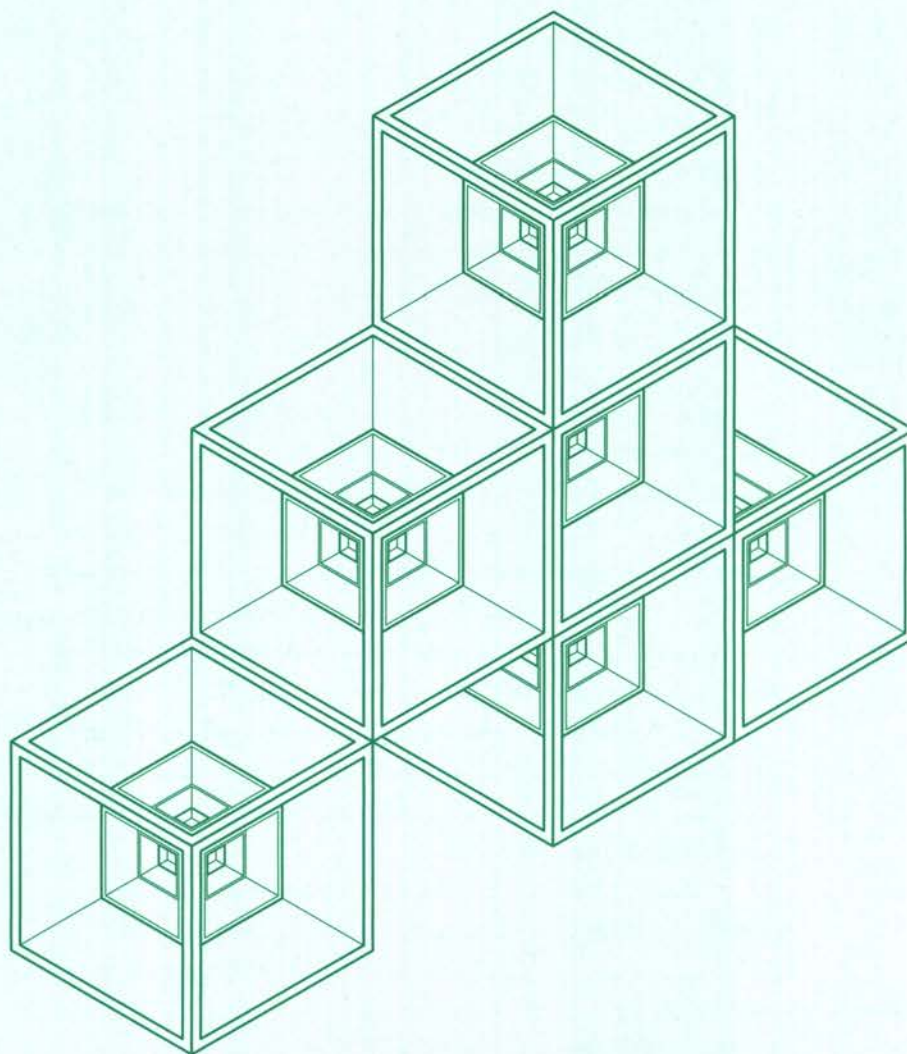


AAMT

Asia-Pacific Association for Machine Translation

Journal



June 2010

No.47

アジア太平洋機械翻訳協会

目 次

巻頭言：	機械翻訳国際連盟会長に就任してー今、挑戦すべきことー.....井佐原 均..... 1
特別寄稿：	韓国・朝鮮語翻訳の易しさと難しさ..... 蓮池 薫..... 2
シンポジウム報告：	機械翻訳シンポジウム「機械翻訳研究の課題と新方式の展望」参加報告.....住田 一男..... 4
追悼池原悟先生：	井佐原 均..... 6
	再録：非線形言語モデルと重文複文の意味類型パターン化 (No.40 June 2007).....池原 悟..... 7
レポート：	日本語からアジア諸言語や手話への機械翻訳ー古典的方式での構築ー.....池田 尚志..... 15
	A comparison of the two major English-Japanese free translation sites, Excite Translation and Yahoo! Translation.....杉浦 滋子..... 20
連載企画(第三回)：	言葉と人材ーグローバリゼーションの再考と日本ー.....九門 崇..... 29
新サービス紹介：	地方自治体が本格的に導入を始めるホームページ自動翻訳サービス「My サイト翻訳」 (株)高電社..... 32
委員会活動報告：	China Workshop on Machine Translation (CWMT09) 参加報告 AAMT 機械翻訳課題調査委員会 WG1... 36
	LISA China Focus 発表報告..... AAMT 機械翻訳課題調査委員会 WG3... 39
	ユーザー辞書共通フォーマット UTX (英語版)..... AAMT 機械翻訳課題調査委員会 WG3... 40
会員紹介：	AAMT 会員のひろば (第 7 回)..... 46
活動報告：	協会活動報告 (2009 年 11 月～2010 年 5 月)..... 53
お知らせ：	事務局からのお知らせ..... 56
編集後記：	 宇津呂 武仁... 57

CONTENT

Foreword:	Greeting from the president of AAMT..... <i>H. Isahara</i> 1
Special Issue:	Translating works from Korean Language: What is difficult and what is easy.... <i>K. Hasuike</i> 2
Report:	Report on Symposium "Problems of Machine Translation Studies and Prospects for the New Methods"..... <i>K. Sumita</i> 4
In memory of Professor Satoru IKEHARA:	 <i>H. Isahara</i> 6
	Non-compositional language model and semantic pattern dictionary for complex and compound sentences (No.40 June 2007)..... <i>S. Ikehara</i> 7
Report:	Machine translation from Japanese to Asian languages and Japanese sign language by a classical method..... <i>T. Ikeda</i> 15
	A comparison of the two major English-Japanese free translation sites, Excite Translation and Yahoo! Translation..... <i>S. Sugiura</i> 20
Serial Lecture:	Language and HR・Rethinking globalization and how Japan should react..... <i>T. Kumon</i> 29
New Services:	"My Site Translation", Machine translation service for website that local governments buckle down to introduce.... <i>KODENSHA</i> .. 32
Committee Report:	Report on the 5th China Workshop on Machine Translation..... <i>MT Committee for seeking future direction of MT</i> 36
	Report on LISA China Focus 2009..... <i>MT Committee for seeking future direction of MT</i> 39
	Common Format for User Dictionary UTX..... <i>MT Committee for seeking future direction of MT</i> 40
AAMT Members:	AAMT Agora..... 46
AAMT Activities	AAMT Activities (November 2009～May 2010)..... 53
Information:	Information from AAMT..... 56
Editor's Note:	Message from the Chair of the Editorial Committee..... <i>T. Utsuro</i> 57

機械翻訳国際連盟会長に就任して ー今、挑戦すべきことー

アジア太平洋機械翻訳協会会長
豊橋技術科学大学 井佐原 均

昨年、オタワで開催された機械翻訳サミットで機械翻訳国際連盟(IAMT)の会長に就任いたしました。2年間の任期の間、アジア太平洋機械翻訳協会(AAMT)の会長職に加え、微力ながらも、さらに国際的観点から機械翻訳の発展に尽くしたいと思っております。皆様のご指導・ご協力をよろしくお願いいたします。

世界的な経済不況の中、どのように経済を成長させていくかが焦眉の急です。情報産業は、(1)その成長が直接的に経済成長に寄与する、(2)その成長が他の産業の競争力を強化し、間接的に寄与する、の二つの面で経済成長を牽引すると言われています。機械翻訳をはじめとする翻訳支援技術は、多言語での情報提供による市場獲得など、特に(2)の観点において重要なものであり、言語を制するものが国際市場を制すると言えるでしょう。

このような状況に日本は十分に対応できているでしょうか？もちろん、AAMTはヨーロッパ機械翻訳協会(EAMT)やアメリカ機械翻訳協会(AMTA)と協力してIAMTの一角をなすものであり、アジアだけを考える場ではありませんし、ましてや日本だけを考える場ではありませんが、90年代には機械翻訳をはじめとする情報技術で世界を席卷していた日本の現状を考えることは、有意義なことだと考えます。

ここ10年ほどの間に、いくつかの欧米のグローバル企業が機械翻訳システムを使って、自社製品の日本でのローカライゼーションを始めています。日本でのビジネスのためには、大量の文書を高速に日本語化することが必須であると考えてのことでしょう。このような企業は、もともと欧米言語間での翻訳に機械翻訳システムを使っていたという経緯もあるでしょうが、欧米発のシステムを使うところが多いようです。このような、ビジネスのための企業内文書の翻訳、また翻訳

ビジネスにおける機械翻訳システムといった動きのほかに、ネット上での無料の翻訳サービスも行われています。ここではコーパスを用いた機械翻訳が使われ、翻訳結果を収集して、翻訳精度の向上に利用するという動きもあるようです。

このような状況の下、日本のグローバル企業は、どのようにして自社の文書を翻訳しているのだろうか？日本発の機械翻訳システムは、どの程度使われているのだろうか？など、AAMTとして、きちんと状況把握をしておかなくてはならないと考えています。

機械翻訳はローカライゼーションに有効ですが、ここでは単体のシステムだけではなく、辞書作成や後編集といった翻訳過程全体としてのサービスが重要となります。長年に互るルールベースの蓄積を持つ日本企業には厳しい状況かもしれませんが、現在はコーパスを用いた翻訳システムが主流となっています。コーパスを用いたシステムの性能向上には、対訳データの蓄積が必須であり、著作権処理も含めたデータの共有の枠組みがますます重要になるでしょう。

ここ10年ほどで、言語処理技術、言語資源、翻訳市場など、機械翻訳を取り巻く環境が大きく変化し、新しい流れが起こっています。日本をはじめ、多くのアジア諸国のように、英語を母語としない国が自国の製品で国際競争を勝ち抜いていくには、母語から多言語への高速かつ低コストでの翻訳が必須となるでしょう。もちろん今でも企業や大学でご努力いただいておりますが、さらに新しい技術開発に向けた国を挙げての取り組みが求められます。統合的な翻訳支援環境の構築が重要であり、母語に着目した研究開発も考えられます。技術文書などに対象を特化した制限言語や、入力文を翻訳しやすい形に変換するとともに曖昧な部分を指摘する支援システムなども重要になってくるでしょう。

韓国・朝鮮語翻訳の易しさと難しさ

蓮池 薫

日本語と韓国・朝鮮語は、語順がまったく同じであると同時に、漢語系の単語がほぼ90%セント以上一致している。これらによって韓国語・朝鮮語を日本語に翻訳するのが、比較的容易であるのは事実だが、似ているがゆえに落とし穴もある。

そのまま単語を置き換えていけばかなり意味が通じるため、いわゆる直訳になりやすいのだ。その結果、訳文が韓国語式の日本語になりがちである。

1つは漢語の動詞が多くなることである。

日本語の語彙に占める漢語の比率は50%前後なのに対して、韓国語は70%に達すると言われる。そのため、直訳していると漢語の多い日本語となり、かなり固い文章になりがちだ。

「14～16世紀にかけて西ヨーロッパの各地ではルネサンス運動が展開された」

韓国語を直訳した文章だが、一目見た限り悪い文章ではないし、意味もはっきりと伝わる。ところが多少固いイメージを否めない。

「展開」という漢語の代わりに「ルネサンス運動が起きた」と訳せばなんら違和感を与えないままやさしい表現になる。

「会社では人員整理を進行させている」

「進行」という漢語も韓国語ではよく使われるが、注意すべき単語である。日本語では「行う」「行われている」といった表現に訳したほうがこなれた感じを与える。

韓国では「〇〇化」という表現や「〇〇的という」表現がかなり多いが、これらも日本語らしく訳すべきだろう。

「蒸気機関の使用は英国の産業発展を加速化させた」という韓国語の文章は、「蒸気機関の使用は英国産業の発展を加速させた」もしくは「発展を促した」と訳すべきだろう。

ほかにも「人間の姿を描写した作品」では、「描写」という漢語の代わりに「描いた」という固有語を使う

と柔らかくなる。「継承する」を「受け継ぐ」、「統制を強化する」を「統制を強める」、「国が滅亡した」を「国が滅んだ」、「地中海の制海権を掌握して」→「地中海の制海権を握って」なども同様の例である。訳で漢語を使ってはならないということではなく、日本の固有語とうまくミックスして使うべきだ。

2つ目は日本語に多い受動態が使われなくなることだ。韓国語は日本語ほど受動態表現が多くなく、能動態になる。そのため表現は明確で、簡潔で力強い文になりやすい。それをそのまま訳す場合もあるが、主語を変えて受け身表現にすることも必要だ。

「1941年、日本がフランス領インドシナを占領すると、アメリカ、イギリス、中国、オランダは日本に対する経済封鎖を行った」

このような韓国語文は、「1941年フランス領インドシナを占領した日本は、アメリカ、イギリス、中国、オランダによって経済封鎖を受けた」などと訳すことができる。

「子どもが泣いたので、仕方なく父は家族を動物園に連れて行った」は「子どもに泣かれて仕方なく父は家族を動物園に連れて行った」と直すことができる。

3つ目は、尊敬語の扱いである。

韓国語と日本語では尊敬語の使い方が大きく異なる。韓国語の場合、絶対尊敬語と言って誰の前でも尊敬すべき人のことは尊敬語を使って話す。たとえば、取引相手から「社長さんは今、会社にいらっしゃいますか」と聞かれたら、たとえ自社の社長でも「はい、いらっしゃいます」と答える。また「お父様、お元気ですか」と他人に尋ねられたら、自分の父でも「いいえ、うちの父さんは今入院されています」と言う。これは話す相手によって尊敬語が普通語や謙譲語に変わる、日本の相対尊敬語とは違う。

このような文章の翻訳は当然のことながら日本語にならって直されなければならない。

「わが社の社長がご挨拶をされたあと、貴社の代表

の方からご講演をいただきます」は「わが社の社長がご挨拶申し上げたあと、貴社の代表の方からご講演をいただきます」とならなければならない。

韓国語の翻訳で苦労するところは、ひとことで韓国語・朝鮮語に慣れてしまうと、その訳文が日本語として多少不自然なところがあっても、意味が十分通じるためそのまま見過ごしてしまうことが多いことだ。そのため、たえず純粋な日本語に多く触れることで、韓国語と日本語の表現の違いにいつも敏感になることが大切である。

たとえば、「八方美人」という四字熟語は、日本語では「だれに対しても如才なく振る舞う」ことを意味するが、韓国・朝鮮語では、「多芸多才である」ことを表現する言葉だ。このような違いを忘れてそのまま直訳したら、意味の通らないおかしい文章になってしまう。

以上のことを踏まえ、韓国語翻訳のあるべき品質レベルを考えると、やはり日本語らしい表現であることを前提とし、それぞれの翻訳分野で求められる基準をクリアすることだと思う。たとえば、科学技術・歴史・社会など非文芸分野の文章では、正確さとわかりやすさが生命になるが、小説や詩など文学作品では、多少原文の表現は変えてでも日本人にとって自然に受け入れられる感情表現にすることが求められる。たとえば、生活において韓国人は日本人より自分の感情をはっきりと表す。肉親が亡くなったときにも、大きな声と身振りで悲しみをいっぱいに表示するのが普通だ。一方、日本では感情を押さえたほうが美德とされ、人を感動させる。周りに他者がいるときはなおさらだ。したがって韓国の文芸作品を訳すとき、日本人が感情移入できるように必要に応じて多少押さえた表現にするほうが、文芸作品の趣旨、つまり読者に感動を与えるという本来の目的に合う。

機械翻訳について、私は政治・経済など非文芸分野の翻訳では利用しているが、文学作品の場合は、いまだ利用できるような翻訳ソフトに出会えずにいるのが現状である。

非文芸ものの翻訳では、機械翻訳でもかなりの正確さと日本語らしい表現が可能になっている。一からすべて自分で訳すより、機械翻訳によって一旦訳されたものを修正して仕上げるほうが、所要時間がかかり短くて済む。機械翻訳したあとの修正では、同単語の修正を文章全体を「選択」して「置き換え」で一挙に処理する機能など、各種テクニックを駆使すれば、さらに時間短縮は可能になる。

ただ表現の自然さ、感情のつながりを重視する文芸ものの場合は、一次翻訳を機械にしてもらいより、自ら一歩ずつ積み重ねていく方が現段階では効率的なようである。

また機械翻訳では、上記した韓国語翻訳での注意事項、つまり漢語を減らすこと、受動態や相対尊敬語への変換なども不十分で、これからの課題と言えるだろう。



機械翻訳シンポジウム「機械翻訳研究の課題と新方式の展望」参加報告

株式会社東芝 住田 一男

福岡大学理学部柴田勝征研究室が主催した掲記シンポジウムが2009年10月31日と11月1日との2日、福岡大学セミナーハウスにて開催された。報告者は、パネルディスカッションのパネリストとして参加した。本報告では、シンポジウムの概要について述べる。

