



AAMT

The Asia-Pacific Association for Machine Translation

Journal

No.22

April 1998

アジア太平洋機械翻訳協会

目 次

エッセイ	21世紀の機械翻訳に期待するもの	1
成果報告	UPF：異なる機械翻訳システム間でユーザ辞書データを 交換するための標準フォーマットおよび電子環境	2
技術早分かり	用例を用いた翻訳	10
会員投稿	不思議の国のアリスの長文翻訳	14
新製品紹介	● 翻訳サーフィン V4.0	16
	● 日本語でサーチ V1.0	18
	● ロボワード Ver.3.0	20
商品一覧表	機械翻訳関連ソフトウェア一覧表	22
留学生雑感	日本語と韓国語との言語の相違点と機械翻訳における問題点	28
事務局だより	● 1999年シンガポール MT Summit Ⅷの日程決まる	表紙 2
	● 第8回通常総会のご案内	9
	● 協会活動報告	15
	● AAMT 事務局スタッフの交替	19
	● AAMT の E-mail アドレス変更	32
	● 長尾 真先生に祝賀行事	表紙 3
	● 役員・委員の交替	表紙 3

事務局だより

1999年 シンガポール MT Summit Ⅷの日程決まる

機械翻訳に関する国際会議 MT Summit は、2年毎にアジア太平洋 (AAMT)、ヨーロッパ (EAMT)、アメリカ (AMTA) の順で開催されておりますが、AAMT 主催の次回会議の日程が決まりました。

開催期間 1999年9月13日 (月)～17日 (金) の5日間

開催場所 シンガポール (主会場はシンガポール大学)

これは、去る4月17日に、シンガポールに於ける日本側主催者と現地窓口との打合せの結果、正式に決まったものです。

すでに、AAMTでは、辻井副会長を委員長とするプログラムコミッティが発足しておりますが、期間と会場が決まりましたので、これから広報活動を強化していきます。

1年半後に迫ったシンガポールサミットを成功させましょう。

21世紀の機械翻訳に期待するもの

松下電器産業株式会社 取締役 三木 弼一

今世紀も余すところあと2年となり、21世紀を担う技術の開発に世界中が精力的に取り組んでいます。古くから技術開発に取り組み、20世紀に花開いた技術も数多くありますが、機械翻訳技術はその重要性が叫ばれ続けていながら、なかなか実用域に到達していない技術の1つではないでしょうか。

昨今のインターネットの普及やパソコンの普及により、日本人にとっては、世界の情報を知るために機械翻訳ソフトが有効である場面が増えました。しかし、その性能はまだ低く、書かれた内容の概略を知る用途が中心です。インターネットの世界では、現在は英語が広く用いられています。しかしながら、全世界のさまざまな言語を話す人々がインターネットを利用し始めると、さまざまな言語がインターネットの世界を飛び交うようになるでしょう。全世界の人々が円滑にコミュニケーションを図るためにも、機械翻訳技術はますます重要な技術となることは間違いありません。子供達の世界でも、国境をこえたコミュニケーションがさかんになってきていると聞いています。しかしながら、日本人の子供達は、英語が話せないのもので、そのコミュニケーションに参加することができないそうです。英語が話せるように教育していくことはもちろん大切ですが、言葉ができなくてもコミュニケーションがとれるように工夫することも機械翻訳技術の開発にとって重要な要素でしょう。

もう1つ、移動体通信の発展も、機械翻訳の重要性を認識させるものです。携帯電話、PHSが急速な勢いで普及しています。インターネットは主に書かれたテキストを媒体にしていますが、電話は音声はその媒体です。世界中のどこにいても、だれとでも話ができる。すなわち通訳機能が求められるようになるでしょう。さまざまな研究機関で通訳技術の開発が行なわれており、少しずつその成果があらわれてきています。1歩1歩着実に技術成果を積み重ねていくことが大切です。

以上はその用途が比較的明らかなものですが、イ

ンターネット、移動体通信、衛星放送など、デジタルネットワークの世界を流れる知識情報の量は、今後飛躍的に増大し、人手を介して翻訳するという発想では、とうてい処理しきれなくなってしまうでしょう。人による翻訳の精度と機械翻訳の精度はまだまだその差は大きく、21世紀中に機械翻訳が人間に追いつくかどうかは分かりませんが、人と機械とがうまく協調しながら全体の効率を上げる工夫が重要になってくるでしょう。また、21世紀のデジタルネットワーク社会では、従来、機械翻訳技術が目指していた、書かれた文章を翻訳し、翻訳結果も書かれたものとして流通するという状況や、先に例としてあげた通訳装置だけではなく、いままで想像もできなかったようなさまざまな場面で、機械翻訳技術が利用されるようになるでしょう。利用局面を想定し、それに適した技術開発を行なっていくことも、機械翻訳技術の新たな展開のためには必要ではないでしょうか。

21世紀の知識情報社会にとって重要な技術である機械翻訳の技術開発の推進にアジア太平洋翻訳協会の役割は非常に大切であり、世界中の技術者が協力し、1日でも早く実用に耐える翻訳技術を確立されることを期待してやみません。

以上



UPF：異なる機械翻訳システム間でユーザ辞書データを交換するための標準フォーマットおよび電子環境

技術動向調査委員会 委員長 村木一至 委員 亀井真一郎 檜山 努

1. 概要

アジア太平洋機械翻訳協会の技術動向調査委員会では、異なる機械翻訳システム間で「ユーザ辞書」のデータ交換を可能にするための標準となる枠組み（Universal PlatForm；UPF）を世界で初めて開発した。機械翻訳のユーザは、この枠組みを使うことによって、ある会社の翻訳ソフトに登録した対訳ユーザ辞書を別の会社の翻訳ソフトで使うことができるようになる。また辞書作成者は、この枠組みに則って辞書を提供することで多くの機械翻訳ユーザにその辞書データを使ってもらえるようになる。このように、この枠組みを使うことによってユーザ辞書データを広く交換・流通させることができるため、機械翻訳システムの有効活用が促進される。

2. 活動の背景

昨今のインターネットの普及に歩調を合わせ、機械翻訳ソフトウェアに対する注目は急速に高まってきている。パソコン用の翻訳ソフトは年間約50万本が、またパソコンにプレインストール（添付）された翻訳ソフトは年間約200万本が出荷されている。また、インターネットのホームページに掲載された情報を翻訳する必要性から、翻訳ソフトには広範な分野に対応していることが求められている。

機械翻訳システムには、通常数万語から10万語の単語がシステム基本辞書として標準装備されている。しかし実際に利用される単語はユーザの用途に応じて大きく異なるため、標準装備された辞書では全てを網羅することができない。そこで、機械翻訳システムの翻訳品質を高め、その能力を十分に引き出して有効活用するには、ユーザ毎に必要な単語とその訳語を「ユーザ辞書」の形式で蓄積して利用することが必要となる。翻訳時にそのユーザ辞書をシステム基本辞書と共に使用することで適切な訳文を得ることができるようになる。しかしながら、辞書構築は一般に多くの知識（言葉に関する知識、およびそれをどのように辞書に記述するかに関する知識）と

時間と労力を必要とする仕事であり、ユーザ個人個人が辞書を充実させるのには限界がある。また現在までに提供されている各種機械翻訳ソフトはそれぞれ独自の情報記述仕様と辞書形式を持っているため、異なるソフトの間でユーザ辞書交換・利用してユーザ辞書を融通しあって互いの辞書を強化することもできない。このような理由で現在までは、機械翻訳システムの持つ潜在能力が十分引き出されないという問題点があった。

そこで、この問題を解決するため、アジア太平洋機械翻訳協会の技術動向調査委員会では、ワーキンググループ（注1）を作り、異なる翻訳システムを使うユーザ間でもユーザ辞書データを交換・流通させられる仕組み作りに取り組んできた。この活動は、96年度および97年度の2年度にわたって情報処理振興事業協会（IPA）の創造的ソフトウェア育成事業の一環として支援を受けている。

（注1） ワーキンググループ；

NEC、NEC 情報システムズ、東芝、シャープ、富士通研究所、ノヴァ、九州松下、松下電器、インターグループ、日立、沖、キャノン、三洋

3. UPF 構想

辞書が「知識集約型データ」であるという性質を生かして、機械翻訳に有効利用するために、ユーザ辞書データを広く交換・流通させる仕組みを構想した。以下のような環境の開発を行なった。

- （1）機械翻訳システムの機種の違いを超えて、ユーザ辞書データを交換できるようにするためのユーザ辞書共通フォーマット（UPF 言語変換標準）の設計
- （2）共通フォーマットで記述された辞書を蓄積・流通させるための、一般アクセス可能な電子環境の提供

上記（１）の共通フォーマット設計に際しては、言語学的な正確さを求めると同時に、現実のシステムとの互換性を重視した。また対象をユーザ辞書に絞り、辞書作成の簡便さを重要視した。現実には販売され、ユーザに使用されている複数のMTシステム間でそのユーザ辞書の仕様を比較分析して設計を行なうことで、共通フォーマットが現実のシステムから乖離する危険を回避した。さらに共通フォーマットの設定に際しては機械にとっての処理しやすさと共に人間にとっての可読性も重視した。特別なツールなしでユーザが内容を読むことができることを共通フォーマットの条件とし、SGMLで記述することとし辞書のマスタはテキストファイルとした。ただし辞書の構築・参照の効率化のために共通フォーマット専用の辞書エディタを開発した。共通フォーマットの仕様は一般に公開されるので、共通フォーマット辞書と自社システム独自のユーザ辞書の間の変換用に、各社が独自にコンバータを作成できる。

上記（２）の辞書共有環境は、AAMTのWEBホームページ上に構築した。各ユーザは共有環境に置かれているUPF形式の辞書を自分のシステムに取り込んで使用できる。また自分の作成した辞書データをUPF形式に変換して共有環境に置くことで広く他のユーザに提供できる。

このような環境を整えることで、図1に示すように異なる機械翻訳システム間でユーザ辞書のデータ交換が可能となる。

4. 市販ユーザ辞書比較分析

4.1 調査の概要

UPF言語変換標準の作成では、市販の機械翻訳システムからの乖離を避けるために、メーカーが発売している機械翻訳ソフトウェアのユーザ辞書マニュアルに基づきユーザ辞書の仕様の調査を行なった。調査の対象とした機械翻訳ソフトウェアは以下の通りである。

- PIVOT (NEC)、
- ATLAS (富士通)、
- PC-Transer (ノヴァ)
- ASTRANSAC (東芝)
- Duet (シャープ)
- Dr.Surf (九州松下電器)
- Pansee (沖電気)
- HICATS (日立)
- Transporter (三洋電機)

