



AAMT

The Asia-Pacific Association for Machine Translation

Journal

No.19

June 1997

アジア太平洋機械翻訳協会

目 次

委員会報告	機械翻訳ユーザ辞書共通フォーマット UPF 設定活動報告	1
ヒアリング	放送分野における機械翻訳の利用	6
ヒアリング	国連大学におけるネットワーク言語の研究開発	9
技術早わかり	対話の機械翻訳	11
研究機関紹介	慶応義塾大学環境情報学部・政策メディア研究科 石崎研究室	18
留学生雑感	日本語の勉強と翻訳	21
学会報告	言語処理学会 第3回年次大会 参加報告	25
関連団体情報	CICC 近隣諸国間の機械翻訳システムプロジェクト	29
	ネットワーク上での英語の使用について	34
新製品紹介	用語と訳語の統一用ソフト HT ³ の開発	36
	インターネット翻訳の王様	41
	翻訳支援システムツール TRANSIT	44
	日英翻訳ツール「これと英! Ver. 1.0 for Windows95」	46
	パーフェクト日英 訳せ!! ゴマ	48
関連団体情報	(社)日本翻訳連盟の平成9年度事業計画(案)	51
協会便り	AAMT 5周年記念シンポジウムの開催案内	4
	第7回通常総会開催案内	20
事務局便り	協会活動報告	8
	研究会メンバーの募集	10
	MT Summit VI開催近づく!	17
	委員交替のご連絡	17
	MT Summit VIの展示会出展案内	43—表3
	AMTA、EAMTもホームページを開設	45
イベント紹介	COM JAPAN 1997	5
	(社)日本翻訳協会設立十周年記念 第8回翻訳フェア	10
	「テクニカルコミュニケーションシンポジウム'97」の案内	50
	JTF 翻訳環境研究会 年間カリキュラム	52

機械翻訳ユーザ辞書共通フォーマット UPF 設定活動報告

技術動向調査委員会 委員長 村木 一至
委員 亀井真一郎
UPF 実行連絡会 幹事 桧山 努

1 背景

商用の機械翻訳ソフトウェアが製品化されて10年余が経過した。当初、機械翻訳ソフトウェアは翻訳の専門家向けに設計・開発され、主に技術翻訳に用いられてきた。しかし、近年のインターネットの発達とパソコンの普及により、機械翻訳ソフトウェアは一般ユーザにも急速に浸透し始めている。機械翻訳ソフトウェアを有効活用するためには、各ユーザ毎に、頻繁に使用する語彙を「ユーザ辞書」として蓄積し、ソフトウェアの基本辞書と合わせて使う必要がある。しかし辞書作成は一般に時間と労力がかかる仕事であり、個人ユーザー一人一人が辞書を個々に充実させるのには限界がある。この問題を解決するために、個人が個別に蓄えている辞書データを流通させ、相互利用するための機構を開発する活動を開始した。我々はこの機構をUPF (Universal Platform) と名づけている。本稿では、このUPFの開発活動について述べる。

2. UPF の概要

現在、機械翻訳ソフトウェアはそれぞれ独自の辞書を持っている。それぞれの辞書は各機械翻訳ソフトウェアに特化したものであり、データ構造や必要な辞書内容は統一化されていない。そのため、機械翻訳ソフトウェア間で相互に辞書を利用することはできない。そこで、異なる機械翻訳ソフトウェア間で独立に定義されている辞書形式のうち、利用者に仕様が公開されているユーザ辞書を対象として、各メーカーのユーザ辞書と相互変換が可能な「共通辞書」を想定し、記述内容と記述形式を設計する。各社の機械翻訳ソフトウェアのユーザ辞書をこの「共通辞書」に一度変換し、変換した共通辞書を個々の機械翻訳ソフトウェアのユーザ辞書に変換することにより、異機種 of 辞書情報が相互に利用可能となる。

ここで設計した共通辞書の記述仕様の実用性を保証するために、実システムによる評価実験を行なう。

評価実験では、記述仕様に準拠した対訳辞書を、各機械翻訳ソフトウェアのユーザ辞書に変換して翻訳実験を行ない、翻訳品質の評価を行なう。また、辞書作成を効率化し、流通を促進する辞書エディタ等のツールを開発し、さらにデータ交換、蓄積、流通のためのWWWサーバを用意する。

3. 開発体制

UPF は情報処理振興事業協会 (IPA) 殿の創造的ソフトウェア育成事業の一つとして、平成8年度と平成9年度の2年間で開発を行なう。実際の開発は、アジア太平洋機械翻訳協会技術動向調査委員会を母体として、UPF 仕様検討委員会を設置し、その指導の下、(株) NEC 情報システムズ内にUPF 実行連絡会を置き、開発を行なっている。参加企業は以下の通りである。

UPF 仕様検討委員会

NEC、キャノン、東芝、富士通研究所、ノヴァ、日立製作所、沖電気工業、シャープ、松下電器産業、カテナ、インターグループ、三洋電気、NEC 情報システムズ

UPF 実行連絡会

NEC、富士通研究所、ノヴァ、シャープ、インターグループ、東芝、NEC 情報システムズ

4. 活動項目

UPF 設定活動では、以下の項目について開発を行なう。

- (1) UPF 言語変換標準記述仕様の設計 (平成8、9年度)

UPF 言語変換標準記述仕様は、「基本言語変換標準記述仕様」と「拡張言語変換標準記述仕様」に分けて開発を行なうことにした。一般に機械翻訳ソ

ソフトウェア用の対訳辞書には「品詞」や「活用」といった最低限の翻訳を可能とする情報が記述されている。これらは異なった機械翻訳ソフトウェア間で値まで共通化できる情報であると考えた。これを「基本言語変換標準記述仕様」と呼ぶ。

これに対して、「意味分類」や「英語の動詞を持つ格文型（英語でSV、SVO、SVC、SVOO、SVOCのどの文型をとるか）」の詳細のように翻訳の質の向上に必要な情報が存在する。これらの情報は複数の機械翻訳ソフトウェア間で値まで共通化することは困難であると考えた。そこでこのような情報に対しては、情報を記述するための汎用的な記述形式だけを設計する。これを「拡張言語変換標準記述仕様」と呼ぶ。

（２）対訳辞書開発支援ソフトウェアの開発（平成 8、9 年度）

UPF 言語変換標準記述仕様に準拠した対訳辞書の作成を支援する「対訳辞書開発支援ソフトウェア」を開発する。対訳辞書開発支援ソフトウェアとしては、基本言語変換標準記述仕様に準拠した辞書データを入力・修正するための「基本言語変換標準準拠対訳辞書エディタ」、拡張言語変換標準記述仕様に準拠した辞書データを入力・修正するための「拡張言語変換標準準拠対訳辞書エディタ」、複数の対訳辞書を管理する「対訳辞書管理ソフトウェア」を開発する。

（３）UPF 言語変換標準準拠対訳辞書（平成 9 年度）

UPF 言語変換標準に準拠した辞書の実例として、ビジネス分野の語彙を対象に、日英、英日各々 2 万語の「ビジネス分野日英対訳辞書」と「ビジネス分野英日対訳辞書」を開発する。

（４）実証評価支援システムの開発（平成 9 年度）

UPF 言語変換標準記述仕様の実用性を検証するために、WWW サーバを用いて、UPF 言語変換標準記述仕様、対訳辞書管理ソフトウェアなどを一般にも公開するとともに、利用者間で UPF 言語変換標準に準拠した対訳辞書を流通、交換できる機能を開発する。

（５）評価実験作業（平成 9 年度）

UPF 言語変換標準に準拠した対訳辞書が、複数の機械翻訳ソフトウェアのユーザ辞書に対して相互変換が可能であることと、変換した辞書を実際に機械翻訳ソフトウェアのユーザ辞書として用い、翻訳実験を行なうことにより、UPF 言語変換標準の実用性の評価を行なう。

５．これまでの活動状況

平成 8 年度に開発を行なった「基本言語変換標準記述仕様」と「基本言語変換標準準拠対訳辞書エディタ」について、その活動状況、および成果について述べる。

５．１ 基本言語変換標準記述仕様の開発

（１）作業概要

基本言語変換標準記述仕様の開発では、品詞や活用といった辞書情報を規定する「記述内容」と辞書情報を記述する形式を規定する「記述形式」に分けて開発を行なった。「記述内容」に関しては、各社機械翻訳ソフトウェアのユーザ辞書仕様を調査し、比較結果を基にして、基本言語変換標準の記述仕様案を作成した。記述仕様案を基にして、4 回の UPF 実行連絡会を開催し、基本言語変換標準記述仕様の詳細検討を行なった。

第 1 回（10/31）

議題：UPF 言語変換標準の開発方針
基本言語変換標準の品詞設定

第 2 回（11/18）

議題：格パターンの記述形式・意味分類について

第 3 回（11/29）

議題：格パターンの記述形式の再検討

第 4 回（12/13）

議題：格パターンの記述形式の再検討
品詞毎の辞書情報の検討

基本言語変換標準の記述形式

基本言語変換標準日英・英日対訳エディタの仕様検討

（２）方針

基本言語変換標準においては特に言語的な厳密さより機械翻訳における情報流通性を重視し、また機械翻訳用のユーザ辞書としての有効性を得るために

以下の方針とした。

- ユーザ辞書登録では登録語彙の9割が名詞・固有名詞であり、特に固有名詞が多いことから、固有名詞を名詞の下位分類の一つとして登録しやすくする。
- 動詞・形容詞も登録できるようにするが、記述の繁雑さを避けて流通を促進するため、記述できる詳細情報は典型的な型に制限する。詳細情報の記述は拡張言語変換標準に従って行なうこととする。

- いわゆる「処理する」のようなサ変動詞も終止形登録として動詞の一つの型として扱う。例えば、「導入」「導入する」といったいわゆるサ変名詞・サ変動詞の場合には、名詞と動詞の二つのエントリに分けて登録する。つまり「導入」という見出しを名詞として作り、訳語として「introduction」のような英語の名詞を与える。これとは別に「導入する」という見出しを動詞として作り、訳語として「introduce」のような英語の動詞を訳語とする。

- 日英・英日の登録を1回で済ませるために、英日日英は共通形式とする。

この方針に基づき、基本言語変換標準では以下の品詞を扱い、これ以外の品詞は拡張言語変換標準で扱うこととする。

日本語：名詞、動詞、形容詞、形容動詞、副詞

英語：名詞、動詞、形容詞、副詞

一般に日本語の品詞と英語の品詞の対応は複雑であるが、記述を簡素化するため、日英の品詞の組み合わせは以下に示す内容に限定した。

日英

日本語品詞	英語品詞
名詞	名詞
固有名詞	固有名詞
動詞	動詞
サ変動詞	動詞
形容詞	形容詞
形容動詞	形容詞
副詞	副詞

英日

英語品詞	日本語品詞
名詞	名詞
固有名詞	固有名詞
動詞	動詞、サ変動詞

形容詞

形容詞、形容動詞

副詞

副詞

(3) 辞書記述例

基本言語変換標準の記述形式としては、SGMLのようなマークアップ言語を用いることとした。以下に名詞の辞書記述例を示す。

```
<entry>
  <japanese>
    <jentry>本</jentry>
    <jpos>名詞</jpos>
    <jnountype>普通名詞</jnountype>
    <sem>その他の具体物</sem>
  </japanese>
  <english>
    <eentry>book</eentry>
    <epos>noun</epos>
    <enum>c</enum>
    <epl>books</epl>
    <eheadpron>consnant</eheadpron>
  </english>
</entry>
```

5.2 基本言語変換標準準拠対訳辞書エディタ

辞書記述者の負担を軽減し、辞書情報の流通を促進するため、Windows95とUNIX上で動作する、基本言語変換標準の記述仕様に準拠した記述項目を持つ専用の辞書開発用のエディタの開発を行なった。基本言語変換標準の記述仕様は、記述項目が有限であり、言語ごと、あるいは品詞ごとにその記述項目のセットが異なる。そのことを利用して、基本言語変換標準準拠対訳辞書エディタは記述者の入力に制約を与えることで、ミスタイプや、形式エラーを防止することができる。つまり、基本言語変換標準の記述仕様に準拠した対訳辞書を作成する際に、基本言語変換標準の記述仕様の記述形式および記述内容の整合性が保証されるため、対訳辞書を作成するユーザは記述内容と記述形式の整合性等を意識することなく基本言語変換標準記述仕様に準拠した辞書を開発することが可能となる。

基本言語変換標準準拠対訳辞書エディタは以下の機能を持つ。

- (A) 対話的対訳辞書情報入力機能
- (B) 辞書情報実時間登録、修正、削除機能
- (C) 基本言語変換標準整合性チェック機能
- (D) 対訳データ印刷機能

5.3 学会発表

昨年度は、以下の学会においてUPFの活動報告行なった。

- (1) 情報処理学会第54回全国大会
1997年3月13日 千葉工業大学
「商用機械翻訳ユーザ辞書の共通フォーマット設定に向けて」発表者 亀井真一郎
- (2) 言語処理学会第3回年次大会
1997年3月27日 京都大学
「機械翻訳ユーザ辞書の共通フォーマットの設定」発表者 伊藤悦雄

6. 今後の予定

今後、継続して拡張言語変換記述仕様の開発を行ない、その記述仕様に準拠したUPF言語変換標準準拠日英、英日対訳辞書の開発を行なう。

次いで、このUPF言語変換標準準拠対訳辞書を用いて、UPF言語変換標準記述仕様の有用性を実証するための評価実験作業を行なう。

さらに、辞書開発を支援する拡張言語変換標準準拠対訳辞書エディタ、対訳辞書管理ソフトウェアを開発する。また、本プロジェクトの成果は、人工知能学会第11回全国大会(6/25早稲田大学)、MT-SummitVI(10/29~11/1 San Diego)で報告を行なう予定である。

UPF開発活動の概要はAAMTのホームページ(<http://www.jeida.or.jp/aamt>)に掲載している。成果に対して多くの皆様からコメントをいただき、仕様をより良いものとしてゆきたい。

協会便り

AAMT 5周年記念シンポジウムの開催案内

アジア太平洋機械翻訳協会として活動を開始してから5年が経過したことを記念して、5周年記念シンポジウムを開催します。

テーマ 「次代の翻訳を拓くコーパスベース技術」
 日時 '97年6月23日(月) 9:30 ~ 15:50
 場所 機械振興会館 地下大ホール(定員250名)
 (東京都港区芝公園3-5-8 電話 03-3434-8211(代表))

《スケジュール》

- 9:30 開会
- <特別講演>
- 9:30~9:50 「コーパスベースの自然言語処理技術の展望」
… 田中穂積会長 (東工大教授)
- <委員会・研究会報告>
- 9:50~10:10 技術動向調査委員会報告 (IPAプロジェクト開発内容)
… 村木一至委員長 (日本電気株式会社部長)
- 10:10~10:30 市場動向調査委員会報告 … 石崎 俊委員長 (慶応義塾大教授)
- 10:30~10:50 例文評価研究会報告 … 横山晶一 座長 (山形大教授)
- <招待講演>
- 11:00~11:30 「用例と規則の融合翻訳」
… 古瀬蔵氏 (ATR 音声翻訳通信研究所 第3研究室主任研究員)

- 11:30~12:00 「遺伝的アルゴリズムを用いた適応型機械翻訳手法」
 … 荒木健治氏（北海学園大学工学部 電子情報工学科 助教授）
- <一般講演>
- 13:00~13:15 「事例ベース翻訳システム TOPTRAN の現況」
 … 仁井正治氏（凸版印刷 株式会社 マルチメディア事業部部長代理）
- 13:15~13:30 「知識と翻訳…MTのはたす役割」
 … 堂野前進氏（長瀬産業 株式会社 総務部文書課 翻訳担当次長）
- 13:30~13:45 「文書走査を利用した複合名詞解析」
 … 久光 徹氏（日立製作所 基礎研究所研究員）
- 13:45~14:00 「コーパスからの言語知識獲得技術と機械翻訳への応用」
 … 北村美穂子氏（沖電気工業 株式会社 研究開発本部 関西総合研究所）
- 14:00~14:15 「英語ニュース作成支援のための用例提示システム」
 … 浦谷則好氏（NHK 放送技術研究所 先端制作技術研究部）
- 14:15~14:30 「パターンベース翻訳」
 … 渡辺日出雄氏（日本アイ・ピー・エム 株式会社 東京基礎研究所副主任研究員）
- 14:45~15:00 「アナロジーに基づく構文解析」
 … 安藤真一氏（ATR 音声翻訳通信研究所 第三研究室）
- 15:00~15:15 「知識ベース増殖のためのソフトウェア」
 … 杉山健司氏（株式会社 日本電子化辞書研究所 研究部長）
- 15:15~15:30 「定形文の用例ベースの適用」
 … 畠中伸敏氏（キヤノン 株式会社 情報通信システム本部担当課長）

<特別講演>

- 15:30~15:50 「言語知識のグローバル共有」… 辻井潤一副会長（東大教授）

15:50 閉会

《参加費》 会 員：無料 非会員：2,000円

イベント紹介

COM JAPAN 1997

情報・通信を融合した総合展示会

会 期：1997年11月4日（火）～7日（金）

会 場：東京ビッグサイト（有明：東京国際展示場）

主 催：通信機械工業会、日本電子工業振興協会、日本パーソナルコンピュータソフトウェア協会

入場料 ； 無料

予定来場者数：約300,000人

幹事事務局 ； 〒105 東京都港区芝公園3-5-8 機械振興会館 日本電子工業振興協会内

電話 03-5405-7256 FAX 03-3433-2003

放送分野における機械翻訳の利用

NHK 放送技術研究所 先端制作技術研究部 主任研究員 江原暉将

1. はじめに

放送業界での機械翻訳の市場規模はそれほど大きいものではないが社会的なインパクトは非常に大きい。しかし、今のところ NHK 以外での機械翻訳の開発や利用の話はきいていないので、ここでは NHK での現状について2つのシステムを説明する。

2. 日本語字幕作成支援システム

2. 1 本システム開発の背景

番組の国際化が進み、日本で放送される海外番組が増加してきた。特に放送衛星で英語番組が多用されるようになった。それにともない翻訳量が增大してきた。使用言語は大半英語であるため、英語から日本語に翻訳して字幕を作る「日本語字幕作成支援システム」の研究を、1988年4月に開始した。そして、1989年8月から試験使用を始めた。現在、平日は毎日、17時20分から3分間衛星第一放送でこのシステムを用いた放送を行っている。

2. 2 システムの概要

ニュース文には次のような特徴がある。

- ① 内容がさまざまに広範囲である。
- ② 固有名詞や新語が頻出する。
- ③ 話しことばあるいは、読みことばであり、文字化されていない。(最近は聴力障害者向けに付加されたクローズドキャプションを利用している)

これに適応するように開発されたこのシステムは、下記のような特徴をもっている。

- ① 文法のカバーする範囲が広い。
- ② 放送用語辞書を利用している。
- ③ 固有名詞や数量表現の処理に強い。

システムは英文入力機能、機械翻訳機能および字幕編集機能をもっている。英文入力機能は英文ワープロによる入力であり、スペルチェックも行う。機械翻訳機能は辞書検索、英文解析、英日変換、日本

語生成および局所翻訳からなる。字幕編集機能には、文型選択および訳語選択の機能があり、日本語ワープロで出力される。

英語ニュースはまず翻訳者によってクローズドキャプションを利用して、あるいはキーボードを用いて入力される。次いで、機械翻訳にかけられテロップ原稿が作成される。原稿はデスクのチェックを経た後電子テロップに作成され放送される。

翻訳対象が広いと、充実した用語辞書が必要である。現在、基本語5万語と放送用語7万語からなる辞書を使っているがまだ不足である。また、適用範囲の広い英文解析文法が必要で、現在約3千の文法規則を利用している。局所翻訳としては例えば、

<U.S. President Bill Clinton>

の場合は、President を「大統領」と、

<Chinese President Jiang Zemin>

の場合は、「国家主席」と訳されるようになっている。数量表現においても、同様の工夫がなされている。

使いやすさを考えて、ヒューマンインターフェースとしてパソコンを利用した。マッキントッシュ2 CI で、時間当たり約1万語の翻訳が可能である。

このシステムによる翻訳例を次に示す。

《原文》The revolutionary British hovercraft is on course to win the blue ribbon for the fastest Atrantic crossing.