シンポジウム概要

1日目には、日本大学・新田義彦先生による基調講演に引き続き、パネルディスカッション(パネリスト：奈良先端大学院大学・乾健太郎先生、福岡大学・柴田勝征先生、日本大学・新田義彦先生、新潟大学・宮崎正弘先生、筆者、司会：佐良木昌先生)が行われた。2日目には、福岡大学・首藤公昭先生ならびに桜美林大学・青山文啓先生の招待講演に引き続き、機械翻訳システムの紹介とデモが、岐阜大学・池田尚志先生、鳥取大学・徳久雅人先生、新潟大学・宮崎正弘先生、福岡大学・柴田勝征先生から行われた。参加者は40名弱と小規模なシンポジウムではあったが、活発な質疑応答が行われ、有意義なものであった。

基調講演(新田先生)

「機械翻訳の新方式の展望」という表題で、先生が20数年前に試作された翻訳システムの訳出と、現在の商用ソフトウェアとの訳出とを対比し、現在の商用ソフトが善戦しているもののさらなる改善が必要であると指摘された。また、現在の商用ソフトは下訳や参考訳、外国文献の通覧などには有効であるが、今後さらによいものとしていくには、翻訳実務に携わっておられる方の翻訳過程を地道に研究し、定式化するとともに、それら翻訳実務家の知見や要望の取り込みが必要であると主張された。

パネルディスカッション

今回のシンポジウムのテーマである「機械翻訳研究の課題と新方式の展望」というテーマで、各パネリストがプレゼンテーションを行った後、フロアからの質疑を交え議論を行った。

乾先生は、「次の10年?」という表題で機械翻訳の外側から見るという立場で、自然言語処理、統計翻訳におけるフレーズ解析の利用や Semantic Role Labeling などの意味処理の研究について技術動向を示された。これらの技術は単一の言語を対象としたものであるが、統計翻訳を含む今後の機械翻訳は、この進化した意味解析技術を取り込んで行くことで高精度化が図れるのではと主張された。

柴田先生は、「英和機械翻訳の困難性」という表題で、英単語の多品詞の問題、英文中の関係詞に起因する品詞解釈や構文の曖昧性の問題、英語熟語と日本語訳との間での語順や構造の不一致の問題など、事例に基づいて英和翻訳での課題を列挙されるとともに、統計処理や機械的な推論だけでは訳せず文脈の考慮が必要な事例を示された。

報告者は、「機械翻訳の現状と展望」という表題で、企業の研究者という立場から、実用システムとして製品化されている規則翻訳技術に基づく機械翻訳ソフトの適用領域とその課題について述べるとともに、新しい応用としての音声翻訳システムについて述べた。今後の技術の、従来の規則翻訳とともに、統計翻訳とをそれぞれ有効な特徴を使い分けてアプローチしていくことが重要と主張した。

宮崎先生は、「高品質な機械翻訳システムを目指して」という表題で、日本語文の書き換えに基づく日英翻訳について紹介された。英訳が困難な日本語文の書き換えを行うことにより、適切な英訳を行う日英翻訳試作システムの課題として、訳語選択や多

義性解消において、連想型の知識ベースの構築と利用をあげられた。

以上のポジション発表に引き続き、フロアからのコメントや質疑に基づく議論では、大規模なタグ付きコーパスや対訳コーパス構築の重要性和正確性についての問題点、統計翻訳と規則翻訳のそれぞれの得失など活発な議論があった。今後の翻訳精度の向上には意味解析の深耕を地道に進めていくことが必要であり、近年の統計翻訳中心のトレンドに批判的な主張があった一方、統計処理や強力な機械学習により単言語での意味解析についての近年の進化は積極的に活用すべきとの主張があった。

招待講演(首藤先生)

「処理単位としての言語表現—NLP への遥かなる道—」という表題で講演された。「赤の他人」「油を売る」「警鐘を鳴らす」「風前の灯」など、日本語の慣用句や決まり文句を継続的かつ精力的に収集するとともに、それら表現の構造を分析し(読み、構成する単語の境界、異表記、文法や意味的な機能、格構造など)、その構造情報を記述した知識ベース化を進められてきた。

現状の市販の機械翻訳では、このような慣用的な表現が入力された場合は直訳となる。あるいは、その表現が用例に登録されていたとしても構造として認識していないため、語の挿入などの変形に対応できない場合が多く、不適切な訳が出力される。慣用的な表現を複単語表現(Multiword Expression)として認識することにより、機械翻訳が適切な訳出が可能となる。

慣用的な表現の意味の記述は、統計処理では扱いがたい問題であり、これを継続的に収集・整備されてきた先生のご努力には頭が下がる思いである。今後ぜひ公開されることを期待したい。それにより、機械翻訳にとどまらず、仮名漢字変換、大語彙連続音声認識など様々な自然言語処理アプリに適用され、それらの精度向上につながるものと考えられる。

招待講演(青山先生)

「単語として形容詞が持つ情報」という表題で、IPAL 辞書の編纂を行われてきたご経験に基づき、述部が形容詞である場合の文について、形容詞が支配する語との統語的關係に関して、その関係を記述する上で問題のある多くの事例をご紹介された。IPAL は、動詞、形容詞、名詞についての統語的性格に焦点をあてて開発された辞書であり、これまで自然言語処理の多くの研究で活用されている。意味分析の深さを再認する機会となる内容であった。

翻訳システム紹介

①「日本語からのアジア諸言語への機械翻訳システム jaw」(池田先生)、②「意味パターン翻訳システム」(徳久先生)、③「日英機械翻訳システム Laurel」(宮崎先生)、④「US 式英日機械翻訳」(柴田生成)、の 4 つの試作システムについての説明がデモを交えて行われた。

①では、日本語から中国語を初めとした複数のアジア言語への翻訳に取り組まれている。②は、整備・蓄積した用例パターンによって日本語の大域的な構造を認識し、英語に翻訳することを特徴としている。③は、訳出が難しい日本語の表現を等価な表現に書き換えることにより安定した英訳を行うことを特徴としている。④は、英語教科書の基本的な文を中心に分析し、解析文法を構築されている。など、それぞれ特徴のあるシステムであった。

むすび

機械翻訳シンポジウム「機械翻訳研究の課題と新方式の展望」の参加報告を述べた。各セッションともフロアからの質問・コメントのため時間が延長になるほど、たいへん活発な議論が行われた。なお、本シンポジウムの各プレゼン資料については、<http://www1.rsp.fukuoka-u.ac.jp/#MT2009> (2010/1/24 確認)にアップされているので、詳細についてはそちらを参照していただきたい。

池原悟先生を偲んで

昨年末に鳥取大学の池原先生がお亡くなりになりました。まだまだご活躍・ご指導いただけたと思っておりましたのに、残念でなりません。ここに池原先生の最後のご寄稿を再掲し、池原先生のご功績を偲びたいと思います。

池原先生との思い出をお話しするのには、私は適任ではないとは思いますが、数年前にヨーロッパの会議で池原先生と同じホテルに宿泊しており、先生と奥様が街を散策に出られるところにエレベータホールでお会いし、お元気になられたのだなと後ろ姿をお見送りしたことが思い出されます。最後にお会いしたのは鳥取大学で開催された今年の言語処理学会の折で、長尾先生からの「お身体に気をつけて」とのお言葉をお伝えしたところ、長尾先生に気にかけていただけるとは光栄ですね、と喜んでおられました。

日本語語彙大系をはじめ、池原先生の数々の業績を私たちは十分には活かせていないように思います。機械翻訳の技術向上のためにも、先生のご研究を学び、発展させていきたいものです。

アジア太平洋機械翻訳協会会長 井佐原 均

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があるが、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）の意味を結合価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。

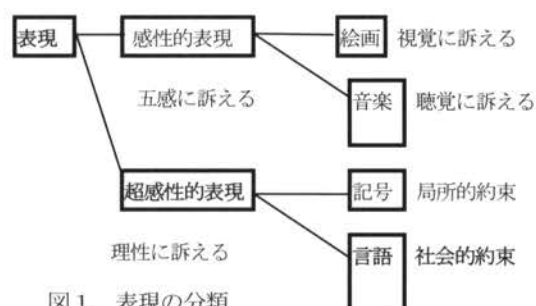


図1. 表現の分類

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峙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でに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。

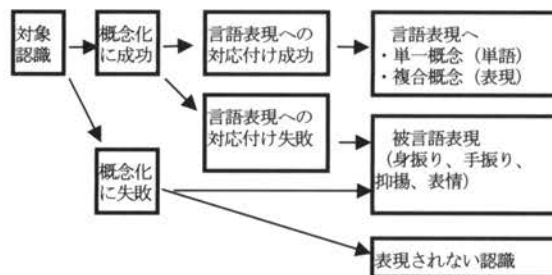


図2. 認識・概念・表現の関係

(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、

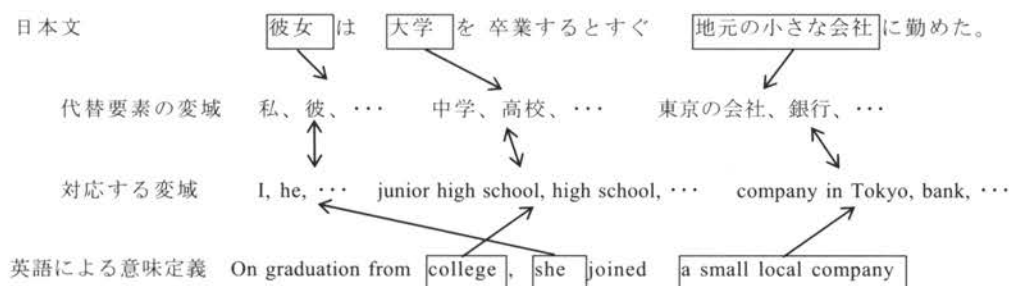


図3. 線形要素の例

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＞線形要素と非線形表現の同時性

「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」でもよいから、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てばよい。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

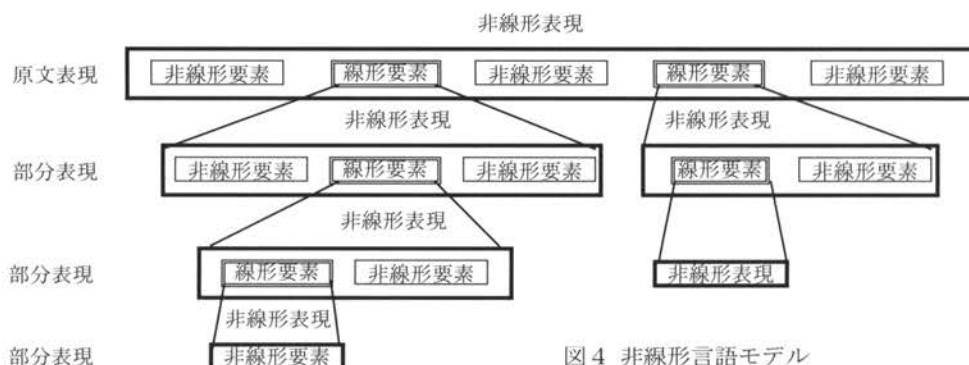


図4 非線形言語モデル

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。

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、日本文が12.9語/文(最大63個)、英文が、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。

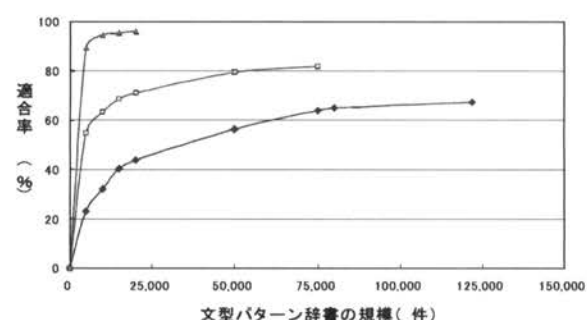


図5. 文型パターン適合率の飽和特性

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た後で結合しても、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、意味を掘くく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。

日本語からアジア諸言語や手話への機械翻訳 — 古典的方式での構築 —

岐阜大学工学部 池田 尚志
(2010 年 4 月から山梨英和大学)

1. はじめに

機械翻訳に関しては研究面では統計翻訳方式が現在の主流である。統計翻訳に触れずして機械翻訳を語ると「なぜ統計翻訳をやらないのか?」と言われる。かつて第5世代コンピュータという大きなプロジェクトがあって、論理型のプログラミング言語prologが流行した。「なぜprologを使わないのか?」と言われた時代があったのを何となく思い起こす。prologはとても面白い興味ある言語であり、筆者も勉強し使った。しかし誰もかれもが何もかもをもprologでという風潮にはいささかうんざりした。情報爆発の時代にあって統計翻訳は面白い手法であるし、探求すべき方向の一つである。しかしそれだけがすべてか。そうでない方法だって面白いしもっと探究してみてもいいではないか。

筆者には、意味理解の視点を欠いた統計的な機械処理だけで、人工知能的に翻訳規則が形成できるとは思えない。例えば“旅行の際のホテルの予約”、“観光案内文”、“天気予報文”といった、対象とする文の形と語彙がおおむね限定できるような単純な表現の領域では統計的手法が可能であり有効かもしれない。しかし、そうでない一般的な表現の領域では難しいと考えている。なによりも統計的な手法を実施するためには、“適切な”対訳データが大量に準備されていなければならない。しかし文の数、その対応の数は、文字どおり組み合わせ論的な爆発の領域である。そのような大量の適切な対訳データが現実存在しているのだろうか?

筆者は大学に勤めていて留学生達との出会いもあって、日本語からベトナム語やシンハラ語、ミャンマー語、中国語等への機械翻訳システムの構築を試みる機会があった。そこでは上記のような考えもあり、また人間の頭で言語規則(翻訳規則)を考えるのも案外楽しいものだ、といったような思いで統計的な手法ではなく古典的なトランスファー方式&パターン翻訳規則の手法でのシステム(jaw)を試作した。その詳しい経緯は[1]に書いたが、本稿ではjawの手法の概要と手話への翻訳を中心に紹介する。{以下で図表等 [1]からの抜粋を

含む。}

2. Jawの翻訳方式

図 I と図 II にjawによる処理の概要を示す。jawでは相手言語のテキスト(翻訳文)を出力する前に“表現構造”(図 I の★)という段階を置いて、入力文から“表現構造”への変換、“表現構造”から相手言語の文の生成(線状化)というステップで翻訳する。この“表現構造”を介した翻訳手法がjawの大きな特徴である。

表現構造は、相手言語での言語表現に必要な言語表現上の部品と情報が書き込んであるデータ構造であり、相手言語の言語的構造を表現するデータ構造である(図 II -③)。“表現構造”から先は相手言語の世界である。たとえば図 II -③にある“接続構造”は、中国語の接続表現のための表現構造として設計したものである。中国語では一般に二節を接続する場合、接続を表現する語が二つあり得て、それを配置する場所も節頭/節末/動詞の直前などいくつかあり得るので、そういった情報を書き込めるようにしている。

表現構造にはプログラミング言語C++のオブジェクトをそのまま利用しており、jawの翻訳規則はC++のオブジェクトを作り出すプログラムである。所定の記述方式で翻訳規則を書くと、対応する表現構造を作り出すC++のプログラムが自動的に出来上がる仕組みにしてある。また、表現構造から相手言語の翻訳文を生成する線条化の処理はC++に備わっているクラスメソッドの仕組みを利用したプログラムで行っている。線条化は、“表現構造”に書き込まれた言語表現の部品に必要な変形(複数形、格変化、時制変化など)を加えたり必要な補充(冠詞、前置詞など)をした上で、適切な語順(SVO, SOV, 冠詞+形容詞+名詞, 関係節表現, 等々)に並べて文を作り出す処理であり、相手言語の言語構造をそのまま反映したプログラムである。当然言語毎に備えておくプログラムである。このようにプログラミング言語C++のパラダイムを利用した表現構造を介して翻訳を行う手法がjawの大きな特徴である。jawによる翻訳処理の手順は次のようになる。

- ① 日本語テキストを解析して構文構造を得る(図Ⅰ-a, 図Ⅱ-①)。
 - ② 表現パターン翻訳規則の日本語パターンと照合して対応する翻訳規則を得る(図Ⅰ-b)。この処理で日本語テキストの構文構造に対応してそれをちょうど覆うような翻訳規則の木構造ができる(図Ⅱ-②)。
 - ③ 複数の翻訳規則が対応可能な場合には、覆うパターンの数や種類による条件、より個別的な規則を優先するといった条件など、最適解を選択するメカニズムを備えている。
 - ④ 翻訳規則を実行して相手言語の表現構造を作成する(図Ⅰ-c, 図Ⅱ-③)。この段階で、“機能語(助詞、助動詞など)に関する翻訳規則”が実行され、必要な情報が表現構造中に書き込まれる。
 - ⑤ 線状化のプログラムを呼び出して、翻訳文を生成する。
- 以上のような手法で、各国からの留学生たちとも共同して、日本語からベトナム語、シンハラ語、中国語、ミャンマー語等への機械翻訳システムの構築を試みた。この研究に関連して5名の留学生が論文を書き博士の学位を取得した[2][3][4][5][6]。

3. “日本の手話”という言語への翻訳

手話(Sign Language)は視覚言語であり音声言語とは異なるが、同じく“言語”である。“言語”であるならば“言語処理”の対象であっていいし、“機械翻訳”の対象であっていい。IT技術による手話通訳支援は待望される技術であるし、手話に関する言語処理はそのための基礎研究にもなるはずだ。このような思いで、中国語やベトナム語などと並んで日本語から“日本の手話”への機械翻訳を試みた。