調査は以下を中心に行なった。

(1) ユーザ辞書に登録可能な品詞

調査対象とした機械翻訳ソフトウェアのユーザ辞書に関する資料から、登録可能な品詞に関する部分の調査を行なった。

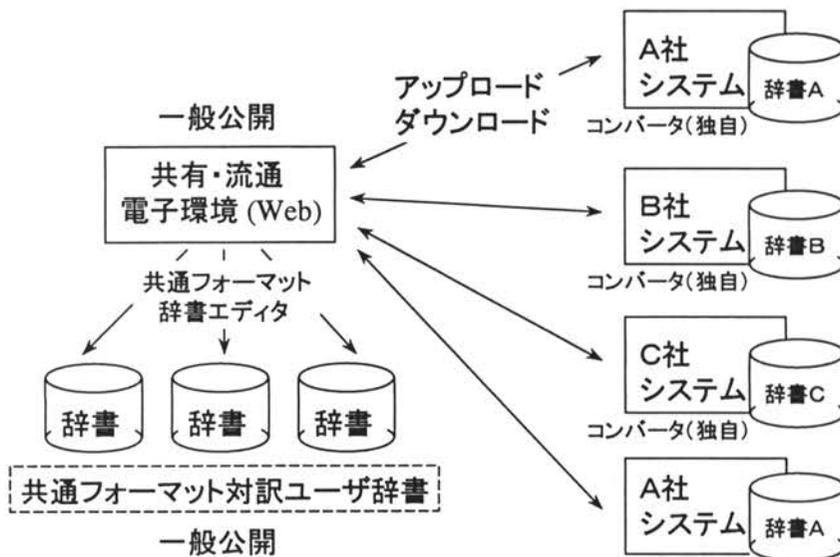


図1 UPFの全体構成

(2) 各品詞毎に必要な情報

調査対象とした機械翻訳ソフトウェアのユーザ辞書に関する資料から、各品詞毎に、必要な辞書項目、およびその辞書項目に記述が許される辞書項目値の調査を行なった。

4.2 品詞の設定

各社のユーザ辞書の登録可能な品詞は表1、2の通りである。

4.3 分析の結果と問題点

各ユーザ辞書の調査分析の結果、UPF言語変換標準の記述仕様を検討する上で、問題点としてあがった項目と、検討結果について述べる。

4.3.1 日本語見出しの登録単位

活用する品詞(動詞(サ変動詞)、形容詞、形容動詞)に関しては、日本語の見出し語の表記は「語幹登録」と「終止形登録」に分かれている。UPF言語変換標準記述仕様では、登録単位としてどちらを採用するか検討する必要があった。例えば、動詞に

おいて語幹登録を行なう場合、五段活用で力行、サ行というように活用する行を明記する必要がある。これに対して、「終止形登録」では五段、一段といった活用の種類のみを明記すればよく、活用する行の情報には必要ではない。

また、終止形登録を行なった方が見出し語が明確に理解される。例えば語幹登録を採用すると、見出し語が「行」である場合、「行く」なのか「行う」なのか見出し語を見ただけでは判断できない。また、「語幹登録」では語幹に切るという判断をユーザに強いることにもなる。

以上の理由から終止形登録が妥当であると考えた。

4.3.2 品詞定義の相違

各社のユーザ辞書で扱う品詞は、同一の品詞名を使用しているも、その内容には相違がある。例えば、「静か(だ)」という語に対して、形容動詞と定義しているソフトウェアもあれば、形容詞の一つとして定義しているソフトウェアがあった。これについては、ユーザにわかりやすい記述仕様を目指すという方針から、機械翻訳に詳しくない人でも親しんで

表1 各社ユーザ辞書に登録可能な品詞(日本語)

Pansee	名詞		動詞	サ変動詞	形容詞	形容動詞	連体詞	副詞		助動詞	格助詞	接続助詞	
PIVOT	名詞		動詞	サ変動詞	形容詞	形容動詞		副詞					単位
ATLAS	名詞		動詞	サ変動詞	形容詞	形容動詞		副詞					
ASTRANSAC	名詞	固有名詞	動詞	サ変動詞	形容詞	形容動詞	連体詞	副詞		接続助詞	機能詞		
Duet	名詞		動詞	サ変動詞	形容詞	形容動詞	連体詞	副詞	接続詞		感動詞		単位
PCTranser	名詞		動詞		形容詞			副詞					
HICATS	名詞												
Transporter	名詞		動詞		形容詞	形容動詞	連体詞	副詞	接続詞				
Dr. Surf	名詞		動詞		形容詞	形容動詞	連体詞	副詞					

表2 各社ユーザ辞書に登録可能な品詞(英語)

Pansee	名詞	動詞			形容詞	副詞	接続詞	前置詞					
PIVOT	名詞	動詞			形容詞	副詞			単位				
ATLAS	名詞	動詞			形容詞	副詞							
ASTRANSAC	名詞	動詞(句)			形容詞	副詞		前置詞(句)			従属副詞句		
Duet	名詞		自動詞	他動詞	形容詞	副詞	接続詞	前置詞		限定詞			
PCTranser	名詞	動詞	自動詞	他動詞	形容詞	副詞	接続詞		単位		関係詞	助動詞	その他
HICATS	名詞												
Transporter	名詞	動詞			形容詞	副詞							
Dr. Surf	名詞	動詞			形容詞	副詞							

いるであろう学校文法を中心とした品詞設定とした。上記の「静かだ」は形容動詞として扱うこととし、よって形容動詞という品詞を設定した。

4.3.3 固有名詞の扱い

固有名詞が単独の品詞として設定されているものと、名詞の下位分類に設定されているものが存在する。固有名詞はユーザ辞書に登録される頻度が非常に高い品詞である。したがって、一つの独立した品詞として設定する方法も妥当性があるが、名詞や動詞といった他の品詞とレベルが違っている（一段細分類化された品詞である）。そこで、他の品詞とレベルを合わせるため、固有名詞を品詞としては設定せずに、名詞の下位分類とした。

4.3.4 サ変動詞（サ変名詞）の扱い

サ変動詞が単独の品詞として設定されているものと、動詞の活用の一つ、あるいは、名詞の下位分類の一つとして扱っているものがある。サ変動詞も固有名詞と同様にユーザ辞書に登録される品詞としての頻度が高い。しかし、動詞の活用の一つとして扱うことが可能である。

サ変動詞についても、他の品詞とレベルを合わせるため、単独品詞ではなく、動詞の活用の一つとして扱うことにした。

このような調査・分析・検討をふまえて、UPF 言語変換標準の記述仕様を決定した。

5. UPF 言語変換標準

5.1 概要

共通フォーマット UPF を介すことで、異システムユーザを含む他ユーザとユーザ辞書データが共有・交換でき、個々のユーザの辞書構築コストが軽減できる。そのためには、UPF 形式から各システムのユーザ辞書への変換（ダウンロード）と各システムのユーザ辞書から UPF 形式への変換（アップロード）との双方向変換が可能となるように UPF を設計する必要がある。そこで UPF では、以下の2種のフォーマットを設定することにした。

(1) 基本言語変換標準（M シート）

全 MT システムのユーザ辞書で取り扱うことができ、UPF との間で相互変換（アップロード・ダウ

ンロード）が可能であることを推奨する語彙情報記述形式

(2) 拡張言語変換標準（K シート）

各 MT システムで記述する可能性のあるすべての語彙情報を記述するための形式

5.2 基本言語変換標準

現実に利用されている複数機械翻訳システムのユーザ辞書で扱える語彙（品詞）には相違があるから、狭い意味で上記の双方向条件を満たすためには、各システムに共通して記述できる語彙（記述可能な語彙の「AND」(=intersection) だけを対象とする仕様が必要である。それが基本言語変換標準である。

基本言語変換標準（M シート）では取り扱う言語情報を制限した。制限に当てはまらない辞書情報は拡張言語変換標準（K シート）で記述することを前提にして、基本言語変換標準ではユーザが辞書登録する時の不要な迷いを減らし、頻繁に辞書登録する場合の辞書作成の効率化を図り、データの流通性に重点を置いた。

基本言語変換標準では登録する品詞を以下の品詞に限定した。

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

また、日英の各品詞間の対応にも制限を設けた。日本語と英語とで、名詞は名詞に、動詞は動詞に、形容詞・形容動詞は形容詞に、副詞は副詞に対応する辞書のみが作成できるように制限した。このような制限を加えることで一般のユーザにとっての辞書作成のわかりやすさと効率化を図った。このような品詞の限定は、MT システムが商用になってから現在まで10年間に、現実に MT ユーザが登録した辞書の約9割が固有名詞を含む名詞であるという各社に共通した経験的データに基づいている。なお固有名詞は名詞の下位分類とし、いわゆるサ変動詞は動詞の下位分類とした。また活用語の登録単位は、わかりやすさを重視して終止形に統一した。また用言の格フレームは簡易なパターン化を行うことでユーザの

不要な迷いを減らし辞書作成の効率化を目指した。
具体的には基本言語変換標準では日本語の場合、
以下の情報を記述する。

[名詞]

種類：普通名詞、固有名詞
意味分類：人、組織、その他具体物、場所、時間、
その他抽象物

[動詞]

活用型：一段、五段、力変、サ変(含：サ変動詞)
格パターン：が、を、に
格要素意味制限
英語の格パターンへのマッピング情報

[形容詞]

(辞書見出しと品詞以外に記述情報なし)
(格要素は「が」のみに制限)

[形容動詞]

活用型：なだ型、のだ型
(格要素は「が」のみに制限)

[副詞]

(辞書見出しと品詞以外に記述情報なし)

以下に基本言語変換標準での記述例を示す。

```
<entry>
<japanese>
  <jentry>    食べる    </jentry>
  <jpos>      動詞      </jpos>
  <jinfl>     一段      </jinfl>
  <jcase>     が, を    </jcase>
</japanese>
<jetrans> ((が=主語; 名詞句; 人;)
(を=目的語; 名詞句; その他の具体物;))
</jetrans>
<english>
  <entry>    eat      </entry>
  <epos>    verb     </epos>
  <eheadpron> vowel   </eheadpron>
  <evpresent> eats    </evpresent>
  <evpast>   ate      </evpast>
  <evpp>    eaten    </evpp>
  <eving>   eating   </eving>
  <ecase>   svo      </ecase>
```

```
</english>
</entry>
```

UPF 言語変換標準は、さしあたり日本語と英語の
2 ヶ国語を対象としているが、その記述形式は上記
のように、日本語部分、日本語英語対応部分、英語
部分とに分離しており、将来の多言語辞書への拡張
への対応を考慮している。

5.3 拡張言語変換標準

拡張言語変換標準 (K シート) では、基本言語変
換標準 (M シート) における登録の制限をなくし、
MT システムのユーザ辞書に登録する可能性のある
辞書情報はすべて登録可能な仕様とした。拡張言語
変換標準は、あらかじめ設定していない新規辞書情
報をユーザが設定して記述できる枠組みを有してい
るところに特徴がある。新規辞書情報は以下の形式
で新規なタグを定義して記述する。

```
<tagdefine>
  <tag_name>   タグ名   </tag_name>
  <tag_descript> 説明   </tag_descript>
  <parent_tag> 親のタグ名 </parent_tag>
  <value_sets> 値のリスト </value_sets>
  <value_sets_descript> 値の説明
                                     </value_sets_descript>
  <tagdefine_comment> コメント
                                     </tagdefine_comment>
</tagdefine>
```

6. 評価実験

6.1 概要

UPF 言語変換標準に準拠した辞書が、実際に各
社の機械翻訳ソフトウェアのユーザ辞書に変換可能
かどうか、また、変換した辞書をユーザ辞書として
用いることにより、翻訳結果にどのような結果を及
ぼすかといった評価実験を行なった。評価実験の対
象とした例文数は、日本語、および英語各 1 0 0 0
文である。機械翻訳ソフトウェアは以下のものを使
用した。

日英

ASTRANSAC for Windows V3.0 (東芝)

ATLAS JE V4.0for Windows95/NT (富士通)
PC Transe/je for Windows Ver4.0 (ノヴァ)
英文名文メイキング (NEC)

英日

ASTRANSAC for Windows V3.0 (東芝)
ATLAS EJ V4.0for Windows95/NT (富士通)
PC Transe/ej for Windows Ver5.0 (ノヴァ)
Power E/J ver2.1 (シャープ)
翻訳アダプター TMCROSSROAD (NEC)

