see
24	プリントをください	Gives me the print
25	ヒントを教えてください	The hint _ _ please tell me
26	コンピュータの発展	The development of the computer
27	触覚などの感覚	The touch sense _ _ _ the sense
28	複数の情報	A plural information
29	デジタルの技術	Digital _ _ technology
30	別の媒体	Other _ _ the media
31	移動の向上	Increases the movement
32	情報処理の技術	An information disposition _ _ the fundamental
33	八つの章	Eight _ _ the chapter
34	誤りの検出	The error _ _ the detection
35	現在の通信	The communication in now

36	社会のグローバル化	The society _ _ global _
37	汎用のコンピュータ	The general purpose _ _ the computer
38	多数のコンピュータ	Many computer
39	情報の交換	Information exchange
40	外部のコンピュータ	The computer the outside
41	異なる業種間のビジネス	Is different _ _ _ _ the business the period
42	世界の最新情報	Latest of the world information
43	ネットワークの間を	The network _ _ the period _
44	自分専用のコンピュータ	An exclusive computer one
45	同軸ケーブルのイーサネット	The composed of coaxial cable _ _ the ethernet
46	百メガビット／秒のイーサネット	One hundred _ _ the mbit_ _ the communication for the second , the freeway
47	広域のネットワーク	The wide area _ _ the network
48	映像のメディア	The image _ _ the media
49	性能の向上	The performance _ _ increases
50	教育の分野	The education field
51	携帯型の端末	Portable _ _ the terminal
52	ユーザ好みの形	The user _ _ the shape of the taste
53	サービスの機能	The function of service
54	ネットワークアクセスのモデル	The network access _ _ the model
55	帯域の制限	The bandwidth _ _ limitation
56	個別の通信網	Separately _ _ the communications network
57	均一の形	The shape during uniform
58	端末の接続	The terminal _ _ the connection
59	音声の伝送	The speech _ _ the transmission
60	呼制御用のDチャンネル	For the call control _ _ the channel
61	一定の間隔	The interval constant
62	データの一部	A data the part

3.4. 変換知識の追加

意味辞書に登録されていない単語を登録し、文の構造を決定できるようにした。また、単語の合成を行い二つの日本語の単語を一つの英単語に変換できるようにした。その結果、正しく翻訳される率が高まった一方で、問題点も表面化した。

変換知識追加を施した後の出力結果

1	もっとゆっくり言ってくれませんか	Slowlier _ _ will you tell?

2	これは何の意味ですか	What is this meaning?
3	これは何ですか	What is this?
4	この単語はどこにありますか	Where is there this word?
5	著作はだれですか	Who is the author?
6	この雑誌はどこにありますか	Where is this magazine?
7	これはいつ発明されたのですか	When has this been invented?
8	なぜ情報は増大するのですか	Why is information increasing?
9	いつこの計算は終了するのですか	When _ _ dose this calculation end?
10	この式はどういうふうに導かれるのですか	How am I introduced this formula?
11	計算量はいくらになるか	How much is the calculation required?
12	計算時間はどのくらいになるか	How is the time in the calculation?
13	誤りは何個あったか	How many an error were there?
14	どちらが正しいか	Which one is right?
15	どうすればとけますか	If what does , is solved?
16	この値は間違いですか	You the value a mistake?
17	私の結論は奇妙ですか	Am I the conclusion strange?
18	数字の順序は逆ではないか	Isn't the number opposite?
19	質問してもいいですか	Could I ask?
20	参考書を紹介してくれませんか	Will you recommend the reference book?
21	私の名前は山本です	My name is Yamamoto
22	おっしゃることがわかりません	I don't know to say
23	わかりました	I see
24	プリントをください	Gives me the print
25	ヒントを教えてください	Please tell me to hint
26	コンピュータの発展	The development of the computer
27	触覚などの感覚	The sense for the touch sense
28	複数の情報	A plural information
29	デジタルの技術	A digital technology
30	別の媒体	Another media
31	移動の向上	Increases the movement
32	情報処理の技術	The fundamental of information processing
33	八つの章	Eight of the chapter
34	誤りの検出	The detection of the error
35	現在の通信	The communication in now
36	社会のグローバル化	The globalization of the society
37	汎用のコンピュータ	The general purpose computer
38	多数のコンピュータ	Many computer
39	情報の交換	Information exchange
40	外部のコンピュータ	The computer the outside

