AAMT

The Asia-Pacific Association for Machine Translation

Journal

No.24
October 1998

アジア太平洋機械翻訳協会

目　　次

エッセイ	21世紀にむけ「機械翻訳の土壌」を育む	1
論文／論説	機械翻訳の現状と今後の動向	3
技術早分かり	文法形式と語彙化	12
業界動向	産業翻訳業界概観	18
国際会議報告	COLING-ACL'98参加報告	23
新製品紹介	●The 翻訳インターネット	28
	●英日・日英翻訳支援ソフト「PENSEE V3.0」	30
	●国際契約書システム V1.0	32
専門誌の紹介	専門誌「Machine Translation」定期購読のお知らせ	2
MT Summit	MT Summit Ⅶ September13—17, 1999, Singapore	27
事務局だより	委員の交替	11
	'98年3月～9月にご入会された方々	17
	協会活動報告	26

CONTENTS

Essay :	Develop the Fundamentals of Machine Translation for the 21st Century	1
Paper :	Machine Translation in the Present and the Future	3
Technological Writings :	Grammar Formalisms and Lexicalization	12
Translation Industry :	Brief Report on Technical Translation Industry in Japan	18
International Meeting :	COLING-ACL'98 Report	23
New Products :	● The Internet Translation	28
	● Translation Support Software E/J「PENSEE V3.0」	30
	● International Contract SystemV1.0	32
Information of Technical Journal :	「Machine Translation」 Subscription	2
MT Summit :	MT Summit Ⅶ September13－17, 1999, Singapore	27
AAMT News :	Replacement of Committee	11
	New Members joined from Mar. to Sept, '98	17
	AAMT Activities	26

エッセイ

21世紀にむけ「機械翻訳の土壌」を育む

株式会社インターグループ代表取締役　小谷泰造

コミュニケーションを生業とする当社が機械翻訳の業務に取り組んでから、既に十数年の月日が経ちます。メーカでない当社が、様々な電子化辞書のプロジェクトに参加させていただき開発のお手伝いをさせていただいたり、いち早く実際の翻訳業務にて機械翻訳ソフトを応用して参りました。また並行して、広く一般に啓発を図るため、MTサミットをはじめとして、数々の機械翻訳関連の国際・国内会議の運営をお任せいただき、内外に機械翻訳の啓蒙につとめて参りました。

決して長いとはいえないこの期間の間に、コンピュータや情報処理技術の飛躍的な進歩、普及には目を見張るものがありました。特に、パーソナルコンピュータの驚異的な普及や、インターネットによるネットワーク世界の拡大、携帯情報端末等によるモバイルコンピューティングなど、個人ユーザを取り囲むコンピュータやネットワークの状況は、本当に大きく変わってきました。

かつては、キーボードというものを叩き、画面上の呪文のような英文字を操ることができなければ、何もしないただの箱であったコンピュータなるものが、「マウスとクリック」によって、とりあえずは誰でも簡単に動作させることができるようになり、今や、人間に語りかけるようにコンピュータに向かって自然な日本語で発声すると、その言葉をコンピュータが理解して、画面上に文字が入力されたり、基本的な機能が働いたりするようになってきました。いわば、コンピュータ、機械と人間がコミュニケーションできる一歩手前まで来ている感があります。

そういう状況の中で、機械翻訳自体もずいぶんとその様子を変えてきたように思われます。かつては、主に業務用でしか使用されていなかったそのシステムが、パソコンと同じく安価なソフトとして、一般のユーザ層をターゲットとする商品になってきました。今まで外国語が壁になって海外とのコミュニケーションができなかった人たちにとっては、それを自由に可能にする、まるで夢のような商品に思われるでしょう。

別な見方をすれば、上記のようなパソコン自体の技術的発展や機械翻訳ソフトの商品的変貌が、巷の電化商品店街や書店に溢れるコンピュータ関連雑誌の話題で多く取り上げられる一方で、肝心の「機械による翻訳文の品質」については、割に「無関心」になっているのではないかという気もします。それは二つの方向で「無関心」になっているように思われます。

ひとつは、「機械翻訳は実際の仕事上の正確な翻訳には使えない」という業務判断からくる無関心であり、もうひとつは、「英語が日本語にかわっていればそれである程度満足する」という安心感からくる無関心です。

確かに、趣味やプライベートで翻訳ソフトを購入する一般消費者にとっても、日常業務での利用を検討している企業、団体にとっても、機械翻訳ソフトほど、購入時にどこのメーカの商品にすべきか、判断の基準がむつかしいものはありません。現在は「翻訳の質」というよりは、ブランド名の他、パッケージの裏面に印刷されている情報、「専門用語辞書が何十万語だ」とか、「翻訳スピードが一分に何語だ」とか、「見やすい画面だ」とか、「操作性が抜群！」などという、うたい文句に購入動機が左右されているのではないでしょうか。（場合によっては、パッケージそのもののイメージや価格のみで！）しかし、いつまでもこうであっては、翻訳ソフトにとっても国際化の流れにとっても、流れに逆行するものといえます。

例えば当社の業務で言えば、国際会議の現場で行

われる同時通訳というコミュニケーションにおいて、そこには２つの要素があると思われます。それは、その分野の専門知識と「ことば」の知識です。いくら言葉だけを変換したところで、専門知識のない人が聞いてもその内容は理解できないでしょうし、逆に言葉の変換（通訳）が100％完璧でなくても、専門知識の背景のある人が聞けば、話の筋は理解できるというものです。残念ながら、今の機械翻訳のユーザには、まだその当りのことが欠けているようです。外国語でコミュニケーションする為には、コミュニケーションする内容自体についての知識やその文化に対する理解もなくてはなりません。

もちろん、機械翻訳の研究開発の現場では真摯に、かつ精力的に、新技術の探求に力が注がれているのはいうまでもないことです。しかし、今後機械翻訳の開発をするメーカ側にも、そういった意識は必要になってくるでしょう。開発者と利用者が同じ目的意識を持って研究に携わり、最良のものを生み出す環境・意識が揃えば、21世紀に向けた「機械翻訳の土壌」を育むことにつながるのではと思います。