《機械翻訳出力》革命のイギリスのホバークラフトは、最も早い大西洋の横断に一等賞を得させるため、コースに存在する。

《実際の字幕》英のホバークラフトは、大西洋の横断で一等賞を得るため出航。

字幕は字数を多くできないため、約半分に縮めなければならない。これも重い後編集の理由の一つになっている。

2. 3 システムの評価

3分の字幕を作成する場合に、このシステムを使ったときと、すべて人手で行ったときの作業時間を比較した。

	機械支援	全人手
英文聴取・文字化	10分	—
機械翻訳・内容分析・再翻訳	60	—
日本語字幕原稿作成	50	60分
日本語字幕データ作成	30	30
タイミング入力	30	30
合計	3時間	2時間

現状では、機械を使用することによって作業時間は50%長くなっている。

翻訳精度については次のような実験を行った。

《使用した評価文》

テスト文名称	略称	文数	備考(作成・出所)
衛星放送	BS	201	BS ワールドニュース
JEIDA	JD	309	日本電子工業振興協会
NARITA	NRT	228	大阪大学(成田先生)
AP	AP	179	AP 通信社のニュース

《評価者》衛星放送の翻訳者（95年度は3人、96年度は1人）

《評価方法》文単位に下記の5段階の主観評価基準を用いて文単位の正翻訳率で評価した。

「評価基準」

文意忠実度に基づく、以下の5段階で評価する。

5. 入力文の文意は出力文に忠実に再現されている。
4. 入力文の大意は出力文に再現されているが、細部に問題がある。
3. 入力文の大意は出力文にほぼ再現されているが、誤解を受けやすい部分がある。
2. 入力文の文意は出力文にほとんど再現されていない。
1. 入力文の文意は出力文に全く反映されていない。

《正翻訳率》

正翻訳率＝

(評価5 + 評価4 + 評価3) / (評価5 + …… 評価1)

《評価結果》

正翻訳率

	JD	NRT	BS	AP
'95年実験正翻訳率	74.6%	61.4%	61.0%	52.5%
'96年実験正翻訳率	75.9	64.5	63.5	53.6
正翻訳率の向上	1.3	3.1	2.5	1.1

2. 4 翻訳精度の向上

実用という面から、どのくらいの翻訳精度が最低必要であろうか。筆者は、自分のために自分が理解できればよいという場合には60%ぐらい、翻訳して他人に情報提供するという場合には80%以上の精度が要求されると試算している〔1〕。

英日翻訳の場合80%以上の正翻訳率を実現するためには以下のことが有効であると考えている。統計的な手法による構文解析を用いることにより、英語構文解析精度を90%まで向上させる。名詞句と動詞句などの共起データを利用して、英日翻訳の訳語選択精度を90%まで向上させる。これによって80%の翻訳精度を確保する。

3. 日英対訳記事データベースを利用した翻訳支援システム

「日本語字幕作成支援システム」について、「日英対訳記事データベースを利用した翻訳支援システム」を開発した。本システムは、日英翻訳を支援するシステムである。

翻訳者が求める支援には次のようなものがある。

- ① 固有名詞は過去にどのように訳されているか。
- ② ニュース特有の用語や表現は過去にどのように訳されているか。
- ③ 記事の構成は日英翻訳においてどのように変わるか。

そこで、図-1に示すようなシステムを開発した〔2〕。現在、8,000の日英対訳記事をデータベース化して利用している。

記事の検索には「類似表現検索」と「日英単語間の共起度を利用した表現対応推定」を用いている。前者は、図-2に例示されているように重要語を抽出し、その活用語尾展開も含めて、重要語を含む用例を検索する方法である。後者は図-3に示すように、日英単語が記事の中でペアになって共に現れた回数から共起度を計算し、文の対応を推定するもの

国連大学におけるネットワーク言語の研究開発

国連大学高等研究所 UNL プロジェクトディレクター 内田裕士

国際連合の使命の一つは、国家間の理解・平和のための調和のとれた情報交換を促進することである。この情報交換のための言語的な橋を提供することを目的に、1996年4月にUNLプロジェクトがスタートした。

UNLとは、Universal Networking Languageの略称で、United Nation Languageにも通ずる。

「エンコンバータ」と「デコンバータ」という2つのソフトウェアを用いることによって、異なる言語を使う人々のコミュニケーションを可能にする一種の中間言語である。この言語を国際連合の加盟国が共通に用いることによって、インターネットにアクセスできる人は誰でも、加盟国すべての母国語からテキストをUNLに“エンコンバート”し、同様に、どんなUNLのテキストも母国語に“デコンバート”することができる。

国連大学高等研究所内にあるUNLセンターでは、各国のパートナーと共に一般的なネットワークサーバーや、コンピュータプラットフォームと互換性のあるUNLソフトウェアを開発し、普及させようとしている。現在すでに、ブラジル、中国、エジプト、フランス、ドイツ、インド、インドネシア、イタリア、ヨルダン、スペインおよびロシアの研究機関や企業が開発に協力している。

UNLプロジェクトの開発期間は10年の予定である。最初の3年間は、国際連合の公用語であるアラビア語、中国語、英語、フランス語、ロシア語、スペイン語の6つの言語と、使用人口が一億人を越えるドイツ、ヒンズー、インドネシア、イタリア、日本、ポルトガルの6つの各言語の変換モジュールの開発に専念する。これによって、世界人口の75%~80%の人が利用可能になる。実際には、さらに3言語を加えて15言語を平行して開発する予定である。1998年には試験運用を開始し、1999年には本稼働に入る。残り7年間では、最初の3年間で開発された技術をもとに、185カ国すべての加盟国の言語モジュールの開発を行う。

UNLシステムは、いずれの国の人も母国語で他国の人々とコミュニケーションできることを目指したものであり、UNLは多言語ネットワークの中の世界共通の言語である。

UNLシステムは、ネットワークとUNL-母国語間変換プログラムから構成されている。母国語の情報をUNLに変換するシステムをエンコンバータ(enconverter)、UNLから母国語の情報を生成するシステムをデコンバータ(deconverter)と呼ぶ。情報はネットワーク上でUNLの形式でやりとりされる。

UNLの文書を作成するには、一般にUNLエディタが使用される。UNLエディタは対象言語の“エンコンバータ”と“デコンバータ”を内蔵しており、ユーザが母国語で文書を入力すると順次UNLへの変換を行い、インターラクティブにUNL文書を作り上げていく。UNLに変換された文書は、その変換の正しさを確認するために、“デコンバータ”を用いて再び母国語に変換、表示される。これによりユーザは意図した文書を正確にUNLで作成することができる。

UNLで記述された情報は、ホームページにおかれたり、ファイルの形で管理される。又ネットワークを流れwwwサーバーに格納される。ファイル転送プログラムで転送することもできる。これらのUNL文書は、デコーダの組込まれたブラウザで各国の言語に変換され、読まれることになる。UNL情報は、それがUNL情報であることを明示するタグを導入することでHTML文書の中に埋め込まれる。これによってHTML文書の利点を活かしてUNL情報を扱うことができる。

UNLは、単語の意味を示すUW(Universal Word)と、文を構成するUW間の関係を示す関係記号と、UWの意味や使われ方を限定する属性記号からなる。

UNLの語彙は、UW(Universal Word)と呼ばれる。これは各国のことばで表される意味を表現するための語彙であり、コンピュータの内部で情報を表

現するために使用されるものである。UW は基本的には英単語を用いている。英語を用いるのは、英語は現在地球上で最も広く用いられ、各言語との間で辞書が作られていてそれを利用できるメリットがあるからである。UW は、ある単語が表す意味を、その意味を示す英語の単語と、その英単語の意味を特定するための他の英単語との可能な関係を suffix として記述したものである。

語彙は、UNL システムの開発を通して膨大な量になっていくと予想される。各言語のことに含まれる意味は膨大であり、ごくわずかな意味の違いも取り上げれば無限に増大していくおそれがある。システムの運用には、ユーザは必要に応じて新たな UW を自由に登録できる必要がある。その新たな UW がシステム全体にとって有用かどうか判断することは大変で、とても人手で UW を管理することは不可能である。UNL システムでは、UW を自動管理することを目指している。ユーザは必要と思われる UW を定義することができるが、その際、UW と母国語との対応および UW の意味体系のどこに

位置づけられるものかを定義しなければならない。UNL システムはこの情報をもとに UW を自動管理し、各言語のデコンバータに適切な情報を与えて新しい UW に対応できるようにする。もちろん、登録されたあと、滅多に使用されない UW は淘汰されていく。

UNL プロジェクトに関する情報はホームページで流している。

ホームページアドレス：<<http://www.unl.ias.unu.edu>> 現在は、日本語と英語を用いているが、'97 年末には、UNL でも掲載する予定である。

問合せ先：

〒150 東京都渋谷区神宮前 5-53-67
国際連合大学 高等研究所 UNL センター
UNL プロジェクトディレクター

内田裕士

電話 03-5467-5547

FAX 03-5467-2324

E-mail uchida@unl.ias.unu.edu

事務局だより

‡ 研究会メンバーの募集 ‡

例文評価研究会は'96年度をもって終了し、'97年度新たに「ネットワーク翻訳研究会」が発足することになりました。

この研究会では、ますます盛んになっていくネットワークやパソコン通信で行われている翻訳について、評価や指針を研究します。だれもが受信者に、だれもが発信者になることができるインターネット時代における機械翻訳のありかたを追求します。

ご希望の方は、所属、電話番号、FAX 番号、E-mail アドレスを添えて、7月15日（火）までに FAX 又は E-mail で、事務局へお申し込みください。（書式は自由です）

イベント報告

（社）日本翻訳協会設立十周年記念 第8回翻訳フェア

上記フェアが'97年2月28日、東京一ツ橋の如水会館で開催された。

- 特別講演 「翻訳とインターネット」……………東大教授 石田晴久氏
- パネルディスカッション 「翻訳者はインターネットに何を期待するか」
司会 図書館情報大学教授 石川徹也氏
- セミナー 「翻訳と CALS」 他5件

懇親会および展示会も開かれ、AAMT も参加した。

対話の機械翻訳

ATR 音声翻訳通信研究所 美馬秀樹

1. はじめに

近年、インターネット上のホームページの英語で書かれた内容を閲覧する目的などで、機械翻訳システムを利用する人が増えてきました。ホームページなどのように紙面等にかかれた文書の翻訳は“書き言葉の翻訳”と呼ぶことができます。そのような書き言葉の翻訳とは別に、海外旅行におけるホテルフロントでのやりとりや、電話での外国人との会話のような、いわば対話状況にある言葉の翻訳も機械翻訳の可能性として考える必要があります。一見、そのような対話に対す翻訳も、書き言葉の翻訳と同じ技術で処理可能なように思われますが、外国文献の翻訳と、国際会議の通訳がそれぞれ異なる課題を持ち、それぞれに相応しい技術を必要とするように、機械翻訳においても、対話の翻訳では書き言葉の翻訳とは異なる課題に対する技術や、より高度な処理を必要とします。

例えば、

例(1)「支払いは銀行振り込みをお待ちしております」

のように、書いてみると表現的に少しおかしいような文でも、話している場合には自然に使われます。また、

例(2)「今日、シングル二つありますか？」

のように、助詞や、疑問の終助詞が脱落した表現も対話場面では多く現れます。このような、いわゆる表現のねじれや、書き言葉としては非文と判断される表現に対しても対話の翻訳ではうまく扱う必要があるのです。

対話の翻訳において、書き言葉の翻訳と比較して、特に重要だと思われるものを次に挙げます。

1) (文法的な処理が難しい) 話し言葉の扱い

話し言葉とは、単純には音声媒体とする言葉だと言えます。しかし、実際には、話し言葉と書き言葉とを明確に区別することは簡単にはできない場合が多いので、ここでは、日常会話のよう

な口語的性格の強い表現に現れる特有の問題を扱うための技術ということで話しを進めます。先の例にあるような文法的な処理が難しい表現や、話し言葉に特有の表現などを適切に扱うことが要求されます。

2) 対話の状況に応じた表現の扱い

状況の扱いとして、従来の文脈処理も含めて考えなければなりません。対話翻訳に特有の問題としては、冠詞の扱いのような言語的情報に基づいた訳文の生成ばかりではなく、対人関係のような非言語的要因に応じた訳文の生成も考慮する必要があります。

3) 円滑な対話の実現

対話の利点である実時間性や効率性を阻害しないためにも、人間の行う通訳のように、少なくとも発話終了とともにその訳文が提示される程度の即時性が望まれます。さらには、同時通訳に見られるような発声の途中段階でも確定した句や節の訳を出力できる処理技術が望まれます。

本稿では、対話に対する機械翻訳の実現に要求される特有の課題に対し、現在の処理技術や、取り組まれている研究を紹介するとともに、今後どのような成果が望まれているのかを考えていきたいと思えます。

2. 話し言葉の扱い

2.1 頑健な処理

話し言葉においては、自分の頭の中以外に発話内容をモニタリングする機構を持つことができないため、発話が長くなるにつれ、表現のねじれなどの現象が起こりやすくなると考えられます。先の例(1)は、ヲ格をもつ述語がさらに提題の助詞「は」を備えている文ですので、その構文構造的な解析は容易ですが、意味的には2文が融合しているものであるため、意味解析を適切に行うのは困難です。このように、

話し言葉に対しては、書き言葉の翻訳で用いられているような文法中心の処理だけでは多くの場合十分には対応しきれません。したがって、文法によらず、どのような入力に対しても翻訳システム全体として失敗しないような頑健な処理が必要となります。

文法的に説明することが難しい表現や、文法的には非文であると思われる文の扱い方としては、次のような処理方法が考えられています。

a) 意味主導の翻訳

システムが対応可能な文形や意味表現に（内部的に）変換し翻訳する方式

b) 表現主導の翻訳

文法的な難しさや誤りに対しても、表現そのものを目的言語の対応する表現に翻訳する方式
例えば、助詞が欠落した表現に対して、a) では、尤もらしい助詞を補い翻訳するといった方法を用いることを指し、b) は、助詞が欠落した表現に対応する目的言語の表現を利用して翻訳することを指します。

a) の手法では、入力文に対し、常にシステムが理解できるように、デフォルト値などを用いて意味の補足を行うため、非常に強力な頑健な解析が可能となります。しかし、場合によっては入力文には含まれない意味まで付加する可能性があることも否定できません。

一方、b) の手法では、文の意味よりも表現そのものに重点があります。助詞が省略されるのにも、ある種の意味がある（聞き手の解釈のあいまい性を増すことで間接的に丁寧さを与える）のような考えが基本にあるため、状況を含めたニュアンスのようなものまで伝えられる可能性があります。異なる言語間では、文化の違いなどもあるため、すべての言語現象について対応する表現が容易に見つかるとは限らないという問題もありますが、大規模な対訳コーパスが利用可能となり、表現対のような経験的知識の獲得が容易になれば、このような問題も徐々に解決していくと考えられます。これについては2.3項でさらに説明します。

2.2 決定的処理

音声対話の翻訳では、音声入力に始まり、音声による出力で訳を伝えるのが処理の基本的な流れです。

また、常に実時間性が要求されますから、前編集や後編集のような人間による支援の要素が入る余地はありません。したがって、入力発話の音響的な尤もらしさや、言語的な妥当性を最大限活用し、理解度の高い訳文の生成、あいまい性解消のための決定的な処理が期待されます。

2.3 話し言葉を扱う技術

我々人間が外国語を習熟しようとする場合、文法的な知識を学習することももちろん大切ですが、それだけではなく、実際に使用されている言語表現にできるだけ多く接し、基本表現や頻出表現を大量に学習することも必要となります。特に会話場面においては、経験的知識として蓄積した基本表現や頻出表現を活用することが非常に多いと思われます。近年、大規模コーパスの利用が可能になり、機械翻訳においても、用例に基づく手法、統計的手法など経験的知識に着目した手法が盛んに研究されています。文法に基づく枠組みの問題点等については、本誌第16号の技術早わかりの記事で詳しく述べられています。それに対して、経験的知識に基づく枠組みでは、

- ・実際の言語表現から獲得した経験的知識を活用できるため文法的に説明が難しい表現を扱える

→頑健性の確保

- ・基本的に、用例に対し単純な類似性の計算だけで翻訳処理が行えるため高速な処理が可能

→決定性の確保

- ・知識の記述や追加が容易

→拡張性、知識管理の容易性の確保

など、話し言葉の翻訳に対して多くの利点があります。

●経験的知識に基づくアプローチ

このような枠組みの一つとして、コーパスなどから収集した原言語表現とその翻訳例との対応から、「～をお願いしたいんですが」のような頻出表現やフレーズ、慣用句等の意味的にまとまった単位で対訳用例を収集し、このような表現パターンを組み合わせることで文全体の翻訳を行う方式が考えられてい

ます。また、同じ表現パターンに対して複数の訳し方がある場合にも、入力文と用例がどれくらい似ているかを判断し最も類似したものの対訳表現を使って翻訳結果を作り出すといった、ある意味で、人間が経験的に妥当性を判断し利用するような枠組みに類似した処理方式も考えられています。

例として、「京都のホテル」のような「XのY」の表現パターンに対する英訳を考えてみます。「XのY」という表現は日本語では非常に頻度の高いパターンであるとともに、この表現により多くの意味内容を表すことができる（つまり、あいまい性が多い）のですが、それ故に、英語のような比較的リジッドな言語に翻訳する場合にはXとYの部分の語やフレーズの意味内容により複数の訳し分け候補が存在します。“~of ~”や“~for ~”の表現を使ったり、「江戸時代の面影」では“the image during the Edo Period”のように“during”を使って訳すことも考えられ、機械翻訳で適切な翻訳を行うのは非常に難しい問題とされています。これに対し、用例を基にした枠組みでは、「ホテルの住所」に対しては“the address of the hotel”、「英語のパンフレット」に対しては“a pamphlet in English”などのように翻訳された対訳用例を知識として持っていれば、「旅館の住所」の入力に対しては“the address of the hotel”の表現例を修正して、“the address of the inn”と訳したり、「日本語のパンフレット」に対しては、同様に“a pamphlet in Japanese”のように訳し分けることができます。また、「Xをお願いしたいんですが」のようなパターンに対しても、

X' please (Xが“人数”など)、
could you give me X' (Xが“名前”など)、
I would like X' (Xが“パンフレット”など)、

:

のように単に“please”だけで用が足りる場合もあれば、“I would like ~”のような表現を用いるべき場合もあります。このような用例を多く集めることで、目的語の意味内容により、ある程度妥当な表現へ訳

¹ 現在、ある表現に対し、類似した用例を検索する方式として、ソーラスや類語辞典を用い、語と語の意味的な近さを基に表現間の類似度を計算するものや、統語的構造の類似度計算を基にする方法、また、頻度などの統計的情報を用いて単語の多義性を解消した上で、類似した表現を検索する方法、などが考えられています。

し分けることが可能となります。

さらに、

例(3)「日本語のパンフレットをお願いしたいんですが」

のような文に対する英訳においては、先の例にある「XのY」、「Xをお願いしたいんですが」のパターンから構成される、

((XのY)をお願いしたいんですが)

のような言語構造に対し、各パターンにおける妥当な対訳表現を組み合わせて文を生成することで、

“I would like a pamphlet in Japanese.”

のような翻訳結果を得ることができます。

●話し言葉に特有な表現の扱い

このような方式では、2.1項で説明したような表現主導の翻訳が処理の基本と考えられますから様々な言語現象に対応できる可能性があります。また、文法的には説明が難しかったり、処理を失敗させないためには事細かく情報を記述してやる必要があった文に対しても比較的簡単に処理できることが考えられます。例えば、

例(4)「安くて清潔な政治ならいいんですが」

のような換喩的な表現に対しては、単純に文法の枠組のなかで処理をしようとする、“政治”という語に対して、特殊な用法のような形で、“安い”とか“清潔”などの語との関連をあらかじめ記述しておく必要が生じます。

一方、用例に基づく枠組みでは、「安くて清潔なホテル」等のよく似た表現とその対訳例“inexpensive and clean hotel”をうまく利用することによって、少し苦しい訳かも知れませんが、

“Inexpensive and clean politics would be fine.”