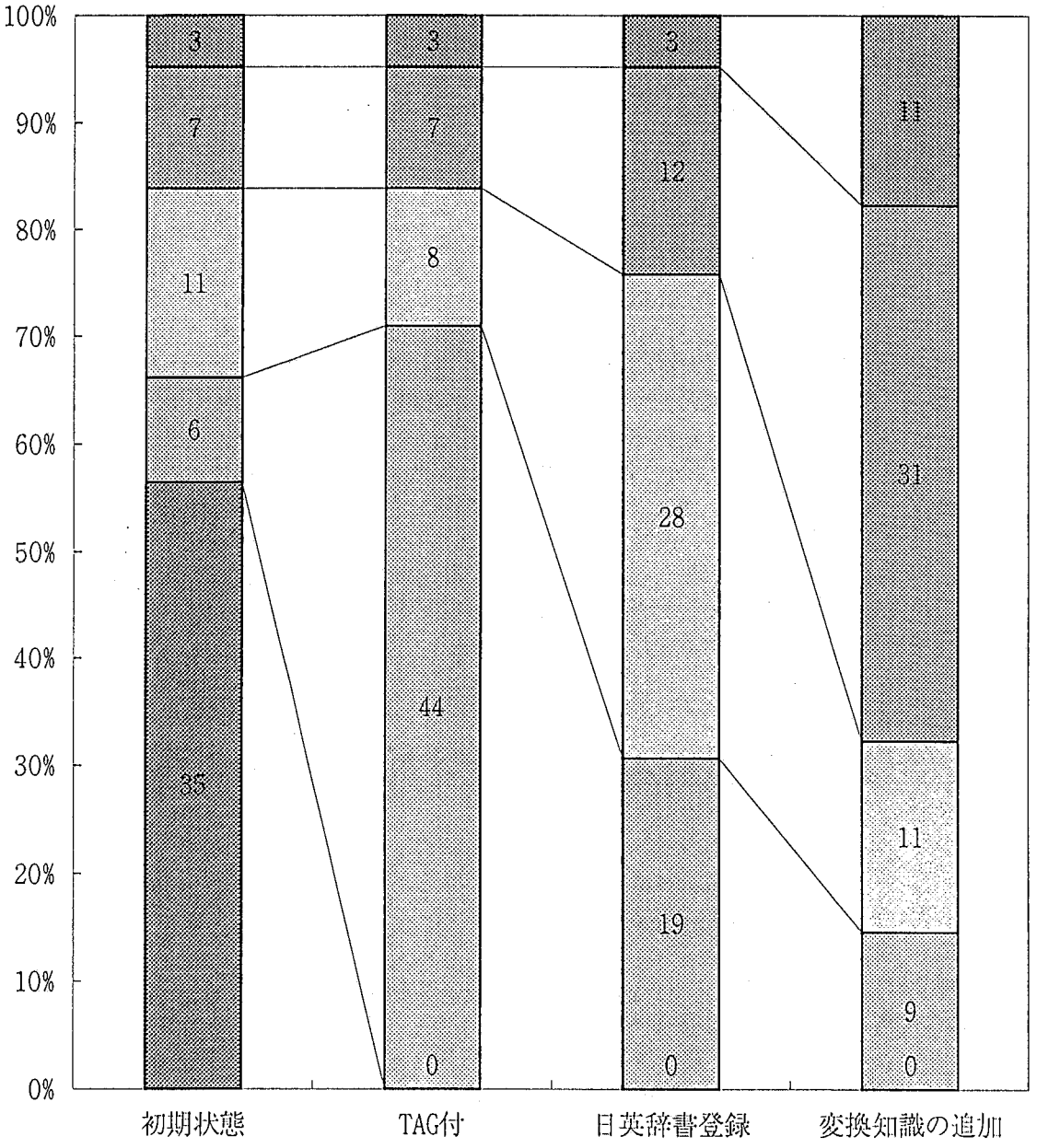
41	異なる業種間のビジネス	The business of interdisciplinary
42	世界の最新情報	Latest of the world information
43	ネットワークの間を	Between the network _ _
44	自分専用のコンピュータ	A private computer
45	同軸ケーブルのイーサネット	The ethernet in the composed of coaxial cable
46	百メガビット／秒のイーサネット	One hundred mbits _ _ the communication for the second , the freeway
47	広域のネットワーク	The network the wide area
48	映像のメディア	The image media
49	性能の向上	The increases the performance
50	教育の分野	The education field
51	携帯型の端末	A portable the terminal
52	ユーザ好みの形	The shape of the user taste
53	サービスの機能	The function of service
54	ネットワークアクセスのモデル	The model the network access
55	帯域の制限	The limitation the bandwidth
56	個別の通信網	Separately communications network
57	均一の形	The shape during uniform
58	端末の接続	The connection of the terminal
59	音声の伝送	The transmission in the speech
60	呼制御用のDチャンネル	The d channel for the call control
61	一定の間隔	The interval constant
62	データの一部	A data the part