6.2 評価実験の手順

評価実験は、次のような手順で行なった。

(1) 辞書の変換

UPF 言語変換標準記述仕様に基づき、各社独自に各機械翻訳ソフトウェアのユーザ辞書形式に変換するソフトウェア (コンバータ) を作成する。変換する対象とした辞書は、本プロジェクトで開発した UPF 言語変換標準準拠対訳辞書である。

(2) 翻訳実験

用意した日本語・英語各1000文を入力文として、変換した辞書を各機械翻訳ソフトウェアのユーザ辞書 (以下 UPF 辞書と略す) に設定した時の出力結果と、ユーザ辞書を設定しないときの出力結果を得る。

(3) 評価

ユーザ辞書を設定した時の出力結果と、設定しないときの出力結果比較し、○△×の主観評価を行なった。

・評価結果○

UPF 辞書を用いて翻訳した結果、UPF 辞書の訳語が使用されており、翻訳文が同等かそれ以上の品質になった場合。

・評価結果△

UPF 辞書を用いて翻訳した結果、UPF 辞書の訳語が使用されているが、評価対象の訳語が UPF の単語に入れ替わっているため、UPF 辞書を使用する前の翻訳文とは文の構造などが変化しており、多少の問題が生じているが、文の意味は分かる程度で

ある場合。

・評価結果×

UPF 辞書を用いて翻訳した結果、UPF 辞書の訳語が使用されているが、翻訳文の品質が明らかに低下した場合。または、評価対象の訳語がコンバータによって変換されているのにも関わらず、何らかの理由で、翻訳文に使用されず、UPF 辞書の使用前と使用後で翻訳文に変化がなかった場合。

6.3 評価結果

上記のような評価の結果、名詞では90%以上、その他の品詞においても85%以上が評価結果○となった。この数値は、一般にユーザ辞書を使用した翻訳結果の数値と同等以上である。この評価結果から、UPF 言語変換標準は、機械翻訳ソフトウェアのユーザ辞書に変換した際、有効に機能することが実証された。

各翻訳実験の評価結果を表3、4に示す。

6.4 評価、考察

翻訳実験の結果、UPF 辞書を用いた場合、8割以上の翻訳結果が良好との結論を得た。この結果から、UPF 言語変換標準は、複数の機械翻訳ソフトウェア間のユーザ辞書に変換可能な辞書記述形式として、有効であることが実証された。

今回の翻訳実験で、評価結果△あるいは評価結果×のものには、以下のものがあった。

(1) 副詞の位置に関する情報

今回、実験に用いた UPF 言語変換標準準拠対訳辞書は、各品詞毎の情報基本言語変換標準に準拠して記述されている。基本言語変換標準では、副詞生成の位置については情報を記述していないので、問題が生じている。しかし、拡張言語変換標準に採用しているため、この問題は解決できている。

(2) oneself 等の文中の他の要素との照応関係に関する情報

例えば、「tidy oneself up ; 身支度する ; If I am going to visit another company this afternoon, I had better go and tidy myself up.」のように UPF 言語変換標準準拠対訳辞書には“tidy oneself up”として記述されている

表3 日英翻訳実験の結果

	評価結課○	評価結課△	評価結果×	評価不能
ASIRANSAC for Windows V3.0	94.0%	1.0%	4.0%	1.0%
ATLAS JE V4.0for Windows95/NT	90.1%	5.4%	3.9%	0.4%
PC Transer/je for Windows Ver4.0	93.0%	0.0%	6.0%	1.0%
英作文メイキング	92.1%	0.6%	4.3%	3%

表4 英日翻訳実験の結果

	評価結課○	評価結課△	評価結果×	評価不能
ASIRANSAC for Windows V3.0	92.0%	4.0%	3.0%	1.0%
ATLAS EJ V4.0 for Windows95/NT	87.4%	4.4%	6.5%	2.2%
PC Transer/je for Windows Ver5.0	94.0%	1.0%	4.0%	1.0%
翻訳アダプター TMCROSSROAD	97.0%	0.0%	2.0%	1.0%
Power E/J ver2.1	92.9%	1.5%	4.6%	1.0%

が、例文中では“tidy myself up”といったように文中の他の要素と照応して変化するものがある。こういった語が各社のシステムでどのように扱っているかの検討がなされていなかった。これは、今後の課題である。

(3) 個々の機械翻訳ソフトウェアとの相性の問題

UPF 言語変換標準準拠対訳辞書に含まれている単語が、機械翻訳ソフトウェアのシステム辞書に登録されているため、UPF 言語変換標準準拠対訳辞書の訳語が反映されなかったり、あるいは逆に UPF 言語変換標準準拠対訳辞書の辞書情報を採用したために、翻訳品質の低下を招いている場合があった。一般にシステム辞書は、ユーザ辞書より詳細な辞書情報が可能であるため、これは、UPF 言語変換標準の問題というより、ユーザ辞書そのものの扱いかたに関わってくる問題である。辞書変換ソフトを開発する際に、システム辞書に登録済みの単語は、変換しないと工夫が必要になってくる。

7. おわりに

本稿では、これまで行ってきた機械翻訳ユーザ辞書の共通フォーマット設定活動について報告した。活動の成果として、以下のものを AAMT のホームページ上 (<http://www.jeida.or.jp/aamt/index.html>) で公開している。

- ・ UPF 形式の記述仕様

- ・ UPF 形式専用辞書エディタ
- ・ UPF 形式の日英・英日辞書（各2万エントリ；ビジネス分野用語）
- ・ UPF 形式辞書のアップロード、ダウンロード環境

また、以下の市販の翻訳ソフトウェア用の辞書変換ソフトが用意されている。下記の辞書変換ソフトはホームページから自由にダウンロードできるようになっている。今月発売の「Dr.SURF Deluxe」の最新版には、共通フォーマットとの間の辞書変換ソフトが標準装備されている。

- ・ 英作文メイキング (NEC)
- ・ 翻訳アダプター TMCROSSROAD (NEC)
- ・ ATLAS JE V4.0for Windows95/NT (富士通)
- ・ ATLAS EJ V4.0 for Windows95/NT (富士通)
- ・ PC Transer/je for Windows Ver4.0 (ノヴァ)
- ・ PC Transer/je for Windows Ver5.0 (ノヴァ)
- ・ Power E/J ver2.1 (シャープ)
- ・ Dr.Surf Deluxe (メディアヴィジョン、九州松下)
- ・ 翻訳革命 (九州松下)

また、東芝、沖電気工業、日立製作所においても同社の翻訳ソフト用のコンバータを開発中である。

我々の活動の成果が、今後の機械翻訳システムの有効利用に寄与し、ひいては日本人の外国語情報の受発信を促進することを願っている。

参考文献

- 亀井 他 : 「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」情報処理学会第54回全国大会
- 伊藤 他 : 「機械翻訳ユーザ辞書の共通フォーマットの設定ーアジア太平洋機械翻訳協会における活動中間報告ー」言語処理学会第3回年次大会
- 赤羽 他 : 「機械翻訳ユーザ辞書データ流通のための共通フォーマットーアジア太平洋機械翻訳協会の活動報告ー」人工知能学会第11回全国大会
- kamei et al : “Sharable fomats and their supporting environments for exchanging user dictionaries among different MT systems as a part of AAMT activities” MT Summit VI 1997
- 松山 他 : 「機械翻訳ユーザ辞書データ流通・相互利用のための共通フォーマット設定活動ーアジア太平洋機械翻訳協会の活動報告ー」情報処理学会第56回全国大会
- 亀井 他 : 「UPF: 機械翻訳ユーザ辞書の共通フォーマット」言語処理学会第4回年次大会

事務局だより

第8回通常総会のご案内

AAMTの方針を決め、運営状況をご報告する通常総会が間近になりました。会員の皆様には5月上旬に案内状をお届けいたしますが、概略、次のような予定で準備を進めております。

今回は、通常総会のあと、成果報告会、講演会、懇談会など、会員相互の交流、懇親、情報アップを目的とした行事を計画しておりますので、ぜひご参加ください。

1. 日 時 6月17日(水) 13:30~19:00
2. 会 場 機械振興会館 6階会議室
(東京都港区芝公園)
3. プログラムの予定

行 事	内 容	時 間
通常総会	(1) '97年度の事業、決算報告 (2) '98年度の予算、事業計画 (3) その他	13:30~14:15 (45分)
成果報告会	(1) 技術動向調査委員会報告 (2) 市場動向調査委員会報告 (3) ネットワーク翻訳研究会報告 (4) MT Summit、その他報告	14:30~16:20 (110分)
講演会	東京大学、西垣 通教授 「大翻訳時代を迎えて」(仮題)	16:30~17:30 (60分)
懇親会	立食パーティ形式による交流懇談	17:40~19:00 (80分)

なお、講演会は会員外の聴講者は有料とし、懇親会は会員の方からも若干の会費をいただく予定ですが、詳しくは5月中旬にお知らせいたします。

用例を用いた翻訳

日本 IBM(株) 東京基礎研究所 渡辺日出雄

1. はじめに

今回の「技術早分かり」では、用例を用いた翻訳手法について解説します。(第7回と話がダブってしまう点がありますが、ご容赦ください。)

一般に機械翻訳は、ソース言語の解析、ソース言語からターゲット言語への変換、ターゲット言語の生成という3つのプロセスから構成されています。それぞれのプロセスの動作は、規則(ルール)により記述されているというのが普通でした。翻訳処理で一番難しいのは、曖昧さの解消です。規則できれいに記述しただけでは、通常多数の曖昧な解釈が得られます。この中から、もっともよさそうな解釈を選び出さなければならないわけです。従来は、これを更に詳細な規則により制御しようとしていたのですが、場合の数の多さのため規則の数も非常に多くなり人間が適切に制御できる複雑さを超えつつあります。このような状況で、多くの曖昧さの中から、人間の言語的直感に頼らずに、うまく最適な解釈を選び出す方法がないかということで、統計的に解釈を選択する統計ベース手法や、用例を基に選択する用例ベース手法等が考案されてきました。

用例ベース手法は、簡単に言うと、翻訳例を大量に用意しておき、入力文に似ている翻訳例をいくつか選び出し、それらを参考にして翻訳をするということです。これは、人間が行っている翻訳処理と大変似ていると言えます。例えば、

(s1) He bought a car. =彼は車を買った。

(s2) She saw him. =彼女は彼を見た。

という翻訳例を知っているなら、

(s3) He saw a car. =彼は車を見た。

と訳すことは、人間には割と簡単なことです。かなり多くの人が英語の授業を長年受けてきたと思いま

すが、あの長年の授業は我々の頭に多くの用例を溜め込むものだったとも言えると思います。(もちろん、それだけではありませんが。)

用例ベース手法は、80年代初頭に京都大学の長尾教授により提唱された「アナロジーに基づく機械翻訳」というアイデアがベースとなっていて、主に日本がこの分野の研究の中心的な役割を果たしています。80年代の計算機の能力では用例ベース処理を実現するのは困難でしたが、90年代になって計算機の能力の飛躍的拡大に伴って現実的な手法として注目を集めるようになりました。

2. ルールベース翻訳

用例ベース翻訳の良さを理解していただく意味でも、従来からのルールベース翻訳に関して簡単に説明しておきます。もちろん、ルールベース翻訳の全てを詳細に解説するわけにはいかないので、簡単に雰囲気だけでも分かってもらうために、ここでは「かける」という動詞の英語への訳し分けの問題を考えてみましょう。

以下に、「かける」という動詞を用いた幾つかの日本語の例文を載せておきます。これらの文の英訳を考えてみてください。

- (1) 部品に税金をかける
- (2) やかんをストーブにかける
- (3) 絵を壁にかける
- (4) ランプを天井にかける
- (5) 容疑者を裁判にかける
- (6) めがねをかける
- (7) 屋根にはしごをかける

これらの英訳がすぐに分かったでしょうか。以下に翻訳例を載せておきます。

- (1) impose a tax on goods.