のような翻訳を得ることができます。

このように、用例に基づく枠組みは、我々が学校の英作文の授業で習った“よく似た表現例をまねて新しい文を作る”という考えに似ていると言えます。それ故に、入力文の詳細な構造や意味記述などが得られなくとも、翻訳処理全体として失敗しないような頑健性を持つと考えられます。

実際に、日韓翻訳に対し、無作為に選んだ初見の文を対象に行った実験では、69対話(延べ1300文)の内、90%以上の翻訳結果が正解となりましたが、その内の約20%は言語の構造的な解析が成功してい

なかったのにもかかわらず翻訳処理全体としてはリカバーされていた、という頑健性を示す結果を得ました。

その他にも、終助詞「よ」「ね」「な」を使うなどの、書き言葉にはほとんど用いられない表現等に対しても適切な目的言語への翻訳や、妥当な訳し分けが期待されます。コーパスから得られる用例や表現の分析が進み用例などの知識が充実するにしたがい、多くの問題が解決すると考えられます。

●経験的知識と文法的処理との協調

ここまでの話からすると、用例を利用すれば全く文法が必要なくなるような印象を持たれたかもしれませんが、しかし、より実用的なレベルまで突き詰めて考えると、本誌第16号の技術早わかりの記事でも述べられているように、たとえ用例を基にした手法と言えども全く文法的処理が必要なくなるわけではないのも確かです。先の例の「～をお願いしたいんですが」のようなボタンにしても“-”の場所にどのような表現が現れてもよいわけではありません。また、いくら各々のフレーズの翻訳が成功したとしても、それぞれのフレーズを文法的にうまく組み合わせなければ自然な翻訳文は得られません。したがって、実用的なシステムでは、用例に基づく枠組みと文法に基づく枠組みとを協調的に融合させることが重要だと考えられます。現在、このような協調的枠組を利用した機械翻訳のプロトタイプシステムが構築され、旅行会社を対象とした日英間、日韓間、日独間の翻訳実験が行なわれています。評価実験において、良好な翻訳正解率を得た結果も報告されています。

3. 状況に応じた表現の扱い

翻訳を行う際、人間の場合は、ごく普通に文脈から得られる手がかりを使って文の解釈を行っていると考えられます。さらに、対話場面においては、人間は文脈などから得られる言語的手がかりだけではなく、自分の置かれた状況や社会的役割、対話相手との社会的関係などの非言語的な情報も手がかりとして、文の解釈や対話相手に対する言葉の選択などを行っていると思われます。敬語などの待遇表現の扱いはこれを象徴する例でしょう。したがって、対話

を機械翻訳で扱う場合には、状況における非言語的な情報に対しても、うまく獲得し利用することで、文の解釈に役立てたり、状況に応じた適切な表現を使って訳し分けるといった処理も考える必要があります。

3.1 社会的役割と対話場面

海外旅行先のホテルでの対話場面においては、食事をしたり、旅行を目的として飛行機などの交通手段を利用したりするのは、多くの場合旅行者です。したがって、例えば、“eat”の日本語訳として、

“召し上がる”(話者がフロント)

“いただく”(話者が旅行者)

などと訳し分けることが考えられます。

また、話し手や聞き手の社会的役割、対話場面の情報がうまく利用できれば、日本語では省略されている主語を補うことができる可能性もあります。例えば、

例(5)「もしお召し上がるのでしたらすぐ作ります」

とレストランのコックがしゃべったのなら、「お召し上がる」の主語には“you”を、「作る」の主語には“I”を補って英訳するのが妥当だと判断できます。

3.2 社会的地位

日本語のような待遇表現が重視される言語では、話し手の社会的地位にふさわしい言葉づかいや、聞き手との社会的関係を考慮した翻訳が期待されます。特に英日翻訳などでは、英語においては社会的地位の差による表現の違いは非常に少ないため、聞き手と話し手の社会的関係をシステム側で考慮し、適切な訳語を選択する必要があります。例えば、“Will you help me?”のような文を日本語に翻訳する場合、

「手伝ってもらえる?」(聞き手が友人など)

「手伝っていただけるとありがたいのですが」(聞き手が会社の上司など)

と訳し分けるのが妥当だと思われます。

3.3 韻律(発声の高低・強弱など)の情報

非言語的な要因として、話し手の発話の韻律などの音声的な特徴も考慮せざるを得ません。

例えば、例(2)のような文では、文末を上がり調子にして発声すると質問になりますし、下がり調子な

ら話者側の状況を述べていることになります。

このように、音声的な情報を考慮することで、適切な意図を持つ表現へと翻訳する必要があります。

3.4 状況に応じた表現を扱う技術

このような状況に関する情報を活用するためには、
・辞書に対する、状況に応じた訳語（言葉使い）の情報の記述

を行い状況パラメータなどの情報により、適宜、適切な言葉使いを選択することが考えられます。さらに、

- ・状況別の表現や言葉使いの統計的情報
 - ・話し手、聞き手の情報を含めた用例
- などの経験的知識を基に、妥当な表現や訳語の選択を自動的に行う枠組みも考えられます。

また、状況に関する情報をどのように効率的に獲得するかという問題も重要です。これに対しては、次のことが考えられます。

1) システムの要求項目として、事前に設定

ホテルでの応答システムのように、ある程度固定的な使用環境に置かれるもの等が対象。運用開始時にプッシュボタン等で設定しておくほうが効率的だと考えられる。→社会的役割や対話場面

2) 統計的情報などから自動的に推定

- ・状況を表すパラメータ
- ・話者などの情報が付加されたコーパス

などが整備されれば、状況と関連した形での言葉の頻度情報が得られるため、対話中よりある程度話者の特徴を自動的に推定することが可能。

→性別や社会的地位など

さらに、韻律などの情報を活用するためには、音響的特徴の情報も把握できるコーパス等の整備も期待されます。いずれにしろ、状況の情報を効率的に獲得することは今後の重要な研究課題と言えます。

4. 円滑な対話の実現

翻訳を交えた対話においては、可能な限りの即時的な翻訳が期待されます。少なくとも、人間の行う通訳のように、発話終了とともにその訳文が提示され

る程度の実時間性が望まれます。人間の行う翻訳の場合には、たとえ逐次通訳であっても、話し手が発話するのと同時にその内容の理解や翻訳がある程度進行しているものと思われます。長文の場合には、適宜、通訳メモをとったりすることにより、少なくとも原言語の内容の理解は同時に進行させていると考えてよいでしょう。しかしながら、現在の対話翻訳システムのほとんどは、図1-(a)にあるように、一文毎でかつ発話終了後に処理を開始するのが通常です。また、現在のコンピュータの処理能力をもってしても、音声認識等を含む翻訳処理に要する時間を無視できるレベルにすることは難しいのが現状です。したがって、特に一文が長くなった場合などには、話し手の話している時間と翻訳処理に要する時間の間、聞き手はずっと待たされることになるわけですから、対話としての自然さを著しく損ねてしまいます。

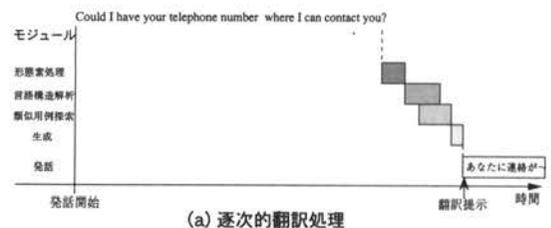


図1. 翻訳処理における各モジュールの処理タイミング例

4.1 漸次的翻訳処理

上記の問題に対して、話し手の発話と同時に翻訳処理を進行させ、発話終了と同時に翻訳結果を提示することが考えられます。このような処理方式にすることで、あいまい性がある発話途中での句や節の確定が十分ではなかった場合でも、少なくとも発話終了時には即座に翻訳結果の提示が開始できると考えられます。さらには、講演などのように話し手が連続的に発話する場合や瞬間性を要求されるような対話場面では、図1-(b)のように、発声の途中段

階でも確定した句や節の訳を次々と出力するような同時通訳的な処理方式が望まれます。このような技術が実現できれば、円滑な対話翻訳の実現に向けた一つの大きな解決策になるでしょう。

以下では、このような翻訳を実現する「漸次的翻訳処理」について、その基本的な要素技術を説明します。

1) 確定する句や節の情報単位

翻訳処理では、一般に発話の単位となる部分構造を大きくとればとるほど正確な翻訳結果を得る可能性が高くなりますが、逆に円滑な対話の実現という面では不利になります。したがって、どの程度の情報単位をもって句や節を確定し発話を行うか、という発話のタイミングが重要な問題となります。この問題は、原言語と目的言語との言語的距離にも関わるものであると考えられます。例えば、英日間の翻訳では、述語位置や修飾方向などの言語構造がかなり変化しますので、比較的大きな単位でしか確定できませんが、韓日間では、語順等にかなりの類似性がありますので、場合によっては単語単位で次々と翻訳を提示できる可能性もあります。したがって、

- ・意味的にまとまった単位での言語パターン対の利用
「I would like～」、「～をお願いします」のような比較的長単位で、かつ常に表層語句を含んだ形の言語対の利用→作られる言語構造の階層が少なく、情報単位の効率的な切り出しが可能。
- ・目的言語との言語的距離に相応しい情報単位の設定

それぞれの言語対に応じた単位を翻訳処理毎に設定→目的言語にふさわしい原言語解析の実現などにより部分構造の基本単位を設定することで、効率的な漸次処理が可能になると考えられます。

2) 句や節に対する翻訳処理の決定性

途中の句や節がうまく切り出せたととしても、それに対する翻訳のあいまい性が同時に解消できなければ、少なくとも同時通訳の処理の実現には至らないこととなります。この問題に対しては、

- (i) 句や節を構成する表現の依存関係を先の情報単位に基づきボトムアップに解析
- (ii) (i)の解析と同時に対訳用例を適用
- (iii) 統計的情報から尤もらしい訳語や表現を確定

などの処理で部分構造に対しても解析の進行と同時に、決定的な翻訳が可能になると考えられます。

現状では、このような同時通訳にふさわしい翻訳の単位の情報や、用例などもシステムに学習させる必要がありますが、今後、大規模な同時通訳コーパス等が整備されれば、自動的に抽出し、最適化することも可能になると考えられます。

4.2 同時通訳システムの実現に向けて

先にも述べたように、日本語と英語のような語順や構造が大きく異なる言語間では、入力される単語やフレーズの列に従って、頭ごなしに訳語を出力していくだけでは入力文に忠実な訳を生成できるとは限りません。例えば、日英翻訳では、述語表現は文末まで決定できない場合が多いため、通常は、文末まで翻訳が確定できないこととなります。これに対して、人間の通訳者では、能動態から受動態への訳し変えや、長い疑問文の付加疑問文への変形、話題化変形というような、同時通訳におけるノウハウや、文脈やドメインからの推定による先読み等のテクニックを活用し、巧妙に翻訳を行っているようです。しかしながら、機械翻訳システムでは、まず、文脈等の情報を取り扱うのには限界があります。知識や論理によりゴールの推定を行うようなプロトタイプモデルによる会話の先読みの研究は行われてはいますが、一般的な枠組みまでは扱いきれていないのが現状です。また、同時通訳システムでは、出力は音声なのが通常ですから、既に発声された内容は物理的に消去することや、上書きすることは不可能です。したがって、既発話の内容に誤りがあった場合や、次に発話しようとする内容との間に意味的な不整合が生じた場合などには、前後のつじつまを合わすなどの処理により速やかに文全体としての発話内容を調整する仕組みも必要となります。

このような発話内容の調整に関しても、同時通訳の事例やコーパスなどを研究することにより、態の変化などを行うための言語パターン対を学習させたり、同時通訳特有の用例を集めたりすることで、より人間のやり方に近い処理やつじつまの合わせ方などが実現できると考えられます。

例えば、

例(6)「何も問題はありませぬ— (pause) —でし

ようか?」

のような発話に対しては、通常では“Is everything alright?”と翻訳すべきですが、ポーズの途中で、“Everything is alright”と翻訳結果を出力してしまう場合も考えられます。このような場合でも、

(Xでしょうか? => X', isn't it?)

のような同時通訳に都合のよい対訳パターンを使って付加疑問文へ変形することで、言い直しを行わずに、文全体としての翻訳内容を調整することができます。

さらに、今後、文脈や状況に対する処理が確立され、次発話の内容等に対する強力な推論が実現可能となれば、より迅速で、誤りの少ない同時通訳処理が可能になると考えられます。

5. まとめ

対話の翻訳において、通常の書き言葉の翻訳とは違った難しさに対し要求される技術を中心に紹介してきました。現状では、まだまだ人間と同じような自然な会話ができるまでには至りませんが、一つの有用な方法として、コーパスから得られるような用例や言語パターンなどの経験的知識を活用することで、より人間に近い自然な表現への翻訳を実現する可能性があることを分かっていたかと思います。

また、文脈依存の訳し分け等の現状の機械翻訳では難しい問題に対しても、文脈や状況に依存した訳し分けに必要な情報の分析を詳細に行い、文脈パラメータなどを新たに設定して注釈付対訳用例データベース等を構築し、利用できる枠組みを整えることで、多くの問題が解決できると考えています。

事務局だより

委員交替のご連絡

'97年度のスタートにあたり、委員の交替がありました。

1. 運営委員会	沖電気工業株式会社	退任	羽下雄之輔
		後任	大宅伊久雄 研究開発本部 メディアネットワーク研究所長
	松下電器産業株式会社	退任	吉原通年
		後任	高木友博 東京支社 技術渉外部 副参事
2. 市場動向調査委員会	沖ソフトウェア株式会社	退任	石川友隆
		後任	中村 裕 開発第3部第3グループ係長
3. 編集委員会	シャープ株式会社	退任	福持陽士
		後任	奥西稔幸 情報商品開発研究所 主任
	日本電気株式会社	退任	亀井真一郎
		後任	田村真子 情報メディア研究所 音声言語研究部

事務局便り

MT Summit VI 開催近づく!

すでにご承知のとおり、'97年10月29日から11月1日まで、米国の San Diego で MT Summit が開かれます。参加申込の受付が始まりました。詳細は、MT News International No. 16をご覧ください。

また、今後ホームページで情報が流されると思います。ご覧ください。

AMTA <<http://www.isi.edu/natural-language/AMTA.html>>

本年は機械翻訳の50周年です。1947年3月4日に、Warren Weaver (Director of the Natural Science Division of the Rockefeller Foundation) が Norbert Wiener に、コンピューターによる翻訳の可能性について手紙を送ったのが機械翻訳のはじまりとされています。この意義深い年の Summit に多数ご参加ください。

慶応義塾大学環境情報学部・政策メディア研究科 石崎研究室

慶応義塾大学環境情報学部 教授 石崎 俊

1 はじめに

慶応義塾大学湘南藤沢キャンパス（SFC）は、神奈川県藤沢市に1990年に設立された比較的新しいキャンパスです。ここには、環境情報学部と総合政策学部があり、さらに1994年にスタートした大学院の政策メディア研究科があります。両学部ともに1学科構成（学部名と同じ）のため通常は学科名を省いています。

私は1992年4月に通産省工業技術院電子技術総合研究所から移りました。当時は最上級生が学部の3年生でしたので、研究室の教育と研究の体制を一から作ることになりました。ラッキーだったのは1993年度から始まった文部省科研費の対話に関する重点領域に参加させていただいたことです。コンピュータ設備や学生へのアルバイト謝金などを賄うと同時に、毎年数回ある研究集会で関連する最新の研究成果を聞くことができました。

また、赴任当初から、学内で先生たちの交流を図るためにメタファー研究会が作られ、哲学、社会学、言語学、心理学、それに情報科学などの先生たちが多数集まって、メタファーを多角的に議論する機会がありました。従来の自然言語処理によるアプローチが、ある意味で一面的であり、学際的に考えると大変興味深い問題を多く含んでいることに気づくことができました。

1994年からはアメリカのノースウエスタン大学のアナロジーで有名なгентナー教授の元で学位を取って来た今井むつみさんが助手として一緒に研究できるようになり、認知心理学と自然言語処理や学習・発達に関する共同研究が始まりました。今井さんは今年の4月から専任講師に昇任しています。

また、1996年の1月からは、NEDOの提案公募型研究開発の資金が使えるようになり、「自然言語を中心にしたマルチモーダル情報の統合」を推進するための研究設備が格段に整備されました。後述するような最新の画像処理用のコンピュータや音声処理用のコンピュータと入出力設備などが研究室で

使えるようになりました。

また、このキャンパスでは研究コンソーシアムという制度があって、大学の先生たちでグループを組織して研究テーマの提案を行い、それに興味を寄せる会社が参加して共同研究を行います。我々のグループも、言語学と社会学の先生を加えて「対話型知識ベース」に関するコンソーシアムを1994年度から3年計画で実施しており、1997年度から2期目に入る予定です。

現在の研究体制は、私と今井さんを中心に、大学院生が博士課程2年生を筆頭に10名くらいです。学部の3年生と4年生がゼミ（ここでは研究会と称しています）に参加しますが、3年生は今井さんと共同で運営して10数名、4年生は私が単独で運営して15名くらいです。

ゼミには、当キャンパスのどちらの学部からも参加できます。また、理工学部と違って物理や化学などの技術的な教育科目がないので、技術系といっても情報処理に特化されています。また、私たちのゼミに参加する学生の半数くらいは技術系ですが、残りはいわゆる文系の学生です。したがって、コンピュータと心理学、言語学などを統合した学際的な研究が中心テーマになっています。

このような経緯でスタートした研究室は、まだかなり新しいものですが、下記のようなテーマについて研究を進めていて、最近は少しずつ研究成果が始めています。論文発表が定期的に行えるようになり、動くシステムもできてきました。

2 研究テーマと研究の概要

2-1 自然言語処理

1980年頃に米国イェール大学に客員フェローとして長期滞在して以来、自然言語処理の研究を続けています。慶応義塾大学に来てからは、応用システムというよりも認知科学的に人間の言語情報処理機能を探求し、そのモデル化に基づいた自然言語処理システムの構築を目指しています。このような方向

性を選んだ理由としては所属する学部の性格もありますが、むしろ、最近の脳科学の目覚ましい進歩に触発されたことのほうが大きいでしょう。

従来の自然言語処理の研究が、実用化の直前で足踏みしているような印象を受けますが、この根源には意外に深いものがある、人間の言語機能に関する基礎研究が重要かも知れないと思っています。文脈を扱うことの重要性には誰も異存はないと思いますが、従来のアプローチでは汎用の文脈処理は難しいかも知れません。人間が幼児のころから実行しているマルチモーダルなアプローチがコンピュータで実現できるかどうか興味深いところです。

(1) メタファーと多義語の意味変動モデル

「AはBだ」というメタファー表現において、AとBに共通な顕著な属性を抽出するメカニズムをニューラルネットワークで実現する研究を進めています。また、動詞とその目的語の組み合わせに関して、それぞれの属性が動詞の意味の変動に与える影響を、多くの被験者による評定実験を行って調べています。目的語の名詞の属性の記述法や、具体名詞と抽象名詞の場合の違いなどを重回帰分析で調べています。

(文献)

今井豊、石崎俊、「属性情報を用いた比喻表現における概念の類似性の発見手法」情報処理学会自然言語処理研究会、116-12 (1996-11)

岩田崇、今井むつみ、石崎俊、「多義動詞の意味変動の定量的モデル」電子情報通信学会言語理解とコミュニケーション研究会、NLC 96-58 (1997-03)

(2) 連想実験と概念辞書

概念の体系と概念のもつ属性、概念に関係する動作とその動作が生じる環境(状況)などを連想実験を行って多量に収集し、そのデータから概念辞書を作成することを試みています。連想実験では、連想頻度、連想順位などの情報が使えるので、刺激概念から連想概念までの距離を定量的に表現し、それを自然言語理解や情報検索に使用することを目指しています。

(文献)

岡本潤、内山清子、石崎俊、「オンライン連想実

験システムと学習基本語彙の概念辞書化」情報処理学会自然言語処理研究会、118-18(1997-03)

(3) 感情解析

感情に関する研究は心理学的なアプローチは多いけれども、言語データを集める困難さなどのために、自然言語処理からのアプローチはほとんど見られない状況でした。しかし、我々のキャンパスでは学生同士の連絡に、電話代わりに電子メールを駆使している、感情表現を含む多量の電子メール文を収集することができました。

そこで、それらのデータに基づいて、感情の分類をおこない、自然言語処理の一環として感情解析を行うシステムを試作しました。

(文献)

井上みづほ、藤巻美菜子、石崎俊、「電子メール文における感情表現の解析システムについて——感情表現の収集、分類、解析——」電子情報通信学会思考と言語研究会、96-608(1997-03)

(4) 機械翻訳

NTTのCS研との共同研究で、英単語と日本語の連語の対応づけなどの基礎的なレベルの研究を進めています。

2-2 音声情報処理

私は自然言語処理の研究を始める前は、音声認識の研究を行っていました。学位論文も音声情報処理でしたが、昨年にNEDOの資金で音声関係の装置が購入できたので、ようやく音声情報処理の研究をはじめることができて、主に音声合成の研究から始めています。

(1) 音声合成

疑問文における文末のピッチパターンの制御法を工夫して、規則合成方式で自然な疑問文の音声合成を研究しました。最近では感情を音声で表現する方法の基礎的な研究を進めています。

2-3 画像情報処理

画像の研究も音声と同じく、先端の研究をするためには高額の装置が必要です。NEDOの予算でINDIGO2の最新機種や周辺機器、関連ソフトをそ

ろえて研究を始めています。

(1) 仮想空間と空間指示語

3次元の仮想空間中で前後左右という4つの空間指示語が表す領域を、参照物体自身が向きをもつ場合とまたない場合に分けて認知実験を行い、空間指示の曖昧性を解消するための方法について検討しています。また、空間指示語が指し示す領域の定量的な同定も、同様な3次元仮想空間中で実験し、人間の空間指示モデルの構築に向けて研究を進めています。

(文献)

中西卓哉、今井むつみ、石崎俊、「仮想空間における空間指示の研究」電子情報通信学会言語理解とコミュニケーション研究会、NLC 96-60 (1997-03)

(2) 顔画像の合成と認識

マルチモーダルなインタフェースとして、コンピュータディスプレイ上で表情をもつ顔を合成したり、ディスプレイの前に座っている人間の顔を認識して人物を同定する研究を、電総研と共同で進めています。自然言語に含まれる意味内容や、音声のもつ言語外情報をこれらの画像情報と統合することによって、マルチモーダル対話システムを構築しようとする試みです。

2-4 マルチモーダル情報の統合

前項の顔画像の合成と認識もマルチモーダル情報の統合を目指した研究ですが、そのような研究の基礎的なデータ収集として、対話におけるマルチモーダルコーパスを作成するための研究を進めています。これは、(社)日本電子工業振興協会の対話処理技術専門委員会の調査研究活動に対応していて、世界的なレベルでの対話コーパスの標準化を推進すると同時に、マルチモーダル性を対話にもたせるための研究です。

3 おわりに

当研究室では、自然言語処理研究の中でも認知科学的な側面や、脳科学的なアプローチも取り入れて基礎的な研究を中心に進めています。

上で挙げた研究のほかにもいくつかの研究テーマを進めています。まだ基礎的段階でこれから発展させる必要のあるものが多い状況です。

当協会の市場動向調査委員会の委員長などの役職をしばらく前から勤めています。PC上のMTシステムでインターネット情報を和訳して概要を把握することができるようになって来ています。人間とコンピュータの対話や情報検索などを含めて、近い将来に我々の研究成果が機械翻訳システムに役立つことを期待しています。