専門誌の紹介

専門誌「Machine Translation」定期購読のお知らせ

オランダの出版社 Kluwer Academic Publishers が、年４回機械翻訳の専門誌「Machine Translation」を発行しています。当協会の個人会員で定期購読の申込をされた方は、Kluwer 社から購読料の約３割引のサービスを受けることができます。購読申込および代金支払いは AAMT が一括して受付けますので、ご希望の会員の方は事務局へお申込ください。なお、本の送付は代金をお支払いいただきました約１カ月後となります。

１．次回発行月
　1999年３月　（年４回発行）

２．割引対象者
　AAMT、AMTA、EAMT の個人会員（法人会員は除く）

３．購読代金（４冊分の価格）

1998年度発行	一般購読料：＄110	会員対象割引：＄80
1999年度発行	一般購読料：＄122	会員対象割引：＄90

４．申込方法
　会員氏名・住所（送付先）を日本語・英語で併記し、日中の連絡先を明記して手紙、FAX、E メールでお申込ください。お申込後、事務局から購読代金をご連絡いたしますので、日中の連絡先をお忘れなくお知らせください。

５．代金支払い
　お申込いただきました方に事務局から購読代金をお知らせいたします。代金の計算には、月の第一営業日に東京三菱銀行が提示している米ドル為替レートを適用いたしますのでご了承ください。お支払いは、申込された月の末日までに銀行振込でお願いいたします。手数料はお振込人様にてご負担ください。なお、お申込人様から代金のお振込がなく、事前に何もご連絡がなかった場合には、購読申込は無効となりますのでお気を付けください。

論文／論説

機械翻訳の現状と今後の動向

－日本の言語資源の充実のために－

山端　潔[1]　大倉清司[2]　亀井真一郎[1]　松井くにお[2]

1　はじめに

情報社会の到来により、今や機械翻訳はグローバルネットワークの必須アイテムとなっている。世界中の「情報」は自然言語で記述されたものがほとんどであるといってよい。また、ネットワーク社会の到来と共に必要とする情報が自国語以外の言語で書かれているケースが増大しており、「情報」の受信／発信には機械翻訳技術の活用がなくてはならない時期を迎えている。1994年頃を境に企業の事業環境にも変化が起こり、最近の主要な機械翻訳ベンダの機械翻訳事業には黒字を達成しつつあるところも多い。一時は停滞気味であった新製品開発や翻訳品質の向上作業もここへきて再び非常に活発に展開されている。

現在の機械翻訳の訳出品質は、一般テキストを対象とした場合、約60%と言われている。この品質は決して十分とは言えないが、それでも例えば英語でかかれたWEBページを日本語に翻訳し概要をつかむといった需要に応え、パソコンに機械翻訳ソフトが装備されているのは今や常識となっている。今後、翻訳品質が90%を越えるようになれば、海外旅行での本格的利用や、テレビ、電話などのメディアと接続することにより大幅に利用分野が拡大し、爆発的に普及することが期待される。

機械翻訳は単独のシステムとして魅力があるだけではない。自然言語処理分野の要素技術を総合したシステムとして多くの隣接技術分野に対する波及効果が大きく、広範な技術を牽引していくという点でも、非常に重要な技術である。技術的に関連する分野としては、「情報」「知識」を扱うあらゆる技術分野、例えば、テキスト処理技術（検索、要約、分類、等）、クロスランゲージ情報検索技術[3]、音声処理技術、文字認識技術、対話技術、自動通訳、Electronic Commerceなど広範囲に及ぶ。

従来技術の改良だけでは翻訳品質90%はまだ遠い目標であるため、研究者・技術者は技術的ブレークスルーを求めて研究活動を行っている。有望と思われる方向の一つにデータ指向のアプローチがある。これは、言語の複雑さに対して、文法記述者の能力だけに頼るのではなく、大量のコーパスとそれを処理するマシンパワーで立ち向かおうとするアプローチである。このアプローチの前提となるのが、言語の振舞いを記述したコーパスとそれらを扱うソフトウェアツール類からなる言語資源である。

「知識」「情報」を記述するための自然言語は、人間が意図的に設計構築した人工物ではないという点で、物理や化学の対象と同じく、一種の自然現象と見なすことができよう。自然現象である以上、それを取り扱うためには、まず言葉の振舞いを記録したコーパスを大量に収集し、その分析を行うことが必須である。コーパスおよび処理ツール類は言語資源と総称されるが、欧米では古くから整備が行われている。これら言語資源は、機械翻訳を始めとする自然言語処理関連分野で利用されるばかりでなく、言語教育など他の分野でも有効に用いられている。実際、欧米では、大規模コーパスから抽出した、実例文や頻度情報を記載して評判の高い学習者用辞典類も発行されている。それに比べて、言語技術の重要性の認識および言語資源の整備の点で、欧米に比べ日本は決定的に遅れている。大規模コーパスの構築には多大な人的・金銭的・時間的コストがかかるため、共通インフラとしての重要性を認識し、民間と公的機関の力を集めて収集と構築を進める必要がある。

1 NEC C&C メディア研究所
2 富士通研究所マルチメディアシステム研究所
3 日本語キーワードで英語テキストデータベースを検索するなど、問い合わせの言語と異なる言語のデータベースを検索する情報検索手法

2　技術動向

2.1　概論

機械翻訳システムにおいて言語知識をもっぱら格納するのは辞書と文法であり、これらの量と質が翻訳精度を決める。例えば、SYSTRAN 社の露英翻訳システムは翻訳精度90%とも言われるが、翻訳精度を支えているのは基本語だけで40万語にのぼる辞書である。この豊富な言語知識は、機械翻訳の黎明期以来、数十年にわたり人手で収集したものであり、単語対訳にとどまらず複合語や表現の対訳をも多量に含めることで高品質な翻訳が可能になったと言われている。それに対し、現在の商用の英日・日英の機械翻訳システムの基本辞書の登録語数は多くて20万語程度である。さらに、日本語と英語のように構造と語彙が大きく異なる言語の間では、辞書のみを強化しても同様な精度は期待できない。

