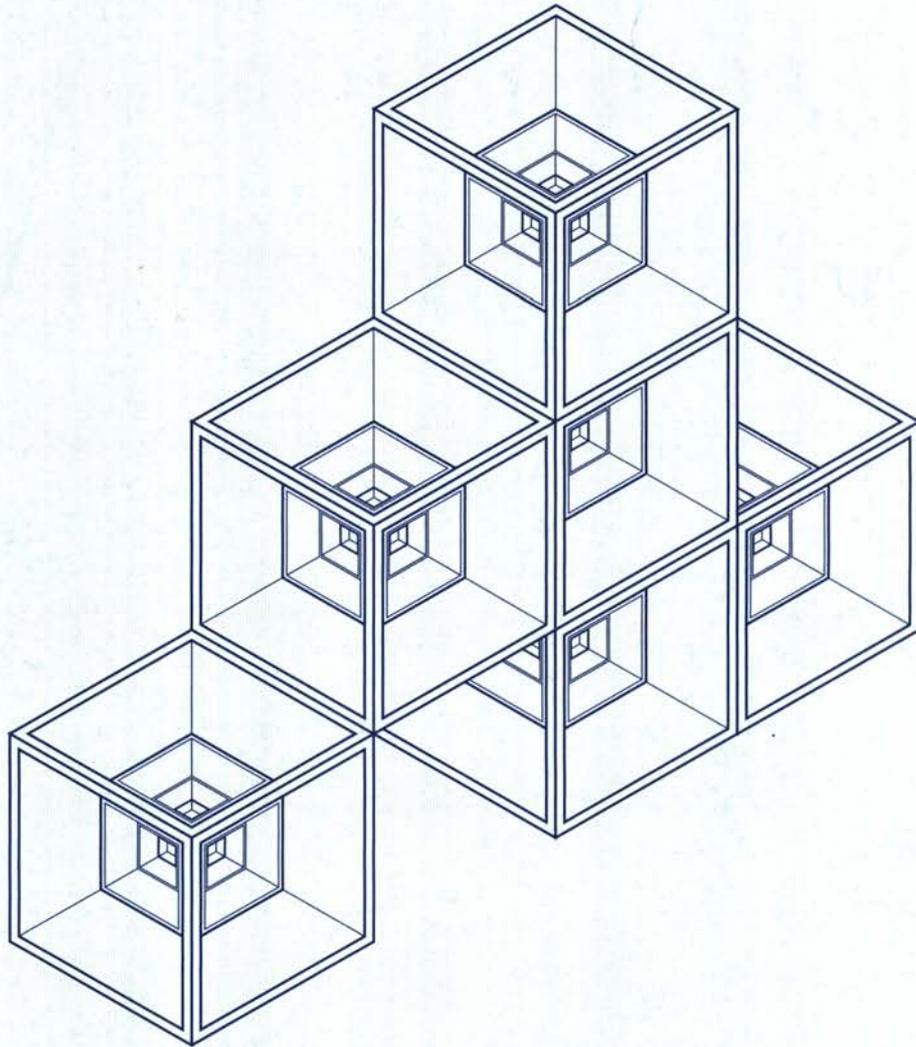


AAMT

Asia-Pacific Association for Machine Translation

Journal



June 2007 *No.40*

アジア太平洋機械翻訳協会

目 次

| | | | |
|-------------|--|-----------------------|----|
| 巻頭言： | 特許翻訳ワークショップへ向けて | 横山 晶一 | 1 |
| 研究報告： | 非線形言語モデルと重文複文の意味類型パターン化 | 池原 悟 | 2 |
| 業界報告： | 翻訳業務におけるMT/TMの利用実態 | 久徳省三 | 10 |
| お知らせ： | 第17回通常総会開催のご案内 | | 14 |
| 最新プロジェクト紹介： | 世界の言語をつなぎ、人と人とをつなぐ言語グリッドの取り組み | | |
| | 稲葉 利江子、村上 陽平、重信 智宏、灘本 明代、石田 亨 | | 15 |
| 新製品紹介 | 「The 翻訳® 2007 ビジネス」 | 東芝ソリューション (株) | 20 |
| 最新プロジェクト紹介： | 「翻訳者を支援するオンライン多言語レファレンス・ツールの構築」(椎茸プロジェクト) について | 影浦 峽、阿辺川 武 | 23 |
| 最新プロジェクト紹介： | 科学技術文献を対象とする日中機械翻訳システム開発プロジェクト | 井佐原 均 | 28 |
| シンポジウム報告： | 『機械翻訳のイノベーションシンポジウム』報告 | 内元 清貴 | 30 |
| MT 関連サービス： | インターネット機械翻訳サービスリスト | AAMT インターネット WG | 32 |
| 活動報告： | 協会活動報告(2007年3月～5月) | | 37 |
| お知らせ | MT SUMMIT XI Workshop on Patent Translation 開催案内 | | 39 |
| 編集後記： | | 宇津呂 武仁 | 41 |

CONTENTS

| | | | |
|------------------|--|--|----|
| Foreword: | MT Summit XI WS on Patent Translation | <i>S. Yokoyama</i> | 1 |
| Report: | Non-compositional language model and semantic pattern dictionary for complex and compound sentences | <i>S. Ikehara</i> | 2 |
| Report: | Actual conditions of CAT tools (MT/TM) in translation work | <i>S. Kyutoku</i> | 10 |
| Information: | 17 th AAMT General Gathering | | 14 |
| New Project: | Combining language resources to overcome language barrier in intercultural collaboration | <i>R. Inaba, Y. Murakami, T. Shigenobu, A. Nadamoto, T. Ishida</i> | 15 |
| New Product: | “The Honyaku® 2007 Business” | <i>Toshiba Solutions Corporation</i> | 20 |
| New Project: | Shiitake Project: Constructing a translation aid system for online volunteer translators | <i>K. Kageura, T. Abekawa</i> | 23 |
| New Project: | Development of Japanese/Chinese MT system for scientific and technological documents | <i>H. Isahara</i> | 28 |
| Symposium: | Report on “The Symposium on Innovation in Machine Translation Technologies” | <i>K. Uchimoto</i> | 30 |
| MT Web Links: | List of MT services on the Web (the Asia-Pacific Region) | <i>AAMT IWG</i> | 32 |
| AAMT Activities: | AAMT Activities (March ~ May 2007) | | 37 |
| Information: | MT Summit XI Workshop on Patent Translation | | 39 |
| Editor's Note: | Message from the Chair of the AAMT Journal Editorial Committee | <i>T. Utsuro</i> | 41 |

特許翻訳ワークショップへ向けて

山形大学大学院、AAMT/Japio 特許翻訳研究会副委員長

横山 晶一

AAMT の中に、AAMT/Japio 特許翻訳研究会（委員長：辻井潤一東大大学院教授）ができて3年以上たちます。総会等で何度か紹介申し上げた通り、この研究会では、特許翻訳に焦点を絞って、1年間に8回の会議を開き、特許翻訳に関するさまざまな問題点について討論し、委員各々の研究成果も含めて、Japio の方から毎年報告書を出しています。また、総会でも、各年度の成果について、簡単な報告を行っており、今年も副委員長の中川裕志東大教授が報告なさる予定です。

ご承知のように、特許の請求範囲や詳細は、長い一文で書く習慣があり、複雑な構造を持っています。また、新規性、進歩性が要求されるだけに、新しい概念や用語が次々に出てきて、即応的に対処する必要があります。世界的には、特許を国際的な知的財産にする方向へ進んでおり、ある国で特許を出せば、世界中で特許として認めようという最終目標に向かって、各国の取り組みや協議が行われている状況です。

このような状況においては、機械翻訳の必要性はいうまでもなく、その質の向上が緊急の課題となっています。上記研究会もこの趣旨のもとで、機械翻訳辞書の研究、特許翻訳における統計的機械翻訳の効果の調査、翻訳用対訳辞書の問題の調査などを行ってきました。今年度も、引き続き、統計的機械翻訳の効果の調査、特許翻訳向け対訳辞書の構築法や曖昧性解消方法の検討を行うとともに、特許文の翻訳における言語学的考察なども行っていきたいと考えております。

2005年9月にタイのプーケット島で開かれた MT Summit X では、本研究会が中心となって、付随するワークショップとして **Workshop on Patent Translation** を開催し、招待講演や一般発表など、30人ほどの参加者でしたが、親密な情報交換と討論を行うことができました。これについてもジャーナルでご報告した通りです。

これらの実績と成果を踏まえて、本年9月にコペンハーゲンで開かれる MT Summit XI においても、本研究会が中心になって、**Workshop on Patent Translation** を開催することを計画し、本会議に提案したところ、他の2つのワークショップとともに認められました。辻井先生と私が共同議長になり、本会議後の9月11日に一日かけて開催される予定になっております。この会議の内容や目的、**call for papers** については、AAMT の HP の下にある <http://www.aamt.info/patent/> にまとめてありますのでご参照下さい。6月15日に投稿の〆切、6月末までに可否の通知、7月20日までに最終原稿の受付というスケジュールで進行予定です。論文は、本会議論文とともに CD-ROM に収録される予定です。

前回同様、招待講演やユーザからの報告なども計画しておりますので、ふるって論文投稿や参加、討論をいただければと思います。コペンハーゲンのワークショップへ向けての皆様のご協力、ご支援を心からお願いする次第です。

非線形言語モデルと重文複文の意味類型パターン化

鳥取大学工学部知能情報工学科 池原 悟

1. 研究の背景とあらまし

機械翻訳では、最近、統計翻訳方式が注目されているが、日英言語のように異なる言語族間で品質の良い翻訳を実現するのは難しい。実用システムの多くは、依然として要素合成法を基本とするトランスファー方式が中心で、統合構造と意味を分離して翻訳するため、原文の意味が失われる点に問題がある。

これに対して、パターン翻訳と用例翻訳は、統語構造と意味を一体的に扱う方式であり、高品質の翻訳が期待できる。このうち、パターン翻訳では、あらかじめ、パターン間の意味的な排他性を考えた大規模なパターン辞書を開発する必要のあるが、用例翻訳は、構造的に一致する対訳用例の中の意味的に類似する単語や表現を置き換えて訳文を得る方法であり、あらかじめパターン辞書を準備する必要はない。しかし、用例中の要素置換の可否は用例毎に異なるため、その判定を自動化することが難しい。この問題を解決するには、各用例に対してあらかじめ置換可能な要素を指定しておけばよいが、その方法は、結局のところパターン翻訳に帰着する。

また、最近、認知言語学のCognitive Grammar (ラネカー等) やConstruction Grammarの方法 (フィルモア等) が注目されるが、構造的な意味の単位を決定する基準が曖昧で、工学的適用が困難である。

ところで、表現構造と意味を一体化した方法としては、すでに、日英機械翻訳システムALT-J/Eにおいて、単文構造を対象とした表現意味辞書「日本語語彙体系」が開発されている。これにより、日英機械翻訳における単文の意味的な訳し分けの問題はほぼ解決できたと見られるが、この辞書は、動詞と名詞を中心とした表現 (客体的表現) の意味を結合値パターンの形式にまとめたもので、助詞や助動詞の表現 (主体的表現) の持つ非線形な意味が失われることが問題であった。また、従来の機械翻訳では、重文や複文のような長い表現の翻訳が特に不得手であるのに対して、これらの文に対する表現意味辞書は存在せず、依然として要素合成法に頼らざるを得ない点が問題であった。

そこで、筆者等は、これらの問題を解決するため、言語表現の構造と意味の關係に着目した「非線形言語モデル」を提案し、重文と複文に対する「意味類型パターン辞書」を開発した。この研究は、科学技術研究機構 (JST) が推進する創造的基礎研究推進事業 (CREST) の一つで、5年計画で実施したものである。

この研究では、日英対訳文 (100万件) から、重文複文の対訳文 (15万件) を抽出し、それに含まれる線形な表現要素を単語レベル、句レベル、節レベルの3段階で汎化することにより、合計22.7万件の対訳パターン辞書を実現した。得られたパターンは、一つ一つが、日本語表現の意味を掬い取る網のようなものである。そこで、重文複文の表す意味の体系を構築し、各パターンに該当する意味コードを付与することによって、パターンとそれによって掬い取られる意味の關係が明らかになるようにした。

以下、本稿では、この研究の概要について紹介する。まだまださまざまな問題を残しているが、既に80%近い被覆率が達成されており、実用的価値が期待される。

2. 言語表現の持つ意味とは何か?

本研究では、構成部分の意味からだけでは全体の意味が説明できないような表現を「非線形表現」という。非線形表現は、それを構成要素に分解する過程で元の表現の意味が失われる。既に述べたように、従来の言語処理における最大の問題は、実際の言語では、非線形表現が多いことである。

このような非線形表現は言語に限らずさまざまな表現に存在する。心理学では、このような現象はゲシュタルトと呼ばれ、古くから主要な研究対象とされてきたが、言語処理では、非線形表現はやっかいなものとして避け、依然として要素合成法を前提とした研究開発が行われている。

ところで、分解する過程で意味が失われると言うことは、「表現の構造が意味を持つ」と言うことであるから、非線形表現の処理に挑戦するには、「表現構造の持つ意味とは何か」について明らかにする必要がある。そこで、まず、

この問題についての本研究の考え方を説明する。

(1) 言語の本質は何か？

まず、「言語の本質は何か？」について考える。この問題はあまりにも初歩的で、今更と思われるかもしれないが、現実には、言語と記号の区別さえ明確に意識されない研究が見られる。

さて、言語は表現の一種であるが、絵画や音楽も表現の一種である。そこで、表現を図1に示すように、「感性的な表現」と「超感性的な表現」に分類する。前者は五感に訴える表現であり、後者は理性に訴える表現である。視覚に訴える絵画や聴覚に訴える音楽はいずれも前者の表現であるのに対して、記号（地図や道路標識等）や言語（自然言語）は、後者の表現である。

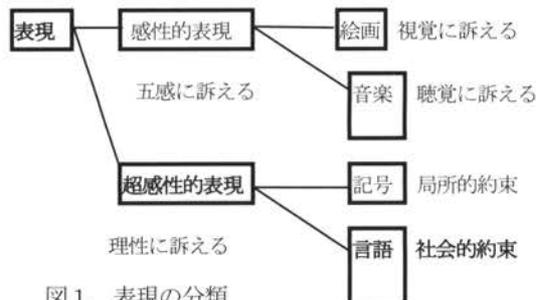


図1. 表現の分類

ここで、問題となるのは、記号と言語との違いである。いずれも、表現と意味の関係についての約束を背景とする表現であるが、記号の場合、約束の適用範囲は局所的であり、原則として使用するに先立って定義を必要とする。これに対して言語の約束は、当該言語集団の中において自然発生的に成立した慣習であり、原則として定義することなしに使用できる。すなわち、記号と言語の本質的な違いは、社会的に共有された約束に支えられた表現であるか否かにあると考えられる。

ところで、言語が社会規範としての言語規範に支えられた理性的表現である以上、言語処理研究の第1の課題は、言語規範の特徴と性質を解明し体系化することだと言える。そこで、言語規範を「意味的な約束」（表現と意味の関係に関する約束）と「文法的約束」（語の並びに関する約束）に分けると、後者は前者の約束に共通した構造上の特長を取り出したものであり、二次的な約束と言える。

従来の言語処理は、二次的な約束を一次的な約束とみなし、本来の一次的な約束の体系化を棚上げしてきた嫌いがある。

(2) 言語表現の意味とは何か？

一次的な約束を体系化するには、言語表現と意味の関係を明確にしておく必要があるが、言語表現の意味については、言語哲学的にも混乱した状況にあり、定説が存在しない。

このような現状の中で、三浦つとむは言語過程説において「関係意味論」を提唱している。この意味論は、「言語表現とそれに対応づけられた話者認識の関係」を表現の意味とするものである。「対象」、「話者の認識」、「表現」、「聞き手の解釈」など言語実体のいずれも意味とはせず、それらの関係を意味としている点でユニークな説と言えるが、「実際に使用された表現に対してのみ定義されること」、「意味は、それぞれの表現に固有で、客観的であり、解釈によって変化するものでないこと」が、計算機処理にとって好都合である。そこで、以下では、「関係意味論」に基づき意味的約束の性質を明らかにする。

(3) 概念とはどのようなものか？

関係意味論に基づき、言語表現の意味は、言語表現とそれに対応づけられた話者の認識との関係であるとする。表現に対応づけられる認識は概念だと考えられるが、そのように断定して良いかについては、解決しておくべき問題がある。そこで、概念と表現の関係について検討する。

さて、概念は一般に、「対象の持つ特殊性を普遍性の側面から採り上げた認識の単位」である。言い換えると「対象の持つ個別的特徴を捨象し、必要不可欠の要素を統一体として捉えたもの」と言うことができる。

言語表現と概念の関係を考えるには、「認識としての概念」と「言語規範としての概念」の違いを明確にする必要がある。前者は、話者の認識の中で形成された概念であり、後者は、各言語において規範として成立している概念である。言語表現では、話者認識の中で概念化に成功した内容が必ずしもそのまま言語で表現できるわけではない。言語規範として成立している概念を介してのみ初めて言語で表現することができる。この違いは、三浦つとむが「概念の二重性」と呼んでいるものである。

言語意味処理の研究において、どちらの処理を目指すかによって大きな違いが生じる。前者の処理では、「世界知識に関する知識ベース」が必要となるのに対して、後者は、「表現と概念に関する知識ベース」があれば実現可能と見られる。筆者はこの違いに着目して、前者の処理を「意味理解」、後者の処理を「意味解析」と呼んで区別しているが、本研究では、後者の概念を意味処理の対象とする。

(4) 言語表現は概念を表すか?

