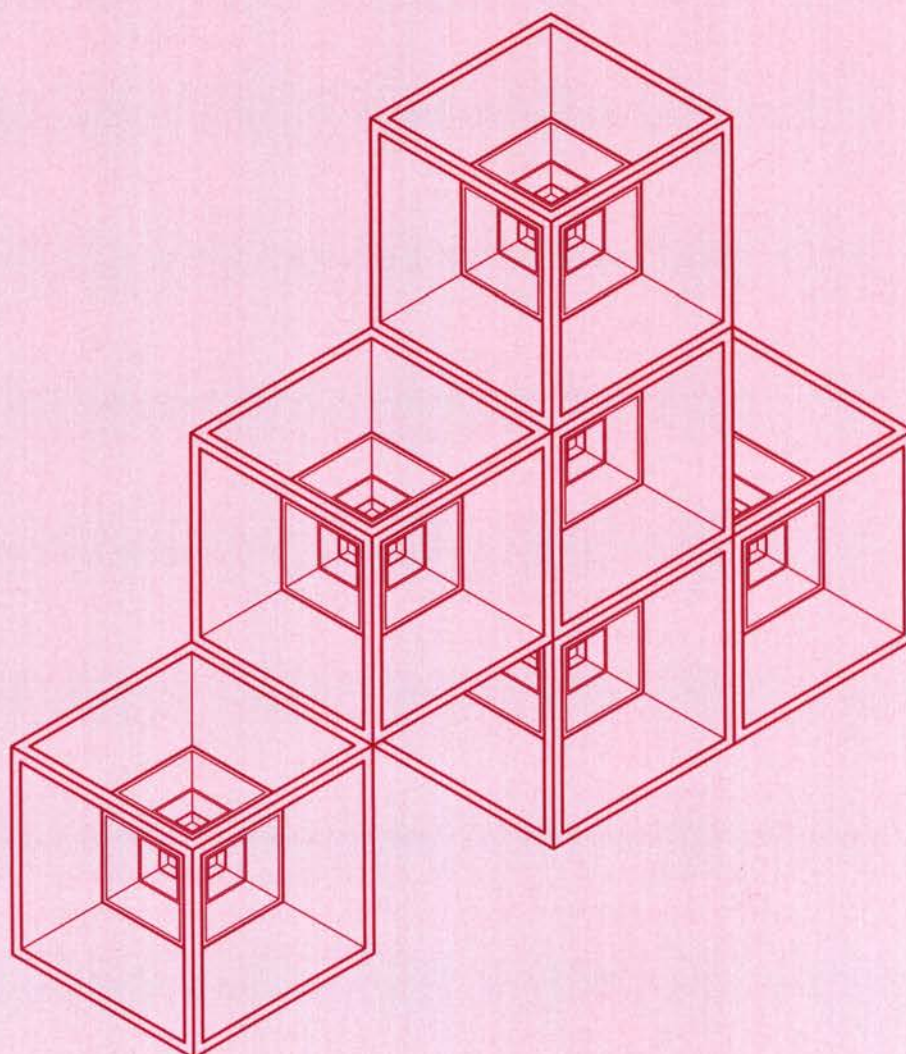


AAMT

Asia-Pacific Association for Machine Translation

Journal



June 2009 *No.45*

アジア太平洋機械翻訳協会

目 次

巻頭言：	特許技術と機械翻訳	守屋 敏道	1
レポート：	サン・マイクロシステムズでの機械翻訳実用化への取り組み	俣野 宏子	2
	翻訳品質の定量評価ツールによる納品検査の成功／挫折と現状	脇田 早紀子	4
	高度言語情報融合フォーラム(ALAGIN)について	鳥澤 健太郎、木俣 豊、隅田 英一郎、中村 哲	9
	みんなの翻訳	内山 将夫	11
	自由に翻訳！	エラリー ジャンクリストフ	15
プロジェクト：	人に理解しやすく 機械にも理解しやすい 産業日本語	渡邊 豊英、大塩 只明	17
連載企画(第二回)：	言葉と人材ーグローバル時代における企業内の言語とコミュニケーション	九門 崇	19
新サービス紹介：	中国・韓国特許検索システムの紹介	株式会社クロスランゲージ	21
新製品紹介：	「The 翻訳 [®] 2009 プレミアム」	東芝ソリューション (株)	26
委員会活動報告：	AAMT 機械翻訳文テストセット (第 1 版) 公開に向けて	AAMT 機械翻訳課題調査委員会 WG1	29
	英日・日英機械翻訳に関するアンケート結果報告	AAMT 機械翻訳課題調査委員会 WG2	31
	ユーザー辞書共通フォーマット UTX	AAMT 機械翻訳課題調査委員会 WG3	42
MT 関連リンク集：	機械翻訳研究プロジェクトリンク集	AAMT 機械翻訳課題調査委員会 WG2	45
	翻訳ソフト活用のためのリンク集	AAMT 機械翻訳課題調査委員会 WG2	48
会員紹介：	AAMT 会員のひろば (第 4 回)		50
活動報告：	協会活動報告 (2009 年 2 月～2009 年 5 月)		62
MT Summit XII：	MT Summit XII 開催のご案内	AAMT 事務局	64
編集後記：		宇津呂 武仁	68

CONTENTS

Foreword:	Patent Information and Machine Translation	<i>T. Moriya</i>	1
Report:	Where We Are with Machine Translation at Sun Microsystems.....	<i>H. Matano</i>	2
	Quantitative Assessment of Translation Quality Acceptance Check 'by Tool' vs. 'by Human'.....	<i>S. Wakita</i>	4
	ALAGIN- Advanced LAnGuage INformation Forum -	<i>K. Torisawa, Y. Kidawara, E. Sumita, S. Nakamura</i>	9
	Minna no Hon'yaku(Translation for everyone by everyone)	<i>M. Uchiyama</i>	11
	Translation, freely.....	<i>H. J-Christophe</i>	15
Project:	Technical Japanese – the language for human and computer	<i>T. Watanabe, T. Ohshio</i>	17
Serial lecture:	Language and HR – Corporate Language and Communication in the Global Era.....	<i>T. Kumon</i>	19
New Services:	WORLDPATENT.JP.....	<i>Cross Language Inc.</i>	21
New Product:	The Honyaku 2009 Premium.....	<i>Tbshiba Solutions Corp.</i>	26
Committee Report:	Toward Publication of AAMT Test-sets for Quality Evaluation of Machine Translation Sentences.....	<i>MT Committee for seeking future direction of MT</i>	29
	User unquiet for 2009	<i>MT Committee for seeking future direction of MT</i>	31
	UTX	<i>MT Committee for seeking future direction of MT</i>	42
MT Web Links:	Links of MT Projects	<i>MT Committee for seeking future direction of MT</i>	45
	Links of Translation Tips.....	<i>MT Committee for seeking future direction of MT</i>	48
AAMT Members:	AAMT Agora.....		50
AAMT Activities	AAMT Activities (February 2009～May 2009).....		62
MT Summit XII:	Announcement of MT SUMMIT XII.....		64
Editor's Note:	Message from the Chair of the Editorial Committee.....	<i>T. Utsuro</i>	68

目 次

巻頭言：	特許技術と機械翻訳	守屋 敏道..... 1
レポート：	サン・マイクロシステムズでの機械翻訳実用化への取り組み.....	俣野 宏子..... 2
	翻訳品質の定量評価ツールによる納品検査の成功／挫折と現状	脇田 早紀子... 4
	高度言語情報融合フォーラム(ALAGIN)について	
	 鳥澤 健太郎、木俣 豊、隅田 英一郎、中村 哲.....	9
	みんなの翻訳.....	内山 将夫..... 11
	自由に翻訳！	エラリー・ジャックリストフ... 15
プロジェクト：	人に理解しやすく 機械にも理解しやすい 産業日本語.....	渡邊 豊明、大塩 只明... 17
連載企画(第二回)：	言葉と人材ーグローバル時代における企業内の言語とコミュニケーション	九門 崇..... 19
新サービス紹介：	中国・韓国特許検索システムの紹介	株式会社クロスランゲージ... 21
新製品紹介：	「The 翻訳 2009 プレミアム」	東芝ソリューション (株) ... 26
委員会活動報告：	AAMT 機械翻訳文テストセット (第 1 版) 公開に向けて.....	
	 AAMT 機械翻訳課題調査委員会 WG1..	29
	英日・日英機械翻訳に関するアンケート結果報告	AAMT 機械翻訳課題調査委員会 WG2.. 31
	ユーザー辞書共通フォーマット UTX.....	AAMT 機械翻訳課題調査委員会 WG3.. 42
MT 関連リンク集：	機械翻訳研究プロジェクトリンク集.....	AAMT 機械翻訳課題調査委員会 WG2.. 45
	翻訳ソフト活用のためのリンク集	AAMT 機械翻訳課題調査委員会 WG2.. 48
会員紹介：	AAMT 会員のひろば (第 4 回)	50
活動報告：	協会活動報告 (2009 年 2 月～2009 年 5 月).....	62
MT Summit XII：	MT Summit XII 開催のご案内	AAMT 事務局 .. 64
編集後記：		宇津呂 武仁... 68

CONTENTS

Foreword:	Patent Information and Machine Translation	T. Moriya 1
Report:	Where We Are with Machine Translation at Sun Microsystems	H. Matano..... 2
	Quantitative Assessment of Translation Quality Acceptance Check 'by Tbol' vs. 'by Human'	
	 S. Wakita.....	4
	ALAGIN- Advanced LAnGuage INformation Forum -	
	 K. Tbrisawa, Y. Kidawara, E. Sumita, S. Nakamura	9
	Minna no Hon'yaku(Translation for everyone by everyone).....	M. Uchiyama..... 11
	Translation, freely.....	H. J-Christophe .. 15
Project:	Technical Japanese – the language for human and computer.....	T. Watanabe, T. Ohshio... 17
Serial lecture:	Language and HR - Corporate Language and Communication in the Global Era...	T. Kumon ... 19
New Services:	WORLDPATENT.JP.....	Cross Language Inc..... 21
New Product:	The Honyaku 2009 Premium	Tbshiba Solutions Corp... 26
Committee Report:	Toward Publication of AAMT Test-sets for Quality Evaluation of Machine Translation	
	Sentences.....	MT Committee for seeking future direction of MT..... 29
	User unquiet for 2009	MT Committee for seeking future direction of MT..... 31
	UTX	MT Committee for seeking future direction of MT..... 42
MT Web Links:	Links of MT Projects	MT Committee for seeking future direction of MT..... 45
	Links of Translation Tips	MT Committee for seeking future direction of MT..... 48
AAMT Members:	AAMT Agora	50
AAMT Activities	AAMT Activities (February 2009～May 2009)	62
MT Summit XII:	Announcement of MT SUMMIT XII	64
Editor's Note:	Message from the Chair of the Editorial Committee	T. Utsuro. .. 68

特許情報と機械翻訳

財団法人日本特許情報機構 専務理事
兼 特許情報研究所長 守屋 敏道

特許は、新たに開発された斬新な技術（発明）を他人による無断使用から守るために、出願者に与えられる排他的独占権です。研究開発の成果である発明について特許を取得するには、出願者は、発明の詳細を説明する文書（明細書）と、特許を求める発明を定義する文書（特許請求の範囲）などの所定の書類を整えて各国特許庁にその国の公用語で特許出願をし、特許審査を受けなければなりません。これらの文書の内容が特許情報の重要な部分を構成します。特許情報は、発明技術の他人による無断使用を禁止する範囲を示す権利情報の性格と新しい技術を公開する技術情報の性格とを持っています。技術開発に携わる研究者や国内外でビジネスを展開する企業にとって、自らの研究の先進性や事業の競争力を見極めるため、また特許侵害を避けるために、特許情報は大変重要な情報です。特許審査における先行技術情報として、特許情報の利用が欠かせないのは言うまでもありません。

世界知的所有権機関(WIPO)の 2008 年統計によれば、2006 年一年間の世界の特許出願総数は実に 176 万件もあります。その内訳を見ますと、各国国内へのオリジナル出願の総数が 106 万件であり、残り 70 万件が国境を越えて重複して出願されているものです。2006 年の日本人のオリジナル出願が 35 万件で、外国から日本への出願が 6 万件ですから、65 万件のオリジナル出願は日本語以外の言語で公表されています。日本語の出願書類を英語等の各国公用語に翻訳して日本から海外に特許出願するものは約 16 万件です。

経済のグローバル化が進み、技術開発の国際競争が激しくなっている状況下、研究開発投資の増大を反映して、世界全体として特許出願は増加傾向にあります。海外への特許出願や他国特許情報を利用するための翻訳に掛かるコストは膨大なものとなっています。

特許庁に提出する特許出願書類は特許権利書の原本になるため、その翻訳を機械翻訳のみに任せることは極めて難しいですし、適切でもありません。翻訳を人が最終的に精査することが必要です。しかし、翻訳精度の高い機械翻訳があれば、特許出願書類の翻訳負担はかなり軽減されと考えられます。

Japio は、特許分野での機械翻訳の利用拡大と精度向上を目指して、AAMT のご協力を得て、2003 年度から AAMT/Japio 特許翻訳研究会を発足させ、研究活動に参加しております。

Japio は長年、翻訳事業において機械翻訳を実際に利用しておりますので、機械翻訳結果に触れる機会が多くあります。機械翻訳技術が大きな進歩を遂げていると称賛する一方、機械翻訳の一層の精度向上には、翻訳用辞書の充実とともに、人と機械との協働作業（コラボレーション）が不可欠ではないかと感じております。機械翻訳利用者のストレスを少なくし、人と機械のコラボレーションを進める観点から、今後の機械翻訳研究で次のようなことが実現できることを期待しております。

- (1) 機械翻訳システムの能力では手に負えないほどに非明晰な日本語文は、翻訳を拒否され、修正を勧告されるか、望ましくは、機械翻訳可能な明晰な文章が機械翻訳システム側から提案されること
- (2) 機械が機械翻訳をした場合には、機械自身が日本語原文をどのように理解したかが分かるように、その理解内容を示す日本語文が表示されること
- (3) 機械翻訳での翻訳文は、文法に従った文章でアウトプットされること
- (4) 機械翻訳しやすい日本語文を書く意識とその書き方が広く普及すること

Japio は、機械翻訳の一層の実業利用を目指して、機械翻訳研究に携わる皆様と共に活動を続けたいと考えております。

サン・マイクロシステムズでの機械翻訳実用化への取り組み

サン・マイクロシステムズ株式会社

俣野 宏子

1. はじめに

サン・マイクロシステムズでは、数年来、製品をローカライズする際の機械翻訳技術の最適な使用方法について評価検証を行い、実用化に向けて取り組んできました。現在、いくつかの言語では、実用レベルに達したと考えています。

本稿では、当社でのこれまでの機械翻訳への取り組みと現状、および今後の展望についてご紹介します。

2. これまでの取り組み

2.1. 背景

近年、グローバリゼーションの進展に伴い、英語以外の言語にローカライズされた製品の需要は一層増大し、大きなビジネスチャンスをもたらしています。

一方、wiki やブログ、動画などの新しいメディアが台頭し、日々膨大なコンテンツが世界中で量産されています。多くのコンテンツは翻訳しきれていないため、有望なビジネスチャンスを逃している可能性があります。

その中で当社は、ローカライズ製品の製造コストを削減し、製品のリリースサイクルを短縮し、新しいメディアに柔軟に対応しつつ、お客様に迅速にコンテンツを提供するための最も有望かつ実用化可能な技術の1つとして機械翻訳技術に着目し、下記のプロセスで実用化に向けて取り組み始めました。

- 機械翻訳技術を当社の翻訳プロセスに実装可能かどうか検証
- 機械翻訳エンジンを決定
- フィージビリティスタディ

2.2. 検証

近年、SMT、RBMT、ハイブリッド型などのさまざまな機械翻訳エンジンの多言語での研究開発が進み、実際の製品として使用できるようになりました。

そこで当社では、プロジェクトチームを結成し、下記の観点から、機械翻訳技術を当社の翻訳プロセスに実装可能かどうか各種検証を行いました。

- 機械翻訳技術のベストプラクティス是何か。
それは当社に適用可能か
- 製品のローカライズコストを削減し、翻訳作業効率を向上しつつ、現在と同レベルの翻訳品質に管理できるか
- 従来ローカライズ対象外だった製品や言語にも拡張可能か
- 現在使用している翻訳メモリー、用語集ツール、翻訳管理ツールと組み合わせで自動化できるか
- 十分なROIを得られるか
検証の結果、下記の翻訳プロセスで、機械翻訳後にポストエディットし使用することに決定しました。
- 従来から製品ドキュメントを翻訳している製品群
- 従来翻訳していなかったポータルサイトやオンラインコンテンツ群

2.3. 評価方針

プロジェクトチームでは、下記の方針で言語ごとに最適な機械翻訳技術の定量評価を行い、機械翻訳エンジンを決定することにしました。

- 翻訳量の多い言語で利用できるか（例：日本語、簡体中国語、フランス語、スペイン語）

- その言語でベストの機械翻訳エンジンは何か
 - SMT vs RBMT
 - ポストエディットしやすいか
- 最適な翻訳対象コンテンツは何か（例：テクニカルドキュメント、マーケティング資料、ユーザサポート用コンテンツ、wiki、ブログ）
- フェーズに分けて、言語・製品ごとに順次実用化できるか

2.4. フィービリティスタディ

実際の製品ドキュメントから生成した英文と翻訳対象言語のペアで構成されたテストセットを使い、言語ごとに複数のレビューアが下記3項目の評価を行いました。

- オリジナル（ポストエディット前）の機械翻訳の品質
- ポストエディット後の翻訳品質
- ポストエディットを含む翻訳作業の効率

翻訳品質の評価は、セグメントごとに読みやすさを数値基準で評価し、問題がある場合は、その理由をカテゴリから選択します。

翻訳作業の効率評価で使用したテストセットには、機械翻訳、翻訳メモリーからの再利用、英文（翻訳なし）が混在しています。レビューアが翻訳もしくは編集を行い、次のセグメントに移動するまでの時間（秒）をセグメントごとに計測し、単語またはカテゴリごとの所要時間と編集距離から作業効率を算出し、定量的に評価しました。

これらの評価結果から、言語ごとに機械翻訳エンジンを決定し、適用可能な製品群から順次使用することにしました。

3. 現状（フェーズ1）

本年5月にリリースされた最新の Solaris 製品のフランス語とスペイン語のテクニカルドキュメント（1）と、システム管理者向けポータルサイト（2）の5言語の翻訳プロセスで機械翻訳を使用しはじめています。

フェーズ1では、機械翻訳の使用により翻訳作業効率が約20%以上向上しました。その結果、翻訳コストを削減できました。

機械翻訳を使用しても、翻訳以外のステップ（たとえばレビュー）に与える影響は限定的で、翻訳プロジェクト全体は従来通り、もしくはそれ以上のスピードをキープできました。

また、翻訳の品質も実用に十分なレベルを保証できました。

一方、機械翻訳エンジンを、従来から使用している翻訳ワークフローシステムに統合したため、機械翻訳を特に意識せずに利用できます。

以上のことから、プロジェクトマネージャの目から見ると、機械翻訳を使ったプロジェクトでも、コスト以外は、従来の翻訳プロジェクトと同じように見えます。

4. まとめ

当社では、言語ごとに異なる機械翻訳エンジン、評価方法、カスタマイズ、周辺コンポーネントを統合した現在の機械翻訳プロセスは、全体として十分実用レベルに達していると考えています。

今後は、他の言語や製品群にも順次適用範囲を広げていく予定です。

当社では翻訳対象の99%は英語から英語以外の言語ですが、英語以外の言語間での機械翻訳の実用性についても現在評価中です。

5. 関連資料

- (1) <http://docs.sun.com>
- (2) <http://www.sun.com/bigadmin/hubs/multilingual>

翻訳品質の定量評価 - ツールによる納品検査の成功/挫折と現状

日本アイ・ビー・エム株式会社

脇田 早紀子

筆者は、IBM製品を日本語化(マニュアルや画面を英語から日本語に翻訳)する部門に所属している。翻訳作業自体は翻訳会社に依頼して行うが、その品質を日々確認し、向上を図るのが筆者の仕事である。

品質について説得力を持って語るために、翻訳に携わる、立場の違う多数の人間が納得できる形で品質を定量的に表すことが強く求められてきた。本稿では、当部門で使用している品質の定量化の仕組みについて述べるとともに、品質評価のプロセスの移り変わり、特にツールによる評価と人による評価の比較について述べる。

以下、「評価」と「納品検査」を似た意味で使用しているが、細かくいえば、翻訳会社からの翻訳納品物を「評価」して、問題のあるものについては見直しと修正を依頼する営みが「納品検査」である。また、「納品検査」の結果はそれぞれフィードバックとして送付し翻訳会社の品質向上の材料としてもらうとともに、集計してグラフ化し、品質状況の把握に役立てている。

1. 品質の定量化、基本の考え方

仮に、翻訳物を詳細にレビューして、「誤り」を漏れなく指摘したとする。この「誤り」の「密度」を100点満点から引いて、これを品質の評価値としよう。

このシンプルな、そして誰でも思いつくアイデアが、私たちが便宜的に使用している評価システムの根幹となっている。品質というものが、実際には誤りの多寡だけで測ることができない多面的なものであることはもとより承知している。しかし現状は、多すぎる誤りの海に日々苦戦しているところであり、とりあえずこれが減ってくれたらうれしいので、正確性に偏ったこの評価を日常的に用いている。以下、そのつもりでお読みいただきたい。

ここで「誤り」と呼んでいるのは、次の表のようなものである。

表：誤りの分類と重要度

重要度	誤り分類
大	誤訳(意味が原文と異なる)
大	訳飛び(原文にある内容が邦文では抜けていて意味が伝わらない)
中	製品名、資料名、資料番号、特記・商標などが定められたものと異なる
中	訳漏れ(原文にある内容が邦文では抜けているが軽微)
中	不正確(複数の意味にとれるなどあいまいな)
中	キーワードのタイプミス(重要語句のタイプミス)
中	ヘルプまたはドキュメントと画面訳の不整合
小	タイプミス(意味は通じる程度)
小	変換ミス(意味は通じる程度)
小	用語(不統一または指定表記に沿わない)
小	訳重複(コピーミスなどによる重複)
小	英語残り(消すべき英語が残っている。「a」「the」「」など)
小	原典ミス(資料上つじつまが合わない)
微	体裁(半角スペースの有無、個条書きリスト中の不揃い、括弧や句読点の違い)
微	言い回し(日本語として不自然)
微	漢字(漢字/仮名の使い分けや送り仮名がルールに沿わない、表外字、表外読み)

上表の「誤り」の出現数をただ数えてしまうと、感覚的な「誤りの量」とまったく合わないことになる。例えば、まったく意味が違っている誤訳はかなり困るのに比べて、用語不統一の困り方は「少し」と考えられる。そこで、上記の表にある重要度に従って「大」は5点、「中」は3点、「小」は1点、「微」は0.1点として重み付けして合計したものを「誤り」のポイントとする。さらに、そのポイントを10000ワードあたりに直したものを100点から引いて評価値とする。以上をまとめると、次のような式になる。

$$\text{品質評価} = 100 - (「大」 \times 5 + 「中」 \times 3 + 「小」 \times 1 + 「微」 \times 0.1) \times 10000 \div \text{ワード数}$$

品質の表し方に、ただひとつの正解があるわけではない。上記の分類や式は、試行錯誤を繰り返しつつ、作業しやすく、かつ、感覚に合う評価値が出るように微調整を繰り返してできたものである。だから、ひとつの現実的な近似の仕方といってもいい。

分類作業をしてみるとすぐにぶつかる問題が、分類の難しさ、不安定さである。これは「誤訳」にあたるものか、「不正確」程度でいいのか、評価者によって違ってしまふことはよくあるし、自分が昨日つけた分類に違和感があることさえよくある。まさに「分類はお天気次第」。

完璧な分類ができるようになってきているわけではないが、なるべく安定性を持たせる工夫はする。各分類と例文、考え方をまとめたガイドを作成して評価者の間で共有することはもちろん、お互いに「これは『タイプミス』ではあるけれど、意味がまったく違ってから『誤訳』じゃない?」というようなフィードバックをまめに行い、すり合わせを続けていくことは欠かせない。

また、単純な問題のようだけれどやっかいなのが、「誤り」の繰り返しである。あるひとつの概念が、望ましくない用語で表現されていたとして、それが資料に100回出現していたらどうだろうか。これをそのまま100個と数えてしまったら、ほかの誤訳やタイプミスの問題は消し飛んでしまう。一方、繰り返しは全部合わせて1個と数えるというのも極端な話で、たくさん出てくる用語が間違っていれば、少ししか出てこない用語が間違っているより影響が大きいと感じるのは当然だ。結局、出現頻度が25ページ当たり 3 ～ 10 回程度ならダブルカウント、それを超える場合はトリプルカウントというルールにしているが、これが理屈の上から正しいのかどうかわからない。なんとなく、人間の感覚はlog関数的かなと思うので(視覚聴覚などをお手本にすると)、だいたいこんな感じでいいような気がする。

「気がする」という言葉を論文で使う人はあまりいないと思うが、ここではけっこう大切なことなのだ。

安定した分類も評価も、理屈に合うことと感覚に合うこと、この両面をバランスよく取り入れていかなければ説得力のある評価にはならない。

さて、このような考え方に基づいて、曲がりなりにも評価の作業を始めたのが2002年。このときの活動は、評価者、評価対象ボリューム、作業時間を固定にすることによって評価の安定を図ろうとしたもの。確かに、ある月に評価した10件の資料の中で、どの資料の品質が良く、どれが悪いといった相対的な把握はできるようになり、それだけでも画期的なことではあったのだが、一月に評価して80点だったものと、九月に評価して80点だったものの品質が同じかということ、これは非常に心もとない状況だった。

明らかに、一月の評価より九月の評価のほうが甘くなっているのである。これは、評価者個人がいくらまじめにやっているつもりでも、単調な作業で適切なフィードバックもなく、慣れてくればいわゆる「流れてしまう」面がどうしても出てくるということで、今から考えればすごくあたりまえなことなのだが、そのころはよくわかっていなかった。このあと、評価者によるばらつきを見る実験や、そもそも評価者の選定、育成などの課題に否が応でも分け入っていくことになるが、ひとまず2002年、2003年前半くらいの成果としては、相対的な品質の良し悪しがわかるようになったこと、そして、人による評価の難しさがわかったことが挙げられる。

2. ツールによる納品検査、夢の実現

人による評価は、その評価者の能力や状態によって大きく左右されるし、時間もお金もかかる。これが、翻訳物をちょいちょいと機械に通して、「78点です」なんて出てきたら、なんて幸せなことだろうか。

無茶な上司がいるもので、そういう機械を作ってよ、というリクエストがあった。そんなこと、できるわけがないじゃないですかといっていたんはあしらい、でもなんとかなるかもと思い直したのは2004年前半のことだった。