手話を機械翻訳や言語処理の対象とすると、まず“手話にはそれを書き留めたテキストというものがない”という大きな問題にぶつかる。つまり手話は今のところ書記言語ではない。これではこれまでのような“言語処理”や“機械翻訳”はできない。

そこでまず手話を書き留める方法を考えなければならない。筆者等は“日本語援用手話表記法(jjs-notation)”と名付けた手話の表記法を考案した[7](図Ⅲ)。jjs-notationにおける手話単語名は手話単語を識別するための識別子であるが、日本語と手話の語彙は一対一に対応しているわけではない。たとえば図Ⅳの1は「二人で行

く」に対応する手話単語であるが、『二人で』に対応する表現は手話では語彙として分化していない。図Ⅳの2は「あなたが私を見る」に対応する手話単語であるが、『あなた』、『私』は語彙として分化していない。便宜上その手話表現の意味に近い日本語の語句を用いることにしたので、「日本語援用手話表記法」といういい方をしている。

機械翻訳システムjaw/SLは、日本語のテキストからjjs-notationによる手話のテキストへの機械翻訳を行う。翻訳の手法は中国語やベトナム語などの音声言語を相手言語とする場合と全く同じである。ただし日本語と手話の構造には隔たりの大きい場合も多く、通常の音声言語を対象とする場合よりも機械翻訳の困難さは大きい。たとえば、「ビールを飲まされる」を、手話では「コップを渡されていやいや飲む様子」として表現したり、「将来飼ってみたいのは大きな犬です」を分裂文「将来ほしいのは何かというと、大きな犬が欲しいです」(jjs-notationでは『<clef>将来 好き 何, 犬 大きい 好き』となる)に言い換えて翻訳したり、といったことがある。我々の研究では適切に言い換えた文を入力として与えたが、さらに研究を進めていく上では、省略や言い換えといった意味理解の領域を深く追求していくのか、あるいは場面をうまく制限して適用するなど工学的な対応を考えていくのか、判断が必要である。

4. おわりに

機械翻訳エンジンjawとそれを用いた日本語からベトナム語、シンハラ語、中国語等への翻訳システムについて述べた。さらに日本の手話に関して、手話を書き留める表記法の提案を含めて日本語から日本の手話への翻訳システムについて述べた。jawは古典的なトランスファー方式でパターン翻訳規則の手法によるものである。試作したシステムはパイロットシステムであって実用レベルのサイズの翻訳規則を備えるまでにはいたっていない。しかし、ベトナム語やシンハラ語などいわゆるマイナーな言語に関わることができたのも、手話というこれまで機械翻訳の分野ではほとんど対象として考慮されていなかった言語に足を踏み込めたのも、jawの翻訳の方式によるものだと考えている。統計翻訳にのみ注目していたらこのようには進まなかっただろうと。

1. 日本語からアジア諸言語への機械翻訳システムの構築 奮闘記 -- 中国語へ、ベトナム語へ、シンハラ語へ、日本の手話へ --, 池田尚志, 日本語学, 2009年10月号, vol.28-12, pp.62~71, 2009
2. 日中機械翻訳システムに関する研究 --変換・生成の方式およびテンス・モダリティの翻訳処理, 謝 軍, 岐阜大学工学研究科博士論文, 2003
3. 日中機械翻訳に関する研究 --取り立て表現、否定表現の翻訳規則を中心に--, ト朝暉, 岐阜大学工学研究科博士論文, 2004
4. 日本語ーベトナム語機械翻訳に関する研究 --連体修飾構造を中心に--, Nguyen My Chau, 岐阜大学工学研究科博士論文, 2005
5. A Study on Machine Translation System from Japanese to Sinhalese on MT Engine *Jaw*, Samantha Thelijagoda, 岐阜大学工学研究科博士論文, 2006
6. 日中機械翻訳システムに関する研究 --存在表現及び軽動詞構造に関する翻訳処理を中心に, 王 軼謳, 岐阜大学工学研究科博士論文, 2008
7. 日本語テキストから手話テキストへの機械翻訳に関する研究, 松本 忠博, 岐阜大学工学研究科博士論文, 2008

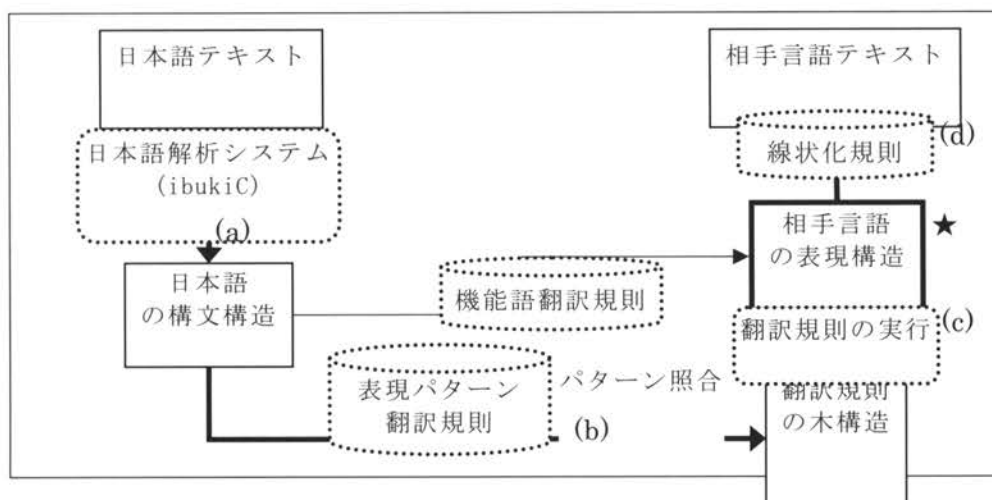


図 I . jaw による翻訳処理の概要

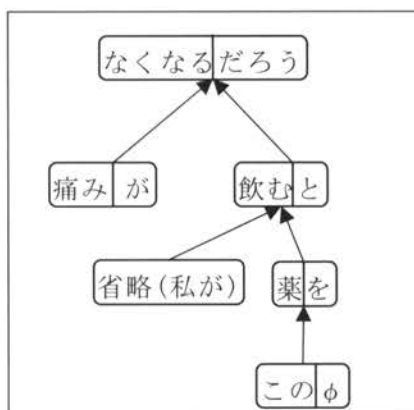


図 II -① 日本語の構文構造の例

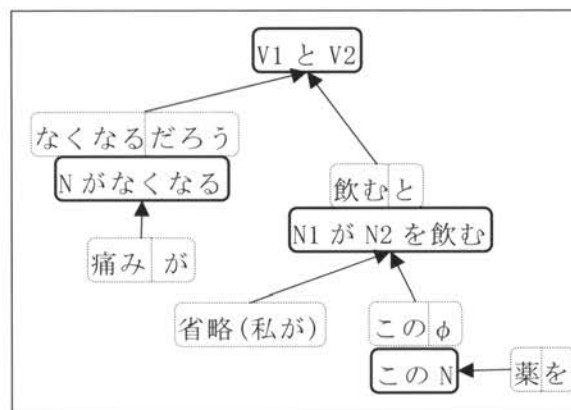
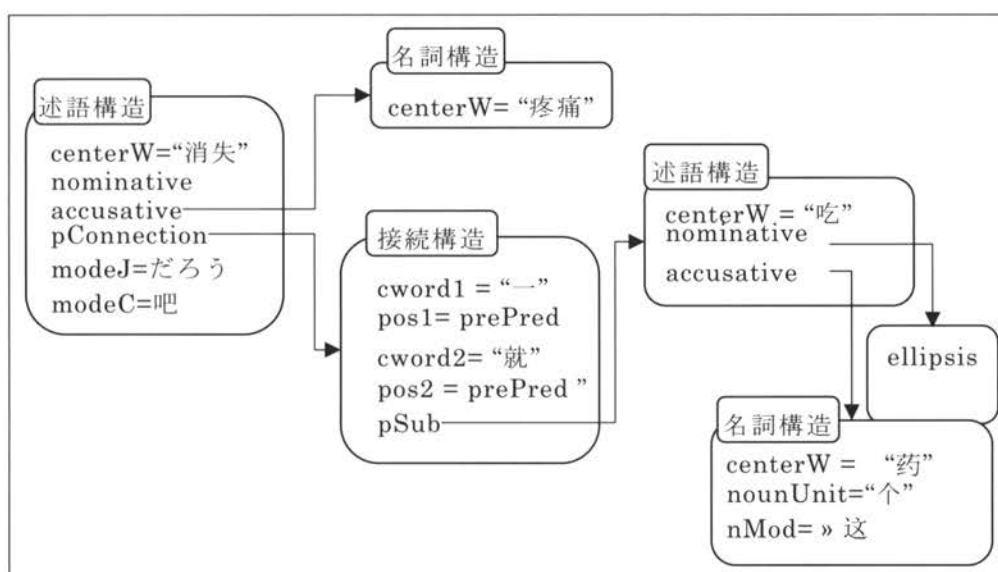


図 II -② 翻訳規則の木構造の例



図Ⅱ-③. 中国語の表現構造の例
「この薬を飲むと痛みがなくなるだろう」
「一吃这个药, 疼痛就消失吧」

jjs-notation の形式 ; 手話単語名[手形] (空間 ; 修飾)	
jjs-notation	日本語での表現
行く	; (一人で) 行く
行く [2]	; 二人で行く
見る (1→2;)	; 私があなたを見る
見る (2→1;)	; あなたが私を見る
過去 (;とても)	; ずっと前
あなた 母(x;) 話す (x→1;)	; あなたのお母さんから聞いた
{<t> 妻 仕事} , 例-仕事。	; 妻の仕事はパートです



図Ⅲ 日本語援用手話表記法 (jjs-notation)
({<t> x}は x が topic であることを表す)

図Ⅳ. 手話単語と日本語

A comparison of the two major English-Japanese free translation sites, Excite Translation and Yahoo! Translation

Shigeko Sugiura

1. Introduction

In Sugiura (2009)¹, I compared two major English-Japanese free translation sites, Excite Translation (<http://www.excite.co.jp/world/>) and Yahoo! Translation (<http://honyaku.yahoo.co.jp/>). The reason I chose these two sites was that according to a questionnaire conducted on 330 company workers of both sexes (age 20's to 60's) on June 29, 2008 by Internet Com and JR Tokai Express Research (<http://japan.internet.com/research/20080702/1.html>), these two sites were by far the two most popular free translation sites. The users who answered this questionnaire overwhelmingly used the sites to translate between English and Japanese. I compared the two sites by translating two different types of text each for English-to-Japanese translation and Japanese-to-English translation. The types of text were (a) newspaper articles with a human interest factor², and (b) encyclopedia entries³. The results and further investigation showed that errors in translation could be attributed to the following factors.

- A. Differences in punctuation used in written English and Japanese
- B. Differences in English and Japanese as language systems
- C. Ambiguity and complexity of sentence structure

Some of the difficulties caused by these factors were shared by both sites, while others were treated differently in the two sites. Difficulties due to A will be discussed in section 2, difficulties due to B in section 3, and difficulties due to C in section 4.

2. Differences in punctuation used in written English and Japanese

To begin with the basics, in English words are separated by spaces but not in Japanese. Also, in English, the beginning of a sentence is marked with a capital letter, and the end of a sentence is marked with a period. In Japanese, there is no way to mark the beginning of a sentence. The end of a sentence is usually marked with a *kuten* (句点, literally 'phrase point'), but the use of the *kuten* is less rigidly required than the English period. These facts do not cause difficulties by themselves,

¹ 'Muryoo hon-yaku saito-no hon-yaku nooryoku-no genzitu (The actual translating ability of free translation sites)', *Nihongogaku* October 2009.

² An article on car repair in Ghana from the Los Angeles Times website (<http://articles.latimes.com/2009/jun/16/world/fg-ghana-car-yard16>) and an article on educational efforts from the Yomiuri newspaper website (<http://www.yomiuri.co.jp/kyoiku/renai/20090725-OYT8T00259.htm>).

³ The article on Japan from Encyclopedia Britannica (<http://www.britannica.com/EBchecked/topic/300531/Japan>) and the article on statistics from Yahoo! Japan Encyclopedia (<http://100.yahoo.co.jp/detail/%E7%B5%B1%E8%A8%88%E5%AD%A6/>).

but may combine with other aspects of the two languages and lead to difficulties.

2.1. Quotation markers

Both languages have punctuation to express quotation. English uses single quotation marks (‘ ’) and double quotation marks (“ ”) to mark quotation of whole utterances and marked usage (literary styles may differ regarding usage). Japanese uses *kagi-kakko* (「 」, literally ‘key brackets’) to mark quotation of whole utterances and marked usage. There are, however, several differences in the use of quotation markers. One difference is that in Japanese, the use of the closing *kagi-kakko* is often regarded as enough to signal the end of a sentence as well as the end of an utterance, and thus no *kuten* appears. This causes problems, since the absence of a *kuten* makes it difficult to judge whether a pair of *kagi-kakko* marks a quoted utterance or marked usage of an expression. Both sites fail to recognize this double function of the closing *kagi-kakko*, and accordingly did not analyze the quotation and the following sentence as two different sentences in Japanese-to-English translation. This happened even when the quotation ended a paragraph. Such errors are due to the combined effect of the optionality of the *kuten* in the presence of a closing *kagi-kakko* and absence of marking for the beginning of a sentence in Japanese. However, the two sites treat the Japanese *kagi-kakko* differently. Excite Translation treats the material between an opening *kagi-kakko* and the closest following closing *kagi-kakko* as a unit (although, curiously, when the material was one sentence the Japanese *kagi-kakko* were rendered as double quotation marks in the English translation, but when the material was more than one sentence the *kagi-kakko* were kept in the English translation). Yahoo! Translation, in contrast, do not admit quoted utterances of more than one sentence long. Therefore, when a sentence containing an opening *kagi-kakko* ended, a closing quotation mark was inserted even when there was no closing *kagi-kakko*, and the corresponding closing *kagi-kakko* was ignored in the English translation. Yahoo! Translation also made errors not made by Excite Translation in that when there was no overt subject to the Japanese sentence with an opening *kagi-kakko* at its head, an overt subject was amended in the English translation but this subject appeared (wrongly) outside of the opening quotation mark. (See section 3.2.1 for null pronouns.)

2.2. Punctuation in English which pose difficulties

The colon (:) and semi-colon (;) are used in English but not in Japanese. In English-to-Japanese translation, Yahoo! Translation simply reproduced them. Excite Translation renders the semi-colons as *kuten*, and treats colons differently according to whether the preceding sentence is a complete sentence, as in (1a), or an incomplete sentence, as in (1b). In the former case, the colon is rendered as colon, and in the latter case, as *kuten* (the incomplete English sentence, moreover, is translated as a complete sentence, with the expression *ika-no toori-desu*, meaning ‘as follows’, amended).

- (1) a. Sometimes Alex Foley sits at home sifting through the pieces of a life-size puzzle in his mind: how to put that car back together.
 b. Examples of familiar sequences are: one, two, and three; a, b, and c; and first, second, and third.

The majority of Japanese readers would not really understand what the colons and the semi-colons express if they appeared in the Japanese translation.

2.3. Punctuation in Japanese which pose difficulties

The *nakaguro* (・), a black dot in the middle of the line, is a Japanese punctuation meaning ‘and’, so the two expressions divided by it are in a coordinating relation⁴. Both sites gave very unsatisfactory results with sentences containing the *nakaguro*. Yahoo! Translation did better, since it reproduced the *nakaguro* as slash (/) when the two expressions coordinated were nouns (Excite Translation ignored the *nakaguro* in this case). When the two expressions were verbs, both sites translated the first verb as gerund, and the second verb as an ordinary tensed verb, i.e. treated the first as being syntactically subordinate to the second.

3. Differences in English and Japanese as language systems

Differences in two language systems lead, of course, to difficulties in translation. The most obvious difference is the difference in the meaning of words and word forms. In section 3.1, a special case of such difference will be considered, where one form in one language corresponds to more than one form in the other language. In section 3.2, difficulties due to the difference in which element may be null will be considered. In Section 3.3, difficulties due to the difference in the marking of grammatical categories and topichood will be discussed.

3.1. Differences in the meaning of word forms

3.1.1 English verb past participle

With some verbs, the English verb past participle is ambiguous. On one reading it is part of a passive construction, and on the other reading the verb past participle is adjectival. Thus in (2a) *broken* is part of a passive construction, and the sentence expresses that the breaking of the window happened in the past, and in (2b) *broken* is adjectival, and the sentence expresses that the state of the window was ‘broken’ in the past.

- (2) a. (The video shows that) the window was broken at three o'clock.
 b. (They inspected the room and noted that) the window was broken.

The two meanings should be translated into Japanese as follows.

- (3) a. Mado-wa san-zi-ni warareta.