- (2) put a pot on the stove.
- (3) hang a picture on the wall.
- (4) suspend a lump from the ceiling.
- (5) take a suspect to court.
- (6) wear glasses
- (7) place a ladder against the roof.

さて、これらの訳し分けを機械に行わせるための規則を考えてみましょう。多分以下のような規則を考えることができますと思います。

- (1) 「を格」が「税金」のようなものだったら、「impose」を使う。
- (2) 「に格」が「ストーブ」のようなものだったら、「put」を使う。
- (3) 「に格」が「壁」のようなものだったら、「hang」を使う。
- (4) 「に格」が「天井」のようなものだったら、「suspend」を使う。
- (5) 「に格」が「裁判」のようなものだったら、「take」を使う。
- (6) 「を格」が「めがね」のようなものだったら、「wear」を使う。
- (7) 「を格」が「はしご」のようなものだったら、「place」を使う。

これらの規則の中で「. . . のようなもの」の部分は実際にはシソーラス（単語を類似なものどうしをまとめて木構造にしたもの）上のコードや意味的な属性コードを使います。これらの規則を見て、「うーん、もっと他の訳し方も見てみないと、ちょっとおかしいかもしれないなあ」と感じられた方も多いと思います。実際、機械翻訳のルールを書くというのは、このような作業の繰り返しによって規則を精密化していくことなわけです。

今は「かける」という動詞について見てみました。他の動詞についてもこのような規則を記述しなければならないわけで、大変な労力のかかる作業であるということが、少しは実感していただけたでしょうか？

3. 用例ベース翻訳

さて、上記のような作業を延々と続けるのは非常に面倒だと思いませんでしたでしょうか？これらの翻訳例をシステムに入れると自動的に適切な訳し分けが出来るようになればいいのになあと思いませんでしたか？そうなんです。これをやろうというのが、用例ベース翻訳なのです。

3.1 用例の表現形式

用例ベース翻訳では、大量の用例を集めておくわけですが、どのような形式で保存しておくのが問題となります。理想をいえば、

This is a pen. = これはペンです。

のようにソース言語の文字列とターゲット言語の文字列のペアの形で持っているということになるのですが、残念ながら、このまま使うというのはまだ可能にはなっていません。

現在可能なのは、もう少し内部的な形式にまで解析した形で翻訳例を持つというものです。例えば、上記の例の場合、

英語：

((root 2 "is" "be" verb)
 (subj 1 "This" "this" pron)
 ((comp 4 "pen" "pen" noun)
 (spec 3 "a" "a" det)))

日本語：

((root 4 "です" "だ" copula)
 (subj 1 "これ" "これ" pron)
 (comp 3 "ペン" "ペン" noun)))

対応関係：

(1 1)(2 4)(4 3)

というような感じになります。この例では、() が一つの単語を表し、(親ノードとの関係、単語のID、単語の表層形、単語の原形、品詞) という並びになっています。対応関係のところでは、(ソース単語ID ターゲット単語ID) となっていて、それぞれの単語が対応していることを表しています。

3.2 マッチングと類似度計算

翻訳する文が与えられた時に、用例データベースにまったく同じものがあれば、そのターゲット部分をそのまま翻訳としてしまえばいいのですが、そういうケースは非常に希なわけです。大抵は入力文の幾つかの部分に対して、類似な翻訳例を探すこととなります。その際、入力文の断片より翻訳例の方が大きいということもありますから、類似な翻訳例の部分を探すということになります。例えば、入力文が

(s4) アメリカが日本からのコンピューターに
税をかけた。

であり、用例データベースに、

(s5) 日本政府は自動車の部品に税金をかけた。
= Japanese government imposes a tax
on goods of cars.

(s6) アメリカからの製品の輸入が増えている。
= Import of products from US is
increasing.

というのがあった場合、入力文の「日本からの」以外の部分が (s5) の「自動車の」以外の部分と類似であり、入力文の「日本からのコンピューター」の部分が、(s6) の「アメリカからの製品」の部分と類似であることとなります。以後、入力文の部分と翻訳例（あるいはその部分）が類似であることを「マッチする」と呼ぶことにします。

この様に、入力文の部分とマッチする翻訳例（あるいはその部分）を探して、幾つかの翻訳例がちょうど入力文全体とマッチするような翻訳例の組を探すわけですが、この様な組合せは当然複数得られますので、どの組合せが一番いいのかを決める必要があります。

ここで登場してくるのが、類似度です。すなわち、翻訳例の組の内、それぞれの翻訳例の類似度の和が最小のものが一番類似な翻訳例の組となるわけです。

さて、類似度はどのように決まるのでしょうか？
基本的には単語の意味的な近さが指標となります。

この意味的な近さは、前述の意味マーカーやシソーラスなどから計算されます。それ以外にももちろん文法的な制約も加味されます。例えば、意味的に近くても名詞と動詞では類似度は低くなるということです。それ以外に、入力と用例の構造的な類似度も反映されます。例えば、上記の (s4), (s5), (s6) のケースで、(s4) の「日本からのコンピューター」に対して、(s5) の「自動車の部品」と (s6) の「アメリカからの製品」のどちらが類似かということになるわけですが、対応する単語（「日本」と「自動車」あるいは「アメリカ」）の類似度や構造的な類似度（「A の B」と「A からの B」の違い）などを考慮して決定されるわけです。

3.3 ターゲット構造の作成

さて、上記の入力文 (s4) に対して (s5) と (s6) がマッチした場合に、どうやって翻訳出力を生成するかというのが次の問題です。(s5) と (s6) のマッチした部分に対応したターゲット言語（英語）部分をつなげてみると、

Japanese government imposes a tax on
goods from US.

というおかしな英語になってしまうことから分かる様に、ターゲット側を修正する必要があります。この現象は、使用している翻訳例は似てはいるけれども入力文と完全に一致しているわけではないことから起こります。そこで、マッチした翻訳例中の単語で入力と同一でないものは、対応するターゲット側の単語を辞書引きをして入力単語の正しい訳語に置き換えてから、ターゲット側を組み合わせるといふ処理をします。上記の例の場合だと、(s5) のターゲット側のマッチした部分

Japanese government imposes a tax
on goods

で、入力文と完全に一致していない日本語部分に対応している単語を正しい訳語に置き換えて、

US imposes customs on computers

となります。次に、(s6)でも同様に、

products from US

が

computers from Japan

となります。ここで、変更されたターゲット側を組み合わせることで、

US imposes customs on computers from Japan.

という訳が生成されることとなります。

この翻訳処理の例を図1に示します。上記の説明では、用例や入力の部分を直線的に表現してきましたが、図1に示すように実際には解析が済んだ木構造の状態で行います。

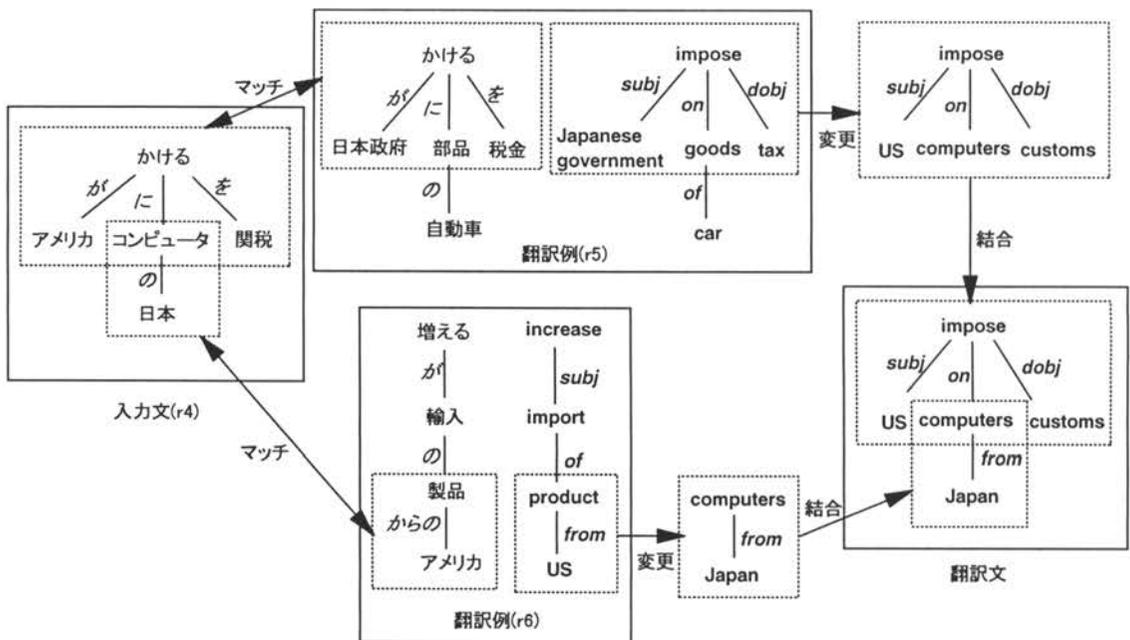


図1

4. おわりに

ここで、用例ベース翻訳手法の意義について考えてみたいと思います。従来のルールベース翻訳システムは、翻訳知識が人間の直感に大いに依存したルールという形で構築されていて、ルール間の相互作用が非常に微妙なレベルで均衡を保って動いているというシステムでした。翻訳ルールの体系を構築した人以外には修正が難しいという問題点があり、効率的なシステムの修正・向上を妨げてきました。(これは、ソフトウェアのソースコードと似た問題で、人の書いたプログラムを維持し修正していくのは非常に大変なのと似ています。)極端な言い方をすれば、従来の翻訳システムは、まだ手作り作業に大いに依

存しているという意味で、産業革命前の手工業の段階にあるという感じでしょうか。本稿で説明した用例ベース手法や以前のこのコーナーで取り上げられた統計ベース手法などは、人間の直感に頼っている部分の置き換えによって工学的なシステム化を目指したものと言えます。この様な手法を大いに取り入れていくことにより、翻訳システムの効率的な開発・向上を見込むことが出来ると思います。

非常に簡単ですが用例ベース翻訳について説明してきました。この手法を用いた実際のシステムを構築するには、パフォーマンス等の解決すべき問題点はまだ幾つか残っていますが、従来のルールベース翻訳を補助する形で今後使用されていく機会が増えていくことと思われます。

不思議の国のアリス中の長文翻訳

会員 翻訳家 佐伯 正剛

はじめに

ここに紹介致します、短・長文一対の文章は、ルイス・キャロル氏の著書「不思議の国のアリス」(Alice in Wonderland, Lewis Carroll)からの抜書きで、この文は上記の著者の中の前後12の長短の詩のうち、17番目に当たります。

- さて、上記の短・長文のうちの長文は、否定語が多い為、全体の意味が把握出来ませんでした。そこで、上記否定語の数が6であることに思い当たりました。これら初歩の数学又は文法によれば、偶数個の否定語を持つ文は、肯定文であることに思い当たりました。
- そこで、本文に戻り関係の部分を読み直したところ、何とこの長文の前に10字前後の文によって、長文に先立って要約文があることに気が付いたのです。この後に、上記長文が書かれているのです。しかも、この短文は勿論否定語のないものでした。
- 従って、著者は先ずこの短文を示し、その後長文が書かれています。このことは、著者はわざわざ前記長文の直前に明解な結論を書き、その後にこの長文を書いたということです。このことは、著者の意図によれば、十歳前後の少女に対して難しい長文で応答していることは、読者を成年と見立てて書いたものと思われまます。
- ところで、筆者が実際に日本語の訳文による図書5冊を読んだ範囲では、著者4名が肯定文として訳されており、残る1名が否定文として訳しておいでです。なお、この内、2名の方々は、口語文ではなく文語文となっております。
- そこで、早速上記短文と長文を紹介します。

(1) 短文：

“Be what you would seem to be.”