(ishizaki@sfc.keio.ac.jp)

協会便り

第7回通常総会開催案内

日 時 1997年6月23日 午後4時～5時
場 所 機械振興会館 地下大ホール
《東京都港区芝3-5-8 ☎ 03-3434-8211 (代表)》

議 題 第1号議案 '96年度事業報告
第2号議案 '96年度決算報告
第3号議案 '97年度事業計画案
第4号議案 '97年度予算案
第5号議案 その他、会員提案事項

日本語の勉強と翻訳

九州工業大学情報工学部中国研究員 周玉龍

私は中国政府派遣研究員として半年前に日本の九州工業大学情報工学部の野村自然言語処理研究室に入りました。野村研究室は近年自然言語処理技術を応用していろいろな研究課題を進んでいきました。

実はコンピューターで自然言語を処理するのは非常に難しいことだと思っています。コンピューターどころか人間さえ自然言語をよく理解して上手に応用できるのもやはり難しいと思っています。

たとえばある自然言語で書いた文章をはかの自然言語で書いた文章に翻訳することは自然言語処理分野の中の一つの重要な領域である。人間は何年間か外国語を勉強してもその外国語を母国語に上手に翻訳できない人は少なくないでしょう。

これから、日本語の勉強と翻訳（人間の勉強と翻訳及びコンピューターでの翻訳）について自分の感想を書きます。実は私は十年前から日本語を勉強し始めましたがなかなか上手にできませんでした。

中国人にとって日本語の勉強は
ちょっと易しいと思う

私は中国人にとって日本語を勉強するには別国の中国語の出来ない人より易しいと思います。なぜなら、一方中国人が漢字の書き方はよくできたからです。この点だけでたくさんの時間が省けるのは当たり前のことです。中国人にとって漢字がすぐ書けるだけではなく、他方のもっと優れた利点があると思います。それは日本語漢字及び漢字からなった単語の意味です。実は半分以上の日本語漢字及び漢字からなった単語の意味は中国語の同じ形のと殆んど同じです。そういう利点があって中国人は日本語の単語を覚えるのに速いのです。

中国人は日本語を勉強する時に利点があるにもかかわらず欠点もあります。たとえば、以下の漢字と単語が日本語と中国語の両方にもあります。しかし両方の意味は全然違います。中国人にとってそのような単語を見てから、すぐあたまの中に反映してきたのは

もちろん中国語単語の意味で、間違えることは発生しやすいはずですが。そのような単語はあまり少ないのです。

日本語単語	中国語単語及び意味
娘	娘：母親
床	床：ベッド
手紙	手紙：トイレトペーパー
春雨	春雨：春に降った雨
麻雀	麻雀：スズメ（小鳥）
丈夫	丈夫：夫、主人
勉強	勉強：強制する、やっつことである
文句	文句：文章の語句
怪我	怪我：私を責めるべきだ
馬鹿	馬鹿：アカシカ（動物）

以上の単語は形が同じで、しかし意味は全く違います。実はもっと多いのは一部分の意味が同じで、他の部分は違います。このような単語を勉強するのに力を入れなくてははいけません。

日本語の助詞と助動詞は難しい

日本語を勉強することについて私にとってもっと難しいのは日本語の中に特有な助詞と助動詞の使い方です。助詞と助動詞は日本語の中で一番重要なものだと思って、それを使って単語と単語、文節と文節の間を関係を決められます。即ち、助詞と助動詞は日本語センテンスを作るのに「粘着剤」の役割をしています。普通の単語は一つ一つの石ころのようで、もし助詞と助動詞のような「粘着剤」がなければ一緒のことに粘着できないのです。

中国語の“我吃饭”という文は日本語で“私はご飯を食べる。”という意味です。

その中に「は」という助詞は「食べる」という動

作をする動作主が誰かを提示しています。「を」という助詞は「食べる」という動作の対象が何かの印です。

しかしながら、常用助詞と常用助動詞の別々の使い方は十数種類か数十種類もありますから、センテンスの中の各助詞と各助動詞の役割がどのように判断するかそれはセンテンスを正しく理解するのに一番重要なものです。

たとえば、助詞「に」の使い方はほんとうに難しいと思っています。

山に住む。
五時に出かける。
福岡駅に着く。
大人になる。
映画を見に行く。
親に似ている。
先生は学生に作文を書かせる。
先生に叱られる。
.....

場所、時間、目的、状態変化、比較の基準、使役の対象などいろいろな機能があって、日本語助詞の重要性がこれから見られます。

各助動詞も上手に使えるのはやはり難しいです。助動詞もセンテンスの中に重要な役割をしています。日本語の中に動詞のような用言部分は非常に重要な部分で、なにををするという動作を指定します。助動詞は主に用言の後に繋がって用言部分の本来の意味を変えたり新たな意味を追加したりします。

たとえば、打ち消しの助動詞の「ない」、希望の助動詞の「たい」、推量の助動詞の「よう」、伝聞の助動詞の「そうだ」、使役の助動詞の「せる(させる)」、受身の助動詞の「れる(られる)」、可能の助動詞の「れる(られる)」、自発の助動詞の「れる(られる)」、尊敬の助動詞の「れる(られる)」など、たくさんありますね。その中に、たとえば、なん種類の助動詞の「れる(られる)」は出てきました。それは上手に身につけて正しく応用できるのは確かに困難なことです。実は助詞と助動詞が上手に使えるかどうかは日本語のレベルが高く達するかどうかの試金石だと思います。

普通語、丁寧語、謙譲語、尊敬語、

同じ意味の別々の表現方式

中国語の中で断定の文は“.....是.....”の一つの形式で表現できますが日本語の中で「だ体」、「です体」、「である体」、「でございます体」などがあります。それに日本語の中に謙譲語とか尊敬語とかありますから、同じ意味のことばで別々の表現方式があつてうまく使えるのは確かに難しいです。

日本は礼儀正しい国ですから日本語の中にこのような礼儀待遇を正しく反映しているのは当たり前のことだと思います。

実は日本語を話す前に話し手はまず聞き手に対する関係を把握しなければなりません。話し手は、常に、相手は自分にとって、親しい身内の人間か、あまり親しくないよその人間か、上司か、同僚か、部下か、といたような人間関係を判断してから表現方式を決めてきます。

たとえば、中国語の“这是什么？那是词典”とは日本語で下の別々の表現方式があります。

- | | |
|-----------------|--------------|
| (1)これは何か。 | それは辞書だ。 |
| (2)これは何ですか。 | それは辞書です。 |
| (3)これは何であるか。 | それは辞書である。 |
| (4)これは何でございますか。 | それは辞書でございます。 |
| (5)これは何ですか。 | それは辞書でございます。 |
| (6)これは何でございますか。 | それは辞書だ。 |

中国語の“读过这本书吗？不，还没读过”とは日本語で表現方式はなん種類もあります。

- | | |
|----------------------|----------------|
| (1)この本、もう読んだか。 | いいえ、まだ読んでいない。 |
| (2)この本、もう読みましたか。 | いいえ、まだ読んでいません。 |
| (3)この本、もうお読みになりましたか。 | いいえ、まだ読んでいません。 |

上の例のような表現方式の多様であるセンテンスは翻訳される時に面倒くさいと思います。もし中国のセンテンスから日本語へ翻訳すれば1文対多文ですからどのように選ぶか問題が生じます。反対に中国語に翻訳すればもし同じな訳文にすると確かに不十分だと思っていて、日本語の中の人間関係の礼儀待遇の情報がなくなりましたからです。

一番重要な用言部分はセンテンスの最後に置く

日本語のもう一つの特徴は動詞ぐらいの一番重要な用言部分はいつでもセンテンスの最後に置くのです。この語順の点について日本語は中国語とも英語とも違います。それに最後の用言部分は常に補助動詞などを繋いで授受関係を表します。その授受表現や恩恵の授受の認定は中国人にとってもう一つの非常に困難な問題になります。

日本語のセンテンスを読むあるいは聴く時、最後まで進まない文の中の格関係は全然決められません。たとえば、「私は彼に本を送って、. . .」という文を聴くと本は私から彼に送るのか彼から私に送るのが全然決められません。もし最後に「送ってあげました」だったら私から彼に送るので、最後に「送ってもらいました」ならば彼から私に送るという意味です。

“時計を買ってやった。”とか“時計を買っていたきました。”とか中国人にとって同じような理解の難しいセンテンスです。それはもちろん補助動詞の理解と使用に出てきた問題です。

「. . . てやる」「. . . てあげる」「. . . てさしあげる」「. . . くださる」「. . . てくれる」「. . . てもらう」「. . . ていただく」などのような補助動詞は強い方向性がある人間礼儀待遇表現も含んでいます。しかし中国語の中にそのような表現はしな

いです。補助動詞を一つだけ使ってもかなり複雑になるのに、それにこれらの語の敬語の問題が加わり、さらにこれらの補助動詞が二つ、三つ重ねて用いられて、「連れて行ってあげてください」とか、「聞いても

らってやってくれませんか」とかになるとそれはもっと難しいでしょう。私にとって今でもそのようなセンテンスはよく理解できません。

機械翻訳はもっと難しいはずだ

人間が別国の言語を勉強し理解してから母国語に翻訳するのは上に述べたようないろいろな困難に遭いました。コンピューターは人間よりどうでしょうか。実はもっと難しいはずです。

私は日本に来る前に中国の南開大学のコンピューター学部で研究グループと一緒に簡単な試験用の「日中機械翻訳システム」を作りました。つまりコンピューターで自動的に簡単な日本語のセンテンスを中国語に翻訳できたシステムです。

機械翻訳は人間翻訳とちょっと違いますけれども、翻訳の手順と翻訳に遭った主な問題は人間翻訳と大体同じだと思っています。

まずコンピューターの中に「電子辞書」が必要です。コンピューターは人間のようにその辞書を引ながら単語と文の意味を決めます。その上で原文(日本語)の形態素解析をするプログラムとか文の構造を解析する構文解析プログラムとか変換や訳文(中国語)の生成をするプログラムなどの翻訳ソフトウェアが必要です。

人間は目で見て考えてから一つのセンテンスの中にいくつの単語があるかすぐ決められます。コンピューターは辞書によって形態素解析をしますから多分たくさんの解析結果を出してくるかもしれません。たとえば、「くるまではこをはこぶ。」という文を解析して、いくつかの結果が得られます。

くるま で／はこ を／はこぶ。 (車 で／箱を／運ぶ)

くるま では／こ を／はこぶ。 (車 では／子 を／運ぶ)

くる まで／はこ を／はこぶ。 (来る まで

／箱 を／運ぶ)

くる までは／こ を／はこぶ。 (来る までは／子 を／運ぶ)

.....

実際は人間こそコンテキスト (CONTEXT) を参照しないでこの単なる文を見てから同じ問題になるでしょう。

上の文は曖昧さがたくさん出てくる原因の一つはすべての単語が平仮名からなるのです。もちろん、もし輸入文の中に出てくる漢字を使えば人間もコンピューターも易しく解析できるはずですが。

実は曖昧さは形態素解析段階で出てくるのではなくて、構文解析段階も意味分析変換段階も出てくるのです。

例えば、「私は橋で寝ている人を見た。」という文は二つの解析結果が得られます。

- (1) 私は／橋で／「寝ている人」を／見た。
- (2) 私は／「橋で／寝ている人」を／見た。

「橋」という所は私が見た人の寝ている場所か私の立っている場所か両方のどちらもいいですから曖昧さが出てくるのです。

まだ意味分析の段階も曖昧さがよく出てきます。

例えば、「私は村山で、かつて三年ぐらい日本語の勉強をしたことがある。」という文の中の「村山で」は村山という人が山形県の村山市という所かの曖昧さがあります。

しかし、「私は村山で、かつて日本国の政府総理大臣をしたことがある。」という文の中の「村山で」は確かに村山富市という日本国の元首相に違いありません。

人間は「総理大臣」と「村山」という人に繋が

ってから正しく理解出来るのです。このような文はコンピューター翻訳システムにとって本当に難しいことです。辞書の中に単語意味の関連情報をたくさん入れなければなりません。

実はコンピューターで人名と地名の処理は人間の処理よりもっと難しいです。世界中のすべての人名と地名を全部辞書の中に登録する考えは不可能です。

人間は辞書を引かずになぜ人名と地名の識別はできるのでしょうか。勿論関連情報が必要です。例えば、「. . . さん」「. . . 氏」「. . . ちゃん」「. . . 君」などの直前に人名の可能性が高いです。ほかに人名と関連する単語はたくさんあります。「社長」「会長」「部長」「次長」「課長」「議長」「市長」「所長」「委員長」「記者」「容疑者」「俳優」「作家」「選手」「医師」「大学生」「高校生」などいろいろあります。このような単語の直前か直後かに人名の出現率が高いです。

コンピューターもこのような関連単語を考えるべきです。地名の処理も勿論関連情報が必要です。

また、下のような文を読んでから「高い」という単語の意味は何かすぐ分りますか。

“この机はその机より高いです。”

「高い」とは机の価格か机の高度かこの文だけ見たら本当に分からないのです。

“この机は高いですから私はお金が足りなくて買えません。”

この文の中の「高い」は価格の高さを指すものです。

“この机は高いですから私にとって使いにくいです。”

この文の中の「高い」は高度の高さを指すものです。

人間は以上の文の中の単語間の関連情報を用いて考えてからすぐ理解できて翻訳出来るのです。コンピューター翻訳システムにとってやはり難しいことでしょうね。

言語処理学会 第3回年次大会 参加報告

NEC 情報メディア研究所
通信・放送機構 (TAO)
ATR 音声翻訳通信研究所

亀井真一郎、福島俊一、山田洋志
若尾孝博
安藤真一

1 全体概要

3月27日28日両日にわたり京都大学において言語処理学会の年次大会が開かれた。言語処理学会は工学系、言語学系、心理学系など広い学問分野の研究者が集った学際的な学会である。第3回の今年の年次大会はポスタ発表も含め150件近くの論文が発表された。発表件数は第1回の約100件、第2回の約120件から一層増加しており、言語処理研究における多分野交流の活発化を物語っている。

2 招待講演、パネルディスカッション

1日目の午後には神田外語大学の井上和子氏による招待講演「ミニマリストプログラム (MP) と日本語研究」が行なわれた。従来の生成文法の枠組みである統率束縛理論 (GB) との違いを通して、理論言語学の最新研究の枠組みであるミニマリストプログラム (MP) の概説がなされた。

2日目の最後にはNTT 島津明氏の司会で「言語理解研究への回帰」と題したパネルディスカッションが行なわれた。鳥取大学の池原悟氏は、言語表現の中に潜むまだ使われていない多くの情報の活用必要性と、体系的な意味解析の枠組みの研究の重要性を強調し、統計的手法だけにのめり込むことの危険を指摘した。東工大の楠見孝氏は、認知心理学の立場から、今後の言語研究において工学 (システム構築)、言語学 (言語の解明)、心理学 (認知の解明) が相互に影響し合うことの重要性を強調した。九州大学の坂本勉氏は心理言語学の観点から、人間の言語理解過程としての統語解析 (parsing) 研究を論じた。東工大の徳永健伸氏は情報検索における自然言語処理の必要性に関して論じた。情報要求のレベル (1: 直観的要求、2: 意識化された要求、3: 具体化された要求、4: 調整された要求) のうち、

検索式で表される「調整された要求」以外のレベルでは言語処理 (言語理解) が必要であることを指摘した。横浜国立大学の中川裕志氏は、言語処理研究を推進する力として、研究を経済の観点から考えることの重要性を指摘した。経済的に有望な言語処理の適用分野研究として、言語理解を実世界の物事と対応をとることととらえて画像処理と融合させた、ドラマシーン検索におけるト書き文処理を例示した。

3 発表論文の傾向

論文発表は以下のセッションに別れて行なわれた。セッションの名称からも、言語処理学会が広い研究分野をカバーしていることがわかる。

形態素解析、構文解析、構文意味解析、意味分析、意味解析、生成、談話、対話、マルチメディア、言語学応用、辞書、コーパス (構築、設計)、ク (知識抽出)、ク (言語分析)、ク (共起関係)、機械翻訳 (1)、(2)、情報検索 (1)、(2)、ポスタ (1)、(2)

中でも特徴的なのはコーパス関連で4つのセッションが設けられていることである。これは、電子化された大量のテキストが容易に入手可能となり、言語の実運用場面から言語知識を抽出整理しなおそうという気運の現れであると考えられる。また「意味分析」「意味解析」「構文意味解析」などのセッション名称から、意味に関係する研究も盛んになりつつあることがうかがえる。形態素解析、構文解析に関する技術が確立されてきており、今後は意味分析や談話分析の深耕が期待される。

さらにより広い視点からの言語の研究も広がりつ

つある。「マルチメディア」セッションでは画像、音楽等と言語との融合の試みがなされた。「談話」「対話」「生成」等のセッションでは、従来の処理単位である「文」を越えた言語現象の処理に挑んでいる。

4 各セッションの概要

以下にはいくつかのセッションの概要を記す。

[機械翻訳 (1)(2)]

機械翻訳システムは、近年のインターネットの発達により一般ユーザへも急速に普及してきている。このような状況を反映して、本大会でも二つのセッションで機械翻訳関係の研究が活発に討議された。

より一層の機械翻訳システムの普及を目指す研究としては、アジア太平洋機械翻訳協会で行なっている機械翻訳ユーザ辞書の共通フォーマット設定活動の報告があった。機械翻訳の実運用にその充実が必須であるユーザ辞書に関して、異なるメカの異機種間でユーザ辞書フォーマットを共通化し、辞書データの流通を促進させることを目指している。また、翻訳技術そのものではなく訳出品質の評価に関する発表も行なわれた。一般に翻訳品質の評価は難しく、これまで評価に関する研究は十分でなかったが、今後の品質向上には欠くことのできない分野である。

現在広く一般ユーザに普及しているのは英日システムである。インターネットの普及期にあってWWW ホームページや電子メールという形で英語を受信するケースが急増しているのがその理由である。しかし今回の発表では日英に関する件数が英日に関する件数を大きく上回っていた。これは近い将来、日本からの外国語の情報発信が活発化することを先取りしてのことと推察される。日英翻訳におけるテンス・アスペクトの扱い、副詞の訳出、用例利用方式、意味理解のための辞書構成などの広い分野の研究が発表された。また、日英翻訳を「情報発信」ととらえて英日翻訳の「情報受信」と対比し、情報発信に適したマン・マシン・インタフェースとしてユーザが翻訳の途中過程に介入して訳出を確認しながら結果を得る方式の発表もあった。

また将来の方式として、話し言葉の多言語翻訳に関する研究も報告された。原言語と目的言語が言語

的に遠い日英、英日などの翻訳では、言語的により近い日韓などの翻訳と比較して、システム構築コストも多く必要であり、結果の品質向上も困難であることが改めて示された。

[形態素解析]

形態素解析は解析精度が上がってきており、さらなる精度向上には一層の工夫が必要になっている。改良手段として、多くの発表が、誤り例やコーパスからの学習を利用している。解析と生成に同一の文法規則と辞書が使えるような工夫や、異なる複数の応用分野で共通に使えるシステムの開発などの報告があった。コーパスからの確率学習によるシステムでも正解率が90%を超え、実用域に近づいてきている。

[構文解析]

音声認識、形態素解析、構文解析、意味解析など各々のレベルの処理だけでは曖昧性が解消できない場合が数多く存在する。このため本セッションでは、構文解析そのものよりも、各レベルの処理を融合するための枠組みへの取り組み比重が高かった。より多くの知識や曖昧性を扱うために、GLR法（一般化LR文法）による複数レベルの制約統合、オブジェクト指向による漸進的な解析など、パーザの処理効率やメモリ効率を向上させるアルゴリズムが提案された。また、用例主導翻訳の一環としてアナロジー関係を利用して構文解析を行なうアプローチや、異音間接続表と形態素間接続表を組み合わせる手法が提案された。

[構文意味解析]

構文解析は品詞に基づいて行なわれるのが普通であるが、それは必ずしも必然的ではない。品詞という枠組みは、統語的振る舞いを同じくする語群に付けられた名称であるから、むしろ統語的振る舞いの方が第一義的である。このような観点に立ち、単語を統語的振る舞い（係り受け関係）からクラスに分ける研究が発表された。また、日本語の複合語の構造解析のような場面では、助詞・助動詞といった機

能語を介さず語と語がじかに接するので、通常の構文解析だけを用いたのでは、組み合わせの上で可能な解釈の中から正しい解釈を得ることが困難である。この問題への対処として規則と用例を融合させる方式が発表された。

構文意味解析では、単語の多義性解消、文解釈可能性の優先づけが重要なテーマである。今回は、構文解析に確率を取り入れて、複数の可能な解釈の間に優先度づけする研究として、一般化 LR 構文解析に確率を組み込む方法が発表された。また、システムが単語の多義性解消を誤った場合、その原因を特定する研究の一環として、その誤りの主たる原因が、そこで用いているシソーラスの細かさによるのか否かを判定する評価方法が提案された。

[意味解析]