⁴ The *nakaguro* is also used to mark the boundary between first and last names of foreigners. This is necessary because Japanese does not use spacing between words. Excite Translation used the *nakaguro* in this function in English-to-Japanese translation, but Yahoo! Translation did not.

window-TopicMarker three-time-at break(transitive)-Passive-Past

b. Mado-wa wareteita.

window-TopicMarker break(intransitive)-*teiru*-Past

Both translation sites translated *the window was broken* as adjectival. With other verbs, generally Yahoo! Translation gave passive sentence translations while Excite Translation was more versatile: in some cases the verb past participle was translated as a passive construction and in others as adjectival (also, in some cases, as an active construction). However, the choice seems to depend on the verb used, and so the result may or may not be satisfactory as translation.

3.1.2 Tenseless predicate forms in Japanese

In English, when sentences are coordinated, the predicates are all tensed, but in Japanese, only the last predicate is tensed, and the other predicates appear in tenseless forms, either the *renyoo*-form or the *te*-form. One type of difficulty which showed up very clearly was due to the combined effect of this characteristic of Japanese and the fact that quite a few number of predicates in such tenseless forms is part of postpositional and adverbial expressions. Examples of adverbial expressions which include the *renyoo*-form are *NP-o hajime* ‘starting with NP’, *NP-ni tuki* ‘due to NP’, *NP-ni kakawarinaku* ‘regardless of NP’⁵, and examples which include the *te*-form are *NP-o hukumete* ‘including NP’⁶, *NP-ni kurabete* ‘compared with NP’⁷. It is formally not possible to distinguish between these tenseless forms which are coordinated with tensed predicates (and should be translated as tensed in English), and those which are part of postpositional and adverbial expressions (and should be translated as such expressions in English), and this leads to errors in Japanese-to-English translation.

3.1.3 Japanese *teiru*

Maybe it is not appropriate to treat *teiru* in, for example, *wareteiru* (see 3b), as a unit, since *te* is part of the *te*-form of the verb *wareru*, and *iru* is originally a verb, but it is convenient to do so. *Teiru* may express progressive aspect (i.e. that the action is in progress) or perfect (i.e. that the action has been completed). Thus *Taroo-wa kimono-o kiteiru* may mean either ‘Taro’s putting on of the kimono is in progress’ or ‘Taro’s putting on of the kimono has been completed, and as a consequence, Taro is wearing the kimono’. In practice, with predicates denoting actions such as *hasiru* ‘run’, the meaning is usually progressive, with predicates denoting changes such as *otiru* ‘fall’, the meaning is usually perfect, while with some predicates, such as *kiru* ‘put on’, the possibilities are more evenly split. In principle, though, ambiguity is possible with the majority of non-stative predicates. *Teiru* also poses another kind of problem, in that when habitual meaning is

⁵ See (8).

⁶ See (8).

⁷ Some expressions allow the predicate to be either in the *renyoo*-form or the *te*-form, while some expressions appear only with one predicate form. Of the five expressions cited, the first three only allow the predicate in the *renyoo*-form but the latter two allow both the *renyoo*-form and the *te*-form.

expressed, *teiru* may or may not appear, without difference in cognitive meaning. All this means that the following four combinations should in the majority of cases be translated as given in the examples, although other possibilities exist.

- (4) a. Predicate in present tense e.g. *hasiru* ‘runs’, *otiru* ‘falls’
- b. Predicate+*teiru* in present tense e.g. *hasitteiru* ‘runs, is running’, *otiteiru* ‘has fallen’
- c. Predicate in past tense e.g. *hasitta* ‘ran’, *otita* ‘fell’
- d. Predicate+*teiru* in past tense e.g. *hasitteita* ‘ran, was running’, *otiteita* ‘had fallen’

Yahoo! Translation translates (4a) and (4c), (4b) and (4d) identically for all predicates, while Excite Translation seems to divide predicates into three classes, action predicates, change predicates, and others. With action predicates, (4a)(4c) are translated as simple predicate, i.e. ‘run(s)’, ‘ran’, (4b)(4d) are translated as progressive, i.e. ‘is running’, ‘was running’. With change predicates, (4a)(4c) are translated as simple predicate, i.e. ‘fall(s)’, ‘fell’, (4b)(4d) are translated as perfect, i.e. ‘has dropped’, ‘had dropped’. With others, (4a)(4c) are translated identically as present tense, and (4b)(4d) are translated identically as past tense. Moreover, there is also a certain problem with the *teiru* perfect, in that it is often more suitable to translate it using English expressions of state rather than perfect. For example, it is often suitable to translate *otiteita* as ‘lay on the ground’ rather than ‘had fallen’.

3.2 Null elements

3.2.1 Japanese null pronouns

Japanese has null pronouns, or, in other words, what would be expressed in English as pronouns are not expressed at all. This means that in Japanese-to-English translation, appropriate pronouns need to be amended. There is a certain case, however, in which the subject does not appear in both languages, which is a tenseless sentence with an arbitrary subject, such as *to live in Japan* and *living in Japan* in English. Since the subject may be null in Japanese, and the tenseless form of the verb is identical to the present tense form of the verb, a tensed sentence with a null subject (which should be translated with an amended subject) is formally identical to a tenseless sentence with a null arbitrary subject, which should be translated as a clause with a *to*-infinitive or gerund.

In translating such Japanese sentences into English, the strategy taken by Yahoo! Translation is basically to amend the first person singular pronoun *I*. Excite Translation basically amends *it* unless the sentence without an overt subject is a complement phrase headed by *no* or *koto*, in which case it is translated as a clause with a gerund or *to*-infinitive. This means that both sites give unsatisfactory results to a certain degree.

3.2.2 English null predicates

In English, part of a predicate may be null (or omitted) when it is identical to a preceding

predicate. In Japanese, such predicates must not be omitted. This means that in English-to-Japanese translation, the appropriate verb form must be supplied in the Japanese translation. As an example, Yahoo! Translation gave (6a) and Excite Translation gave (6b) as translation for (6).

(5) Even if the passengers couldn't be saved, the car always can be.

(6) a. Tatoe jookyaku-ga sukuwareru koto-ga dekinakatta-to sitemo, kuruma-ga tune-ni aru-koto-ga dekimasu. (Yahoo!)

b. Jookyaku-o sukuu-koto-ga dekinai demo, kuruma-wa itumo sukuu-koto-ga dekimasu. (Excite)

Excite Translation amended the verb *sukuu* 'save' correctly in the second predicate⁸, but Yahoo! Translation did not, translating *be* as the existential verb *aru*, as in (6a). However, this happens because there are two auxiliary verbs in the main clause of (5). When there is only one auxiliary verb, as in (5'), Yahoo! Translation amends *soo-suru* 'do so' as in (6'), which is not perfectly satisfactory but is an improvement over (6a).

(5') Even if the passengers couldn't be saved, the car always can.

(6') Tatoe jookyaku-ga sukuwareru koto-ga dekinakatta-to sitemo, kuruma-wa tune-ni soo suru-koto-ga dekimasu. (Yahoo!)

3.3 Definiteness, number, topichood

In English, noun phrases are expressed as either definite or indefinite, but in Japanese no such expression exists. Similarly, in English, countable nouns are expressed as either singular or plural, but no such expression exists in Japanese. Thus, the Japanese sentence 'Watasi-wa ringo-o katta', which in most cases should be translated as 'I bought some apples,' were translated as 'I bought an apple' by Yahoo!, and as 'I bought the apple' by Excite. On the other hand, Japanese expresses the topichood of an NP with the topic marker *wa*, but English does not. In translating two Japanese sentences differing in whether the subject is marked with *ga*, the subject marker, or *wa*, the topic marker, both sites gave the same English sentence as translation. However, in translating two Japanese sentences differing in whether the object is marked with *o*, the object marker, or *wa*, the two sites differed: Yahoo!, which translated the σ -marked object as indefinite, translated the *wa*-marked object as definite. Excite translated both the σ -marked object and *wa*-marked object as definite. Since topics are always definite, objects which are topics should be translated as definite, but σ -marked objects can be definite or indefinite.

4. Ambiguity and complexity of sentence structure

Sentences can be ambiguous, but this can be due to ambiguity of words used in the sentence and ambiguity due to possibilities of sentence structure. Ambiguity of words are rather easily

⁸ This strategy backfires sometimes, amending predicates in unnecessary cases.

detected (and maybe corrected), but the other type is more difficult to cope with. To give one example, the adverb *together* in (7) may be interpreted either as modifying the action of the subject (we did something together), or as modifying the resulting state of the object (the car became whole). In Japanese, the former would be translated as *issyo-ni* ‘together’, and the latter as, maybe, *moto-doori-ni* ‘as before’. In this case the ambiguity of the English sentence can only be resolved by referring to the context.

(7) We put the car back together

Regarding complexity, take the translation of the Japanese sentence (8), which is based on a sentence from the article on statistics. (9a) is my translation, (9b) the translation given by Yahoo!, and (9c) that given by Excite.

(8) Tookeigaku-wa, sizen-gensyoo-to syakai-gensyoo-no betu-ni kakawari naku, sarani tyuusyooteiki-na suuti-no syuudan-o-mo hukumete, syuudan-kara tyuusyutu-sareta suuryoo-teki zyohoo-tositeno tookei-suuzi-o bunseki-suru tookeiteki-hoohoo-o kenkyuu-suru.

(9) a. Statistics studies statistical methods for analyzing statistical figures seen as quantitative information extracted from groups, whether of natural or social phenomena, and also including groups of abstract values.

b. The statistics can include an abstract numerical group regardless of the distinction of a natural phenomenon and the social phenomenon and study a method of statistics to analyze statistics number as the information of amount extracted from the group into.(Yahoo!)

c. Statistics researches a statistical method of analyzing the statistical figure as quantitative information to which the natural phenomenon and the social phenomenon were extracted by the group including the group of a more abstract numerical value without separately relations.(Excite)

Semantically, the single underlined part modifies the dotted underlined part, and the double underlined part, while modifying the dotted underlined part, takes second place to the single underlined part. If these relations are not clearly expressed, the sentence would not be comprehended. As can be seen, Yahoo! Translation misinterpreted *hukumete* ‘including’ as tensed, but aside from that, it fails to present the single underlined part as modifying the dotted underlined part. Excite failed to interpret the other postpositional expression *kakawarinaku* ‘regardless of’ and placed it apart from the noun phrase connected with it, although it successfully translates the double underlined part as modifying the dotted underlined part. This example illustrates that complex sentence structure leads to grave errors in translation.

I will give another example to illustrate a different kind of complexity in structure, complexity in informational structure. In the Japanese sentence (10), also based on the same sentence, the single underlined part states the background of the rest of the sentence, the double underlined part states the time point at which the rest of the sentence is relevant, and the dotted underlined part modifies the main predicate of the sentence, as seen in (11a), my translation. Thus these

three parts should appear in this sequence in both English and Japanese. However, English and Japanese have different constituent order in many constructions, and different ordering may be desirable. In other words, syntactically speaking English and Japanese have different constituent order, but from an informational viewpoint, the same constituent order might be preferable. The translations (11b) and (11c) illustrate the conceptual difficulty of sentences with inappropriate information structure.

(10) Suuzi-deeta-o bunseki-suru tookei-teki hoo-hoo-no kyuusoku-na sinpo-to ooyoo-ryooiki-no kakudai-nitomonatte, genzai-dewa ippan-ni, tookeigaku-wa, syuudan-kara chuusyutu-sareta suuryoo-teki zyoo-hoo-tosite no tookei-suuzi-o bunseki-suru tookei-teki hoo-hoo-o kenkyuu-suru gakumon-to kangaerareteiru.

(11)a. With the rapid progress of statistical methods for analyzing numerical data and the expansion of domains of application, at present statistics is generally considered as the study of statistical methods for analyzing statistical figures seen as numerical information extracted from groups.

b. Generally, with rapid progress and the expansion of the applied domain of the statistical method to analyze number data into, as for the statistics, it is thought with study to study a method of statistics to analyze statistics number as the information of amount extracted from the group into by the present.(Yahoo!)

c. In general, statistics is thought to be study that researches a statistical method of analyzing the statistical figure as quantitative information extracted by the group now as a statistical method of analyzing the figure data advances rapidly and applied area expands.(Excite)

Lastly, I would like to point out an interesting difference between the two sites in translating from Japanese to English. When unable to analyze sentence structure, Excite Translation gives isolated English words separated by two dots. In contrast, Yahoo! Translation gives out translations in all cases, and gives out untranslatable pieces as they appear in the original, that is, as Japanese.

5. Conclusion

In this article, I compared two free translation sites regarding English-to-Japanese and Japanese-to-English translation. It was shown that these two sites differ in several ways, but they also share many difficulties. I am unable to judge whether and when such difficulties can be technically overcome, but in the meantime I recommend users to employ their linguistic knowledge to obtain better results, assuming that users translate to or from their native language (or at least a language they can confidently handle).

In translating from your native language,

- a. avoid the more uncommon punctuation marks and make sure the appropriate punctuation marks are there. In Japanese-to-English, make sure there is a *kuten* at the end of every

sentence⁹.

- b. avoid complex sentence structures. Divide a long complex sentence into short simple sentences.
- c. avoid expressions which could lead to trouble. For example, in Japanese-to-English translation, avoid adverbial and postpositional phrases including a verb form, and avoid the use of null pronouns. In English-to-Japanese translation, avoid null predicates, use active sentences over passive sentences, and use numerals (*one, many*) and determiners (*that, those*) with nouns instead of singular and plural forms of nouns, the indefinite article *a/an* and the definite article *the*.
- d. avoid all colloquial and dialectal expressions. In Japanese-to-English translation, render all proper nouns in *katakana*.

In translating to your native language,

- a. employ more than one translation site and use the translation that makes the most sense. Also try feeding shorter pieces of the original text.
- b. if a certain word seems out of context, see if this is not due to an ambiguous word for which the wrong meaning was selected.
- c. if the translation does not make sense, try assuming that it is due to one of the possible pitfalls discussed here. To give an example, if in Japanese-to-English translation the verb form seems out of place, try changing a progressive form (*be + -ing*) into a perfect form (*have + past participle*) and vice versa.

⁹ A *kuten* following a closing *kagi-kakko* gives better results than the opposite.

(連載第三回)

言葉と人材 ― グローバリゼーションの再考と日本

東京大学工学部 特任研究員

九門 崇

グローバリゼーションと知識経済化が進む21世紀の流れは、今後ますます“多様化する世界（ワールド・オブ・ダイバーシティ）”に向かっていくと考えられる。グローバル化とインターネットにより、「モノ・金・情報」の流れがよりスピードを増し、世界のフラット化が進んでいけばいくほど、より企業や個人も競争力によって線引きがされやすくなり、様々な格差も生じる。これまでの国家という概念だけで図れない、様々な階層や社会が重なり合っている世界になっていく。本稿では、その多様性を大きく、1)「新興国と先進国」、2)「富裕層と貧困層」、3)「グローバル都市と地方都市」の3つのカテゴリーに分類し、日本に求められるグローバル感知力について論じる。

1) 多様化する世界

第1のカテゴリーは、欧米などの「先進国」と、成長著しい中国、ベトナム、インドなどの「新興国」という分類である。先進国については、今後の大幅な経済成長や市場拡大は期待できない傾向にある。これに対して、日本企業にとって重要性を増しているのが新興国でのビジネスで、中国以外の国にもさらに注目していく必要がある。国際協力銀行が実施している「わが国製造業企業の海外事業展開に関する調査」の中で、中期的（今後3年程度）有望事業展開先について03年度調査と比較すると、中国は03年度から09年度まで第1位を維持し、インドは第5位から第2位に、ベトナムも第4位から第3位に上昇した。さらに、回答企業の得票率で見ると中国が93%から74%に減少する一方、インドが14%から58%に、ベトナムも18%から31%に増加している。

第2のカテゴリーは、富裕層と貧困層である。富裕層について米ウォール・ストリートジャーナル紙の記者であるロバート・フランク氏は、米国のニューリッチを取材した結果、富裕層が「リッチスタン (Richistan)」というバーチャル国家の住民とも呼ぶべき状況にあり、

今後は世界各国の富裕層が自国の文化とは異なる富裕層独自の文化を形成するようになると述べている。つまり、車・食事・衣服・旅行など消費の傾向も自国内よりも富裕層の間でより似かよる傾向にあるということだ。このような富裕層の世界で、中国やインドなど新興国の富裕層は今後ますます存在感を増していくであろう。

これに対して貧困層は、地域によっては貧困度合いの偏りがみられる。世界銀行の「世界開発指標 (WDI) 2007」によると、1日1ドル未満で生活する最貧困層（2004年に推定9億8500万人）の割合や1日2ドル未満で生活する貧困層（2004年に推定26億人）の割合は低下しているが、合計すると世界の人口の半分以上が貧困下に置かれている。また、一部の国や地域では、貧困層は雇用機会に恵まれず、教育の機会も限定的で、健康状態にも問題があるなどの理由から経済成長の恩恵を受けることができず、その結果格差が広がったとしている。2008年のWDIによると、南米・カリブ諸国、サハラ以南アフリカでは、上位20%の富裕層が占める所得の割合は下位20%の貧困層の18倍にも上っている。

富裕層の例を見ると、フランスの5大シャトーを始めとする高級ワインに中国やロシアの富裕層が関心を示すという現象はまさに消費傾向の相似を表している。アートに関心を示す新興国の富裕層や投資家も増えている。日本の現代アートの作品に台湾人や韓国人が非常に関心を示して、日本で開催されるオークションに参加するケースが数年前から増えている。中国でも自国の現代アートがバブルと言われるほど投資家などに人気である。