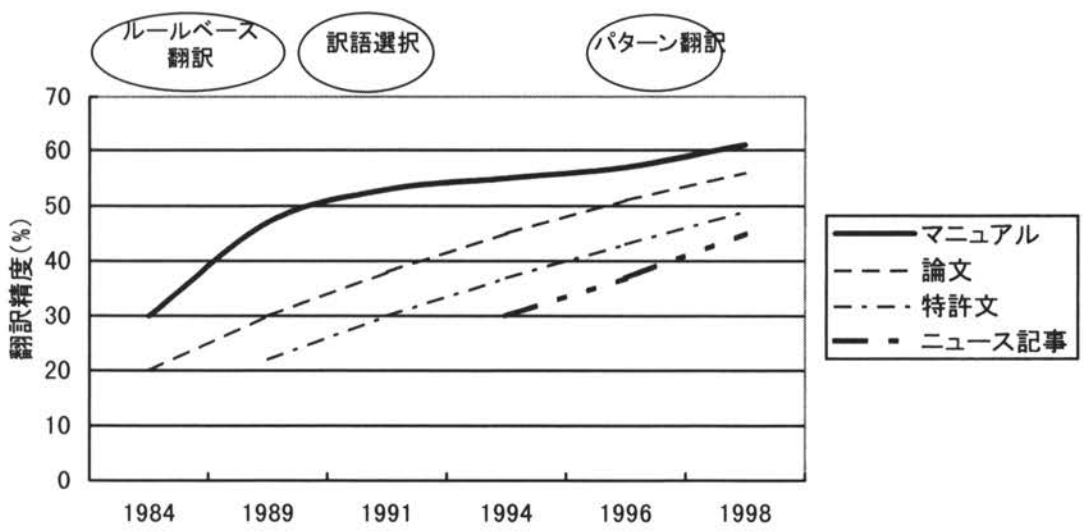
2.2　商用システム

現在の商用機械翻訳システムのほとんどは、いわゆるルールベースシステムである。ルール（文法規則）の表現形式はシステムにより異なるが、日本での商用システム開発開始以来十数年を経た現在、どのシステムでも、文法書にあるレベルの一般的な現象はほぼ記述が終わっていると思われる。

翻訳システムの訳質向上のネックになっているのは、言語知識の多様性と個別性、すなわち言語の複雑さである。実際、人手で辞書と文法を一つ一つ記述していくだけでは、現実の言葉の複雑さに対応しきれないことが経験的にわかってきた。例えば、一つの単語には多数の訳があり、状況によって訳し分けが必要だが、訳し分けのルールをすべての語の組合せに対し個別に手で書いていくのはコストがかかりすぎる。また、複数単語からなる複合語が適切に訳されることが翻訳訳質の点からはきわめて重要であるが、毎年数十万語ずつ増え続ける複合語に対応して対訳辞書を増強し続けることは現実的に難しい。さらに、訳質向上のためには単語だけでなく、表現や言い回しを適切に訳せることが重要だが、こうした表現は書き言葉／話し言葉、新聞／雑誌等の媒体、分野や世代により意味や用法が異なるため、現実には人手で記述しきれない。さらに、ルールを増やしていってもそれだけで訳質が向上するわけではない。ルール数が増えると、ルール間の相互干渉が増大し制御が困難になり、結果として訳質向上が飽和してしまう。

このように、ルールベースシステムの上に人手で辞書・文法を記述していく現在のやりかたは、もっぱら言語の量と複雑さに対応しきれないという点において、訳質向上の壁につきあたっている。図１に示すように、現在も徐々に訳質は向上しているものの、一般のテキストに対して翻訳品質90%を達成するには技術的ブレークスルーが必要である。

図１翻訳精度の推移



2.3 語彙化文法

言語処理システムの開発経験が蓄積するにつれて、語彙から分離できる一般的な文法知識よりも、語彙に特有の個別の文法知識の方が圧倒的に多く重要であることが認識されるようになった。語彙に個別の文法知識を見通しよく記述するための枠組が語彙化文法であり、すべての文法規則は原則として単語の辞書中に記述する。語彙化文法は、次に述べるデータ指向アプローチとの相性がよいので、この観点からも欧米での文法構築プロジェクトの多くが語彙に多くの翻訳用の情報を組み込んだ語彙化の枠組を採用している。

2.4 データ指向のアプローチ

言語のあまりの複雑さに人間である開発者が対応しきれないという問題は、機械翻訳に限らず言語処理一般に共通する問題である。この問題に対し、ここ10年ほどでメジャーになってきたパラダイムがデータ指向のアプローチである。これは、複雑さに対して、人間の能力だけに頼るのではなく、大量のコーパスとそれを処理するマシンパワーで立ち向かうアプローチである。このアプローチはとくに音声認識分野で成功をおさめた。

このアプローチに必須の要素が二つある。一つは大量のコーパス（文例集）であり、もう一つはコーパスデータから知識を抽出／獲得／学習するアルゴリズムである。特に、前者のコーパスはアルゴリズムの如何に関わらず、例えば人間が知識を抽出する際にも共通に利用できる資源であり、処理パラダイムを超えて汎用性が高い。

2.5 事例ベース・パタンベース

対訳事例を多数収集し翻訳に利用する事例ベース翻訳方式はデータ指向アプローチの典型例である。80年代前半に（当時の）京大・長尾教授により唱えられて以来、有望なパラダイムとして多くのプロジェクトに影響を与えてきた。

商用システムで直接この方式を採用したものはまだないが、類似した発想に基づくものとしてパターンベースの翻訳がある。これは、具体的な対訳事例をもとに一部を人間が抽象化して翻訳ルールとする方式だが、複数の商用システム中に組み込まれ有効性が確認されている。

対訳事例から動詞の格フレーム周りの対訳規則を抽出する研究も精力的に行われており、研究レベルでの成功をみている[4]。しかし、対訳事例を大量に集めるのが困難なことがこの方式を進める上でボトルネックとなっている。