さて、「果たして、言語表現は概念化された認識を表すと言って良いか」の問題に戻ることにする。言語表現において単語が概念を表す点については異論のないところであるが、問題は、「複数の単語からなる表現も概念を表すか」について、従来、明確な議論が見あたらないことである。

時枝誠記は、言語過程説において本居宣長の「玉の緒の理論」を引き継ぎ、日本語表現を「客体的表現」と「主体的表現」に分類したが、前者を「概念化された客体の表現」、後者を「概念化されない主体の感情や意思の表現」と説明している。これは、言語において概念化されない認識の表現を認めたものである。三浦つとむは、これを修正し、どちらの表現も概念化された認識の表現だとしている。

言語が理性的な表現であることを考えると、概念化に成功しない認識が言語で表現できるとする考えには矛盾がある。筆者の考えによれば、「客体的表現」と「主体的表現」との違いは、対象と主体との間に対時関係が存在するか否かの違いによって説明できる。そこで、本研究では、「客体的表現」、「主体的表現」共に概念化された認識の表現であると考え、概念化されない認識は言語では表現されないと考える。

以上の議論により、複数単語から構成される言語表現も概念を表すと考え、本研究では、言語表現から見た概念を以下の2種類に分類する。

- 単一概念＝単一の単語で表現できるような概念
- 複合概念＝複数単語の表現（句、節、分など）によって表される概念

この分類は、言語表現の形式に着目した分類でありあくまで相対的である。「単一概念」は、内部構造を意識しないままで抽象化された概念であるのに対して、「複合概念」は、内部構造を意識した概念、すなわち適切な単語で表現できないため、その内部構造まで意識して複数単語の表現に対応付けざるを得ない概念である。

言語では、新語の意味を文中で定義して使用したり、一度述べた内容（複合概念）を代名詞（それ、これ、等）で表現したりするが、これらは、複合概念を単一化する仕組みの一部である。

以上から、認識、概念、表現の関係は図2のようにまとめることができる。

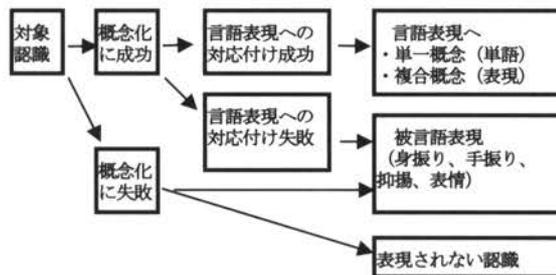


図2. 認識・概念・表現の関係

(5) 言語表現の構造も概念を表すか?

前節では、複数単語からなる表現が表す概念を「複合概念」と呼ぶことを述べた。そこで、いよいよ、「表現構造も概念を表すと言えるか？」について考える。

既に述べたように概念は、「対象の特殊性を普遍性の側面から採り上げた認識」である。対象の持つ普遍性をどのレベルで捉えるかによって、概念間に階層関係が生じる。例えば、「太平洋は日本海より広い。」、「富士山は大山より高い。」に共通する概念を「相対比較」と呼び、「相対比較」と「絶対比較」に共通する概念を「比較」と呼ぶことができる。

これに対して、言語表現の構造は、表現の形式を表すものであり、個別的な言語表現を抽象化したものである。個別的な概念を表す表現を抽象化すれば、より上位の概念を表す表現になると考えられる。このことから、それぞれのレベルの概念にはそれに対する表現の形式が存在し、逆に、表現の形式には、それによって表される概念が対応すると考えることができる。

以上から、個々の言語表現と同様、表現構造も概念を表すと言って良いことが分かる。但し、以上の議論は、概念化に成功した認識のみが、言語で表現されることを前提としている。概念には、表現や表現構造が対応するが、逆に、言語表現や表現構造のどのような部分を切り出してもそれに対応する概念が存在する訳ではないので注意が必要である。

3. 非線形言語モデルとパターン化

本章では、このモデルの概要と重要な性質のいくつかについて説明する。

なお、前章で述べたように、表現の表わす内容は概念である。従来の言語処理では、表現の表わす内容を意味と呼んでいる場合が多いので、それに倣って、以下では、表現の表す概念を意味と呼ぶが、それは本来の意味でない点に注意する。

3. 1 言語表現の線形性と非線形性

「人間は対象を概念化する過程である種のフレームワークを使用していること」、また、「そのフレームワークとしては話者の母国言語の表現の枠組みが用いられること」が指摘されている(有井1987)。

ところで、話者が思い浮かべた表現のフレームワークには、他の要素に置き換えると全体の意味が損なわれる要素と他の代替要素に置き換えても全体の意味は損なわれないような要素の2種類が存在する。

そこで、前者を「非線形要素」、後者を「線形要素」と区別し、「表現要素」と「表現」に対する「線形性」、「非線形性」を以下のように定義する。

<定義1>線形要素と非線形要素

一つ以上の代替要素が存在し、その要素に置き換えても「表現構造の意味」が変化しないような要素を「線形要素」、それ以外の要素を「非線形要素」と定義する。

<定義2>線形表現と非線形表現

「線形要素」のみから構成される言語表現を「線形な表現」、1つ以上の非線形要素を有する言語表現を「非線形な表現」と定義する。

但し、定義1の「表現構造の意味」は、前章の結論から、「抽象化された複合概念」である。上記の定義を現実の言語表現に適用するには、これを記述する方法を明確化する

必要がある。

ところで、計算機では、どのような意味記述も単に記号にしかすぎないから、意味論的に矛盾のない記号体系で表現できればよい。本研究では、日英機械翻訳を考え、日本語表現の意味を英語表現で定義する。このように、言語表現の意味を他の自然言語で定義する場合、定義に使用した言語側での意味的な多義が問題となる。しかし、機械翻訳の場合は、翻訳結果の意味を理解するのは目的言語側の人間であるので、あまり問題にはならないと考えられる。

図3に日英対訳文の例を示す。原文は、「何かの事象の直後、誰かが何かの行為をする」と言う「事象間の関係」（「抽象化された複合概念」）を表す日本語の表現で、その意味は英語表現で定義されている。単一概念である「彼女」、「大学」などには、代替可能な要素が存在し、それを置き換えても「抽象化された複合概念」（英語表現構造）は変化しないから、これらは線形要素である。

3. 2 非線形要素の基本的特徴

前節の定義では、線形要素は以下の5つの重要な特徴を持つことが指摘できる。これらの特徴から、パターン化のための重要な指針が得られる。

<特徴1>線形要素の言語ペア依存性

言語表現の意味を別の言語表現で定義しているため、

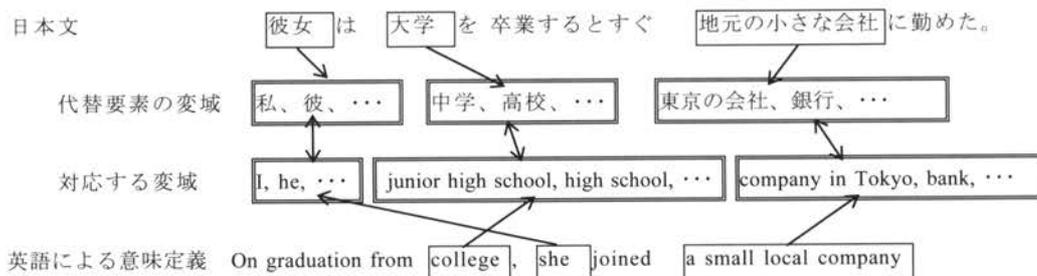


図3. 線形要素の例

線形要素の数や範囲は言語ペアに依存する。同族言語の場合は線形要素の範囲が増大し、異なる言語族間では減少することが予想されるが、これは言語による翻訳の難易性の違いを反映している。

<特徴2>線型要素の相互独立性

線型要素は代替要素に置き換えても全体の意味が変わらないことが条件である。この条件は各線型要素毎に成立しなければならないから、一つの表現構造の中に複数の線型要素がある場合は、それらの線型要素間の関係は相互独立でなければならない。

<特徴3>代替要素の有限性

線形要素は置換可能だと言っても何に置き換えても良い訳ではない。置き換え可能な範囲は、文法的、意味的に制限されるから、パターンではそれが「変域」として明示される必要がある。

<特徴4>線形要素と非線形要素の相対性

線形要素のみからなる表現が線形な表現であるが、要素の範囲は任意に決められるから、表現全体の線形性、非線形性は要素の選び方に依存する。従って、汎用的なパターンを得るには、線形要素が多くなるように工夫すればよい。

<特徴5>線形要素と非線形要素の同時性

「線形要素」が線形であるのは、あくまで表現全体から見たときの話であり、それ自身は非線形な表現であっても良い。

3.3 非線形言語モデル

前項で示した特徴から導かれる言語表現モデルについて述べる。定義1によれば、言語表現は「非線形要素」と「線形要素」から構成されるが、特徴4によれば表現要素の範囲は任意に選択できるから、意味のまとまる範囲の表現（例えば、単語、句、節など）の中から「線形要素」を抽出することとする。このようにして抽出した「線形要素」は、特徴5により、それ自身「非線形表現」でもよいため、言語表現は、一般に図4のような言語表現モデルで表すことができる。

この図から分かるように、非線形表現に含まれる線形要素を取り出していくと、最終的には線形要素を持たない非線形表現に帰着する。帰着した非線形表現は、単一の単語の場合もあるが、置き換え可能な要素を持たない慣用句のような場合もある。

以上から、非線形言語モデルでは、言語表現は、非線形

要素と非線形要素から構成される。

3.4 なぜ非線形な表現をパターン化するか

図2の言語モデルで大切なのは、分解の各段階で出現する「非線形表現」が意味のまとまる表現の単位だということである。要素分解によって元の意味を失わないようにするには、各段階の非線形表現に対する意味辞書を持つべき。例えば、言語表現を文、節、句、単語のレベルに分類し、その中の非線形な表現を対象とする表現意味辞書を構築すれば、文全体の意味をすくい取るための仕組みは一通り揃うことになる。

ところで、非線形表現を記述する方法であるが、

- (1) 非線形要素は通常字面表記が適していること
- (2) 線形要素と非線形要素の出現順序は固定的な場合が多く任意性が少ないこと

を考えると、パターンが適している。そこで本研究では、意味のまとまる表現を対象にパターン化する。

4. 文型パターン辞書の開発

4.1 パターンの記述方法

本章ではパターン記述方法の概要を説明する。詳

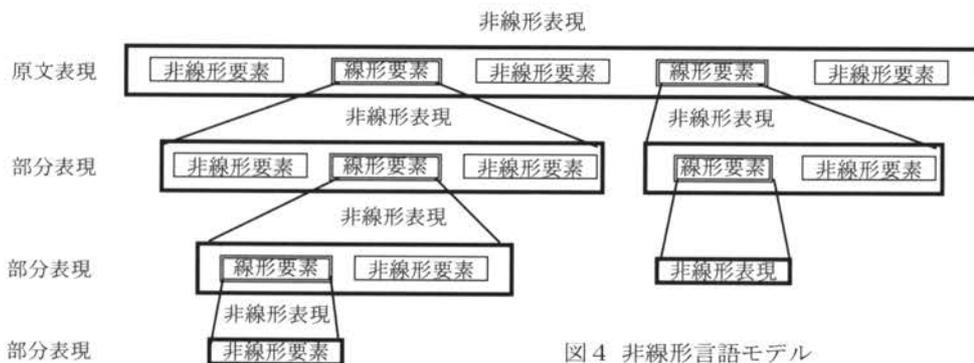


図4 非線形言語モデル

細は、「意味類型パターン記述言語仕様書」をご覧ください

(1) パターン化の原則

前章では、日本語表現の意味を英語表現を用いて定義することを述べた。この典型的な例は、日英対訳文である。そこで、本研究では、日英対訳例文の汎化によってパターン対を作成することとし、パターン記述言語を開発した。

ところで、汎化の対象となるのは、線形な表現要素である。前章の定義から、日英対訳例文において、線形要素には下記のものがある。

- (1) 英語側に対応する要素があるもの
- (2) 英語側に対応する要素はないが、削除しても

英語表現は変化しないもの

- (3) 元の対訳表現にはないが、それを挿入しても、倦関係が変化しないもの

- (4) 語順など構造的な変更を行っても英語側の構造が変化しないもの

そこで、日英対訳表現を対象に、これらに該当する部分表現を抽出して汎化することによってパターンを生成することとした。

(2) パターン記述の基本的枠組み

日英対訳パターンを記述するため、

- (1) 汎化の程度に応じた記述能力を持つこと、
- (2) パターン間で意味的な排他性が得られること、
- (3) 大規模な対訳コーパスから半自動的にパターンが生成できること

を目標にパターン記述言語を設計した。パターンの記述に使用される要素は、表1に示すように字面、変数、関数、記号の4種類である。

表1. パターン記述言語

| # | 要素 | 種類用途 |
|---|---------------|--|
| 1 | 字面 | 日本語文字、英語文字 |
| 2 | 変数 (17種類) | ①単語変数(9種類)、②句変数(5種類)、③節変数(1種類)・意味的な制約条件あり |
| 3 | 関数 (157種類) | ①語形関数、②時制、相、様相関数、③品詞変換関数、④文型生成関数、⑤その他 |
| 4 | 記号 (10種類) | ①表記の揺らぎ吸収、②任意要素の指定、③語順任意指定措、④位置変更可能指定、⑤省略要素補完指定、⑥その他 |

表現要素のうち、自立語的な要素(単語、句、節)では、線形な要素は変数を使用して記述し、非線形な要素は、字面によって記述する。

これに対して、付属語的な要素(助詞、状動詞など)では、線形な要素は関数を用いて記述し、非線形な要素は字面または関数で記述する。

関数には、非線形な要素を指定するもの(字面の代わりに使用される)と線形な要素を指定するものがあり、両者は区別して使用される。例えば、表現には現在形でしか使用できないもの(非線形)や現在形、加工形の双方で使用できるものなど(線形)がある。このような時制、相、様相などの表現の線形性と非線形性は、使用する関数によって識別される。

また、表記の揺らぎや語順の任意性など、特殊な線形要素は、記号を用いて記述する。

素は、記号を用いて記述する。

4. 2 文型パターンの作成方法

(1) 手順の概要

まず、日英対訳例文100万件を収集し対訳コーパスを作成した。次に、その中に含まれる述部2つまたは3つ重文、複文を取り出し、それを汎化することによって文型パターン辞書を作成した。対象対訳例文の平均単語数は、日本語が12.9語/文(最大63個)、英語が、10.3語(最大59語)である。

なお、本研究では、汎用性の高い文型パターンを網羅的に収集することを目指している。重文と複文を対象としたのは、すでに、単文では、「日本語語彙大系」が実現されているためである。また、述部数を2と3に限定したのは、述部数4以上の重文、複文は、述部数を3以下の文に分解して訳せる場合が多いと考えたためである。

以上の過程で、線形要素であることが自動的に判定できるものについては、機械的な置き換えを行い、自動的判定の困難なものは、言語アナリストの判断にゆだねた。

(2) 汎化のレベル

表現要素の線形性と非線形性を判断するには、あらかじめ表現要素の選び方を決める必要がある。そこで、言語表現の文法的な構成単位に着目して以下の3レベルの汎化を行った。

<単語レベル>:線形な自立語(名詞、動詞、形容詞、副詞など)を変数化したレベル。

<句レベルの汎化>:線形な句(名詞句、形容詞句、動詞句、副詞句など)を変数化したレベル。

表2. 作成された文型パターンの例

| レベル | 言語 | 文型パターン | 例文 |
|-----------|-----|--|---|
| 単語 レベル | 日本語 | #<N1(G4)は>/V2(R3003)て/V3(G932)を/ N4(G447)に/V5(R1809).tekita. | うっかりして定期券を家に忘れてきた。 |
| | 英語 | I was so AJ(V2) as to V5 #[N1_poss] N3 at N4. | I was so careless as to leave my season ticket at home. |
| 句 レベル | 日本語 | NP(G1022)1は/V2(R1513).ta/V3(G2449)に/ V4(R9100).teiruのだから/V5(N1453).dantei | その結論は誤った前提に基づいているのだから誤り である。 |
| | 英語 | NP1 is AJ(N5) in that it V4 on AJ (V2) N3. | The conclusion is wrong in that it is based on fales premise. |
| 節 レベル | 日本語 | CL1(G2492).tearuので、N2(G2005)に当たっ ては/VP3(R3901).gimu. | それは極めて有毒であるので、使用に当たっては十 二分に注意しなくてはならない。 |
| | 英語 | so+that(CL1, VP3.must.passive with subj(CL1) _poss N2) | It is significantly toxic so that great caution mus t be taken with its use. |

<節レベルの汎化>:線形な節(連体節と連用節)を変数化したレベル。

4. 3 作成した文型パターン

(1) 生成された文型パターン数

文型パターンの記述例を表2に示す。また、得られた文型パターンの種類と異なりパターン数を表3

表3. 文型パターンの種類と異なりパターン数

| | 単語レベル | 句レベル | 節レベル | 合計 |
|-----|---------|--------|--------|---------|
| 重文 | 61,171 | 39,243 | 18,173 | 118,587 |
| 複文 | 48,123 | 32,049 | 5,778 | 85,950 |
| 重複文 | 12,610 | 8,146 | 1,524 | 2,280 |
| 合計 | 121,904 | 79,438 | 25,475 | 226,817 |

に示す。単語レベルでは、元の標本文とほぼ同じだけのパターンが作成されたが、句レベル、節レベルになるにつれ汎化が困難となり、得られたパターンは急速に減少した。これは大半の対訳例文の節は非線形要素であり、汎化困難であることを示している。