一方、貧困層の例では、08年の世界的な食糧価格の高騰などにより、食糧があっても購入できない層が出てきており、今後の状況を注視する必要がある。08年のハイチでは、米や豆類など日用食料品の価格高騰が引き金となり国民による暴動が1週間以上続き、首相が解任された。エジプトでは小麦価格が高騰して、貧困

層向けに政府が補助金を出していた配給パンが供給不足になり、配給を巡って死者が出るなどの騒ぎとなった。

第3のカテゴリーは、グローバル都市と地方都市である。グローバル化に対応している大都市とそうでない地方都市との格差が、国内はもとより国際間でも生まれている。2006年の中国の地域別GDPを見ると、1位の上海市（20,668元）と最下位の貴州省（1,985元）では約10倍もの所得格差が出ている。中国の著名な経済学者である清華大学の胡鞍鋼教授は、中国には「四つの世界」が存在していると論じている。

上海・北京などの沿海部大都市、広東省・江蘇省など沿海部、中部、そして貴州省など内陸部を中心とする貧困地域である。こうした地域間の所得格差は中国だけではなく、インドやロシアなど他の新興国でも、そして日本や米国など先進国でも起こっている。

グローバル化が進んだ世界では、都市や地域の発展は世界中から資金・企業・優秀な人材をどう誘致できるかが大きな要素となっている。中国の各地域・都市が外国からの投資誘致を競い合っているのは有名な話だ。また、先に話したフォーチュン・グローバル500社の本社がある都市を都市毎の売上で上位10位までランキング（同率順位を含む）すると、1位は東京で、以下パリ、ニューヨーク、ロンドン、北京、ソウル、トロント、マドリッド、チューリヒ、ヒューストン、ミュンヘン、大阪、ローマ、アトランタとなる。1位の東京と10位のアトランタの売上額を比較すると、約8.7倍の差となり、上位都市の間でもこれだけの差がついているのであるから、上位都市と下位都市の間は推して知るべしだ。

最後に、組織と個人の関係性の中でこれまで以上に個人の重要性が増し、働き方もボーダーレス化することを考えると、今後は企業や個人がビジネスをする上でも共通の「価値観」でつながることが大事になるだろう。特に08年秋の米国発金融危機以降、米国流の資本主義が見直され、人々の価値観が大きく多様化を始めた。その1つが、「ソーシャル」という価値観だ。「ソーシャルビジネス」は環境や貧困などの社会問題をビジネスを通じて解決していくという取り組みを指す。

2) グローバル感知力の必要性

こうしたグローバル化は日本とはどう関係しているのだろうか。日本に長くいるとグローバル化を身近に

感じる機会は少ないかも知れない。しかし、これからグローバル化に対応するのに必要なのは、「グローバル感知力」とも言うべき力である。日本と世界の間で起こっていることに敏感になり、違和感を感じることでできる力がより求められるのである。日本では、日本語で全ての生活が完結し、仕事も基本的に日本語で行われるため、一見海外の人たちと自分のポジションを巡って競争が繰り広げられているとは思えない。英語が共通語で様々な国から留学生や移民が来て働いているアメリカなど英語圏では、「フラット化する世界」に象徴されるように、既に自分たちの仕事がインドのIT技術者やコールセンターに移るのではなどの危機感が早くから募っていた。日本はこれまで日本語という言葉が一種の参入障壁になっていたといえる。

しかし、現状では、IT分野のオフショア開発を日本語が話せる中国やインドの技術者が担うようになって、プログラミングや中流程度の開発を日本人が担当する必要がなくなってきた。今後は、プロジェクトの品質管理やITコンサルティングなどより付加価値が高い業務を担当しなければ、仕事がなくなっているのである。ファストフードやコンビニなどの店員で外国人率が非常に高くなっているのは周知の事実だが、デパートなどの小売業でも、最近の中国・香港・韓国など東アジアからの観光客増加によって、接客も中国人などが担うようになってきた。銀座のファンケルストアの1階では、サプリメントを求める中国人観光客が大半で、中国人の店員が対応している。秋葉原の大手家電量販店は中国人観光客のツアールートに入っているため、1つのフロアの店員ほとんどが中国人になり、残る日本人店員も顧客の大半が中国人のため、中国語ができない自分達の仕事がなくなっているという危機感を持っている。

グローバル化の3つめのポイントに述べた「グローバル都市と地方都市」という点では、東京・大阪・名古屋などの大都市以外の日本の地方都市は人口減少、経済停滞などの影響を受けているところが多い。そのため、地域活性化のために観光に注力する都市も多いが、そこでも訪れる観光客は東アジアを中心とする外国人である。JNTOの調査によると、2009年の日本への観光客は韓国、台湾、中国で全体の5割以上を占める。彼らに対して土産物の販売や観光名所の案内、旅館や宿でのサービスなどを提供するには

やはり日本語以外の言語が必要になってくる。地方都市もグローバル化の余波から逃れることはできないし、むしろそれを好機に変えていく努力が求められている。

参考文献：

国際協力銀行「わが国製造業企業の海外事業展開に関する調査」2009年

世界銀行「世界開発指標（WDI）2007」2007年

トーマス・フリードマン『フラット化する世界』日本経済新聞社 2006年

日本政府観光局（JNTO）プレスリリース（平成22年1月25日）

ロバート・フランク『ザ・ニューリッチーアメリカ新富裕層の知られざる実態』ダイヤモンド社 2007年

地方自治体が本格的に導入を始めるホームページ自動翻訳サービス

「My サイト翻訳」

株式会社高電社

ホームページを英語・中国語・韓国語にリアルタイム翻訳

株式会社 高電社（本社：大阪市阿倍野区、代表取締役社長 岩城 陽子）は、ホームページに翻訳開始のリンクボタンを貼付するだけで、日本語のホームページをリアルタイムに外国語へ翻訳するASPサービス「My サイト翻訳」（マイサイトホンヤク）を提供しています。2006年11月、倉敷市役所が日本国内の地方自治体で初めて導入した当サービスは、その後2年間のシステム改良を行ったことで、2008年4月以降、導入する地方自治体が急激に増加しています。



公開順	自治体名	都道府県	導入開始日
1	倉敷市	岡山県	2006年11月1日
2	越谷市	埼玉県	2008年4月1日
3	佐世保市	長崎県	2008年4月1日
4	綾瀬市	神奈川県	2008年7月1日
5	喜多市	福島県	2008年10月1日
6	東海村役場	茨城県	2008年12月1日
7	所沢市	埼玉県	2008年12月1日
8	飯田市	埼玉県	2009年1月30日
9	日立市	茨城県	2009年2月1日
10	豊島区	東京都	2009年2月12日
11	大阪狭山市	大阪府	2009年2月27日
12	鹿嶋市	茨城県	2009年3月2日
13	岸和田市	大阪府	2009年3月3日
14	大田区	東京都	2009年3月31日
15	北区	東京都	2009年4月1日
16	浦安市	千葉県	2009年4月27日
17	南砺市	富山県	2009年5月1日
18	一宮市	愛知県	2009年6月1日
19	市川市	千葉県	2009年7月1日
20	春日部市	埼玉県	2009年7月1日
21	足立区	東京都	2009年7月1日
22	小金井市	東京都	2009年8月1日
23	江東区	東京都	2009年8月3日
24	登米市	宮城県	2009年9月1日
25	板橋区	東京都	2009年9月1日
26	つくば市	茨城県	2009年9月30日
27	寝屋川市	大阪府	2009年11月1日
28	蒲都市	愛知県	2009年11月1日
29	君津市	千葉県	2009年12月1日
30	宇都宮市	栃木県	2010年2月1日
31	芦屋市	兵庫県	2010年2月1日
32	岐阜市	岐阜県	2010年3月1日
33	大田区 子育て支援課	東京都	2010年4月1日
34	清瀬市	東京都	2010年4月1日
35	荒川区	東京都	2010年4月1日
36	東広島市	広島県	2010年4月1日
37	観音寺市	香川県	2010年4月1日
38	*さいたま市	埼玉県	2010年4月15日
39	*浜松市	静岡県	2010年4月19日
40	伊東市	静岡県	2010年5月12日

*印 政令指定都市

■ホームページのすべての情報を外国語に自動翻訳

日本語のホームページを自動翻訳するには、同ホームページの任意ページ上に翻訳開始のリンクボタンを貼付するだけで、即日公開ができるASPサービスであるため、当サービスの利用者にシステムの組み込みなどの負担がかからないところが、第1の利便性といえます。また翻訳開始のリンクボタンは、ホームページ内のどのページにも貼付でき、そのリンクボタンをサイト訪問者が選択した際にどのページから翻訳を行うかを任意に指定できる柔軟性も兼ね備えています。

第2の利便性は、翻訳対象の範囲が同一ドメイン内の下層ページやリンク先のページに渡り、すべて自動的に翻訳されるところにあります。これまで外国語ページを設置している地方自治体の場合でも約十数ページの静的ページが用意され、しかも、多くが英語のみの対応に止まっていますが、これにより在住外国人向けの情報提供量は飛躍的に増加することになります。

第3の利便性は、ホームページ上の複雑なレイアウトやページタイトル、CMS（コンテンツ管理システム）でボタン化したテキスト、画像にカーソルを置いたときに表示される代替テキスト（alt属性）をも翻訳して表示する解析技術の向上により細部にわたり外国語に対応していることです。

第4の利便性は、日本語のホームページをリアルタイムに翻訳することです。

外国語の静的ページは、日本語のホームページと同時にリニューアールすることが、都度の翻訳作業とコストを必要とするため対応が困難といえるでしょう。当サービスは、サイト訪問者が翻訳開始のリンクボタンを選択、または以後のリンクを選択する度に翻訳処理を行うため、ニュースやイベント情報など情報の鮮度を落とすことなくタイムリーに提供できるほか、緊急・災害時の情報基盤としても活用できるようになります。

その他、市役所と同一ドメインで翻訳ページを表示したり、外国語に翻訳している状態で同一言語によるサイト内検索や翻訳したページが海外検索サイトに登録されるような機能も提供しています。

■導入検討のきっかけは、ホームページのリニューアルと翻訳精度

ある関東の自治体では、外国人登録者数が最近10年で1.6倍に増加しており、在住外国人への情報提供が必要になってきているといえます。外国人登録者のうち約30%は中

国系、約20%が韓国及び朝鮮系、同じく約20%がフィリピン系で、その他にもブラジル系、タイ系、パキスタン系、バングラデシュ系など多岐にわたりますが、話される言語は英語・中国語・韓国語が上位を占めています。

当自治体でも従来、在住外国人向けに英語・中国語・韓国語・ポルトガル語版の「ごみカレンダー」や「防災マップ」などを配付し、英語版ホームページを開設して行政情報の提供に努めてきたが、翻訳作業はボランティアの外国人などに頼る部分が多かったため時間的、情報量的にも限界があることは否めず、とくにホームページに新しい情報を掲載することが難しかったといえます。

またMyサイト翻訳の翻訳精度について「自治体の導入実績が最も豊富な自動翻訳サービスとしてMyサイト翻訳のことを知りました。」と話す自治体広報の担当者は、「国際交流担当部門の協力を得て翻訳精度の検証を行ったところ、あらかじめ自動翻訳であることを理解していれば、翻訳結果は非常に満足です。」と評価したことも導入の決め手となっています。

■オンラインユーザー辞書機能に高い評価

外国人訪問者は、日本サイトをどのような方法で閲覧しているのでしょうか。

すでに「無料翻訳サービス」で最も評価が高く、広く利用されている「excite翻訳」では、高電社の中国語や韓国語の翻訳エンジンが採用されていますが、外国人訪問者がこの「無料翻訳サービス」を利用するには同サービスを開いてから翻訳したいページのURLを打ち込む必要があるため、「ワンクリック」サービスとはいかず、また辞書登録など訳質向上の機能が提供されていないため、地名などの固有名詞や行政用語など誤訳する箇所も多かったはずでした。

2006年11月の倉敷市役所導入以降、機械翻訳の結果のみで提供を開始した「Myサイト翻訳」は、提供先ごとに訳質向上が可能なオンラインユーザー辞書登録の機能追加へと進化を始め、2008年4月以降の急激な導入実績増加の要因となりました。

このオンラインユーザー辞書登録は、単語単位、文章単位の登録が可能です。文章単位で登録した場合は、翻訳と同時にネイティブ文章が表示され、ページ上で機械翻訳とネイティブ翻訳の結果を混在させることが可能で、市長のページなどニュアンスをきっちり伝える必要があるページなどに有効です。また、HTMLソースレベルで登録が可能のため、翻訳と同時に画像やFlashなどで表示する箇所

を外国語版の画像やFlashなどに差し替えて表示することも可能になりました。



高電社は、導入サイトを定期的にチェックすると共に単語登録や日本語テキストの文章表現に関するアドバイスを行い、必要に応じて辞書登録の作業代行や研修会の開催を行っているほか、これまで導入した自治体の訳質向上で登録した辞書を基本辞書として一元化し、新規導入時に初期登録を行っています。

新型(英) インフルエンザ	NH1 influenza 新型インフル	感染症
新型(英) インフルエンザの感染予防について	about infection prevention	感染症
新型(英) インフルエンザの注意	be alert to NH1 influenza	感染症
新型(英) インフルエンザ	NH1 influenza	感染症
新型(英) インフルエンザの感染経路	NH1 influenza infection	感染症
新型(英) インフルエンザの感染	NH1 influenza infection	感染症
新型(英) インフルエンザの感染について	about the infection	感染症
新型(英) インフルエンザに関する情報	information of NH1	感染症
新型(英) インフルエンザに関する最新情報	new information	感染症
新型(英) インフルエンザの感染状況	infection status	感染症
新型(英) インフルエンザの感染状況について	status and our response	感染症
新型(英) インフルエンザの感染状況について	new comprehensive	感染症
新型(英) インフルエンザの感染状況について	What's new	感染症
新型(英) インフルエンザの感染状況について	Registration Period	感染症
新型(英) インフルエンザの感染状況について	Application form	感染症
新型(英) インフルエンザの感染状況について	Application form	感染症
新型(英) インフルエンザの感染状況について	Download Application	感染症
新型(英) インフルエンザの感染状況について	Applications service	感染症
新型(英) インフルエンザの感染状況について	Application form	感染症
新型(英) インフルエンザの感染状況について	list of applications	感染症
新型(英) インフルエンザの感染状況について	How to apply	感染症
新型(英) インフルエンザの感染状況について	Application form	感染症
新型(英) インフルエンザの感染状況について	Identification card	感染症
新型(英) インフルエンザの感染状況について	Policy Division	感染症
新型(英) インフルエンザの感染状況について	Policy Planning & Coordination Section	感染症
新型(英) インフルエンザの感染状況について	Policy Management Department	感染症
新型(英) インフルエンザの感染状況について	Policy Planning Evacuation Office	感染症
新型(英) インフルエンザの感染状況について	Policy Formulation Office	感染症
新型(英) インフルエンザの感染状況について	Refuge Management Section	感染症

■ 自動翻訳の訳質向上に取り組む地方自治体

一般的に機械翻訳による訳質は、原文の日本語の状態に依存して変わります。訳質向上のポイントとして、次のようなことが考えられます。

■ 訳質低下を招く日本語原文の要因

- (1) 理解しにくい長文
- (2) 文節区切りの位置
- (3) 目的語の位置
- (4) 造語、新語、専門用語など
- (5) 比喩的、口語的、専門的な表現
- (6) 文の途中の改行
- (7) 誤字、脱字など

■ 訳質向上のポイント

- (1) 長文を避け、複数の短文で書くようにする。
- (2) 動詞の区切りで文を区切り、助詞や構文を整える。
- (3) 箇条書きを利用する。
- (4) 記述しなくても意味が伝わる表記は省略する。
- (5) まわりくどい言い回しを避ける。
- (6) 同じ単語や同義語が、文中に重複するのを避ける。
- (7) ある語句がどの語句と関連しているかをはっきり示すようにする。
- (8) 主語・目的語はできるだけ省略せずに明記する。
- (9) 同音異義語・ひらがな表記に注意する。
- (10) 比喩的な表現や造語の記述を避ける。
- (11) 前後の文脈がはっきり分かるようにする。
- (12) 多義的用語を適切な動詞に変える。

2009年11月に「Myサイト翻訳」を導入した大阪府寝屋川市役所では、オンラインユーザー辞書登録の活用はもとより公式サイトに記載される日本語について、上記内容に気を配り、可能な限り機械翻訳で良質の翻訳結果が表示されるための試みを行っています。

また同市役所では、「日本人と外国人の間であっても、コミュニケーションを分け隔てなくできるということが最も大切なことと考え、そのためにも、誰が見てもわかるようなホームページを作っていきたい。」と考えられています。



■進化する日中・中日翻訳エンジン

最近の新聞報道では、既に中国のネット利用者数が米国を抜き世界一になったと報じられ、急激な経済成長とともに中国語が英語を抜き「世界最多の言語」となる日も近いのかもしれない。

当社の日中・中日翻訳エンジンは、1998年中国語翻訳ソフトウェアを発売して以来、十数年に渡る改良を続けてきました。この改良の歴史において、当社エンジンは飛躍的な進歩をとげました。