2.6 統計モデルと学習

統計言語処理は、文法をはじめとする言語モデルを統計モデルとして構築し、大量コーパスからの学習により言語データの統計的性質を反映するパラメータを学習するというパラダイムである。例えば、コーパスから単語 n 個の並びに関する統計量をあらかじめ計算しておくことにより、文中の n－1 語の並びを見て直後の単語を統計的に予測することができるが、この統計モデル（n-gram モデル）は形態素解析技術などに幅広く応用されている。最近の音声認識技術は完全に統計モデルと学習のパラダイムに基づいており[5]、データとモデルの改良により劇的に性能が向上したことは周知のとおりである。

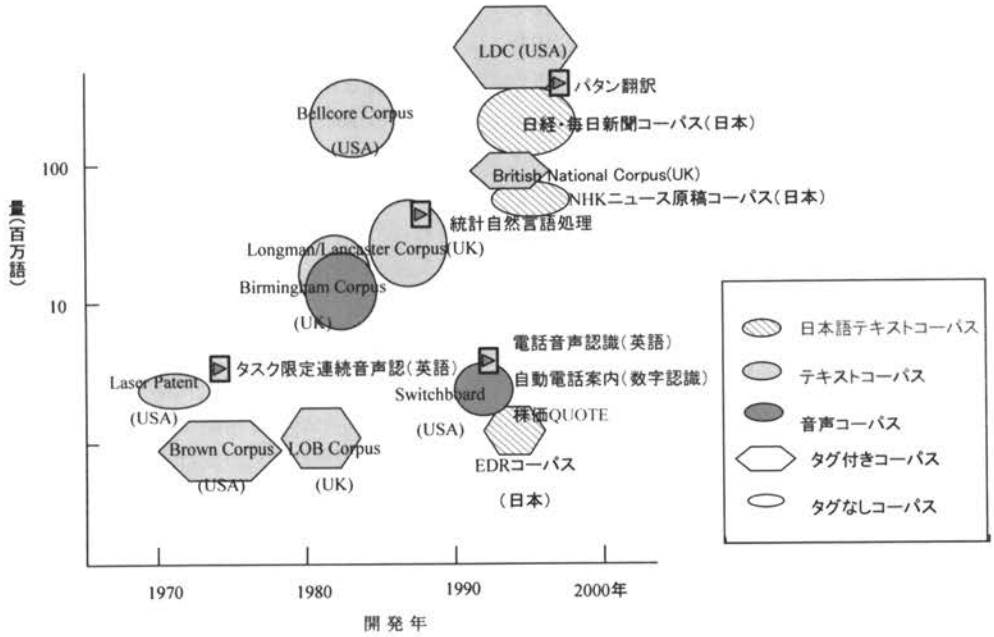
言語処理すべてを統計モデルベースで行うのは、音声認識と異なり複雑度が大きすぎて今のところ困難だが、例えば英語の品詞推定処理などでは、統計モデルによる手法が最も高精度かつ（データさえ用意できれば）容易に構築可能であるスタンダードな手法として確立されている。さらに、最近では、実際的な文法規則への確率量の導入についても、構造的曖昧性解消を主たるターゲットとして研究が進んでいる[6]。また、訳語選択への確率量の導入も有望な応用として研究が進んでいる。ただし、これらの研究は、英語のコーパスが整備されている事情を反映して英語をターゲットとした研究開発が先行しており、日本語に対する適用は遅れているのが実情である。

2.7 シソーラス

大量の言語データから知識を獲得するにあたって

---

4 奈良先端大など
5 ほぼすべてが HMM（Hidden Markov Model）を用いている
6 確率的 GLR（東工大など）や確率化依存文法など



は、語と語の意味関係をツリーの形で記述したシソーラスがきわめて重要である。この分野で有名なのは、米 Princeton 大学で構築された WordNet である。我が国の日本電子化辞書研究所で開発された EDR 概念辞書もこの範疇に入る。前者は無償で公開されていることもあって、（英語をターゲットとした）多くの言語処理研究で標準のツールとして使われている。

2.8 コーパス

上述したデータ指向のアプローチの前提となるのは大量のコーパスである。単に文を大量に集めただけでは有用性は限られており、品詞や構文構造など言語的な情報が付与されていることが重要である。欧米には、LDC（Linguistic Data Consortium）のコーパス群、British National Corpus（英）など、公的機関の支援を受けて作成された大規模（数億語規模以上）かつ高品質の言語的情報が付与されたコーパスがあり、広く提供されている。その成果は言語処理にとどまらず広く利用されている。例えば最新の英語辞書の語義配列はこうしたコーパスの分析に基づいており、重要語の統計データは直接辞書の説明に反映されている（BNC はもともと辞書のプロジェクトであった）。

一方、日本では、単に文例集としての新聞記事コーパスが入手できるにすぎない。最近、京都大学などを中心に新聞記事の単語分割と品詞付与作業を行う動きもあるが、資金面の裏付けが十分でないため、活動規模が小さいのが実情である[7]。

2.9 まとめ

機械翻訳システムは、開発開始より10年以上の歳月を経て、ようやく安定した評価と改良のサイクルに突入しつつある。訳質向上に重要なのは言語知識の強化であり、このために、従来のような開発者の直感に頼った職人的なやりかただけでなく、現実のデータを反映した知識を機械支援により獲得する工学的方法を本格的に導入すべき時期に来ている。次のブレークスルーはこの延長線上で起きる可能性が高い。これは、機械翻訳に限らず、検索、要約など言語処理技術一般にも言えることである。

その前提となるのが現実の言語データ、特に形態素区切りや品詞などの言語情報が付与された大規模なコーパスである。この種のコーパスは、欧米では、

7 その他、JEIDA による対訳コーパス作成プロジェクトなど数種のプロジェクトが存在する

人文科学と工学分野、また学界と産業界に共通のインフラとして公的な援助をベースにここ10年で大きく開発が進み、その成果が広く還元され、英語を対象とした研究開発が大きく進歩しつつある。一方、日本では、コーパス収集の環境整備がやっとはじまった段階であり、残念なことだがその立ち後れは明らかである（図２参照）。