2002年からの評価活動の中で、いろいろな誤訳が蓄積でき、それをつらつらと眺めてみると、結局のところかなり単純な取り違いによる誤訳が多いことがわかった。「deployment」と「development」、「detect」と「delete」のような見た目が似ている語、あるいは、見た目は似ていなくても、「server」と「client」のような対語は脳みそのなかで近い部分にあるらしく、間違っていることがけっこうある。こんな間違いだったらツールで見つかるのではないだろうか? と思いついて、ツールを作ってみたのが2003年。

ツールの仕組みはいたって簡単である。翻訳者用の技術用語対訳集は、いつも使っているものがすでにあるので、「server」の訳語は「サーバー」、「client」の訳語は「クライアント」といった調子で何万語も登録されている。一方、翻訳はTM/Winという翻訳専用エディターで行っているの、セグメントと呼ばれる「文」のような単位で、英日がペアになって保存されている。まず原文のほうを文字列検索して、対訳集に載っている語を探す。例えば「server」があったとすれば、今度は訳文のほうに「サーバー」があるかどうか探してみ、あればよし、なければ少し怪しいというわけだ。(実際には、ひとつの英語に複数の訳語が登録されていることが多いので、その訳語のどれかがあればよいという探し方をする)

文字列検索以外の解析はしていないので、朝思いついて午前中にperlでツールを作り、午後はテストができたくらいの単純なものだ。それでも非常に強力で、これまでうっかりそのまま出していた大量の誤訳を明らかにすることができた。誤訳だけでなく、訳漏れの場合もあり、タイプミスであるために検索できなかったという場合もある。それらはいずれも間違いだから修正しなければならないが、省略や意識により訳語が見つからないということも多い。だからツールが警告してきた個所については、その後に人がチェックしなければならないが、それにしても目の子でチェックしているのに比べたら超速で確認できる¹⁾。このツールは翻訳会社にリリースし、ツールでわかる誤訳などを

つぶしてから納品してもらうようにした(2003年後半)。誤訳チェッカーで見つかる「誤り」の例を以下に示す。

[見た目似ている] conversion/conversation, now/no, procedure/processor, filler/filter, image/main, initiate/initialize, detect/delete, data/date, program/problem ellipses/ellipse, build/bind, service/server, description/destination
[対語など] offline/online, first-level/second-level, server/client, reactivate/inactivate, caller/callee, valid/invalid, descending/ascending
[英数字] DVHPROFA/DVHPROFM (コマンド), CWVRR5000E/CWVRR5001W (エラー番号)
[否定/肯定]
- The server has been renamed.
サーバーは名前変更されていません。
- OMPROUTE does pass tags created by others.
OMPROUTE は、上記以外で作成されたタグは渡しません。
- The lock file must not reside on the DOSRES or SYSWK1 SCSI disks.
ロック・ファイルは、DOSRES または SYSWK1 SCSI ディスクになければなりません。
- Incoming packets that had unknown protocols
踏めプロトコルを持っていた着信パケット

図：誤訳チェッカーの検出例

いうまでもなく、このツールによって検出できる「誤り」は全体の「誤り」のごく一部である。それでも、その一部の「誤り」が多いということは、やはり全体にいい加減な翻訳をされた資料である可能性が高いということではないだろうか。そう考えて、このツールで検出できる「誤り」を使って評価した結果と、人がレビューしてわかった「誤り」を使って評価した結果を比較する実験を行った³⁾。

その結果、高い相関があることがわかり、ツールによる品質定量評価への道が開けた。2004年後半からは、翻訳会社からの納品物を全数検査し、品質の観測をしてグラフが描けるという、夢の納品検査プロセスが実現したのだ。

ここで、注意深い読者はすでにお気づきかもしれないが、このストーリーには大きな欠陥がある。「相関が高い」といってもそれはあくまで、ツールのかかってないまっさらの翻訳物に対してツールで検査を行った場合であって、ツールを翻訳会社に開示してしまっている以上、検査前に既に、ツールで見つかる「誤り」

については出がらし状態になっているはずである。

そうなのだ。ツールによる検査を思いついたとき、すでにツールを開示してしまったことについて後悔しきりだった。開示してしまったツールを使って品質の観測をすることはできない。

しかし、この思いつきは、ただ捨てるにはあまりにももったいないほど魅力的だった。開示済みのツールでは見つからない、重大な誤訳を検出する新アイデア(否定語のチェック)もひねり出したが、これだけでは検査に使えるほどの出現頻度にはならないようだった。そこで私は、姑息な方法を思いついた。

開示済みのツールから、ハッシュ値をつかってランダムに2割の警告を「ぶっこぬいた」ツールを作り、これを「バージョンアップ」と称して翻訳会社にリリースする。「自称」バージョンアップ版には、対訳集の変更なども盛り込まれているから、翻訳会社としては当然新しいバージョンを使用する。そしてこちらは、抜いた2割と、新機能である否定語のチェック部分を利用して、品質の観測を行う。

このような苦渋紛れの手で、2004年後半から一年ちょっとならば、まさに品質チームの夢の実現という状況であった²⁾。専任の検査者が一人と、私がワークロードの半分を割くくらいの体勢で、納品物の全数検査が行えて、大量の「誤り」の修正と品質の観測を同時に行うことができたのだ。

このプロセスで観測した品質は、人による評価と異なり、体調や気分やスキルに左右されにくい。月による品質の変動も把握できたし、会社ごとの品質の特徴もわかってきた。例えば、ある会社は翻訳ボリュームが一定限度を超えると急に品質が落ちるとか、またある会社はコーディネーターの体勢が変わって品質ガタ落ちになったものの、少しずつ持ち直しているとかそういうことである。

3. ツールによる納品検査、挫折

この幸せが長く続かないことは、スタート時の無理やりな方法からしても、明らかなことだった。もっとも、仮にツールを開発した当初からそれを秘匿し、検

査プロセスを実現していたとしても、あまり本質的な違いはなかったかもしれない。なにしろ、ツール自体がとても単純な、底の浅いものであるから、人間様の学習能力にかかったらひとたまりもないのである。

各翻訳会社は地道に品質改善に取り組み、当然のことながら、納品検査でもらうフィードバックの蓄積と解析に努めた。その結果、納品検査でわかる「誤り」は非常に偏っており、何かツールっぽい...ということとはほどなくわかってしまう。よく指摘される事例を集めて、独自ツールの開発をしようとする会社もあったし、ある会社に至っては、旧バージョンのツールを利用すると、検査で指摘される個所のいくつかが出てくることまで突き止めた。

こうなってくると、ツールがごく一部の「誤り」しか見つけられないということがとても痛い。翻訳会社がどうしても人情として「検査対策」に走ってしまうので、あるべき品質改善の取り組みから比べるとどうしても歪みが出てしまうのだ。結局、プロセス開始から二年たったころには、この検査が利くのは新参の翻訳会社だけ、という状態になり、観測の役を果たさなくなってしまった。

4. 人による納品検査

このように、ツールによる納品検査は中断せざるを得なくなったが、このころには、品質の観測ができるという状態にすっかり馴染んでしまい、この快適と安心を手放すことはできなくなってしまっていた。そこで、ものすごく手間がかかることはわかっていながら、人が読むことで納品検査をするプロセスを開始したが、2007年だった。原始的な方法に逆戻りしたとマイナスに捉えることもできるが、ツールによる検査プロセスを踏み台にすることで、品質観測の本来の形に進化したとプラスに捉えることもできる。ツールによる検査をやってみていなかったら、とても人手による大規模な納品検査には踏み切れなかっただろうと思うからだ。

ツールによる検査は、品質の捉え方を歪ませた面もあったが、とにかく単純な取り違い誤訳とタイプミスを駆逐する効果は絶大だった。今度は、抜いて隠した2割も戻し、否定語チェックも追加してフルバージョ

ンのツールをリリース。翻訳会社には、そのツールをかけてわかる「誤り」をつぶしてから納品するように依頼し、そこからの再スタートである。

品質チームのレビューアーは増員し、しかも検査対象は減っている。全数検査を行っていた前プロセスと比較すると、対象資料は全体の3割程度、しかも各対象資料の2000ワード分をサンプルしてチェックということになり、その点から見ると力弱い。

とはいえ、検出できる「誤り」のバラエティーやきめ細かさはやはりツール検査とは比較にならないので、結果として得られる品質の観測データの質としては、同等かやや良いくらいのものと感じている。特に、あるひとつの資料について「良い」か「悪い」かを判断する材料としては、格段に信頼がおけるようになった。

結局のところ、品質の観測という点でいえば、数で稼いでもいいし、深さで稼いでもいいのだ。ある程度のバランスは必要だけれども。

5. 評価精度向上のために

人による評価の精度は、ひとえにその人材の質にかかっている。ザルな評価者にかかれば、どんな資料も高品質との評価が得られる。

ある評価者が挙げた「誤り」を確認し、その指摘が正しいかどうかを判断するのは比較的やすい。一方、挙げなかった「誤り」があるかどうか確認するのは...結局、全部読んでみるしかない。というより、読んでみたとして、今度は自分の評価精度に自信が持てるかという話になってしまう。

人による評価について、実験を繰り返すうちに、いやおうなしに辿り着く結論は、「漏れの無いレビューというのは幻想だ」ということである。二人の評価者が同じ資料の同じ箇所をレビューして、見つけた「誤り」のポイントが倍半分違うなどということは珍しくないし、ポイントとしてはだいたい合っている場合でも、挙げている「誤り」の内容があまり重なっていないなどということは日常茶飯事だ。いったい「誤り」は全部でどのくらい眠っているのだろうか？

まず、評価者の選定の問題として、いつも「誤り」

ポイントがとて低くなってしまふような(ザルな)評価者では困るし、あるときはとても詳しく、あるときはとてもあつさりしているような、ムラのある評価者でも困る。

また、評価者の育成の問題として、自分が見つけた「誤り」のほかにどれくらい「誤り」があったのかを知ることが重要であるし、そもそも自分の仕事の質が見られているという緊張感がやはりどうしても必要だ。

そこで、二人の評価者が同じ資料の同じ箇所を、同じ時間をかけてレビューするという「ペア評価」の仕組みを重視している。ローテーションで組み合わせを次々変えて行う「ペア評価」を軸に評価者の選定と育成を行うほか、現在では評価者によるばらつきを軽減するための補正值算出にも使用している。

評価者としては是非、しっかり誤訳などを見つける眼力のある実力者が望ましく、そういった人材は常に払底しているので、品質チーム内の担当者のほかに、翻訳会社で活躍中の、優秀な翻訳者さん・校閲者さんに依頼することもある。もちろん、その翻訳会社の翻訳物の納品検査をしてもらうことはできないので、他者の翻訳物の検査を依頼する。翻訳者さんとしては、普段なかなか得られない、自分の仕事に対するフィードバックを得られる機会でもあり、翻訳の仕事の山谷を埋めることができるのでおむね歓迎のようである。

これからも、人材育成と品質向上と効率化、二兎も三兎も追いながら、バランスよく進めていきたい。品質を測定する営みは、たいへんコストのかかるものなので、あの手この手で元をとっていくのが正しい態度だと思っている。

参考文献

- 1) 誤訳/訳漏れチェッカー・翻訳者向け対訳集の利用
脇田早紀子, 松山香子(第67回情報処理学会全国大会,2005)
- 2) 誤訳/訳漏れチェッカーの利用・翻訳納品物の受け入れ検査プロセス
脇田早紀子, 前川伸子(第68回情報処理学会全国大会,2006)
- 3) 翻訳品質定量評価作業への校正支援ツール導入効果
脇田早紀子(2004年前期IBM大和技術論文)

高度言語情報融合フォーラム(ALAGIN)について

鳥澤 健太郎[†], 木俣 豊[‡], 隅田 英一郎[†], 中村 哲[†][†]MASTAR プロジェクト (独)情報通信研究機構, [‡]知識処理グループ (独)情報通信研究機構

1. 背景: 情報爆発と Big Science としての音声言語研究

本稿では本年3月に発足した高度言語情報融合フォーラムについて紹介するが、フォーラム自体の説明に入る前に、昨今の音声・言語処理技術を取り巻く環境について概観したい。近年、Web 上におけるいわゆる情報爆発を背景として、音声・言語処理技術にはかつてないほどの注目が集まっている。実際に Web 上の膨大なテキストに、いわゆる corpus-driven な方法論を適用して、実社会にインパクトを与えうる技術が実際に開発されつつある。

また、爆発している Web はコーパスの供給源、つまり、ブレークスルーの enabler として機能するだけではなく、新たな社会問題ももたらしている。例えば、情報の信頼性、違法有害情報、検索におけるロングテールの存在といった諸問題である。社会の進化がクローズアップさせる類似の問題として、言語の壁もある。社会のグローバル化の進展で、資本、物品の国境を跨いでの移動はかつてないスケールに達している。しかしながら、情報の移動は、いわゆる言語の壁に阻まれて資本、物品の流通に釣り合う適切なペースで生じていないのではないだろうか？

例えば、いわゆるリーマンショックは、国境を越えた資本のネットワークによって世界中に大きな衝撃を伝搬させた訳だが、一体どれだけの日本人がその引き金であったサブプライムローンの存在をショック以前に認識していただろうか？ 後知恵ではあるが、言語の壁がなければより大勢の人間が、問題の可能性を認識できたのではないかとも思われる。私見では、このような例は、グローバル化の進展にともない、言語の壁がもたらす諸問題がより深刻になることを示唆するものである。こうした問題に対処するにあたって、音声・言語処理技術が

大きな役割を果たすことは間違いなく、そうした技術の研究開発の加速が望まれるところであろう。

また、ここでもう一つ重要なことは、研究開発が corpus-driven な方法論にシフトしつつあるため、計算の規模が飛躍的に大規模化し、多数の計算機を結合したクラスタの利用が必須となりつつある、ということである。近年、いわゆるテキストからの知識獲得の研究が脚光を浴びているが、実際に Web 中の数億文書に構文解析を適用し、種々の知識獲得手法を適用する研究はクラスタ無しではほぼ不可能である。また、研究で利用される注釈付きコーパスもより大規模になる傾向があることから、研究にあたってはより大規模なグループによってそうしたデータを準備する必要がでてきている。

つまり、音声・言語処理技術の研究は、出口として様々なニーズが生まれてきているものの、世界に伍して研究をするためにはより大規模な研究予算が必要とされる big science になりつつあるということである。特に、研究予算に関しては劇的な増大は困難であり、コミュニティ全体でのリソースの共有によって研究効率を改善するのが、まず最初のステップであろう。また、big science としての音声・言語処理をコミュニティ外の人々に認識させ、予算的な面も含めて研究を促進させることが必要となろう。

2. 高度言語情報融合フォーラム(ALAGIN)の設立

こうした背景を踏まえて、本年の3月に高度言語情報融合フォーラム(ALAGIN - Advanced Language Information Forum)が発足した。会長は辻井潤一東大教授、副会長は喜連川優東大教授と松島裕一情報通信研究機構理事であり、現在企業63社、大学関係者65名がメンバーである。フォー

ラムの目的は、産学官を巻き込んで、いっそうの技術共有、リソースの共有を図り、研究の効率化を図るとともに、出口としての将来のアプリケーション像を明確にするための議論を行なうことである。これ以上のフォーラムの大枠については、フォーラムの Web サイト(<http://www.alagin.jp/>)を参照されたい。

3. 情報通信研究機構からのフォーラムへの貢献

こうした活動の一環として、まず、情報通信研究機構からは音声・言語資源とツールをフォーラム会員にむけて配信を開始する。具体的に H21 年度に予定しているものは以下の通りである。

1. 対訳コーパス
2. 音声対話コーパス
3. 中国語形態素解析器
4. 文脈類似語リスト(100万語)
5. 上位語オントロジー
6. 負担・トラブル表現リスト
7. 単語クラス抽出サーバー, 意味関係抽出サーバー

対訳コーパスと音声対話コーパス、中国語形態素解析器は特に説明を要しないものと思われるが、特に対訳コーパスは今後数百万文オーダーでフォーラムでの配信を進めていく。また、それ以外の言語資源(4-6)は日本語に関するものであり、情報通信研究機構でクロールした億単位の Web 文書から自動獲得したものをベースとしている。4 の文脈類似語リストは文脈が類似している語を集めたもので、高精度な類義語リストとして利用でき、例えば、地魚の一覧を含むクラスなどが入手できる。

5 は、情報通信研究機構が別途一般公開している上位下位関係抽出ツール(<http://nlpwww.nict.go.jp/hyponymy/index.html>)

の出力を補完するものである。ツール自体は Wikipedia から100万語以上をカバーする上位下位関係を抽出するが、この上位語オントロジーを併用することでより高い精度の上位下位関係が入手できる。また、6 は検索支援システム「鳥式改」(鳥澤他、情報処理学会学会誌「情報爆発特集号」, 2008.8)で利用されている負担やトラブルを表す表現のリストであり、これによって意外なトラブルを網羅的にネット上で検索することが可能となる。7 は例えば、Web サービスで特定の単語クラス(e.g., 自動車の部品)や意味の関係(e.g., トラブル対策)が必要になった時に、Web 文書の情報をもとにフォーラム会員がそれらを作成できるサーバーであり、情報通信研究機構が運営する。これにより、自然言語処理を用いた Web サービスの開発を効率化することが狙いである。

4. 今後の展開

フォーラムではまずこうした音声・言語資源、ツールを共有することで会員のアクティビティをサポートすることから始める。本稿では情報通信研究機構からの貢献を中心に述べたが、他会員からの資源共有の申し出もすでにいくつか出ており、今後、コミュニティにおける資源共有はよりスムーズに行なわれるであろう。最終的には、より実質的な議論、共同研究を会員間で開始し、音声・言語処理の領域で、フォーラムの中からブレークスルーが生まれる環境を育むことを狙いたい。また、より遠大な目的としては、音声・言語処理技術の重要性、必然性をコミュニティ外にむけてフォーラムとして訴え、big science としての音声・言語研究を促進させる運動を起こすべきと考える。今後フォーラム会員にむけて、こうした意見を発信し、コミュニティ全体で追求すべきより適切な方向を探ることになろう。関係者の協力を願いつつ、筆を置く次第である。

みんなの翻訳- ボランティア翻訳者の支援と翻訳の共有のための Web サイト

内山 将夫[†], 阿辺川 武^{††}, 隅田 英一郎[†], 影浦 峯[‡]

[†](独)情報通信研究機構, ^{††}国立情報学研究所, [‡]東京大学

1. ボランティアによる翻訳

ボランティアの翻訳者は、様々な文書を翻訳しています。そのなかには、たとえば、オープンソースのマニュアルの翻訳や、ブログの翻訳や、NPO/NGOの文書などの翻訳があります。

ボランティアの翻訳者は、翻訳により、世の中に貢献していると言えます。たとえば、オープンソースのマニュアルの日本語訳は、日本人のユーザーにとっては、大変有難いものですし、ブログの翻訳は、他のメディアが注目しない場所や人々について光を当てるものと言えます。

したがって、ボランティアの翻訳者を支援することは、世の中に貢献することになります。

また、日本にいるボランティアの翻訳者は、現在、数千人程度ですが、外国語特に英語を翻訳できる潜在的なボランティア翻訳者の数は、数十万人程度ではないかと推定されます。

そのため、翻訳をしたい人が、簡単に翻訳ができる環境を提供すれば、現状よりも、もっと多くの人が翻訳をするようになり、より多くの外国の情報を取り込めるとともに、日本の情報を発信することもできるようになると思います。

このような動機から、情報通信研究機構言語翻訳グループと東京大学図書館情報学研究室は、共同で、「みんなの翻訳」という Web サイト (trans-aid.jp、図 1) を開発しました。



図 1: 「みんなの翻訳」サイト

みんなの翻訳の特徴は、(1)高機能な翻訳支援エディタ QRedit を誰もが利用できることと、(2)みんなの翻訳で公開されている翻訳には、「2 次的著作物を作成し、それを公開しても良い」というライセンスが付与されているため、適切な使用であれば、翻訳を利用できるということと、(3)三省堂の協力により「グランドコンサイス英和辞典(36 万項目収録)」が翻訳支援に利用できることです。

2. 高機能な翻訳支援エディタ QRedit

翻訳支援エディタ QRedit の基本設計理念は、以下の 4 点に集約されます。(1)新たな情報・機能を提供するのではなく、翻訳者が現に行っている作業の手間を省く、(2)システムが決めるのではなく翻訳者が決めるのに必要な情報を提供する、(3)翻訳者

の発想を豊かにする情報を表示する、(4)できるだけシンプルにする。これらの方針は、翻訳者へのインタビューおよび現状の翻訳支援技術の水準に基づいて決めました。QRedit では、入力された原文に対し、複数の辞書や翻訳者が登録した用語を対象に辞書引きを行ない、翻訳者は単語をクリックすることで簡単にその訳語を把握することができます。また、高度なイディオム検出機能を備えていて、語間に挿入を許した異形イディオムを検出できます。これらのイディオムに対して、QRedit では、図 2 のように、下線を引いて示すなどして、翻訳者が見落さないようにしています。イディオムは、熟練翻訳者でも誤訳する可能性がありますので、このように、エディタ側から何らかの警告を与えることは、有用です。

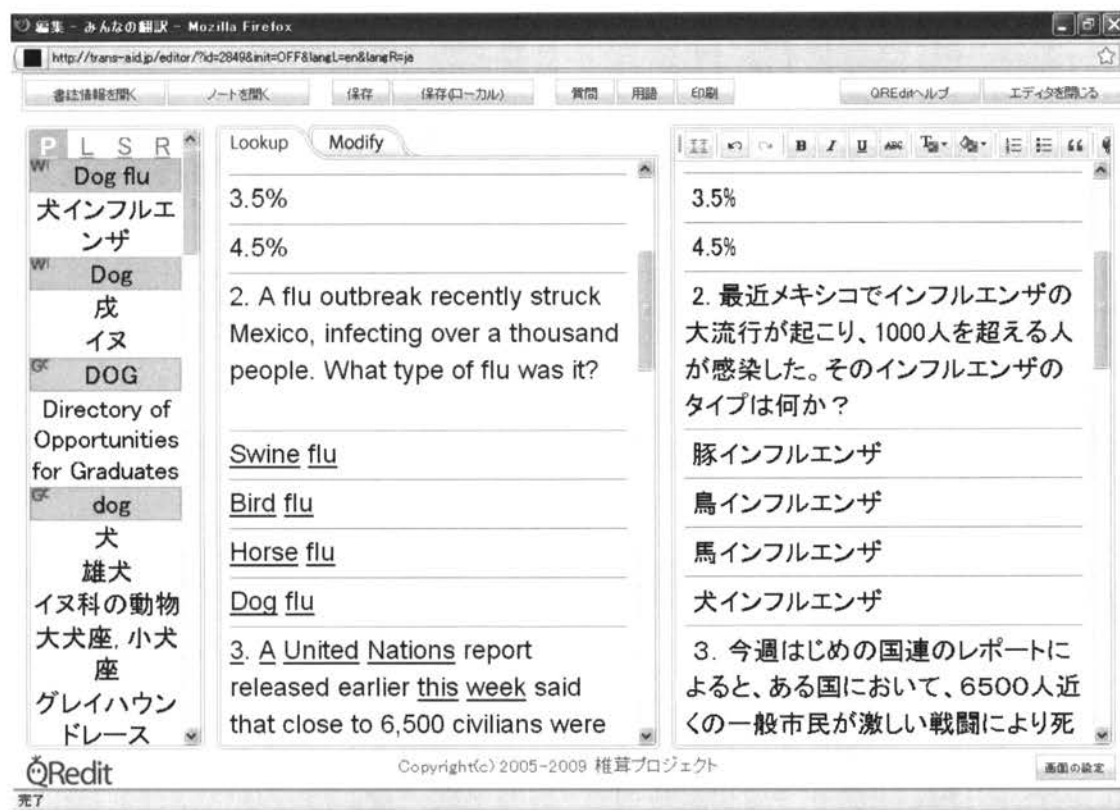


図 2 : 翻訳支援エディタ QRedit

その他にも、QReddit では、Web 検索ができるとか、用語を登録できるとかの機能があります。特に、用語を登録すると、その用語を QReddit 内から検索できます。この検索は、自分が登録した用語だけでなく、他の人が登録した用語も同様にできますので、みんなの翻訳で公開されている用語が増えれば、辞書にのっていない用語であっても、辞書引きができるようになります。

3. 翻訳の共有

翻訳結果を共有するためには、原文と翻訳文の使用許諾について考慮する必要があります。たとえば、当然ですが、原文の著者が翻訳文の公開を許可していない場合には、翻訳文は公開できないので、翻訳結果を共有することはできません。

そのため、みんなの翻訳の利用者には、原文と翻訳

文の使用許諾について確認を求めています。また、みんなの翻訳の利用者には、各自が翻訳した文は、2次的利用ができるように許可することを求めています。そのために、システムは、みんなの翻訳の利用者が翻訳文を保存するときに、以下のようにして、使用許諾などを確認しています(図 3)。

(1)まず、システムは、「あなたが翻訳の対象とした文書(原文)は、原文著者が明示的に許可している場合を除いて、私的な利用その他など、著作権法で認められている範囲でしか利用できません。原文著者は、その翻訳を公開しても良いと(あなたや他の人に)許可をしていますか?」と確認します。(2)それが「はい」の場合には、システムは、クリエイティブ・コモンズ等の「あなたの文書から 2 次的著作物を作成し、それを公開しても良い」という条件に矛盾しない使用許諾条件を設定してもらうようになっています。

公開設定

公開情報を入力して下さい みんなの翻訳における著作権の考え方

あなたが翻訳の対象とした文書(原文)は、原文著者が明示的に許可している場合を除いて私的な利用その他など、著作権法で認められている範囲でしか利用できません。
原文著者は、その翻訳を公開しても良いと(あなたやみんなに)許可をしていますか?