単語間の意味的な類似性は、どのような「視点」(その単語が使われている環境)からその類似性を測るかによって変化する。しかし従来のシソーラスでは特定の視点からの類似性だけを記述しているという問題があった。これに対して本セッションでは「視点」を考慮して単語間の類似性を判別する手法が2件発表された。共に文脈情報を取り入れた手法であり、機械翻訳における意味的曖昧性の解消にも役立つ可能性があるため興味深い。この他に、動詞間時間関係の抽出法や日本語関係代名詞の解析手法についても提案があった。

[意味分析]

言語の「意味」は直感的にはその存在が明らかであるが、万人が共有できる形でとらえることは困難であり、今後の研究の進展が期待される。今回は、意味を視覚化してとらえようとする試みが2件発表された。一つは日本語の慣用表現を、その意味的な成り立ちの観点から意味空間に位置づけ、多種多様な慣用表現の分類・把握を行なう研究である。二つ目は、日本語の古語における待遇表現を、作者、読者、登場人物1、登場人物2の位置関係として理解する試みである。

また意味を日本語と英語の比較という観点から把握する試みも発表された。和英辞書を利用して日本

語の複数文節を英単語に対応づける方法、日本語の再帰形(「自分」など)の英語との対比研究がそれぞれある。また日本語独特の現象の分析として「のだ構文」、名詞化辞「の」の特徴分析に関する発表も行なわれた。

[情報検索(1)(2)]

近年のインターネットの爆発的普及を背景に、自然言語処理研究の重要な応用分野として、情報検索への取り組みが活発化している。今回も2セッション(A3、A4)で15件の発表があった(全文検索システム:2件、検索精度改善:5件、類似検索:3件、テキスト分類・分割:4件、テキスト抄録:1件)。ギガバイト規模のテキストを対象とした高速な全文検索手法、多様なテキスト集合へのアクセスを容易にするテキスト分類や類似検索などは、インターネットを意識した研究テーマであるし、複合語や係り受けを考慮して検索精度を改善する試みなどは、自然言語処理をベースとしてよくとられるアプローチの1つである。しかしながら、伝統的な情報検索研究を踏まえた評価という面では不十分なものも多く見られた。これに関しては、最近、情報処理学会のワーキンググループ活動により、小規模ながらも日本語を対象とした初めてのテストコレクション BMIR-J1 が開発されたので、その利用普及と拡充が期待される。

[コーパス(設計構築)]

ここでは現在設計中あるいは構築中の各種コーパスの紹介が中心であった。まず京都大学で進行中のテキストコーパス・プロジェクトは、高精度な日本語解析システムを利用し、人手により、正しい形態素・構文情報が付与された日本語文章データを作成し、解析システムとともに公開するのが目的である。対象は新聞テキストであり、97年に1万文が実験コーパスとして公開される予定である。次に、日本電子化辞書研究所で進められている知識ベース増殖のための中央データベースの紹介があった。このデータベースには、解析レベルを異にしたコーパスが集められている。後の解析は、対象となるコーパスの解析レベルに応じて行われ、その解析結果がまたコ

ーパスとしてデータベースに蓄積される、つまり知識増殖をもたらすこととなる。また、米国国防省 DARPA の支援で行われた MUC 6 (1995年) において作成された照応関係を明示したタグをつけた英語新聞記事コーパスおよび作成の過程についての紹介があった。対象とする照応関係としては、当初「同等」、「部分と全体」など5種類が提案されたが結局、同等関係のみを対象とし SGML でタグづけを行なった。

WWW 上の言語データにタグづけができれば、大規模な言語データベースができることになる。HTML だけではなくもう少し詳しいタグがつけられれば、機械でのより高度な処理も可能となりうる。このための標準のタグおよび必要とされる意味タグ用の辞書についての考察が行なわれた。本セッションでは、WWW を利用して言語データを収集する提案も見られた。これを実現するためのインターフェースのプロトタイプや、実際の運用にあたっての注意事項などが紹介された。

[コーパス (知識抽出)]

多くの分野にわたった発表がなされた。コーパス研究の基礎としては、形態素解析と構文解析で品詞体系を統一する試みが発表された。日本語は品詞のレベルでも標準がなく、また標準化が困難である。問題点の明確化・共有化と早期の標準設定が望まれる。また、作成者、使用者の意図・想定によってコーパスを分類し、コーパスの概念を明確化すべきであるという指摘もなされた。

コーパスからの知識抽出の研究としては、文字・単語の tri-gram を使用した言語・文学作品のクラスタリングの研究、27項目のもの呼び方を元にした方言地図の作成、10地域で世代の違う10人の音声を取録した音声コーパスの作成、単語の N グラムを

使った音声認識などの研究発表があった。

[コーパス (言語分析)]

コーパスを用いて言語を分析した結果が報告された。文法開発支援、ニュース記事からのキーワード抽出、分野別 OCR 誤りの分析、格の従属関係の学習、口語・古語の扱い、同語反復表現の分析、さらには、新聞における JIS 第一、第二水準漢字以外の出現状況の分析などトピックは多岐にわたった。

ニュース記事を対象としたキーワード抽出では、平均情報量基準に基づいてキーワードを選定し、実際のニュース記事を用いてキーワードの有効性を検証している。格の従属関係の学習の研究では、動詞の格の従属関係をコーパスを用いて統計的処理により学習する手法が提案され、その有効性が検討された。学習の方法としては最大エントロピー法がもちいられた。文節の係り受け解析を行なった結果、格の従属関係を考慮した確率モデルはそれを無視したモデルよりも精度が良いこと、また、最大エントロピー法で推定した確率モデルは、従来の確率モデルよりも良い結果を示すことが示された。

RWC テキストデータベースにおける口語・古語の扱いの研究では、特に新聞記事に現れる文語体、旧仮名遣いおよび口語体特有の表現について、一貫性を持たせるために行なった処理が報告された。同語反復表現の分析では、例えば「約束は約束だ」などを対象として、その範囲・分類などが検討された。同語反復表現としては、トートロジ的なものと慣用的なものがあり、また同一の表現にも意味の曖昧性がみられた。これらの研究は、従来あまり対象とされていなかった興味ある言語現象が広く分析対象として取り上げられるようになってきていることを示す好例である。

CICC 近隣諸国間の機械翻訳システムプロジェクト

財団法人国際情報化協力センター(CICC) 国際情報化研究所長 上金 孝平

はじめに

財団法人国際情報化協力センター(CICC)は、通商産業省からの委託を受けて、1987年度から8年間にわたり政府開発援助(ODA)の一環として「近隣諸国間の機械翻訳システム」に関する共同研究を、中華人民共和国、インドネシア共和国、マレーシア国及びタイ王国の研究機関と実施しました。引き続き、95年度及び96年度はフォローアップ事業を実施し、3月末をもって当該事業も終了しました。本稿では、CICCが関係機関と実施した機械翻訳システムに関する10年間の事業の成果の概要を紹介することとします。

1 背景

近年、日本をとりまく近隣諸国との間で、さまざまな技術交流や文化交流が行われています。技術交流や文化交流を円滑に進めるにあたって、コミュニケーションを図ることが不可欠となっております。実際のコミュニケーションでは、日本語や英語で直接話し合う、通訳を介して会話をする、手紙のやりとりをするなどの方法がとられます。日本語習得の難しさや通訳者の不足から、実際には、英語を使用した直接の会話や文書のやりとりが行われることが多いという状況にあります。

しかし、日本人を含めて、全ての技術者や現場作業者が英語に堪能というわけではありません。意志疎通が不十分のために円滑に技術移転ができなかったり、建設工事が遅れることがあります。言語の壁のために相互信頼が確立できないといった弊害も少なくありません。したがって、いろいろな機会や場において、現地語のマニュアルや各種ドキュメントを作成することが求められております。

多言語機械翻訳プロジェクトは、このような状況を背景にして開始されました。また、アジア諸国に共同研究の可能性を打診した結果、最終的に日本を含めて、中華人民共和国、インドネシア共和国、マレーシア国及びタイ王国の5カ国の共同研究開発プ

ロジェクトとして始められました。

2 目的・目標

機械翻訳システムは、次のような目的を持って研究開発されました。

- ①最新の情報処理システムを共同研究開発すること。
- ②共同研究開発から派生する母国語情報処理技術の基盤を確立すること。
- ③その他さまざまな技術交流や文化交流を促進すること。

また、技術面では、日本、中華人民共和国、インドネシア共和国、マレーシア国及びタイ王国のアジア5カ国間を対象として、これらのどの国の言語にも翻訳できる方式の研究開発を行うことを、経済面では、情報処理分野の文献を実用的な速さで及び正確さで翻訳を行うことを目標として掲げました。

3 組織体制

このプロジェクトは、日本、中華人民共和国、インドネシア共和国、マレーシア国及びタイ王国の5カ国の共同研究開発プロジェクトでした。日本にお

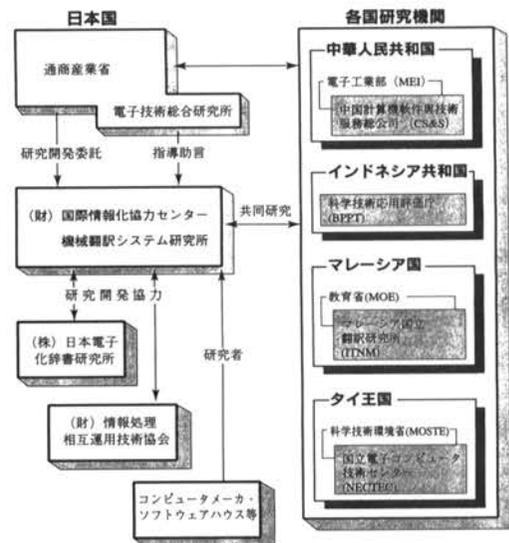


図1 体制図

いては、通商産業省から研究開発の委託を受けたCICCが中心となり、いくつかの企業が機械翻訳システムの研究開発及びプロジェクトへの支援を行いました。近隣アジア諸国においては、それぞれの国の研究推進機関又は研究機関が研究開発を行いました。

機械翻訳システムの研究開発体制を図1に示します。なお、CICCの機械翻訳システム研究所は、1995年4月に国際情報化研究所に改称されました。

4 内容及び成果

このプロジェクトでは、5か国語間の機械翻訳システムにおいて「中間言語方式」を採用しました。したがって、研究開発の内容及び成果としましては、中間言語の仕様の作成、作成した仕様に基づく機械翻訳システムの開発及び電子化辞書・コーパスの作成、開発したシステムによる機械翻訳の実証実験等が挙げられます。

4-1 中間言語の仕様の作成

対象とする5か国語で表現された文章を独自の表現形式で表すための仕様（規則、記述方法等）を作成しました。当初の仕様は、何度も見直しが行われ、改善が図られました。

4-2 システム開発

作成された中間言語の仕様をベースにして、機械翻訳システムが開発されました。このシステムは、入出力システム、翻訳支援システム、文解析システム、文生成システム、電子化辞書システム、文章ファイル管理システム、ネットワークシステム等から構成されています。

4-3 電子化辞書及びコーパスの作成

機械翻訳システムを実際に稼働させるためには電子化辞書が不可欠で、対象とした各国語について、基本語及び専門用語の電子化辞書を作成しました。また、実証実験の対象データとして、各国語で記述された3000の例文（コーパス）を作成しました。

4-4 実証実験

開発された機械翻訳システム及び作成された電子化辞書を使用して、作成したコーパスを対象に各国

語間での翻訳の実証実験を行いました。最終的に、目標とした少なくとも50%の翻訳率をクリアしました。

5 経過

1987年度から1994年度までの8年をかけて、機械翻訳システムの研究開発を進めました。図2に機械翻訳システムの研究開発の経過を示します。1995年度及び1996年度には、8年間の共同研究の成果をもとに、フォローアップ事業を実施しました。

	年度	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996
(1) 機械翻訳システムの基礎研究		—									
(2) 機械翻訳システムの研究開発			—	—							
(3) 機械翻訳システムの評価と改良							—				
機械翻訳システムの研究フォローアップ										—	

図2 研究開発の経過

5-1 機械翻訳システムの基礎研究

最初の2年間で、機械翻訳システムを構成するいくつかのシステムと要素の基本仕様を設計し、試作しました。約5000語の辞書と数十の例文を作成し、文法規則の仕様を設計しました。また、限定された翻訳例文を使って、機械翻訳システムの処理能力を確認しました。

5-2 機械翻訳システムの研究開発

次の4年間で、基本語辞書約5万語、情報処理関連の専門語辞書2万5千語及び約3000の例文を作成し、文解析や文生成を制御する文法規則を試作しました。また、追加した翻訳例文と任意の自然文を使って、機械翻訳システムの翻訳処理能力を確認しました。

5-3 機械翻訳システムの評価と改良

最後の2年間で、(2)段階までに試作された機械翻訳システムを使って、翻訳処理能力実験を繰り返し、辞書、文法規則及びシステム全体を改善しました。

5-4 機械翻訳研究フォローアップ事業

8年間の共同研究の成果をもとに、研究に参加した各国において、辞書・文法の改良・追加、機械翻訳システム実用化に向けた追加研究を独自に実施できるよう支援しました。

6 各国の研究機関の現状

6-1 中華人民共和国

CS&Sは、本プロジェクトにより大きく発展しました。また、本プロジェクトに参加した清華大学等の大学における言語情報処理に関する研究能力も向上しました。

CS&Sは、独自に中一英、中一日の機械翻訳システムを開発中です。

6-2 インドネシア共和国

BPPTは、この分野ではゼロからの出発で、大きく成長発展しました。現在でも、開発した機械翻訳システムを維持しており、インターネットによる電子化辞書、翻訳システムをサービス中です。

また、独自に1997年から2002年までの自然言語処理に関する研究計画を策定し、実施しています。

6-3 マレーシア国

ITNMは、機械翻訳システムの研究から撤退し、マレーシア工科大学へ引き継ぐ予定です。

ITNMの実施していたマレーシア語のスペルチェッカーの開発、電子化辞書生産は、残念ながら中止となりました。

6-4 タイ王国

NECTECは、本事業とともに発展し、設立10周年を迎えました。独自にCD-ROMによる電子化辞書を開発・製品化(タイ語全般、タイ語用例、タイ語同義語・反義語、タイ語—英語、英語—タイ語)し、現在、その改訂版を製作中です。

7 今後の活動

CICCとしましては、これまでの10年間の研究開発の実績、培われた人的ネットワーク等を活用して、関係国との間で、機械翻訳に関する情報交換等を独自に実施していくこととしております。具体的には、①研究者、研究機関リストの作成・提供、②電子メ

ールやインターネット電子会議による情報交換、③各国レポート等のホームページによる情報提供、④本プロジェクトに関する小冊子の作成等を予定しております。

また、機械翻訳に関する新しいプロジェクトの可能性を検討したり、関係機関との連携を図っていくこととしております。

おわりに

本プロジェクトは、近隣アジア諸国との間での情報分野における共同研究としては、初めての試みであり、その成果や貢献に対して各国からも高い評価を受けております。

それは、本プロジェクトが、各国における言語処理研究基盤の向上(研究者・技術者の養成、電子化辞書の整備等)、各国における情報処理技術基盤の向上(文字入力方式の確立、WS・UNIX技術者の養成等)、ネットワークの確立(人的ネットワーク、機関等のネットワーク等)に大きく貢献したからであると思います。

機械翻訳システムや技術は、日々、発展しておりますが、技術的にも難しい課題であり、インターネットやパソコンの普及に伴う情報環境の変化、多言語情報処理の動向等も考慮して、今後とも引き続き取り組んでいくべき課題であると思います。

最後に本プロジェクトに御協力・御支援いただきました関係機関及び多数の関係者の皆様方に心から謝意を表したいと思っております。

(参考) 機械翻訳システムの概要

1 機械翻訳のシステム構成と翻訳の流れ

図3に機械翻訳のシステム構成と翻訳の流れを示します。

機械翻訳システムは、以下の機能システムにより構成されています。

①入出力システム

端末及びフロッピーディスクなどの磁気媒体からデータを入力します。また、機械翻訳されたデータを端末の画面上やプリンタに出力します。

②翻訳支援システム

機械翻訳処理を効率良く行なうために欠かせない

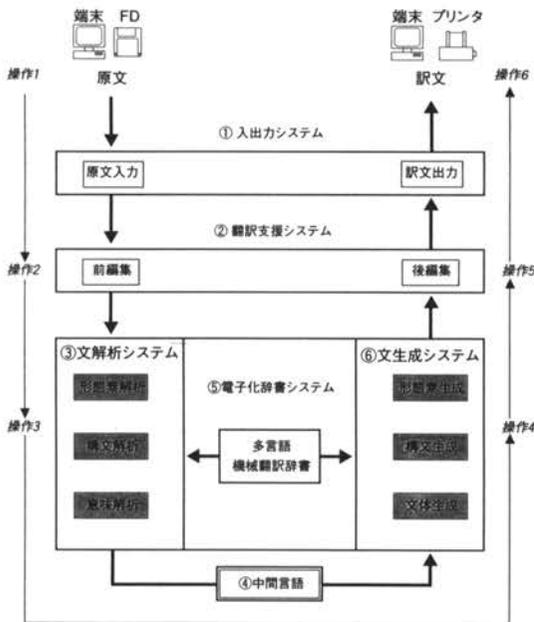


図3 機械翻訳システムの構成と翻訳の流れ

原文の修正（前編集）や訳文の修正（後編集）を支援します。

③文解析システム

入力された文章に対して形態素解析、構文解析、及び意味解析を行い、解析されたデータを中間言語に置き換えます。

④電子化辞書システム

各言語の文法情報や意味情報など機械翻訳に必要な情報を管理します。

⑤文生成システム

中間言語のデータに対して、各言語に対応した文体生成、構文生成及び形態素生成の処理を行います。さらに、各言語対応の「文章ファイル管理機能」を持った「総合システム」があります。総合システムは、機械翻訳システムをひとつのワークステーション（スタンドアロン）上で実現したり、また、ワークステーションどうしをネットワーク上に接続するためのインターフェースを実現します。

以下に翻訳の流れを説明します。

*操作1：原文入力

原文をシステムに入力します。原文はワードプロセッサなどを利用して端末から直接入力するか又はフロッピーディスクやハードディスクなどの磁気媒体から入力されます。

*操作2：前編集

翻訳支援システムの前編集では、翻訳オペレータが原文を端末の画面上で機械翻訳の処理がしやすいように、必要に応じて修正します。処理されたデータは文解析システムに送られます。

*操作3：文解析

文解析システムでは、電子化辞書システムに蓄積されている機械翻訳情報（文法情報や意味情報などの解析文法規則及び各言語と中間言語との対応情報）を基に入力されたデータの形態素、構文及び意味の解析を行ないます。さらに解析されたデータを機械翻訳情報を基に中間言語に変換し、文生成システムに送ります。

*操作4：文生成

文生成システムでは、電子化辞書システムの機械翻訳情報（生成文法規則及び各言語との中間言語との対応情報）をもとに指定された言語に合わせて中間言語から文体、構文及び形態素を生成し、出力システム・翻訳支援システムに送ります。

*操作5：後編集

翻訳支援システムの後編集では、翻訳オペレータが機械翻訳結果と原文を対照しながら画面上で確認し、必要に応じて編集します。

*操作6：訳文出力

編集された訳文は、出力システムを介してプリンタや磁気媒体に出力されます。

2 対象言語

この多言語機械翻訳システムでは、日本語、中国語、インドネシア語、マレーシア語及びタイ語を対象言語としました。各国の言語は中間言語を経由して、相互に翻訳されます。このような方式を中間言語方式といい、図4に示すように、ひとつの言語を一度に複数の他国語に翻訳できるのが特徴です。

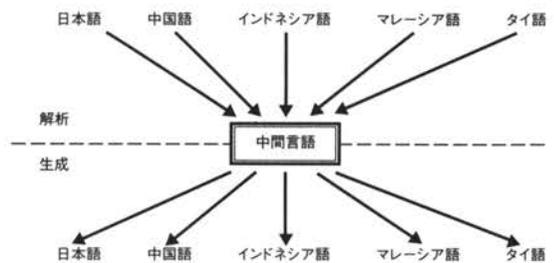


図4 言語構成

3 翻訳方式の考え方と特徴

この機械翻訳システムでは、多言語の翻訳処理を目標としています。日本語から中国語へ、中国語からインドネシア語への翻訳など、機械翻訳システムを共同開発している国の言語であれば、相互に翻訳処理できるというものです。

3-1 中間言語方式

多言語機械翻訳システムでは、翻訳処理を実現する手段として、「中間言語方式」を採用しています。中間言語とは、あらゆる言語の表現、文章の意味内容、表現態様を、それぞれの言語に依存しない共通の表現形式に置き換えた言語です。例えば、日本語から中国語への翻訳過程で、日本語を中間言語に変換し、この中間言語から中国語を生成します。

中間言語方式を採用することによって、どの国の言語であっても、目標とする国の言語への翻訳ができるようになります。また、機械翻訳システムを開発する技術者にとっても、中間言語方式は大きな利点があります。各言語の技術者は、中間言語と自国の言語の翻訳システムだけを実現すれば、どの言語へも翻訳が可能となるからです。

3-2 多言語機械翻訳辞書

中間言語方式を実現するには、各言語の語彙が中間言語の語彙と対応づけられていることが必要です。ここで、中間言語と並んでシステムに重要な要素として辞書があります。中間言語方式を採用している機械翻訳システムでは、各言語の辞書は次のように統一されていなければなりません。