3　市場

3.1　国内の市場動向

3.1.1　システムの普及度

インターネットの発達に牽引されて近年急激にパソコンが普及してきた。現在日本国内で年間約700万台のパソコンが販売されているが、その大部分のモデルに機械翻訳ソフトがプレインストールされている。インターネットによって居ながらにして世界中の情報にアクセスできるようになったため、現実に一般ユーザにとっても外国語に接する機会が急増し、情報を日本語に翻訳して取得する必要性が生じてきたためである。この事実からみても、機械翻訳ソフトはグローバルネットワークと接続された計算機環境にとってすでに必須アイテムとなっている。

3.1.2　市場規模

日本では、現在約20社が機械翻訳システムを商品化している。各社の売上を合計した国内の機械翻訳市場は現在約30数億円から40億円規模と推計されている。パソコンにプレインストールされる場合、一台あたりのソフトウエア単価は小さいが、数百万台のパソコンが対象となるため、それだけでも10数億円から20億円の規模の売上がある。また独立のパッケージソフトは１万円から数万円で販売されているが、各社の売上を合計するとやはり10数億円を越える売上があるものと推定されている。

従って、主要ベンダ各社はMT部門でほぼ単年度黒字を達成しつつあり、その中から開発費を捻出して改良を継続している状況となっている。

今後さらにインターネットが浸透し、パソコンの普及が進むので、2000年には機械翻訳市場は100億円を越える規模に拡大することが予想されている。

3.1.3　対象言語

日本における機械翻訳システムの対象言語は日本語と英語が大多数を占める。これは、文書を英語で書く必要のある割合、英語で書かれた文書を読む必要のある割合が、他の言語の場合と比較して圧倒的に多いためである。しかし日本と経済的に密接な関係がある国は英語圏ばかりではないため、話者人口の多い英語以外の言語の翻訳ニーズも高い。実際、韓国語、中国語を扱う翻訳システムも登場している。またスペイン語やフランス語などを含めた多言語翻訳の研究も進みつつある。

3.2　海外の市場動向

3.2.1　ヨーロッパ

ヨーロッパは多言語使用地域であり、古くから言葉の研究の伝統がある。このような背景により機械翻訳の研究開発も活発で、最近では、ベルギーのL&H社のように音声技術と言語技術を合わせ持つ企業の活動も活発化してきている。インターネットの普及により英語から各国語への翻訳のニーズが高まっているのも日本と同様である。

1996年にEuropean Commissionに提出された報告によると、36言語対に対し154のMT製品が入手可能であり、合計数百万のユーザが低価格製品を購入した。製品の数や言語対が急速に増える傾向にはないが、既存のベンダは言語対の追加を継続している。

European Commission（Telematics Program）では、言語処理技術を、すべての企業がネットワーク経由でグローバルマーケットにアクセス可能なElectronic Commerceの時代における基盤技術の一つと位置づけている。通信や決済等のインフラが整備された後にcriticalな課題として残るのは言語を介した人間同士のコミュニケーションであり、特に異言語間のコミュニケーションを適切にサポートする技術が、多言語経済圏であるヨーロッパ内部でElectronic Commerceを推進するうえで、あるいはヨーロッパの企業がグローバルな競争で生き残っていくために重要であるとの考えからである。ここでいう言語処理技術には、機械翻訳をはじめとして、クロス言語検索・要約などの異言語間情報流通技術を広く含む。

3.2.2　アメリカ

アメリカは世界各国から情報を収集する必要性から機械翻訳に対する大きな国家投資が継続的に行わ

れている。主要な公的支援として軍事関係資金がある。現在の世界の紛争（懸念）地域、例えば、朝鮮半島、セルビアークロアチア、中近東などの情報を収集する必要性から、使用人口の観点からはあまり多くないこれらの言語の研究にも投資が開始されている。これらはいずれも英語以外の言語を英語に翻訳する方向である。

公的機関を離れても、英語を母国語としない住民の増加と、英語圏以外の国へのインターネットの普及という二つの要因により翻訳のニーズが明らかに増大している。前者の要因により、例えばオンラインショッピング用のWebページは、スペイン語など英語以外の主要な言語でも用意することが常識になりつつある。また、Altavista等のWeb検索エンジンが機械翻訳を標準のサービスとして搭載し始めているように、インターネットが機械翻訳の利用を動機づけている状況は日本やヨーロッパと同様である。

4　技術的波及効果

4.1　要素技術のテキスト処理への応用

機械翻訳は自然言語処理の各要素技術を統合して作られる総合システムである。したがって機械翻訳の研究開発の過程で開発された各要素技術は、隣接関連分野に利用可能であり、実際広範な分野に利用され始めている。

例えば、日本語の文を単語に分割する形態素解析技術は機械翻訳の第一フェーズとして開発され、最高品質のソフトウエアでは約99%の精度を達成していると報告されている[8]。この形態素解析技術は、例えば、テキスト要約、テキスト分類、テキスト配信のパーソナル化といった自然言語テキストを扱う広範な技術分野で現在すでに使用されている。

4.2　音声技術、文字認識技術との融合

音声認識技術は、従来は音声波形を人間の音声として認識する音響処理部といわれる部分の研究が主であったが、近年、言語知識を用いて認識率を高める言語処理部といわれる部分の研究が活発化している。音声認識の言語処理部では、対象分野のテキストを単語分割して単語間のつながりやすさを学習しておき認識時の語彙選択に使うが、学習の際の単語分割にも、認識時の語彙認定にも形態素解析技術が活用されている。音声認識の音響的優先度、言語処理における優先度を統一的に扱う枠組も機械翻訳・自然言語処理の研究の中で開発されてきた。

また音声合成技術は、テキストの構造を認定しそれを音声として出力する技術であるが、その第一段階のテキスト解析で用いられる技術は、辞書構築技術、形態素解析技術、構文解析技術など機械翻訳の解析技術と共通する要素が多い。音声合成技術は機械翻訳技術と独立に発達してきたが、その高度化に向けて今後これらの分野の融合が進むものと考えられる。