☒ はい ☐ いいえ

使用許諾条件 あなたの文書(訳文)の使用許諾条件を選択して下さい。
☒ クリエイティブコモンズの場合

☒ 表示	この作品を利用するときには原作者のクレジットを表示して下さい(説明📖)
☐ 表示・継承	「表示」に加えて、2次的著作物をこの作品と同一の使用許諾条件として下さい(説明📖)
☐ 表示・非営利	「表示」に加えて、非営利での使用のみを許可します(説明📖)
☐ 表示・非営利・継承	「表示・継承」に加えて、非営利での使用のみを許可します(説明📖)
☐ 制限なしに利用可能(パブリック・ドメイン)	パブリック・ドメインに属するものとして、世の中に寄贈します(説明📖)

完了

図 3 : 文書の使用許諾権の設定

このようにして、みんなの翻訳では、原著者や翻訳者の著作権を尊重しつつ、翻訳を共有できる仕組みを準備しています。

4. 今後の展開

みんなの翻訳は2009年4月8日に一般公開しました、それから1ヶ月ほどで、500人弱のユーザが

登録しています。また、アムネスティインターナショナル日本を含む4つの翻訳グループにも使っています。現状のみんなの翻訳は、英日および日英の翻訳しかサポートしていませんが、これを多言語に展開するとともに、共有された翻訳や用語を有効利用するための言語処理技術を研究開発していきたいと思っています。

自由に翻訳！

エラリー ジャンクリストフ
 翻訳・通訳（株）DOUBLET 代表取締役
 OmegaT 現地化管理担当／ユーザーサポート担当

翻訳支援ソフトとは？

翻訳支援ソフトは翻訳メモリソフトとも呼ばれ、翻訳メモリ機能によって翻訳支援を行うソフトである。

機械（自動）翻訳との決定的な相違点は、翻訳を行うのがソフトではなく、翻訳者であることだ。

翻訳メモリソフトにはいくつかの共通機能がある。

- 原文を全体として扱うのではなく、部分（分節）に分けて扱う
- 翻訳済みの分節を原文と訳文を合わせたペアとして記憶する
- 記憶したペアを次の分節の参考翻訳として使う

上記の「翻訳済みの分節を原文と訳文を合わせたペア」の集合、及びそれを含むデータファイルは翻訳メモリ（または翻訳メモリファイル）と呼ばれる。

上記の基本的機能に加え、ソフトによって下記のような機能も提供される。

- 様々な原文ファイル形式を直接扱える
- 翻訳者が分節化の規則を設定できる
- 原文の特徴（見栄えなど）を訳文に反映できる
- 分節に含まれる単語訳の参考となる用語集を扱える

また、原文、訳文、翻訳メモリ、用語集などのデータをまとめて効率よく扱うためには「翻訳プロジェクト」という仕組みも必要となる。

OmegaT は上記のそれぞれの機能を提供するだけでなく、使いやすさにも重点をおいたソフトである。

自由なソフトウェア

OmegaT は自分に合った環境で使用が可能だ。Java ベースのソフトウェアであるが故に、OS に拘らず使用できる。通常 Java 実行環境が必要となるが、Java 実行環境のない OS の場合は、Java 実行環境付きの OmegaT を使用することも可能だ。

OmegaT のもつ自由度は OS だけではない。OmegaT はまさに「自由なソフトウェア」なのだ。「自由なソフトウェア：コピー、研究、変更、配付等の扱いに関して、ほとんど、またはまったく、制限が付けられていないソフトウェア。ソースコード

の開示を前提とする。」（Wikipedia、「自由なソフトウェア」）。

OmegaT の場合、使用権利は GPL ライセンスによって与えられる。基本的には、ソフトそのものを自由に使う権利、自由に改良する権利、変更点配布の権利があり、配布するコードに対しては、同じ権利を与える義務がある。言い換えれば、誰でも自由に使用でき、コピーや改良ができるわけである。

もう一つの OmegaT の特徴はボランティアによって開発、多言語化、テスト、サポート、普及化が進められていることである。ボランティアが中心なため、OmegaT に関わる活動もまさに自由だ。優先順は人によって異なるが、ボランティアが同時に翻訳のプロでもあるため、不完全ながら、翻訳者が求める機能を知り尽くした人々たちによって作られたわけだ。

2009 年 5 月現在、開発管理者は 4 代目となり、多言語ユーザーリストの登録者数は 1000 人を超える。開発サイトからの平均ダウンロード数は（12 ヶ月にわたって）月に 4200 を超えているが、殆どの Linux ディストリビューションにも入っているため、実際の使用者数は集計不可能だ。人気のあるフリーランス翻訳者の登録サイトである proz.com でも、他の翻訳支援ソフトと同様に扱われている。

OmegaT の仕組み

ここで紹介するのは OmegaT 2.0.2 ベータ版である。OmegaT の開発過程には、「安定版」、「ベータ版」、「開発版」などがある。

「安定版」とは、公開時に十分に安定し、取扱説明書も更新された版である。「ベータ版」も公開時に十分に安定しているが、取扱説明書が未更新である。「開発版」は上記の二つと違った形で公開されており、開発中のためテスト用のみに使用されているが、特にアクセス制限されていないため、バイナリ構築が可能な環境さえ整えていれば普通に使える。

OmegaT が扱うのはファイルではなく、翻訳プロジェクトである。先ほど述べた様に、翻訳プロジェクトは「原文、訳文、翻訳メモリ、用語集などのデータ」を関連付けて管理する仕組みである。

OmegaT におけるこの仕組みは非常に分かりやすく、使いやすのが特長だ。

なぜなら、それぞれのデータを関連付けるのは単純

にフォルダであるからである。

新規プロジェクトを作ると、「原文」を入れるフォルダ、「翻訳メモリファイル」を入れるフォルダ、「用語集」を入れるフォルダ、そして、翻訳者が求めるときに OmegaT が訳文を入れるフォルダが作成される。

フォルダ単位なので、その中のデータが全て扱われる。つまり、原文ファイルを一つずつ訳すのではなく、原文フォルダ内のファイルを集合として扱われるので、HTML サイトなどのフォルダ階層の複雑なファイル集合も簡単に訳すことができる。

新規プロジェクトにおいて、既に存在しているフォルダを使用することも可能である。たとえば、顧客ごとにデータを整理すれば、新規プロジェクトでもその顧客用の原文フォルダや過去の参考翻訳メモリフォルダを指定することができる。

翻訳プロジェクトの管理作業は OmegaT に依存することなく、Mac OS X ならば Finder、Windows では Explorer などのファイル管理ソフトで可能だ。そして、設定ファイルはすべてテキスト形式で、普通のテキストエディタで編集することもできる。

OmegaT の今

最新のベータ版では「文字列の自動テキスト出力」機能が導入された。

これは、現在の原文分節、記録された訳文分節と選択されている文字列それぞれを、別のテキストファイルへ自動的に出力する機能である。

Java コードはプログラマではないユーザーにはかなり難しい。読むのも、書くのもである。従って、OmegaT の機能拡張は最近までは開発者のみが行っていた。

この新しい機能によって、bash などのシェルスクリプトやスクリプト言語の ruby、python、perl、または、単純に C 言語でも、そのテキストファイルの内容を処理することで新しい拡張機能の開発が可能になる。

例えば、現在の分節を Google Translate へ送って、その訳をクリップボードに自動的に入れることで、訳文に Google の自動翻訳を参考翻訳として使える。

もう一つの「拡張機能」は原文の「見栄えタグ」の

操作。例えば、原文から、タグを削除する、あるいは、原文からタグのみを残す...などの操作も可能になる。下記のスクリプトは「原文からタグのみを残す」一つの例である：

```
on run {}
    do shell script "sed \"s/<[^>]*>/g\"
~/.Library/Preferences/OmegaT/script/source.txt |
/usr/local/bin/pbcopy"
end run
```

sed 命令を Mac OSX の Applescript の中に入れたので、UI からそれを起動することも可能になる。

これくらいの規模の機能開発なら、ちょっとした好奇心と時間をもった方であれば行えるだろう。多言語ユーザーリストには、そういった拡張機能のコードが公開／交換／改良され、少しでもスクリプトの知識のある方ならすぐに使える。OS はそれぞれなので、共通の言語や方法が好まれるが、自分の OS で適用できない方法でも、原理さえ分かれば、数行で書き換えることができる。

API という程のものではないが、その機能によって、使用者のコンピュータリテラシーも高まるだろう。それは、自由なソフトの基本的な役割とも言える。

これから…

OmegaT は日本語化されているので、起動すれば、「お手軽スタートガイド」が表示され、そこで記述された基本操作さえ覚えればすぐに仕事で使える。取扱説明ガイドは 1.6 版から更新されていないままであるが、原理はそれほど変わっていないし、新しい機能はソフトのメニューから日本語で表示されるので、試せばすぐにわかる。分からないときにはユーザーリストに訊くことができる。そして、いずれ、最新版の取扱説明ガイドを日本語に訳すボランティアが現れたら…完成！

さあ、あなたも是非、OmegaT を試してみてください！

開発サイト：

<https://sourceforge.net/projects/omegat/>

ユーザーリスト：

<http://tech.groups.yahoo.com/group/OmegaT/>

オープンソースカンファレンス 2009 関西 in 京都
7月11日(土) 11時15分「翻訳支援ツール
OmegaT 最新情報！」

<http://www.ospn.jp/osc2009-kansai/>

人に理解しやすく 機械にも理解しやすい 産業日本語

財団法人 日本特許情報機構

渡邊 豊英、大塩 只明

1. はじめに

財団法人日本特許情報機構（Japio）は実施事業である特許翻訳事業の効率化の一環として、日本語見直しの必要性を実感し、平成 19 年度より技術用日本語に関する研究に着手し、「産業日本語」という制限言語のコンセプトをデザインし、その開発計画を策定した。

グローバル化する経済社会は事実上の世界標準言語といえる英語でのコミュニケーションをベースに動いており、日本語でのコミュニケーションをベースとする我が国の産業界にとって、言語の相違に基づく負荷は大きく、国際競争力低下のリスク要因になっている。また、表現上の自由度の大きさや曖昧な表現が、時として、言語本来の機能である正確な情報の伝達の妨げとなることや、知識管理の高度化・効率化の妨げとなっていることも否めない。

「産業日本語」はこのような状況の打開策の一つになると考える。

2. 産業日本語とは

産業日本語は産業技術文書の記述に用いられる日本語を規格化ないし標準化することにより、当該文書が伝達すべき情報を正確に伝達、かつ、機械処理しやすくする。これにより、日本語の情報力を強化し、産業技術分野・科学技術分野における知的生産性の飛躍に貢献する。

産業日本語は二つの側面を持つ。一つ目は人間のコミュニケーション能力を強化する日本語である。

- ・情報を的確かつ容易に表現できる
- ・情報を容易かつ正確に読み取れる
- ・産業関連の広範囲の人々に活用される

二つ目はコンピュータ処理に向く日本語である。

- ・コンピュータによる高度な文書処理（作成・検索・翻訳・要約等）に適す
- ・文書処理を知識処理へと発展させる
- ・現状の文書処理技術における蓄積を効果的に継承する

3. 経緯

「産業日本語」のコンセプトは平成 17 年に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（ISeC）から Japio に「明晰日本語」の名称で提案された。ISeC が Japio に提案した理由は 3 つある。一つ目は特許データが大量に使えることである。1993 年以降に公開された文献が全てテキストデータとして整理されている。二つ目は外国語の対訳データを蓄積していることである。特許の国際化の進展で外国から日本への出願が年間 6 万件あり、外国語の対訳テキストデータが入手しやすい。三つ目は正確性を判断する専門家がいることである。従来の書き方と「産業日本語」の間で意味の違いが生じないこと、また違いが生じた場合には違いの詳細を分析することができる。

Japio は ISeC の提案を受け下記の技術文書の問題を解決するため、「産業日本語」を推進することとした。

- ・現在の特許文は機械翻訳できない場合が多く、翻訳コストの低減が難しい
- ・弁理士が作成した特許文書は特殊な言い回しが多く、発明者にとって読みにくい場合がある
- ・翻訳者にとっても同様に読みにくいため、外国で有効な権利が取得できない場合がある
- ・特許は一般の人にとって難解で、技術文書として利用しにくい場合がある

もう一つの要素として、AAMT や言語処理学会等の産官学での研究開発の結果、日本語処理の知見や経験の十分な蓄積がなされてきており、またコンピュータによる日本語文書作成の高度な支援が可能となってきたことがある。

- ・ 言語学における知見の蓄積
- ・ 言語処理における要素技術の蓄積
- ・ 言語資源としての特許テキストの蓄積
- ・ 日本語文章論や日本語ライティングの経験や用途別文書作成におけるノウハウの蓄積

3. これまでの成果

平成 17 年 ISeC は産業日本語を機械処理する際に利用する予定の言語である CDL (概念記述言語)の基本仕様を策定した。CDL は構文木に加えて文間の照応等のネットワーク記述が可能である。

平成 19 年度、平成 20 年度の二年間に Japio は「経済活性化のための技術用日本語プラットフォームの開発に関するフィージビリティスタディ」を行った。この中で開発計画の策定、産業日本語での文書作成を支援するためのオーサリングシステムの動作実験と評価および産業日本語文への書き換えに必要な知識データの実験開発と評価実験を行った。このスタディと並行して Japio は言語処理、知的財産および言語学の専門家に参加をお願いして「明晰日本語委員会 (平成 19 年度)」と「産業日本語委員会 (平成 20 年度)」を開催して、平成 19 年度に明晰日本語基本仕様 (第 0 版) を、平成 20 年度に産業日本語共通基盤仕様 (第 1 版)、特許版 TJ (第 0 版)、翻訳 TJ (第 0 版)、検索 TJ (基本仕様版)、図式 TJ (基本仕様版) を策定した。なお、TJ は (Technical Japanese: 技術用日本語) の略称である。平成 20 年末、AAMT/Japio 特許翻訳研究会において「産業日本語」を説明して協力をお願いした。

4. 今年度の予定

平成 21 年度は特許分野での産業日本語の活用を具現化するため、広く配布可能な特許文書ライティングマニュアルを作成するとともに、産業日本語共通基盤仕様 (第 2 版)、特許 TJ (第 1 版)、翻訳 TJ (第 1 版)、を策定する。また、産業日本語のオーサリングシステムの作成に着手したいと考えている。さらに、「産業日本語」は我が国の産業界全体で取り組むべきテーマであり、特許以外の分野で展開するための総合的な議論・研究・開発を行い、普及を図る場としての「産業日本語フォーラム」の設立を言語処理学会等と調整中である。

5. 世界を睨んだ知的財産戦略の強化

「知的財産推進計画 2008」において、国際競争力を強化するために、より一層の知的創造サイクルのグローバル化が提唱されている。2008 年 5 月に総合科学技術会議から公表された「革新的技術戦略」では、研究開発の初期段階からの戦略的な知財の創造・保護・活用を図る体制の整備の重要性が謳われている。2008 年 9 月に経済産業省から発表された「新経済成長戦略のフォローアップと訂正」等においても、世界を牽引する知財システム構築の必要性が謳われている。特に、2008 年夏に発表された「イノベーションと知財政策に関する研究会」での提言には、特許関連文書において使用される日本語の見直しと関連があるものも多い。

「産業日本語」の導入は、これらの課題の解決策となると確信している。

参考資料

- ・ 産業日本語ホームページ
<http://japio-tjp.org/index.html>
- ・ 平成 19 年度、20 年度シス協報告書要約
<http://www.japio.or.jp/kenkyu/kenkyu01.html>

平成 19 年度、平成 20 年度の「経済活性化のための技術用日本語プラットフォームの開発に関するフィージビリティスタディ」事業は、(財) JKA の機械工業振興事業補助金の交付を受けて行う (財) 機械システム振興協会の委託事業により実施しています。

言葉と人材 ― グローバル時代における企業内の言語とコミュニケーション

日本貿易振興機構(JETRO)

九門 崇

1. はじめに

第1回の原稿では、グローバル化が進展する中、日本企業が「世界標準」を考えるに際して重要な領域になると考えられる日本国内や海外法人におけるグローバル人材(外国人)の活用について概論を述べた。

今回は、こうした日本企業のグローバル人材活用において、日本企業の社内言語とコミュニケーションスタイルについて考察してみたい。

中国では、上海など大都市の市場は様々な国籍の企業が参入していて、すでに「ミニ・グローバル」といえる状況である。欧米系企業では、中国でもオーストラリア人やシンガポール人など様々な国籍の人々が幹部になるケースが見られ、日系企業でも現地化ということで中国人を幹部にする取り組みを始めている。こうした人材登用を進めるにおいても、グローバルな発想をさらに広げていく必要がある。

そこで重要になってくるのが、異なる国籍の社員間でのコミュニケーションをどこまで「世界標準」に合わせしていくのかという問題だ。ここでは言語とコミュニケーションの2つに分けて考えたい。

2. 英語がグローバルビジネスの共通語に

まず、グローバルビジネスの世界では、英語が共通語化しているのが現実だ。作家の水村美苗氏は著書『日本語が亡びるとき―英語の世紀の中で』で、インターネットの普及により、英語が普遍語(universal language)としての地位を確立する時代に入っているとしており、その上での日本語教育や英語教育の必要性を訴えている。冒頭に述べた中国での例のように、どの国の企業であれ、組織で働く人材の国籍が多様化していけばいくほど、共通の言語でコミュニケーションをとる必要が出てくるためだ。

アジアでは中国やその他東南アジアに華僑が多

いことから、英語に次いで中国語の重要度も増すだろう。今後、中国系企業がさらにグローバル展開していく際の社内共通語が英語になるのか興味深いテーマだ。

では、日本企業にとって社内共通語の英語化は必要なのか？ またそれは可能なのか？ 実際に社内オペレーションをすべて英語化しようとする、特に本社で様々な問題が起こるだろう。海外とのやり取りや会議を英語化することにとどまらず、社内の文書や e-mail などすべて英語化した際に、社員や幹部がどこまで対応できるかという大きな課題が生じるためだ。逆に、海外法人においては、ロシアなど市場が拡大するも日本語が話せる人材が不足している国・地域では、社内共通語を英語にしている場合もみられる。その中で、三菱商事は社内コミュニケーションにおいて、日本語と英語の「バイリンガル化」を進めており、今後の動向が興味深い。

完全な英語化に向かう前の現実的な対応としては、外国人を採用する際に、日本語よりも使用人口が圧倒的に多い英語を話し、能力評価の高い人材をより大きな世界中の人材市場から採用することだ。その上で日本語教育を施す方がより優秀な人材を確保できる。

しかし、実際にはいくら英語が流暢に話せても日本語が十分に話せない人材はそれだけで日系企業には採用されにくいのが現状だろう。アジア最大のインターネット求人サイトを運営するジョブストリート(本社マレーシア・登録者数500万人超)日本法人とエーオンコンサルティングジャパンが調査(2008年9月、有効回答数28,679人、対象者の居住国はインドネシア、マレーシア、シンガポール、フィリピン、タイ、ベトナム、インド)したところ、日本企業で働く際の障害として、「言葉(79%)」が「昇進の機会が少ない(26%)」など他の回答を圧倒的に引き離してトップに挙げられ

ている。しかし、日産やソニーのトップに外国人が就任したように、今後は経済成長著しいインドや中国の企業が日本企業を買収してインド人や中国人がトップになるケースが出る可能性もある。その場合の社内共通語が日本語になるという保証はない。

3. 相手を「同化」させない「対話」のコミュニケーションへ

もう一つの課題は、単に言語の問題というよりも、一般的なコミュニケーションの問題だ。言語の障壁は社内で通訳をつけるなどすればある程度解決する。しかし、それだけでは不十分な場合もある。

グローバル人事に詳しいコンサルタントによると、海外の日系企業で、日本人駐在員が英語や中国語が話せないため通訳をつけた時、それでも自分の意図していることが相手に伝わらないことがあったという。日本人の場合、それは相手の理解力が不足していると考えがちだが、実際によく話を聞いてみると、日本人側が相手にわかりやすく話せないケースも多いようだ。

一般的に日本人のコミュニケーションは“あうんの呼吸”や“察する”に見られるように、しばしば「暗黙知」的になる。価値観がお互い異なることを前提にコミュニケーションしている人は少なく、またそういう教育やトレーニングも受けていない。

しかし、これからは、異なる価値観を持っているからお互い「対立」したり、相手を「同化」させるということではなく、それ前提として「対話」に変えていく努力がビジネスの世界でも一層必要になる。例えば、欧米を始めとして、インド人や中国人、ロシア人など新興国も含めた多様な価値観を持つ世界のリーダーとのコミュニケーションができる人材をどう育てるかは大きな課題だ。

英語で、“Let's agree to disagree”という表現がある。「お互い意見が異なるという点に同意しよう」ということだ。ここで重要なのは、同質化するのではなく、個人としての自分のアイデンティティをしっかりと持った上で「対話」していくということだ。日本人が全くアメリカ人のような考え方になった

ら、それは個性が失われるということで、相手もそういう日本人に魅力は感じないだろう。同様に、日本企業がアメリカ企業と全く同じように行動してもその強みを発揮できるかどうかは疑問である。

グローバリゼーションが進んだ世界では、人、情報、資金などが自由に国境を越えて移動する。その本質は、「フラット化」と「多様化」がないまぜになった「多様な価値観を持つ世界」だ。グローバル化が進むことによって世界が全て同質化に向かうわけではなく、より多様性が浮き彫りになる世界になるということだ。その時代には、より自分自身の個性や自国の価値観を強く意識しながらどう異なる価値観の人々と仕事を進めていくかということが重要になる。

例えばシンガポールが東南アジアの地域統括拠点としての地位を維持しているのは、近隣のタイ、インドネシア、マレーシアまで飛行機で1時間程度、インドやベトナムでも同5時間くらいで行けるという地政学的優位性だけが要因ではない。インド人・マレー人・華人など様々な人種が一緒に住んでいる、「ミニ・アメリカ」といえる多様性の社会であることも影響しているのではないか。また、シンガポール政府は海外から優秀な人材を誘致することに積極的で、インドではシンガポールに来ませんかというテレビ広告が流れていたり、世界中から研究者を集めるような努力をしている。

フラット化により「人」の自由な移動が国境を越えて企業間でも起こると、価値観が違うことを前提としたコミュニケーションや企業の人事システムが求められる。そして、外国人であれ、女性であれ、転職者であれ、全て同じプラットフォームの上で同じシステムに基づき評価されていく。これが欧米のグローバル企業が行ってきたことだ。日本企業は、社内共通言語をどこまで英語にするのか、企業システムをどこまで世界で統一していくのかが問われているといえる。

参考文献

「日本語が亡びるとき—英語の世紀の中で」
水村 美苗、筑摩書房、2008年

中国・韓国特許検索システムの紹介

株式会社クロスランゲージ

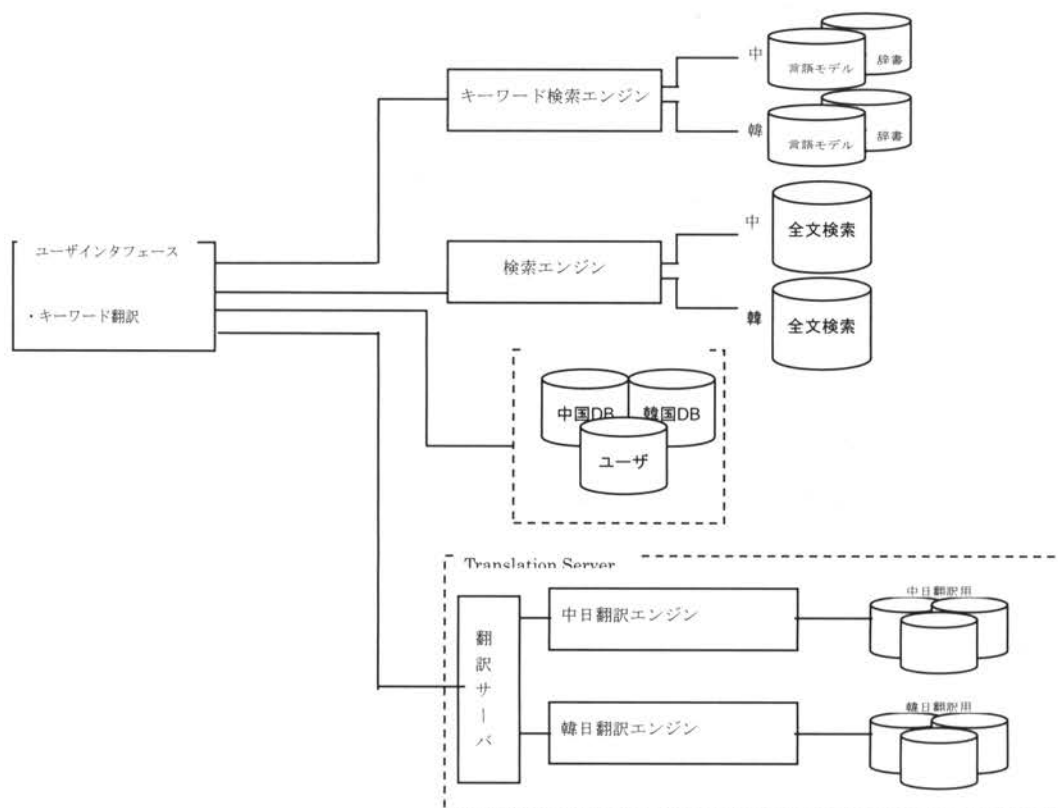
株式会社クロスランゲージにて、2年半以上サービスを提供してきた中国・韓国の特許検索サービス、「WORLDPATENT.JP」の仕組みを紹介する。本サービスは、検索システムと機械翻訳システムを組み合わせた特許情報の検索サービスである。

1. 概要

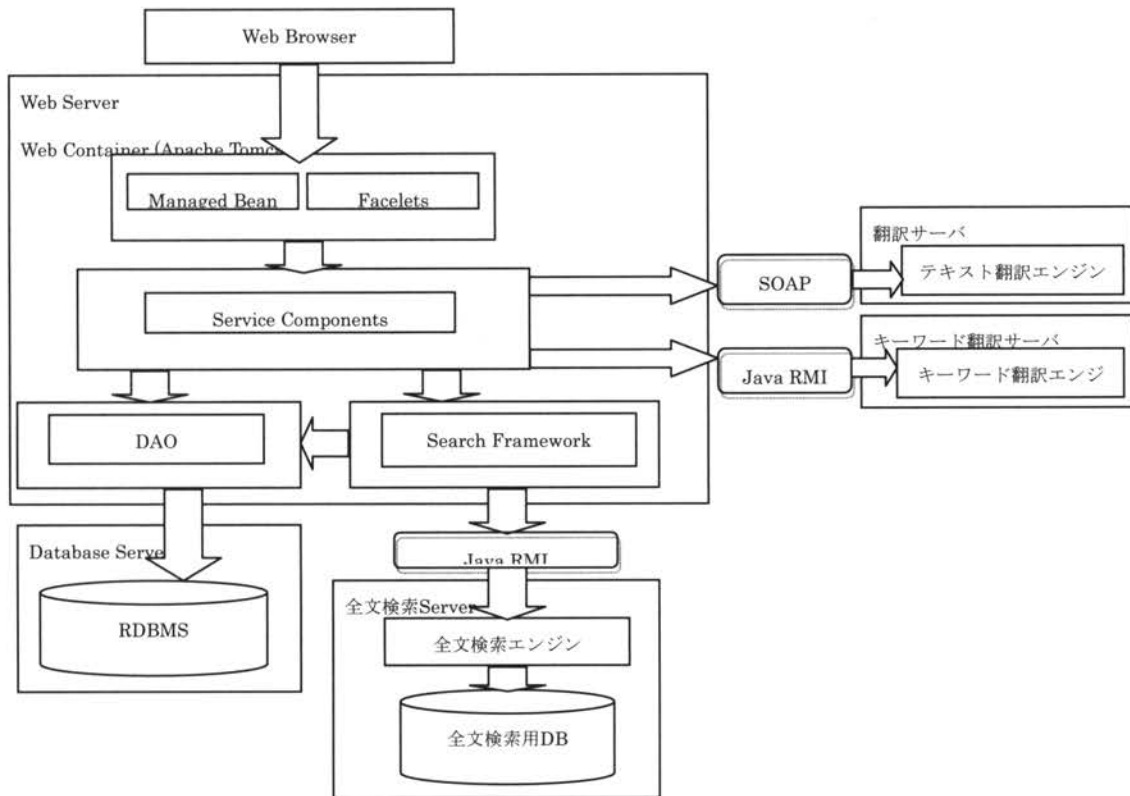
株式会社クロスランゲージは、2007年11月、中国特許・実用新案・意匠に関して検索サービスを開始した。クロスランゲージは、中国特許産権出版局（CNIPR）と、再販契約を締結し、データを購入