- * 言語に依存しない共通の語彙概念との対応関係
- * 辞書の体系、構成及びデータ形式の統一

辞書を作成するときに、どのような語彙を収集し、各言語にどのような情報（文法属性など）を付与するかは、その言語固有の問題です。

4 ソフトウェア構成

機械翻訳システムは、次のようなシステム（ソフトウェア）から構成されています。

①基本システム

*タスク管理システム (TCS : Task Control Sub-system)

*通信制御サブシステム (CCS : Communication Control Sub-system)

*サブシステム情報サービス (SIS : Sub-system Information Service)

*テキストベースサブシステム (TXB : Texts Base Sub-system)

②文解析・文生成システム

*日本語文解析・文生成システム (Japanese Analyser /Generator)

*中国語文解析・文生成システム (Chinese Analyser /Generator)

*インドネシア語文解析・文生成システム (Indonesian Analyser/Generator)

*マレーシア語文解析・文生成システム (Malaysian Analyser/Generator)

*タイ語文解析・文生成システム (Thai Analyser/Generator)

③ユーザーインタフェースシステム

*翻訳者向け翻訳支援システム (TSS 3 : Translation Support System for Translator)

*研究者向け翻訳支援システム (TSS 6 : Translation Support System for Researcher)

*1バイトコードワードプロセッサ (WP 3 : Single Byte Word Processor)

*2バイトコードワードプロセッサ (WP 7 : Double Byte Word Processor)

④ユーザインタフェース統合ツール

*ユーザオペレーションパネル (デスクトップ) (UOP : User Operational Panel)

図5に分散モデルの場合の構成例を示します。

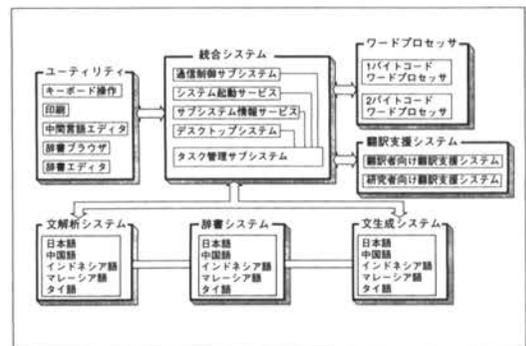


図5 ソフトウェア構成図

ネットワーク上での英語の使用について

(社)日本電子工業振興協会 テキスト処理技術専門委員会 市場動向ワーキンググループ

1. はじめに

パソコンとネットワークの普及によって、ネットワークを介したオンラインテキストデータの利用やメール送受信の機会は様々な場面に広がっている。

(社)日本電子工業振興協会(JEIDA)のテキスト処理技術専門委員会では、これら新しい利用者像が期待するインターネット上のコンテンツとその利用形態、及び支援機能を正確に探ることを目的として、前者に関しては「文書データ利用」「電子新聞」「電子図書館」、後者に関しては「ネットワーク上の英語の使用」「情報検索」をそれぞれ取り上げ、1996年4月～5月にアンケート調査を行なった。214件の回答(有効:212件)のうちの50%が技術/研究職、35%が教職/学生であった。

本稿では、上記5テーマのうち「ネットワーク上の英語の使用」を取りあげ、主なアンケート結果を紹介する。そこではネットワーク上での英語使用の現状及び機械翻訳ソフトの利用状況を把握することを目的として、以下の14設問を設けた。

[ネットワーク上での英語の使用について]

- 1) 使用の有無・場面
- 2) 使用頻度
- 3) 使用目的
- 4) アクセスの仕方
- 5) 不便な点
- 6) 対処方法
- 7) 不便さの解消による英語の使用頻度の変化
- 8) 英語を使用しない理由

[機械翻訳ソフト]

- 9) 使用経験
- 10) ソフト名
- 11) 使用機能
- 12) 満足点
- 13) 不満点
- 14) ネットワーク上での利用

なお、他の4つのテーマを含め全ての設問項目・集計結果・分析結果は次のJEIDAホームページに掲載中であるのでご覧いただきたい。

<http://www.jeida.or.jp/committee/textsryori/sec-0.html>

2. 調査結果

2.1 英語の使用

《使用の有無・場面、使用頻度》

受信・発信別の電子メール、WWW、ネットニュースの利用状況・頻度を尋ねた所、WWWによる情報の受信が回答者の90%と最も多く、更に利用者の80%が週に数回以上と高い頻度で使用しており、WWWの普及によって英語情報に接する機会が増加していることを示している。全体としては、受信型のアクセスにおける使用率が、発信型のアクセスにおける使用率よりもかなり高い傾向にあり、職種によるばらつきも受信型では少ないことがわかった。

《アクセスの仕方》

下記集計に示すように、「斜め読みによる概要把握」と答えた人の割合が高く、ほんやりネットサーフするだけでなく、英語で書かれた大量の情報の中から、必要なものを素早く読み取ろうとする利用状況が表れている。

- 19.3%：文章をじっくり読む
- 68.7%：文章を斜め読みして概要を把握する
- 4.7%：カタログなどリスト形式の情報を読む
- 4.0%：データベースを検索する
- 0.7%：あちこち覗いてみるが、内容を読むことは殆んどない
- 0.0%：イメージを見る・音を聞く

(以下省略)

単一回答。 回答件数=150

《不便な点》

下に示すように「不便だと感じることはない」と

するものはわずか15%弱である。

- 49.2%：精読するのが難しい
 - 55.4%：概要を把握するのに時間がかかる
 - 43.6%：わからない単語が多い
 - 44.6%：書いた内容や表現が適切かどうかわからない
 - 41.5%：書くのに時間がかかる
 - 14.4%：不便だと感じることはない
- (以下省略)
- 複数回答可。 回答件数=195

《対処方法》

下記集計に示すように、不便さを解消する手段としては、紙の辞書や電子辞書を利用するという割合が高く、機械翻訳ソフトを利用する人の割合は10%未満にすぎなかった。

- 58.9%：紙の辞書を引く
 - 51.4%：電子辞書あるいはオンライン辞書を引く
 - 18.9%：市販の参考書や用例集などを参考にする
 - 25.7%：過去のメールなど、保存してある個人情報などを参考にする
 - 9.1%：機械翻訳ソフトを使う
 - 6.9%：人に訳してもらう
 - 12.6%：読むのを諦める
 - 3.4%：書くのを諦める
- (以下省略)
- 複数回答可。 回答件数=175

2.2 機械翻訳ソフト

《使用経験》

回答者の半数近くが使用経験ありと答えていた。本アンケートが、低価格機械翻訳ソフトの発売がまだ数少なかった一年前に行なわれたことを考慮すると予想よりは多いと言える。

《満足度／不満度》

下記集計に示すように、機械翻訳ソフトを使用したことのある人で「不満な点はない」と答えた人はなく、利用者が完全に満足できるレベルには達していないという厳しい結果となった。しかし「概要把握」が満足と答えた人が不満と答えた人を上回っているのをはじめ、「辞書・参考書代わり」など使

方によっては有効なツールになりうることもわかった。

《満足度》

- 5.6%：翻訳された内容がよく理解できた
- 42.2%：概要をつかむのに役立った
- 7.8%：適切な言い回しが得られた
- 35.6%：辞書・参考書などを見る手間が省けた
- 20.0%：処理速度が速い
- 42.2%：満足した点はない

(以下省略)

複数回答可。 回答件数=90

《不満度》

- 42.2%：翻訳された内容が理解できなかった
- 20.0%：概要がつかめなかった
- 62.2%：適切な言い回しが得られなかった
- 52.2%：結局、辞書・参考書などを見なければならなかった
- 25.6%：処理速度が遅い
- 0.0%：不満な点はない

(以下省略)

複数回答可。 回答件数=90

3. おわりに

インターネット上で何らかの場面で英語に接する利用者が多いこと、しかも利用者のほとんどが英語に対し不便さを感じていることなどから、機械翻訳の需要は大きいものと思われる。実際にも、現在の概要把握型の英文読解支援機能に対する満足度は比較的高い。ただ、英語の不便さを解消する手段としては、(機械翻訳ソフトではなく)紙の辞書や電子辞書が利用されていることから、現在の機械翻訳ソフトが「支援」としてはまだ十分ではないことも明らかになった。具体的な英語の使用場面毎に、利用者が何に不満を持ち、何を求めているかを掘り起こすことが早急の課題である。

《参考文献》

奥西、金岡、斎藤、藤田、本間、松井、土井、井佐原：「インターネットにおける情報利用の現状と期待—電子協アンケートの結果から—」(情報処理学会第54回全国大会、1997/3)

用語と訳語の統一用ソフト エイチティースリー HT³の開発

株式会社日立国際ビジネス

1. はじめに

1.1 開発動機

当社の業務のひとつに、マニュアルの日英翻訳がある。

最近、機械翻訳ソフトも廉価で販売されるようになっているが、完全性が求められる業務用翻訳には、機械翻訳の導入は限られた環境を除いては不向きである。やはり、専門分野に通じたプロの翻訳者に翻訳を依頼し、仕上げにネイティブチェックをかける現在の方法が現実的である。しかし、翻訳量の多いマニュアルの場合、顧客から支給された技術用語の日英対訳の用語辞書を完全に翻訳に役立てるのは実際は非常に難しい。

用語辞書にある文字列については、指定された英語を本当に全部使用したかどうかは、明白な証拠がなく、翻訳者とチェッカーの感性に頼らざるを得ない。しかもこれまでは、翻訳者が仮に、文章と用語辞書を丹念にチェックしても、もとの日本語原稿が悪く、文章中の用語が用語辞書にある表現（標準表記）になっていないこと（異表記）も多いので、用語辞書の英訳語が採用されないこともあった。

以上のことから、「表現の統一」を実現する手段としてパソコンで使えるソフトの検討を行った。

機械翻訳ソフト、辞書引きソフトや文章校正ソフトなどを検討したが、どれも本目的にすっきりとは適合しないと判断し、最終的に自主開発に踏み切った。これにより、品質向上と納期短縮の両面において効果が期待される。

なお、ソフトの名称 HT³は Hitachi Terminology Transformation Tool の頭文字からとった。

1.2 標準表記と異表記

ここで HT³の最大の特徴である標準表記と異表記について説明しておく。「標準表記」は辞書の見出し語として記載されている文字列のことである。「異表記」は「標準表記」の同義語である。長音記号の有無、拗音のひらがなの大小、送り仮名の有無のよ

うな「言葉遊び的な異表記」もあれば、業界スラングや自家製用語の異表記もある。例えば、「フロッピーディスク」と言う表現を標準表記としよう。この異表記には、「フロッピー、フロッピ、フロッピディスク、アーカイブディスク、ディスク、FD」などがある。もちろん英語にも標準表記と異表記がある。

「floppy disk」が標準表記とすると、異表記には「diskette、archive diskette、FD」などがある。学術的には「同義語」と「異表記」を区別するようであるが、ここでは、標準表記に対して、それ以外を異表記とした。

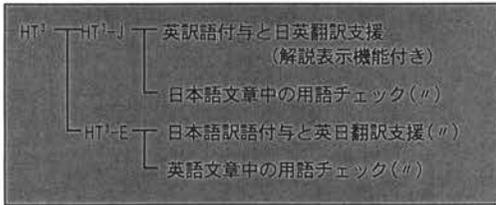
1.3 訳語付与ソフトから用語統一ソフトへの発展

日英翻訳支援用の訳語付与では、日本語原稿へ用語辞書にある訳語を付与するが、日本語原稿の用語は多くの異表記が含まれているのが通常である。そこで標準表記、異表記に拘わらず、正しい英訳語を付与できるようにするため、用語1個につきひとつの標準表記と複数個の異表記を登録できるような電子辞書形式と、それに対応できるソフトを開発した。仮に、「フロッピーディスクと floppy disk」が対訳として出ている電子用語辞書を使用して「フロッピを取出す. . .」という文章と突き合わせてもヒットしないから、当然「フロッピ」という語には「floppy disk」は付与されない。HT³では、フロッピーディスクという標準表記とともに「フロッピー、フロッピ、フロッピディスク. . .」を異表記として辞書登録しておくことによって、これらのうちのどの文字列が文章中にあっても「floppy disk」という訳語が自動的に付与されることになる。このことにより、同じ用語辞書の日本語（または英語）の標準表記と異表記を利用して日本語（または英語）の文章中の、用語の表記チェックおよび異表記の標準表記への置換が可能となった。

2. ソフトの概要

HT³は最もユーザー数が多いと思われる Windows のワープロソフト MS-Word のアドインソフトとして開発した。バージョンは MS-Word6.0 および MS-Word7.0 (Windows95対応版) である。

ソフトには、日英翻訳支援と日本語文章中の用語をチェックするための「HT³-J」、英日翻訳支援と英文文章中の用語をチェックするための「HT³-E」の2種類がある。



辞書作成は Excel5.0 または 7.0 上で、データベースに変換してから文書中の文字列とマッチングさせる。マッチング方式には、前方最長一致を採用した。

電子辞書のサイズはハードディスクに依存する。データベースによる検索を採用しているため、電子辞書のサイズは処理速度に影響しない。処理速度はパソコンのハードの性能に影響される。開発者としては、カラー対応の Pentium の 75Mhz 以上の性能のパソコンでの使用を推奨する。

HT³ソフトをインストールするにはパソコンのハードディスクに 3 MB 以上の空きが必要である。

3. 機能説明

3.1 HT³-J [英訳語付与と日英翻訳支援]

HT³-J の「英訳語付与と日英翻訳支援」機能および「日本語文章中の用語チェック」機能のうち、ここでは「英訳語付与と日英翻訳支援」機能について使用手順を簡単に説明する。

日本語原稿を MS-Word で開く。次に、ツールから HT³-J を立上げる。HT³-J のメインメニューの「辞書管理」で、使用する電子辞書を選択する。「英訳語付与ボタン」をクリックすると英訳語付与が開始される。選択した電子辞書に含まれる日本語文字列は標準表記、異表記の区別なく、日本語原稿の文字列とヒットしたものの全てに標準表記の英訳語が付与される。この場合、電子辞書に品詞情報と共に動詞

と名詞が異なるものの英訳語 (例: 着信=arrival [名] と arrive [動]) を適切に入力してあれば、日本語文章中での用途により付与される英訳語も動詞の場合は arrive、名詞の場合は arrival と出し分けられる。



図1 HT³-J メインメニュー

英訳語付与が完了したら、いったん HT³での作業を中止し、英訳語が付与された日本語原稿をプリントアウトして翻訳者に渡しても良い。こうすれば翻訳者はもはや紙の用語辞書を参照することなく、指定された英訳語を全て利用した翻訳作業がすばやくできる。

もちろん、HT³の翻訳支援機能を使って翻訳しても良い。HT³では、英訳語付与済み日本語原稿画面と翻訳用英文入力画面の2画面を同時に開いておき、付与された英訳語を英文入力画面の英文に取り込むことができる。更に、翻訳支援機能では付与された英訳語が英文で採用されたかどうかのチェックと統計を取ることができるから、感性に頼らない訳語使用およびチェックが可能である。

この機能を使って翻訳した場合には、採用した英訳語が英文中に色分け表示されているので、ネイティブチェッカーにその状態でチェックを依頼すれば、顧客から指定された用語に対する英訳語をチェッカーが勝手に変更してしまう事故も防止できる。

また、HT³は用語辞書に解説欄を持っている。ここに翻訳に参考になるノウハウや、例文あるいは用語の解説なども入力できる。解説欄に何か情報を持っている場合は、日本語原稿に付与された英訳語がイタリックになる。翻訳支援画面でこのイタリックになっている英訳語を、ハイライトしてから「解説ボタン」をクリックすると解説欄にある情報を参照できる。ベテラン翻訳者の翻訳知識を他の人にも共有できるので便利である。

なお、英訳語付与の方式には5つのバリエーションが準備されている。翻訳者が好みにあわせて選択することができる。

3.2 HT³-J [日本語文章中の用語チェック]

HT³-Jでは日本語原稿中の用語が用語辞書の標準表記で表記されているか、異表記かをチェックし必要により異表記を標準表記に置換することができる。統計も取れる。ここでは、「用語チェック」機能について使用手順を簡単に説明する。

日本語原稿をMS-Wordで開く。次に、ツールからHT³-Jを立上げる。HT³-Jのメインメニューの「辞書管理」で、使用する電子辞書を選択する。「日本語文用語確認ボタン」をクリックすると日本語用語確認ボタン群が表示される。ボタン群の「用語確認(色分け)ボタン」をクリックすると「用語確認(色分け)」が開始される。指定した用語辞書の標準表記と完全にヒット(一致)した文字列は「水色」になる。ヒットしたが、英数字やカタカナの全半角が違っている文字列は「黄緑」になる。用語辞書の異表記と完全にヒットした文字列は「ピンク色」になる。異表記にヒットしたが全半角が違っている文字列は「赤」になる。

通常文章校正ソフトでは問題の部分の色分け表示したりハイライトはするが、良い部分は表示しない。その点HT³は問題の部分だけでなく、良い部分も表示するので標準表記が使用されている場所が確認できる。

用語確認(色分け)が終了したら、「置換候補表示/置換」をする。この機能は黄緑、ピンクおよび赤に色分け表示された文字列(すなわち水色以外で標準表記の候補を持っている文字列)が文頭から順にハイライトし、その文字列に対する標準表記を置換候補として表示する。表示された標準表記に置換するときは「はい」を、置換しないときは「いいえ」をクリックしていく。文末までチェックし終わったら完了である。

用語辞書の解説欄に何かデータが記入されている場合は、色分け表示された文字列はイタリック表示されるから、「解説ボタン」をクリックすると解説が表示される。

また、ヒットした用語数、標準表記の数、異表記の数を統計表示することができる。

解説表示機能を利用すれば、用語辞書の解説文を参照表示できる。

3.3 HT³-E [日本語訳語付与と英日翻訳支援]

次に、HT³-Eの「日本語訳語付与と英日翻訳支援」機能および「英文中の用語チェック」機能のうち、「日本語訳語付与と英日翻訳支援」機能について使用手順を簡単に説明する。基本的な使用手順は、前述3.1 HT³-J [英訳語付与と日英翻訳支援]と同様である。

まず、英語原稿をMS-Wordで開く。次に、ツールからHT³-Eを立上げる。HT³-Eのメインメニューの「辞書管理」で、使用する電子辞書を選択する。「日訳語付与ボタン」をクリックすると日本語訳語付与が開始される。選択した電子辞書に含まれる英語文字列は標準表記、異表記の区別なく、英語原稿の文字列とヒットしたもの全てに標準表記の日本語訳語が付与される。



図2 HT³-E メインメニュー画面

日本語訳語付与が完了したら、いったんHT³での作業を中止し、日本語訳語が付与された英語原稿をプリントアウトして翻訳者に渡しても良い。日英翻訳のときと同様、こうすれば翻訳者はもはや紙の用語辞書を参照することなく、指定された日本語訳語を全て利用した翻訳作業がすばやくできる。

もちろん、HT³の翻訳支援機能を使って翻訳しても良い。HT³では、日本語訳語付与済み英語原稿画面と翻訳用日本語文入力画面の2画面を同時におき、付与された日本語訳語を日本語文入力画面の日本語文に取り込むことができる。更に、翻訳支援機能では付与された日本語訳語が日本語文で採用されたかどうかのチェックと統計を取ることができるから、感性に頼らない訳語使用およびチェックが可

能である。

この機能を使って翻訳した場合には、採用した日本語訳語が日本語文中に色分け表示されているので、チェッカーにその状態でチェックを依頼すれば、顧客から指定された用語に対する日本語訳語をチェッカーが勝手に変更してしまう事故も防止できる。また、HT³の辞書の解説欄に、翻訳に参考になるノウハウや、例文あるいは用語の解説などを入力している場合は、英語原稿に付与された日本語訳語がイタリックになる。翻訳支援画面でこのイタリックになっている日本語訳語を、ハイライトしてから「解説ボタン」をクリックすると解説欄にある情報を参照できる。

なお、日本語訳語付与の方式には、HT³-Jの日英翻訳支援と同様、5つのバリエーションが準備されている。翻訳者が好みにあわせて選択することができる。

3.4 HT³-E [英文章中の用語チェック]

HT³-Eでは英語原稿中の用語が用語辞書の標準表記で表記されているか、異表記かをチェックし、必要により異表記を標準表記に置換することができる。統計も取れる。

「用語チェック」機能について使用手順を簡単に説明する。基本は3.2 HT³-J [日本語文章中の用語チェック]と同様であるが、大文字/小文字、語尾変化への対応が日本語と違う。

英語原稿をMS-Wordで開く。次に、ツールからHT³-Eを立上げる。HT³-Eのメインメニューの「辞書管理」で、使用する電子辞書を選択し、「英文用語確認ボタン」をクリックすると英語用語確認ボタン群が表示される。ボタン群の「用語確認(色分け)ボタン」をクリックすると「色分け」が開始される。指定した用語辞書の標準表記と完全にヒットした文字列は「水色」になる。スペルは標準表記とヒットしたが大文字/小文字が違っていたり、語尾変化している文字列は「黄緑」になる。用語辞書の異表記と完全にヒットした文字列は「ピンク色」になる。異表記にヒットしたが大文字/小文字が違っていたり、語尾変化している文字列は「赤」になる。色分け表示が終了したら、「置換候補表示/置換」をする。この機能はピンクおよび赤に色分け表示された文字列(すなわち水色と黄緑以外で標準表記の候補