翻訳結果の推移

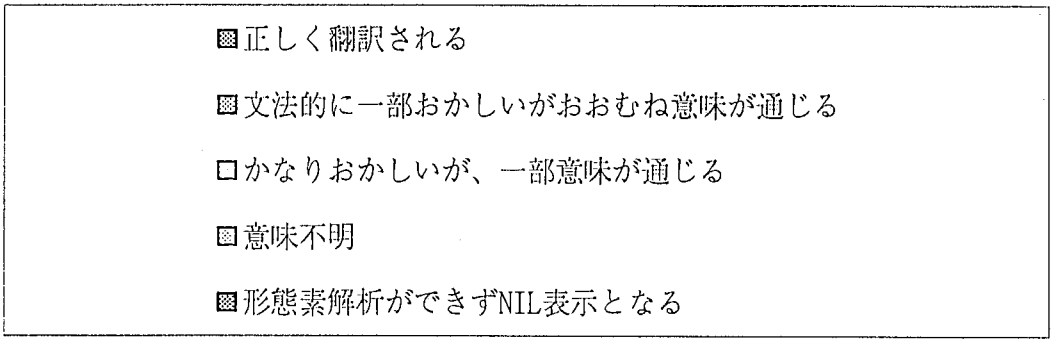
	初期状態	TAG付	日英辞書登録	変換知識の追加
正しく翻訳される	3	3	3	11
文法的に一部おかしいがおおむね意味が通じる	7	7	12	31
かなりおかしいが、一部意味が通じる	11	8	28	11
意味不明	6	44	19	9
形態素解析ができずNIL表示となる	35	0	0	0

単位：文

翻訳結果の推移



注：グラフ中の数値は文の数



3.4.1. 訳し分け

一つの日本語に対応する複数の英語表現が存在し、その会話が発せられた状況や文脈の違い等により、訳語が異なることがある。

旅行会話向けの翻訳システムの影響を受け、情報処理分野の翻訳に支障が出る場合

例

「わかりました」→ "I see"
"I understand"

通常旅行会話では質問などの返答に対して納得したことを表す表現として "I see" と発するケースが多い。旅行会話での翻訳を対象としてきた TDMT では「わかりました」と入力されると "I see" と翻訳するように登録されている。ところが、情報処理分野で沿革教育を想定する場合では、相手が理解したか否かを訪ねるのに、"I see" と返答されたのでは、本当に理解したかどうか不確定要素を残してしまう。この場合、理解したことを示す "I understand" と翻訳する方が正確である。つまり、情報通信分野ではこのようなケースが多く、"I understand" を標準的な訳語と登録する方が適当である。TDMT では状況判断をして、訳を変えることは現段階ではできない。今回は情報処理分野での翻訳を可能とする研究目的であるので、「わかりました」に対応する訳語を "I understand" に変更した。

前後の修飾関係などにより訳し分けが必要となる場合

例

「移動の向上」→ "increase of movement"
「性能の向上」→ "performance improvement"

情報処理分野に翻訳を限定しても、訳し分けが必要となる場合がある。上記例のような場合、local 辞書に用例を追加することにより解決できる。しかし、もっと open な入力に対しても対応できるように、カテゴリーを分けることによる訳し分けの方法や、表現（機能語）に関する訳し分けの翻訳用例を研究し、対応策を考える必要がある。

3.4.2. 文法的に誤った結果を示す場合

単数、複数形の障害

例

「誤りは何個あったか」→ "How many an error were there?"