(2) 汎化された要素の割合

各レベルの汎化において変数化された要素数を表4に示す。汎化された線形要素の割合は、単語レベルでは、73.9%、句では24.0%であったのに対して、節ではきわめて少なく14.83%であった。

非線形要素は、それを取り出して翻訳し、元の文に組み込んでも意味的に適切な翻訳結果は得られない。上記の結果から、従来のような要素合成法では、多くの場合、対訳例文に示されるような質の良い翻訳はできないことがわかる。

表4. 線形要素の割合

| 要素種別 | 全要素数 | 変数の数 | 線形な割合 |
|---------|---------|---------|--------|
| 単語(自立語) | 734,528 | 542,833 | 73.9 % |
| 句 | 463,636 | 111,359 | 24.0 % |
| 節 | 267,601 | 39,714 | 14.8 % |

4. 4 文型パターン辞書の評価

(1) 実験の条件と評価の方法

入力文とのパターン照合実験によって、パターン辞書の被覆率特性を評価した。実験の条件は以下の通りである。

- ①変数の意味的な制約条件は無視する
- ②実験はクロスヴァリデーションの方法とする。

このうち②は、パターン作成に使用した例文を入力文に使用するが、当該入力文から作成された文型パターンへは必ず適合するため無視し、それ以外のパターンへの適合のみを評価する方法である。

文型パターンの被覆率は以下の2つのパラメータによって評価した。

- ・**統語的被覆率 (R)** : 入力した文のうち、1文型パターン以上に適合した入力文の割合
- ・**意味的被覆率 (C)** : 入力文に対して、意味的に正しいパターンが1つ以上存在する割合

(2) 統語的被覆率の飽和特性

文型パターンの数と適合率の関係を図5に示す。曲線は、下から順に、単語レベル、句レベル、節レベルのパターンの飽和特性を示す。

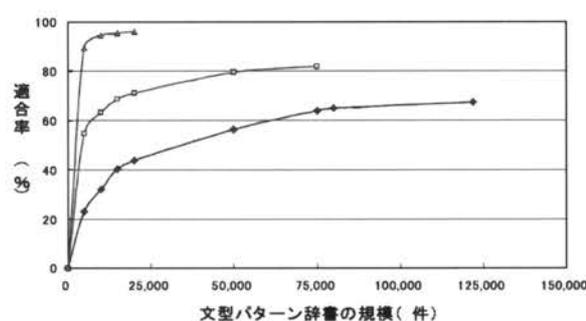


図5. 文型パターン適合率の飽和特性

この図から、適合率の飽和傾向が顕著である。横軸の文型パターンを使用頻度順に並べ替えると、飽和の速度は5倍程度速くなる。過去の研究によれば、単文をほぼ網羅するために必要な結合価パターン数は、約2.5万件であると推定されている。重文、複文の場合も、ほぼ網羅するために必要なパターン数も同じオーダーに収まることが期待される。

(3) 適合率と正解率

次に、文型パターン辞書全体の適合率と正解率を表5に示す。この表から、文型パターン辞書全体では、構文上、入力文のほぼ全体をカバーすることが分かる。しかし、意味的に不適切で翻訳に使用できないパターンに適合する場合も多く、それらを除いたときの意味的な被覆率は78%である。

表5. 文型パターン辞書の被覆率特性

| パターン種別 | 平均適合パターン数 | 統語的被覆率 (R) | 意味的被覆率 (C) |
|--------|-----------|------------|------------|
| 単語レベル | 63.2 件 | 62.6 % | 36.7 % |
| 句レベル | 520.3 件 | 82.7 % | 55.0 % |
| 節レベル | 1241.7 件 | 96.4 % | 67.0 % |
| 合計 | 1825.2 件 | 97.8 % | 78.0 % |

なお、述部数2の重文、複文の意味的な被覆率は、それぞれ、82.5%、71%であった。

5. あとがき

言語表現の非線形性に着目した言語表現モデルとそれに基づく文型パターン化の方法を提案し、日本語の基本的な重文、複文の表現に対して、単語レベル（12.2万）、句レベル（7.9万件）、節レベル（2.5万件）からなる文型パターン辞書（合計22.7万件）を開発した。

その結果によれば、汎化することができた線形要素は、自立語の場合73.9%であったのに対して、名詞句、動詞句などの句では、24.0%、節はきわめて少なく14.8%であった。

非線形要素は、それを取りだして翻訳し、元の文に組み込んでも意味的に適切な翻訳結果は得られない。この結果から、ほとんどすべての複文、重文は、複数の単文などの要素に分けて翻訳し後で結合しても、対訳例文に示されるような品質の翻訳はできないことが分かった。

また、クロスバリデーションによる文型パターン辞書の評価実験では、単語レベル、句レベル、節レベルの文型パターンの意味的な被覆率は、それぞれ、36.7%、55.0%、

67.0%で、それらを組み合わせて使用するときは、78%であった。また、最も基本といえる述部数2の重文、複文の意味的な被覆率は、それぞれ、82.5%、71%であった。

文型パターンは、非線形な言語表現を対象としている。線形な表現の場合は従来の要素合成法が適用できるので、今後、両者を併用した機械翻訳方式を実現すれば、翻訳の品質は大幅に向上することが期待される。また、文型パターンは、意味を掘り取るための網の目である。機械翻訳だけでなく、広く意味解析への利用が期待される。

なお、本稿では、述べなかったが、本研究では、「意味的等価変換方式」（本ジャーナル、No.33、pp.1-7、2002.3を参照）の実現に向けて、重文複文の意味分類体系を構築し、すべてのパターンに意味分類コードを付与した。その詳細については、別の機会にご紹介する予定である。

付記1

本研究の成果の研究利用は原則無料です。詳細は、「鳥バンク」（<http://unicorn.ike.tottori-u.ac.jp/toriban/>）にてご案内する予定です。多くの方々のご利用を期待しています。

付記2

本研究が契機となって平成19年度文部科学大臣表彰科学技術賞（研究部門）を受賞しました。今まで多くの方々に支えられ、意味解析の研究を続けていくことができたお陰です。今後の機械翻訳技術の発展を願うと共に、その活動を支えてこられた多くの方々に深くお礼を申し上げます。

翻訳業務における MT/TM の利用実態

(社) 日本翻訳連盟 翻訳支援ツール委員会 委員長(前理事)

久徳 省三

■ はじめに

私は約 10 年前に最後の職場を定年で退職したときから在宅で翻訳の仕事をしている翻訳者です。当初から機械翻訳（当時 TM は無かった）ソフトを使用していますが、仲間の話を聞くと翻訳分野によって MT/TM（以後機械翻訳）ソフトの効果が違うようです。

本書をお読みになる方は、機械翻訳ソフトの開発販売をしていらっしゃる方、機械翻訳を翻訳の仕事にお使いの方などと推定いたします。

約 10 年の経験をもとに、私がどのように機械翻訳ソフトを使用し、どのように機械翻訳ソフトに支援されてきたかを報告するとともに、翻訳者の立場から、開発者の方がたにご配慮をお願いしたことをお書きます。

■ 業界の現状

残念ながら、IT ローカリゼーション分野以外の仕事では、翻訳支援ツールの利用頻度は少なく、翻訳会社さんも翻訳者も現在の商品に対する評価は良くありません。後述のとおり、ソフトウェアメーカーさんが業界に誤解を与えた結果と思われる。

機械翻訳を使用していると言うと、翻訳会社が訳質に疑問をもち、警戒するので、使用していながらそれを内緒にしている翻訳者もいます。

そのような状態で、訳文の納品形態も、数年前までは翻訳者はテキスト形式で納品すればよかったのが、この数年で MS Office などの形式で納品しなければならない仕事が増えました。

また、工業分野の実務翻訳では、部分（差分）翻訳が主流になりつつあります。つまり、旧製品のマニュアルの訳文があるので、新製品のマニュアルは変更になった箇所だけを翻訳して、経費と時間を軽減することです。

この場合、差分箇所の抽出は翻訳会社が行いますが、翻訳会社が翻訳者に支給する原稿は会社によって異なります。差分箇所だけの原稿を支給された場合は、翻訳者は他の部分がどのような訳文になっているか分からないので、翻訳者が最善の訳文を納品します。

その結果、同じ書類に別の翻訳者が訳した訳文が挿入されることになり、書類の整合性は保てません。このような場合、翻訳会社が旧製品のマニュアルの原稿と訳文を入手し、翻訳メモリを作成し、それを原稿と一緒に翻訳者に支給すれば書類の整合性は保てます。

ほとんどの翻訳会社は、まだこの問題に真剣に取り組んでいませんが、ソースクライアントが差分翻訳の訳文の不整合にクレームをつけるのは時間の問題と考えられます。既に海外の翻訳会社からの依頼の大半は TM 使用が条件になっています。

米国のベストセラー著者のピーター・ドラッカーさえ、自分が書いた文はエディターにリライトさせてから印刷に出していたそうです。品質を売り物にしている翻訳業界が、このように文章を部品として扱う行為をしているのですから、機械翻訳ソフトウェアを利用して整合性を確保することは時代の要求だと思います。

(社)日本翻訳連盟 翻訳支援ツール委員会では連盟のホームページに機械翻訳ソフトのメーカーさんとユーザが情報交換をする掲示板の開設を計画し、二年前に理事会で承認されていますが、まだ実現していません。

TM 製品では、SDL さんが Trados を買収し、ライオンブリッジさんが Logoport を買収した結果、Trados が現在の市場占有率を今後も確保できるのか懸念されます。できない場合の業界に与える影響は極めて大きいので SDL さんに頑張ってもらいたいと期待します。

■ 翻訳者の実態

個人差はありますが、翻訳者は毎日、英訳の仕事で約 12,000 字の原稿、和訳の仕事で約 3,000 ワードの原稿を訳しています。和文原稿の 1 文の字数が 50 字、英文原稿の 1 文のワード数が 20 ワードと仮定すると、毎日英訳で 240 文の原稿、和訳で 150 文の原稿を訳していることとなります。機械翻訳ソフトは 1 文に最低 1 回マウスボタンをクリックしなければならないので、翻訳者は毎日 200 回～300 回マウスボタンをクリックしています。

業界では、1 日に何枚できるという言い方をしますが、英訳は訳文 200 ワードを 1 枚、和訳は訳文 400 字を 1 枚と数え、1 日に最低 20 枚の量をこなさないと翻訳者は商売になりません。

機械翻訳でソフトが自動的に訳文を転記してくれることもありますが、手直しの必要もあり、基本的には訳文はキーボードから手入力します。つまり、毎日英訳では 4,000 ワード、和訳では 8,000 字の訳文を入力しなければなりません。キーボードのキーを叩く回数は毎日 20,000 回以上になります。1 日の作業時間が 8 時間と仮定すると、1 時間に 2,500 回、1 分間に 42 回キーを叩いているわけです。

手入力に必要な時間を減らせば作業効率が上がるので、各翻訳者はローマ字入力を仮名入力にしたり、音声入力を使用したり、自分がやり易い方法を考えています。

従って、たとえマウスのクリックであっても、翻訳者が行う作業が少ない機械翻訳ソフトが好まれています。

一方、翻訳者は翻訳の仕事を受注すると納期に間に合わせるため、睡眠時間を犠牲にしなければならないことがあります。2 週間連続で毎日の睡眠時間が 4 時間という疲労困憊の状態です仕事をすることもあります。つい、うっかりデータを保存せずにアプリケーションを終了して、数時間の作業結果を喪失することもあります。アプリケーションによっては、フリーズすることもあります。その場合の被害を最小限にするため、データを自動的に保存する機能をデフォルトで設定したソフトが望まれます。

■ 機械翻訳ソフトの使用例

私は、翻訳専門になってすぐに機械翻訳ソフトを導入しました。会社勤務時代、約 40 年携わってきた貿易の仕事は、書類作成やコレボンなど、会社にいるときは事務的な仕事をします。そのとき作成する書類は新規作成より前例を参照するケースが多いので、機械翻訳を使って翻訳作業を少しでも効率化させると同時に、対訳文例集を保存したいと考えたからです。

一般的には、機械翻訳は英日翻訳に向いており、日英翻訳には不適といわれています。同様に、分野はマニュアルや仕様書等の翻訳の際に有利とされています。ところが、私の場合は日英翻訳にも、また、契約書などの法務文書の翻訳にも機械翻訳を活用しています。

機械翻訳を使って日本語から英語に翻訳した場合、英訳文をほとんど書き直す必要があり、

機械翻訳に慣れている人でも、あまり使用しないというケースが多いようです。また、日本語は仮名表記と漢字表記の混用や長音符の不整合などが原因で、TMのヒット率が低下します。

但し、機械翻訳を使用して作業をすると、たとえ翻訳メモリに蓄積しなくても、機械翻訳ソフトは対訳データを自動的に作成するので、それをテキスト形式ファイルで保存しておく、対訳文例集としてエディターのgrep検索機能を使用して後日参考にすることができるので、私の場合は英日・日英のどちらの翻訳にも使用しています。日本語から英語の場合でも、やはり用語の統一は同じようにできます。マニュアルなどはもちろん、法務文書用語・文例も登録しておけばいいわけです。特に契約書は非常に長い文章が多いのですが、短いフレーズの集まりとしてフレーズごとに分割して訳すのにも機械翻訳は便利です。

おそらく特許翻訳でも同様の使い方ができるでしょう。専門的な医薬関係の論文は別として、医療機器の仕様書、周辺の資料、社内報などを機械翻訳を使って翻訳したことがあります。

作業手順は、TM付MTソフトの対訳画面で一文ごとにMTに下訳をさせ、リライトした訳文をTMに保存します。TMの一致率は低く設定し、この文は前にあったと記憶している場合【類似文検索】にかけ、ヒットすれば過去訳を利用します。完全一致の場合は、前後関係で不自然でないかぎりそれをマウスのワンクリックで転記します。このような方法で訳した訳文ですから、直訳になりがちですから、必ず訳文だけを見直して不自然な箇所をリライトします。

翻訳会社から用語集が支給された場合は、あらかじめSuperHT3のユーザ辞書に変換し、同ソフトの【訳語付与】機能で原稿に訳語を追記したデータを印刷して、それを見ながら作業を進めます。

これら支給用語集は、単語語尾に余計なスペースがあることが多いので、翻訳ソフトに一括登録する前にExcelのTrim関数などで、削除する必要があります。外国語単語の場合、語尾に半角スペースが2個以上あるとヒットしません。

SuperHT3を使用した仕事の訳文は、時間が許すかぎり同ソフトの【用語確認】機能で指定用語を使用しているか確認します。SuperHT3を使用しない仕事でも、取扱説明書の和訳などはSuperHT3用に作成した「間違いやすい訳語」(通り→とおひ、及び→おひ、など)を収録したユーザ辞書で確認します。

よく、TMは資産だと言いますが、TMが役にたつケースは同じソースクライアントの類似文献の翻訳と、数日にわたり作業する同一文献の前訳です。

■ メーカーさんに反省を促したい点

従来はMT/TMのメーカーさんは、あたかもTMを使用すれば完全に翻訳をするような印象を与えながら製品を販売してきたようにお見かけします。

欧米語どうしであれば良いのですが、日本語と欧米語の場合構文が異なりますから、人工頭脳を備えていない現在のコンピュータに完全な翻訳をさせることは不可能と思われます。

メーカーさんは、機械翻訳が完全な翻訳をするのだという売り方ではなく、翻訳を支援できる機能はこのような機能で、それを使用して翻訳するとこのような利点があるという売り方をした方が賢明だと思います。

機械翻訳ソフトは翻訳をするツールではなく、翻訳をする人に登録した訳語を部品として提供するツールで、部品を組み立てるのは翻訳者の仕事であるという売り方が良いのではないのでしょうか。

人間の記憶力には限界があるので、それをコンピュータに補わせ、訳語の統一、表記の間違いのチェック、対訳データベースの作成など、翻訳の作業をやる人にとって有益な機能があるのですが、「使い物にならない」との印象を与える結果になっているようです。

TM を使って翻訳の作業をしていると、この文は前にあったはずと TM 検索をすると、和文原稿の場合は片仮名表記と漢字表記、長音符の有無、英文原稿の場合は冠詞と前置詞の違いが原因のことが多いようです。

中には、原稿に数字の違いがあるのを見つけ、訳文を納品する際に「訳注」で報告するとソースクライアントに感謝されることがあります。

機械翻訳ソフトの特徴の中には豊富な専門用語辞書がありますが、かならずしも現場に即した訳語が収録されているとはかぎりません。例えば、専門用語辞書には Invoice は、1. 送り状 2. 仕切状 3. インボイス 4. 仕切り書 5. 請求書などが収録されています。貿易会社や流通業者では Invoice は片仮名の「インボイス」を使用しています。貿易会社から翻訳会社が受注した案件の和訳に「仕切り書」という訳語を使用したら、その貿易会社からは次の仕事は来ないでしょう。

一般ユーザーが英文の大意を理解するために安価な機械翻訳ソフトも販売していらっしゃるのに難しい面もあるかもしれませんが、少なくとも業界で販売する場合は、完全に翻訳をするソフトではなく翻訳を支援するツールであるという売り方をしていたきたいと思います。

第 17 回通常総会開催のご案内

第 17 回アジア太平洋機械翻訳協会通常総会を下記のとおり開催いたしますので、ご案内申し上げます。

記

1. 日時 2007 年 6 月 11 日 (月) 13:30 - 14:15
2. 会場 総評会館 204 会議室 (東京・御茶ノ水)
3. 議題 第 1 号議案 2006 年度事業報告 (案) 第 2 号議案 2006 年度決算報告 (案)
第 3 号議案 2007 年度事業計画 (案) 第 4 号議案 2007 年度予算 (案)
第 5 号議案 役員選任 (案) その他、会員提案事項