ソフトウェア的には、翻訳辞書容量が拡大し登録単語数は大幅に増加しました。翻訳技術においては、開発当初主流とされていた文法中心のルールベースに加え、文例翻訳や統計的な手法を取り入れたハイブリッド型へと進化しています。これらの改良により、以前は部分的にしか理解できない翻訳レベルでしたが、文意をほぼ正確に理解することができるレベルに到達しつつあります。また、多様化する情報に対しては分野を特定して専門用語辞書を利用したり、固有名詞についてはユーザー辞書を活用するなど翻訳環境を充実させることで、利用者が望む適切な訳出を実現することができるようになったことも翻訳率向上の大きな要因と言えます。

日本語と中国語では言語的にも相違点が多く、解決が困難な問題も多々あります。特に、原文そのものに曖昧性があつたり、文脈、状況、立場によって解釈が異なる場合、機械翻訳がクリアすべき課題は山積しております。しかし、機械翻訳の強みが発揮される領域、例えばマニュアルや論文などの規則的で明確な文章の翻訳は、スピード・統一性においてはもちろんのこと、翻訳率においてもヒューマン翻訳に劣らない有用性が、近い将来期待できると思っています。

今後も、当社は常に進化しつづける翻訳エンジンを提供して参りたいと思っています。

China Workshop on Machine Translation (CWMT09) 参加報告

株式会社富士通研究所
長瀬 友樹

はじめに

第5回中国機械翻訳研究会 (China Workshop on Machine Translation、以下CWMT09) が、2009年10月13-14日、南京大学で開催された。この会議は一年に一度中国の機械翻訳研究者が一堂に集まって成果を発表する場であり、2005年にアモイで第一回大会が開かれて以来、ハルビン、北京を回って、今回が第5回目の大会である。ちなみに、第3回大会までは「Symposium on Statistics Machine Translation, SSMT」という名の下でSMT中心の研究発表がなされていたが、昨年から会議の名称が改められ、機械翻訳のより広い分野で発表が行われるようになった。

私が会議に参加した主たる目的は、AAMT機械翻訳課題調査委員会WG1の活動成果の発表である。WG1で構築してきた日中機械翻訳文の評価のためのテストセット（以下、AAMTテストセット）の概要説明と、AAMTテストセットを用いて実際に代表的日中エンジンの評価実験を行った結果についての報告を行った。

南京市は古くから長江流域・華南の中心地で、かつて首都が置かれたこともある。人口700万人を超える大都市であるが、最近の北京や上海に比べると、良い意味で中国らしさを残した街だと思う。南京といえば、多くの日本人にとって、過去に不幸な歴史があった場所としての印象が強いかもしれない。私も南京市民の対日感情が悪いのではないかと心配したが、これは杞憂であった。3日間の滞在中、自分が日本人であることから不愉快な思いをしたことは一度もなかった。

会議概要

会場は南京市のほぼ中央に位置する南京大学である。中国を代表する総合大学のひとつで、広い敷地の中に重厚な建築物が点在しており、大学の歴史を感じることができる。会議が行われたのは比較的大きな教室であったが、教室の後ろに予備の椅子を並べてやっと全員が座れるという満員の状態だった（写真1）。とにかく人が多く熱気がムンムンするところも、最近の中国を象徴しているようで面白

かった。並行のトラックはなく、全員が同じ教室で同じ発表を聞く形である。参加人数はおよそ150名でほぼ全員が中国人。海外からは私の他にシンガポールと米国からの発表があったが、発表者はどちらも中国語のネイティブだった。日本からの参加は私一人であり、参加者の顔を見た限りではアジア人以外の参加者を見つけることはできなかった。

会期の1日目は、プログラム委員長の中国科学院Sun教授による開会の挨拶のあと、評価タスクに参加した各チームからのプレゼンが行われた。夕刻には大学のレストランでレセプションがあった。2日目は一般の技術論文発表のセッションにおいて12件の発表が行われた。



写真1 会場風景

評価タスク参加チームからの報告

CWMTは、あらかじめ共通の評価タスクを設定して参加チームを募り、各チームは事前に配布された学習コーパスを使って機械翻訳システムを開発し、その結果を本会議で報告するという形で進められる。ちょうど、IWSLTやNISTなどの評価型ワークショップで行われているやり方と同様である。ただしCWMTの場合はテキストからテキストへの翻訳であり、音声技術の評価は含まれない。今回設定されたタスクは以下の5つである。

- (1) 中英翻訳 新聞記事（単一システム）

(2) 中英翻訳 新聞記事 (融合システム)

(3) 英中翻訳 新聞記事

(4) 英中翻訳 科学技術文献

(5) 中蒙翻訳 日常会話文

参加チームは17で、うち外資系企業としてSYSTRAN、Microsoft、Fujitsu、HKUSTの4団体が含まれる。ただしSYSTRAN以外は中国本土（または香港）にある拠点からの参加であり、実質的な外国企業はSYSTRANのみである。各システムの出力はBLEU、NIST、GTM、mWER、mPER、ICTで評価され、予稿集に評価結果の一覧表が掲載されている。中英翻訳はMicrosoft、英中翻訳では中国技術院計算技術研究所のシステムが、最も良い成績を納めた。評価に用いられたシステムは大半がSMTベースであるが、ルールベースシステムを基本とするSYSTRANが健闘（タスク(1)で3位など）していたのが印象に残った。

一般技術論文発表

会議2日目は一般技術報告のセッションで12件の発表があった。そのうちの1件として、「AAMTテストセットによる日中機械エンジンの評価」と題して、私からAAMT機械翻訳課題調査委員会WG1の成果報告を行った。具体的には日中翻訳のために構築したテストセット（評価例文と評価用設問文）の説明と、テストセットを使って実際に日中エンジンを評価した結果についての報告である。2日間のセッションで英語による発表は私だけだったが、聴衆は熱心に聞いてくれ質疑も活発だった。「設問を作っているのは言語学の専門家か？」という意味の質問を複数の人から受けたことから、テストセットの利用者は設問設定の網羅性や妥当性について関心が高いことがわかった。今後検証していく必要があると感じた。

なお、前日のレセプションで何人かの中国人と会話をしたがAAMTのことを知らない人が意外に多かったため、自分の発表の冒頭でAAMTの全体的な紹介を行ってから本論に入るようにした。AAMTの存在と活動内容について中国の機械翻訳関係者に知ってもらうために一定の効果があつたと考えている。



写真2 昼食風景

言語の壁

余談になるが、CWMTは外国人の参加を前提にしておらず、重要な情報の多くが中国語のみで通知された。そのため、今回CWMTでの発表に纏わる経験から、奇しくも、自分自身が言語の壁を実感することとなった。たとえば、論文の採録通知、投稿原稿の執筆要領、予稿集、会場で配られる参加章（写真2）などがすべて中国語表記であり、読み解くのに大変苦勞した。当日の会場においては、中国語を理解できない私には中国語での発表を聞いても内容を十分に理解することができないため、苦難の連続であった。このような場面では、やはり機械翻訳は便利なツールである。実際に中国語のメールを読む際に何度か機械翻訳のお世話になった。もしも、看板や資料を撮影すると瞬時に文字を認識して翻訳してくれるカメラや、中国語による発表をリアルタイムに翻訳して読み上げてくれるツールなどが使えたとしたら、今回の中国出張がより有意義なものになっていたことは間違いない。



写真3 参加章

おわりに

CWMT09への参加を通じて、中国の機械翻訳研究が想像以上に活発であることを知った。特にSMTの研究では、研究者の層の厚さや研究規模の大きさに日本を凌ぐ勢いを感じた。昨今、中国は経済をはじめとして様々な領域で急成長を遂げているが、機械翻訳研究も例外ではないようである。一方、機械翻訳の社会への普及を考えた場合には、AAMTのように大学とメーカ、ユーザが一緒になって様々な観点から機械翻訳の課題について議論するような場が、今後の中国にも必要かもしれない。次回のMT SUMMITが中国のアモイで開催されることを契機として、AAMTが日中の

機械翻訳関係者の交流を促進することができれば、両国の機械翻訳にとってこの上なく意義深いことである。

今回の会議参加を通じて、中国科学院のSun先生、ハルビン工大のZhao先生をはじめ、中国を代表する機械翻訳研究者との交流を深められたことは大きな収穫だった。また、中国の学生や若手研究者にAAMTの存在と活動についてアピールできたことも良かったと思う。ただ、南京には北京出張の帰りに立ち寄っただけの短い滞在だったため、名所を見て回る時間がなかったことが残念である。次の機会があれば、緑豊かな華南の古都をゆっくりと散策してみたい。

LISA China Focus 発表報告

株式会社富士通研究所
大倉 清司

AAMT課題調査委員会 標準化・共有化ワーキンググループ (WG3) で策定しているユーザ辞書標準フォーマットUTX (※) について、LISA China Focusで発表してきました。

LISA China Focusは、LISA (The Localization Industry Standards Association)が主催するワークショップで、2009年11月16日 (月) から11月18日 (水) まで中国・北京のRadisson SAS Hotelにて開催され、164人が参加しました。

LISAが規定する翻訳メモリの標準フォーマットTMXは広く使われており、日本の機械翻訳メーカーも翻訳メモリの形式としてTMXをサポートしています。LISAは用語集の標準フォーマットとしてはTBXを規定していますが、仕様が複雑であり、事実上広く使われていないのが現状です。

今回、LISAのワークショップに参加した目的は、UTXについて発表し、翻訳に携わる海外の多くの人々にAAMTおよびUTXのことについて知ってもらう他に、会議の参加者と交流を深め、機械翻訳をもっと使ってもらうように促すこと、またLISAと共同で、辞書の標準化に向けて連携できないか打診することでした。

参加者はローカライズの会社、翻訳会社がほとんどでしたが、機械翻訳や翻訳メモリツールなどを開発しているベンダーも参加していました。全体的に、標準化に関して非常に興味が高かったです。かなりの (翻訳) 会社が、10以上の言語を扱っている印象で、自国語と英語だけの翻訳だけという会社はなかったように思います。ローカライズ業界では、多言語を扱うのは当然のようです。

機械翻訳は実際のローカライズにおいてはほとんど使われていないようです。一部導入している会社もありますが、効率化できていないようでした。一方、翻訳メモリは使われています。ローカライズにおいても辞書引きは多用されているが、それは機械翻訳技術を使って簡単に組み込めることを、参加者との交流を通じてアピールしてきました。

私の発表では、標準化のパネルセッションと同時開催だったにもかかわらず40人くらいの聴講者がいました。AAMTのことを知っていたのは7名程度 (ほとんど日本人) でした。

質疑応答では主に以下のようなやりとりがあり、UTXに対する高い関心がうかがえました：

Q. (発表では、英語、日本語、中国語に関してUTX-S辞書の例を示したが) 他の言語でも使えるのか？

A. 全ての現象を扱えるわけではないが、最低限 (名詞など) はできると考えている。

Q. 効率の評価はしたか？

A. していない。

Q. 効率よくUTXをつくるツールなどはあるか？

A. 今のところはない。

発表の後、個人的な質問もありました：

Q. 中国語の翻訳ソフトを業務に使っているが、UTX-Sをそのまま使えるのか？

A. コンバーターを用意する必要がある。将来的に翻訳ソフトがサポートすれば使えるようになる。

LISAとの協力体制については、事務局長のMichael Anobile氏と、標準化を推進するArle Lommel氏と話し合った結果、AAMTとLISAの関係を深め、お互いに協力して標準化をすすめることになりました。帰国後、LISAからAAMTとメンバーシップを交換することについて提案があり、AAMT側は基本的に合意しています。理事会で正式に決定することになりそうです。

今後ですが、これから開発されていく翻訳エンジンもあるでしょうから、国際会議などで継続的にUTXを発表していくことが重要だと感じました。それにより、多くの翻訳エンジンがUTXをサポートすれば、世界中の人々が辞書を共有できるようになります。

※ UTXについては以下のURLをご参照ください。

<http://www.aamt.info/japanese/utx/>

Common Format for User Dictionary UTX (Universal Terminology eXchange)

WG3, MT Research Committee

"Translation software is useless" Are you sure?

Advantages of UTX

- Significantly improve the translation accuracy
- Accumulate, share, and reuse translation assets
- Cut the time and cost to check up terminology

Introduction

UTX (universal terminology eXchange) is a common format for user dictionaries established by the AAMT (Asia-Pacific Association for Machine Translation). It is an open specification, and the AAMT is working towards its standardization. The AAMT is an organization consisting of researchers, developers, manufacturers, and users of machine translation (MT). MT is the core technology of translation software.

Characteristics

UTX greatly improves the accuracy of translation software by sharing the knowledge of terminology in a bilingual format. The goal of UTX is to **create a simple, easy to make, easy to use dictionary** from a user's viewpoint, not a developer's.

Different manufactures' translation software can use the same UTX dictionary. Also, a UTX dictionary is human-readable, and it can be used as a glossary which does not involve translation software at all.

Why UTX?

With UTX, a user can easily share and reuse user dictionaries of translation software. When translation software fails to translate correctly, the problem is often that it **doesn't have sufficient translation knowledge of certain words and phrases** that should be translated. UTX can greatly improve the accuracy of translation software by supplementing translation knowledge as user dictionaries.

Until now, even if a user of translation software makes the effort to prepare user dictionaries, the knowledge is fragmented and dispersed, and thus not effective. Also, even a simple plain text file is difficult to share or to reuse, if its format is not standardized as a dictionary. However, if a single standard like UTX is adopted, dictionaries can be widely shared among various tools, such

as translation software from different manufactures, and they can also be reused quickly.



Types

There are two types of UTX:

UTX-Simple ... A simple tab-delimited format.

UTX-XML ... A XML format which can contain many data (specification under development).

Who uses UTX?

UTX is specifically designed to be used by end users of translation software, or translators. It does not require any advanced technical knowledge of linguistics, grammars, or machine translation software etc. to create or use it. It can be made only from minimum data such as basic parts of speech (noun, verb, etc.), and the plural form, if the entry is a noun.

In which domains?

UTX can be used in any domain, if it is specialized to a certain extent, such as ICT, medicine, law, engineering translation etc. It may not be suitable for translation of non-specialized, general contents.

What kind of words should we include?

Basically, a UTX dictionary should contain **only technical terms** of a specific domain. In most cases, entries are nouns, especially compound nouns. For example, a word like "XML declaration" can be correctly translated into its Japanese equivalent, "XML 宣言" by just registering it in a user dictionary. Basic vocabulary like "window" should not be included, because such words are already contained in the system dictionaries of translation software. Translation accuracy can be improved by collecting, sharing, and reusing the data of fine-tuned bilingual translations which are not included in translation software out-of-box.

Sentences should not be included, except when it is appropriate to treat them as a kind of "words." As a rule, UTX should be separated from translation memory, which is a bilingual database of sentences, rather than words.

Is it multilingual?

Yes. The character code of UTX is Unicode (UTF-8) and all the languages that can be handled by UTF-8 are accepted. However, a UTX-Simple dictionary includes only entries of the single source

language A, and their translations in the single target language B. The dictionary from B to A comprises another dictionary.

How do we make UTX?

UTX-Simple dictionaries can be easily created and edited with any text editor or spread sheet application. With a conversion tool, a UTX-Simple format dictionary can be converted into some proprietary formats for translation software.

URL for the conversion tool

<http://www.aamt.info/japanese/utx/>

A tool is also under development to convert various formats to UTX format.

To open source developers and translators - Why not release your glossary in UTX format, and share it with others?

By releasing and sharing glossaries, you can localize your software more quickly, and allow users to access it from all over the world.

UTX mailing list

Anyone can participate in the discussions on UTX through a mailing list.

Please search the Web with the keywords: "UTX mailing list."

Reference materials

Okura Seiji et al. (2008) "Introduction to UTX, a specification for a shared user dictionary" 13th Annual Meeting of Association for Natural Language Processing

Francis Bond et al. (2009) "Sharing User Dictionaries across Multiple Systems with UTX-S" in Second International Workshop on Intercultural Collaboration (IWIC2009), Stanford

Example

(based on UTX-Simple 1.0)

#UTX-S 1.0; en-US/ja-JP; 2008-03-15T10:00:00Z+09:00; copyright: Francis Bond (2008); license: CC-by 3.0			
#src	tgt	src:pos	src:plural
new	新規の	adj	
fast	高速な	adj	
prosody	韻律	noun	prosodies
save	保存する	verb	

denotes a comment line.

- 1st line (comment line): Basic information on the dictionary. Each item is divided by a semicolon and a space. "UTX-S" means UTX-Simple.

UTX-S <version number>; <source language>/<target language>; <last update timestamp>; copyright: <Copyright holder>; license: <License>; <Additional data> (optional)

2nd line (comment line): The names of properties. Each item is tab-delimited.

In the above example:

<Source word> <Target word> <Part of speech of the source word> <Plural form of the source word>

The 3rd and subsequent lines contain actual entries. Each item is tab-delimited.

The character-encoding method is UTF-8 (with no BOM).