している。このデータは、中国語簡体字によるテキストデータと、公報全図と、代表図の画像データである。これら中国語のテキストを、原文のままデータベースに蓄積してあり、機械翻訳システムにより日本語の検索キーワードを中国語に変換して検索し、検索結果の中国語原文データを日本語に機械翻訳して、画面に併記して表示するシステムである。必要に応じ、代表図、公報全図を表示することもできる。またデータについては、必要データを抽出後、複数データを一括してTSVファイル、PDFファイルへ出力してダウンロードすることができる。

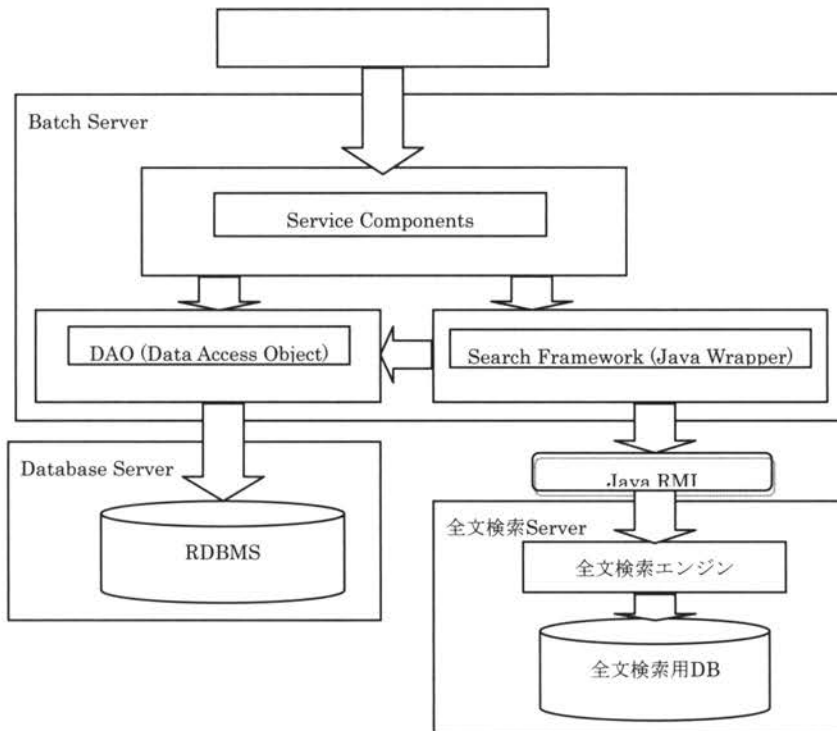
WORLDPATENT.JPシステム概要図



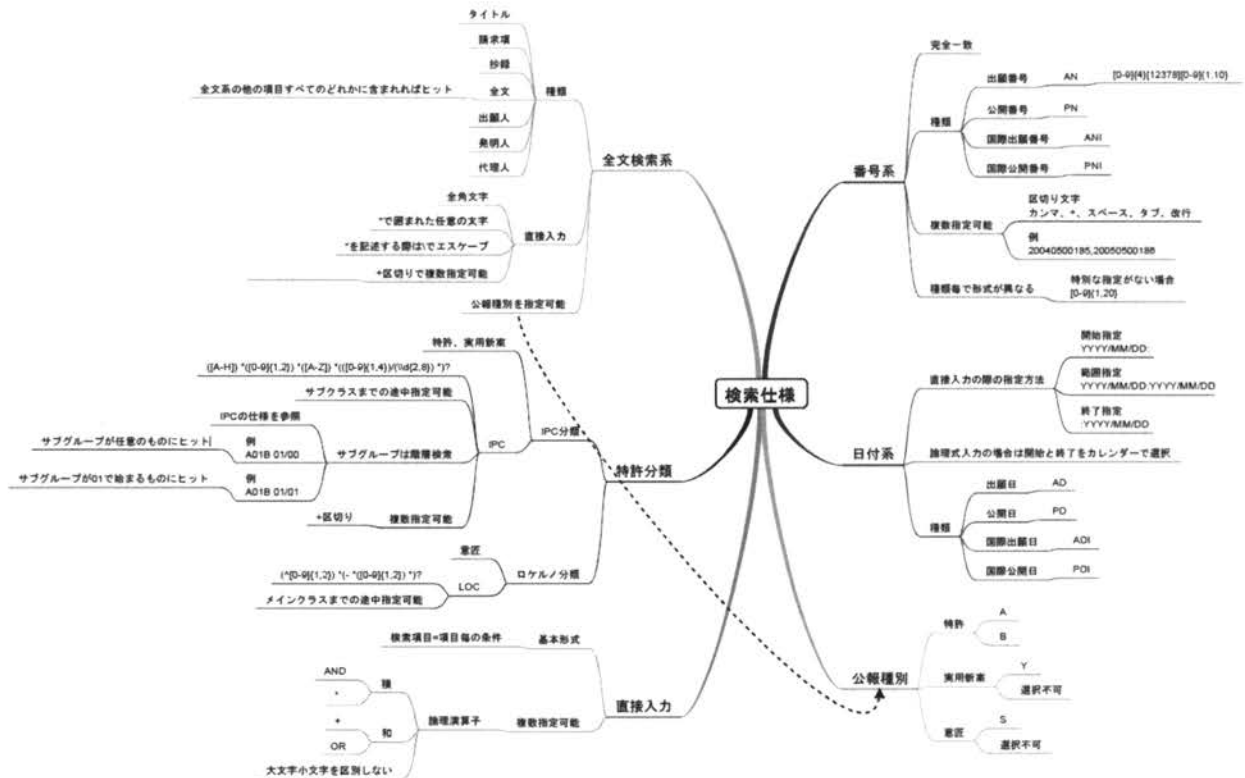
WORLDPATENT.JPソフトウェア構成図【Webアプリケーション】



WORLDPATENT.JPソフトウェア構成図【蓄積系】



検索仕様



2. 中国語の翻訳エンジン

本システムに搭載している中国語・日本語の翻訳エンジンは、クロスランゲージの中国語パッケージ製品である「翻訳ピカイチ中国語」(旧製品名・蓬萊)のエンジンに、特許文章用のチューニングを加えたものである。チューニングを加えないままの訳では、例えば「本発明」と表示すべきところで、「本の発明」などと表示してしまう。これら特許特有の表現をあらかじめ登録して利用している。

韓国語も同様に「翻訳ピカイチ韓国語」(旧製品名・高麗)のエンジンに、チューニングを施して利用している。

3. データの種類

本システムは、中国特許、中国実用新案、意匠についてのデータを蓄積している。特許は、公開と登録と、両方を蓄積している。実用新案と意匠については、登録されたデータのみ。

中国特許、実用新案については(1)書誌、(2)

要約、(3)第一請求項および(4)法的状況をテキストデータで蓄積しており、代表図と公報は画像データとして蓄積している。

意匠は書誌と要約と法的状況、全図を蓄積している。韓国については、特許と実用新案になっていて、それぞれの公開と登録データを蓄積している。意匠はサービス対象外である。

4. 検索の仕様

検索の項目は、公開種別系、番号系、日付系と全文検索系で、下記の内容になる。

発明の名称、出願人、発明人、代理人、要約、第一請求項、フリーキーワード、出願日、出願番号、公開(公告)日、公開(公告)番号、国際出願日、国際公開日、国際公開番号、IPC分類(主分類)、IPC分類、優先権主張番号、優先権主張日、優先権主張国

セッション中は一度検索した内容はセッション集

合として保持されているが、別途保存集合として保
ことも可能である。検索項目として作成した集合を
指定することにより、絞り込みが可能である。

5. 表示の仕様

検索の結果は下記のように一覧表示される。下記の
場合は5件ずつの表示をする設定であり、デフォル
トで翻訳の名称のみ、機械翻訳して一覧表示する。

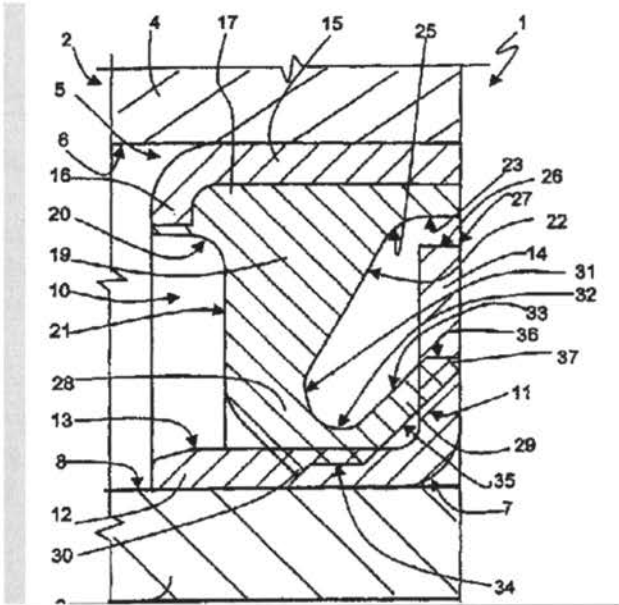
検索		OPEN
一覧から 全て選択 して 選択した文庫を 詳細表示 一時フォルダに追加 選択解除		
☐ [発明等の名称] 滚动轴承的密封装置 転がり軸受の密封装置 [出願番号/日] 200180000021425 2001/09/17	公開 追加 表示	☒ 出願番号/日 ☐ 代理人 ☐ 公開(公告)番号/日 ☐ 主分類 ☐ 国際出願番号/日 ☐ 副分類 ☐ 国際公開番号/日 ☐ 優先権主張 ☐ 出願人 ☐ 元出願 ☐ 発明人 ☐ 参考文献 公報種別: 特許公報 翻訳種別: 対訳 辞書1: 標準 辞書2: 標準 表示件数: 5 件 再表示
☐ [発明等の名称] 将油液填充进入滚珠轴承中的方法和装置 滾珠を油を充填することによって滾珠軸受中に油を注入する方法と装置 [出願番号/日] 200310000103127 2003/10/31	公開 追加 表示	
☐ [発明等の名称] 轴承保持器 軸受フレーム [出願番号/日] 200310000102995 2003/10/31	公開 追加 表示	
☐ [発明等の名称] 滚动轴承定位构造 玉軸受は構造に位置確定する [出願番号/日] 200210000046480 2002/11/11	公開 追加 表示	
☐ [発明等の名称] 轴承装置及利用该轴承装置的马达 軸受装置とこの軸受装置を利用するモータ [出願番号/日] 200280000003272 2002/09/12	公開 追加 表示	
1 2 3 4 5 6 7 8 9 10 次のページ ▶		

ここからひとつのレコードを指定すると、下記のような表示になる。

検索		OPEN
一覧から 全て選択 して 選択した文庫を 詳細表示 一時フォルダに追加 選択解除		
戻る		
☐ [発明等の名称] 滚动轴承的密封装置 転がり軸受の密封装置 [出願番号/日] 200180000021425 2001/09/17 [公開(公告)番号/日] 1483117 2004/03/17 [国際出願番号/日] 0000010709 2001/09/17 [国際公開(公告)番号/日] 0000053955 2002/07/11 [出願人/権利者] SKF工業株式会社 [発明者] 安德烈·戈里塞利 安吉洛·維尼奧托 [代理人] 張金素 [主分類] F16 15/32 [副分類] F16 15/32 [優先権主張] TO2000A001222(IT 2000/12/28) [元出願] [参考文献]	公開 追加 表示	☒ 第一請求項 公報種別: 公開公報 翻訳種別: 対訳 辞書1: 標準 辞書2: 標準 表示件数: 5 件 再表示
文献フォルダ 保存文庫数: 5810件/2000		
セッション 保存		
一時フォルダ (0件) 集合13 (3,262件) (TI=(軸受+提高基本工资)) AN...		

[要約]
滚动轴承(2)的密封装置设置有第一罩(5)和第二罩 轉がり軸受(2)の密封装置は第一(5)を覆い第二が覆
(7), 这些罩由金属材料形成并且各自连接制滚动轴承 いを設置してある(7), これは覆って金属材料から成形
(2)本身的外座圈(4)和内座圈(3)上, 由合成材料形成 してしかもそれぞれ転がり軸受(2)に接続することは自
的密封件(10)与第一罩(5)形成一体并且具有密封唇 身の外座圈(4)と内座圈(3)上にこれを置く。第一(5)が
(11), 该密封唇(11)基本上布置成与第二罩(7)形成接 一体化になるしてしかも金属材料成形した封止材(10)を被覆
触, 第一罩(5)由圆柱形部分(15)来限定出, 该圆柱形 することをするを覆うことと合成する(11), 前記は
部分(15)装配在外座圈(4)上, 密封件(10)固定到圆柱 唇(11)を被覆して基本時に第二と(7)を覆って接触する
形部分(15)上从而从圆柱形部分(15)本身进行自由延 ことを形成することにより仕掛けする。第一(5)が筒形
伸并且突出在圆柱形部分(15)本身之上, 并且它向着第 形部分(15)から限にきて規定はまだ決めることを覆う。
二罩(7)的圆柱形滑动表面(13)在轴向上具有锥形 前記筒形部分(15)組み立てた不在のこの座圈は
形状。(4)上を覆む、封止材(10)は筒形部分(15)上がこれ
によって筒形部分から部分的であること(15)自身自由延
びてしかも突出して筒形部分(15)自身以

代表図を表示するオプションをオンにすると、要約と第一請求項の間に代表図を表示する。



【第一請求項】

1. 一种滚动轴承(2)的密封装置(1), 该滚动轴承包括内座圈(3)和外座圈(4), 这些座圈相互之间共轴线并且与滚动轴承(2)本身的旋转轴线(A)共轴线, 该密封装置(1)包括: 第一罩(5), 它由金属材料形成, 它连接到外座圈(4)上; 第二罩(7), 它由金属材料形成, 它连接到内座圈(3)上; 及密封件(10), 它由合成材料形成, 它与第一罩(5)形成一体并且设置有密封唇(11), 该密封唇基本上布置成与第二罩(7)相接触; 该装置(1)的特征在于: 第一罩(5)由圆柱形部分(15)确定, 该圆柱形部分装配在滚动轴承(2)的外座圈(4)上; 所述密封件(10)固定到圆柱形部分(15)上, 从而从圆柱形部分(15)本身进行自由延伸并且突出在该圆柱形部分(15)本身之上, 并且它向着第二罩(7)的圆柱形滑动表面(13)在轴向上部分上具有锥形形状。

1. 転がり軸受け(2)の密封装置(1)は、前記の転がり軸受け(2)の内座席(3)と外座席(4)を含んで成る(4)、しかも転がり軸受け(2)自身の回転軸線(A)共軸線相互間の共軸線とこれらの座席(3)は囲む、前記密封装置(1)は以下を含む: 第一覆う(5)、それは金属材料から成形する、それは外座席が(4)上を囲むことに接続する; 第二覆う(7)、それは金属材料から成形する、それは内座席が(3)上を囲むことに接続する; 封止材(10)、それは材料を合成することから成形する、それと第一は(5)が一体化になるしてしかも唇を被覆することを設置してあることを覆う(11)、前記は唇を被覆して基本的に第二と(7)相を覆って接触することように仕掛ける; 前記装置(1)の特徴: 第一(5)が筒形状部分(15)から確定を覆う、前記筒形状部分は転がり軸受け(2)の外座席に(4)を囲むことを組み立てる; 前記の封止材(10)は筒形状部分(15)が上に固定する、(15)自身自由に延びてしかも突出してこの筒形状に部分的である(15)自身以上これによって筒形状から部分的である、しかもそれは第二が(7)筒の形スライド面(13)を覆って軸方向部分に錐の形形状を有することに向かっている。

【注的状況】

【状況】

【公開日】

【備考】

【状況】

【公開日】

公開

2004/03/17

証明日:

詳細: 公開

実質審査の生

効

2004/05/26

【状況】

【公開日】

【備考】

【状況】

【公開日】

公開

2004/03/17

証明日:

詳細: 公開

実質審査の効

力発生

2004/05/26

6. ダウンロードの仕様

原文と訳文を TSV ファイルの状態ダウンロードできる。エクセルなど一般のソフトウェアで開いて社内管理可能になる。また、代表図、公報などとも画像のファイルは、PDF 形式でもダウンロードが可能である。

7. システムの今後

現在、本システムは、オンラインで提供しているため、デュープレックス構造になっているが、今後はシステムの構成をシンプルにして、企業・組織向けにシステムごと提供してゆく予定である。

「The 翻訳® 2009 プレミアム」

東芝ソリューション株式会社

1. はじめに

東芝ソリューション株式会社は、「用例ベース自動翻訳」で翻訳精度と翻訳効率を向上させた、英日/日英翻訳ソフト「The 翻訳 2009 プレミアム」を発売しました。「用例ベース自動翻訳」は、登録された用例を高精度な翻訳エンジンが自動的に参照し、過去に訳した表現や言い回しなどを有効活用し、より自然で読みやすい文に翻訳できます。

また、本商品では、「ダブル知識翻訳」を強化し、翻訳精度を向上しました。さらに、訳し方が何通りかある語句の場合、Web 検索サイトを使って、インターネット上で使われている例を参考にして訳語を選択できる「フレキシブル Web 検索」機能を備えました。

2. The 翻訳 2009 プレミアムの主な特長

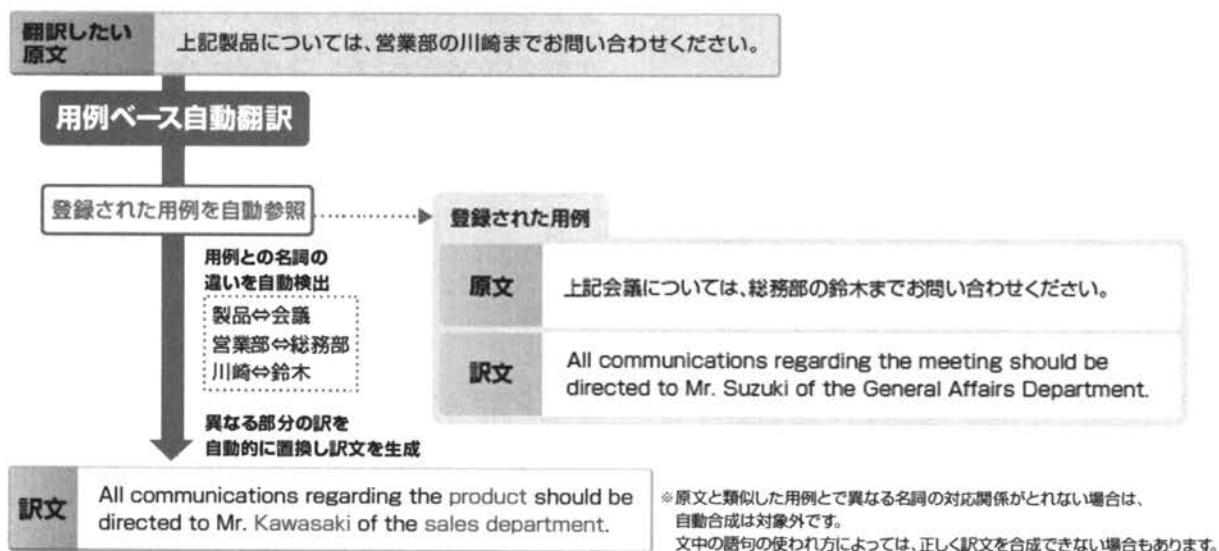
(1) 用例ベース自動翻訳

従来の翻訳メモリ機能を用いた翻訳では、翻訳メモリ辞書に翻訳する文と全く同一の用例が登録されていなくても、一部が異なる用例（類似用例）が登録されていれば、その類似用例の訳文を参照し、異なる部分を変更して完成文を作成することができました。ただし、類似用例の訳文中の編集が必要な語句とその置き換え候補（原文の異なる部分の訳

語候補）をわかりやすく提示するなどのサポート機能はあったものの、ユーザが手動でその部分の訳語を作成する必要がありました。The 翻訳 2009 プレミアムは、こうした編集作業をより効率化し、利用者が過去の翻訳結果を用例として登録した翻訳メモリ辞書を自動検索し、原文の類似用例と原文で使われている名詞の違いを検出し、その名詞を翻訳した結果を類似用例の訳文と自動合成して訳文を出力する「用例ベース自動翻訳」を搭載しました（図1）。過去の訳文例を活用して訳文を自動生成できるため、マニュアルや企業内文書、特許文書などの定型の業務文書を扱うプロフェッショナルな翻訳業務に適しています。ユーザが用例の蓄積を重ねることで、過去の訳文を流用した自然な翻訳結果が得られ、翻訳精度を向上することができます。

(2) ダブル知識翻訳

「ダブル知識翻訳」は、「原文を解析するための知識」と「訳文を生成するための知識」の2つの知識を活用し、より自然で読みやすい訳文を実現しています。The 翻訳 2009 プレミアムでは、これらの2つの知識を約3000ルール増やし、翻訳精度を向上させました。例えば、「騒音寄与スペクトルは、通常1000Hz未満の周波数域に集まっている。」と



過去の訳文を流用した自然な翻訳結果が得られます。

図 1 用例ベース自動翻訳の例

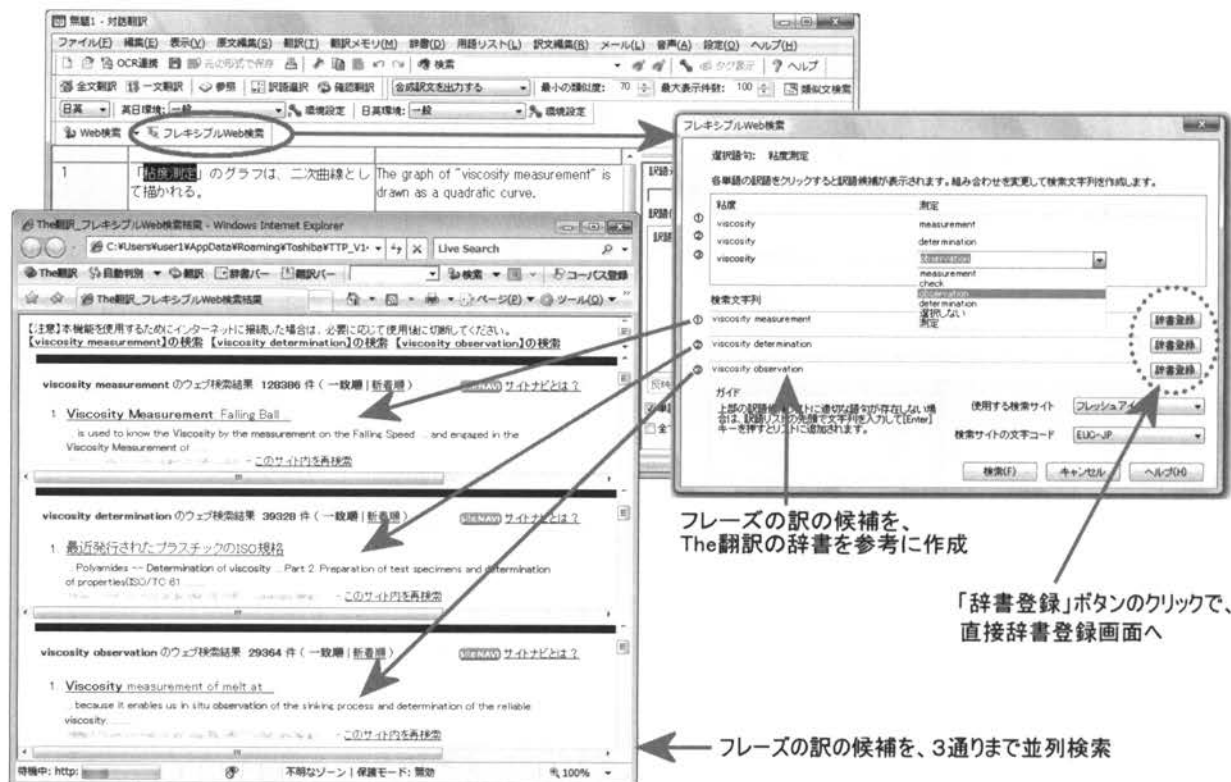


図 2 フレキシブル Web 検索の例

いう原文において、状態を表す日本語の自動詞表現「...に集まる」を英語受動態表現「be concentrated in ...」に変換することによって、下記のように自然な英語に訳出することができます。

●従来の訳文：

Noise contribution spectra **have usually gathered in** the frequency region below 1000 Hz.

●The 翻訳 2009 プレミアムの訳文：

The noise contribution spectrum **is usually concentrated in** the frequency region below 1000 Hz.