音声認識技術と翻訳技術を融合させることで、人間と機械が対話する自然言語対話システムの研究も盛んになってきている。この場合、翻訳技術は人間の言葉を理解しそれを機械の言葉に翻訳する技術として使われる。言語理解の実用化にはまだ課題が多いが、機械翻訳の研究を通じて言語理解技術を発達させることにより、翻訳品質を向上させるだけでなく、対話システムへの応用が期待されている。また対話システムにおいて機械の応答する文の出力技術には、機械翻訳で培われた言語生成技術が応用される。

さらに音声認識技術、機械翻訳技術、音声合成技術を融合させたシステムとして人の話した言葉を別の言語として音声出力する自動通訳システムが研究され、すでに試作システムもいくつか構築されている。自動通訳システムが対象とするのは、従来機械翻訳が対象としてきた書き言葉とは性質の異なる話し言葉であるため、技術の融合と共に、言語現象のさらなる分析研究が必要である。自動通訳システムの実用化にはまだ越えなければならない課題が多いが、一般ユーザの関心の高い夢のある技術である。

また手書き文字を読みとって認識する文字認識技術は現在までは入力図形と標準パタンとの照合に技術開発の中心があったが、認識率が90%を越えつつある現在、今後のさらなる認識率の向上には言語知識との融合が必須である。

今後数年のうちに、音声技術、文字認識技術は真に実用化の時期を迎えると予想されるが、それを支

8 京都大学のJUMAN、奈良先端大の茶筅、NTTのJTAGなど

援するのが機械翻訳を中心とする自然言語処理技術であるといえる。

4.3 多言語辞書の応用

現在の情報検索は単一言語内に限定して行われるのが普通であるが、今後はインターネット上の外国語情報を我々の母国語である日本語で検索・収集するニーズが増大することが予想される。言語の壁をまたがって、検索対象テキストの言語とは異なる言語で情報検索を行う技術はクロスランゲージ情報検索[9]技術と呼ばれ、クロスリンガル情報流通の柱となる技術であるが、そこでは機械翻訳技術、特に多言語辞書が必須の言語資源として重要な位置を占める。

クロスランゲージ情報検索の一般的な構成では、まずユーザが与えるキーワード（検索式）を外国語キーワードに翻訳して検索を行う。このためには多言語辞書を参照する必要があり、キーワードの翻訳の質は辞書で決まると言ってよい。また、検索されてきた結果のテキストを日本語に翻訳することも必要であり、これには機械翻訳システムが使われる。キーワード翻訳の際には正しく意味が伝わるような訳語を選ぶことが重要だが、これには機械翻訳の訳語選択技術が応用できる。もちろん、入力が文とは限らないし、一つの入力キーワードに対して訳を複数与えてもかまわないなど、機械翻訳と異なる部分も多いため、発展のためには情報検索技術と機械翻訳技術という異なる技術分野の融合が必須である。

4.4 コーパスの応用

機械翻訳の高度化のために開発された自然言語コーパスは、言語処理技術一般に適用可能なリソースとして広く応用できる。欧米では古くから言語技術の重要性に対する認識が高く、コーパスを中心とした言語資源の充実が早くから行われている。コーパスは機械翻訳の研究だけでなく、一般の学習用の辞書や自然言語処理技術全般に広く応用されている。実際、古くから欧米でのコーパス研究は人間の学習者が使う辞書記述に応用されている。コーパスを分析して、個々の単語がどのように使われるかを実例文、頻度などで示した辞書が出版されており、学習者の理解の大きな助けになっている。

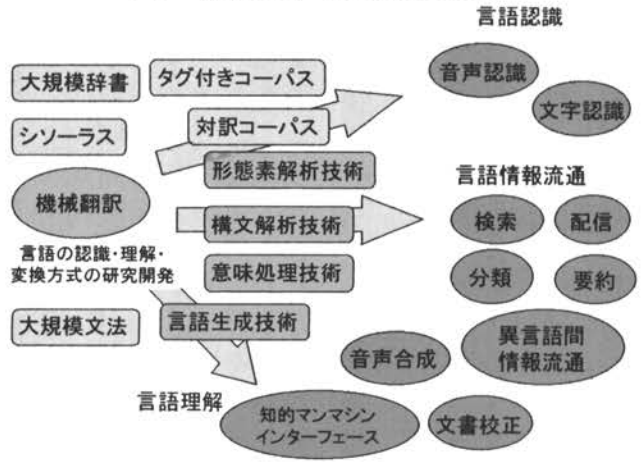
以上のように、機械翻訳は、単独システムとしての利用だけでなく、自然言語処理技術の統合システムとして、他の非常に広範の分野の技術と融合し、応用されつつある。

5 今後の動向と将来への提言

5.1 クロスリンガル情報流通

インターネットに代表されるグローバル情報ネットワークの発達と、パーソナルコンピュータの普及により、世界中の情報が各国の言語できわめて広範囲に流通する時代に突入した。情報ネットワークの普及と経済のグローバル化により国と国を隔てる地

図3 機械翻訳の技術的波及効果



9 Cross- Language Information Retrieval

勢的・政治的な障壁が急速に消滅しつつある中で、今後も残る最大の障壁は言語の壁であり、言語の壁を越えた情報の受信／発信／流通、すなわちクロスリンガルな情報流通をサポートし促進する技術の重要性が高まっている。そのようなニーズに応えるためには様々な自然言語処理技術を統合することが必要であり、そのなかでも機械翻訳技術はクロスリンガル情報流通技術の中心として重要である。機械翻訳技術で培った要素技術と言語変換の基礎技術・資源を、情報検索技術、テキスト要約技術、テキスト分類技術、テキスト集配信のパーソナル化技術などと統合して初めてクロスリンガル情報流通の時代の自然言語処理技術の構築が可能となるのである。

各国の状況を見ると、欧米では機械翻訳を中心とする自然言語処理技術の重要性が認識されており、その基盤研究への公的資金の投入が継続されている。ヨーロッパは多言語地域としての特質から翻訳の需要が非常に高く、それを効率化するため、機械翻訳等の自然言語処理技術の研究が推進されている。米国では、政治、経済、技術、軍事といった種々の分野の世界中の情報をいちはやく収集し、世界に先駆けた戦略を実行するため、多言語処理の研究が盛んに行われている。