Tips for making a UTX dictionary (glossary)

- Define the domain of the dictionary clearly.
- Only use upper case for proper nouns.
- The basic form of word should be entered (singular form for a noun, root form for a verb - as you would see in a commercial dictionary).
- Any comments should be noted separately in the comment field, not as a part of the entry.
- Choose only the single, most appropriate translation corresponding to a source word. If it has multiple DISTINCTLY different meanings, they can be treated as separate entries.

Disclaimer

By using the specifications of UTX, UTX-Simple, and UTX-XML (hereinafter collectively called "UTX Specifications") or the dictionaries based on UTX Specifications (hereinafter called "UTX Dictionaries"), you agree to be bound by the following terms. The invalidity or unenforceability of this disclaimer shall in no way affect the validity or enforceability of any other provision herein.

1. To the authors of UTX Dictionaries and related tools from the AAMT and its members:

(1) UTX Specifications are made public, and anyone can use them. The AAMT, however, does not waive any rights thereof and no one may alter UTX Specifications nor make them public.

(2) THE AAMT AND ITS MEMBERS PROVIDE UTX SPECIFICATIONS "AS IS," WITH NO GUARANTEES WHATSOEVER. YOU SHOULD USE UTX SPECIFICATIONS AND UTX DICTIONARIES AT YOUR OWN RISK.

(3) THE AAMT AND ITS MEMBERS SHALL NOT ASSUME ANY RESPONSIBILITY FOR UTX SPECIFICATIONS AND THE RESULT OF THEIR USE INCLUDING, BUT NOT LIMITED TO, THE EXISTENCE OF INFRINGEMENT OF THIRD PARTIES' RIGHTS AND THE ACCURACY, ADEQUACY AND QUALITY OF THE TRANSLATION.

(4) THE AAMT AND ITS MEMBERS SHALL NOT ASSUME ANY RESPONSIBILITY FOR VERIFYING NOR DO THEY GUARANTEE THE LEGITIMACY OF THE COPYRIGHT FOR EACH UTX DICTIONARY. YOU AND THE ORIGINAL AUTHOR OF EACH UTX DICTIONARY ARE RESPONSIBLE FOR THE LEGAL PROBLEM IF IN ANY CASE THAT THE ORIGINAL AUTHOR OF THE UTX DICTIONARY IS NOT THE LEGITIMATE HOLDER OF THE APPROPRIATE COPYRIGHT.

(5) The AAMT and its members grant you the permission to stipulate the terms and conditions for the use of UTX Dictionaries by their users for commercial or non-commercial purposes as long as you have the appropriate copyright; provided, however, that the author of UTX Dictionary is solely responsible for verifying the legitimacy of the copyright for data used in the UTX Dictionary.

(6) THE AAMT AND ITS MEMBERS SHALL NOT ASSUME ANY RESPONSIBILITY FOR THE RESULT OF USE OF THE TOOLS RELATED TO UTX DICTIONARIES.

2. To the users of UTX Dictionaries from their authors:

The users of UTX Dictionaries may make use of UTX Dictionaries, in accordance with their license terms and conditions. Since the license terms and conditions of UTX Dictionaries are varied, please confirm the license indicated in the UTX file header.

3. To the users of UTX Dictionaries and related tools from the AAMT and its members:

THE AAMT AND ITS MEMBERS SHALL NOT ASSUME ANY RESPONSIBILITY IN CONNECTION WITH THE UTX SPECIFICATIONS AND THE RESULT OF THEIR USE INCLUDING, BUT NOT LIMITED TO, THE EXISTENCE OF INFRINGEMENT OF THIRD PARTIES' RIGHTS AND THE ACCURACY, ADEQUACY AND QUALITY OF THE TRANSLATION. You should resolve such problems between you and the author of the UTX Dictionaries.

MT Research Committee

Members of Sharing/Standardization Working Group (not in a particular order)

YAMAMOTO Yuji (leader)	CosmosHouse
ITOU Hajime	Inter Group Corp.
MURATA Toshiki	Oki Electric Industry Co., Ltd.
Francis Bond	Nanyang Technological University
SHIMAZU Miwako	Toshiba Solutions Corporation
OKURA Seiji	Fujitsu Laboratories Limited
Michael Konin Kato	Learning Consultant

The latest version of this brochure is available from:

<http://www.aamt.info/english/utx/>

Contact: aamt-info@aamt.info

December 2009 edition

AAMT会員のひろば

AAMT 会員の新たな交流の場を AAMT Journal 誌面上で提供するべくスタートいたしました「AAMT 会員のひろば」、会員の皆さまのご助力をいただきまして、第一回の No.41 のスタートから第七回を迎えることができました。今号では、法人会員一社、個人会員二名の皆さまからのご寄稿をいただいております。

独自のお取組みのご紹介、機械翻訳研究への提言、AAMT の活動へのご要望など、今回も貴重なご意見をお寄せいただきました。

AAMT Journal では今後も引き続き、会員の皆さまからのご寄稿を心よりお待ちしております。

ご寄稿・お問い合わせは AAMT 事務局(E-mail: AAMT-info@AAMT.info)まで宜しくお願いいたします。

法人会員（敬称略・50 音順）

会員名

勝田美保子（株式会社十印 会長）／KATSUTA Mihoko（TOIN Corporation）

自己紹介

私は、1963年に事業を立ち上げ、たまたま時流に乗って翻訳、通訳、マニュアル制作、その他、国内向け、海外向けテクニカルドキュメンテーションサービスに取り組んで参りました。今は、更にソフトウェアのローカライゼーションが主役です。

約3年前まで、20年余、私は日本翻訳連盟(JTF)の会長をつとめました。20数年前になりますが、JTFで10数人の会員でルクセンブルグに機械翻訳開発の状況を見学に行ったことがあります。大変興味深い旅でした。今、AAMTとJTFは相互会員でもあり、何かと協力関係にあります。

MT/翻訳とのかかわり

今から30年ほどまえ、京都大学の工学部の教授でおられた長尾先生のご依頼をうけて、弊社の若い社員が、機械翻訳開発の研究室でお手伝いをさせていただく事になりました。先生の機械翻訳開発プロジェクトに多少でも参加できるのは翻訳会社としては大変光栄でした。

また機械翻訳協会が設立され、一流大学の先生方や大企業の研究者の方々が協会の中心でした。たまたま私も理事の末席においていただき、機械翻訳のユーザーとして、勉強させていただきました。

以来、長年にわたって長尾先生とのご縁が出来ました事は大変幸運でした。先生は学者として研究者として、情熱と integrity のある本当に立派な方で、感動する事もしばしばでした。

MTおよび翻訳業界に期待すること

ヨーロッパ共同体 (EU) のように国境を越えた共同体で、機械翻訳は非常に順調に使用されているようです。EU のプログラムの一つである Moses を始めとした英語を含むヨーロッパ言語では、その成果も着実に上がっているようです。また、ヨーロッパの一般企業においても多くの成功事例が紹介されています。日本を含むアジアでも、アジア共同体構想が着実に進んでいるということで、日本の翻訳業界ではアジア言語に対処して行く事が、いま重要な課題となっています。産業翻訳業界では、厳しい経済環境の中で、お客様がビジネスを常に有利に進められるよう、翻訳のコスト、スピード、品質のみならず、更なるご要望に応じて行く事が現在の課題です。

あらゆるビジネスの世界に、次々と大きな変化の波が押し寄せてくる中で、翻訳業界にとって、機械翻訳の開発は、未来を制覇する大きな波ともいえるでしょう。しかし機械翻訳のシステム開発はコンピュータ、人工知能、その他のテクノロジーのレベルの進歩にもよる事で、これからも容易なことではないと思います。これに携わる方々の忍耐力と、あきらめずに開発を続けていかれる情熱に心から期待し、感謝しています。言うまでもなくその成果は翻訳業界だけではなく、人類への大きな貢献である事は言うまでもありません。

AAMT への要望

沢山の情報をメールしていただいて有難うございます。大変参考にして居ります。

個人会員 (敬称略・50 音順)

会員名

守屋 雅樹 (MORIYA Masaki)

自己紹介

1.1. 機械翻訳はプログラミングの楽しい題材

損害保険会社のシステム子会社で SE をしています。勤務先とは無関係に、個人会員として AAMT へ入会致しました。

私の趣味の一つはプログラミングで、その主題の一つとして、エスペラント/日本語の機械翻訳システムを開発しています (<http://ttt.ermitejo.com/>)。目下のところこのサイトでは辞書引きまがいの些末なシステムしか運用していませんが、HPSG による構文解析・意味解析やそれ以降の生成系処理などを順次実装・公開していくつもりです。これをライフワークだと位置付けてしまうと完成までの期間が長くなってしまうので、地道に続けつつも数年でなにがしかの目処を付けたいと考えています。なお、現在は Perl で実装していますが、処理のホットスポットは C で再実装することも考えています。

人工言語であるエスペラントであれば、機械翻訳との相性が良いようにも思えるかも知れません。しかし事實はそうではなく、自然言語の場合とあまり変わらないと感じています。エスペラントは規則の例外が少ないだけで、語彙・構文・文脈のそれぞれにおいて、曖昧性がどうしても介在してしまうからです。ロジバンのようにより緻密な言語設計がなされた言語であれば、また違った世界が見えてくるのかも知れません。

ともかくも、機械翻訳は奥の深い分野であって、計算機における自然言語処理の一つの精華であります。しか

しそれだけに、歯ごたえのある題材とも捉えることができます。また、言語を計算するというのが、初めて“Hello, world!”を印字した時にも似た「動いて楽しい」という快感を私にもたらししてくれます。個人が趣味で楽しんで開発する機械翻訳の水準は専門家の方々からご覧になれば極めて拙いものかも知れませんが、素人だからこそ切り込めるような設計・実装などをできれば良いなと思っています。

また、エスペラントをピボット言語とする機械翻訳システムはともかく、エスペラントそれ自体の翻訳を目的とするシステムはおおよそ商業上の成功が見込めないため、市販製品が対応してくれることは望み薄です。こうしたマイナー言語には個人の趣味でないとなかなか光が当たらないので、日本におけるエスペラントの中心団体である日本エスペラント学会の会員でもある私が何とかしたいという気持ちもあります。

1.2. AAMT への入会動機

大学では歴史を学び、社会に出てからは保険屋のシステムを作っているという機械翻訳門外漢の私には、自然言語処理全般の学術的・理論的な背骨が決定的に欠けています。それらは専門書を漁り読むことで補える余地がありますが、設計ノウハウやシステムの使われ方に至っては全くの五里霧中です。

そこで、それらについてなにがしかの示唆が得られるのではないかと思います、2007 年に AAMT の門を叩きました。目論見はおおむね正しく、本誌『AAMT Journal』や本会ウェブサイトなどで得られる情報は、私にとって大変貴重なものになっています。

1.3. UTX-Simple 用ツールの開発

翻訳ソフト用ユーザー辞書の標準仕様として、AAMT では機械翻訳課題調査委員会 WG3 にて UTX を策定しています。これは大変意欲的な取り組みだと感じており、私も諸手を挙げて賛成しています。上述した拙作のエスペラント/日本語翻訳システムでもユーザー辞書を使う設計をしていて、その辞書を UTX-Simple 仕様に準拠したものとするを想定しています。その内部では処理最適化の為に辞書を独自のデータ構造体にして取り扱うこととなりますが、その為の相互変換器を細々と自作していました。これは UTX-Simple 形式を解析して Perl の内部データ構造に変換した上で拙作のシステムのデータ構造を生成したり、その逆方向の処理を行うものです。

この内、UTX-Simple 形式を解析・生成する部分については普遍性があるのではないかと考え、Perl モジュールとして一般公開しようと思い至りました。本会ウェブサイト上の UTX 解説ページ (<http://www.aamt.info/japanese/utx/>) からお問い合わせ申し上げた縁で、機械翻訳課題調査委員会 WG3 の会合にお邪魔してご相談し、様々なご指摘・ご示唆をいただきました。

その結果、オープンなライセンス (Perl モジュールとして最も一般的な、Artistic License と GNU GPL とのデュアルライセンス) を適用した上での一般公開についてご快諾いただくことができました。関係各位にはこの場を借りてお礼を申し上げます。また、公開時期については当初の目論見を大きく外してしまいました。ご迷惑をお掛けして申し訳ありません。この記事が載った AAMT Journal が発行される頃までには、Perl モジュールのオンラインアーカイブである「CPAN (Comprehensive Perl Archive Network)」 (<http://www.cpan.org/>) で公開できる予定です。また、開発版のソースコード一式はオープンソースソフトウェアのホスティングサービスである「GitHub」でも公開しています (<http://github.com/gardejo/p5-text-utx/>)。

同モジュール「Text::UTX」は UTX 形式の抽象化レイヤーという定義ですが、各フォーマットの解析部 (パーサー) および生成部 (ライター) をプラグインの形で容易に組み込めるように設計しています。現在は UTX-Simple の各バージョンに対応したパーサーとライターを同梱していますが、一通りの実装が終わったら、WG3 と適切に連携させていただきながら、WG3 に参画されている会員メーカー各社様の製品の独自フォーマットに対応したパーサー

とライターも順次開発・公開していく予定です。

しかしながら、Text::UTX がエンドユーザーの方々にとって手軽に親しめるものだと、私自身考えていません。辞書をただ変換したいだけの人に対して「まず Perl 処理系をインストールして、UTX::Simple モジュールをインストールして……」などというお願いが通じないことは明白だからです。それは UTX の「作りやすく、使いやすい」という目標とも相容れません。そこで、あるフォーマットのユーザー辞書をアップロードすると、指定した別のフォーマットに変換されたユーザー辞書をダウンロードできるというような、オンライン変換サービスを将来的に開発・公開したいとも考えています。

機械翻訳および翻訳業界に期待すること

2.1. 機械翻訳の活用法のさらなる広報を望みます

製品は素晴らしいのですから、売り方を改善するとともにユーザーに喜ばれると思います。機械翻訳は万能の器ではありません。漫画『ドラえもん』に登場する 22 世紀の道具「翻訳こんにやく」と違って、人間は機械翻訳との上手な付き合い方を模索する必要があります。そして、機械は人間に歩み寄れませんが、機械を作り出した人であれば歩み寄ることができます。

一般のユーザーが機械翻訳に抱く評価は、残念ながら芳しいものではありません。機械翻訳の使い方の勘所や機械翻訳の活用法を、メーカーの方々はアピールしきれているでしょうか？

製品によってはプロやセミプロ相手の価格設定がなされたものもあり、メーカーにとっては「分かっている人」相手の商売という位置付けがなされているのかも知れません。しかし分からせるための努力を惜しんで、ノンプロのユーザーは「分からない人」のままで終わってしまいます。購入したユーザーは幻滅し、その口コミを見聞きした見込み客が購入を断念してしまい、販売機会の拡大もまた望めなくなるでしょう。

2.2. 無料ウェブ翻訳サイトの功罪

AAMT Journal 46 号のアンケート結果にもありましたように、日本においてはウェブ翻訳が翻訳ソフトの代わりとして一般ユーザーに広く使われています。それらのサイトではユーザー辞書という概念自体が存在せず、訳語や構文の候補も選択できず、翻訳メモリーも使えないなど、現代的な翻訳ソフトや翻訳支援ソフトの特性を十全には活かしかねているようにお見受けします。そして、訳出過程にユーザーが関与する道が事実上閉ざされているため、機械翻訳の可用性に対して一般ユーザーが抱く心象をより深刻なものにしかねません。“無償サイトの場合は「無料なのでこれくらいでも OK」という割り切り”（機械翻訳課題調査委員会 WG2「英日・日英機械翻訳に関するアンケート結果報告」『AAMT Journal』第 46 号(2009 年), p. 35)が生じるにしても、機械翻訳の真髄を知らずして満足していることに対しては、前座の嘶だけ聴いて帰ってしまうような勿体なさを感じてなりません。

提携先ウェブサイトのページビューを減らすようなことはなかなか書き難いかも知れませんが、無料ウェブ翻訳サイトのできることとできないことを明示比較し、より機械翻訳を使いこなしたい人へ向けて、自社製品の購買に繋がるような広報を行っていただきたいです。

2.3. ウェブ翻訳サービスはスタンドアローンソフトのサブセット版とは限らない

上述した拙作のエスペラント/日本語翻訳システムでは、英和・和英翻訳のスタンドアローン製品と同様にユーザー辞書・適訳選択・翻訳メモリー等々の満艦飾の機能を盛り込もうと、夢だけは大きく描いています。これはスタンドアローンソフトのサブセット版のようなウェブ翻訳サービスに対するアンチテーゼの意味合いを含めたものですが、そもそも、ウェブ翻訳サービスはなぜ機能が比較的制限されているのでしょうか。各社の経営戦

略上の理由は窺い知れませんが、そこに技術的な制約がかつて横たわっていたことは確かでしょう。

まず、機械翻訳そのものが計算機資源を莫大に要求する高価な処理であり、長大な原文を与えてしまった場合、クライアント側のタイムアウト時間より先に処理を返せる保証があるわけではありません。一方で、原文を細切れに与えようとしても、セッションを何らかの形で維持するなどの面倒事が増えます。ユーザーインターフェースもこれまでは比較的貧弱でした。