(2) 長文：

“Never imagine yourself not to be otherwise than what it might appear to others that what you were or might have been was not otherwise than what you had been would have appeared to them

to be otherwise.”

6. 前記の短文・長文については、この両文章の間に“-or, if you'd like it put more simply-”の挿入があります。

7. 上記について5名による訳文を紹介します。

(1) 石川 澄子氏訳

「見える通りのものであれ」ということ—もっとわかりやすく言い表したければ—「傍目に見える汝と、汝そのものが違わないと空想することなかれ。かつての汝も、それから後の汝も、それより遥かの昔に、傍目には、汝そのものとは似ても似つかぬものと見えしならん汝に他ならざりしなればなり。」と言いました。(以下略)

(2) 北村 太郎氏訳

「見えるままのものであれ。」—もっと簡単に言いたければ—「自分自身を他の人に見えるかもしれないものであり、それはすなわち、おまえがそうであり、あるいはそうであったかもしれないものであり、それはすなわち、おまえがかつてそうであったところのものが、他の人に違うように見えたであろうところのものに他ならない」って言うんだよ。

(3) 高杉 一郎氏訳

「自分が見えたいと思うものになれ。」—もっと簡単に言ってもらいたければ—「自分は、他の人々の目に映っているものと、他のものだと思うな。自分が自分がそうだったこと、あるいはそうだったかもしれないことは、自分がそうだったことが、他の人々の目には、他のように映っていたことと違うわけではない。」

(4) 福島 正実氏訳

「汝の見えたいと思うものであれ。」と言うのよ。もっと簡単な言い方をしたければ、「汝がありしもの、またはありやもしれざりしものは、他人には、汝のありたるものが、彼らにはさにあらず、と見えしものに他ならず、と見えしやもしれざる

ものに他ならず、と汝自身において考えることなかれ」と言うのよ。

(5) 矢川 澄子氏訳

「見える通りのものであれ」ということ— もっと分かりやすく言い表したければ—「傍目に見える汝と、汝そのものが違わないと空想することなかれ。かつての汝もそれから後の汝も、より遥かの昔に傍目には、汝そのものとは似ても似つかぬものと見えしならん汝に他ならざりしなればなり」

あとがき

以上は、5名の翻訳者による、長文38語の訳出を例として掲げましたが、その内4名の方々が肯定文としての訳をしておられます。

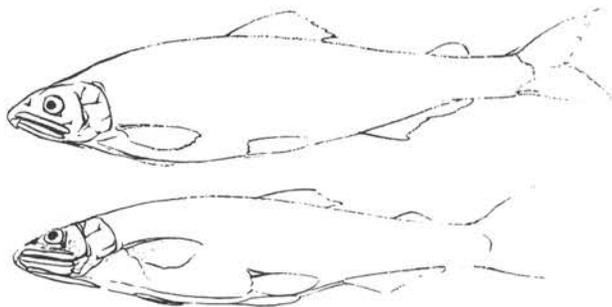
このことは、翻訳を業とされている方々でも否定語を多数含むや長文については、否定語を奇数個または偶数含む場合に、陥り易い比較的単純な誤りをおかすことが、時により有り得ることを示す一例と思います。

事務局だより

協会活動報告

(1998年1月～3月)

理事会	3月30日	①'98年度事業計画案 ②'98年度収支予算案 ③決算理事会および総会日程 ④市場動向調査委員長の交替
臨時理事会	2月10日	①牛尾真太郎氏の理事選任
運営委員会	1月30日	①'98年度事業計画の検討 ②シンガポールサミット企画、その他
	3月2日	①'98年度事業計画の検討 ②総会、成果報告会の企画 ③予算案
	3月19日	①'98年度事業計画案の作成 ②'98年度予算案の作成 ③総会、成果報告会の検討その他
市場動向調査委員会	2月13日	①UPF 活動についてのヒアリング ②来年度の活動方針
技術動向調査委員会	2月13日	①市場動向調査委員会との合同会議とし、UPF 活動のまとめ ②来年度の活動方針
編集委員会	2月23日	①AAMT Journal No. 21の反省 ②No. 22の企画 ③来年度の編集方針
ネットワーク翻訳研究会	2月12日	①ヒアリング(JICST 白木沢佳子氏) ②本年度のまとめと来年度の活動方針
インターネットWG	1月27日	①ソフトウェア一覧表の改訂の進め方 ②'98年度の方針
	2月23日	①AAMT ホームページのメンテナンス ②本年度のまとめと来年度の目標



インターネット英日翻訳ソフト 翻訳サーフィン V4.0

富士通株式会社

1. はじめに

発売以来好評をいただいておりますインターネット連携英日翻訳ソフト「翻訳サーフィン」をこの度V4.0にバージョンアップいたしました。

新製品の「翻訳サーフィンV4.0」は、これまでの「Web ページ翻訳」機能に、英文メールを翻訳する「メール翻訳」、英語のニュースグループ記事を翻訳する「ニュース翻訳」、指定した時刻に指定したページをダウンロード/翻訳を行う「Web 巡回翻訳」の3つの新機能が加わり、さらに使いやすいインターネット翻訳ツールに進化いたしました。

2. 新機能

2.1 Web翻訳

Internet Explorer, Netscape Communicator などのWWWブラウザと連携して、英文ホームページのレイアウトをそのまま日本語に翻訳します。

- ・基本辞書として約15万語の豊富な語彙で精度の高い翻訳を提供します。(語数UP)
- ・Internet Explorer4.01, Netscape Communicator 4.03と連携可能です。(新規対応)



Web翻訳

2.2 メール翻訳(新機能)

Outlook Express, Netscape Messenger などのメーラと連携して、自分宛ての英文メールを日本語に翻訳します。

- ・メール到着と同時に翻訳を実行します。
- ・特定の相手からのメールのみ翻訳を行う指定も可能です。

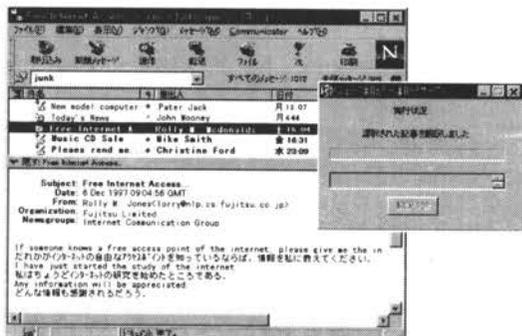


メール翻訳

2.3 ニュース翻訳(新機能)

Outlook Express, Netscape Collabra などのニュースリーダーと連携して、英文のニュース記事を翻訳します。

- ・ニュースリーダーでニュース記事を開くと同時に翻訳を実行します。

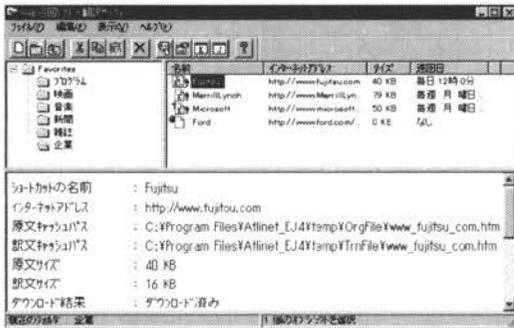


ニュース翻訳

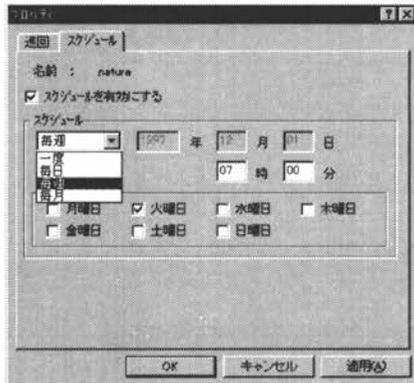
2.4 Web 巡回翻訳(新機能)

あらかじめ指定されたページ (URL) を予約した時刻に自動巡回し、ダウンロード・翻訳します。

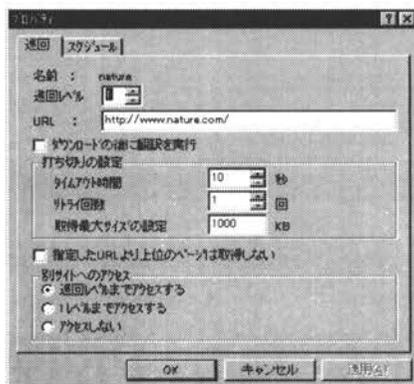
- ・「Web 巡回リスト」でよくアクセスするURLを分類・整理して管理できます。
- ・巡回時刻は、ビデオで番組予約するようなイメージで設定できます。
- ・最大9階層まで自動でリンクをたどりダウンロード/翻訳を行う指定ができます。



Web巡回リスト



巡回時刻の設定画面



巡回条件設定画面

2.5 「日本語でサーチ」連携 (新規対応)

- ・英文ページ検索ソフト「日本語でサーチ」と連携すれば、日本語キーワードで英文ページを検索し、検索結果を日本語で読むことができます。

3. その他の機能

3.1 Web 翻訳

- ・ページ中のヘッダ(見出し)、リスト(項目)、リンク部分のみを翻訳し、翻訳時間が節約できる[アウトライン翻訳機能](#)。
- ・指定した範囲のみ翻訳を行う[部分翻訳機能](#)。
- ・参照しているページのタイトルを翻訳する[タイトル翻訳機能](#)。
- ・クリップボードにコピーされた英文を翻訳する[クリップボード翻訳機能](#)。
- ・キャッシュに保存したページをオフラインで翻訳する[一括翻訳機能](#)。

3.2 チューニング機能

- ・ユーザ独自の利用者辞書が作成可能。
- ・「ATLAS 専門用語辞書 V4.0」の 24 種類の辞書が利用可能。
- ・訳文スタイルが変更可能。
- ・メインウィンドウに表示するツールバー・ボタンの種類、サイズなどが変更可能。

4. 動作環境

動作 OS : Windows 95 または Windows NT4.0

ディスク容量 : 40MB 以上

必要メモリ : 16MB 以上(OS 含む)

必要ソフトウェア : Netscape Navigator 3.X,

Netscape Communicator 4.0X または

Internet Explorer 3.X/4.0/4.01

価格 : 8,800 円(税別)

(「日本語でサーチ」とセットになった「翻訳サーフィン+」は12,800 円(税別))

お問い合わせ先

富士通株式会社 ソリューション販売推進部

〒144-8588 東京都大田区新蒲田 1-17-25

TEL : 03-3730-3136

受付時間 : 月～金(祝日除く)

9:00～12:00/13:00～17:00

URL: <http://jpark.ext.fujitsu.co.jp/ATLAS/>

英文ホームページ検索ソフト 日本語でサーチ V1.0

富士通株式会社

1. はじめに

「日本語でサーチ」は WWW ブラウザ上で、日本語のキーワードを入力することにより海外のホームページを簡単に検索できるインターネットツールです。このソフトを利用することにより、英語の全くわからない方でも、インターネットの膨大な情報の中から、自分の必要としている情報をいち早く検索し、アクセスすることが可能です。