これに比べ、日本は、言語処理技術の重要性に関する認識が希薄である。特に、言語処理技術を基盤技術として中長期的に育成していくという発想に乏しいように思われる。グローバル情報ネットワークの上では、地理的な距離は捨象され、言語的な距離が最大の障壁としてクローズアップされる。この点で日本はもともと不利な状況にあるにも関わらず、課題の認識と対処の両面で欧米に遅れをとっているのが実情である。今後、あらゆる活動がグローバル化する状況下で、自らのプレゼンスを示し世界の中で主要な役割を果たしてゆくためには、日本こそが率先してクロスリンガル情報流通の重要性を認識し、それを支える技術への継続投資をすすめることがきわめて重要であると考える。

5.2 企業・機械翻訳ユーザの立場

オンラインで得られる世界中の情報の９割が英語という現在の情報社会において、企業としては、一般の人が言語の壁を乗り越えて自由に世界中の情報を知ることができるようなシステム構築技術を開発することが課題となっている。十分に実用的な機械翻訳システムを構築するためには、辞書や文法の充実が必要不可欠であるが、現状の開発方法では限界に近づいており、それを打開するための大量の言語資源が必要な状況にある。翻訳のための資源としては対訳コーパスがもっとも有用だが、この種のコーパスはきわめて少ないため、公的機関とも連携をとりつつ、積極的かつ継続的な収集を行う必要がある。

機械翻訳ユーザは、基本辞書の他、ユーザ辞書として自分がよく使う文書に現れる単語などの翻訳知識を蓄積しているが、全国のユーザのユーザ辞書が共有できるようになれば、さらに辞書の充実がはかれるだろう。また、機械翻訳システムへのフィードバックも重要である。

5.3 言語資源の充実

日本語や英語といった自然言語は、人類の歴史と共に自然発生したものであり、人間が意図して設計し構築したものではない。この意味で、物理や化学の現象と同様の、一種の自然現象であると見ることができる。したがって言葉の研究は、自然現象の場合と同じく、まず事実の収集・観察から始めなければならない。実際に言葉が使われている状況を広く観察し、個々の語の振舞いを詳細に記述し、その背後にある規則をモデル化することによって初めて、言葉を技術として取り扱うことができるようになる。この意味で、コーパスの研究は自然言語処理技術にとって最も基本的な研究分野であり、コーパスをはじめとする言語資源は、人文科学と工学分野に共通の、また学術界と産業分野に共通のインフラとして非常に価値の高いものである。

しかし、大規模なコーパスの構築には、膨大な人的・時間的コストがかかる。日本語や英語の場合、数万語から数十万語が日常的に使用される。特に重要で使用頻度の高い数万語の単語の使われ方を調べるために、仮に各単語の使用例を100例ずつ集めたとしても、数百万の例文が必要となる。実際には、日常会話か、学術論文か、株価のニュースか、新聞記事か、といった語られる内容（ドメイン・分野）の違いによって、使用される語彙や表現が大きく異なるので、それぞれに対して百万規模の例文が必要となり、全体として必要なコーパスは非常に大規模となる。コーパス構築のための言語処理技術が高度

化してきたとはいえ、形態素解析、構文解析等の精度はまだまだ100%には達しておらず、したがって生のコーパスに品詞や文構造など必要な情報を付加するためには、一旦、タグ付与ツールや構造付与ツールを使って情報を付与した後、必ず人が目で見てチェックするという人的コストが必要となる。

さらに著作権の問題もある。一般に、一定量以上の質の良い言語データは、自然言語処理の研究機関とは異なる出版社や新聞社で開発される。そのため、言語データを利用したい場合、ユーザは個別にデータ保持者と著作権や価格を交渉しなければならず、多大な労力がかかる。また著作物を言語データとして自然言語処理技術に利用するという使い方が、データ保持者が今まで想定していなかった利用形態であるため、未だデータ提供の一般的ルールが出来上がっておらず、データ提供が円滑に行われないという問題が存在する。

コーパスの構築には上記のような問題点があるため、一つの機関が単独で十分な量の大規模コーパスを構築するのは困難である。そこで自然言語処理に関係する機関が協力して研究環境を整える必要がある。

米国、ヨーロッパはコーパス研究環境の整備の点で大幅に先行している。欧米の電子化コーパスの収集は20年以上の歴史を持っており、公的機関の後押しによって数億語規模のタグ付きコーパスも既にいくつか構築され利用可能な状態になっている。米国にはLDC（Linguistic Data Consortium）、ヨーロッパにはELRA（European Language Resources Association）という会員制コンソーシアムが設立されており、双方とも公的な支援を受けて活動を行っている。各所で開発された辞書やコーパスといった言語データを集積しそれらの利用を希望するユーザに配布するという仲介業務を行って、言語データの流通を促進している。

これに対し、日本ではコーパスを始めとする言語データに関しては、まだ収集・環境整備の段階であり、実際のコーパス収集、タギングは海外に大きく遅れをとっているのが現状である。そこで日本においても現在「言語資源コンソーシアム（GSK）」の設立準備が急ピッチで進められている。今後は、言語資源コンソーシアムが仲介することで、データ保持者との権利、価格の交渉も円滑化されてゆくことが期待される。

コーパスをはじめとする言語資源は、人文科学と計算機科学分野に共通の、また学術界と産業界に共通のインフラである。政府機関等の公的機関、大学等の研究機関、企業がその重要性を認識し共同で整備作業に取り組んでいくことが肝要である。