しかし、RIA(Rich Internet Application)方面の長足の進歩により、技術面での様々な障壁は取り払われつつあります。HTML5 が常用されるようになるのはもう少し先でしょうけれども、JavaScript (jQuery や Prototype などのフレームワークの援用を前提とします)・Adobe Flash・Microsoft Silverlight などの RIA 基盤は既に十分に実用可能な程度にこなれています。例えば、同一の画面で原文・訳文のアラインメントを順次付け足していくことで、ページ遷移なしで細切れに訳出することが可能となります。また、一度に複数の原文を与えても、訳出が終わった訳文から順次画面に表示していくことで、ユーザーの「待たされ感」を低減することもできるでしょう。複数候補からの適訳の選択などにも、スタンドアローンソフトと遜色のないリッチなインターフェースを提供することができるでしょう。そして、それらの開発工数も、旧来と比べて飛躍的に低減されることが期待できます。

2.4. ウェブ翻訳サービスならではの強み

ウェブ翻訳サービスはスタンドアローン翻訳ソフトの機能やインターフェースを焼き直すだけのものでもありません。遠からず、ウェブを「利用」した翻訳サービスだけでなく、ウェブを前提条件として「活用」した翻訳サービスが隆盛となるように予想しています。

例えば、語義曖昧性の解消法としてリコメンドシステムを活用するというような、集合知(CGM)を活用したサービスなどが考えられます。沖電気工業株式会社さんの「訳してねっと」などはその一つの好例ですが、辞書を「追加」という能動的な行動だけでなく、適訳の選択過程などを統計的な情報として収集・活用することで、入力機会の飛躍的な上昇を見込むことすら可能になるでしょう。無論、プライバシーへの配慮は必須となりますが。

また、サーバー側に翻訳エンジンのみを持ち、データ構造体を JSON や YAML などの汎用フォーマットでやり取りすることで、クライアント側には入出力部のみを備える構成を採用できます。そうすると、無理にクライアントをウェブブラウザに限定する必要もなくなります。各 OS のネイティブアプリであれば、既存のアプリケーション資産やノウハウを有効活用できます。興味深いインターフェースを備える Android や iPhone などの人気のプラットフォームへの対応も選択肢の一つとして現実味を増しますし、いわゆるガラパゴス携帯電話向けの Java アプリの開発も比較的容易になることでしょう。翻訳エンジンをトークン付き Web API を介して外販するというところですら考えられます。

そして、SaaS(クラウド)にも大いに期待できます。翻訳ソフトの基盤と部品を用意しておいて、特定のユーザー群に焦点を当ててカスタマイズした翻訳サービスを迅速に提供できるようにすると、特に企業ユーザーには喜ばれそうです。ユーザー群単位で辞書や翻訳メモリーを特段意識せずに共有できるだけでなく、日常業務と翻訳作業をシームレスに連携できるとしたら、素晴らしいですね。CRM(Customer Relationship Management)や SFA(Sales Force Automation)等の営業システムと翻訳システムが連携すると、外資系企業を顧客とした企業ユーザーに対する強烈的な訴求力を発揮できそうです。

AAMT へのご要望

3.1. エンドユーザー向けの存在感の拡大を望みます

AAMT に期待することは、上述の機械翻訳業界に向けた期待と同様に、機械翻訳のエンドユーザー向けの存在感を増して欲しいということです。上述したようなエンドユーザー向けの取り組みは、各社だけでは大変難しいものとお見受けしています。例えば UTX-Simple 仕様にしても、機械翻訳のメーカーさんに対応いただいたとしても、ユーザーがその機能を活用してくれないことには、盛り上がることは難しいでしょう。

また、事務局の多忙さは重々承知しておりますが、報道機関向けのプレスリリース(ニュースリリース)も出していただきたいです。例えば、毎年恒例の「日英自動翻訳サービスに関するアンケート」の母数の規模は、実施する側にとってはいささかの憂慮に値する事態であるように思えます。新聞・雑誌が採り上げてくれることは難しいかもしれませんが、IT 系ニュースサイトであれば、翻訳ソフトの利用者層とも親和性のあるメディアであるため、期待を抱けそうです。

3.2. 今後ともよろしくお願い申し上げます

機械翻訳の研究開発者や製造販売者だけでなく利用者也集うことが、本会の強みだと思います。AAMT を「あゝむ」と読むことすら知らずに「えーえーえむてーい」と読んで恥をかいた私ですが、機械翻訳へさらなる存在感をもたらす為に、そして世界の言語間に架かる橋をより長く・より広くする為に、私も微力を尽くしたいです。今後ともどうぞよろしくお願い申し上げます。

会員名・所属

安原宏(特定非営利活動法人セマンティック・コンピューティング研究開発機構)

/Yasuhara Hiroshi

自己紹介

AAMT の発足時から会員 機械翻訳との関係は、電子協の機械翻訳システム調査委員会、CICC 近隣諸国間機械翻訳システム開発プロジェクト、EDR 電子化辞書等。最近は、一般財団法人日本特許情報機構特許情報研究所の客員研究員として、特許文書に関する言語処理及び機械翻訳に興味がある。人間いつかは数ヶ国語をマスターしたいと思うが、そのような夢をかなえてくれる教育的で軽快な MT システムができないかと思っている。

機械翻訳および翻訳業界に期待すること

機械翻訳は、1980 年以降の官民挙げての本格的な研究開発が実施された。歴史的に見れば本協会の機械翻訳の歴史年表ページにもあるようにコンピュータが生まれると同時にスタートしている息の長いテーマである。ALPAC レポート以降、コンピュータの進歩によって機械翻訳への再挑戦が始まった頃、日本では 1978 年に電子協(現在の JEITA)に機械翻訳システム調査委員会(初代委員長は長尾真先生)が設けられて、1980 年に世界の機械翻訳研究を調査する企画に参加させていただいた。訪問先は、カナダ・モントリオール大学のトムメテオシステム、フランス・グルノーブル大学のヴォクア教授の MT 研究、ドイツ・ザールブリュッケン大学の MT 等であった。特に、モントリオール大学では 1976 年に天気予報の英仏翻訳の実用化に成功していた。サンプル翻訳：

“MOSTLY CLEAR AND COLD WITH PERIODS OF VERY LIGHT SNOW TODAY AND WEDNESDAY.” →
“AUJOURD HUI ET MERCREDI GENERALEMENT CLAIR ET FROID AVEC TRES FAIBLES CHUTES
DE NEIGE PASSAGERES.” 成功の要因は、天気予報は、限定的な語彙や文法で表現できるためにサブランゲージに適していたかもしれない。或は、英仏の言語的な近さかもしれない。いずれにしても、天気予報の MT という閉じた世界ではあるが、MT 実用化の確かな一歩であろう。訪問時、カナダの TAUM METEO チームは航空機のメンテナンスマニュアルの機械翻訳に移行していたがこれが成功したという話は聞かない。天気予報のコーパスから飛躍しすぎたのであろうか。

日本でも、ある種の分野限定的、あるいは用途限定的な機械翻訳システムが幾つか開発されてきた。京都大学の科学技術文献表題の英日翻訳や抄録の日英／英日の自動翻訳、NTT の市況速報日英、NHK の外電ニュース英日等々である。

このような閉じた世界、ドメイン限定の MT を積極的にサービスする方向で、その集大成が、一般的なテキスト機械翻訳と考えるのは邪道であろうか。そのためには、サブランゲージを構成する語彙と文法（或は、用例・統計データ、意味文法）の保守・拡張をどのようにして機械化するかが研究課題であろう。サブランゲージの拡張が連続的に可能な MT 技術が開発できればブレークスルーになるかもしれないという淡い期待を抱いている。例えば、スポーツ MT(ゴルフ、野球、相撲等)、料理 MT、旅行 MT、経済 MT、医療 MT、行政 MT 等々様々なサービスがあるが、これらも更に細分化しなければいけないであろう。一つ一つの発話単位に帰着する。しかし、分野が狭くても正確な翻訳が得られるのであれば利用者は増える。Web 環境が整えば、ドメイン対応 MT サービスの爆発の時代が来て、その背後で機械翻訳及び、その基礎となる自然言語の本質が明らかにされるのではないか。38 億年の歴史を持つ生命には多様性と普遍性が潜んでいる(中村桂子)。言語の多様性と普遍性が解き明かされる日を期待したい。

AAMT へのご要望

創立 20 周年を目指して、以下のような情報発信に着手されることを期待しています：
世界の機械翻訳システムポータルサイト、機械翻訳白書、オンライン機械翻訳百科事典、機械翻訳ハンドブック、
機械翻訳システム入門、世界の言語とアジアの言語、...

協会活動報告

(2009 年 11 月～2010 年 5 月)

理事会 (E-mail にて議論)

2010 年 3 月 24 日～3 月 31 日

- ①第 1 号議案 2009 年度事業報告 (案) ② 第 2 号議案 2009 年度決算報告 (案)
- ③第 3 号議案 2010 年度事業計画 (案) ④ 第 4 号議案 2010 年度収支予算 (案)

2010 年 4 月 7 日～4 月 14 日

AAMT と LISA (Localization Industry Standards Association) との提携についての承認

監事会

2010 年 4 月 22 日 平成 21 年度会計報告

講演会

2010 年 5 月 19 日

「Statistical Machine Translation: Challenges and Opportunities」

Franz Och 博士 (Google)

会場：東京大学 情報学環・福武ホール 福武ラーニングシアター

機械翻訳課題調査委員会

2009 年 11 月 10 日 (2009 年度 第 7 回)

- ①前回委員会の議事録の確認 ②各 WG の活動について (各 WG に分かれて議論)
- 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

2009 年 12 月 18 日 (2009 年度 第 8 回)

- ①今年度事業計画及び来年度活動計画について ②前回委員会の議事録の確認
- ③各 WG の活動について (各 WG に分かれて議論)
- 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

2010 年 1 月 28 日 (2009 年度 第 9 回)

- ①今年度事業計画及び来年度活動計画について ②前回委員会の議事録の確認
- ③各 WG の活動について (各 WG に分かれて議論)
- 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

2010 年 2 月 26 日 (2009 年度 第 10 回)

- ①今年度事業計画及び来年度活動計画について ②前回委員会の議事録の確認
- ③各 WG の活動について（各 WG に分かれて議論）
- 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

2010 年 3 月 26 日（2009 年度 第 11 回）

- ①今年度事業計画及び来年度活動計画について ②前回委員会の議事録の確認
- ③各 WG の活動について（各 WG に分かれて議論）
- 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

2010 年 4 月 26 日（2010 年度 第 1 回）

- ①前回委員会の議事録の確認 ②各 WG の活動について（各 WG に分かれて議論）
- 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

2009 年 11 月 13 日（2009 年度 第 6 回）

- ①前回議事録の確認 ②今年度の研究内容について
- ③次回の開催について

2009 年 11 月 27 日

AAMT/Japio 特許翻訳研究会シンポジウム

会場：総評会館 201 会議室（東京・御茶ノ水）

基調講演 江原暉将（山梨英和大学 教授）

* 講演予定の辻井潤一（東京大学 教授）欠席のため代理講演

招待講演「特許庁の特許情報提供における機械翻訳の活用と今後の展望」

岡崎輝雄（特許庁 総務部普及支援課特許情報企画室調査班長）

招待講演「中国における知財情勢と知財情報の活用」

後谷陽一（特許庁 特許審査第三部半導体機器審査長/前日本貿易振興機構（JETRO）北京センター 知的財産権部 部長）

AMT/Japio 特許翻訳研究会研究活動報告

江原暉将（山梨英和大学 教授）

横山晶一（山形大学 教授）

招待講演「中国における特許翻訳の現状」

浜口宗武（㈱知財翻訳研究所 代表取締役）

董 昭（㈱知財翻訳研究所 顧問中国弁理士）

招待講演「翻訳者を支援するサイト：みんなの翻訳」

内山将夫（情報通信研究機構 主任研究員）

影浦峯（東京大学 教授）

久保順子（津田塾大学 特任助教・筑波大学大学院）

2009 年 12 月 11 日（2009 年度 第 7 回）

- ①前回議事録の確認 ②今年度の研究内容について

③次回の開催について

2010 年 2 月 12 日 (2009 年度 第 8 回)

- ①前回議事録の確認 ②今年度の活動について ③研究報告
- ④平成 21 年度研究報告書作成関連 ⑤AAMT 事務局からの依頼事項について
- ⑥次回の開催について

2010 年 5 月 14 日 (2010 年度 第 1 回)

- ①前回議事録の確認 ②今年度の研究内容について
- ③次回の開催について

編集委員会

2009 年 12 月 AAMT Journal No.46 発行

ML にて最新号の進捗打合せ及び次回委員会の開催日調整。

2010 年 2 月 19 日 (2009 年度 第 3 回)

- ①AAMT Journal No.47 進捗確認 ②No.48 以降の企画について
- ③その他

2010 年 3 月 AAMT Journal 臨時選抜号発行

インターネット WG

- ①AAMT ホームページの更新 (毎月)
- ②AAMT Forum メイリングリストの管理
- ③AAMT Forum への情報発信 (月数回)
- ④委員会活動におけるメイリングリストの管理
- ⑤会員専用ホームページ開設にむけた準備
- ⑥その他事務局ネットワークのインフラ管理全般

＜事務局からのお知らせ＞

～AAMT 第三代会長辻井潤一先生紫綬褒章受賞のお知らせ～

AAMT 第三代会長であり、AAMT 理事を勤められております、辻井潤一先生（現・東京大学大学院情報学環・情報理工学系研究科教授）が、この度、平成 22 年度春の褒章において、紫綬褒章を受賞されました。

今回の受賞は、辻井先生の長年にわたる知能情報学の分野における業績と科学技術への多大な貢献が高く評価されたものであり、特に機械翻訳においては、先駆的な業績によって国際的に高い評価と研究成果を収めておられます。

心よりお慶び申し上げます。

事務局移転のご案内

会員各位

来る 7 月 1 日より、下記の住所に移転することになりましたので、お知らせいたします。

＜所在地＞

〒441-8580

愛知県豊橋市天伯町雲雀ヶ丘 1-1

国立大学法人 豊橋技術科学大学 情報メディア基盤センター内

＜新 TEL & FAX＞

0532-44-6620

＜移転日＞

2010 年 7 月 1 日より

どうぞよろしくお願い申し上げます。

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 47 号をお送りします。

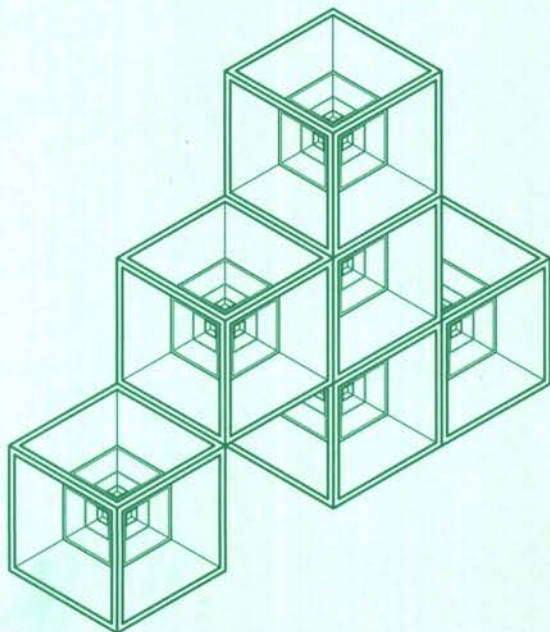
(元)鳥取大学の池原悟先生が昨年 12 月 15 日にお亡くなりになりました。今号では、2007 年 6 月発行の AAMT ジャーナル(No.40)に掲載されました、池原先生の研究報告「非線形言語モデルと重文複文の意味類型パターン化」を再録させていただくとともに、AAMT 現会長井佐原均先生よりご寄稿いただきました追悼文を掲載させていただきました。編集委員一同、池原悟先生のご冥福を心よりお祈りいたします。

今号では、特別企画として、新潟産業大学 蓮池薫先生より特別寄稿をいただきました。蓮池先生は、韓国語、朝鮮・韓国文化をご専門とされ、2008 年には文春新書より『蓮池流韓国語入門』をご刊行されるなど、多方面にわたり活躍されています。このたび、編集委員の一人が先生の講演会に参加させていただく機会があり、その折、先生に AAMT ジャーナルへのご寄稿を依頼したところ、ご快諾いただき、今号にご寄稿いただくこととなりました。

また、雑誌「日本語学」の 2009 年 10 月号の特集「機械翻訳の実際」において大変興味深い記事を発表しておられた麗澤大学 杉浦滋子先生、山梨英和大学 池田尚志先生より、機械翻訳の利用法および研究動向に関して貴重な寄稿をいただきました。

一方、AAMT 内の活動報告として、機械翻訳課題調査委員会より 3 件のご寄稿をいただきました。その他、「AAMT 会員のひろば」の企画におきましては、法人会員の紹介文 1 件、個人会員の紹介文 2 件を掲載しました。

AAMT



AAMTジャーナル No.47

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 言語翻訳グループ内

phone：0774-93-4625 fax：0774-93-4627

編集委員会：宇津呂 武仁 小谷 克則 大倉 清司

熊野 明 三浦 貢 村上 嘉陽

事務局：神崎 享子 喜田 あゆみ

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>