通常総会后、2006 年度活動報告会および講演会を開催いたしますので、多数のご参加をお待ち申し上げます。講演内容は変更する場合がございますので、予めご承知おきください。

| 行 事 | 内 容 | 時 間 |
|----------|--|---------------|
| 総会 | 第 17 回 AAMT 通常総会 開会の辞・会長挨拶・御来賓挨拶・議題 | 13:30 - 14:15 |
| | 休 憩 | 14:15 - 14:30 |
| 活動報告会 | 当協会各委員会からの 2006 年度活動報告: 機械翻訳課題調査委員会・AAMT/Japio 特許翻訳研究会・ インターネット WG・編集委員会 | 14:30 - 15:30 |
| 講演会 | | 15:30 - 16:50 |
| 講演 I. | 「JAL の安全対策とテキスト処理」 | 15:30 - 16:10 |
| 講演 II. | 寺田 昭 氏 (株)日本航空インターナショナル運航本部運航技術部 “KUI to assist online multilingual text development” ソレントラムワニッチ・ウイラット氏 (NICT タイ自然言語学ラボラトリー) | 16:10 - 16:50 |
| AAMT 長尾賞 | 授与式・記念講演会 | 16:50 - 17:30 |
| 懇親会 | 山の上ホテル 本館地下一階「ラヴィ」 立食パーティ形式・参加費 5,000 円 | 18:00 - 20:00 |

会場 (総評会館) までの交通機関のご案内:

JR 中央線・総武線「御茶ノ水」駅聖橋口より徒歩 5 分

東京メトロ千代田線「御茶ノ水」駅 B3 出口より徒歩 0 分、丸ノ内線「淡路町」駅 B3 出口より徒歩 5 分

都営地下鉄新宿線「小川町」駅 B3 出口より徒歩 5 分

世界の言語をつなぎ、人と人をつなぐ言語グリッドの取り組み

(独)情報通信研究機構 言語グリッドプロジェクト

稲葉 利江子、村上 陽平、重信 智宏、灘本 明代、石田 亨

1. はじめに

近年のインターネットの普及と国際化の流れにより、異文化交流や国際交流学習の活動が盛んになっている。しかし、未だ解決できていないのが、「言語の壁」の問題である。英語が「共通語」と思われがちだが、実際インターネットの言語人口は多様化しており、一番多い英語人口でも 35.2%となっている[1]。また、異文化コラボレーション活動の中心となるのは、それぞれの現地の人たちであり、必ずしも共通語が存在するとも限らない。そこで、考えられるのが機械翻訳の利用である。しかし、機械翻訳をそのまま利用するにはいくつかの問題が生じる。①利用したい言語対のサービスがあるのかわからない。②複雑なライセンス契約と高額なコストがかかる。③コラボレーションの現場ではコミュニティ特有の用語が利用される。

これらの問題を解決する取り組みとして NICT 言語グリッドプロジェクトでは、インターネット上にある機械翻訳などの専門家が作成した言語サービスとコミュニティ独自で利用されている用語の対訳を集めたユーザ辞書などの現場で生まれた言語サービスを組合せ、利用できる仕組みづくりを行っている。

本稿では、言語グリッドの仕組みを説明し、実際に異文化コラボレーション活動の現場で利用されているユースケースについて示す。

2. 言語グリッド

2.1 言語グリッドとは

言語グリッドは、インターネット上の既存のサービスを容易に利用でき、ユーザ自らが新たな言語資源や言語処理機能を追加可能な仕組みである。一般

にグリッドは、「分散した資源を連携させ、オープンスタンダードなプロトコルが利用でき、質の高いサービスを提供できるもの」と定義されており、言語に関するグリッドは、この提案が世界で初めてである[2]。言語グリッドは、インターネット上に分散した言語資源を連携させ、容易に利用できるように様々な機能を提供する。特定のツールではなく多様なツールからアクセスできるようにし、インターネット上の言語資源を Web サービスインタフェースで提供する。また、同種の言語サービスの利用を容易にするために、言語資源の種類ごとに利用インタフェースを統一する。さらに、言語資源の権利を保護するため、言語サービスへのアクセス権を管理し、不正な利用を防ぐ。

また、言語グリッドが連携する言語資源としては、対訳辞書や機械翻訳などを組み合わせる「水平型言語グリッド」と、応用に特化し現場が必要とする言語サービスを生み出す「垂直型言語グリッド」という 2 つのグリッドがある。水平型言語グリッドは、アジアの 10 言語程度の対訳辞書や機械翻訳を利用するとともに、欧州の研究機関と連携し世界全体で 20 言語程度をカバーする。一方、垂直型言語グリッドは、防災教育、医療通訳、絵文字コミュニケーションなどの様々な異文化コラボレーション活動のための辞書や用例対訳を利用可能とする。

2.2 言語グリッドのプレイヤー

言語グリッドへは、5 種類の関わり方がある。

① 言語資源提供者

機械翻訳や形態素解析、専門分野の辞書、対訳集などの言語資源の提供者である。我々は、無償・有償の選択が出来るようサービス管理システムに関しても検討を行っている。

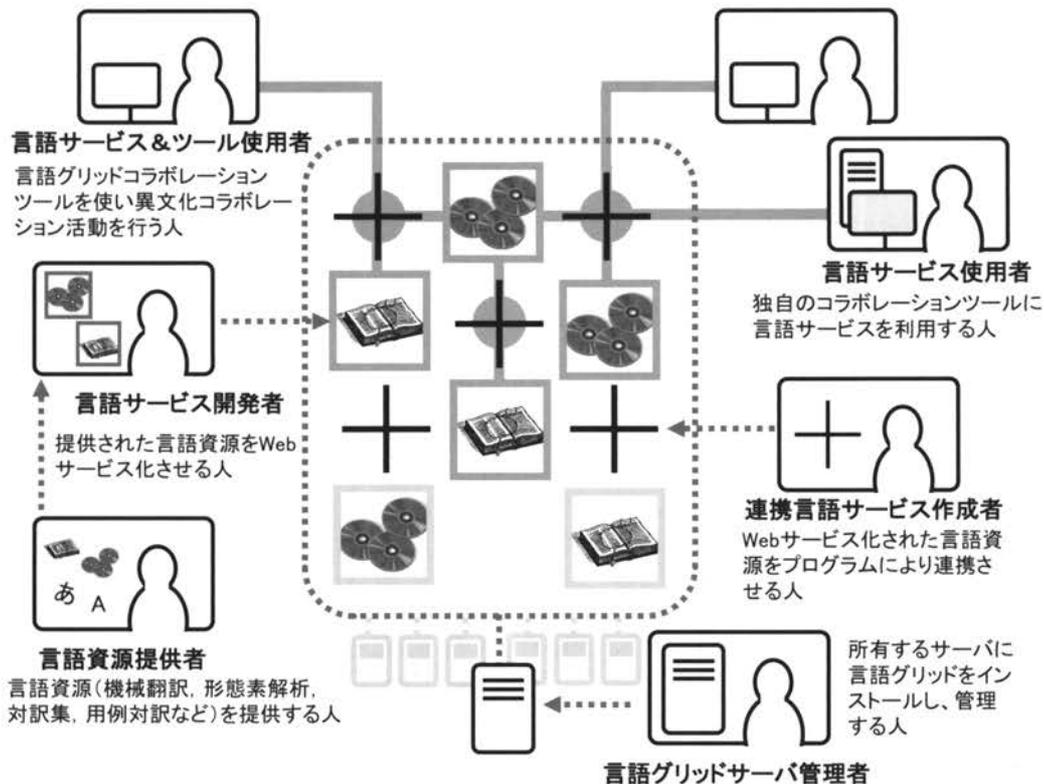


Fig.1 言語グリッド関係図

② 言語サービス開発者

言語資源提供者から提供された言語資源を Web サービス化する開発者である。それぞれの言語資源の形態により言語グリッドに連携するためのラッピング作業がマニュアル化され、提供されている。

③ 連携言語サービス作成者

統一フォーマットにより Web サービス化された言語サービスをワークフローにより連携させ、新しい言語サービスを作成する人である。例えば、日英翻訳と英アラビア翻訳、各言語の形態素解析を連携させることで新たな日アラビア翻訳を可能とすることが出来る。

④ 言語グリッドサーバ管理者

言語グリッドはコアノードとサービスノードからなる[3]。このコアノードを世界中に配備することで分散処理が可能になる。このコアノードを設置し管理する人である。

⑤ 言語サービスユーザ

異文化コラボレーションの活動現場において言語グリッドのサービスを利用する人である。言語グ

リッドは、言語のインフラであるため多言語チャットなどの既存のコラボレーションツールで利用することが出来る。言語グリッドプロジェクトでは、多言語チャット (Langrid Chat)、多言語共有黒板 (Langrid Blackboard)、多言語入力支援 (Langrid Input) の 3 つのコラボレーションツールを研究開発している。

3. 言語サービスをつなぐ技術

機械翻訳をベースとした多言語コラボレーションツールにより多言語間でのコミュニケーションを支援する場合、翻訳の品質、また翻訳に一貫性がないため生じる言語ごとの翻訳の非均質性が問題となる。このような問題を解決するには、従来の機械翻訳のように汎用的な翻訳サービスを利用するのではなく、複数の機械翻訳やコミュニティユーザにより構築されたコミュニティ辞書、そして多様な言語処理機能を連携させて、コラボレーションのタスクに特化した翻訳サービスをユーザ自身で構築できるようにする必要がある。

コミュニティ辞書により翻訳品質を向上させる同様の取り組みの一例として、「訳してねっと[4]」と呼ばれるWeb ベース協調型翻訳環境が既に開発されている。訳してねっとでは、コミュニティ内の人たちが協調しながら専門辞書を作成することで、そのコミュニティに特化した翻訳の精度を向上させることができる。しかしながら、

機械翻訳の一部としてコミュニティ辞書が組み込まれているため、既存の機械翻訳では作成したコミュニティ辞書をそのまま利用できないという問題が生じる。したがって、機械翻訳エンジンの内部でコミュニティ向けにカスタマイズするのではなく、機械翻訳やコミュニティ辞書などの対訳辞書をサービスとして捉え、言語サービスを連携させるワークフローでカスタマイズする仕組みが求められる。これにより、同じインタフェースを備えた言語サービスであれば置き換えが可能になり、翻訳の拡張性を向上させることができる。Fig.2に、NPO パンゲア¹用の翻訳サービスを、形態素解析、コミュニティ辞書、機械翻訳を連携させ、作成したワークフローを示す。この翻訳サービスは、日本語の入力文を専門用語を考慮して英語、韓国語、ドイツ語に翻訳し、返すサービスである。

折り返し翻訳の中で呼び出される各翻訳サービスは、コミュニティに特化した翻訳を実現するために、コミュニティ辞書と複数の機械翻訳を連携させたワークフローで実現される。このワークフロー上では、まず入力として受け取った文章を形態素解析

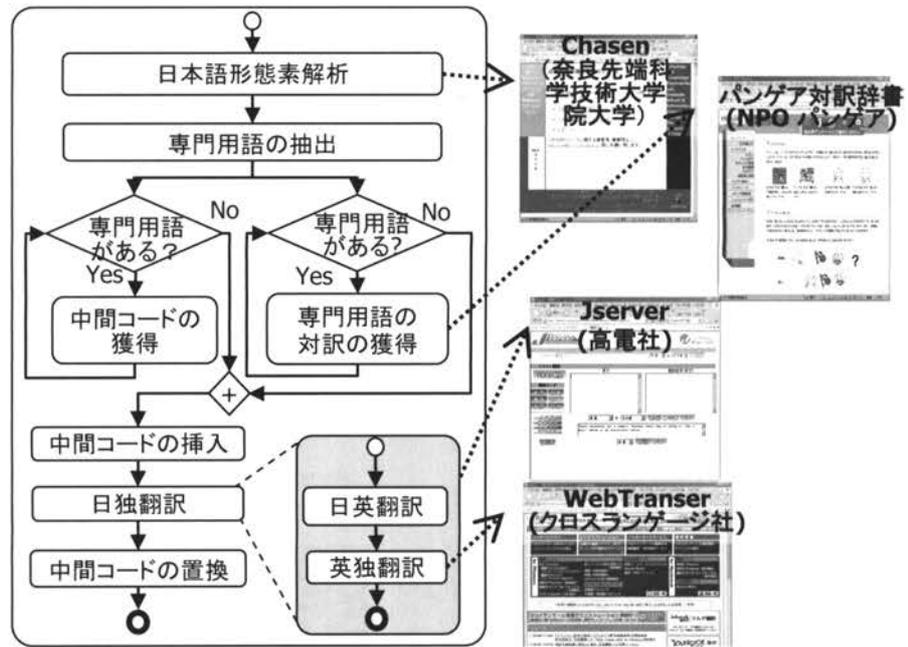


Fig.2 多言語コミュニケーション支援のためのワークフロー

サービスにより、形態素に分ける。分けられた形態素の集合から、専門用語のみを抽出し、専門用語の集合を作成する。次に専門用語の対訳を専門用語対訳辞書サービスから取得し、同様に、翻訳に影響を与えない文字列である中間コードも取得する。取得した中間コードで入力文中の専門用語を置き換え、中間コードを含んだ文を翻訳する。最後に翻訳結果から中間コードの部分の専門用語の対訳で置き換えることで、専門用語に対応した翻訳結果が得られる。なお、指定された言語間の翻訳に対応した機械翻訳が存在しなければ、複数の機械翻訳を連携させてマルチホップ翻訳を実現する。このようなワークフローによりコミュニティに特化した多言語折り返し翻訳サービスを言語グリッド上に実際に配備するには、まず事前に既存の形態素解析、機械翻訳および専門用語の対訳辞書を標準インタフェースに従ってラッピングし、言語グリッド上に配備し連携する必要がある。これにより、各言語サービスは言語グリッド上で利用可能になり、Fig.2に示したワークフローにより連携され、サービスの提供が可能になる。実際に Fig.2 ではサービス実体として、日本語の形態素解析に奈良先端科学技術大学院大学の Chasen を、コミュニティ辞書には NPO パン

¹ <http://www.pangaeaan.org/>

ゲアのコミュニティ辞書を、日英翻訳、英独翻訳にはそれぞれ高電社の J-Server と Cross Language 社の WebTranster を呼び出している。

4. 異文化コラボレーション活動における支援

言語グリッドプロジェクトでは、異文化コラボレーション活動の現場での実証を通じて、人々が使える多言語コミュニケーション環境づくりを推進している。本章では、ユースケースとして、実際に異文化コラボレーションの現場にて言語グリッドを利用している事例を示す。また、プロジェクトで開発している言語グリッドを言語基盤としたコラボレーションツールもあわせて紹介する。

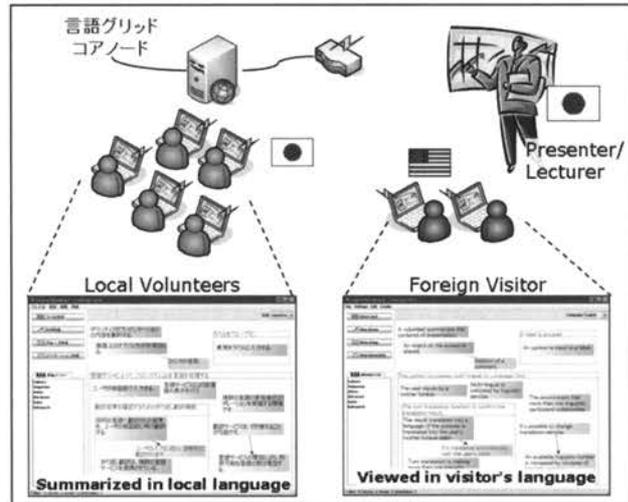
(1) 多言語ミーティング支援

日本では学校や企業での日常の打ち合わせやセミナーなどでは一般的に日本語が利用されている。しかしながら、日本語の不得意な外国人参加者がいる場合、外国人は日本語での議論は難しい。そこで、多言語でのミーティング支援として多言語共有黒板 (Langrid Blackboard) を提供している。多言語共有黒板は、多言語で読み書きが可能なツールとなっており、KJ 法²をもとに議論内容をラベルやグループを用いてまとめることができる。(Fig.3)

(2) 多言語コミュニケーション支援

世界に拠点をもち、異文化コラボレーション活動を行っている組織において問題となっているのが、スタッフ間でのコミュニケーションである。実際、NPO パンゲアでは、日本、韓国、オーストリア、ケニアに活動の拠点をもち、共通のプレイグラウンドを提供することで、世界の子どもたちをつなげる

² 多くの断片的なデータを統合して情報整理を行い、創造的なアイデアを生み出し、問題解決の糸口を探り出す手法。



(a) ミーティング支援構成図



(b) NICT言語グリッドミーティング (外国人研究者を支援) (c) 京都大学研究会 (中国人留学生を支援)

Fig.3 多言語ミーティング支援

ことを目標としている。実際、各拠点にて、アクティビティと呼ばれるパンゲア独自の活動がなされているが、そこでボランティアとして働くスタッフへの研修および日頃のコミュニケーションにおける言葉の壁が問題となっていた。そこで、言語グリッドを用いた多言語チャットを用いることで、母国語同士での打合わせが可能となり、現在、日本-韓国間、日本-オーストリア間で活用されている。

(3) 既存のコミュニケーションツールへの入力支援
インターネット上での異文化交流において特定の言語 (英語など) による掲示板などが使われていることが多い。また、異文化コラボレーション活動を行っている組織が独自で開発しているコミュニケーションツールがある。そこへ多言語での入力を可能とする入力支援として、Langrid Input を開発している。