なお、本製品は富士通研究所で試行してきた技術を製品化したものです。

2. 特長

2.1 高精度なキーワードの日英変換

入力された日本語文字列は、基本辞書により検索すべき英語キーワードに変換します。基本辞書には、インターネットでよく検索される地名・人名・ブランド名・施設名など約 15 万語の豊富な単語を用意しています。

2.2 自然文による検索



日本語キーワードは英語に変換される前に単語に分割されますので、自然な文をキーワードにすることもできます。

2.3 強力な検索エンジンと連携

検索エンジンは、世界中で定評のある 8 つの中から選択して使用することができます。

AltaVista / infoseek / Yahoo! / HotBot
Excite / WebCrawler / Lycos / MAGELLAN

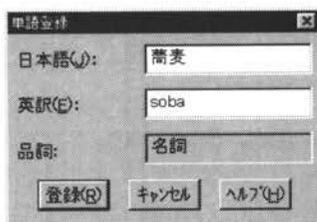
2.4 英訳選択

日本語を英語に変換する際に、訳語が複数ある場合は、複数の英語キーワードから変換候補を選択することが可能です。

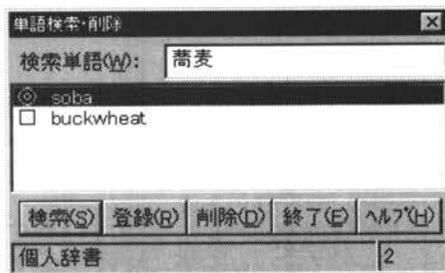


2.5 個人辞書のチューニング

個人辞書に単語を登録することができます。



[単語登録画面]



[単語検索画面]

2.6 インターネット翻訳ソフトとの連携

「翻訳サーフィン」と組み合わせると「日本語でサーチ」で検索した英語ページを即座に翻訳でき、スムーズに海外の情報を収集することができます。その他、以下のインターネット翻訳ソフトとの連携も可能です。

- ・インターネット翻訳の王様 V2.0 (日本 IBM)
- ・EtoJ INTERNET PLUS V4.0 (LogoVista)
- ・Transpad V3.0 for Windows (エー・アイ・ロジック)
- ・Dr.Surf for Windows V2.0 (メディアウェア・イノベーション)
- ・ASTRANSAC for Internet V2.0 (東芝)
- ・インターネットザ・国際人 EJ V4.0 (サンヨー)
- ・PENSEE for Internet V2.0 (沖ワフト)
- ・たちまち翻訳 V2.0 (日立)

※各製品のバージョンは97年12月現在の最新バージョンです。

※各製品名は各社の商標または登録商標です。

「翻訳サーフィン」との連携操作手順

- ①「日本語でサーチ」を起動します。
- ②検索キーワードを入力し、検索ボタンをクリックします。
- ③ブラウザが起動し、検索ページが表示されます。
- ④「翻訳サーフィン」を起動します。
- ⑤翻訳ボタンをクリックし、表示されたページを翻訳します。

3. 動作環境

動作 OS : Windows 95 または Windows NT4.0

対応機種 : 上記 OS の動作する機種

ディスク容量 : 11MB 以上

必要メモリ : 16MB 以上 (OS 含む)

必要ソフトウェア : Netscape Navigator 3.X,
Netscape Communicator 4.0X または
Internet Explorer 3.X/4.0/4.01

価格 : 6,000 円 (税別)

(「翻訳サーフィン」と「日本語でサーチ」
がセットになった「翻訳サーフィン+」は
12,800 円 (税別))

お問い合わせ先

富士通株式会社 ソリューション販売推進部

〒144-8588 東京都大田区新蒲田 1-17-25

TEL : 03-3730-3136

受付時間 : 月～金 (祝日除く)

9:00～12:00/13:00～17:00

URL : <http://jpark.ext.fujitsu.co.jp/ATLAS/>

事務局だより

AAMT 事務局スタッフの交替

AAMT 職員、神野千恵子さんが4月20日付で退職し、後任として木谷 恵 (きや めぐみ) さんが4月1日から勤務しておりますので、お知らせいたします。

事務局の受付時間を朝9時から午後5時半までに広げて、会員の皆様に一層 Quick and Kind Response ができるよう努力して参ります。



ロボワード Ver 3.0

有限会社テクノアウト 第二事業部営業課 高橋信吉

1. はじめに

ロボワードは、辞書・翻訳ソフトにありがちな面倒な入力操作や複雑な処理作業を必要とせず、単語にマウスのポインタを合わせるだけで簡単に言語を訳すことが出来る革新的な翻訳ツールです。

一般の文章翻訳と異なり、ロボワードは単語単位での翻訳となるため、翻訳後に文章が混乱することもなく、また、入力作業を必要としないため、文章を読みながらリアルタイムで次々に単語の翻訳を行うことが出来ます。

2. リアルタイムの簡単翻訳

一般に辞書ソフトで単語を検索・翻訳する場合、対象となる単語のスペルをキーボード、もしくはコピー&ペーストで入力する必要がありますが、ロボワードではそのような面倒な手続きは不要です。調べたい単語の上にマウスのポインタを持っていくだけ…これだけで、即座に単語の訳がポインタの近くに表示されます。また、マウスで単語を指すだけという簡単操作は、マニュアルを読んで操作を熟知するという作業を大幅に軽減します。

3. 多彩な検索方法

ロボワードは、マウスで単語を指すだけの簡単検索以外にも、様々な翻訳機能を有しています。

3.1 前方一致検索

調べる単語の前半部分での検索（ロボ ab と入力すれば、頭が ab の単語を全て表示・翻訳します）

3.2 後方一致検索

調べる単語の後半部分での検索（ロボ ab と入力すれば、語尾が ab の単語を全て表示・翻訳します）

3.3 相似検索

カタカナや、曖昧なスペルで入力した単語をロボワードが判断し、該当すると思われる単語を全て表示・翻訳します。

《※上記の検索・翻訳方法は、キーボードからの入力、もしくはカット&ペーストによる処理となります》

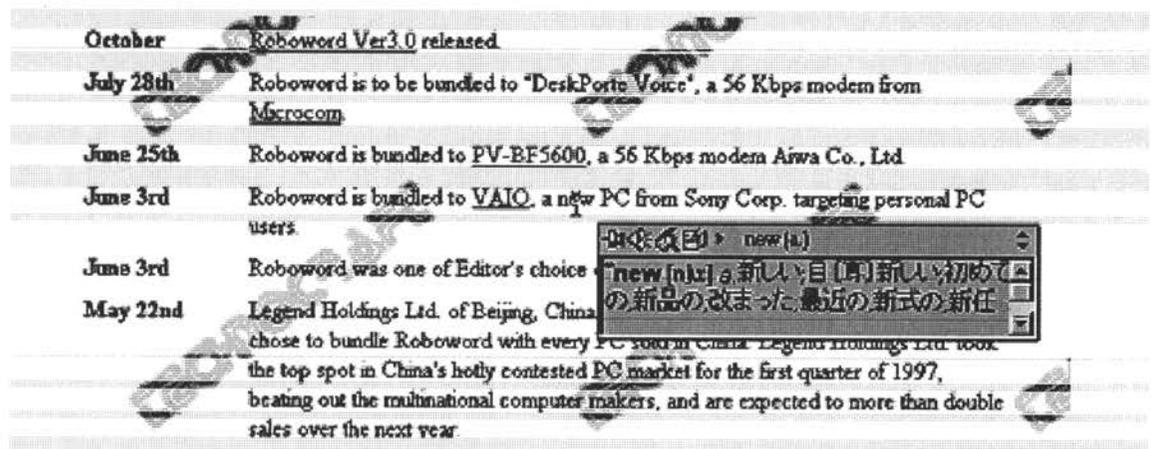


図1 【マウスで単語を指すだけの簡単翻訳】



図2 【多彩な検索・翻訳機能】

4. 辞書拡張機能

ロボワードは、標準で搭載している辞書以外に、各種の辞書を追加することが出来ます。これにより、英和・和英以外に様々な言語や専門用語の翻訳が可能となります。

【現在販売中のロボワード追加辞書】

- 知財翻訳研究所 英和／和英専門用語辞典 標準版（合計約30万語収録）
 - 同プロフェッショナル版（合計約170万語収録）
 - 吉林教育出版社簡体字日中／中日辞典（合計約9万語収録）
- 《以降も、継続して各種の辞書を追加発売していきます》

5. 音声データを収録

豊富な音声データを収録していますので、翻訳以外に、会話やヒアリングの学習にも利用できます。

6. 充実した搭載辞書

ロボワードは用途に合わせ、充実した語彙数と使いやすさを誇る [三省堂デイリーコンサイス英和／和英辞典版] と、豊富な用例と専門ユースにも対応

した [研究社新英和／新和英中辞典版] の2種をご用意致しました。

[三省堂デイリーコンサイス英和／和英辞典版]
（総収録語数約12万語）

価格…¥12800

[研究社新英和／新和英中辞典版]
（総収録語数約16万語用例約19万語）

価格…¥14800

■ロボワード Ver 3.0の動作環境

HDD [三省堂デイリーコンサイス英和／和英辞典版]
25MB 以上（音声データ含む）
[研究社新英和／新和英中辞典版]
40MB 以上、音声データ別途50MB 以上

《以下は共通の動作環境となります》

OS(メモリ) Windows95（8 MB 以上）
Windows NT4.0（16MB 以上）
CPU 486以上（Pentium100MHz 以上推奨）
ビデオ VGA16色以上（256色以上推奨）
ディスクドライブ CD-ROM
（インストール時のみ使用）

【開発元】

有限会社 テクノクラフト

【販売元・お問い合わせ先】

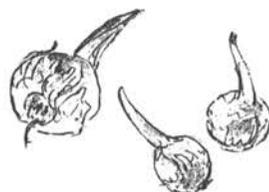
有限会社 サン・オーク

〒5998236 大阪府堺市深井沢町3284安田ビル

TEL 0722-70-6876 FAX 0722-70-6675

http://www.sanoak.co.jp/

(E-mail) sanoak@sanoak.co.jp



機械翻訳関連ソフトウェア一覧表(1)