を持っている文字列)に対する標準表記を置換候補として表示する。

3.5 用語辞書作成

用語辞書は、作成、保守、辞書管理、流通を考慮し、Excel5.0または7.0上で構築することにした。項目としては、日本語読み、日本語標準表記、日本語品詞、日本語異表記、英語標準表記1、英語1品詞、英語標準表記2、英語2品詞、英語異表記、解説、備考、日本語ソート、英語ソート、入力日、分野・出典、入力者を持つ。

電子辞書の使用目的によっては、全項目を埋めなくても構わない。例えば英語のチェックだけのときは日本語関連の項目には入力しなくて良い。品詞コードを省略した場合は全て「名詞+動詞」の扱いとなる。HT³は電子辞書構築の効率を上げるための色々なツールボタンを持った専用のExcelシートを提供している。既存の用語辞書が表形式で電子化されている場合は、そのデータを貼り付けることも可能である。なお、Excelで構築した用語辞書に直接検索をかけると処理時間が非常にかかるため、検索はデータベースに変換して行う。用語辞書は分野を細かく分けて構築する方が便利である。用語辞書を5つまで優先順位をつけて統合することができる。例えば、日本規格協会(JSA)の「JIS工業用語大事典」をベースにし、その上に自社の用語辞書、最後にその製品独特の用語辞書を統合すると効果が大きいであろう。

今後、JSAを始め辞書出版会社、用語集を持っている会社や団体に、HT³用に既存の電子辞書データを使用できる環境作りを呼びかけて行く必要がある。

4. まとめ

最後にHT³の特長を従来ソフトとの違いも含めて整理しておく。

- 用語、訳語の統一を執筆者やチェッカーの感性に頼らずに機械的にできる。証拠が残る。
- ひとつの電子辞書を日英、英英、日英、英日に共通使用できる。
- 解説表示により、電子辞書を引く必要がなくなる。文章中の文字列で電子辞書にあるものはすべて自動的にすばやく色分けされる。(時間の無駄がない、もれがない)

- その色の種類により、標準表記か異表記か一目で判る。異表記には標準表記が提示され簡単に置換できる。
- 従来の電子辞書では、異表記にヒットしなかったが、HT³では異表記にもヒットするので実用性が高い。
- 優先順位をつけて複数の電子辞書を統合使用できる。
- 異表記情報を持たない電子辞書でもファイル形式を合わせれば、それなりに使用できる。
- 従来のワープロソフトにある文章校正でも用語統一はある程度できるが、作り込んだユーザー辞書を一覧表示したり、グループ内で共有したり、既存の電子辞書の流し込みをしたりすることはできなかった。HT³ではこれら全てを可能とした。

なお、HT³の適用業務としては次が考えられる。

用語を正確に使用すべき業務：

各種マニュアル、参考書、辞書、規則書、
教科書、記事、論文などの執筆とチェック

用語辞書がある場合の翻訳：

日英／英日翻訳とチェック

用語辞書、辞典の編集業務：執筆、監修、編集
現在のところ HT³は Macintosh など DOS-V 以外の
パソコンでは利用できない。

[他社商標表示]

Windows 3.1、Windows 95、MS-Word 6.0、MS-Word
7.0、Excel 5.0 および Excel 7.0 は 米 国 Microsoft
Corporation の登録商標、

Macintosh は米国 Apple Computer Inc. の登録商標、

DOS-V は日本アイビーエム社の登録商標、

Pentium は米国 Intel Corporation の登録商標

問い合わせ先：

(株)日立国際ビジネス

開発グループ 貝島良太

〒113 東京都文京区湯島一丁目3-4

TEL：03-5684-4042 (ダイヤルイン)

FAX：03-5684-4013

E-mail：kai@cm.hbi.co.jp

技術文書の翻訳者とチェッカーに朗報！

これは便利！ユニークなソフト、用語と訳語の統一用ソフト HT³発売



あなたは用語集をお持ちですか？ その用語集を翻訳やテクニカルライティングに十分に活かして
ますか？

原稿中の同義語や異表記に困ってませんか？ 用語集や辞典を引いて出ていないときにイライラし
ませんか？

用語集の訳語を全部使ったと思いますか？ これらのなやみを HT³が解決します。

HT³は日本語 MS-Word のアドインソフトです。用語集は Excel で作ります。Word と、Excel の作表機
能をご存じなら、どなたでもすぐに使えます。

一つの辞書を、日日、日英、英英、英日の四通りに利用できます。また、新しいタイプの辞書引きエ
ンジンにもなります。

HT³を日英、英日翻訳に利用すれば、原稿中の用語が標準表記／異表記にかかわらず、相手言語の標準
表記の訳語を付与します。付与した訳語を翻訳文に取り込めば訳語の統一ができます。付与された訳
語を使用したかどうかは日日、英英でチェックできますので、チェッカーも楽です。本当に役立つ辞
書を作り、それを活用する醍醐味を味わってください。

注：HT³ for MS-Word 6.0 と HT³ for MS-Word 7.0 があります。環境にあわせてお選びください。各@
¥150,000 (税別)

問い合わせ先：株式会社日立国際ビジネス 開発グループ [担当：貝島(かいじま) 金上(かねじょう)]

Tel：03-5684-4042 (直通) fax：03-5684-4013、E-mail：kai@cm.hbi.co.jp、kane@cm.hbi.
co.jp]

インターネット翻訳の王様

日本 IBM 株式会社

1. はじめに

インターネット翻訳の王様（以下、翻訳の王様）は、日本 IBM が開発したインターネットユーザーのための英語から日本語への翻訳ツールです。以下の3つのコンポーネントから構成されています。

- ・ Web ページリアルタイム翻訳（キング）
- ・ 一般テキスト翻訳（クイーン）
- ・ 翻訳機能付き巡回ツール（ナイト）

以下それぞれのコンポーネント及び翻訳エンジンについて説明します。

2. Web ページリアルタイム翻訳：キング

キングは Web ブラウザーと Web サーバーの間で一種の Proxy サーバーとして機能するモジュールです。Web ブラウザーからのリクエストに対して、要求されたページを取得し、そのページの簡単な解析を行い、英語のページであれば翻訳ボタンを Web ページの先頭に埋め込んで返すという処理をします。ユーザーが翻訳ボタンを押すと、そのアクションが新たな Web ページ要求となり、キングにその要求が渡されます。キングは、リクエストが翻訳要求であると、翻訳エンジンを使って日本語に翻訳したページを作成し、Web ブラウザーへ戻します。以上が典型的なキングの動作となります。Proxy サーバーとして機能している利点としては、最近多くなってきたフレームを用いたページも翻訳出来るという点です。現在の著名な Web ブラウザーは直接表示しているページの内容を取り出す各種の API を備えています。残念ながらフレーム内で現在表示されている URL を取得することが出来ないため、この Web ブラウザーの API を用いてページ内容を取得する翻訳システムではフレームに対応できません。通信の安定性などを考えるとこの Web ブラウザーの API を用いた方が楽なのは確かなので、この辺のより一層の標準化が待たれるところです。

キングでの翻訳ページの表示は、基本的に元の英



語ページのレイアウトを保存した形式で表示します。また、日本語のみを表示する日本語モードと段落単位で英語と日本語を併記する併記モードの二つの形式で表示できます。後者の併記モードはある程度英語も分かるユーザーにとっては便利でしょう。現在の翻訳技術では完全な翻訳は残念ながら無理ですが、併記モードであれば日本語を読んでいてよく分からない部分は英語を見ることで理解できるようになります。（あるいは、まったく逆に、英語部分を読んでおいて、分からない表現や単語が現れた時に、日本語部分を見るという使い方もあるでしょう。）

分野別辞書は、最初から以下の6つが標準で用意されています。

インターネット(+コンピュータ)、政治、ビジネス、エンターテインメント、スポーツ、芸術

デフォルトの辞書はインターネット分野を含んでいます。また、自動辞書切り替えという機能もあります。これは、上記の6つの分野毎に URL を登録しておくことで、その URL が読み込まれた時には自動的に辞書を切り替えるというものです。

パフォーマンスは、スピード優先から質優先まで5段階で選択することが出来ます。

スピード最優先の場合は、辞書引きをした程度の結果となります。

文末のスタイルとして、「だ・である」調と「で

す・ます」調の切り替えが出来るのはもちろんですが、それ以外に（これはほとんど遊びの機能にすぎませんが）江戸っ子調や女子大生調などの特徴的な文末表現にすることもできます。

3. 一般テキスト翻訳：クイーン

クイーンは、いわゆる一般テキストを対象とした翻訳エディタです。テキストファイル、あるいは、クリップボードからテキストを読み込んで、翻訳することが可能です。

キングは結果を Web ブラウザーに表示するという仕組みのため、入力文の編集、訳語選択や別訳の選択等は出来ませんが、クイーンを用いることでこれらの機能を使うことが出来ます。訳語の選択結果は学習されて、以後のクイーンの翻訳だけでなく、キングの翻訳においてもそれが反映されるようになります。

あと、クイーンでは、辞書ツールを用いて単語の登録が可能です。バージョン1.1では名詞しか登録できませんが、将来のバージョンでは、他の品詞も追加できるようになる予定です。

4. 翻訳機能付き巡回ツール：ナイト

翻訳ソフトでは異色のツールとして、翻訳の王様には巡回ツールであるナイトが付いてきます。日本の一般ユーザーの多くは、インターネットへはダイヤルアップ環境で接続していると思われます。夜中の通話料の割引時間の内に、見たい Web サイトを巡回しておき、後でゆっくり見るという使い方をしているユーザーも多いことでしょう。このナイトを用いれば、ついでに翻訳までしておいてくれるわけです。

5. パターンベース翻訳エンジン

翻訳の王様で用いられている翻訳エンジンは、弊社の東京基礎研究所で開発されたパターンベースの翻訳エンジンです。当研究所では、10年以上に渡って、機械翻訳を中心とした自然言語処理の研究を行ってきましたが、その経験から、翻訳システムに関しては、訳質向上のためのチューニングを如何に効率よく出来るシステムにするかが重要であるとの認識を持っていました。用例やコーパスを用いた手法は、この観点から有望であり、我々も研究を行って

いますが [1, 2]、残念ながらスピード等の点で現在の PC 上の商用システムに使用するまでには至っておりません。そこで、用例翻訳の利点がある程度取り入れ、なおかつ、適度なスピードで処理が可能なパターンベース翻訳手法 [3, 4] を採用することにしました。

我々のパターンベース翻訳エンジンは、パターンを先に適用して残りを従来の翻訳エンジンで処理するというタイプのシステムではなく、パターンと文法的な規則を同時に処理するタイプのシステムです。また、我々のパターンはソース言語とターゲット言語のペアとなっているので、解析が終了すると同時にターゲット側の構造が生成されているという同期文法 (synchronous grammar) の一種となっています。以下にパターンの一例を示します。

【例】

take:VERB:1 a look at NP:2->VP:1

VP:1<-NP:2 を見る:VERB:1

このパターンベース翻訳エンジンでは、基本文法 (パターン) で扱えない表現をパターンの形式で追加していくことでチューニングの大半が済んでしまうので、訳質向上が比較的容易に出来るという利点があります。

6. おわりに

Web ページの翻訳に関しての問題点の一つは、文の認識です。特にテーブルの中のテキストは、一行で文が終わっているのか、それとも数行に渡って一文となるのかの判断をしないと、変な訳になってしまうことが多いという問題があります。しかし、この判定を自動的に行うのは結構難しいのが現実です。HTML のタグとして、文を示すようなものがあると楽になるので、今後は、HTML にこのような処理を助けるタグを追加していくのも一つの方向でしょう。既に、そのような動きが幾つか出てきているので、今後の進展を注視したいところです。

Web ページの翻訳は、翻訳結果が使い捨てであるという点で、現在の自然言語処理技術をうまく生かすことのできるアプリケーションであると言えるでしょう。このような翻訳/言語処理を応用したアプリケーションを開発することにより、英語中心のインターネットの世界での言語的障壁を多少なりとも低下させる役に立てればと願っています。

その他、この製品に関する情報は、
<http://www.ibm.co.jp/software/internet/king>
でご覧になれます。

参考文献

- [1] Watanabe, H. and Maruyama, H., "A Transfer System Using Example-Based Approach," IEICE Trans. on Inf. & Syst., Vol. E 77-D, No. 2, pp. 247—257, 1994
[2] Uramoto, N., "Example-Based Word-Sense

Disambiguation," IEICE Trans. on Info. & Syst., Vol. E 77-D, No. 2, pp. 240—246, 1994

[3] Takeda, K., "Pattern-Based Context-Free Grammars for Machine Translation," Proc. of 34th ACL, pp. 144—151, 1996

[4] Takeda, K., "Pattern-Based Machine Translation," Proc. of 16th Coling, Vol. 2, pp. 1155—1158, 1996

東京基礎研究所 渡辺日出雄

事務局便り

MT Summit VIの展示会出展案内

Exhibits Coordinator の K. Belvin さんから下記の出展案内が来ました。申込用紙と27ブース(広さ:8'×10')の配置図も事務局に届いています。必要な方は事務局にご連絡下さい。

ブース使用料 会員：US\$ 500. 非会員：US\$ 650.

June 2, 1997,

Re : Exhibit Space

MT Summit VI—"Machine Translation : Past, Present, Future"

San Diego, California, October 29–November 1, 1997

Dear Colleague :

You don't want to miss the chance to exhibit your system at MT Summit VI ! This major international forum comes back to the Americas after six years of rotation across the continents. Activity in the field of machine translation has never been greater. The current demand is surpassing all predictions. As Exhibits Coordinator, I am writing to invite you to take part in this exciting event.

The Summit will bring together more than 400 participants—academic, commercial, and government interests—to view the state of the art from Asia, Europe, and the Americas. This is a unique opportunity for you to reach a preselected clientele of serious MT customers or users from all parts of the world. Many of the attendees will be representing companies and institutions that are looking to incorporate translation software into their working environment or even make MT part of their mission. We know this because of the impressive turnout of serious shoppers at the AMTA regional conference in Montreal last October. The exhibits offer these prospective MT users an opportunity to view and test the full gamut of products available. Given the current level of activity in the field (and the pleasant site that has been chosen for the conference!), a record turnout is expected.

MT Summit VI celebrates the 50th anniversary of machine translation, and the occasion is being honored at a very special site, the beautiful Catamaran Resort Hotel, which lies narrowly wedged between its own private beach on Mission Bay and the Pacific Ocean just a few steps away.

The exhibit area and the meeting rooms will be located together in an area reserved for the exclusive use of the Summit. Refreshments, special events, and balconies overlooking the bay will provide additional enticement to draw the attendees to the exhibit area. In addition, exhibitors will have the opportunity to present theater-style demonstrations as part of the regular conference program. A two-page summary of their demonstration and a half-page summary of their exhibit will be published in the conference proceedings.

→ (表3へつづく)

翻訳支援システムツール TRANSIT

株式会社シュタール ジャパン

自動車、コンピュータ、電気製品、etc. 多種多様な外国製品が氾濫するなか、特に機械製品の取扱説明書、いわゆるマニュアルを日本語に翻訳するニーズが増大しています。

このマニュアルを翻訳する作業には、その分野の技術用語の知識が必要になってきます。この技術用語のボギャブラリーを持っている人を、必要に応じて探し出すのは並大抵の苦労ではありません。この用語を辞書としてコンピュータに入れることができれば、翻訳者は技術用語を訳す煩わしさから開放されます。また一度翻訳されたマニュアルも、その製品の仕様が変更になると改訂されます。これをまた一から翻訳するのは非常に無駄な作業です。前回の翻訳とほとんど変わらない部分が当然ありますし、同じ表現であるべきところを変えてしまう危険性もあります。同じ過ちを繰り返すこともあるかも知れません。既に翻訳された内容を新たな翻訳作業に生かすことができれば、リピートオーダーにはこのうえもない効率の向上になるでしょう。

このようにプロフェッショナルの翻訳者の作業を大幅に支援するツールソフトウェアがTRANSITです。

TRANSITはシュタール社のスイス本社とドイツのグループ会社によって開発され、株式会社シュタール ジャパンが協力して、日本語環境で動作するようにしたソフトウェアです。

このシステムは次のような特徴を備えています。

・フィルタリング機能

多くのDTPやワープロソフトの固有のレイアウト情報とテキストを分離して翻訳作業を容易にします。

・プリトランスレーション機能

過去の翻訳文を参照して、一致する部分を自動的に置き換えます。

・ファジーマッチング機能

マッチ率の設定により、指定の文の固まりの単位で、過去の翻訳文との置き換えを行います。

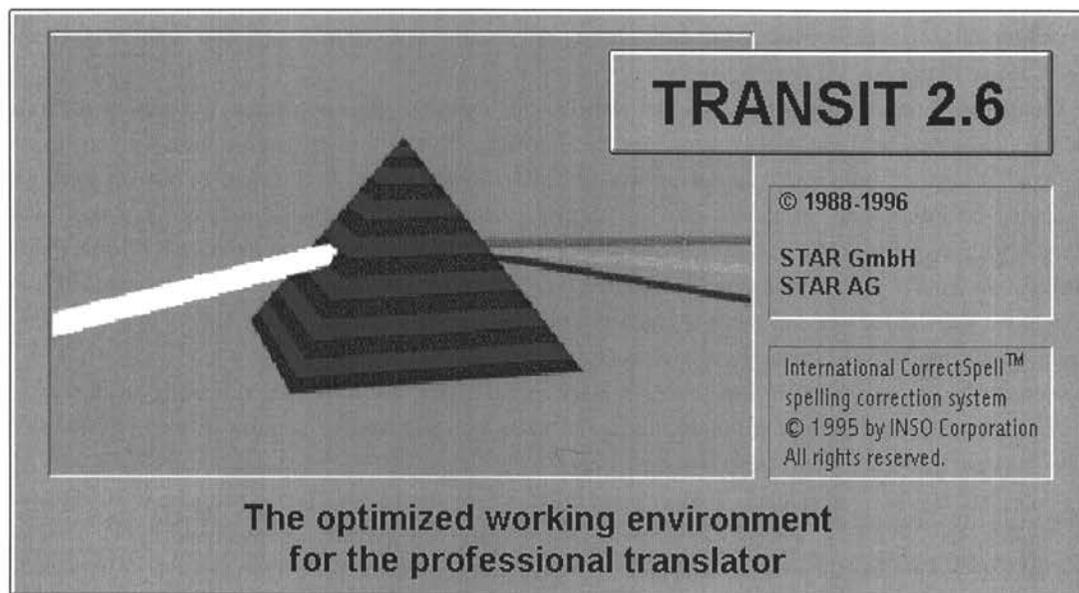


図 1

・マルチ言語

日本語、英語、ドイツ語、フランス語、中国語、韓国語、タイ語、その他、各国多数の言語に対応しています。

・スクラッチパッド

各翻訳部分に、別のウィンドウ画面でメモを残すことができ、見たいときに簡単に呼び出すことができます。

・マルチウィンドウ

翻訳元の言語を表示する画面、翻訳先の言語を表示する画面、辞書画面、メモを表示する画面、ファジーマッチ画面、この5つの画面を同時にしながら作業ができます。

・TermStar

用語を登録しておく辞書を管理します。TRANSITはこの辞書を検索します。辞書はいくつでも作成可能ですし、勿論、変更することもできます。

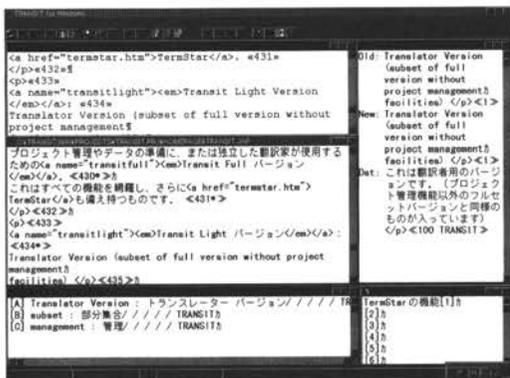


図2 TRANSIT 編集画面

TRANSITはこのように翻訳とDTPの生産性を高め、TermStarとのダイナミックリンクにより翻訳品質を向上させる強力な翻訳支援ソフトウェアです。

<動作環境>

対応機種：IBM PC/AT 互換機 (DOS/V 機)

OS：Microsoft Windows 95, 3.1
または Windows NT4.0

必要メモリ：16MB 以上

(Windows95の場合は24MB 以上)

必要ディスク：80MB 以上

供給媒体：3.5インチFD (1.44MB)