Fig.4 は、NPO JEARN³で利用されている事例を

³ <http://ndys.jearn.jp/>

示す。JEARN では、世界の子どもたちの被災体験や防災安全マップなどに関する意見交換を独自の英語掲示板を利用している。これまでは日本の子どもたちにとって英語の書き込みは非常に困難であった。そこで、Langrid Input を利用することにより、日本語で入力した文章を英語に翻訳し容易に書き込みが出来るようになっている。

Langrid Input には翻訳した文を再度入力言語に翻訳しなおす「折り返し翻訳機能」があるため、どのような内容が翻訳されたかを確認することができ、書き込みを促進している。また、防災関係の用語や JEARN で利用されている造語などをコミュニティ辞書にあらかじめ登録しておくことで、コミュニティに適した翻訳結果を得ることができる。コミュニティ辞書の有無での翻訳結果の適正に関しては、コミュニティに依存したテストデータを用い、TIDES⁴で利用されている 5 段階評価での内容理解度の評価実験を行い、コミュニティ辞書の有効性を確認している。

5. おわりに

異文化コラボレーションにおける多言語コミュニケーションを支援するには、複数の機械翻訳を単純に利用するだけではなく、コラボレーションのタスクに特化したコミュニティ辞書との連携が必要である。本プロジェクトでは、言語サービスへのアクセシビリティおよびユーザビリティの向上を目的とした言語グリッドを利用して、多言語コミュニケーションを円滑にするために 2008 年 4 月の非営利目的での公開を目指し、研究開発に取り組んでいる。

⁴ <http://www ldc.upenn.edu/Projects/TIDES/Translation/TransAssess02.pdf>

JEARNの情報交換用掲示板(英語)

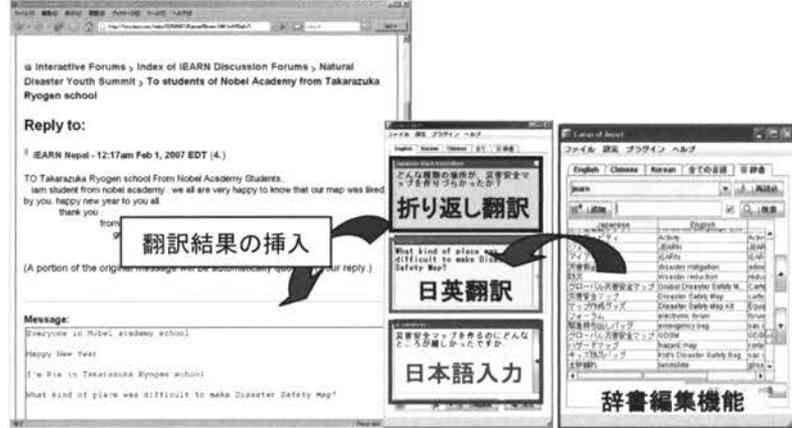


Fig.2 既存の掲示板への入力支援

謝辞：本研究にご協力いただく言語資源提供者の皆様に感謝します。また、言語グリッドの利用・高度化のためにご協力いただくユーザグループの言語グリッドアソシエーション⁵の皆様に感謝します。

参考文献

- [1] T. Andrews, F. Curbera, H. Dolakia, J. Golland, J. Klein, F. Leymann, K. Liu, D. Roller, D. Smith, S. Thatte, I. Trickovic, and S. Weeravarana, "Business Process Execution Language for Web Services," 2003.
- [2] T. Ishida, "Language Grid: An Infrastructure for Intercultural Collaboration," in Proceedings of the IEEE/IPSJ Symposium on Applications and the Internet (SAINT-06), pp. 96-100, keynote address, 2006.
- [3] Y. Murakami, T. Ishida, and T. Nakaguchi, "Infrastructure for Language Service Composition," Second International Conference on Semantics, Knowledge, Grid (SKG-06), 2006.
- [4] T. Murata, M. Kitamura, T. Fukui, and T. Sukehiro, "Implementation of collaborative translation environment: Yakushiutenet," in Proceedings of the Ninth Machine Translation Summit System Presentation, pp.479-482, 2003.

⁵ <http://www.langrid.org/>

「The 翻訳 2007 ビジネス」

東芝ソリューション株式会社

1. はじめに

東芝ソリューション株式会社は、同義語や類義語といった言葉の概念を利用して翻訳精度を向上させたビジネス向け英日/日英翻訳ソフト「The 翻訳 2007 ビジネス」を商品化し、販売を開始しました。

The 翻訳シリーズは、文書の文脈や書式を利用する翻訳方式(CFエンジン)や、文書全体を先読みして得られた情報を利用する翻訳方式(2段階翻訳)を翻訳エンジンに実装することで、高い翻訳精度を実現してきました。

このCFエンジンと2段階翻訳に加え、The 翻訳 2007 ビジネスは、同義語や類義語といった言葉の概念を利用して文意に合った訳語を選択する「概念翻訳」を実装した最新の翻訳エンジンを搭載しています。

翻訳エンジンとともに訳質を左右する翻訳辞書は、最新の時事用語も含め、語数を470万語に増強しました。

また、今後ビジネスにおいて主流のOSになると見込まれるWindows Vista™にも対応しています。

2. The 翻訳 2007 ビジネスの主な特長

(1) 概念翻訳の搭載

概念翻訳では、同義語や類義語といった言葉の意味関係を体系づけた概念辞書(図1)を使用して、ある用語の用法をそれに意味的に近い語にも適用することができます。

例えば、「検査」と「人間ドック」は表面上は似ていません。しかし、概念辞書を参照すると「人間ドック」や「健康診断」は「検査」の一種であるということが分かり、「人間ドック」は「検査」と意味的に近い類義関係にあると判断します。一方、幅広い意味を持つ動詞「受ける」は特に指定がなければ receive のような訳語が使用されています。

receive が適切でない場合は、目的語などの共起する語により訳語を使い分けます。一例として The 翻訳は以下の規則を持っています。

- ・「検査を受ける」→“undergo an inspection”

「検査」と「人間ドック」が類義関係にあることから、この規則を「人間ドック」にも応用し、「人間ドックを受ける」という句を“undergo a complete

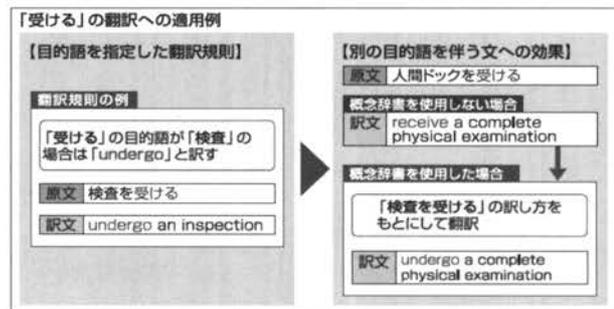
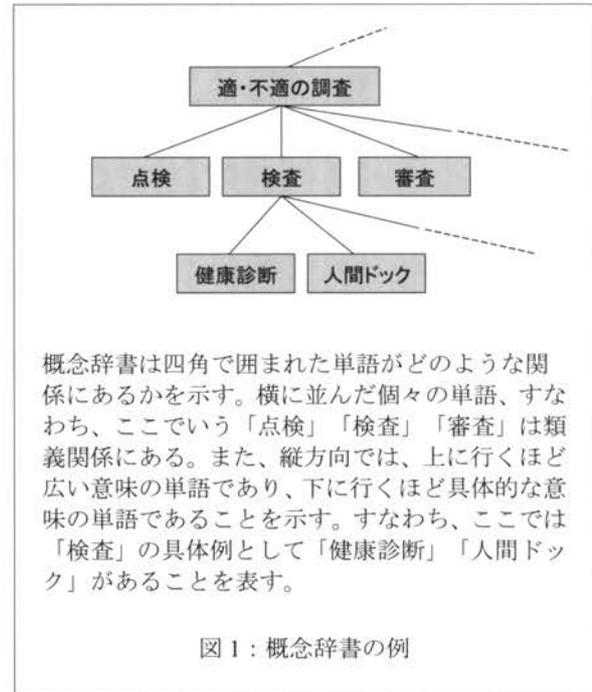


図2：「受ける」の翻訳への適用例

physical examination”と正しく訳すことができます(図2)。概念辞書を使用しない場合は、「人間ドック」を目的語に伴う「受ける」には receive のような訳語が使用されることとなります。

英日翻訳では例えば以下ようになります。The 翻訳は“deal with”の訳語を主語や前置詞 with の目的語によって決定する翻訳規則を持っています。この中には主語が report であるときは“deal with”は「取り組む」とする翻訳規則があります。一方、英語概念辞書では report は study という語と類義関係にあると位置づけられています。これらにより、

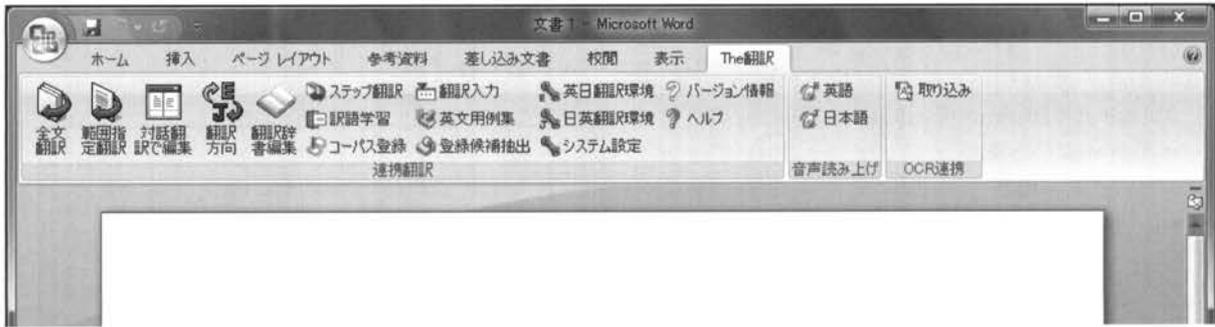


図 3 : Microsoft® Word 2007 のリボンインターフェースに対応した The 翻訳 2007 ビジネスのインターフェース

“Your study of civics will deal with citizenship in the family.”

という文は、下の対比に見られるように、より自然に訳出できるようになりました。

概念辞書不使用時：「市政学に関するあなたの研究は、家族の中の市民権に対処するでしょう。」

概念辞書使用時：「市政学に関するあなたの研究は、家族の中の市民権に取り組むでしょう。」

このように、The 翻訳ではひとつの翻訳規則の効果を他の類似表現の翻訳に応用して翻訳精度を高めています。

(2) 辞書語数の充実と翻訳規則の増強

新商品は、従来商品「The 翻訳オフィス V10」と比較し、179 万語を増強した 470 万語（標準辞書 202 万語+専門用語辞書 268 万語）の辞書を搭載しています。この中には、最新の用語も含まれ、日英には「クラウドソーシング」「情報発信力」といった用語が、英日には“thumb drive”“economic populism”といった用語が標準辞書に追加されています。

また、翻訳知識の強化により、翻訳精度の向上を図りました。

英日翻訳の例

原文：We should produce and use goods in a way that does the least amount of damage to the environment.

訳文： 私たちは、環境への被害を最小限にする方法で品物を生産し使用するべきです。

強化前の訳文： 私たちは、環境に損害の最小量を行う方法で品物をもたらし使用するべきです。

日英翻訳の例

原文：全ての食品添加物について、その発がん性の可能性を評価する。

訳文： All the food additives are evaluated for the carcinogenic possibilities.

強化前の訳文： The carcinogenic possibility is evaluated about all the food additives.

(3) 2007 Microsoft® Office、Windows Vista™ に対応

2007 Microsoft® Officeから採用されたリボンインターフェースに対応しました。これにより、Microsoft® Office上での翻訳作業効率を向上させました（図 3）。

また、Windows Vista™ の UAC(ユーザーアカウント制御)に対応し、ビジネスで重視されるセキュリティ面での強化を実施しています。

3. その他の機能

The 翻訳 2007 ビジネスは上記に加え、以下のような多彩な機能をそろえています。

- ①アプリケーション連携翻訳：Microsoft® Word、Microsoft® Excel、PowerPoint®、一太郎、秀丸エディタなどのアプリケーションに翻訳機能をアドインし翻訳作業ができます。
- ②メール翻訳：Outlook®や Outlook® Express、Windows Vista™ の Windows®メール、Lotus Notes®、EUDORA などのメールツール上で、The 翻訳の機能を組み込んで英文メールを簡単に翻訳できます。また、メールの受信と同時に翻訳を行う自動メール翻訳機能もあります。
- ③PDF 翻訳：Adobe®Reader®でテキストを取り出して PDF 文書を翻訳できます。

- ④ホームページ翻訳：ブラウザ翻訳を使用すると Microsoft® Internet Explorer®, Firefox などのインターネット閲覧ソフトでホームページを表示しながら翻訳することができます。特に Internet Explorer®では The 翻訳の機能を組み込むことにより、The 翻訳の各機能を使って翻訳することもできます。
- ⑤対訳翻訳：原文と訳文を見比べながら一文ずつ翻訳できます。また、サイドウィンドウ、ボトムウィンドウは使い方に合わせて配置できます。
- ⑥翻訳 OCR：印刷文書や画像データを文字認識して翻訳します。*
- ⑦セレクトコーパス翻訳：頻繁に使う分野の訳語を自動学習することができます。
- ⑧中日訳ふり機能：中国語を単語ごとに訳して表示します。ピンイン表示で発音の確認もできます。
- ⑨自動アップデートサービス：The翻訳 2007 ビジネスを常に最新の状態に保つことができます。

*特殊文字、特殊フォントが使用されている場合や、イメージ部分の元原稿の状態によっては、正しく認識されない場合があります。

4. おわりに

以上述べたように The 翻訳 2007 ビジネスは、(1)概念翻訳、(2)語彙・翻訳知識の増強、(3)2007 Microsoft® Office、Windows Vista™ 対応の三点の特長を持っています。第一の概念翻訳と第二の語彙・翻訳知識の増強では、単に単語の数といった量の側面にとらわれず、質を重視し、訳文をよりよくしていくことを目指しました。また、第三の 2007 Microsoft® Office、Windows Vista™ 対応は、これらがビジネス界でよく使われているツールであることから、使い勝手の向上のために重要な意味を持つと考えています。今後も翻訳の精度を向上させるとともに、利用者のニーズを十分に取り入れ、使いやすい商品を目指して開発を進めていきます。

5. 動作環境

| |
|---|
| 対応OS(全て日本語版) |
| Windows® 2000 Professional SP4 |
| Windows® XP Home Edition/Professional SP1/SP2 |
| Windows Vista™ Home Basic(32 ビット版) |
| Windows Vista™ Home Premium(32 ビット版) |
| Windows Vista™ Business(32 ビット版) |
| Windows Vista™ Ultimate(32 ビット版) |

| | |
|--------|---|
| 対応機種 | PC/AT 互換機 |
| CPU | Pentium®以上/Celeron®以上推奨 |
| メモリ | 128MB 以上(256MB 以上を推奨) (The 翻訳必要分) OS 分を含む推奨値 Windows® 2000 Professional ...320MB 以上 Windows® XP Home Edition/Professional ...384MB 以上 Windows Vista™...768MB 以上 |
| ディスク | 英日/日英双方向...940 MB 研究社新英和・和英中辞典...550MB |
| ディスプレイ | 解像度 800×600 以上(1024×768 以上を推奨) |
| スキャナ | TWAIN 対応のスキャナ |
| 音声環境 | SoundBlaster®互換のサウンドボード、マイク、スピーカ |

■製品情報 URL

<http://hon-yaku.toshiba-sol.co.jp/>

- Microsoft、Windows、Windows Vista、PowerPoint、Outlook、Internet Explorerは、米国Microsoft Corporationの米国およびその他の国における登録商標または商標です。
- Adobe、Adobe Readerは、Adobe Systems Incorporatedの登録商標または商標です。
- The翻訳、CFエンジンは東芝ソリューション株式会社の登録商標です。
- 記載されている商品の名称はそれぞれ各社が商標として使用している場合があります。
- 本製品の仕様は予告なく変更される場合があります。

「翻訳者を支援するオンライン多言語レファレンス・ツールの構築」 (椎茸プロジェクト) について

東京大学大学院 教育学研究科 影浦 峽、阿辺川 武

1. はじめに

我々は現在、主にオンラインの文書を翻訳し、オンラインで翻訳文書を公開しているボランティア翻訳者を支援するシステムを構築している。プロジェクトは4年計画で、現在は3年目を迎えたところである。これまで、人間の翻訳者を支援する様々なシステムが提案され、実装され、提供されており、TRADOSをはじめとする[14]翻訳者向けデータ管理システムは比較的良好に使われているが、これらはコンテンツよりもコンテンツを管理するメカニズムの側面で翻訳者を支援するものである。高度な処理メカニズムを備え、次の訳語候補を推定して表示するシステム等も提案されているが[3,5]、利用範囲は限られている。この一因は、翻訳者の要求とシステムが提供するものが一致していないことにある[1,6,7]。

そこで我々は、対象となる翻訳者のニーズを読みとり反映させるかたちで、オンライン翻訳者向け翻訳支援システムの構築を進めている。もちろん、翻訳者の現在の要望を忠実に反映させるだけでは、言語処理技術の可能性を十全に展開すればありえるかもしれない翻訳者の潜在的な要望を引き出せない可能性はあるが、それを考える一歩としても、まずは翻訳者の要望を十分に考慮する必要があると考えるためである。