(3) フレキシブル Web 検索

訳し方が何通りかあるフレーズを Web 検索サイトを使って並列検索し、インターネット上で使われている例を参考にして訳語を選択できます。辞書登録機能と連動しているので、選択した訳語を簡単に辞書登録でき、ユーザ辞書の構築作業の効率化が図れます(図2)。

3. その他の機能

The 翻訳 2009 プレミアムは、上記に加え、以下のような多彩な機能を揃えています。

①対話翻訳(高機能翻訳エディタ)：様々な形式の文書を読み込み、原文と訳文を比べながら翻訳できます。メモリ辞書の活用を強力に支援する絞り込み検索・編集機能、文や単語単位での訳文情報の色分け表示機能、任意の文字列や数字を含む表現を“*”や“#”を使って検索できる「ワイルドカード検索」機能、一文ごとにメモや付箋が付けられるメモ機能など様々な支援機能があります。補助ウィンドウとして、サイドウィンドウ、情報ウィンドウ、ボトムウィンドウがあり、訳語リストや翻訳メモリの検索結果などを閲覧しながら効率的に作業が行えます。

②アプリケーション連携翻訳：Microsoft® Word、Microsoft® Excel、PowerPoint®、一太郎、秀丸エディタなどのアプリケーションに翻訳機能をアドインして翻訳作業が行えます。

③Acrobat®/Adobe® Reader®連携翻訳：PDF 文書を Adobe® Reader®でテキストを取り出して翻訳できます¹。さらに、Adobe® Acrobat®があれば PDF 文書のレイアウトを保ったままでも翻訳でき^{1,2,3}、イメージで作られた PDF 文書も文字認識し

1 ファイルのセキュリティの設定によっては翻訳できない場合があります。

2 Adobe® Acrobat® 9.0/8.1/8.0/7.0 が必要です。無償配布の Adobe® Reader® 9.0/8.1/8.0/7.0 では、連携テキスト翻訳のみとなります。レイアウトは保たれません。

3 翻訳対象文書のレイアウトや作成方法によっては、レイアウトが保たれない場合があります。

て翻訳できます⁴。

- ④ ホームページ翻訳：Microsoft® Internet Explorer®、Firefox などのアプリケーションでホームページを表示しながら翻訳できます。
- ⑤ セレクトコーパス翻訳：翻訳したい分野の文書を登録すると、登録文書で使用されていた語句が優先して使用され、その分野に適した翻訳結果を得ることができます。
- ⑥ 豊富な辞書構築・編集機能：辞書登録候補の抽出、登録用属性の自動付与、辞書編集エディタなど、ユーザ用辞書を構築・編集するためのツールが整備されています。
- ⑦ 高機能 OCR ソフト「Express Reader PRO LE」を搭載：スキャナで読み取った紙文書のテキスト認識結果はテキスト、HTML、RTF の各形式で保存できます。また、直接対話翻訳に取り込み、すぐに翻訳できます。

4. 商品ラインアップ

The 翻訳 2009 プレミアムシリーズは、以下のラインアップとなっています。

- The 翻訳 2009 プレミアム
 - ・搭載辞書語数：484 万語
 - ・搭載翻訳メモリ用例数：20 万例
- The 翻訳 2009 プレミアム 専門辞書パック
 - ・搭載辞書語数：745 万語
 - ・搭載翻訳メモリ用例数：60 万例
- The 翻訳 2009 プレミアム 特許エディション
 - ・特許文書向け翻訳機能を搭載した最上位商品
 - ・搭載辞書語数：766 万語
 - ・搭載翻訳メモリ用例数：62 万例

また、ビジネスユース向けに、ビジネスで利用頻度の高い機能を選びすぐって構成した商品を用意しています。

- The 翻訳 2009 ビジネス
 - ・搭載辞書語数：484 万語
 - ・搭載翻訳メモリ用例数：15 万例

⁴ Adobe® Acrobat® 9.0/8.1/8.0/7.0 が必要です。レイアウトは保たれません。特殊文字、特殊フォントが使用されている場合や、イメージ部分の元原稿の状態によっては、正しく認識されない場合があります。

5. 動作環境

対応 OS (すべて日本語版)
Windows Vista® Home Basic SP0 / SP1 Windows Vista® Home Premium SP0 / SP1 Windows Vista® Business SP0 / SP1 Windows Vista® Ultimate SP0 / SP1 Windows XP® Home Edition/Professional SP2 / SP3 Windows® 2000 Professional SP4 (Windows Vista®は 32 ビット版)

対応機種	PC/AT互換機
CPU	Pentium®以上/Celeron®以上推奨
メモリ	■The翻訳必要分：128MB以上 (256MB以上を推奨) ■OS使用分を含む推奨値 Windows Vista®：768MB以上 Windows XP®：384MB以上 Windows® 2000：320MB以上
ディスク	■The翻訳 2009 プレミアム：1.1GB ■The翻訳 2009 プレミアム 専門辞書パック：1.9GB ■The翻訳 2009 プレミアム 特許エディション：2.0GB 各商品では、下記を必要に応じてインストールできます。 ◎OCRソフト (Express Reader PRO LE)：60MB ◎研究社 リーダーズ英和辞典 (第2版)：110MB ◎研究社 新英和中辞典 (第7版) / 新和英中辞典 (第5版)：550MB ■The翻訳 2009 ビジネス：950MB 下記を必要に応じてインストールできます。 ◎研究社 新英和中辞典 (第7版) / 新和英中辞典 (第5版)：550MB

■製品情報 URL

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

- Microsoft、Windows、Windows XP、Windows Vista、PowerPoint、Internet Explorerは、米国Microsoft Corporation の米国およびその他の国における登録商標または商標です。
- Adobe、Acrobat、Adobe Reader、Adobe Acrobatは、Adobe Systems Incorporatedの登録商標または商標です。
- FreshEyeおよびフレッシュアイは、株式会社ニューズウオッチの登録商標です。
- The翻訳、セレクトコーパス翻訳は東芝ソリューション株式会社の登録商標です。
- 記載されている商品の名称はそれぞれ各社が商標として使用している場合があります。
- 本製品の仕様は予告なく変更される場合があります。

AAMT 機械翻訳文テストセット（第1版）公開に向けて

機械翻訳課題調査委員会ワーキンググループ 1

1. はじめに

アジア太平洋機械翻訳協会の機械翻訳課題調査委員会（以下、AAMT 調査委員会）では、昨今の日中機械翻訳への関心の高まりを背景として、現状の日中翻訳システムの品質を数値化し、さらに、具体的な問題点を明確化することを目的として、(1) 日本語・中国語の基本文対訳コーパス（テスト文）、(2) 各基本文の訳文チェック項目（設問）から成る AAMT 機械翻訳文テストセット（以下、AAMT テストセット）の構築を行ってきた[1]。この度、AAMT テストセット第1版が完成し、これを公開するためのダウンロード・サイトの立ち上げを行ったので報告する。

2. 日中機械翻訳テスト

機械翻訳の評価にテスト文と設問を用いる評価法は、社団法人日本電子工業振興協会の機械翻訳技

術分科会が提案した日英翻訳システム用のテストセット[2]（以下、JEIDA のテストセット）で提案されたものである。AAMT テストセットは、JEIDA テストセットを日中翻訳に拡張したものとして位置付けることができる。AAMT テストセットのサンプルを図1に示す。テスト文は合計 377 文であり、このうち 33 文については2種類の参照訳が付けられている。また、各テスト文には、Yes/No で回答可能な設問が、中国語とその日本語訳で付与されている。

AAMT 調査委員会では AAMT テストセットおよび JEIDA テストセットを使って、インターネットで無料公開中の日中、日英各7種の翻訳システムの評価を実施した。設問、流暢さ、正確さの3つの観点で、中国語ネイティブの評価者2名により1文毎に採点を行った。

例文の分類		日英評価テストセットにおける文番号		設問（日本語、中国語）		流暢さ、正確さによる 5段階評価の平均値				
文番号：1～377		日本語例文		中国語参照訳(2種類)		設問の解答（Yes=1, No=0 の平均値）				
A	B	C	D	E	F	G	H	I	J	K
1	カテゴリー	文番号	日本語(原文)	中文1(正解1)	中文2(正解2)	設問(日本語)	設問(中国語)	設問評価スコア	Fluencyスコア	Adequacyスコア
1	(1) 述部	1	JEG11 1001 彼は多くの研究者を集めた。	他使很多的的研究者聚集起来。	他吸引了很多的研究者。	「集めた」の部分の自動詞／他動詞用法の訳し分けは正しいですか？	“集めた”部分的自动词／他动词的译法是否正确？	0.24	3.00	3.01
2	(1-1) 述部の訳し分け	2	JEG11 1002 彼は標本を集めている。	他在收集标本。		自動詞／他動詞用法の訳し分けは正しいですか？	自动词／他动词的译法是否正确？	0.46	3.29	3.57
3		3	JEG11 1003 彼は論文を集めて本にした。	他把论文收集成册。	他把论文收集成书。	自動詞／他動詞用法の訳し分けは正しいですか？	自动词／他动词的译法是否正确？	0.21	1.54	2.05
4		4	JEG11 1004 彼らは会議室に集まった。	他们在会议室集合。		自動詞／他動詞用法の訳し分けは正しいですか？	自动词／他动词的译法是否正确？	0.60	2.69	3.27
5		5	JEG11 1005 学生が教室に集められた。	学生在教室里集合。		自動詞／他動詞用法の訳し分けは正しいですか？	自动词／他动词的译法及被动句的翻译是否正确？	0.15	1.69	2.11
6	(1-2) 断定文	6	JEG12 0001 この装置はバッテリー駆動だ。	这个装置是电池驱动的。		判断文の訳文は正確ですか？	判断句的翻译是否正确？	0.36	2.63	2.48
7		7	JEG12 0002 手順は左右同一である。	程序是左右相同的。	手续是左右相同的。	判断文の訳文は正確ですか？	判断句的翻译是否正确？	0.37	2.80	3.12
8		8	JEG12 0003 プッシュボタンは簡易操作に最適である。	按钮最适合简单操作。		判断文の訳文は正確ですか？	判断句的翻译是否正确？	0.38	3.01	3.27
9	(1-3) 体言述語	9	JEG13 0001 委員会は彼らの訴えを却下。	委员会否決了他们的上诉。	委员会拒绝了他们的请求。	体言述部の表現がきちんと訳されていますか？	体言谓译语的翻译是否正确？	0.58	4.01	3.44

図1 AAMT機械翻訳文テストセットのサンプル

3. ダウンロードサイトの立上げ

ダウンロードサイトは AAMT の公式サイトのサブドメインとして新しく作成した。図 2 がサイトのトップページであり、図 4 はその URL である。AAMT の公式ホームページからこのページへのリンクが張られる。

今回公開するコンテンツは以下の 3 種類である。

- (1) AAMT 機械翻訳文テストセット (日中)
- (2) AAMT 機械翻訳文テストセット (日英)

※JEIDA テストセットの設問改善版

- (3) 機械翻訳エンジン評価報告

(1)(2) のテストセットには、インターネットで無料公開中の日中、日英各 7 種の翻訳エンジンを対象に、実際にテストセットを使って評価した結果が含まれる (図 1 の I~K 列)。(3) の報告は、評価結果について、翻訳方向、評価者、評価観点等の切り口で比較・検討した結果をまとめたものである。これらテストセット等の公開範囲には、特に制限を設けない。基本的にはインターネットにつながる環境があれば、誰でもダウンロードして利用することができる。以下にダウンロードまでの手順を示す。

1. ユーザ登録を行う (図 3)
2. ユーザ ID とパスワードが登録したメールアドレスに届く



図 2 ダウンロードサイトのトップ

3. ユーザ ID とパスワードでログイン

4. テストセットを選択してダウンロード実行

4. 今後の予定と展望

AAMT テストセット第 1 版は JEIDA テストセット (日英用) の日本語例文を流用したため、中日翻訳の評価に必要なチェック項目の抜けが存在する。例えば、中国語特有の重要な表現である数量詞や介詞の妥当性を問うための例文は、第 1 版には含まれていない。そこで、09 年度の AAMT 調査委員会の活動として日中例文の追加を行うとともに、設問評価で評価者による回答のズレが起きないように設問の見直しを行う予定である。

ダウンロードサイトについては、用途がテストセットの公開に限定されるものではなく、今後 AAMT が様々な言語資源を公開する際に広く利用することが可能である。AAMT 調査委員会の他の活動においても、是非このダウンロードサイトの活用を検討していただきたい。

参考文献

[1] AAMT WG1: AAMT 機械翻訳テストセット構築に向けて, AAMT Journal, No42, pp.18-24, 2008.

[2] Hitoshi Isahara: JEIDA's Test-Sets for Quality Evaluation of MT Systems Technical Evaluation from the Developer's Point of View, Proceedings of MT Summit V, Luxembourg, July 19-22, 1995.

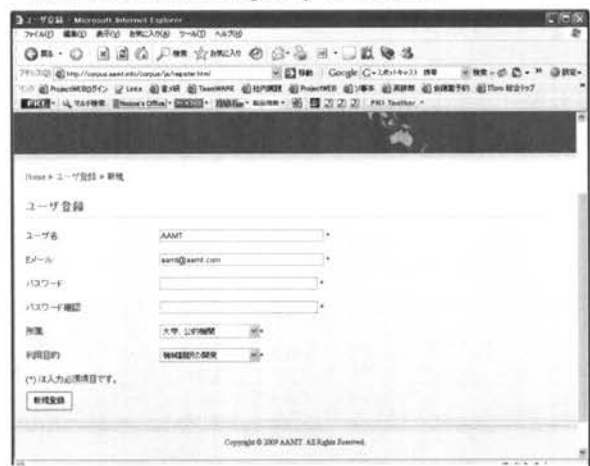


図 3 ユーザ登録ページ

<http://corpus.aamt.info/>

図 4 ダウンロードサイトの URL

英日・日英機械翻訳に関するアンケート結果報告

機械翻訳課題調査委員会ワーキンググループ 2

1. はじめに

AAMT 機械翻訳課題調査委員会 WG2（調査・広報・啓蒙）では、その前身である市場動向調査委員会の時代から調査活動の一環として、年一回インターネット上で一般の機械翻訳ユーザを対象としたアンケートを行っている。AAMT の Web サイトに特設ページを設け、機械翻訳に関するいくつかの設問に回答してもらうという形式である。本稿では、2008 年（2007 年度）のアンケート調査結果について報告する。

2008 年の調査は、従来に比べていくつかの変更を加えた。主な変更点は次の 3 つである。

- 1) 従来「英日」機械翻訳¹を対象にしていたものを「英日・日英」機械翻訳とした。
- 2) 機械翻訳の利用形態（Web 上の機械翻訳サービス翻訳、PC 標準搭載の機械翻訳ソフト、市販の機械翻訳ソフト）毎の設問と、機械翻訳一般に関する設問を明確に分離した。
- 3) 従来は進呈品を翻訳ソフトのみとしていたが、今回は幅広い層から多くの回答を得られるように進呈品を機械翻訳ソフトと図書券（3,000 円分）とした。そして、機械翻訳の必要性を問う設問を追加し、機械翻訳を必要としている回答者のみで分析を行った。

以下に 2008 年のアンケート実施概要を示す。

- アンケート期間
 - 2008 年 2 月 5 日～2 月 25 日
- アンケート進呈品
 - 抽選で 10 名に機械翻訳ソフト、20 名に図書券（3,000 円分）を進呈。
 - 回答者は進呈品として機械翻訳ソフトか図書券かを指定可能。また、進呈品の機械翻訳ソフトは 4 種類で、それらの商品名は回答前に確認可能だが、商品の指定は不可。
- アンケート回答者数
 - 1,495 人（うち、機械翻訳を必要としていると回答した人は 946 人）
 - なお、過去のアンケート回答総数は、2003 年 1,413、2004 年 1,720、2005 年 1,018、2006 年 1,061、2007 年 835。

2. 設問項目

主に以下に示す項目について質問した。

- ・ Web 上の機械翻訳サービスについて（使用経験、利用頻度、使用したことがある翻訳サイト等）
- ・ 機械翻訳ソフトに関して（使用経験がある機械翻訳ソフト、購入価格、これから購入する場合の予算上限等）
- ・ 機械翻訳全般に関して（機械翻訳で利用する言語の方向、機械翻訳を必要とする理由、機械翻訳に追加して欲しい機能、欲しいユーザ辞書の分野、機械翻訳で必要としている言語等）
- ・ 回答者の基本属性について（年齢層、性別、職業、英語力）

¹ 本来英日方向のみで日英方向は対象外だったが、「英日機械翻訳」で日英、英日両方向を指す場合も普通にあるため、回答者がそう理解したかどうかは不明

3. Web上の機械翻訳サービスに関する調査結果

本節では、Web上の機械翻訳サービスに関する調査結果を示す。

図1に示すWeb上の機械翻訳サービスの使用経験の結果では、「知らない」「使ったことがない」は、各年とも各々20%程度にとどまり、「使用したことがある」は約60%に到る。また、図2に示す利用頻度の結果では、「使ったことがある」との回答者のうち、半数は毎月1回以上使用していることがわかる。

図3に使用したことがあるWeb上の翻訳サイトを示す。一般的なポータルサイトのアクセス状況（情報通信白書平成20年版）とは傾向が異なるため、「いつも使用するポータルサイトの翻訳機能を使用する」だけでなく、「翻訳機能を使用するためにそのサイトにアクセスする」という使い方もされていることが分かる。

図1: Web上の機械翻訳サービスの使用経験

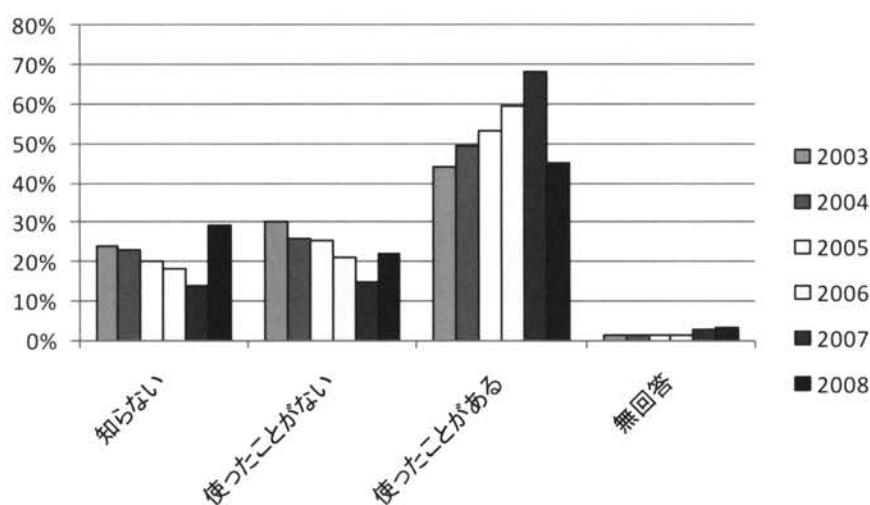


図2: Web上の機械翻訳サービスの利用頻度

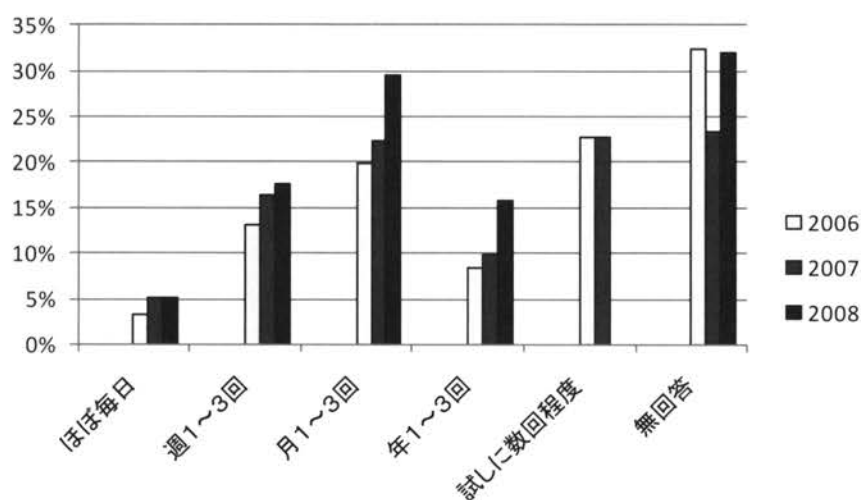
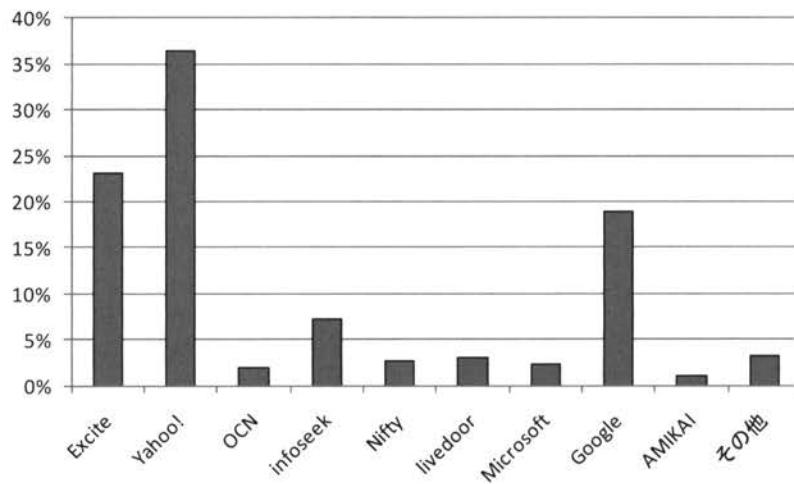


図 3: 使用したことのあるWeb上の機械翻訳サイト（複数回答可）



4. 機械翻訳ソフトに関する調査結果

本節では、市販翻訳ソフトに関する調査結果を示す。

図 4に使用した経験がある翻訳ソフトの結果（横軸は翻訳ソフト名）を示す。3種類の翻訳ソフトの割合が高かった。また、図 5に示す購入価格帯（横軸の価格帯の単位は円）では、パーソナルユース用の3,000円～10,000円が最も多く、ビジネス用の10,000円～20,000円が次に多い。ただし、図 6に示す市販翻訳ソフトをこれから購入する場合の予算の上限（横軸の予算の単位は円）は1万円に境界があり、翻訳ソフト価格に対するユーザのニーズが明確に読み取れる。

図 4: 使用したことのある市販翻訳ソフト (%)（複数回答可）

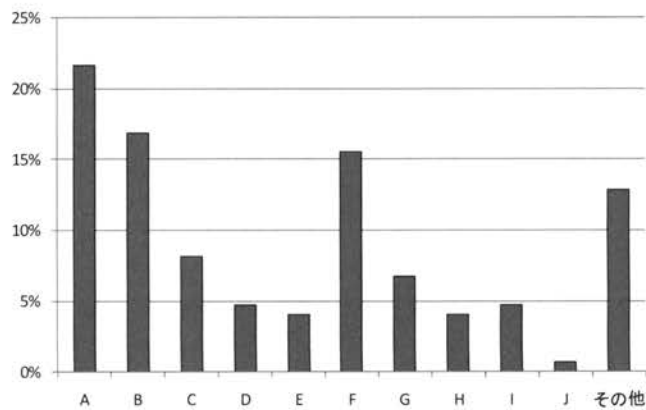


図 5: 翻訳ソフトの購入価格帯

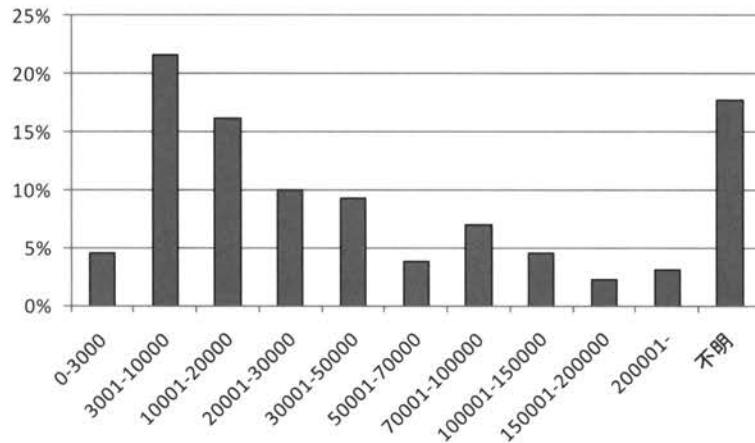
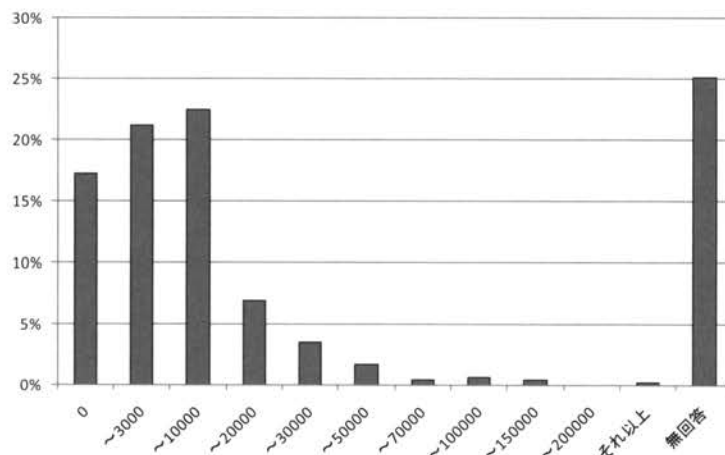


図 6: 市販翻訳ソフトを購入する場合の予算の上限



5. 機械翻訳全般に関する調査結果

本節では、機械翻訳全般に関する調査結果として、Web サービスと機械翻訳ソフトに共通する質問、機械翻訳に追加してほしい機能、欲しいユーザ辞書の分野、機械翻訳で必要としている言語についての調査結果を示す。

5.1. Web サービスと機械翻訳ソフトに共通する調査結果

3種類の機械翻訳（Web 上の機械翻訳サービス、PC 標準搭載機械翻訳ソフト、市販機械翻訳ソフト）に共通する質問事項についての調査結果を示す。

図 7 に翻訳の方向の調査結果を示す。日本語から英語への翻訳のみに利用する人よりも英語から日本語への翻訳のみに利用する人の方が多くことが分かる。ただし、どちらの方向への翻訳にも利用する人の割合も多く、日本語から英語への翻訳のニーズもあることが分かる。

図 8 に機械翻訳を必要としている理由（英語能力別）の調査結果を示す。翻訳者や英語力が高い人が下訳とし

て機械翻訳を必要としていることが分かる。ただし、この割合は、機械翻訳を必要としている人における使い方の割合である。英語力が高い人はあまり機械翻訳を必要としない可能性が考えられるため、図 9 に示す結果に限り、機械翻訳を必要としないと回答した人も含めた分析も行った。機械翻訳を必要としないと回答した人も含めた回答者全体における機械翻訳を必要としている理由の割合を図 9 に示す。図 9 から、英語力が高い人や翻訳者のうち、ある程度の割合で下訳に機械翻訳を必要としていることが分かる。なお、図 8 および図 9 で英語力 0-199 とは TOEIC の点数が 0~199 点程度の英語力を意味する。ここでの翻訳者は英語力とは別の観点による分類であり、英語力と翻訳者の分類は排他関係ではない。選択肢の項目「手がかり」は、「英語がまったくわからないので手がかりが欲しい」を示しており、選択肢の項目「概要」は、「英語は読めるが、まず日本語で概要が知りたい」を示している。

図 10 に利用分野毎の利用率を示す。Web 翻訳サービスによる翻訳では、日常的な連絡のメールなどの「生活情報」での利用率が高く、市販の機械翻訳ソフトでは「自然科学」（工学、物理学など）や「産業」（航空機、自動車、コンピュータなど）での利用率が高いことが分かる。

図 11 に対象文書のタイプ毎の利用率を示す。Web 翻訳サービスは Web ページの翻訳と辞書引きでの利用割合が標準搭載機械翻訳ソフトと市販機械翻訳ソフトより高い。市販機械翻訳ソフトはドキュメントの翻訳での利用割合が Web 翻訳サービスと標準搭載機械翻訳ソフトより高いことが分かる。

図 12 に機械翻訳を必要としている理由毎の利用率を示す。英語がまったくわからないので手がかりが欲しい（手がかり）ため機械翻訳を必要としている人は、市販機械翻訳ソフトに比べて標準搭載機械翻訳ソフトや Web 翻訳サービスを利用する率が高いことが分かる。英語から日本語への下訳作成（英日下訳）や日本語から英語への下訳作成（日英下訳）に機械翻訳を必要としている人は、市販機械翻訳ソフトを利用する率が高いものの Web 翻訳サービスによる翻訳も利用していることが分かる。