	ソフトウェア名	対象言語	基本辞書	OS	メモリ
1	訳師如来 英日版	英→日	20万語	Windows95, NT3.51	12MB-
	訳師如来 英日版	英→日	20万語	Mac7.1	10MB-
	訳師如来 日英版	日→英	20万語	Windows95, NT3.51	12MB-
	訳師如来 英日/仏日/独日	英仏独日		Windows95, NT3.51	12MB-
	訳師三尊	英日		Windows95	16MB-
2	パーフェクト日英訳せ!!ゴマ	日→英	12.4万語	Windows95, NT4.0	20MB-
	訳せ!!ゴマ プロフェッショナルツール	日→英	30分野, 21万語	Windows95, NT4.0	20MB-
3	TransLand/JE Ver2.0	日→英	5.5万語	Windows95, NT4.0	16MB-
	TransLand/JE Ver2.0	日→英	5.7万語	MacOS8	12MB-
	TransLand/JE Ver2.0 for EUDORA	日→英	5.5万語	Windows95, NT4.0	16MB-
	TransLand/JE Ver2.0 for EUDORA	日→英	5.7万語	MacOS8	12MB-
	TransLand/JE 専門用語辞書 Ver2.0	日英	38分野, 81万語	ハイブリッド	
4	ワードザ・国際人 EJ	英→日	12万語	Windows95	12MB-
	インターネットザ・国際人 EJ	英→日	12万語	Windows95	12MB-
	ワードザ・国際人 JE	日→英	8万語	Windows95	8MB-
5	WD-01 SW	英→日	10.6万語	ワープロ書院	
	PowerE/J for Text	英→日	10.9万語	Windows	
	PowerE/J for WWW	英→日		Windows	
6	日韓機械翻訳システム	日→韓	10万語	MS-DOS5.0	3.8MB-
	韓日機械翻訳システム	韓→日	10万語	MS-DOS5.0	3.8MB-
7	PIVOT/JE	日→英	9万語	EWS-UX/V	40MB-
	PIVOT/EJ	英→日	7万語	EWS-UX/V	32MB-
	英日翻訳アダプタ II CROSSROAD	英↔日	10万語	Windows95, NT4.0	32MB-
	英日辞書引き君	英和	9万語	Windows	8MB-
	英人名文メイキング (Ver. 1.0)	和英		Windows95, NT4.0	16MB-
8	ATLAS EJ for Windows95/NT	英→日	13.6万語	Windows95, NT4.0	16MB-
	ATLAS JE for Windows95/NT	日→英	13.6万語	Windows95, NT4.0	16MB-
	翻訳サーフィン	英→日	15万語	Windows95, NT4.0	16MB-
	電辞海	英和・和英	32万語	Windows	4MB-
9	Pensee/V 英日版	英→日	7万語	Windows95, NT4.0	16MB-
	Pensee/V 日英版	日→英	9万語	Windows95, NT4.0	16MB-
	Pensee/V 日英版	英↔日		Windows95, NT4.0	16MB-
	Pensee for Internet	英→日		Windows95, NT3.51	16MB-
	Pensee WWW Server	英→日		Solaris1.x	32MB-

1998年4月現在

販売価格 (円)	販売会社	問合わせ先		
		電話	FAX	販売担当
5,825	CSK	(03)5956-9377	(03)5956-9378	生産技術部、商品サポート部 yakushi@mtmng.csk.co.jp http://www.csk.co.jp/products/nyorai/
5,825				
5,825				
14,800				
8,000				
19,800	エー・アイ・ソフト	(03)3376-7122	(03)3376-5901	エー・アイ・ソフト (株) http://www.aisoft.co.jp/
19,800				
39,800	ブラザー	(052)824-3420	(052)825-0311	ブラザー販売 (株) 情報機器事業部 soft.sales@brother.co.jp http://www.brother.co.jp/transland/
69,800				
12,800				
12,800				
29,800				
29,800	三洋インフォメーションビジネス	(06)443-5144	(06)448-4659	営業統括部国内営業部 http://www.sanyo.co.jp/SIB/
9,800				
25,000				
150,000	シャープ			情報システム事業部OAシステム事業部商品企画部
15,000				
15,000				
298,800	日立情報ネットワーク			営業本部NS事業部
298,800				
500,000	NEC	(03)3456-8343	(03)3456-6348	NEC EC推進本部、 汎用アプリケーション事業部 meshplus@apinfo.mt.nec.co.jp http://meshplus.meshnet.or.jp/adp2/
500,000				
9,800				
9,800				
19,800				
58,000	富士通	(03)3730-3136	(03)5462-2344	富士通 (株)、 ソリューション販売推進部 http://jpark.ext.fujitsu.co.jp/ATLAS/
58,000				
8,800				
12,800				
38,000	沖ソフトウェア&オージス総研	(03)5970-9952		沖ソフトウェア (株) 営業部、 オージス総研システムプロダクツ部 プロダクツ営業部 pkg@okisoft.co.jp http://www.okisoft.co.jp/
38,000				
58,000				
9,800				
148,000				

機械翻訳関連ソフトウェア一覧表(2)

ソフトウェア名		対象言語	基本辞書	OS	メモリ
10	X-EJ 2/W	英→日	6万語	Windows	4MB-
	X-EJ	英→日	5万語	MS-DOS3.1	640KB-
	X-DIC/W	英↔日	6万語, 5.5万語	Windows	4MB-
	X-DIC 専門用語辞書	英和・和英	3分野, 3万語	Windows	
11	Logo Vista EtoJ Pro	英→日	13万語	Windows95, NT4.0	16MB-
	Logo Vista EtoJ Pro	英→日	13万語	Mac	16MB-
	Logo Vista EtoJ Personal	英→日	13万語	Windows95, NT4.0	16MB-
	Logo Vista EtoJ Personal	英→日	13万語	Mac	16MB-
	Logo Vista EtoJ Internet Plus	英→日	13万語	Windows95, NT4.0	16MB-
	Logo Vista EtoJ Internet Plus	英→日	13万語	Mac	16MB-
	コリヤ英和 98!	英→日	13万語	Windows95, NT4.0	32MB-
	コリヤ英和	英→日	5.5万語	Mac	7MB-
	これ英和	日→英	5.5万語	Windows95	16MB-
	これ英和	日→英	5.5万語	Mac	16MB-
	QuickTrans View	英→日	13万語	Windows95, NT4.0	32MB-
12	ASTRANSAC Sun WS	英↔日	19万語, 8万語	UNIX	32MB-
	ASTRANSAC C/S	英↔日	19万語, 8万語	UNIX	32MB-
	ASTRANSAC C/S	英↔日	19万語, 8万語	Windows95, NT	32MB-
	ASTRANSAC for Windows	英→日	21万語	Windows95, NT	16MB-
	ASTRANSAC for Windows	日→英	8万語	Windows95, NT	16MB-
	ASTRANSAC for Windows	英↔日	21万語, 8万語	Windows95, NT	16MB-
	ASTRANSAC for Internet	英→日	21万語	Windows95, NT3.51	16MB-
	ASTRANSAC for Internet	英↔日	21万語, 8万語	Windows95, NT3.51	16MB-
	SurfMate	英→日	8万語	Windows95	16MB-
	ASTRANSAC/Web	英↔日	19万語, 8万語	UNIX, NT	32MB-
13	J・London/EJ	英→日	18万語	Windows95, NT4.0	32MB-
	J・London/JE	日→英	18万語	Windows95, NT4.0	32MB-
	専門用語辞書	英日	31分野, 63万語	Windows	
	専門用語辞書	日英	34分野, 67万語	Windows	
	J・ソウル日韓版	日→韓	10万語	Windows95, NT4.0	32MB-
	J・ソウル韓日版	韓→日	17万語	Windows95	16MB-
	WorldNet/EJ	英→日	10万語	Windows95, NT4.0	32MB-
	J・北京	中→日	16万語	Windows95, NT4.0	32MB-

1998年4月現在

販売価格 (円)	販売会社	問合わせ先		
		電話	FAX	販売担当
48,000	スピリット	(03) 3234-8041	(03) 3234-8089	スピリット (株)
29,000				
48,000				
8,000				
97,000	カテナ&ロゴヴィスタ	(0423) 38-1118	(0423) 38-1079	カテナ (株) ソフトウェアプロダクツ事業部, ロゴヴィスタ (株) sp_cs@catena.co.jp http://www.catena.co.jp/sp/ http://www.logovista.co.jp/
97,000				
39,800				
39,800				
12,800				
12,800				
7,800				
9,800				
12,800				
14,800				
6,500				
700,000				
340,000				
340,000				
63,000				
63,000				
98,000				
12,800				
16,800				
8,400				
200,000				
49,800	高電社	(06) 628-8880	(06) 628-2351	本社 http://www.mesh.ne.jp/KODENSHA/
49,800				
29,800				
29,800				
198,000				
198,000				
12,800				
49,800				

機械翻訳関連ソフトウェア一覧表(3)

	ソフトウェア名	対象言語	基本辞書	OS	メモリ
14	Dr. Surf Deluxe for Windows	英→日	43万語	Windows95	32MB-
15	Net Surfer/ej	英→日	13.5万語	Windows95, NT4.0	16MB-
	Net Surfer/ej	英→日	13.5万語	MacOS8	21MB-
	英語 de めーる	日→英	8万語	Windows95, NT3.51	16MB-
	英語 de めーる	日→英	8万語	Mac	16MB-
	PC-Transer/ej	英→日	20万語	Windows95, NT3.51	24MB-
	PC-Transer/ej	英→日	20万語	Mac	10MB-
	PC-Transer/je	日→英	22万語	Windows95, NT3.51	24MB-
	PC-Transer/je	日→英	22万語	Mac	11MB-
	全分野専門語辞書パック	日英	19分野, 95万語	Windows95, NT3.51	
	全分野専門語辞書パック	英日	19分野, 95万語	Mac	
	特許翻訳拡張キット	英→日	2万語	Windows95	
16	Power Translator Pro.	独・仏・西		Windows	
	Power Translator Pro.	独・仏・西		DOS	
	Power Translator Pro.	独・仏・西		Mac	
	Power Translator Pro.	露↔英		DOS	
	Power Translator	独・仏・西		Windows	
	Power Translator	独・仏・西		Mac	
	Power Translator	独・仏・西		DOS	
	Power Translator 中国語	英→中		DOS	
	Power Translator 6.0	独・仏・西・ 伊・英		Windows95, NT	16MB-
17	たちまち翻訳	英→日	6万語	Windows95, NT4.0	24MB-
	HICATS 英日翻訳支援システム	英→日	6万語	UNIX	
	HICATS 日英翻訳支援システム	日→英	6万語	UNIX	

(注1) この資料は、1998年3月から4月にかけて AAMT 事務局が手紙および電話で調査し、回答のあったものを整理掲載しました。

(注2) 掲載順は不同。

(注3) 資料が左右両頁にまたがりませんが通してご覧ください。

1998年4月現在

販売価格 (円)	販売会社	問合わせ先		
		電話	FAX	販売担当
34,800	メディアビジョン	(03) 3222-3908		(株) メディアビジョン http://www.mvi.co.jp/
29,800	ノヴァ	(03) 5287-7570	(03) 5287-7571	(株) ノヴァ sales@nova.co.jp http://www.nova.co.jp/
24,800				
29,800				
29,800				
198,000				
198,000				
198,000				
198,000				
98,000				
98,000				
80,000				
148,000				
148,000				
148,000				
148,000				
58,000				
58,000				
38,000				
28,000				
58,000				
9,800	日立製作所	0120-55-0504	(045) 865-9001	日立オープンミドルウェア WWW-mk@soft.hitachi.co.jp http://www.hitachi.co.jp/soft/trans/
1,000,000				
1,000,000				

日本語と韓国語との言語の相違点と機械翻訳における問題点

東京大学大学院 総合文化研究科 言語情報科学 姜 龍熙 (kang yong hee)

私は'92年に来日以来、最初は言語学を学び、現在は、主に韓国語を日本語に機械翻訳させる際の誤訳について研究を行っている。韓国では、国語国文科（韓国語は私の学部の専攻である）を卒業した。韓国語は私にとっては、母語でありまた研究の対象でもある。

日本語と韓国語の言語間の類似点と相違点に関しては、言語学者の研究も多い。しかし、本稿では、只の相違点と類似点ではなく、機械翻訳を行う際の問題点と弁別特徴 (distinctive feature) について述べたい。

まず翻訳方式としては、両言語の類似点を生かしたダイレクト方式が主流である。私が韓国のある学会で発表した時のことである。私が携ったプログラムとは別に、韓国のある研究所のプログラムが似た機能と分類方法を採用していて、結果的には名称が異なる点以外は殆ど似ていたのが驚いたことがある。このようなことは、盗作とかの問題ではなく、両言語の特徴からその様になるざるを得なかったことの象徴的なことかも知れない。