<お問合せ先>

株式会社シュタール ジャパン

〒105 東京都港区芝3-14-13 松田ビル

TEL 03-3457-1549 FAX 03-3457-6957

e-mail: transit@starjapan.co.jp

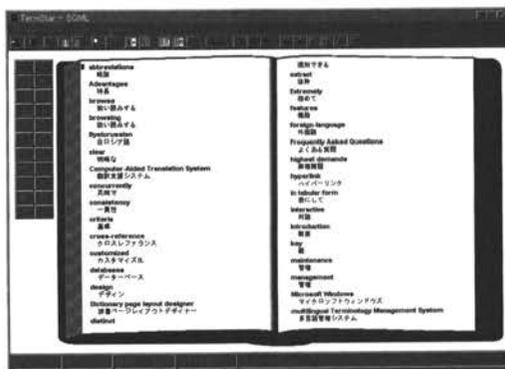


図3 Term Star の画面

事務局便り

AMTA, EAMT もホームページを開設

AMTA, EAMT があいついでホームページを開設しました。

AMTA <http://www.isi.edu/natural-language/AMTA.html>

EAMT <http://www.lim.nl/eamt/>

なお、AAMT のホームページ http://www.jeida.or.jp/aamt/との間で、リンクを張りました。

日英翻訳ツール 「これと英！ Ver. 1.0 for Windows 95」

カテナ株式会社

1、これと英！とコリヤ英和！について

日英翻訳ツール「これと英！」は、低価格英日翻訳ツールの先駆けとして発売された「コリヤ英和！」の姉妹品にあたるものです。インターネット利用による急増した英文翻訳ニーズに対応し、手軽にすばやく翻訳ができる実力と中身にマッチした「コリヤ英和！」という軽妙なネーミングにより、累計出荷本数が40万本を超えました。Windows 3.1版で'95年日本ソフトウェア大賞ビジネスソフト部門「優秀賞」を受賞し、Windows 95版でも、'96年日本ソフトウェア大賞ビジネスソフト部門にノミネートされました。

今インターネットの世界では単に海外の最新情報を入手する段階から、ホームページやメールを通じて自分のメッセージを海外に情報発信する、というような使い方をする方が増えてきました。しかし、いままで日英翻訳ソフトは高価で、個人の方が気軽に手の出せる商品ではありませんでした。

そこで、実用に耐えるソフトウェアを目標に、実際に使う人の身になって考えた結果「これと英！」が完成しました。

「これと英！」は低価格もさることながら、価格に不釣り合いなほどの高精度を誇り、しかも「コリヤ英和！」ゆずりの使いやすさを実現しているまさにユーザが求めている日英翻訳ソフトと言えます。

主な用途としては以下のような場面での使用に適しています。

- ・ 英語の電子メールの作成
- ・ 英文ホームページ作成支援
- ・ 英語マニュアル作成支援
- ・ 英語の学習補助など

2、主な特徴と機能

翻訳方式として「意味トランスファー方式」を採用し、55,000語の基本辞書を使用することで高い翻訳精度を実現しました。

翻訳ウインドウは「コリヤ英和！」ゆずりの使いやすいシンプルなレイアウトです。



図1 翻訳ウインドウ

翻訳方法は「全文一括翻訳」「単文翻訳」「部分翻訳」と用途に応じた翻訳が可能です。

単文翻訳では、文章内の訳語の意味を適切な訳に変更する「訳語置換」や、選択した和文に対応する英文の単語がハイライトで表示される「単語対応」などの編集機能を使って、より精度の高い英文を作成することが可能です。

また、辞書引き機能も搭載していますので意味の知りたい単語を選択して辞書引きボタンを押すだけで簡単に辞書を引くことができます。

基本辞書のほか、ユーザ辞書を作成することが出来ますので、基本辞書にない未知語が出てきた場合や専門的な単語を扱う場合には、ユーザ辞書にその単語を登録することによって翻訳に反映させることが出来ます。さらには挨拶文や慣用句・ことわざな

ど、よく使う定型分を登録しておくことが出来るユーザ定型分辞書が作成可能です。

日本の心というメニューには、英文電子メール作成の際によく使う、お礼・お詫び・四季の挨拶などの定型分が64文登録しており、英文にワンタッチで挿入できます。

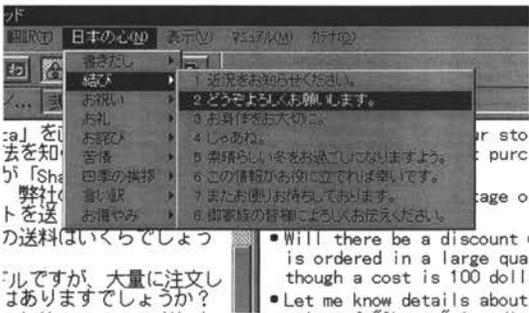


図2 例文メニュー

Windows95のMAPI規格※に対応しているため、翻訳した文章をワンタッチで対応メールソフト(MSExchange・EudoraProなど)にすぐさま送信することが出来るので、いちいちコピー・ペーストをすることなくスピーディに電子メールとして送信することが出来ます。

(※MAPI...MessagingAPIの略。Windows上でメールメッセージを対応アプリケーション間で通信するためのインターフェイス。)

インターネットブラウザやワープロなどでの翻訳に便利なクリップボード翻訳機能を搭載しています。翻訳したい英文を選択し、コピーコマンドを実行するだけで簡単に翻訳することが出来ます。また、翻訳した英文をポップアップウインドウに表示する設定や、翻訳した結果を再度クリップボードにコピーしてもとの日本語にペーストすることが出来る機能

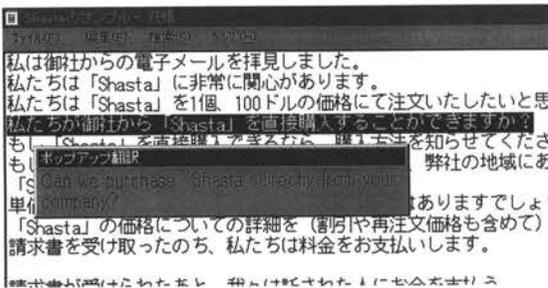


図3 ポップアップ翻訳

もあります。

○動作環境

- 対応機種 : 日本語 Windows95動作機種
- 対応 OS : Windows95日本語版
- 必要メモリ : 16MB 以上
- 必要 HD : 22MB 以上
- メディア : CD-ROM
- 標準価格 : 12,800円

○お問い合わせ先

- カテナ株式会社
- ソフトウェアプロダクツ事業部
- 〒206 東京都多摩市落合1-15-2
- TEL 0423-38-1118
- FAX 0423-38-1079
- URL : <http://www.catena.co.jp/sp/spindex.htm>

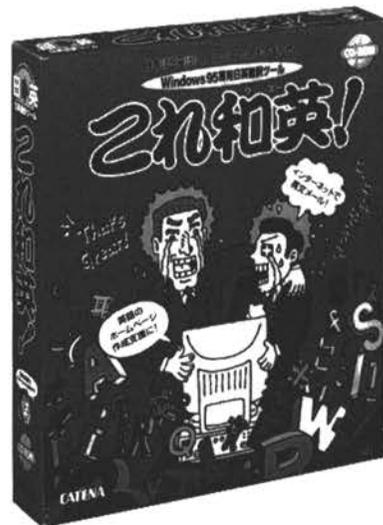


図4 パッケージ

パーフェクト日英 訳せ!!ゴマ

エー・アイ・ソフト株式会社

1. 概要

「パーフェクト日英 訳せ!!ゴマ for Windows」は、高品質な翻訳精度を誇る「ブラビス」シリーズで名高いエムティラボ (株) とエー・アイ・ソフト (株) との技術提携により実現した最も完成度の高い日英翻訳ソフトです。高品質な翻訳を可能にする「用例パターン翻訳」は、単語の訳を並べるだけの従来方式に比べ、より自然できめ細かな英文を作成します。電子メールやビジネスレターの翻訳は勿論、英文ホームページ作成にも威力を発揮します。

2. 特徴

● 英文ホームページ作成支援

レイアウトやリンクは生かしたまま、インターネットの日本語ホームページ (HTML ファイル) を英語に翻訳することが可能です。また、翻訳した結果も同様にレイアウトやリンクを生かしたまま HTML ファイルで保存することができます。

● 日英インターネット・ブラウザ翻訳

現在ブラウザで見ていたページをダイレクトに英訳し、その結果をブラウザ上で見ることができます。

[対応ブラウザ]

Netscape Navigator Ver. 2.0X, 3.0日本語版
Microsoft Internet Explorer 2.0, 3.0日本語版



図1. ブラウザ翻訳

● MS-Word/一太郎に翻訳機能を提供

Microsoft Word Ver.6/95と一太郎 Ver.6上からダイレクトに翻訳することができます。またその他のアプリケーションからも、Drag and Drop機能を使って翻訳することが可能です。

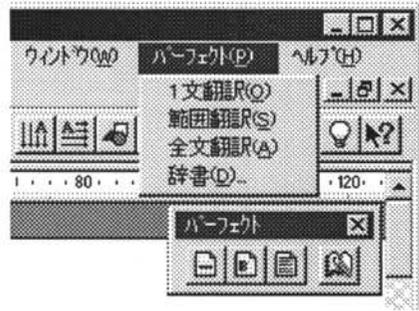


図2. MSワードで翻訳

● クリップボード監視機能

テキストをクリップボードにコピーすると翻訳が自動的に開始され、翻訳結果が再びクリップボードに返されます。これにより、Drag and Drop機能に対応していないアプリケーションでも翻訳機能が提供できます。また、訳せ!!ゴマ上に日本語をダイレクトに入力して翻訳することも、もちろん可能です。

● 実務で使える翻訳品質

高度な用例パターンデータベース (翻訳困難な表現と適切な訳のパターン集) を活用した「用例パターン翻訳」により、電子メールやビジネスレター用表現など、実務に強い翻訳ソフトに仕上がっています。

[翻訳例]

～を私にお送りいただけませんか。

Could you send ~ to me ?

～をご確認のほどお願い致します。

Please confirm ~.

● 専門用語辞書パック

27分野/約20万語の専門用語を1枚のCD-ROMに収録。専門分野の翻訳に力を発揮する専門辞書を低価格で提供しています。

[分野]

医学基礎/外科・整形外科/脳神経・薬理・生理学/内科・小児科/麻薬科・救急医療/薬学/歯科/バイオテクノロジー/自動車/コンピュータ/化学基礎/プラスチック/金属・セラミックス/原子力/物理学/電子・通信工学/機械工学/土木工学/建築工学/電気工学/管理工学/造船工学/航空・宇宙工学/特許/経済/金融・証券/商業・貿易/

● 翻訳速度

1時間に3万語～8万語 (Pentium100 MHz, RAM 24MB, Windows95の場合) の高速翻訳

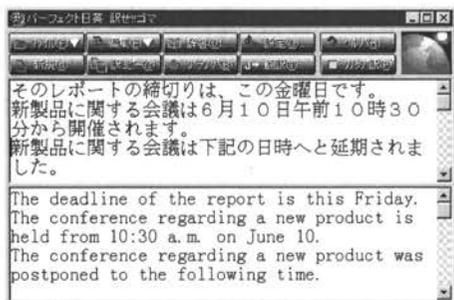


図3. 翻訳エディタ



図4.

3. 動作環境

● OS

日本語 MS-Windows95, NT4.0または
日本語 MS-Windows3.1(16Bit版と32Bit版を同梱)

● メモリ

MS-Windows95→16MB 以上
MS-Windows3.1→12MB 以上
MS-WindowsNT4.0→20MB 以上

● ハードディスク

22MB 以上の空き容量が必要

● 対応ブラウザ

Netscape Navigator Ver. 2.0X, 3.0日本語版
Microsoft Internet Explorer2.0, 3.0日本語版

4. 価格

● パーフェクト日英 訳せ!!ゴマ

標準価格19,800円 (税別)

● 専門用語辞書パック

標準価格14,800円 (税別)

製品お問い合わせ先

イー・アイ・ソフト株式会社

〒151 東京都渋谷区初台1-53-6初台光山ビル3F

TEL 03-3376-7122

FAX 03-3376-5901 (月～金10:00～12:00

13:00～17:00)

本社〒390 長野県松本市中央2-1-27本町第一生命ビル

* Microsoft(マイクロソフト)のロゴ、MS、は米国 Microsoft Corporations の登録商標です。

* Windows は米国 Microsoft Corporations の米国およびその他の地域における登録商標です。

* Netscape Navigatorは、Netscape Communications Corporation の商標です。

* その他の会社名、製品名は、各社の商標、または登録商標です。

(注 1997年4月現在)

『テクニカルコミュニケーション シンポジウム'97』の案内

“かたち、色、音、動き……電子空間でのテクニカルコミュニケーション”

■日時：1997年9月4日（木）・5日（金） 10:00～17:00

■場所：工学院大学（JR 新宿駅から徒歩約4分）

■主催：テクニカルコミュニケーター協会 後援：通商産業省（予定）

■協力：アジア太平洋機械翻訳協会の他、新聞社、出版社、学会など30諸団体

■プログラム：

◎基調講演 「かたち誕生・文字と図像のインタラクション」 杉浦康平 グラフィックデザイナー

◎分科会（パネルディスカッション）

[コーディネーター]

・設計者とライターによるコミュニケーションデザイン

小松原明哲 金沢工大

・マスプロのためのマニュアル作成の標準化 どうつくる?どう使う?

田中祥子 ウインズ

・メディアに適した表現とは?

・Acrobat によるドキュメント電子化～Part 1 その基礎から実際まで～

黒田 聡 ISE

・Acrobat によるドキュメント電子化～Part 2 電子化と配信における課題～

黒田 聡 ISE

・電子空間におけるコミュニケーションマナー～人にやさしい文章表現～

稲垣長利 HTC

・バリアフリーのマニュアル作りを目指して

細山雅一 松下雄器

・アジアで通用するマニュアルを考える

堀江政勝 FLM

・なんでテクニカルライターなの?

雨宮 拓 オフィスクリプタ

◎特別セミナー

[講師]

・ネットワーク社会におけるマルチメディア著作権

藤原宏高 ひかり総合法律事務所

・マルチメディアタイトルの制作現場から

森田祐治 富士プレゼンテック(株)

・Word で作るドキュメント

井上幸司、北村典子、黒沢 勝

～Word ドキュメントとテンプレート制作の事例紹介～

マイクロソフト

・実践テクニカルライティング講座

倉本義介 (株)シーディーエス

・ホームページにみる色・かたち・空間

木寺士郎(ウェブプロデューサー)

・インターネットを翻訳に活用する

(株)コスモ・インタラクティブ

安藤 進 (トランスレーター)

青山学院大学理工学部

◎事例研究・発表（一般）

・機械翻訳の効果的な活用事例 ・インターネット時代の自然言語処理 等々

◎展示会・情報プラザ

・展示会……マニュアルコンテスト受賞作品（これまでの受賞作品も展示）

機械翻訳等マルチメディアコンテンツ

・情報プラザ…テクニカルコミュニケーションに関する書籍や参考資料等の出品

発表者からの作品やツール等の出品 TC 協会活動に関する作品や資料

◎参加費用：入場券（2日間） TC 協会会員 15,000円、非会員 20,000円、学生 31,000円

特別セミナー券 TC 協会会員 10,000円、非会員 15,000円

◆問い合わせ先：TC 協会事務局 TEL03-3368-4607 FAX：03-3368-5087

以上

社団法人日本翻訳連盟の平成9年度事業計画（案）

社団法人日本翻訳連盟(JTF)は、本年度の事業計画(案)を理事会でまとめた。その概要を連絡する。

- 重点課題 1) インターネット・ホームページの運営展開ならびに機関誌等の充実など情報発信の充実化
2) JTF<ほんやく検定>の内容の充実と拡大
3) 会員拡大

1. 翻訳事業に関する調査研究

重点課題1) の取組みとして翻訳業界内外の効果的情報収集とその提供を推進する。

2. 翻訳事業に関する研修会及び講演会等の開催

(1) 月例研究会の開催

「JTF 翻訳環境研究会」を年10回実施し、内容をさらに充実する。

(2) シンポジウム・つどいの開催

1) 「JTF 翻訳祭」(東京)

「国際コミュニケーションフォーラム」を総合テーマに、外国語による異文化理解の基本的な考え方・手法・留意点などの啓蒙を目的に、プロの翻訳者による事例を使った講演と翻訳支援環境をテーマとした展示会からなるシンポジウムを秋に実施。また、懇親会を含むイベントを同時期に開催。

2) 「JTF 翻訳のつどい」(大阪)

今日的テーマのセミナー形式のシンポジウムと懇親会で構成されるイベントを2月に実施する。また関西地区を中心に研修会・講演会を含む交流会を数回実施する。

3. 翻訳事業に関する人材育成および資格能力審査の実施

(1) JTF<ほんやく検定>の実施

年2回、東京・大阪およびNIFTY上の特設会場で実施。仙台、名古屋、福岡等に新設会場を増設することを推進する。

(2) JTF<ほんやく検定>直前対策セミナーの実施

実用レベル(全6分野)と基礎レベルを各2回実施する。

(3) 解説集の刊行

JTF<ほんやく検定>各回につき、解説集を、年2回刊行する。

(4) その他検討

通信講座、評価サービス、合格者インターン制度、中国語などアジア言語の追加などを検討・推進する。

4. 翻訳事業に関する情報収集および提供

(1) 刊行物の編集発行

1) 機関誌「日本翻訳ジャーナル」

顧客の情報も含め、連盟ならではの情報を盛り込み、毎月一回会員に無料配付する。

2) 専門誌「JTF 翻訳研究」

翻訳に関する専門性の高い研究紀要として年一回刊行する。

(2) JTF ホームページの運用展開

JTF 翻訳ネットワーク、産業翻訳フォーラムなどの充実、他誌との提携の展開によって、当連盟の告知、インタラクティブな情報活用の展開を図る。

5. 翻訳事業に関する内外関連機関との連絡及び協調

当連盟事業に協力の意向をもつ国内外の関連団体に焦点を合わせ事業展開を図る。

6. 知的所有権保護事業

二次的著作権をテーマとした研究会を一回実施する。

7. 表彰事業

表彰委員会を設置し、制度・内容の見直しを早急に行い、年度内もしくは総会懇親会時に表彰する。

8. 会員拡大他

賛助会員、法人会員、個人会員の入会促進を精力的に推進する。

以上

イベント紹介

JTF 翻訳環境研究会

年間カリキュラム

ソースクライアント、翻訳関連業界経営者、有能翻訳者の厳しい提言から
ニューウェーブ時代の翻訳サービスで成功する鍵を探り当てる

- 日 時：第1回 6/17(火) 「契約している翻訳者とはこんな人<現役契約者と共に明かす“稼げる産業翻訳者”の優れた特徴>」
- 第2回 7/15(火) 「人間ネットワークを忘れたインターネットは滅亡する！」
- 第3回 8/19(火) 「各翻訳会社こそ自社のインターネットサーバを活用すべし」
- 第4回 9/16(火) 「発想の転換と果敢な決断が翻訳会社の発展の鍵である」
- 第5回 10/21(火) 「“デジタル翻訳者”こそ新時代の産業翻訳のスターである」
- 第6回 11/18(火) 「稼げる翻訳者は、コンピューター環境で差をつけている」
- 第7回 12/16(火) 「最新の情報技術を駆使して年収千五百万以上を確保しよう！」
- 第8回 1/20(火) 「新進翻訳者に活躍の場を用意して数兆円の翻訳需要に応えよう！」
- 第9回 2/17(火) 「あなたの翻訳スタイルは間違っていないか」
- 第10回 3/17(火) 「日本の翻訳業界はこのままでは国際的競争から脱落する！」

■場 所：翻訳会館 東京都港区赤坂8-5-6 (03-3475-5811)

■定 員：30名。申込み順に登録。

■参加料：JTF会員 3,500円 (AAMT会員はJTF会員扱い)

(各1回) 一般 5,000円

■お問合せ：社団法人 日本翻訳連盟 事務局

お申込先 〒104 東京都中央区八丁堀2-8-1 牧野ビル3F

TEL 03-3555-2905 FAX 03-3552-1784

→ (P43からつづく)

As a new feature, the conference registration form includes a box for participants to check if they would like to schedule meetings with software exhibitors during the event. You will be given the list of registrants who have checked this box well in advance. In addition, we can provide you with fliers to mail out to your customers and, upon request, passes to the exhibit area.

I am enclosing the booth space application form and a map of the available exhibit spaces. There are only 27 booth spaces available. It is our policy to allocate spaces on a first-come first-serve basis, so be sure to respond quickly. Full payment must be remitted by August 1 if you wish your exhibit to be included in the program and Proceedings.

Exhibit setup will start at noon on Wednesday, October 29, to allow time to prepare for the opening reception at 6 : 30 p. m. Takedown will be on Saturday, November 1, from 3 : 00 to 5 : 00 p. m, the same time that the conference sessions end.

I will be present at the conference to assist the exhibitors. In the meantime, please feel free to contact me at the coordinates indicated on the first page if you have any questions. I look forward to seeing you there!



Kimberly Belvin
Exhibits Coordinator

連絡先

EXHIBITS COORDINATOR KIM BELVIN 659 DELL STREET SOLANA BEACH, CA 92075 PHONE : 619/481-8446 FAX : 619/350-8613 E-MAIL : kbelvin@ucsd.edu

AAMT
ジャーナル

No. 19
(June 1997)

発行 アジア太平洋機械翻訳協会
所在地 〒105 東京都港区芝公園3-5-12 芝公園真田ビル3階
TEL: 03-5473-7135/7136 FAX: 03-5473-0569
E-mail: KYN02317@niftyserve.or.jp (NIFTY)
ホームページ: <http://www.jeida.or.jp/aamt/>
編集委員会 野村浩郷(委員長) 富士 秀 奥西稔幸 田村真子
事務局 田中 孝 神野千恵子
印刷所 伸光写植印刷株式会社