図 7: 翻訳の方向

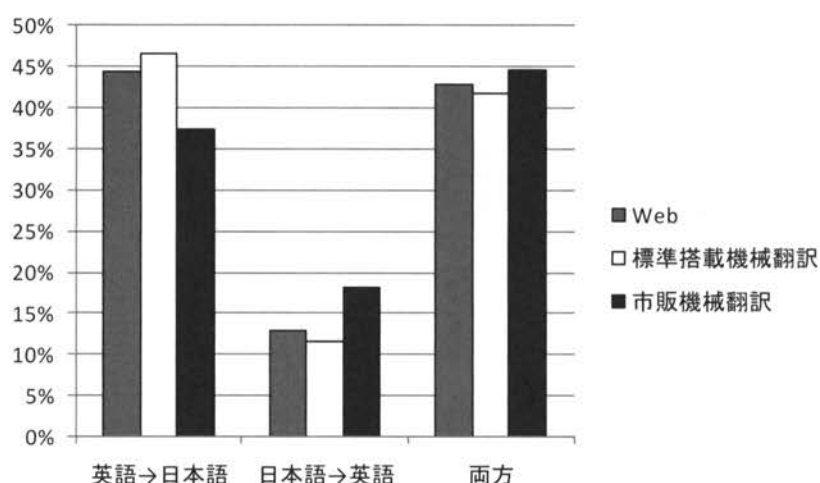


図 8: 機械翻訳を必要としている理由（英語能力別）

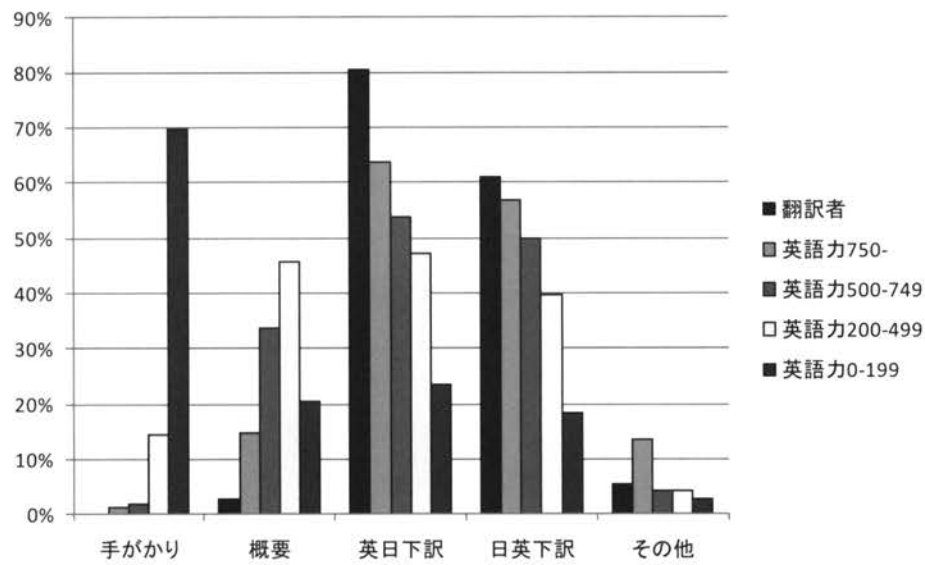


図 9: 機械翻訳を必要としている理由（英語能力別）（機械翻訳を必要としない人も対象）

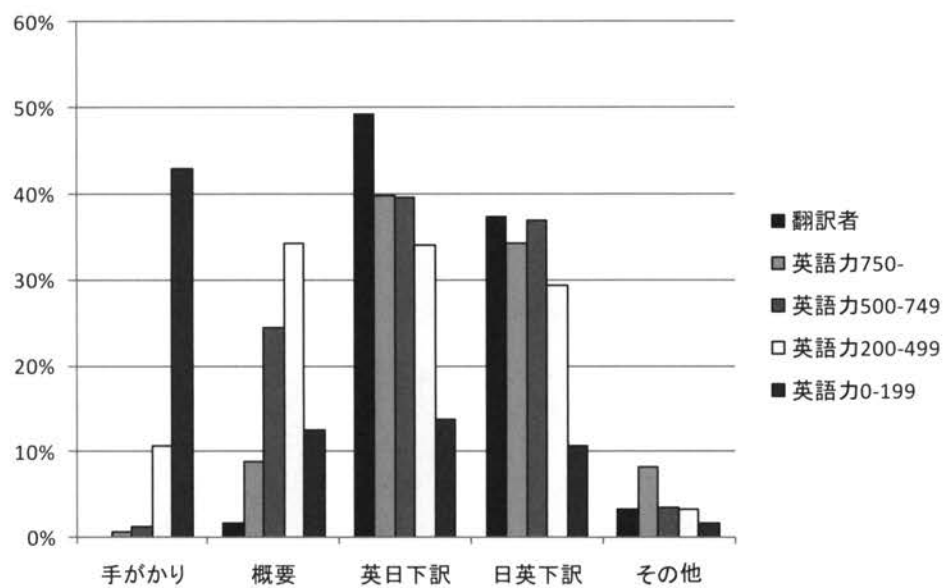


図 10: 利用分野（複数回答可）

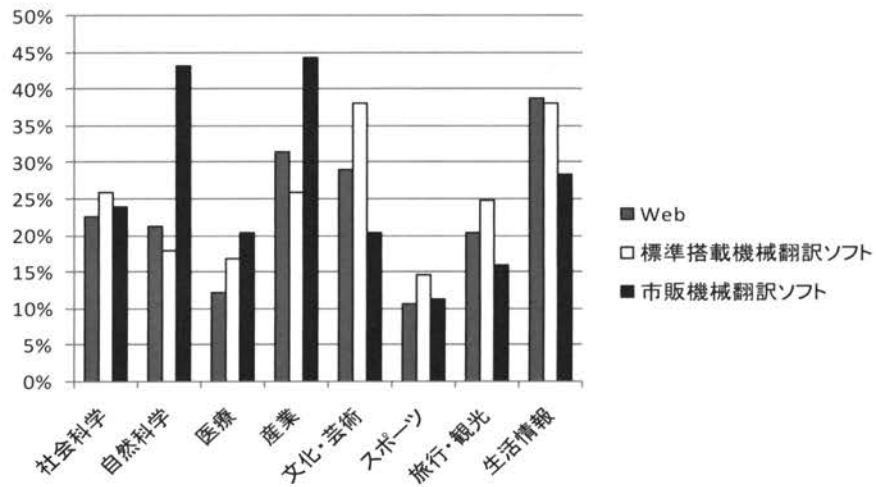


図 11: 対象文書のタイプ（複数回答可）

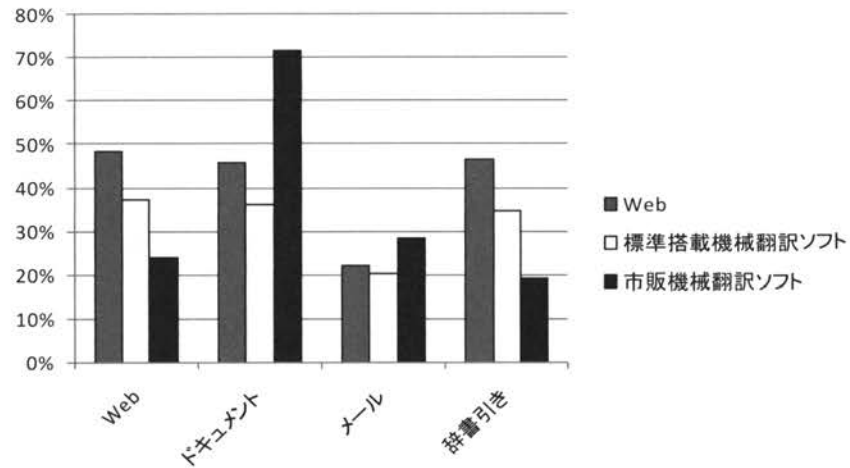
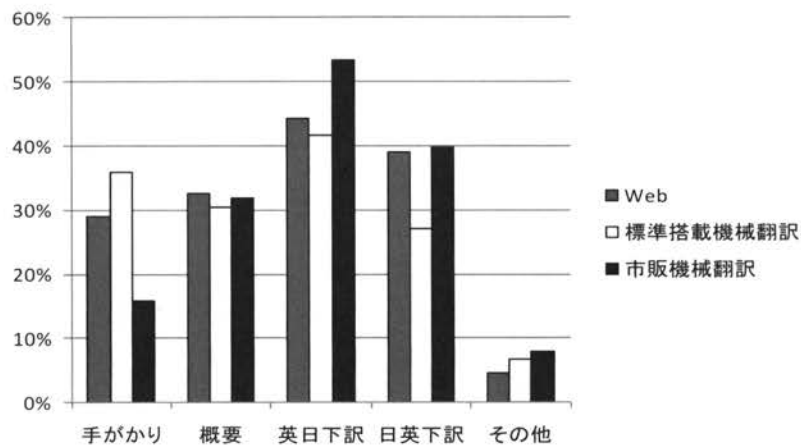


図 12: 機械翻訳を必要としている理由（複数回答可）



5.2. 機械翻訳に追加してほしい機能

機械翻訳に追加して欲しい機能についての調査結果を示す。機械翻訳に追加して欲しい機能に対する回答形式は、最大 200 字までの自由記述とした。のべ 278 件（複数の要望が記載された回答もある。有効回答者数 261 人）の要望が得られた。

表 1 に、内容による分類と要望数の内訳、記載内容の代表例を示す。追加して欲しい機能として多くの要望があったものとして、「音声認識による音声入力機能（31 件）」、「他の翻訳候補を提示して修正できる機能（28 件）」、「新語や専門用語への対応（23 件）」、「翻訳メモリ機能（22 件）」、「辞書引きや辞書登録と機械翻訳との連携機能（11 件）」などがあげられる。この他、翻訳品質を向上させて欲しいという要望も多かった。

表 1: 追加してほしい機能

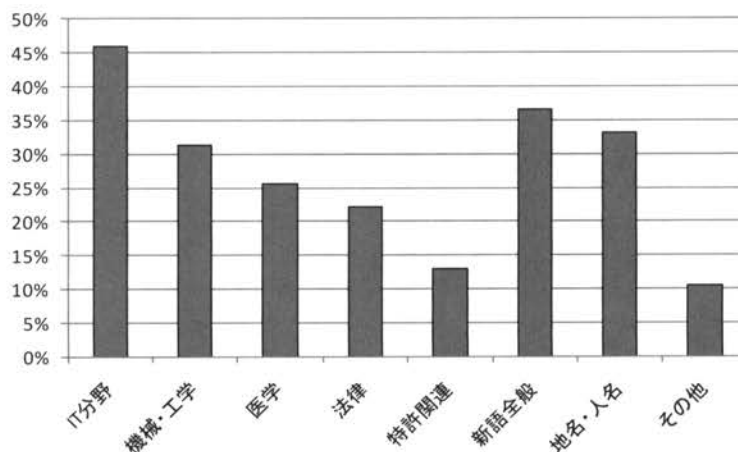
大分類	小分類	件数 (重複有)	例
[品質]	[一般]	33	正しい日本語にすこしでも近づいてほしい
	[文脈、ニュアンス]	9	文脈に応じて訳語を使い分ける程度の機能が欲しい。
	[専門分野]	9	ソフトウェア開発や特許などの専門翻訳機能
	[会話・口語・スラング]	31	状況にあわせた翻訳をできるようにして欲しい。
[機械翻訳結果に対する訂正指示、自動学習]	[学習]	7	翻訳修正を記憶して、次回から反映させる機能
	[他候補の提示]	28	翻訳候補を出して変更できる機能
	[修正再翻訳]	5	構文解釈が間違っていた時に、訳語の順番をマウスでドラッグ&ドロップでビジュアルに入れ替えて、できるだけ正しい翻訳をなおす機能
[辞書]	[一般]	10	新語にも多く対応できる機能
	[専門辞書]	13	専門分野別の辞書、用語集のような機能
	[ことわざ、スラング]	5	語彙を増やして欲しい。
	[利用者辞書作成]	5	用語集を簡単に自作できる機能、手持ちの用語集を利用する機能など。
	[辞書連携]	11	機械翻訳と辞書引き機能や単語登録ソフトとの統合
[例文・翻訳メモリ]	[例文・翻訳メモリデータ]	14	翻訳メモリのような対訳データベース
	[翻訳メモリ機能]	8	対訳がデータベースとして蓄積されていく機能
[文法・スペルチェック]		5	翻訳元言語のスペルチェック、辞書中の訳語有無を事前に調べ、翻訳前に注意を出すなどの機能。
[添削・校正]		8	たとえば、翻訳元の言語を翻訳先の言語に翻訳しやすくなるような下処理をする機能。
[音声]	[認識]	31	音声入力の自動翻訳
[スキャナ連携]		6	紙上の文章の認識から翻訳。
[翻訳支援]		3	名詞、動詞、形容詞、のみ、翻訳をポップアップで出す。文脈から適切な翻訳単語を決定してそれだけ表示してくれる。
[他言語対応]		5	英語以外の言語を英語を経由させないで翻訳するソフト。重訳なら手でやった方が早い。
[原文と訳文の対応関係]		2	原文と訳文との対応関係の表示
[外国語学習支援]		5	英語学習機能、例えば、英単語や英文法をテストできるようなもの
[検索系機能]		3	ブックマークのように自分のページをもって、過去の検索した言葉を保存できる機能
[使いやすさ]		4	日常での使われ方などの外国語初心者にも優しい機能
[価格]		2	安価で操作簡単使い勝手が素晴らしく良い製品に期待します。
[その他]		16	PDF 文章で原文の翻訳文と同じ文章をすぐ作れる機能

5.3. 欲しいユーザ辞書の分野

どの分野のユーザ専門辞書が欲しいかという質問に対する結果を示す。回答形式は複数選択可能な選択肢と自由記述可能な「その他」とした。

図 13 に各ユーザ辞書分野で欲しいと回答した人の割合を示す。専門分野の中では移り変わりの早い「IT 分野」への要望が目立つ。また、「新語全般」、「地名・人名」のように一般文書・ニュースの翻訳品質に影響する辞書への要望も高い。「その他」に記載された分野では、金融・経済、スラングなどの他、各種分野があがっていた。

図 13: 欲しいユーザ辞書の分野（複数回答可）

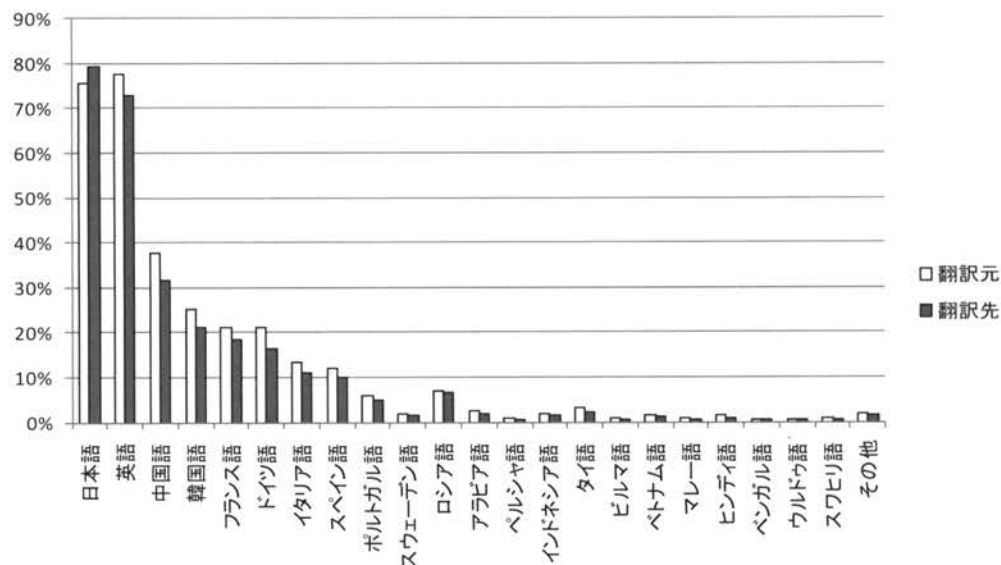


5.4. 機械翻訳で必要としている言語

機械翻訳で必要としている翻訳元と翻訳先の言語についての調査結果を示す。回答の形式は、翻訳元言語と翻訳先言語について、それぞれ独立に、複数選択可能な選択肢と自由記述可能な「その他」とした。

図 14 に結果を示す。英語のニーズが最も高く、次いで中国語、韓国語、フランス語、ドイツ語、イタリア語、スペイン語などのニーズがあることが確認された。

図 14: 機械翻訳で必要としている言語（複数回答可）



6. 回答者に関する調査結果

本節では、回答者の基本属性（年齢、性別、職業、英語力）を示す。

図 15 に年齢、図 16 に性別、図 17 に職業、図 18 に英語能力（横軸は TOEIC スコア）毎の回答者の割合を示す。今回のアンケート調査はこれまでとは異なり、進呈品に従来の翻訳ソフトの他に新たに図書券を加えた。これに起因すると思われる回答者の基本属性の変化が少し現れている。しかし、全体傾向としては大きな変化はなかった。

図 15: 年齢

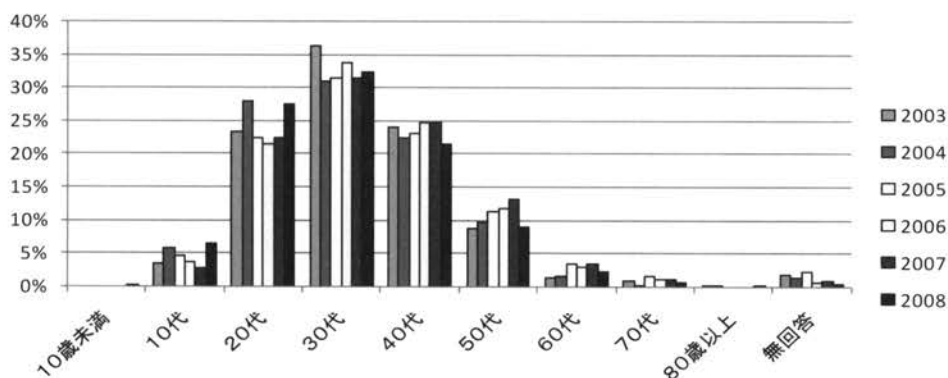


図 16: 性別

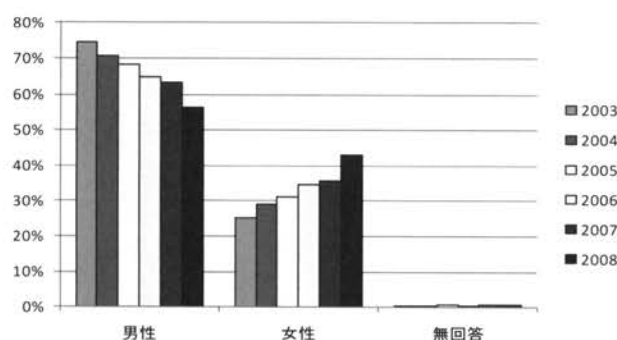


図 17: 職業

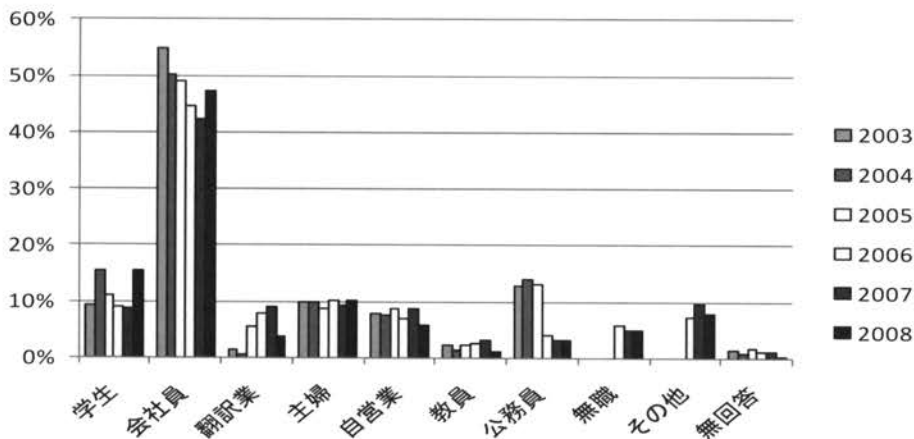
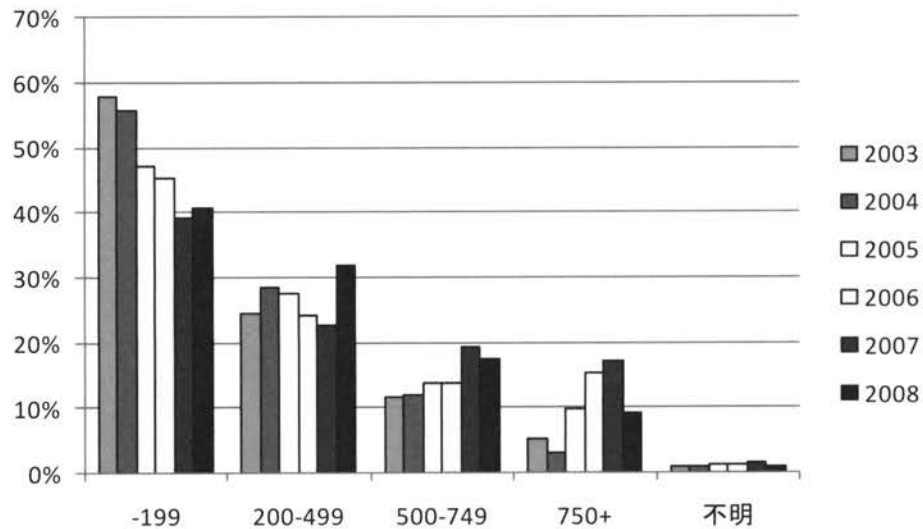


図 18: 英語能力



7. まとめ

機械翻訳に関するアンケートを実施した。本年は新しい設問項目を追加し、下記のような新しい知見を得ることができた。

- ・ 翻訳ソフトの用途は英日方向だけでなく、日英方向の翻訳も需要がある
- ・ 翻訳者または英語能力が高い人にとっては、機械翻訳は下訳としてニーズがある
- ・ 翻訳品質の向上、および「音声認識による音声入力機能」、「他の翻訳候補を提示して修正できる機能」、「新語や専門用語への対応機能」、「翻訳メモリ機能」などの機能の追加が期待されている
- ・ 中国語、韓国語、フランス語、ドイツ語、イタリア語、スペイン語もある程度、機械翻訳のニーズがある

今後、機械翻訳研究者／ベンダーに上記のアンケートに基づいた研究・開発・販売を期待すると共に、今後も機械翻訳技術利用の現状やニーズに係る調査を続け、機械翻訳をより役立つものとしていきたい。

ユーザー辞書共通フォーマットUTX

機械翻訳課題調査委員会ワーキンググループ3

「翻訳ソフトは役に立たない」
そう思っていないですか？

UTX の利点

- 翻訳ソフトの精度を大きく向上
- 翻訳資産を共有・再利用
- 単語を調べる時間と労力を削減

概 要

UTX (Universal Terminology eXchange) とは、アジア太平洋機械翻訳協会 (AAMT) が策定した、翻訳ソフトのユーザー辞書共通フォーマットです。オープンな仕様として標準化を目指しています。

AAMTは、機械翻訳の研究開発者、製造販売者、利用者の三者から構成される団体です。

特 徴

UTXは、対訳形式の用語の知識を共有することで、翻訳ソフトの翻訳精度を大きく向上します。ユーザーの視点から、シンプルで「作りやすく、使いやすい」ことを目指しています。

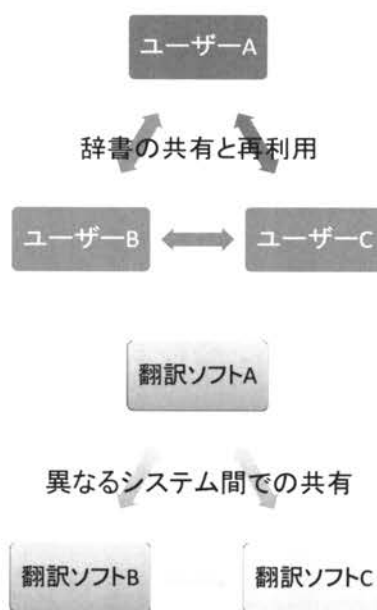
異なるメーカーの翻訳ソフトで、同じ辞書が使えます。また、翻訳ソフトをまったく介さない用語集としても使えます。

なぜ使うのか

UTXを使えば、翻訳ソフトのユーザー辞書をかんたんに共有・再利用できます。翻訳ソフトがうまく訳せないのは、ある語や句をどう訳すべきかという**翻訳知識が不足している**からです。UTXは、翻訳知識をユーザー辞書として補うことで、翻訳ソフトの翻訳精度を大きく改善できます。

また、これまで翻訳ソフトのユーザーは、細切れのユーザー辞書を個人で作成しても、有効活用でき

ませんでした。単純なテキスト ファイルでも、形式が揃っていないと共有や再利用は困難です。しかし、UTXのような単一の規格を使えば、異なるメーカーの翻訳ソフトなど、さまざまなツールで広く共有し、すばやく再利用できるようになります。



種 類

以下の2種があります。

UTX-Simple もっともシンプルなタブ区切り形式。

UTX-XML 多くの情報を含められるXML形式
(仕様を検討中)。

だれが使うのか

翻訳ソフトのユーザー、翻訳者が使うことを想定しています。作成や使用にあたって、文法や言語学などの高度な専門知識は不要です。複数形、名詞や動詞など品詞の区別など、最低限の情報のみで作れます。

どのような分野で使うのか

IT、医療、法律、工学など、一定の専門性がある分野の翻訳であれば、どのような分野でも使えます。

どんな語を含むのか

基本的に、UTX辞書は特定分野の**専門用語のみを含みます**。多くの場合は名詞、特に複合名詞がほとんどです。たとえば「XML declaration」のような語は、辞書に登録することではじめて「XML宣言」などと正しく訳せます。「window=窓」のような基本的な語彙は、翻訳ソフトのシステム辞書に含まれているため含めません。翻訳ソフトに購入時に付属しない、きめ細かい対訳の情報を集約して共有・再利用することにより、翻訳精度を向上できます。

文(センテンス)については、一種の「単語」として扱うのが適切な場合にのみ含めることもできます。ただし、原則としてUTXは、翻訳メモリー(文単位の対訳データベース)とは区別して使います。

多言語に対応しているか

UTXの文字コードは、Unicode (UTF-8) であり、UTF-8で扱える言語はすべて扱うことができます。ただし、1つのUTX-Simple辞書には、起点言語A(原文)から目標言語B(訳文)への方向の訳のみが含まれます。BからAへの辞書は、別の辞書になります。

どうやって作るのか

UTX-Simpleは、**テキスト エディタ**や**表計算ソフト**で**かんたんに編集**できます。変換ツールで、UTX形式から各社の翻訳ソフトの形式に変換して使用できます(現在はUTX-Simpleのみ)。