" errors "と翻訳されるべきところが "an error"となった。"How many" のあとに続く名詞は複数形にするルールを定義すれば解決すると思える。しかし、この場合は、
1.「誤り」は「ある」
2.「何個」　「ある」
と構造的分割され、それぞれのパターン内で「誤り」→"error" 「何個」→"how many"と」翻訳されている。つまり、ルール内で「誤り」と「何個」の間にはつながりはなく、「何個」のあとの名詞は複数形にするといったルールでの解決法は、この場合できない。構造決定する段階で「何個」と「ある」が結びつくように用例、ルールを工夫することでこの場合は解決可能かもしれないが、疑問文における文法的な誤りは今後多数発生する事が予想されるため、何らかの新たな対応策が必要と思われる。単数、複数の問題点としては、特に用法によって加算、不可算になる名詞においての冠詞の付け方を誤るケースが多い（例、calculation)。翻訳における冠詞の付け方は困難な部分である。この場合は類語辞書に名詞を登録する際に文法的な意味を示すコードを付与することで対応できるのではないかと思う。しかし、新たな意味コードを付与することは困難であるため、別の解決法を模索する必要がある。

時制（テンス）の誤り

例
「これはいつ発明されたのですか」→　"When has this been invented?"

過去形で変換されるべきところが、現在完了形となっている。時制を意識せずに翻訳用例に当てはめたため、このような誤った翻訳結果を出力した。

アスペクトの誤り

テンスと同様に翻訳が難しいとされているものが、アスペクトである。「～ている」「～てある」といった語の翻訳である。今回のサンプル文では存在しなかったが、テンスと同様に今後発生が予想される問題であるので、参考までにする。

他人にお願いする場合等の敬語表現

例
「もっとゆっくり言ってくれませんか」→　"Will you speak slowlier?"

正しく翻訳すると、"Could you speak a little slower?"　となる。(slowlier は more slowly となるべきであり、活用辞書の対応が今後必要)　お願いの表現であるにもかかわらず、通常の未来形で翻訳された。慣用句で決まった用例が登録されている場合を除いてこのように翻訳される。意味は伝わるが、目立つ箇所であるので今後「お願い」の概念の導入、又は「お願い」のパターンの強化をする必要がある。

3.4.3. 誤った用例の影響を強く受ける場合

似通った単語を含む誤ったパターンの影響を受け、誤った翻訳結果を出す場合がある。

用例を追加することで解決可能な場合

例
「どちらが正しいか」→ "Which one is righter?"

"righter"は"right"と翻訳されるべきである。これは、“「どちら」が「良い」”という用例(”Which one is better”)を基にパターンが決定されたため、「正しい」と翻訳する部分に比較級が適用された。

今回は、比較級を用いないパターンに “「どちら」が「正しい」” を登録して解決した。意味辞書では「良い」は"173"、「正しい」は"175"で登録されている。また、「良い」と同じパターンには「安い」と「早い」が登録されている。意味コードは、「安い」は"171"、「早い」は"153"である。パターンに登録されていない単語に対して、この番号の差でパターンを選択することは困難と思われる。また、文法的な視点から誤りを矯正することも困難と思われる。この場合は用例を多数追加する事で対応する事が望ましい。

用例を追加することでは解決困難な場合

例
「いつこの計算は終了するのか」→ “When and this calculation ends?”

ここで、問題なのは ”When and this calculation” という ”X and Y” というパターンが用いられたことである。これは、「いつどのようなパーティが開かれますか」という文を翻訳するためにつくられた、”X and Y” つまり “When and what kind of” というパターンを用いていた。Y が連体詞＋名詞のパターンのため、「この計算」との意味距離が極めて近くなってしまった。参考までに、「いつ計算は終了するのか」を入力すると”When will the calculation end?” となる。変換でこのパターンが用いられるように、「いつ」「計算」「終了」といった単語を登録しても、”X and Y”のパターンが用いられた。

3.4.4. 英語に翻訳する際に新たな表現を付与する必要がある場合

例
「計算量はいくらになるか」→ “How much calculation is required?”