まず、誤訳のタイプを分けて分類して見ると次のようになると思う。

・韓国語の表記方法の問題

韓国語の中には、日本語と同様に固有語、漢語（韓国では漢字語と言う）外来語などがあり、その表記方法も漢語は漢字とハングルの両方が可能である。しかし、新聞、雑誌の漢字とハングルの割合はハングルの割合が増えている状況である。同じテキストをハングル専用と漢字とハングルの交じり文を機械翻訳にかけて結果を見ると漢字交じり文の方が成功率が高いのは言うまでもない。

日本語の表記で漢語、外来語をひらがなで書いていることを想像してほしい。

ハングル表記は文節単位で分かち書きされていること以外はひらがなのみの表記と同じである。つまり、ハングルのみで書かれている CORPUS を翻訳させる際の一番難しい問題は、言うまでもなく形態分析のレベルの問題である。このような問題は、音声を基本とする日本語から英語への機械翻訳システム（または、日本語から韓国語の音声機械翻訳システム）を考えている方であれば、その弁別が難しいことが分かると思う。

日本語の入力は、入力する者が漢字に変換するのが基本であるので、テキストの上では区別することができる。

しかし、韓国語では、表記の次元では、ハングルのみ書いても問題はない。

・分かち書きの問題

韓国語にも英語のような分かち書きのルールがあるが、段落または紙面の事情によって任意的である。また、英語のように形態素ごとに分かれている形式ではなく、区切り単位は文節である。

ハングル式の表記（日本語の表現の韓国式）

키셔의 키셔가 키셔で 키셔した。

貴社の記者が汽車で帰社した。

きょうの にゅーすは おもしろい。

今日のニュースは面白い。

更に、外来語の表記には、複数の表記が許される例が多い。

例：초코, 초콜렛, 초코렛, 초코레트；

チョコレート

텍스트 ; テキスト、テクスト (逆のタイプ)

その上、新聞の中で分かち書きの誤例を見つけるのは、希ではない。それゆえ、ルールとその変形の形についても対策を練る必要がある。同じく日本語の新聞や雑誌から「ら抜き言葉」を検索してみると、一般言語使用者のルールと決められたルールの間に隔たりがあることに気が付くであろう。

・用言の活用の問題

韓国語の用言には、日本語とは、違ういくつかの問題を抱えている。次は韓国語用言の活用形である。

例：* 교육이야 말로…… : 教育こそ・・・
(교육이야말로……)

	走る	食べる	する	くる	行く	分かる	作る	聞く	手伝う	違う
						ㄹ変則	ㅅ変則	ㄷ変則	ㅂ変則	ㄴ変則
S-1	@달리	먹	@하	오	가	@알	@짓	@ 들	@ 돕	@다르
S-2	달리	먹으	하	오	가	알/@아	@지으	@들으	@도우	다르
S-3	@달려	먹어	@해	와	가	알아	@지어	@들어	@ 도와	@ 달라

図1. 韓国語の用言の語基タイプ

%丁寧	달립니다	먹습니다	합니다	옵니다	갑니다	압니다	짓습니다	듣습니다	돕습니다	다릅니다
%過去	달렸	먹었	했	왔	갔	알았	지었	들었	도왔	달랐
#現在連体	달리는	먹는	하는	오는	가는	아는	짓는	듣는	돕는	다른
#過去連体	달린	먹은	한	온	간	안	지은	들은	도운	다른
#未来連体	달릴	먹을	할	올	갈	알	지을	들을	도울	다를
%現在	달린	먹는	한	온	간	안	짓는	듣는	돕는	다른
%名詞	달림	먹음	함	옴	감	압	지음	들음	도움	다름

図2. 韓国語用言の語形タイプ

ラジオを (聞く: 持ち上げる)

図1の@の印は語幹とそれに付く語尾が分かれていることや文字単位では、簡単に解析できないことを示す。また、基本形が別の用言でも、ある特定の環境において別の同じ文節を成す。

図2%の印は、語幹と語尾の区別が文字単位ではなく、音素単位なので、その処理は、別の扱いをしなければならないことを示す。

例) 라디오를 들어라.

図2の#印は、助詞と一部の用言の連体形が同じ綴であることを示す。則ち助詞に処理される誤訳が生じる語形である。また、単独形(副詞、冠容詞)が用言の連体形と同じ形になる場合がある。

ラジオを (聞け/持ち上げろ)。

cf: 라디오를 듣다: 듣다

- 3) 体言 (V) + 助詞 (V) (와: と; 1、2の逆)
 4) 体言 (C) + 助詞 (C) (과: と; 1、2の逆)
 5) 体言 (R) + 助詞 (로: で; 子音 (R) で終わった場合のみ1) のような選択をする。)

上のようなことを助詞の相補分布 (Complementary Distribution) という。しかし、問題は格助詞の位置及び綴が用言の一部の連体形と全く同じだということにある。

例: 나는 나는 새를 잡았다

- 1 私は、私は鳥を捕った。
 2 私は飛んでいる鳥を捕った。

・形態素の前に来る語の品詞と時制の区別を必要とする。

먹습니다 먹었습니다 : 食べます 食べました
 춥습니다 추었습니다 : 寒いです 寒かったです

「食べる」については、韓国語では過去を含めない形態素であるが、日本語では過去を含む表現になる。「寒い」については、韓国語と日本語の対応は同じである。しかし、韓国語では、一つの形態素であるものが、日本語では、品詞または時制によって訳を選択しなければならない。また、「です」、「ます」の前の語幹は paradigm によって綴が違ってくるのでそれに対応できるプログラムを作らなければならない。

・前にある対象語の有生 (animate) と無生 (unanimate) の判断を必要とする例の場合は、意味マーカ (semantic marker) で全ての名詞に関してその情報を入れなければならない。

例: 있다 있습니다
 いる、ある います、あります
 없다 없습니다
 いない、ない いません、ありません
 책상위에 책이 있다. / 개가 있다.

机の上に本がある。 / 犬がいる。

前にある「本」と「犬」が有生か無生かというのは辞書の中にその情報が入っていればできる。しかし、前にある「机」の情報は要らないので、前にある全ての体言の中で対象を当てること自体が難しいことである。

・共起情報による選択は非常に難しく、韓国語では、区別しない概念を持つ場合がある。更に言語知識だけではなく、言語使用者が特定の感情を表すなどの知識を弁別とする例である。

例: 알다 남다
 分かる、知る、思う 残る、余る

・曖昧性を持っている表現で選択肢の中でどちらを選んでも正解であるが、その決定には前後の語との知識を必要とする語がある。

例: 차를 타다
車に乗る。 : お茶を入れる。

・韓国語の造語形態素と日本語に相当する体系が異なる場合がある。

(-시- ; s-2 に付く; 尊敬
 -겠- ; s-1 に付く; 意志、推測、緩曲)
 먹다 드시다 : 食べる 召し上がる
 가다 가시면 : 行く いらっしゃる
 내가 가겠다, 내일 비가 오겠다
 私が行くぞ、明日は雨が降るだろう

・格助詞の変化がある語型の場合がある。

남자는 사진을 찍었다.
 男は写真を撮った。
 남자는 꽃을 사진을 찍었다.
 男は花を写真に撮った。
 男は花の写真を撮った。
 韓国語では、対格助詞が TOPIC として用いる場合

がある。その時は、一つの文の中で2回使われても非文ではない。

・可能、否定の形態素と活用形の差

개는 하늘을 날지 못한다.

개는 하늘을 못 난다.

犬は空を飛べない。

・動詞によって対応する動詞の形が異なる。

漢語+되다

1>自動詞と他動詞の区別(される)

；開催される。整理される。

2>～になる(本動詞)；先生になる。

3>～する；矛盾する。完成する。

4>～ある；5mもある。

5>～できる(本動詞、補助動詞)整理できる。

・ゼロ格の問題

내 친구 아버지 직업은 학교 선생님이다.
私の友達のお父さんの職業は学校の先生である。

日本語では、属格の数と表示にはあまり制限がない。しかし、韓国語では、1つの文の中で2つ以上あると非文になる。また、支配関係が明確な場合は、一般的に省略可能である。

・複合語の問題：韓国語と日本語の複合語の対応(語順、意味)が合わない例が多い。

例：꿀 + 물 = 꿀물

蜂蜜+水=?蜂蜜水

갈다 + 타다 = 갈아타다

替える + 乗る = ?替えり乗る

・語中に形態素が挿入する。

韓国語では複合名詞の場合、形態素が語中に入る例がある。

빨래 + 줄 = 빨랫줄

洗濯物+紐=洗濯物の紐

複合名詞について全て辞書に登録するかまたは、文字単位ではなく形態素単位が分かるプログラムを作らなければならない。

上に挙げた例以外にも韓国語と日本語の相違点から生じる誤訳が多い。上の例は、簡単に処理できるものもあるし、簡単には処理できないものもある。今の私の夢を述べさせてもらえるのなら、2002年の日韓共同開催のワールドカップの韓国語の文字放送を機械翻訳による日本語放送で見たいということである。まだ夢のような話であろう。

事務局だより

AAMT の E-mail アドレス変更

当協会の E-mail アドレスが、この4月から JDW05467@niftyserve.or.jp に変更となりましたので、お知らせいたします。

(従来の KYN02317@niftyserve.or.jp は3月31日で終了。)

なお、並行して aamtool@infotokyo.or.jp も活用できますので、ご利用ください。



長尾 真先生に祝賀行事

当協会理事（前会長）長尾 真先生の紫綬褒章受章および京都大学総長就任を祝うふたつの行事が2月28日（土）に京都で開催されました。

①教授退官記念講義

当日の午後2時から2時間、京都大学工学部の大講義室いっぱいの聴衆を前に「人間の情報処理を目ざして」というテーマで講義があった。機械翻訳の歴史にもふれながらのロマンとユーモア豊かな話しに盛大な拍手が送られた。

②祝賀パーティ

会場を京都ホテルに移して、当日の午後6時から祝賀会が行なわれた。300名を越える参加者が待つなかに、長尾先生ご夫妻が登場、多くの方々からの祝辞、写真撮影、記念品や花束の贈呈など賑やかな行事がくりひろげられた。

なお、記念論文・資料集として「教授退官記念誌」が発行され、AAMTにも届きましたので、ご覧になりたい方はどうぞ事務局にお立ち寄りください。

役員・委員の交替

次のとおり役員および委員の交替がありましたので、ご連絡いたします。

運 営 委 員 会	松下電器産業株式会社	退任	高木 友博	
		就任	菅野 淳	東京支社技術渉外部 副参事
技術動向調査委員会	株式会社富士通研究所	退任	斉藤由香梨	
		就任	富士 秀	ドキュメント処理研究部
	株式会社日立製作所	退任	森本 康嗣	
		就任	小林 義行	中央研究所 マルチメディアシステム研究部
市場動向調査委員会		退任	委員長 石崎 俊	
			委員長就任	副委員長 坂本 義行

AAMT
ジャーナル

No. 22
(April 1998)

発行	アジア太平洋機械翻訳協会
所在地	〒105-0011 東京都港区芝公園3-5-12 芝公園真田ビル
	TEL : 03-5473-7135 FAX : 03-5473-0569
	E-mail : JDW05467@niftyserve.or.jp
	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 山端 潔
事務局	中島泰雄 木谷 恵
印刷所	伸光写植印刷株式会社