変換ツールの入手先

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

さまざまな形式からUTX形式への変換をするツールも開発中です。

オープンソース開発者・翻訳者の皆様へ—UTX形式で用語を公開・共有しませんか？

互いの用語を公開・共有することで、ソフトをすばやく多言語化し、世界中の人に使ってもらえます。

UTXメーリング リスト

どなたでもUTXの議論にご参加頂けます。

「UTX メーリング リスト」でウェブを検索してください。

参考資料

大倉清司他(2008)「共有ユーザー辞書仕様UTXの現状と今後の展開」言語処理学会第13回年次大会(東京)

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

実 例

(UTX-Simple 1.0版に基づくサンプル)

#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	

#はコメント行を表す。

●1行目(コメント行):辞書の基本情報。各項目はセミコロンと半角スペースで区切る。UTX-SはUTX-Simpleを示す。

#UTX-S <バージョン番号>; <起点言語>/<目標言語>; <最終更新日付時間>; copyright: <著作権者>; license: <ライセンス>; <追加の情報> (必要に応じて追加)

●2行目(コメント行):属性の規定。各項目はタブ区切り。

上記の例では

#<原語> <訳語> <原語品詞> <原語複数形>

UTX 辞書（用語集）作りのポイント

- 辞書の分野を定義する
- 固有名詞以外は太文字で始めない
- 原形を記載する(市販辞書の見出しの形式)
- 原語、訳語以外の情報はコメント欄に記す
- 原語に対応する訳語は、最良のものを1つだけ選ぶ

免責事項

AAMT および AAMT 参加者は、UTX、UTX-Simple、UTX-XML の仕様(以下 UTX 仕様)を現状有姿のまま提供するものであり、UTX に関する一切の事柄を法的に保証せず、また UTX 辞書を使用した結果に対しても一切の責任を持ちません。また UTX の各種ツールの作成者は、変換の正確さについて保証はしません。UTX 仕様は、ユーザー各位の責任においてご使用ください。

AAMT は UTX 仕様の著作権を放棄していませんが、UTX 仕様は公開されており、どなたでもご使用頂けます。辞書の作成者は、商業使用を含め、その辞書のライセンスを規定できます。ただし、辞書の基盤となるデータの著作権については、個別にデータの提供元に確認してください。AAMT および AAMT 参加者は、AAMT が作成した辞書を除いて、個別のユーザーが作成した辞書の著作権の正当性について確認をせず、保証もしません。

ユーザーは、UTX 仕様に基づいて作成された辞書を、その辞書のライセンスに応じて使用できます。実際に配布されている辞書のライセンスはそれぞれ異なりますので、UTX ファイルの先頭部分に含まれる使用条件をご確認ください。

機械翻訳課題調査委員会

共有化・標準化ワーキンググループ メンバー (順不同)

山本ゆうじ	秋桜舎
伊藤肇	株式会社インターグループ
村田稔樹	沖電気工業株式会社
鈴木次良	株式会社クロスランゲージ
Francis Bond	独立行政法人 情報通信研究機構
島津美和子	東芝ソリューション株式会社
谷口裕子	東芝ソリューション株式会社
大倉清司	株式会社富士通研究所
加藤マイケル孝仁	ラーニング コンサルタント

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

問い合わせ先：aamt-info@aamt.info

2009年5月版

機械翻訳研究プロジェクトリンク集

AAMT 機械翻訳課題調査委員会 WG2

- 本ページでは、AAMT 機械翻訳課題調査委員会「調査・広報・啓蒙」ワーキンググループの活動成果として、「機械翻訳研究プロジェクト」をリンク集として紹介しています。
- 複数の組織からなるプロジェクトの場合、「主な参加団体」として代表的な一組織のみ掲載しています。
- AAMT は以下のサイトにアクセスしたことによる損害を一切補償しません。
- 本リストに掲載されていない「機械翻訳関連研究プロジェクト」をご存じでしたら、AAMT事務局までご連絡いただければ幸いです。
- 順不同

日本のプロジェクト

名称	機械翻訳ユーザー辞書共通フォーマット制定活動
主な参加団体	アジア太平洋機械翻訳協会
URL	http://www.aamt.info/japanese/utx/
期間	2006 年～

名称	MASTAR プロジェクト
主な参加団体	情報通信研究機構
URL	http://www2.nict.go.jp/pub/whatsnew/press/h20/080509/080509.html
期間	2008 年～5 年間

名称	第7次欧州研究開発フレームワーク計画
主な参加団体	名古屋工業大
URL	http://www.emime.org/
期間	2008 年～3 年間

名称	おもてなし de 開国プロジェクト
主な参加団体	情報通信研究機構
URL	http://www2.nict.go.jp/pub/whatsnew/press/h20/080602/080602-2.html
期間	2008 年～

名称	NTCIR-7 *Patent Translation*
主な参加団体	国立情報学研究所・情報学資源研究センター
URL	http://research.nii.ac.jp/ntcir/
期間	2008 年～

名称	NHK自然言語処理研究プロジェクト
主な参加団体	NHK放送技術研究所
URL	http://www.nhk.or.jp/strl/group/ningen_joho/ningen_joho08.html
期間	2008 年～

名称	言語グリッド
主な参加団体	情報通信研究機構
URL	http://langrid.nict.go.jp/jp/
期間	2006 年～2008 年

名称	A-STAR
主な参加団体	ATR 音声言語コミュニケーション研究所
URL	http://www.atr.jp/html/topics/press_061114_j.html
期間	2006 年～

名称	日本法令英訳プロジェクト
主な参加団体	内閣官房室
URL	http://www.kl.i.is.nagoya-u.ac.jp/told/
期間	2006 年～

名称	日中機械翻訳プロジェクト
主な参加団体	情報通信研究機構
URL	http://www2.nict.go.jp/x/x161/research_project/project.html
期間	2005 年～2011 年

名称	椎茸プロジェクト
主な参加団体	東京大学
URL	http://panflute.p.u-tokyo.ac.jp/~kyo/transaid.html
期間	2005 年～2008 年

名称	ライフサイエンス辞書プロジェクト
主な参加団体	京都大学
URL	http://lsd.pharm.kyoto-u.ac.jp/ja/about/about_lsd/index.html
期間	1996 年～

名称	鳥バンク
主な参加団体	鳥取大学
URL	http://unicorn.ike.tottori-u.ac.jp/toribank/
期間	—

名称	音声言語メディア処理プロジェクト
主な参加団体	NTT サイバースペース研究所
URL	http://www.ntt.co.jp/cclab/ccsouken/sp/p_onsei.html
期間	—

海外の著名なプロジェクト

名称	The EuroMatrix Project
主な参加団体	Edinburgh University
URL	http://www.euromatrix.net/
期間	2006 年～

名称	Prospect Project
主な参加団体	Dublin City University
URL	http://www.nclt.dcu.ie/mt/prospect.html
期間	2005 年～2009 年

名称	LOGON Project
主な参加団体	Oslo University
URL	http://www.emmtee.net/
期間	2002 年～

名称	GALE
主な参加団体	DARPA
URL	http://www.darpa.mil/ipto/programs/gale/gale.asp
期間	—

名称	TC-STAR
主な参加団体	Istituto Trentino di Cultura
URL	http://www.tc-star.org/
期間	—

翻訳ソフト活用のためのリンク集

AAMT 機械翻訳課題調査委員会 WG2

- 本ページでは、AAMT 機械翻訳課題調査委員会「調査・広報・啓蒙」ワーキンググループの活動成果として、「翻訳ソフトの活用に役立つサイト」を紹介しています。
- AAMT は以下のサイトにアクセスしたことによる損害を一切補償しません。
- 本リストに掲載されていない「翻訳ソフト活用に役立つサイト」をご存じでしたら AAMT 事務局までご連絡いただければ幸いです。
- 順不同

翻訳ソフト・ツールの活用法紹介

翻訳しやすい文章の作り方	http://www.crosslanguage.co.jp/oyakudachi/edit_je/
知って得する豆知識	http://pf.toshiba-sol.co.jp/prod/hon_yaku/service/case/tips/index_j.htm
翻訳のコツ	http://software.fujitsu.com/jp/atlas/hint1.html
Yahoo!翻訳のコツ	http://honyaku.yahoo.co.jp/guide/kotu_honyaku.html
秋桜舎	http://transpc.cosmoshouse.com/index.htm
バベルMT(機械翻訳)研究会	http://www.babel.co.jp/mtsg/wiki/wiki.cgi
英文法に興味を持たせる インターネットの自動翻訳サイトの活用法	http://www.nicer.go.jp/eltt/english/high/en_h05_05.html
クマゾーの何でも実験室	http://bgrass.halfmoon.jp/translation/
翻訳サービス利用ガイド	http://xn--vcki1fxhz200bq2m.eeyo.biz/
翻訳ソフトで遊ぼう 上手な利用法	http://abc-abc.net/soft.html
機械翻訳 しっかり入門	http://www.marlin-arms.co.jp/jpn/column/mt101/index.html
翻訳ソフトのススメ	http://aucds.com/
翻訳ソフトをつかって正確な英文を書くコツとは？ 翻訳ソフトで英文を書く！	http://allabout.co.jp/study/english/closeup/CU20020415/index.htm
ちょっとした工夫で翻訳ソフトは使いやすくなる！ 翻訳ソフトで英文を読む！	http://allabout.co.jp/study/english/closeup/CU20020412/
機械翻訳を活用するための 100 のヒント	http://www.ispl.jp/~oosaki/research/tips-translation/

翻訳に関するブログ・メールマガジン

秋桜舎ブログ	http://cosmoshouse.air-nifty.com/main/
TranRadar 電子辞書 SHOP blog 「読んで得する翻訳情報マガジン」	http://tran.blog.shinobi.jp/
英日翻訳ソフト＝チューニングすればこんなに良くなる	http://archive.mag2.com/0000185648/index.html

複数翻訳サイトの比較

10の翻訳エンジンから一括翻訳 翻訳くらべ	http://7go.biz/translation/
Crosstranslation	http://sukimania.ddo.jp/trans/trans.php

翻訳ソフト用無料辞書ダウンロード

UPF ビジネス辞書	http://www.aamt.info/japanese/upf/dictdataidx.html
J 北京用	http://www.kodensha.jp/olympic/
The 翻訳用	http://pf.toshiba-sol.co.jp/prod/hon_yaku/service/case/dctnr/index.j.htm
ATLAS 用	http://software.fujitsu.com/jp/atlas/update/start_psnl.html
翻訳の王様用	http://www-06.ibm.com/jp/software/internet/king/download/index.html

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」、第一回の No.41 のスタートから早くも一年半を迎えようとしています。

第五弾の今号では、法人会員 2 社、個人会員 3 名の皆さまからのご寄稿をいただきました。

学識経験者として、研究者として、翻訳者として、日本で、或いは海外で、機械翻訳、翻訳、教育と各分野において、会員各位のこれまでの歩み、その中での貴重な経験談をご紹介いただいております。

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

（社）日本翻訳連盟／Japan Translation Federation

自己紹介

当連盟は 1981 年に設立され、1990 年に通商産業省（現経済産業省）より許可され社団法人となりました。団体の目的は、翻訳事業に関する調査、研究、研修会、人材育成等の実施及び翻訳関連の国際会議等への参加を通じて、翻訳事業の発展を図り、わが国経済発展に寄与することです。円滑に事業を遂行するために国際翻訳家連盟（FIT）、American Translators Association (ATA)、アジア太平洋機械翻訳協会（AAMT）、日本工業英語協会等と相互会員関係をもち協調関係を築いてきました。

AAMT に入会した理由は、翻訳の高品質・コスト削減・効率化を極めて行くには、従来の翻訳者による翻訳に加えて、必要とされる改善点も多々あると思うが、翻訳ソフト・翻訳メモリ・その他翻訳支援ツールの活用を前向きに考えていく必要があると考えたからです。

機械翻訳および翻訳業界に期待すること

翻訳には 2 種類あると思います。一つは大まかなあらすじが分かれば良い。もう一つは商品価値が求められる文芸翻訳と産業翻訳です。前者の場合は現状でも翻訳ソフトがかなり使えると思います。文芸翻訳の場合は翻訳者に文才がないと難しいですし、翻訳ソフトでは困難だと思います。私達は後者とくに産業翻訳での機械翻訳に注目しています。産業翻訳の場合、語学力はもちろん高度な専門知識も要求されます。翻訳ソフトも分野毎個別にユーザーとコミュニケーションをとりながら、きめ細かく開発していく必要があると思います。

現在市販されている翻訳ソフトは個人ベースでうまくお付き合いしながら、ビジネス或いは翻訳に役立ていけるとは思いますが、実務翻訳者や翻訳会社の現場で利用するにはまだまだ改善の余地があります。

AAMT へのご要望

AAMT の事務局は東京から京都に移転されましたが、翻訳ソフト会社および翻訳を必要とする多くの企業が首都圏に集約しているので、AAMT の活動拠点も首都圏におく必要があると思います。そうすれば AAMT の活動範囲および実績ももっと向上するのではないのでしょうか。

翻訳ソフトの開発担当者とユーザー間のコミュニケーションがもっと必要だと思います。それをベースに AI の導入等も進めたら良いと考えます。

翻訳ソフトになじみやすい形にする原文校正ソフトを開発していただきたいと思います。基本的に原文が悪いと人間も機械も良い翻訳ができません。

会員名

山根ドキュメンテーション／Yamane Documentation Corporation

自己紹介

東京の西新宿に拠点を置いております翻訳会社 株式会社山根ドキュメンテーションにおきまして、IT ソリューション事業部の統括マネージャーを務めております吉沢と申します。ローカライズ業務全般、ライティング、ライフサイエンス系、特に医学、薬学、更に CSR、アニュアルレポート、金融系で実績を積んでいる翻訳会社です。AAMT に入会させていただいてから早 1 年近くたちます。多面的な多くの情報が発信され、皆様の活発な活動に日々刺激を受けている次第です。

機械翻訳および翻訳業界に期待すること

機械翻訳の出力に対する手動の修正を行うという品質管理の業務におきまして私が時折考えておりますことは、「機械翻訳は真にユビキタスとなりえるか？」というテーマです。

「ユビキタス」とは、つまり「機械翻訳であることを意識する必要もなくなる」という意味のつもりですが、たとえば、Web のようなインタフェースを通して機械翻訳された契約書をエンドユーザーが読み、その結果、契約の意味がよく理解できなかったり、不幸にも誤解してしまったりした場合、まるで従来の固定電話（つまりユビキタスな通信手段）が故障したときに電話局に問い合わせたような程度のことで済むものなのか？ということです。

しかしながら、私どもの実際の翻訳現場におきましては、日々、機械翻訳をどれだけ「人間が翻訳したものに近づけるか、区別できないほどの品質に上げるか」という作業に追われているのが実情です。今後は、私どもの現場における「機械翻訳実務」につきましても、AAMT のメンバーのみなさまと情報交換をさせていただき、業界の発展に寄与できればと切に願っております。

山根ドキュメンテーション IT ソリューション事業部
吉沢茂夫

個人会員（敬称略・50 音順）

会員名

仲佐 博裕（NKS 技術士事務所（社）日本工業技術振興協会）／NAKASA Hiroyasu

自己紹介

何十年後に偉大なるノーベル賞物理学者が続出することになるとは夢にも思わず、大学院時代の余りにも役に立たない中間子発生を目指した粒子加速器の研究に見切りを付けて、少しはましな実務的な研究ができることを期待して、(財)電力中央研究所に入所し水力・火力・原子力と諸研究歴を積み重ねて 30 有余年、しかし結局は、特に目立つような研究成果も得られぬままに何時か 60 歳定年を迎え、退職してからも既に 10 有余年も経た老科学技術者？というのが自己紹介です。その研究人生では、1985 年の筑波万博以降、陰に陽にパソコンを利用した機械翻訳のソフト開発とその利用に明け暮れてきました。その中には、開発当初の機械翻訳技術を導入するためにメーカーを支援したスポンサー経験、その後機械翻訳の実用化を目標としてソフト開発をしたデベロッパー経験、さらに定年退職後、開発した機械翻訳ソフトを技術情報文献等の翻訳実務に有効に利用してきたユーザー経験、の 3 通りの経験をささやかながら身を持って味わってきました。これらの諸経験について何か会員の皆さまにお役に立つことでもあればと思い、本文の筆を執りました。

機械翻訳との関わり

1985 年の夏の暑い盛りに、中学生だった息子にせがまれて筑波万博の見学に出かけた折、富士通パビリオンの機械翻訳 ATLAS の展示の人気の驚きでしたが、大学時代のその道の友人達が開発に熱中していた大形計算機の威力に過ぎないのだろうとしか思いませんでした。一方、当時ようやく世に出始めたパーソナルコンピュータを使っても機械翻訳が可能であるという話を、電力中央研究所内の仲良しクラブである PC 同好会の席上で友人から聞き驚かされました。その話の出所は、故・上野俊夫氏の名著「パーソナル・コンピュータによる機械翻訳プログラムの制作(1986)」でした。よくもまあ、この時代にここまで機械翻訳の仕組みの種明かしをされたものか、と畏敬の念を抱いたものでしたが、多分、世の多くの諸氏もこの書物を基にしてパソコンによる機械翻訳に熱中されたことと思われます。筆者もその中の一人に過ぎませんが、それにしてもその後随分と長い間、急速な発展を遂げたパソコンの進歩にも後押しされましたが、パソコン翻訳とお付き合いすることになるとは思ってもいませんでした。パソコンに関しては、それまでの 10 数年間、非破壊検査関係の研究開発に従事していた際に、NEC パソコンの国民的名器 PC8001 を計測監視用に利用してきておりましたので、早速に N88BASIC を使った上野氏のプログラムを導入し、いろいろ改良を加えながら機械翻訳の仕組みの解明を楽しんでおりました。

筑波万博での機械翻訳の人気を目の当たりにした電力中央研究所のトップ筋からの指示で、富士通の機械翻訳システム ATLAS-1 を電気事業界でも導入しようとして、1986 年の秋から研究所本部に「機械翻訳実用化検討会」が発足することになりました。その会では経済関係と原子力関係の諸文献について機械翻訳性能を評価することになり、原子力部門でぶらぶらしていた筆者もその筋の専門家として参加し、機械翻訳結果を評価して性能向上に協力することになりました。しかし、富士通のご関係者の方々には大変申し訳けないことながら、その評価結果は散々なものでした。例えば、単純な "It is ... to ..." 構文などもまともな訳ができず、それらのソフト改

良開発には半年単位の時間が掛かるという回答を受けたものでした。ただ後に自分もソフト開発をする身になって、逆の立場から同様な批判を受けることになり、当時の関係諸氏のご苦勞が身に染みて判るようになっていきました。大きな組織で多くの顧客相手の業務では、なかなか小回りの効く即効性ある対応は困難なことであつたであらうと思っています。

ただ、PC 同好会でお遊びで楽しんでたパソコン機械翻訳では逆に、小回りの効く翻訳処置は容易に行うことができました。その成果は「盲蛇に怖じず」的ではありましたが、たまたま論文募集中の横浜開催の知識情報国際集会 ICIK87 で講演発表することにしました。この講演発表は、今でも考えられないことですが、正式な研究題目の設定もせずに PC 同好会仲間と連名でどさくさ紛れに行つたものでした。この国際集会では機械翻訳関連の論文発表が 2 件あり、もう 1 件は、当時京都大学の長尾真教授による本格的な機械翻訳の開発成果のご発表でありました。内容的に格段の差がある発表内容ではありましたが、機械翻訳プロセスを判り易く説明していたためか、筆者の発表には今でも忘れ難い質問を幾つか会場から頂きました。その一つは、米国人の質問者からのものでしたが、当方は英語から日本語への一方向の機械翻訳を発表したのにも関わらず、質問内容は「英日の翻訳結果から日英翻訳すれば、きちんと元の英文に戻れるのか？」というものでした。

この米人からの質問に対して、「現発表は英日翻訳のみであるが、将来の課題として日英翻訳、更に英日と日英での同文復帰も検討したい」と一応その場を繕つた返答をしましたが、これがその後の機械翻訳の泥沼研究に入り込むきっかけになってしまいました。この双方向翻訳課題の解決は、その後の約 10 年間の研究開発作業の後でも達成できませんでしたし、多分、その時点から 20 年余も経た今現在でも、翻訳の質と量のレベルに依るかも知れませんが、世の一般の機械翻訳システムにとってもなかなか難しい永遠の共通課題であらうと思っています。

機械翻訳研究開発の辛い点

勤務していた研究所は、スポンサーである電気事業界に貢献可能な新しい技術があれば、それを評価して速やかに導入する道筋を付けるか、またそうでなければ新しい技術をユーザーサイドから自ら研究開発することを責務としていました。しかし当時の機械翻訳技術は、多大な翻訳実務を抱えている電気事業にとって、すぐに使えてユーザーとして満足できる状況ではないものと判断され、ユーザーの視点から自ら機械翻訳システムを開発して、原子力などの専門分野や特定の対象物を具体的に持っている人達が扱いこなせるようにすることが要求されました。

パソコンでの機械翻訳遊びに夢中になっていた筆者は、専門の原子力分野での限られた対象の翻訳ならばと、これも「盲蛇」的に簡単に考えて、窓際族の優雅な研究対象が見つかったものと思い、1988 年秋に新規の研究テーマを設定することにしました。当時の考え方は、後に取材を受けた電気新聞の月曜インタビュー記事「翻訳は探偵小説を読むような面白さがあります」と大胆にも大げさに見得を切つて始めたものでした。それでも何とか定年までの約 10 年間も機械翻訳のテーマでお給金を頂いて、サラリーマン人生だけは全うすることができました。

この開発研究では、当時ようやく世に出た 32 ビットのパソコンと、英文資料を光学的に読み取つてパソコンに入力するための OCR・イメージスキャナーとを購入する予算獲得から出発しました。接続詞・関係代名詞などを多く含む長文や、特殊な文字・記号を含むクセのある英文、等にも対応可能なように原文入力が簡便に行えかつ訳文もまともに出力できるようにする作業と共に、特殊な専門用語や表現法にも対応可能な辞書の整備や、諸文法規則の追加・修正を進める作業、の連続でしたが、これらの作業は結構面白く楽しみなものでした。しかし、このような研究開発過程は、綺麗事の学術論文を書いたり、威張れるような研究成果を鼓舞できるような代物で

はなく、遅々として埒らない泥沼の研究作業の連続でもありました。周りからは、年度末や期末ごとに研究成果を問われ、次の予算獲得のストーリー作りに苦労し続けたものでした。

特に当時のパソコン翻訳のソフト開発で厳しい問題点は、当初からの使用装置が 32 ビットマシンであったのにも関わらず、16 ビットの OS ソフトしか使えなかったことでした。長文の翻訳対象や辞書内容の作成ではすぐに容量制限にぶつかってしまい、その回避処置に苦労したものでした。この問題は、1990 年代後半以降の 32 ビットパソコンソフト(WINDOWS 95 以降)の急速な進展と普及で解消することになりますが、残念ながら筆者の研究期間の終了後のことでありました。そうこうする中に定年までの持ち時間も乏しくなり、外部業者に委託して英日ソフトについてのみウィンドウ化対応をしたところで定年退職となり、苛酷な研究開発の泥沼と苦労からは抜け出せることになりました。

原子炉故障トラブル情報の機械翻訳

勤め先の研究所は、米国電力研究所(EPRI)との協定により、米国の原子炉故障トラブルに関する多数の英文書類が衛星通信回線で送られてきており、その英文翻訳の担当部門があり、そこでは翻訳にあたっての労力と時間の節約や外部翻訳委託費の節減のために、英日機械翻訳の実用化を図る研究ワークが進められていました。その中には、既述の ATLAS-1 の実用性能向上の努力が進められていましたが、技術的・経済的な限界のために適用柔軟性が期待通りでないと評価され、その際の問題点を解明し解決方法を探求するためのクロスチェックとして、ユーザー自らが翻訳ソフト内容の改良を自由に行えるという特色があるパソコン翻訳の実用性が評価されることになりました。

このため、筆者のパソコン翻訳を原子力情報の実用翻訳に適用し、多様な原文に対応できる機械翻訳性能や操作性能の更なる改善と、適用可能な辞書の整備を図ることになりました。その実運用化の際、米国の原子力現場の作業労働者からの生に近い間違いだらけの英文やラクセのある長文・悪文やらが満ち溢れている英原文の扱いには苦労しました。関係部門からは、「たとえ語句のスペルミスや文法的に間違った英文に対しても、機械翻訳の方できちんと訳してくれなければ困る」といった無茶苦茶な要望にも対処しなければならない苦労もありました。簡単な辞書の作成・整備はアルバイト女性に頼めても、特殊な専門用語や表現法に対応できるような辞書の整備や、上記のような長文・悪文の英原文に対する諸文法規則等の追加・修正は、どうしても専門分野に通じかつ対象外国語と日本語の両方に通じた開発者の能力が必要であることを痛感しました。

モンゴルにおける日－蒙－英の機械翻訳

定年退職の直前に、当時のモンゴル技術大学から学長名で日－蒙－英の機械翻訳の開発についての国際協力依頼の私信を受け取りました。退職前後の暇に任せてモンゴル語について少し勉強をしてみたところ、日本語のルーツとも言えるような共通点が多々あること、従って比較的容易にパソコン機械翻訳の適用が可能であることが判りました。その旨をモンゴル側へ返答したところ、日本の名立たるメーカーや大学・研究機関に送信したのに返事は筆者のみであったとのことで、モンゴルの国を挙げての祭典ナードムの時期にモンゴル国への招待を受け、大学での講義と引換に、周辺の景勝地観光やナードム見物など、定年退職後の思い出深い良き体験を得ることができました。

その後の機械翻訳開発の国際協力の進展につきましては、定年後の個人プレーは人・物・金の点で困難であり、国際研究プロジェクト立ち上げなどいろいろと努力はしましたものの結局は物にならずに、当時新しく組織化さ

れた東北大学アジア研究センターの方にお任せすることで終局しました。このような国際的な研究開発では、「人」として相手国の語学に長けた人材の確保、「物」として良き辞書や諸参考文献の入手、「金」として潤沢な国際交流資金の必要性、などの諸点と共に「時間」が重要な要素であることを実感した次第です。

翻訳業務への機械翻訳の実用

定年退職後は、趣味の客船クルーズを楽しみながらインターネットでの送受信を使った翻訳作業で費用稼ぎをすることを夢見て、手元に残っていた N88BASIC 時代に自主開発したパソコン翻訳を使いこなして、種々の翻訳業務をすることにしました。翻訳実務に機械翻訳を適用する主な利点は、たとえ機械翻訳の性能が低レベルであっても翻訳作業が飛躍的に速まり、数字の桁落ち、翻訳文の段落抜けや行抜けのような致命的な翻訳ミスが無く、また翻訳語句が統一化されるばかりでなく、機械翻訳出力がどうしてこのようになるのかを考えながら、訳文を推敲していると不注意な翻訳ミスも激減すること、などです。