機械的に翻訳すると ”How much is the calculation?” となる。どのようにして日本語の表現に表れない ”be required” を付加するかという問題が生じた。今回は「計算量」という単語に対する local 辞書を作成したが、今後オープンな対応策が必要である。

3.4.5. 単位、数字に関する問題

現段階では数字を一つ一つ単語として認識している。数学的な翻訳分野が増えるにつれ、数字の羅列、分数、少数、変数の扱いが問題となる。また、単位についても、どこまでが単位なのか、単位の合成、数字とのつながりといった面で充分なルールが存在しない。複雑な場合にも対応可能なルールが必要である。

3.4.6. 「の」の訳し分けについて

機能語の訳し分けとして、「の」の変換に注目した。専門用語に対して、「の」がどの程度正しく翻訳できるかという実験であったが、結果はあまり良くない。原因は、専門用語に意味的に近い用例が登録されていないためである。今後は、専門用語を用例に多数登録し、再度検証する。

4. 結論

今回のサンプル文に対してＴＤＭＴの適用実験を行った結果、

- 辞書登録等の作業を施すことで、大幅に正しく翻訳される率が上昇し、今後もさらに基本的な処置を施すことで、さらに上昇が見込まれる。
- 現段階で、致命的な問題点は存在していない。今回表面化した問題点は、今後研究を進めることで改善できる可能性が高い。

と言える。

このことより、現段階では質議応答文に対する協調融合翻訳の適用は可能と判断した。しかし、サンプル文６２文に対して行った実験であり、今後１０００～２０００文のサンプルに対して、単語変換辞書、意味辞書、用例の追加等を行い、基本となるデータベースが作成された時点で改めて判断する必要がある。

5. 今後の研究

5.1. 基本対訳コーパスの作成

サンプル文を大量に増やし、問題点をさらに明確化し、解決して行く必要がある。ここで、全問用語の意味辞書への登録ルールが確立されておらず、サンプル文を増やした時に問題となることが予想される。現在のＴＤＭＴでは角川の類語辞典に登録されているコードを基本に意味辞書が登録されているが、専門用語は類語辞典にはなく、今後ルールを作成し、的確に登録する必要がある。

5.2. 機能語（特に「の」）の訳し分けの研究

機能語の訳し分けは難しいとされている。しかし、質議応答文では質問を正しくために機能語を正しく訳す必要性は高い。その中で、特に「の」の翻訳に注目して研究を進める。今回の翻訳結果でも、「の」が正しく翻訳されなかった率は高い。「の」の訳し分けについての大量のサンプルを解析する事で、訳し分けの手法を検討する。

5.3. 生成ルールの追加・修正

実験で用いたサンプル文の翻訳で、文法的な誤りが多数存在した。これは、最終的に文を生成する際に文法的な誤りをあまり修正していないためである。文の生成に関してルールを追加、修正をすることが望ましいと思われる。

5.4. 英日翻訳の研究

今回は、日英翻訳に的を絞って研究を行ったが、英日翻訳も重要な課題である。日英翻訳の研究がある一定レベルに達し次第、英日翻訳の研究に着手する。

5.5. 音声認識への対応

自動翻訳電話を完成させるためには、音声認識技術とのリンクは欠かせない。音声認識の技術で認識された文に対して、いかに協調融合翻訳を対応させるかという点をふまえて、今後研究を進める。

参考文献

- [Furuse94] Osamu Furuse and Hotoshi Iida. Constituent Boundary Parsing for Example-Based Machine Translation. In *Proceeding of Coling '94*, Vol.1, pp. 105-111, 1994.
- [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. 情処学論、Vol.35,No3,pp.414-425,1994.
- [美馬 96] 美馬秀樹、古瀬蔵.日英間変換主導型翻訳の中間時評価. TR-IT-0189,1996
- [美馬 97] 美馬秀樹. 対話の機械翻訳. AAMT Journal No.19, June, 1997, pp.11-17