本稿では、最初に翻訳とはどういう行為かを簡単に述べ、次いで伝統的に翻訳者が使っているレファレンス・ツールと参照すべき言語単位との関係を、翻訳者への聞き取りも交えて整理し、それを起点としたレファレンス・ツール/メカニズムの高度化方針を定義する。最後に、我々が実際に開発している翻訳支援システムのプロトタイプを紹介する。なお、本稿では、基本的に英日翻訳を想定する。したがって、英語が原言語、日本語が目標言語である。

2. 翻訳という行為

言語学や自然言語処理の領域では翻訳は言語的行為と見なされることが多いが、翻訳者への聞き取りおよび翻訳論[9,15]の検討からは、翻訳がX文Y訳

ではなくまさに翻訳であるための基本要件は、翻訳が言語的な行為にとどまるものではないという点にある。簡単に要約すると、第一に、翻訳の単位は文ではなくテキストであり、第二に、翻訳者が対象言語テキストを作る際の意志決定は原言語テキストに基づくだけでなく、むしろそれよりも一外部世界を捨象したとしても一原言語テキストが置かれている原言語のテキスト集合、それに対応する対象言語のテキスト集合、それらに対応する想定読者集合のテキスト的親和性に依拠してなされる。端的に言う、翻訳は言語的行為であるよりも社会的な産物を創り出す社会＝メディア的行為である。

ここから、翻訳支援における基本方針が定まる。

1. 意志決定は翻訳者にゆだねること。
2. 参照情報は言語のサンプル・コーパスからではなく、翻訳テキストと関連したテキスト、あるいは一応すべてを含むと見なしてよいアーカイヴから提供される必要があること。

実際、翻訳者の多くが、翻訳の際に唯一使えるレファレンス・ツール/メカニズムは、良質の辞書と図書館（オンラインの世界ではGoogleも図書館に相当する）であると語るの、この第二点目に対応している[1]。また、我々が進めているプロジェクトが「レファレンス・ツールの構築」という図書館学的なコンセプトを強く打ち出しているのもこの点に関わっている。

3. 翻訳者のニーズとツール

そこで、レファレンス・ツールの観点から翻訳者の要望を整理することにして、翻訳者によって細かい違いはあるものの、言語単位の観点からは、概ね、(1) 一般語、(2) 熟語・慣用句、(3) 固有名、(4) 専門語、(5) 引用句、(6) 簡単な連語、の6種類が区別される。

それらに対して、(1) どんな情報を参照するのか、(2) どんなツール/メカニズムを利用するのか、および既存のツールの内容と活用機能に対する満足度を三段階で整理したものを、表1に示す。ここで「図書館」としてある部分は、分野のテキストや百

表 1: 言語単位と参照情報・ツールおよび利用者の満足度

| 言語単位 | 参照する機能 | 参照するツール | 内容 | 活用機能 |
|--------|---------------------|-------------------|----|------|
| 一般語 | 原言語の意味・対象言語の表現 | 辞書 | ○ | ○ |
| 熟語・慣用句 | 原言語の意味・対象言語の表現 | 辞書 | ○ | × |
| 固有名 | 対象言語の表現・事実情報 | 人名辞典等、図書館・Google | △ | △ |
| 専門語 | 原言語の意味・事実情報・対象言語の表現 | 専門用語辞典、図書館・Google | △ | △ |
| 引用句 | 既存の訳 | 図書館・Google | × | × |
| 簡単な連語 | 対象言語の表現 | 図書館・Google | × | × |

科事典等、個人で所有できるレファレンス・ツールに収まらない様々な資料を参照することを意味する。

個別のレファレンス・ツールと図書館（あるいはGoogle）のような総合的なレファレンス・メカニズムとは異なるが、これらを一様なものととらえるならば、ツールに対する翻訳者の満足度は、内容と活用（検索）機能の2点から、表1のように整理できる。

一般語や熟語・慣用句については、各翻訳者がお気に入りを使っているいくつかの辞書で内容的にはほぼ満足している。ただし、熟語・慣用句については、検索機能が不十分である。固有名と専門語については、既存の辞典類が充実している領域ではそれなりに使えるものの、全体としては、辞典類では十分でない場合が多い。引用句や連語についても同様である。対応して、これらの言語単位を対象とするツールの活用機能については、辞典等のレベルでは、内容の不十分さから「求めている情報があるのかわからないか判断しにくい」ことが影響しており、図書館における探索では、文献探索は体系的にできても、言語単位の探索が体系的にできないことが、求める情報の発見に困難を引き起こしている。

4. ツールの高度化方針

前節での整理から、ツールの高度化方針を、参照言語単位の種類ごとに定義することができる。

4.1 一般語

一般語について、翻訳者は冊子体および電子辞書に満足している。したがって、翻訳支援システムが既存の優れた辞書を提供できるかどうかは内容的な問題となるが、これは、ビジネス・モデルの問題となる。活用機能について、オンライン文書の翻訳では、“理解.com”¹のような自動的な辞書引き機能が必要となるが、技術的な実現は容易であるため、

¹<http://www.rikai.com/>

問題はインタフェースの使い勝手に集約される。

4.2 熟語・慣用句

多くの翻訳者が熟語・慣用句の検索機能向上を求めている。比較的熟練した翻訳者でも、熟語・慣用句を相対的に苦手としている場合が少なくなく、頻繁に辞書を参照するが、その際、見出し語を推定して引かなくてはならないことが大きな問題となる。また、翻訳者は、熟語・慣用句を明確に認定する場合ばかりではなく、「このあたりは熟語と考えないと意味が通じない」といったかたちで消極的に認定することも多い。

したがって、熟語・慣用句については、既存の高品質辞書のエントリーとテキストとの自動的なマッチングが課題となる。とりわけ、辞書に掲載されている標準形とテキストに出現する異形との柔軟なマッチングが重要な言語処理技術上の課題である。実際、テキストに出現する異形は言語学が述べる異形の範囲よりもはるかに広い。

なお、少なからぬ翻訳者が、ただ一つの正解をシステムが提示するのではなく、いくつかの候補を適切に提示することが望ましいと指摘する。翻訳者は、最終的には翻訳の正解として採用しなくても、正解への確信度を高めるために可能な解釈を複数検討することが少なくない。翻訳者が検討する候補範囲から大きく外れない複数の候補を翻訳者の検討を促すかたちでうまく提示することが、熟語・慣用句検索高度化の具体的な目標になる。現在、我々は岡山大学を中心に挿入や置換による異形を検索し表示するシステムを構築し、翻訳支援システム上での評価を行っている[12]。

4.3 固有名

固有名は、レファレンス・ツールとしての内容的拡充が求められている。とりわけ重要なのは、エントリー量の拡充である。翻訳者は、人名に関する古典的なレファレンスとしては、いくつかの人名辞典

および図書館の各国概説テキストの巻末索引等を参照することが多い。オンライン上での参照は、原綴から音表記を類推して Google を参照し、周辺情報から該当する情報を判断するのが標準的な翻訳者の手続きである。

したがって、固有名辞典を拡充させるとともに、翻字規則等を用いた Google による固有名対訳検索システムを実装することが有益である。現在、名古屋大学を中心にこのモジュールの実装はほぼ終了し評価している段階である。

4.4 専門語

専門語は、固有名と同様、何よりも、見出し語の拡大が求められている。翻訳者は、該当分野の専門用語辞典を参照するほか、図書館でテキストの巻末索引等を参照する。オンライン上では、原語から語構成要素の訳を手がかりに Google を参照する手法がよく用いられる。したがって、専門用語辞典を拡充させるとともに、ために、語構成要素の訳関係なども積極的に利用した用語収集のメカニズムが求められる。これについては現在筑波大学を中心に開発が進められており[13]、予備的な評価も進んでいる。

4.5 引用句

引用句については、冊子体でもオンラインでも、現れうるものの全体を探すよりも、翻訳者が過去に見たものを探す探索が多い。このため、一般に、最初からテーマや分野で探索空間はしばられている。したがって、我々は、翻訳者が訳している文書と関連した文書で、既に訳された文書の翻訳対集合をオンラインで探し、それを対象範囲として引用句を検索するシステムを開発中である[11]。オンラインで公開されている翻訳文書を自動的に収集するシステムは、すでに予備的な実装ができています。

4.6 簡単な連語

簡単な連語とは、たとえば、“successful vote”のように、誰にでも意味はわかるが、いざ翻訳しようとすると日本語で自然な表現を探すことが難しいものであり、翻訳者の熟練度に大きく関わっている。これについては、とりわけ初心者翻訳者のニーズは高いものの、翻訳者自身も情報探索パターンを体系化できていない。現在、我々は、言語表現上の訳語の不自然さを評価し代替を提案する枠組みで、下訳と完成訳と原文の三対の比較分析を進めているが、これが上手く行くなれば、連語の翻訳候補検索もその中に取り込まれる可能性はある。ただし、このように定義された支援は、これまで述べてきた

いわゆるレファレンス・ツールの拡充という問題定義とは少し異なったところに位置づけられる。

5. QRedit: 翻訳者向けエディタ環境プロトタイプ

我々は、現在、上記のモジュールについて分担で開発を進めると同時に、インタフェースを含めた総合的なオンライン翻訳者支援エディタ環境プロトタイプ“QRedit”の構築を進めている。システムは、サーバ・クライアント方式で実装され、ユーザは Web ブラウザを通して利用する。図 1 に、QRedit の画面イメージを示す。エディタは原文書領域と翻訳文書領域とに分割される。利用者は環境により縦の分割と横の分割から作業しやすい分割を選ぶことができる。原文書領域と翻訳文書領域のスクロールは同期しており、原文書のパラグラフごとに同期ポイントが設定されている。

5.1 原文書領域の機能

翻訳者は原文書の URL を指定する（その場合システムはタグ解析を行いテキスト領域を抽出する）かテキストを貼り込み、辞書引きをオンにする。システムは単語列に対し辞書エントリとの照合を行なう。

辞書引き機能

翻訳者が原文書の単語にマウス・カーソルをあわせると、システムは翻訳候補をポップアップで表示する。辞書は現在三省堂『グランドコンサイス英和辞典』[10]と『英辞郎』[4]とが組み込まれている。辞書引きの際には、単純語の辞書引きだけでなく、挿入や置換を考慮した熟語のさまざまな異形まで引けるようになっている（図 1 のポップアップ画面は“with his big fat tongue in his big fat cheek”という出現形が、“with one's tongue in one's cheek”という熟語の辞書エントリと上手くマッチしているところを示している）[12]²。

翻訳表現候補の提示

システムは、2 種類の候補表示機能を備えている。一つは、原文書領域の中でのポップアップによる簡略化された表示であり、ここでは訳語候補だけが表示される。もう一つは、ウィンドウ内の大半を占有するポップアップ領域を生成し、辞書に含まれる見出し語の情報をすべて表示するものである。前者は

² この機能を十分なかたちで組み込んだ機械翻訳システム・翻訳支援システムは我々が 2006 年夏までに調べた限りでは存在しなかった。

ちょっとした確認に、後者は複数の辞書の記述を総合的に考慮しながら訳語や意味を深く検討する際に適している。

5.2 原文書領域と翻訳文書領域の関係

翻訳者は、マウスクリックによって、原文書領域中で表示させた訳語候補を翻訳文書領域に自動的にコピーできる。また、数値表現を自動的に日本語の表現に変換（“Seventeen thousand three hundred fifty”や“1.3 billion”を「1万7350」や「十三億」等に）しコピーする機能を有している。これらの機能により長い固有名詞や数値表現のミスタイプを防ぐことができる。一般に Web ブラウザ内では、マウスクリックを行なうとキーボードのカーソルもそのオブジェクトへと移動するが、本システムでは原文書領域内でどのようなマウス操作を行っても、キーボードカーソルは常に翻訳文書領域に留まるような工夫をしている。そのため、マウスとキーボードの独立性に慣れれば、翻訳文書作成作業のリズムを崩さずに辞書引きやコピーを行うことができる。

5.3 翻訳文書領域

翻訳文書領域では、オープンソースの WISWIG 型 Web エディタ FCKeditor³を用いている。文書は簡単な文字装飾も含めた HTML 形式あるいはテキスト形式で保存することができる。あらかじめ原文書と同じパラグラフ構造に分割されているため、翻訳者は好きな箇所から翻訳できる。

5.4 翻訳者の「気付き」を利用した強調表示

レファレンス・ツールの高度化という点では、複数の辞書や百科事典などシステムが提供できる情報源は、原則として多ければ多いほどよい。しかし一方で、膨れあがる情報を等価な方法で提示するだけでは、いたずらに画面が複雑になり、システムの使い勝手を悪化させる原因となる。また、4.2 で少し述べたことに関連するが、翻訳者の主体的な探索行動に受動的に応えるべき場合と、翻訳者の「気付き」を促すことが有効な場合とがある。例えば、単語は読み違いを除けば見落としようがないが、熟語に気付かなかつたり、ある表現が専門的な意味を持つことを見落とすといった場合もある。

そこで我々は、情報の性質と重要度に応じてユーザに「気付き」を促す機能考えている[2]。熟語・慣用句を例にとると、出現頻度が少なくその存在を知らないと訳出できないような難しい語句に関し

ては、「気付いて」もらうように強調表示の度合を高めたり、中高生レベルの簡単な語句では強調表示を行わないといった調節を可能にする。現時点では、語句の難易度、語句の構成、語句の専門性、情報源といった要素に基づき強調表示の度合を決定している。

6. おわりに

本稿では、レファレンス・ツールの高度化という観点から、翻訳者を支援するために必要となる基本的な要件を整理し、それに基づき我々が構築している翻訳支援機能とシステムの概要を述べてきた。

翻訳者への聞き取り調査から明らかになった重要な点は、第一に翻訳が言語的なレベルではなくテキスト＝メディア的なレベルでの要因に基づく言語上の意志決定行為であるということであり、第二にそれに対応して翻訳者は宙に浮いた構造的に参考になる言語表現例を求めているのではなく、具体的な言語表現に対して、これまで具体的に特定の場所で特定のかたちで与えられた表現や情報を求めているという点である[8]。その意味で、レファレンス・ツールの基本となるのは、探索の対象となっている単位そのものについての具体的な先行存在である。既存の高品質辞書と図書館が基本的な参照の枠組みとなっているのは、まさにこれがゆえであり、Google も、一つの歴史的な世界として、「そこになければあきらめる」という位置づけがなされている限りにおいて、参照に足るものと見なされている。

したがって、システムに求められるのは翻訳者に代わる意志決定ではなく翻訳者の意志決定を支援することとなる。また、内容の拡充が求められている固有名や専門語については、それなりに明確に規定された範囲の中で、歴史的に表現されてきた固有名や専門語を包括的にカバーすること、あるいは一定の有用な方針のもとで体系的かつ一貫した語彙を収集し提示することが求められる。これに対応して、評価方法も従来の自然言語処理的な評価とは異なってくる。具体的にどのような視点からどう評価を行うかについても、システム全体および個別のモジュールの開発とともに、今後、より明晰に定式化して配慮して行く予定である。

なお、本研究は、日本学術振興会科学研究費補助金基盤(A)「翻訳者を支援するオンライン多言語レファレンス・ツールの構築」(課題番号 17200018)の支援を得て行われている。

³<http://www.fckeditor.net/>

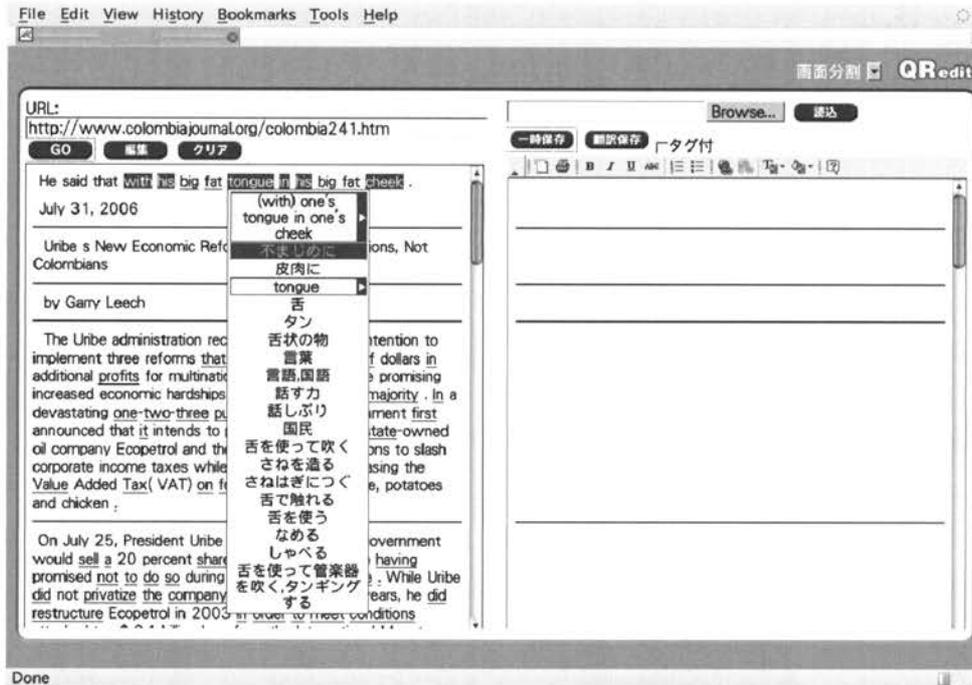


図 1: QRedit のスクリーンショット

参考文献

[1] 専門翻訳者・ボランティア翻訳者 8 名への第一著者による聞き取り。また、アンケート調査も行い 12 名から基本的な翻訳プロセスにおけるレファレンス・ツールの位置づけについて回答を得ている。

[2] Takeshi Abekawa and Kyo Kageura. A translation aid system with a stratified lookup interface. In Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Poster/Demos, 2007. 発表予定。

[3] Jorge Civera et al., A syntactic pattern recognition approach to computer assisted translation. Structural, Syntactic, and Statistical Pattern Recognition, Springer Berlin, pp.207-215, 2004.