最初の頃の翻訳業務は、主に国際規格ドラフト文の英日翻訳でした。以前から放射線計測関係の J I S 委員として国際規格には関わり合いを持っていましたが、その後も別のルートから ISO 関連の国際規格の作成過程でのドラフト文の翻訳依頼があり、英日機械翻訳を使つての翻訳業務に携わりました。当初の ISO 9000 番台の品質保証の諸規格から始めて、その後の ISO 14000 番台の環境管理の諸規格まで、多くのドラフト規格案の間違いや癖の多い英文について、機械翻訳に掛けて時間短縮を図りながら翻訳したものでした。これら一連のドラフト文の翻訳作業中に気付いた驚くべき点は、国際規格がいつも簡単に出来上がって行く魔法のような規格作成プロセスでした。例えば、当初、出来合いの品質保証の国際規格から単に「品質」を「環境」に文字を代えた程度のお粗末な環境管理の国際規格の初期ドラフトを作り、それを関係諸国に回覧して修正させ、その結果の修正案を基にして何回か同じ手順で修正を重ねて、環境管理の規格の最終案を作り上げて行くというものでした。各段階での修正案ドラフト文の翻訳作業は機械翻訳を利用して行いましたが、英語が母国語でない幹事国が殆どなので、初期のドラフトは間違いだらけの変な英文が多く、機械翻訳以前の英文理解に苦勞し戸惑ったものでした。

その後、翻訳元請け組織の影響もありますが、世の中の景気変動に応じて栄枯盛衰を受ける種々の産業分野の翻訳をしてきましたが、最近では病院関係のデジタル化等のエレクトロメディカル分野の翻訳が多くなってきたようです。この医療・コンピュータ関係の翻訳分野では、多くの専門用語が原語句の読みをカタカナ書きにしたものが多く、本格的な機械翻訳者の立場からは、こんな訳文で理解できるのか、との思いが強く、頭の硬い年寄りにはなかなか受け入れ難い世の中になってきたようです。またそのような世の傾向は、ワープロソフトの進歩にも依るものでしょうか、翻訳の性能やレベルの問題よりも、翻訳結果についての表現の体裁を重んじる傾向が強まってきていることです。世の翻訳レベルがそんなにも高くなったのかどうか知りませんが、翻訳原文の体裁通りに翻訳結果も合わせて表示することが要求されるようになってきています。これらの事柄からは、良好な翻訳をする意気込みが殺がれるような気がして、年齢のせいもあって徐々に翻訳実務から遠ざかりつつある昨今です。

AAMT との関わりあいと要望

AAMT には、1992 年 1 月に個人会員として入会して以来、既に 17 年余りもの間お世話になっています。この間には既に前記のモンゴル国における日－蒙－英の機械翻訳技術の開発支援の報告（本誌 20 号）や、AAMT 主催のシンガポール MT サミット VII への参加報告（本誌 28 号）、更に今回の本文と 3 回、本誌への投稿をさせ

て頂いております。また AAMT を通じてのご紹介を受けて、英国マンチェスター大学での原子力用語辞書の開発提案に参画する貴重な経験もさせて頂いたこともありました。この外にも AAMT 会員による新著の紹介として、拙書「エネルギーのしくみ／池田書店」のメール紹介(2001.5.17) もして頂き、時宜を得た事務局の会員対応に今でも感謝しております。

今後とも AAMT のご発展を期待すると共に、機械翻訳の現状についての判り易い解説や紹介、AAMT 会員間の国内外の交流、などは今後とも期待したいところです。また、大会社や大学関係の法人会員ばかりでなく、種々な状況下にある個人会員に対しても、きめ細やかな気配りの程を今後ともお願いする次第です。

会員名・所属

成山 重子 (The University of Melbourne) /NARIYAMA Shigeko

機械翻訳および翻訳業界に期待すること

言語学の研究、通訳翻訳者としての仕事、そして大学での通訳翻訳の授業や日本語教育に携わった経験から、今後の機械翻訳に期待する点など思いつくまま書いてみたい。機械翻訳の研究を垣間見たため、これらの点が困難を極める事は承知で、あえて今後の目標として述べさせていただきたい。

【プロ翻訳と機械翻訳】

機械翻訳の研究分野では、プロの翻訳は完璧という思い違いがある。何十年と通訳翻訳をしているベテランでも納得のいく「完璧」な通訳翻訳はそうめったに出てくるものではない。普通は大まかな意味が伝わっている程度のものであるというのが、意識の高いベテラン通訳翻訳者の共通の認識だ。また、プロといってもピン・キリで、さらに、ベテランでも人により訳はまちまち。まちがいもたくさんある。実際に、ベテランの翻訳者でも時間が許す限り幾度も修正を加え提出するのは普通で、まして MT の訳が 1 度に良い訳がでたらそれは脅威だ。機械翻訳も 2・3 段階の異なる修正をしてから訳出してはどうだろうか。

機械翻訳の質の悪さは、ある種素人の翻訳と同じで、文字通り、1 語 1 句機械的に訳す事に大きく起因している。では、これではどうしていけないのか、どこがプロ通訳翻訳者と違うのか。現在、機械翻訳の研究に携わる方は日本語母語者が多いようなので、和訳の質に関しては判断しやすいが、英訳の質の判断はよりむずかしいと思われる。そこで、英訳の観点からどういう点が不自然で改良が必要か考える。以下に、英語を母語とする 28 名の私の学生が自分の翻訳と機械翻訳(無料の 10 サイト)の訳を比較分析したレポートと私見から主な点を挙げる。

日英機械翻訳の長所

- まずい英訳でも要点はつかめる事があるので翻訳の前に読む価値はある。原言語がわからない人に大いに役立つ。
- 正確性が重要でない、文字通りの訳で足る (マニュアルなどの) 文章には使える。
- 速い・安い・いつでも使える
- 数字の訳の間違いがない (人間はうっかりやタイプミスがあるが、その訂正は文脈から判断できないことも多いので大変重要)

日英機械翻訳の改良点

- 機械翻訳は長文に弱い
- 英語は日本語に比べ、概して1文が短い：つまり、1つの日本語文を2-3英語文に分けた方が自然でわかりやすい文になる。また、日本語はやたら接続詞が多く、これを全部英訳すると意味をなさなかったり、不自然になる。
- 日本語には機能句が多く、それを英訳すると意味をなさない。例、「…となっています」「と思われま
- す」（これは婉曲表現として日本語ではよく使われるが、英語では、話者の「判断の不確かさ」を意味する。）
- 日本語は間接表現を好み、英語／欧州語は直接表現を好む。これを文字通り訳すと、押しの強い文章になり、ひんしゆくをかう事にもなる。つまり、和訳の時は婉曲表現を使ったり省略したりするが、英訳の時は、明瞭な表現を使うべき。このような調整は、新米の通訳翻訳者や機械翻訳には困難。
- 基本的な日英文法の違いも克服できていない。非文が多い。例 a/the, 時制やテンス（日本語と英語では一致しない、つまり、直訳すると間違いになることが多々ある）、数（例：20 person）など。
- [最大の問題であるがその解決は最も困難] 文脈情報を一切考慮していない。という事は、
 - 適切な単語の選択ができない。例：「オーストラリア最大」は 'the highest in Australia'が正しいのだが、機械翻訳は単語単位でみているため、ほとんどの機械翻訳は'Australia maximum'と訳している。
 - 文のつながりや流れが悪く、意味がつかみにくい。何をいいたいのポイントがつかめない。
 - 話者の意図や強調点を考慮していないので、誤ったニュアンスを醸し出してしまう。

つまり、以上の点から、文脈情報を考慮しない語や句単位の訳やパラレルコーパスの有効性はいささか疑問。簡単な例では、Thank you は文脈や場や相手によってその和訳は様々に変化する。「ありがとう、すみません、よろしく、失礼しました、おじゃましました、では、(会釈)、etc」 また、この点については、以下に引用する日本語文とその英訳が如実に表している。これは AAMT 事務局発行の「journal 原稿フォーマット」だが、日本語も英語も自然で違和感のない文。伝えている内容や概念や丁寧さも同じで、日本語をみた人も英語を見た人も同じ理解に達する事ができる。しかし、日本語と英語の文や単語はほとんど対応していない。(挨拶文という点も大きく関与しているが、傾向としての一例。)

「AAMT ジャーナルへの記事／論文提出について」

次号への寄稿のご承諾を賜り、厚くお礼申し上げます。

以下の要領で原稿をお送りいただければ幸甚です。どうぞよろしくお願いいたします。」

これに対応する英語

Article/Paper Submission Instruction for AAMT Journal

When you submit your paper for the upcoming AAMT Journal No.44, please comply with the following instructions. We thank you very much for your kind attention and cooperation in this matter.

この意味でも、用例翻訳の方がより自然な訳出が可能だといえる。それともなるべく大きな単位で処理できれば精度と自然さが増し、訳文調で意味が伝わらない違和感のある文が減るだろう。

【プロ通訳翻訳者が目指す訳とは】

「完璧／理想的」な訳とは、各ことばの訳ではなく、文章全体が醸し出す「概念」を他の言語に置き換える事である。つまり、話者の伝えたい意味や意図がその言語を使わない人たちにも、忠実に伝わり同じ理解ができる事といえる。例えば、日本語を英訳した時、日本語話者が意図した意味や概念が、日本語話者の聞き手に伝わるように、英訳から英語話者も同じ意味や概念として伝わる事。よくいわれる事だが、通訳（翻訳）者は黒子であり、訳を通し、同じ言語を共有しない人たちが同じ言語を使って意味を伝えた時と同じ理解に達せれば、それはよい訳といえる。そのため、日英／英日訳の場合、マニュアルなどはともかく、単語や文や対応しない事の方はよくある。

具体的にプロ通訳翻訳者がよりよい訳のため重要としている主な点は次の通りだ。どれも今のところ、機械翻訳が最も苦手とするところだ。

1. 文化、背景、場面、常識、前提などの言語外知識を考慮し、どこまで訳にその違いからくるギャップを補足するか、またどれを削るべきか適切な判断をする。
2. 読み手／聞き手は誰で、どこまでの知識があるか考慮する。つまり、読み手／聞き手により訳は変えるべき。
3. 話者の意図は？
4. 分野や場に特化した適切なことばの選択

【辞書の重要性：ニュアンスや使用場面や文脈に関する記述のある辞書の考案】

これらの点を改善すべく新しい辞書作成を考案している。研究資金がいただければ本格的に着手したいと思っている。どこが「新しい」のかというと、辞書というのは本来母語者用に書かれているため、母語者にとって当たり前な「暗黙の了解」事項はほとんど記載されない。つまり、細かいニュアンスの違いや使用場面や文脈に関する記述はすでに共有されているとの前提から書かれていない。既存の英和・和英辞書も、ほとんどが日本語母語者のためのものであり日本語母語者が必要な情報だけ記述されるため、日本語を外国語として学習する者に重要な記載がかけられている。例えば、know を英和辞書で引くと、「知る、分かる、理解する、…」など複数の言葉が挙げられているが、その細かいニュアンスの違いや使用場面に関する記述は不十分である。このため、仕方なく、または事の重大さを知らずに日本語学習者は無作為に1つの言葉を選び、しばしば誤文を作ってしまう。例えば、「この問題どうしたらいいでしょうか」の答えとして「分かりません」と言うべきところを「知りません」と言い相手を不愉快にさせても、学習者は何が原因かわからずにいる。

逆の英和の例も然り。英語の先生に‘Would you like to get started?’ ‘始めていただけますか」と言われ ‘I don't care’ と答える日本人が非常に多く、聞いていて冷やっとさせられる。‘I don't care’ は「私の知ったことじゃない／どうでもいい」の否定的ニュアンスがあり、‘I don't mind’ ‘かまいません／いいですよ」というのが正しい。しかし、care/mind に同じような意味があるので、辞書にはその微妙な（しかし重要な）意味のちがいは載っていないことが多い。

また、フレーズになるとより使用場面やニュアンスが重要になる。例えば、英語の How are you? は初めて会っ

た人にも毎日顔を会わせる人にも使われるため、学習者は「お元気ですか」を英語と同様に連発する。初めて会った人に「お元気ですか」と面と向かって言われた日本人はしばし困惑し違和感を持ってしまう。

これらの問題は、共起情報では回避できない違いで、このような単語やフレーズの細かいニュアンスや使用場面や文脈に関する違いを範疇化し辞書に記述することで、日本語学習のみならず機械翻訳に役立てるのではないかと考えている。

【Transtype】

以前 AAMTJournal に掲載された研究報告と同じだが、先の学生のレポートにも、ある程度の語学力がある場合は、機械翻訳より充実した信頼できる辞書の方が有益と報告している。通訳翻訳者にとり質のよい辞書は、情報工学研究にとって最新／高性能／高 CPU のコンピュータを使い言語処理する事に当たる。

そこで思い出すのが、Transtype という翻訳者向けの EC Project が開発中のシステムだ。
<http://tt2.atosorigin.es/summary.htm> そのデモを4-5年前に学会でみたが、その時「使える」「こんなのがほしい」とつくづく思った。ベテラン翻訳者でも一番手間取るのは辞書引きだが、このシステムは見事にこの問題を解決している。ヨーロッパ語間の翻訳システムなので、日本語を介した時どのくらい有効かはわからないが、日韓などの似た言語であれば十分有効であると考えられ、またシステム自体、他の言語に応用できるだろう。このシステムは入力された文章から MT を訳出するが、基本的には翻訳者が中心。翻訳者がタイプする訳語毎に、それにあわせて MT の訳出をもとに、その語に続く訳の選択儀を表示し、翻訳者はそれを選んでもいいし、自分で訳を作っても良い。この MT はできればハイブリットで、用例翻訳、翻訳メモリー、統計翻訳などによる訳であれば、より【使える】だろう。

【機械翻訳：科学の最後のフロンティア】

機械翻訳の究極は言語理解だ。言語理解は、認知学に大きく関与し、人間の心や頭脳の働きの解明が必須である。この研究は科学の最後のフロンティアとも言われており、困難を極める。ことばは1万のルールがあってもその使用法は網羅できない、つかみどころのない、話者や世代や地域や時代でも変化し様々なバリエーションをもつ曲者だ。この曲者を、常識や直感をもたない、0と1の2つの選択肢しかもたないコンピュータを用いて処理する訳だ。油と水を融合させるようなもので、むずかしいのは当たり前！すでに機械翻訳の技術は、ことばの壁を緩和し世界をつなぐことに着実に貢献しているので、今後も長期展望で研究が進むことを切望する。

AAMT へのご要望

もっとプロ通訳翻訳者との意見交換の場があるとよいと思う。

会員名・所属

横山知幸（広島市立大学国際学部）／YOKOYAMA Tomoyuki

自己紹介

私はもともと教育学部・教育学研究科の英語専攻の出身で、前任校までは英語教育関係のことを中心に教えておりました。機械翻訳の研究者でも翻訳家でもありません。ですが、英語教育分野の主たる流れが、いわゆる「コミュニケーション」なるものに向かう中で、長い間英語教育の中心であった英文和訳の研究がしたかったものですから、現在の勤務校に赴任する際に、翻訳関係のことを教えてほしいという依頼をありがたく受けて、今では学部でも大学院でも主に翻訳関係のことを教えています。もっとも、翻訳実務が教えられるわけではありませんので、英文和訳などの歴史や理論的なことを扱っています。自分の研究としては、最近では、「独案内」という明治期のリーダーの自習書を使って、原文と訳文の語順の問題を考えていたりします。今でも色々あって、二単位ほど英語科教育法も教えていますし、普通の英語の授業に近いものも担当していますが。

私自身がこのような経歴なので、これまでゼミで論文を書ってくれた院生のテーマも様々です。『ジキル博士とハイド氏』という多重人格的な小説における代名詞の翻訳の問題を扱ってくれた人もいれば、『アルジャーノンに花束を』という知的障害のある主人公が手術を受けて天才になるのだけれども結局はもとの状態に戻っていくという SF 小説を扱い、原文と訳文における文体の変容とその現れ方の違いを論じてくれた人もいますし、はたまた、機械系の特許翻訳を何年もやっていて、原文に添付されている機械翻訳の問題を考察してくれた人もいます。機械翻訳については、私自身が理解できることは僅かですが、色々な意味で身近に感じてはいます。

AAMT に入会することになったのは、こちらで出していच्छる書籍をお願いしようと連絡したら、いっそ会員になりませんかというお話をいただいたからだったと思います。何とも場違いな会員で申し訳ありません。

機械翻訳および翻訳業界に期待すること

私自身は、機械翻訳の研究者でも翻訳者でもありませんので、主として教育に携わる者の視点から書いてみます。

現在の勤務校には語学センターがあり、300 台ほどの PC があります。私は、ほとんどの授業をここで行っています。その中で、割とよく学生がウェブ上の機械翻訳サービスなどを利用しているように見える英作文関係の授業（英語で読書感想文を書かせています）を例に取りましょう。授業中に教卓の学生 PC 表示画面を見てみると、機械翻訳を頻繁に利用するのは、一割から二割程度の学生です。使い方も様々です。長い日本語文を先に書いて、それをまとめて訳させている学生はあまりいません。むしろ各英文の構文の中心になりそうな表現を機械に翻訳させ、それをもとに枝葉の部分を自分で書き足している学生の方が多い気がします。場合によっては、本当に語句レベルで英和・和英辞典代わりになっている学生もいます。自分が作った英文の方を和訳させ、ちゃんと意図する意味の日本語に訳されるか確認しているのではないかと思われる学生もいます。英作文の方には全く使わず、感想文のもとになる英文を直訳して構造を確認しているような感じの学生もいます（意識気味の全訳は授業の最初に配布済）。多くの学生は、機械翻訳はほとんど使わず、自分の知識・経験とポータブルな電子辞書で何とか作業を進めます。みんな、自分の使いたいものを使いたいように使っているようです。

英語の学習をしながら多少の英語の運用もするような普通のユーザに対しては、機械翻訳だけと言うよりは、

むしろ、機械翻訳・電子辞書・ウェブ上での表現検索（安藤進先生のこの分野のご著書は素晴らしいですね！）などが一体となって、言語の学習者・運用者の理解・表出を支援していただけるような仕組みがもっと発展していけばとてもうれしいなと思います。私自身、機械翻訳ではありませんが、何年もかけて苦労しながら、Jammingで複数の辞書を一括検索できるようにしたり、Emacsでlookupを使ったり、Answers.comやYahoo!百科事典を使ったり、XOOPSで授業用のサイトを作ってみたりと、自分にとって都合の良い言語運用環境を整えてきましたが、学生にそのような努力を期待するわけにはいきませんから、使い勝手の良いまとまった仕組みがあるといいなと思います。最終的な目標は、自分である程度の英語の読み書きが出来るようになることであり、そこに行き着くまでは、各自の都合やレベルにあわせて色々なものを利用し、言語の運用レベルを本来の実力以上に引き上げてやり、その結果として、言語能力の方も少し上がって行けばいいわけです。

私自身よくわかっていないので、変なことも色々書いてしまったかもしれませんが、人間の言語運用を機械に支援してもらうことに関しては、現在でも既にかかなりの程度までそうになっていますし、今後ともますます期待したいと思います。

AAMT への要望

特に要望を書かせていただくほど、AAMT のことがよくわかっているわけではありません。

協会活動報告

(2009 年 2 月～2009 年 5 月)

理事会 (E-mail にて議論)

2009 年 3 月 24 日～3 月 31 日

- ①第 1 号議案 2008 年度事業報告 (案) ② 第 2 号議案 2008 年度決算報告 (案)
③第 3 号議案 2009 年度事業計画 (案) ④ 第 4 号議案 2009 年度収支予算 (案)

2009 年 4 月 1 日～4 月 8 日

AAMT 海外支部創設に関する会則改正に関する承認審議

2009 年 4 月 1 日～4 月 8 日

機械調査課題委員会委員長交代に関する承認審議

機械翻訳課題調査委員会

2009 年 2 月 23 日 (2008 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

2009 年 3 月 16 日 (2008 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

2009 年 4 月 14 日 (2009 年度 第 1 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

2009 年 5 月 8 日 (2009 年度 第 2 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

AAMT/Japio 特許翻訳研究会

2009 年 3 月 13 日 (2008 年度 第 8 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 4 月 24 日 (2009 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

2009 年 5 月 29 日 (2009 年度 第 2 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 3 月 AAMT Journal No.44 発行

2009 年 5 月 22 日 (2009 年度 第 1 回)

①AAMT Journal No.45 進捗確認

②No.46 の企画について

③その他

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。

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

MT Summit XII 開催のご案内

26-30 August 2009
Ottawa, Ontario, Canada

Association for Machine Translation in the America (AMTA)主催による第12回機械翻訳サミット (MT Summit XII) が来る 2009 年 8 月、カナダのオタワにて開催されます。MT Summit XII は機械翻訳分野の研究者、開発者、ユーザ等の幅広い関係者を対象としており、主要な学術、民間、政府機関等による最新の研究発表やデモンストレーションが予定されています。

サミットの本会議においては、研究論文発表、ユーザレポート、招待講演、パネル討論等が用意されており、更に、本会議前後の拡張プログラムとして、チュートリアル、ワークショップ、デモセッション等が予定されています。

AAMT 関連といたしましては、今回も前回に引き続き、本会議後の 8 月 30 日 (日) 午後より AAMT-Japio 特許翻訳研究会主催によるワークショップが開催を予定しております。

MT Summit XII とあわせてまして皆さまからの多数の参加をお待ち申し上げております。

【日時】 2009 年 8 月 26 日 (水) ～30 日 (日)

【会場】 The Fairmont Château Laurier Hotel

1 Rideau Street, Ottawa, Ontario, Canada, K1N8S7

TEL (613) 241-1414

FAX (613) 562-7030

URL <http://www.fairmont.com/laurier/>

【チュートリアル】 2009 年 8 月 26 日 (水)

【本会議】 2009 年 8 月 27 日 (木) ～29 日 (土)

【ワークショップ】 2009 年 8 月 30 日 (日)

プログラムの詳細及び参加申込等は、公式ウェブページをご参照ください。

<http://summitxii.amtaweb.org>

MT Summit XII 3rd Workshop on Patent Translation

August 30, 2009

Ottawa, Ontario, Canada

<http://summitxii.amtaweb.org/>

Workshop Theme

Patent documents are one of the major application areas of machine translation. The workshop aims to foster research and development of the technology for patent translation by providing a forum in which researchers and practitioners can exchange their ideas, approaches, perspectives, and experiences from their work in progress. We have successfully made the same workshops in MT Summit X and XI, and will make deeper discussions about this theme.

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

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

Areas of Interest

Topics of interests include, but are not limited to:

- Analysis and classification for patent documents,
- MT and translation aids for patent documents,
- Contrastive studies for multilingual patent documents,
- Language resources for patent translation,
- Dictionaries and terminology databases for patent translation,
- Parallel or comparable corpora for patent translation,
- Information extraction from patent documents,
- Evaluation techniques for patent translation,
- Multilingual patent classification and retrieval.

Submission of papers

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

Authors are requested to keep their papers anonymous and submit separate cover pages with the following information:

- paper title,
- author's name(s), affiliation(s), address(es), and e-mail address(es),
- 100-200 word summary,
- some keywords for classifying the submission.

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

shimohata245@oki.com

Important Dates

- June 19, 2009: Deadline for paper submission
- July 10, 2009: Notification of acceptance
- July 31, 2009: Camera ready submission deadline

Organization

Workshop Chair:

Terumasa Ehara, Prof. of Yamanashi Eiwa College, Japan

Co-Chair:

Shoichi Yokoyama, Yamagata University, Japan

Program Committee

Bente Maegaard, University of Copenhagen, Denmark

Svetlana Sheremetyeva, Lanaconsult, Denmark

Munpyo Hong, Sungkyukwan University, Korea

Jong-Hyeok Lee, Pohang University of Science and Technology, Korea

Jun'ichi Tsujii, University of Tokyo, Japan and Manchester University, UK

Hiroyuki Kaji, Shizuoka University, Japan

Eiichiro Sumita, ATR, Japan

Shinichiro Miyazawa, Shumei University, Japan

Sadao Kurohashi, Kyoto University, Japan

Takehito Utsuro, University of Tsukuba, Japan

Takashi Ninomiya, University of Tokyo, Japan

Hiroshi Echizen-ya, Hokkai Gakuen University, Japan

Takashi Tsunakawa, Shizuoka University, Japan

Toshimichi Moriya, Japan Patent Information Organization, Japan

Toyohide Watanabe, Japan Patent Information Organization, Japan

Tadaaki Oshio, Japan Patent Information Organization, Japan

Sayori Shimohata, Oki Electric Industry Co. Ltd., Japan

Akira Ushioda, Fujitsu Labs., Japan

Akira Kumano, Toshiba, Japan

Mitsugu Miura, NEC, Japan

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45号をお送りします。

今号の巻頭言は、財団法人日本特許情報機構(JAPIO) 専務理事兼 特許情報研究所長であります、守屋敏道様にご執筆頂きました。

特許翻訳は、機械翻訳の最も重要な適用事例の一つであり、JAPIOにおいて長年にわたり、特許文書の機械翻訳に取り組んでこられた様子について、ご紹介頂きました。

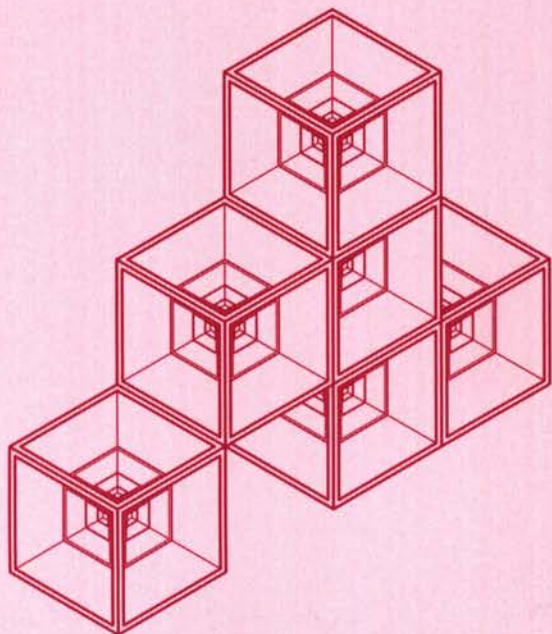
また、今号では前号に引き続いて、計算機上における日本語の取り扱いを容易にするために日本語を規格化することを目指した試みとして、「産業日本語」プロジェクトの紹介をご寄稿頂きました。その他にも、研究機関や企業における各種プロジェクトや取り組みの紹介をご寄稿頂きました。

情報通信研究機構(NICT)からは、音声・言語資源と各種ツールを配信する高度言語情報融合フォーラム(ALAGIN)の立ち上げの紹介、およびボランティア翻訳者の支援と翻訳の共有のためのWebサイト「みんなの翻訳」の紹介をご寄稿頂きました。また企業からは、サン・マイクロシステムズ、および日本IBMにおける翻訳業務への取り組みの紹介をご寄稿頂きました。

一方AAMT内の活動報告としては、機械翻訳課題調査委員会から、日中機械翻訳文テストセット(第1版)の公開、および英日・日英機械翻訳に関するアンケート結果報告についてご寄稿頂きました。その他、今号におきましては、通常の号と同様の企画を掲載しております。

「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2件、個人会員の紹介文3件を掲載しました。

AAMT



AAMTジャーナル No.45

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 後藤功雄 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>