[4] 英辞郎. <http://www.eijiro.jp/>, 2007.

[5] George Foster, Philippe Langlais, and Guy Lapalme. User-friendly text prediction for translators. In Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing, pp.148-155, 2002.

[6] Heather Fulford and Joaquin Granell Zafra. The uptake of online tools and web-based language resources by freelance translators: implications for translator training, professional development, and research. In Proceedings of the Second International Workshop on Language Resources for Translation

Work, Research and Training, pp.37-44, 2004.

[7] Heather Fulford. Translation tools: An exploratory study of their adoption by UK freelance translators. Machine Translation, Vol. 16, No. 3, pp.219-232, 2001.

[8] 影浦峽. 人間の翻訳におけるコーパスの位置づけ. 言語処理学会第 12 回年次大会発表論文集, pp.452-455, 2006.

[9] Jeremy Munday. Introducing Translation Studies: Theories and Applications. London: Routledge, 2001.

[10] 三省堂編修所. グランドコンサイス英和辞典. 三省堂, 2006.

[11] 品川哲也, 森辰則, 影浦峽. オンライン対訳文書対からのテキスト領域抽出とアライメント. 言語処理学会第 12 回年次大会発表論文集, pp.520-523, 2006.

[12] 竹内孔一, 金平昂, 平尾一樹, 阿辺川武, 影浦峽. 置換・挿入を考慮した異形イディオム検索システムの構築. 言語処理学会第 13 回年次大会発表論文集, pp.396-399, 2007.

[13] 外池昌嗣, 宇津呂武仁, 影浦峽, 佐藤理史, 阿辺川武. ウェブを用いた専門用語翻訳支援における多様な情報源からの信頼度情報の提示. 言語処理学会第 13 回年次大会発表論文集, pp.400-403, 2007.

[14] TRADOS. <http://www.trados.com/>, 2006.

[15] Lawrence Venuti. The Translation Studies Reader. London: Routledge, second edition, 2004.

科学技術文献を対象とする日中機械翻訳システム開発プロジェクト

独立行政法人情報通信研究機構 井佐原 均

科学技術が発展し、世界中が瞬時に情報を共有できるようになった現在、世の中に存在するさまざまな知識を活用できるかどうか、個人の生活の充実に大きな影響を与えるようになっていきます。知識の多くは言葉によって表現されており、日本語・英語・中国語といった言語の違いが、知識の流通と利用の大きな妨げになっています。このような言語障壁の克服をめざし、情報通信研究機構 (NICT)、科学技術振興機構 (JST)、京都大学 (黒橋禎夫教授)、東京大学 (辻井潤一教授、中川裕志教授)、静岡大学 (梶博行教授) の5機関は協力して、平成18年度からの5年間で科学技術文献を対象とする実用的な日中・中日機械翻訳システムを開発します。

この研究開発の一部は、科学技術振興調整費・重要課題解決型研究等の推進「日中・中日言語処理技術の開発研究」として実施されます。主として解析・翻訳エンジンの開発、プロトタイプシステムの開発・改良を情報通信研究機構および京都大学が担当し、言語資源の収集、構築を東京大学、静岡大学、科学技術振興機構が担当します。

このプロジェクトでは翻訳手法として、言語の構造をより深く考慮した用例ベース翻訳を用います。基盤となる技術として、日本語や中国語の解析システムの性能向上を図ります。また、用例翻訳の手法を科学技術文献等の長く複雑な文にも対応できるように改良を進めます (図1)。

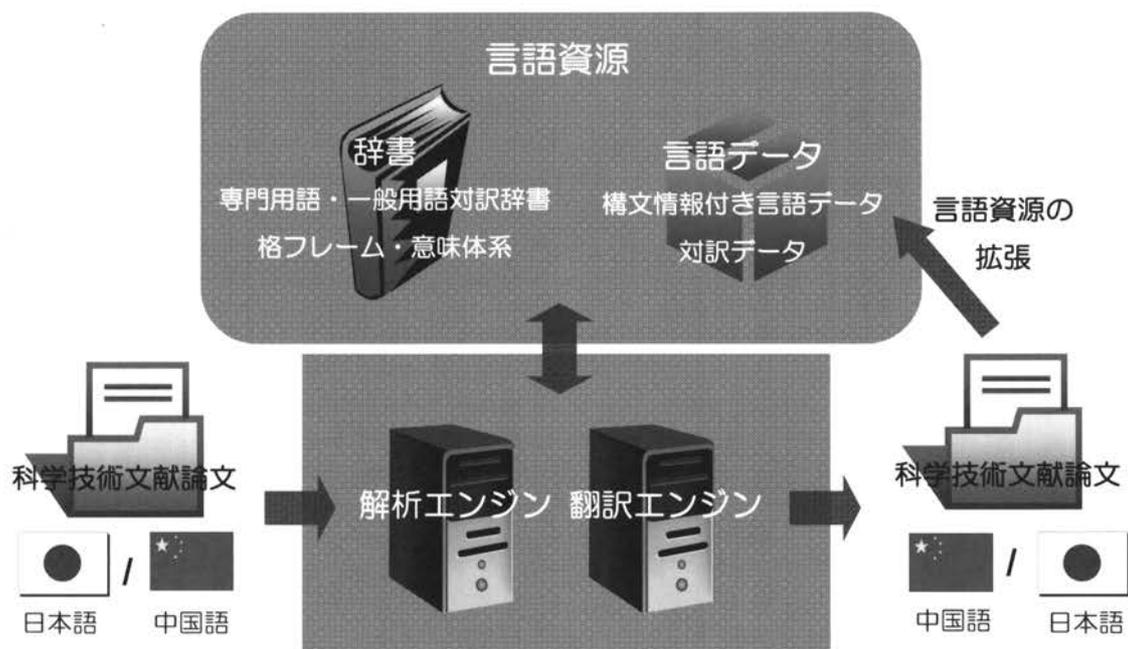


図1 日中・中日機械翻訳システムの概要

この手法の実現のためには、大量の用例（対訳コーパス）を蓄積する必要がありますが、我々は数百万文規模の日中对訳コーパスを開発する予定です。また、科学技術文献の翻訳・情報検索用辞書を対訳コーパスから半自動作成する手法を確立します。このようにして開発される大規模な言語資源を用いた機械翻訳プロトタイプシステムを開発して実証実験ならびに評価を行ないます。プロジェクトの途中段階でも、コーパス等の言語資源は可能な限り研究用に公開する予定です。

今回、開発するシステムの概念図を前頁に示します。実際のシステムは日中・中日双方向のシステムです。

欧米諸国と比べて、特にアジアにおいては英語での情報流通には困難が伴います。今回の研究開発においては、アジア諸国の一員としてのわが国の責務として、アジア言語の機械翻訳の実現を目指し、その第一歩として、特に科学技術の進展が顕著である中国を対象に科学技術文献を主たる対象とする機械翻訳システムの開発を行うこととしました。

この実現により、言語障壁のために中国国内のみで流通している有益な科学技術情報を、我が国の研究者・技術者・事業者が容易に活用できるようになり、共同研究事業の設立など大きなビジネスチャンスにも繋がるでしょう。また、日本が最先端を担う科学技術分野の文献が中国国内で流通することにより、中国における科学技術の発展も期待できます。

さらに、他のアジア言語の言語資源を整備すれば、解析・翻訳エンジンを大幅に変更することなく、それらの言語への展開が可能となります。各国語の言語資源の整備も本研究の成果の利用により効率良く行なえるようになることが期待できますので、日本語や中国語以外のアジア言語についても比較的容易に母国語で科学技術情報を検索・閲覧できるようになり、アジア各国の科学技術文献情報がアジア地域全体で流通しやすくなると期待されます。

これまで、機械翻訳の実装には、変換方式や中間言語方式などが用いられてきました。いずれの方式も人手によるさまざまな知識（文法規則、単語辞書、意味辞書など）の作成が必要であり、そのような知識の一貫性を持った作成は非常に困難です。

一方、人間が翻訳をするときには、このような知識を適用しているのではなく、過去に読み聞きしたことのある類似した文の訳を組み合わせて、翻訳をしているだろうという考え方から、用例翻訳の考えが AAMT の初代会長の長尾真先生により 1981 年に提案されました。当時はコンピュータの能力が十分ではなく、この手法を実用的なシステムに実現することは出来なかったのですが、近年、コンピュータの能力が向上し、また、文をそのまま例として使うのではなく、文法的に解析した上で、入力文と蓄積された用例文の類似性を判定する手法が開発されたことにより、実用的な用例翻訳システムを開発する基盤が整ってきました。

用例翻訳では、膨大な対訳コーパスに含まれる例文と入力文の類似性を利用して翻訳します。ここで必要となるのは用例となる対訳コーパスと、文同士の類似性を判定する方法であり、大規模な規則を作成する必要はありません。また、用例を追加することによって翻訳の質が向上する、また、用例の訳には前後関係による訳の違いが自然に含まれており、機械翻訳の訳文にもそれが反映される、といった特徴があります。

科学技術の目的は、あまねく人々に能力や地位に関わらず平等にすばらしい生活を提供することにあります。私たちはコンピュータに言語を処理する能力を与えることにより、世界中の人々が能力（語学力）や地位（通訳を雇えるか）に関わらず、言語障壁を意識せずすむ環境の実現を目指しています。今回の機械翻訳システムの研究開発により、そのような目標の一端が達成できると期待しています。

『機械翻訳技術のイノベーションシンポジウム』報告

独立行政法人情報通信研究機構 内元 清貴

機械翻訳技術のイノベーションシンポジウムは、科学技術振興調整費《重要課題解決型研究等の推進》「日中・中日言語処理技術の開発研究」プロジェクトのアウトリーチ活動の一環として、2007年3月5日、東京のJSTホールで開催されました。参加者は、主として、大学や企業、研究機関の研究者・技術者でした。当日の参加者数は定員の120名を超え、シンポジウムは大盛況のうちに終了しました。

プログラムは、NICTの長尾理事長とJSTの沖村理事長の挨拶に始まり、プロジェクトの研究代表者であるNICTの井佐原グループリーダーの概要紹介、海外からの招待講演3件、プロジェクト紹介の講演4件が続き、最後にクロージングという流れでした。

最初の招待講演者は南カリフォルニア大学(USC)情報科学研究所(ISI)のEduard Hovy教授で、「Innovations in MT: Enabling 'Shallow' Semantics-Based MT」というタイトルで講演してくださいました。現状の機械翻訳では統語構造までの処理を行なうのが主流ですが、Hovy教授は、さらにその先の意味処理の重要性を説き、成功例を挙げて意味処理の機械翻訳への利用可能性を示されました。その成功例というのは、古くから人工知能の



Hovy 教授の講演の様子

分野で研究されてきた知識表現、アブダクションなどの推論機構、および、最新の構文解析技術を用いることにより、ブートストラップ的にテキストから知識を自動獲得・拡張する枠組み（「Learning by Reading (LbR)」）が実現可能であるというものでした。現状で、教科書レベルなら高校の読解試験(AP Exam.)に合格できるような実用規模の意味処理システムが実現できるそうです。この枠組みが機械翻訳に適用できれば、テキストから自動的に意味知識を獲得・拡張しどんどん賢くなってくれるような機械翻訳システムができることになります。まさに夢のような話です。

二番目の招待講演者はマイクロソフトリサーチ(Microsoft Research)のWilliam Dolan博士で、「MSR-MT: Microsoft's MT Effort」というタイトルで講演してくださいました。マイクロソフトリサーチは世界に先駆けて用例ベース翻訳システムを実用化し、英語のマニュアルを他の言語に翻訳する際に利用して翻訳コストを抑えてきました。当初のシステムは、Logical Form (LF)と呼ばれる言語学的単位を用いることによりうまく用例を汎化していましたが、用例間の優先度をどう決めるかが大きな問題でした。ところが、近年、Treeletと呼ばれる依存構造木の部分グラフを用いることにより、言語学的に適切な単位をなるべく保持しつつ、統計翻訳の手法もうまく採り入れ、優先度の問題を解消することに成功しました。さらに、原言語の依存構造を用いて目的言語の構造を推定することにより、目的言語の構文解析システムがなくてもTreeletを獲得できる仕組みを考案し、英語から他の言語への翻訳のコストを大幅に削減することにも成功しました。このようなマイクロソフトリサーチの成功例には学ぶことが多いように思います。

Dolan博士は、インターネット上のテキストは英語以外の言語で書かれたものが3分の2を占めていることや、アジア言語の比率が増していること、アジア経済の発展によるコミュニケーションの増大などを背景に、今後、翻訳のニーズが増し、機械翻訳の重要性が高まっていくだろうと予測しています。また、機械翻訳の技術が機械翻訳に留まらず言い換えや文法誤りの訂正など広く応用できることを例示し、機械翻訳だけでなく機械翻訳「技術」の需要も増える可能性があることを示唆しました。しかし、そのためには学習データとなる言語資源が不足していることを指摘し、言語資源の重要性も強調しておられました。



Dolan 博士の講演の様子

三番目の招待講演者は中国科学院自動化研究所(CASIA)の Xu Bo 教授で、「Convergence of MT Methodology and Its Perspective」というタイトルで講演してくださいました。近年、統計翻訳の手法が活発に研究されており、その成果は目覚ましいものがあります。その着実な性能向上の様子は、HTRDP や NIST、IWSLT といった評価型ワークショップの報告より知ることができます。しかし、その発展の過程で機械翻訳の方法論は淘汰され、近年はどれも同じ方向に向かって収束していつているようにも見えます。Xu 教授は、講演の中で、統計翻訳の歴史を概観しつつその点を指摘しておられました。中国でも国をあげて同様の方向に向かっているようです。これまで、独自に評価型ワークシ

ョップや統計翻訳についてのワークショップを開催し、研究者が集まって方法論を吟味し、主要テーマを決めて皆で取り組むことにより、着実に機械翻訳の性能を向上させてきたようです。近年の主要な研究テーマとしては、例えば、固有表現などに対する前処理の手法、句に基づく統計翻訳の手法、統語構造に基づく統計翻訳の手法、複数の機械翻訳結果の融合方法などを例として挙げられていました。さらに、現在、その成果をベースとして、多言語環境における人間と人間あるいは人間と機械との対話、多言語情報検索、多言語資源の構築と共有などに関するプロジェクトが進行中とのことです。近年の中国における機械翻訳技術の発展には目を見張るばかりです。

シンポジウムの後半はプロジェクト関係者の発表で、専門用語対訳辞書、中国語の専門用語、用例ベース翻訳システム、タグ付き対訳コーパスの構築およびその基盤技術について、最新の研究内容を紹介するというものでした。このプロジェクトでは、100 万文対規模の日中对訳コーパス、日中専門用語対訳辞書などの言語資源や、中国語解析システム、用例ベース翻訳システムなどを開発し、順次公開する予定です。それらが様々な形で日本における機械翻訳技術の底上げに繋がることを期待しています。

インターネット機械翻訳サービスリスト

AAMT インターネットワーキンググループ

- ★ 本リストでは、AAMT 独自調査による機械翻訳サービスを行なっている Web ページを記載しています。辞書引きサービスは原則として非掲載です。本リストに記載されていない機械翻訳サービス、あるいは記載されているサービスの URL 変更/停止などがございましたらご連絡頂けると幸いです。
- ★ 言語欄は、翻訳前あるいは翻訳後の言語として指定可能な言語を表し、翻訳方向は表していません。
- ★ 以下のサイトにアクセスしたことによる一切の損害を当協会は補償いたしません。
- ★ 順不同

最終更新日：2007 年 5 月 7 日

| サイト(サービス)名 | サービス内容 | 対象言語 |
|----------------------|---|--------------|
| AltaVista Babel | Web/Text | 英仏独西伊葡中日韓露 |
| Fish Translation | http://world.altavista.com/ | |
| AMIKAI | Text | 英仏独西葡中日韓 |
| | http://www.amikai.com | |
| @nifty | Web/Text | 英日韓中 |
| | http://www.nifty.com/globalgate/ | |
| ATLAS | Text | 英日 |
| | http://software.fujitsu.com/jp/atlas/sample.html | |
| BIGLOBE | Web/Search | 英中日韓 |
| | http://honyaku.biglobe.ne.jp/ | |
| clickq Japan | Web | 日韓 |
| | http://www.clickq.com | |
| Cyber G-BOC | Web/Text | 英日中 |
| | http://www.g-boc.com/j/cyber/search/index.html | |
| 英語ナビ | Web/Text | 英日韓 |
| | http://eigo-navi.com/ | |
| ePal.com | Text | 英仏独西伊葡中 |
| | http://www.epals.com/translation/ | |
| Excite | Web/Text/MS Office/Search | 英日中韓 |
| | http://www.excite.co.jp/world/ | |
| Foreignword.com | Text | 25 言語 |
| | http://www.foreignword.com/Tools/transnow.htm | |
| Free Translation.com | Web/Text | 英仏独西蘭葡日ノルウェイ |
| | http://www.FreeTranslation.com/ | |

| | | |
|---------------------------------|---|---------------|
| Infoseek | Web/Text | 英中日韓仏独伊西葡 |
| | http://www.infoseek.co.jp/Honyaku | |
| 石川新情報書府 | Web | 英日 |
| | http://shofu.pref.ishikawa.jp/ | |
| J-Server | Web/Text/Chat/Mail | 英中日韓 |
| | http://www.j-server.com/index.shtml | |
| 金沢 100 万石まつり | Web | 英日 |
| | http://www.100mangoku.net/ | |
| 関西コンシェルジェ | Chat/Mail/BBS | 日韓 |
| | http://www.my-concierge.org/kansai/index.htm | |
| Language Engineering | Web/Text/File/Mail | 英仏独西伊葡中日 |
| | http://www.lec.com/ | |
| LogoVista e-Trans | Web/Text | 英仏独西伊葡日韓 |
| | http://e-trans.logovista.co.jp/ | |
| Microsoft Office Update Home | Text | 英日 |
| | http://software.fujitsu.com/jp/honyaku/eservice/index.html | |
| Newstran.comm | Web/Text/News | 英仏独西伊葡中日韓アラビア |
| | http://www.newstran.com/ | |
| Paran Go Japan! | Web | 日韓 |
| | http://gojapan.paran.com/index.html | |
| OCN | Web/Text | 日英韓 |
| | http://www.ocn.ne.jp/translation/ | |
| ParSit | Web/Text | 英泰 |
| | http://suparsit.com | |
| Poltran.com | Text | 英葡 |
| | http://www.poltran.com/ | |
| ProMT | Web/Text/Mail | 英仏独西伊露 |
| | http://www.online-translator.com/text.asp?lang=en | |
| Q 州マイーコンシェルジェ | Web/Chat/Mail/BBS | 日韓 |
| | http://www.my-concierge.org/kyushu/ | |
| ReaD | Text | 英日 |
| | http://read.jst.go.jp/ | |
| Reverso | Web/Text | 英仏独西 |
| | http://www.reverso.net/ | |
| Rustran.com | Text | 英露 |
| | http://www.rustran.com/ | |

| | | |
|-----------------------------|---|-------------|
| Sakhr | Text | 英アラビア |
| | http://mason.gmu.edu/~aross2/sakhr.htm | |
| 瀬戸内・夢海道コンシェルジェ | Chat/Mail/BBS | 日韓 |
| | http://www.my-concierge.org/shikoku/syodoshima/ | |
| Sister Cities International | Text | 英仏独西伊葡日 |
| | http://www.sister-cities.org/sci/resources/translation | |
| So-net | Web/Text | 英日 |
| | http://www.so-net.ne.jp/translation/ | |
| SPORTS CHANBARA | Web | 英仏独西伊葡中日韓露泰 |
| | http://www.internationalsportschanbara.net/entry.html | |
| sScience.com | Text | 英日 |
| | http://www.sscience.com/index_J.htm | |
| SYSTRANBox | Web/Text | 英仏独西伊葡中日韓露 |
| | http://trans.voila.fr/ | |
| Systranet.com | Web/Text/File | 英仏独西伊葡希 |
| | http://www.systranet.com/systran/net | |
| 特許庁 | 特許公報 | 英日 |
| | http://www.ipdl.ncipi.go.jp/homepg_e.ipdl | |
| TransWalker JC | Text | 日中 |
| | http://www.kodensha.jp/jis/soft/pdamt/index.html | |
| TransWalker JK | Text | 日韓 |
| | http://www.kodensha.jp/jis/soft/pdamt/index.html | |
| TransWalker JET | Text | 日英 |
| | http://www.kodensha.jp/jis/soft/pdamt/index.html | |
| VIL Net | Web/Text | 英日 |
| | http://www.vil-net.jp/hon/ | |
| 金橋訳港世界通 | Web/Text/Mail | 英中日 |
| | http://www.netat.net/ | |
| マジカルゲート | Web/Mail | 英中日韓葡仏伊西独 |
| | http://www.magicalgate.net/ja/index.html | |
| Al-Misbar | Web/Text | 英アラビア |
| | http://www.almisbar.com/salam_trans.html | |
| コネクト・インタナショナル | Mail | 英日 |
| | http://www.connect-i.co.jp/flextran/index.html | |
| Tour2Korea | Web | 日韓 |
| | http://japanese.tour2korea.com/12Home/translator.asp?konum=7 | |

| | | |
|------------------------------|---|--------------|
| 韓日文化交流 | Web/Text/Mail | 日韓 |
| | http://www.kjclub.com/jp/ | |
| alittera translation service | Text | 英仏独伊西葡露 |
| | http://www.alittera.com/Auto.asp | |
| livedoor Translate | Web/Text | 英日中韓 |
| | http://translate.livedoor.com/ | |
| livedoor ギガメーラー | Mail | 英日中韓 |
| | http://docs.livedoor.com/personal/ | |
| Worldlingo | Web/Text/Mail | 日中蘭英仏独希伊韓露葡西 |
| | http://www.worldlingo.com/ja/products_services/computer_translation.html | |
| Tango Town | Web | 日英 |
| | 携帯電話専用 | |
| NewsWatch | Text | 英日 |
| | http://www.newswatch.co.jp/translate/ | |
| LISA | Text | 英独仏蘭希西伊葡中日韓露 |
| | http://www.lisa.org/ | |
| 訳してねっと | Web/Text | 英日中 |
| | http://www.yakushite.net/ | |
| Google Language Tools | Web/Text | 英独仏西葡 |
| | http://www.google.com/language_tools?hl=en | |
| CJT Survey | Mail | 日英中韓仏伊葡 |
| | http://www.cjt-survey.com/contents/mail-trans.htm | |
| Dictionary.com | Text | 日英韓中仏独蘭希伊葡露 |
| | http://dictionary.reference.com/translate/text.html | |
| DreamWiz | Web/Text/Search | 日韓 |
| | http://japan.dreamwiz.com/ | |
| Fresh eye | Web/Text | 日英中 |
| | http://www.fresheye.com/ | |
| Babelplex | Search | 英中蘭仏独希伊日韓葡露 |
| | http://www.babelplex.com/ | |
| Yahoo! Japan | Web/Text | 英中日韓 |
| | http://honyaku.yahoo.co.jp/ | |
| メールピア | Web 変更情報 | 英日 |
| | http://www.asahi-net.or.jp/service/mail/mailpia/ | |
| Mitaka - the big word | Web/Text | 英仏独伊西葡日 |
| | http://www.thebigword.com/MachineTranslation.aspx | |

| | | |
|------------------------|---|--------------|
| YAMATO 英和 | Text | 英日 |
| | http://ogma.atnifty.com/ogma.htm | |
| 地球の歩き方 | Text | 英日韓中独葡 |
| | http://www.arukikata.co.jp/translation/ | |
| GoKorea 日韓フレンズ | Web/Chat | 韓日 |
| | http://www.gokorea.jp/ | |
| モバイル Yahoo! | Text | 日英 |
| | http://honyaku.mobile.yahoo.co.jp | |
| プラザトウキョウ・ドット・コム | Text | 日英西伊葡仏露 |
| | http://www.plazatokyo.com/translate.php | |
| e 翻訳ドットネット | Search | 日中韓英仏伊西独葡 |
| | http://ehonyaku.net/ | |
| Yahoo! Babel Fish | Web/Text | 13 言語 |
| | http://babelfish.yahoo.com/ | |
| 寶貝魚 | Web/Text | 13 言語 |
| | http://fanyi.cn.yahoo.com/ | |
| 熟考 | Web/Text | 英日 |
| | http://www.rozetta.jp/index.html | |
| エキサイト韓たび | Web/Text | 日韓 |
| | http://media.excite.co.jp/Travel/world/kantabi/ | |
| IBM Translation Server | Web | 英日葡仏独中伊韓西 |
| | http://www-306.ibm.com/software/pervasive/tech/demos/translation.shtml | |
| EnjoyKorea | Web | 日韓 |
| | http://www.enjoykorea.jp/translation/translation_main.html | |
| ImTranslator | Text | 英中独蘭仏希伊日韓葡露西 |
| | http://freetranslation.paralink.com/ | |
| Voila | Web/Text | 英独蘭仏伊葡露西 |
| | http://trans.voila.fr/ | |

協会活動報告

(2007年3月～2007年5月)

決算理事会

2007年3月 (E-mailにて協議)

- | | |
|---------------------|---------------------|
| ① 平成18年度活動報告(案)について | ② 平成18年度決算報告(案)について |
| ③ 平成19年度活動計画(案)について | ④ 平成19年度収支予算(案)について |

監事会

2007年4月23日 平成18年度会計報告

機械翻訳課題調査委員会

2007年3月12日

- | | |
|----------------|-------------------|
| ① 前回委員会の議事録の確認 | ② WG③の活動内容についての議論 |
| ③ 各WGの活動について | ④ 活動内容の報告(各WGから) |
| ⑤ 活動内容についての議論 | ⑥ まとめと次回委員会について |

2007年4月11日 無料翻訳サイト運営会社へのヒアリング

2007年4月16日

- | | |
|------------------|---------------|
| ① 前回委員会の議事録の確認 | ② 各WGの活動について |
| ③ 活動内容の報告(各WGから) | ④ 活動内容についての議論 |
| ⑤ まとめと次回委員会について | |

AAMT/Japio 特許翻訳研究会

2007年3月15日

- | | |
|----------------------|--------------------|
| ① 平成18年度活動報告書の作成について | ② MTサミットでのWS開催について |
| ③ 来年度の活動について | |

2007年4月20日

- | | |
|-------------------|--------------------|
| ① 平成18年度活動報告書関連報告 | ② MTサミットでのWS開催について |
| ③ 今年度の活動について | |

2007年5月25日

- | | |
|--------------|--------------------|
| ① 今年度の活動について | ② MTサミットでのWS開催について |
|--------------|--------------------|

編集委員会

2007年3月5日

- ① No.39 発行前の最終確認
- ② No.40 企画の決定
- ③ No.41 以降の企画について

インターネット WG

- ① AAMT ホームページの更新（毎月）
- ② AAMT Forum での情報発信（月数回）
- ③ 会員専用ページの準備
- ④ 委員会活動におけるメーリングリストの管理
- ⑤ その他ネットワーク関連のインフラ管理全般

MT SUMMIT XI

September 10-14, 2007

Copenhagen Business School, Copenhagen, Denmark

European Association for Machine Translation (EAMT)主催による第11回機械翻訳サミット(MT Summit XI)が2007年9月、デンマークのコペンハーゲンにて開催されます。MT Summit XIは、機械翻訳分野の研究者、開発者、ユーザ等の幅広い関係者を対象としています。主要な学術、民間、政府機関等による最新の研究発表やデモが予定されています。サミットの本会議では、研究論文発表、ユーザレポート、招待講演、パネル討論等が用意されています。更に、本会議前後の拡張プログラムとして、チュートリアル、ワークショップ、デモセッション等が予定されています。このように機械翻訳にかかわる幅広い層の参加者が一堂に会することにより、言語障壁やデジタルディバイドを克服するための議論が活発に行われるものと期待しております。

チュートリアル (2007/9/10)

Morning session 9:00 AM - 12:30 PM / Afternoon session 14:00 AM - 17:30 PM

- T1 (full day) "Statistical MT" by Kevin Knight (University of Southern California, USA) and Philipp Koehn (University of Edinburgh, UK)
- T2 (full day) "EBMT for cross-lingual IR" by Cristina Vertan (University of Hamburg, Germany)
- T3 (morning) "Context-based evaluation of MT systems: Principles and Tools" by Maghi King, Andrei Popescu-Belis and Paula Estrella, (ISSCO, University of Geneva, Switzerland)
- T4 September 10 (afternoon) "Using free online MT in multilingual websites" by Federico Gaspari and Harold Somers (University of Manchester, UK)
- T5 September 10 (afternoon) "Workflow and business information management" by Adriane Rinsche (Language Technology Centre, Ltd, UK)
- T6 (morning) "How to Successfully Integrate MT Tutorial" by Jen Doyon (MITRE, USA) and Laurie Gerber (Language Weaver, USA)

ワークショップ (2007/9/11)

Morning session 9:00 AM - 12:30 PM / Afternoon session 14:00 AM - 17:30 PM

- W1 "On patent translation" by Jun'ichi Tsujii, (University of Tokyo, Japan, Manchester University, UK) and Shoichi Yokoyama, (Yamagata University, Japan)
- W2 "The Chinese Room Experiment: The Self-Organizing Feng Shui of MT" by John S. White (SYSTRAN Software, USA)
- W3 "Using Corpora in Natural Language Generation and Machine Translation (UCNLG+MT)" by Anja Belz (University of Brighton, UK) and Sebastian Varges (Dipartimento di Informatica e Telecomunicazioni, Italy)

本会議 (2007/9/12~14)

- 口頭発表・ポスター発表・デモ展示など

MT Summit XIへの皆さまのご参加をお待ちしております。プログラムの詳細や参加申込み等に関しては、公式ウェブページをご参照ください。 <http://mtsummitcph.ku.dk>

MT Summit XI Workshop on Patent Translation

September 11, 2007

Workshop Theme

Patent documents are one of the major application areas of machine translation. The workshop aims to foster research and development of the technology for patent translation by providing a forum in which researchers and practitioners can exchange their ideas, approaches, perspectives, and experiences from their work in progress.

Double submission of papers to both the workshop and the main conference will also be welcomed.

The workshop will consist of 1-3 invited talk(s), presentation of submitted papers, and one panel discussion.

Areas of Interest

Topics of interests include, but are not limited to: Analysis and classification for patent documents, MT and translation aids for patent documents, Contrastive studies for multilingual patent documents, Language resources for patent translation, Information extraction from patent documents, Evaluation techniques for patent translation, Multilingual patent classification and retrieval.

Submission of papers

The workshop invites original submissions in any of the topics mentioned. Any types of papers, including research papers, survey papers and reports on user's experiences, are welcomed. Submissions should be A4, single-column format and not be longer than 3,500 words including figures, tables and references.

Authors are requested to keep their papers anonymous and submit separate cover pages with the following information: paper title, author's name(s), affiliation(s), address(es), and e-mail address(es), 100-200 word summary, some keywords for classifying the submission.

Electronic submissions in PDF format are strongly preferred. The two electronic files should be attached to an email and sent to the following address: shimohata245@oki.com

Important Dates

Paper submission deadline: 15 June 2007

Notification of acceptance: 30 June 2007

Camera ready submission deadline: 20 July 2007

※ 本ワークショップは、AAMT/Japio 特許翻訳研究会のメンバーが中心となって開催されます。

AAMT ジャーナル編集委員会委員長
筑波大学大学院システム情報工学研究科
知能機能システム専攻
宇津呂 武仁

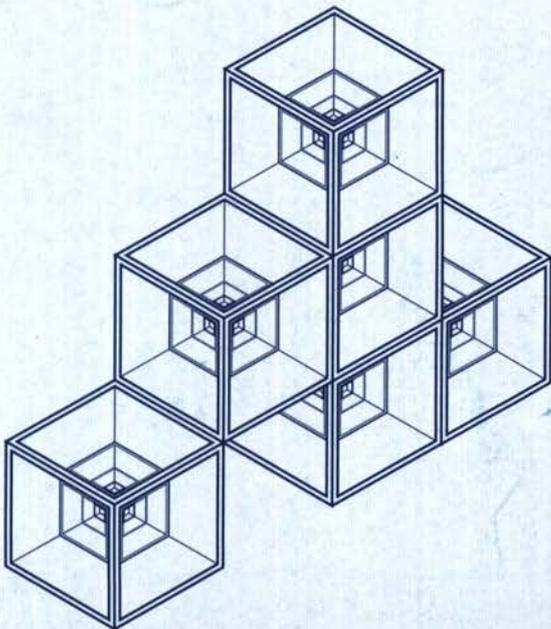
AAMT ジャーナル 40 号をお送りします。

ここ数年、自然言語処理分野におきましては、機械翻訳に関連する研究活動が活発になってきており、大型の研究プロジェクト等もいくつか行われるという状況になっています。そのような動きをとらえようと、今号では、いくつかのプロジェクトの紹介や関連技術報告、イベント報告などが掲載されています。また、インターネットワーキンググループからは、インターネット上の機械翻訳サービスリストに関する調査報告を掲載しています。

一方、AAMT が関係する機械翻訳関係のイベントとして、本年9月にデンマークのコペンハーゲンにおきまして、第11回機械翻訳サミットが開催され、また併設ワークショップとして、AAMT/Japio 特許翻訳研究会のメンバーが中心となって、特許翻訳ワークショップも開催されます。委員の皆様方の多数のご参加をお待ちしております。

また、前号の編集後記におきまして、新企画の芽として、翻訳サービス提供側の会員と翻訳サービス享受側の会員がもっと相互に知り合うことを目指して、会員の皆様に、自己紹介・機械翻訳および翻訳業界に期待すること・AAMT へのご要望等をご自由に執筆頂いた「AAMT 個人・法人会員特集（仮題）」なるものを企画致しまして、会員の皆様方にご寄稿のお願いをさせて頂きました。しかし、残念ながら、現在までのところでは、特集号を発行するに足るご寄稿を集めることができず、苦戦しております。引き続き、会員の皆様方からのご寄稿をお待ちしておりますので、ご寄稿のほどをお願い致します。

AAMT



AAMTジャーナル No.40

発行：アジア太平洋機械翻訳協会（AAMT）

ホームページ：<http://www.aamt.info>

住所：〒619-0289 京都府相楽郡精華町光台3丁目5番地

独立行政法人 情報通信研究機構 自然言語グループ内

phone：0774-93-4625 fax：0774-93-4627

編集委員会：宇津呂 武仁 大倉 清司 熊野 明

松川 淑子 村上 嘉陽

事務局：神崎 享子 和泉 絵美

印刷所：株式会社ナビックス

Asia-Pacific Association for Machine Translation

c/o Computational Linguistics Group, NICT

3-5 Hikaridai, Seika-cho, Soraku-gun, Kyoto, Japan 619-0289

Phone:+81-774-93-4625 FAX:+81-774-93-4627

URL:<http://www.aamt.info>