事務局だより

委員の交替・新任

［市場動向調査委員会］

株式会社インターグループ　退任　寺尾　卓

就任　国友　昭彦　翻訳営業部　課長

［UPF会議メンバー］

株式会社ノヴァ　退任　岡　俊行

就任　藤井美樹子　研究開発2部　主任

［ネットワーク翻訳研究会］

新任　江原　暉将　NHK放送技術研究所　先端制作技術研究部　主任研究員

新任　宮澤信一郎　秀明大学　国際協力学部　教授　コンピューター長

［インターネットWG］

新任　檜山　努　株式会社NEC情報システムズ　技術システム事業部第2技術部

文法形式と語彙化

NEC　C&Cメディア研究所　山端　潔

1．はじめに

機械翻訳システムの核になるのは言葉に関する知識ですが、この知識は通常、辞書と文法という形でシステム内部に格納されています。一般に、辞書には、各々の単語に対して、綴り・品詞・意味コード・対応する訳語など語彙レベルの情報が格納されており、文法には、単語の並びをどのように構造化し、変換・生成をどのように進めるかといった一単語を超えた部分の情報が格納されています。

このように、言語知識を辞書と文法に分けることは当然のことのようですが、その境界を具体的にどこに置くかは立場によって異なります。例えば、変形生成文法では、言語知識は専ら基本となる句構造規則と多数の変形規則を持った文法部門により担われていました。決して辞書情報が重要でないわけではないのですが、語の統語的振舞いはもっぱら個々の単語とは独立な文法部門の知識として一般化して記述されていたわけです。これに対して、変形生成文法の後継として登場したGB理論では、変形はmove-αに極小化され、言語知識の多くは、許容される構造を制約する複数のサブモジュールと語彙部門へと移されています。特に、語彙項目が拡大投射原理を通じて文の基本的句構造を支配することが明確化されています。さらに、LFGやHPSGなどの文法理論でも語彙部門が中心的な役割を果たしています。

以上は理論言語学の状況ですが、より工学的なアプローチを取る計算言語学でも、特に近年、辞書の相対的地位が高まってきています。その中心にあるのが文法の語彙化という概念です。文法の語彙化により、文法を辞書の中に取り込む形で二つを統一することができます。

本稿では、文法の語彙化についてインフォーマルな解説を試みたいと思います。多少テクニカルになる点があるかと思いますが、ご容赦ください。なお、機械翻訳システムには言語変換用の文法も生成用の文法もありますが、本稿では、主に構文解析を目的とした文法を想定して話を進めます。

まず第2節で一般的な文法形式について解説して準備とします。第3節で語彙化について解説し、第4節で具体的な語彙化文法形式をいくつか紹介します。

2．機械翻訳システムの辞書と文法

2.1. 文脈自由文法

文法形式としてもっとも単純なのは文脈自由文法（Context Free Grammar、CFG）でしょう。CFGは、VやVP、Sなどの品詞に相当するシンボル（非終端記号）、単語に相当するシンボル（終端記号）、シンボルをシンボル列で書き換える書き換え規則から構成されています。文法適用は、スタートシンボルと呼ばれるシンボルからはじめて書き換え規則を繰り返し適用し、右辺のシンボルを終端記号だけに還元することによりおこなわれます。図1は、自動詞文と副詞の組合せ" n v"、"n adv v"、"n adv adv v" など n adv＊v という単語列を生成する文脈自由文法の例です（＊は0回以上の繰り返し）

S→NP VP　（1）
VP→adv VP　（2）
VP→v　（3）
NP→n　（4）

図1　文脈自由文法の例

ここで、S、NP、VPは非終端記号、adv、v、nは終端記号です。スタートシンボルはSです。例えば" n v"という単語列は、

S → NP VP → n VP → n v　（5）
（1）　（4）　（3）

という順序で書き換え規則を適用して得ることができます。番号は矢印で適用される書き換え規則の番号です。

文脈自由文法の書き換え規則は、左辺をルート、右辺を葉として持つ高さ1のツリーで表現することができます。図2は図1の文法に対応するツリーの

集合です。書き換え規則を繰り返し適用する操作は、ツリーを葉から生やしていく操作に対応します。例えば（5）の書き換えは、図2の右側のツリーに対応します。書き換えのステップは、左側のツリーをシンボルが一致するノードで結合する操作に対応します。

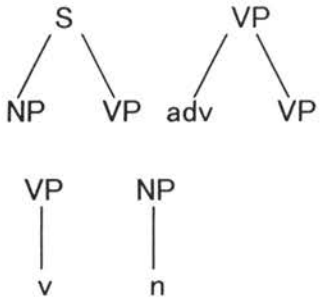


図2　文脈自由文法のツリー表現

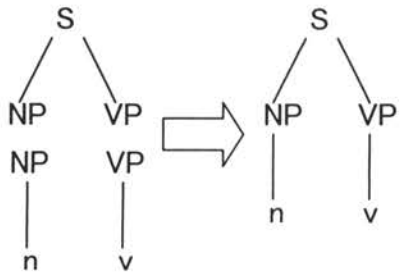


図3　ツリーの合成

このように、文脈自由文法は、ツリーとツリーの結合操作で定義することもできるわけですが、この定義は自然に任意の高さのツリーへと拡張できます。一般に、シンボルをノードに持つ任意の高さのツリーを基本構成要素とし、シンボルが一致するノードで二つのツリーを結合する操作を基本操作とする文法が定義できます。この文法は、Tree Substitution Grammar（TSG）と呼ばれ、最も基本的なTree Grammarです。シンボル列の生成能力の点からは、任意のTSGはCFGと強い意味で（構造的に）等価です。

もちろん、実用的な文法を書くにはTSGでは不充分で、ノードにシンボルではなくもっと複雑な情報、例えば素性構造を持たせるのが普通です。例えば、ユニフィケーション文法は、シンボルの代わりに素性構造を持ち、ツリーの結合のチェックには、素性構造間のユニフィケーションを用います。

2. 2. Tree Adjoining Grammar

TSGを拡張する方向として、シンボルを素性構造に置き換えるのではなく、ツリーどうしの結合演算を拡張するというアプローチがあります。このような拡張に関しては、ペンシルベニア大学のJoshiらにより開発されたTree Adjoining Grammar（TAG）が有名です。

TAGも、TSGと同様、ノードにシンボルを持つツリーを基本構成単位とし、ツリーの結合を基本演算とする文法形式ですが、TSGのようにシンボルを媒介にしてツリーを上下に連結するだけではなく、一方のツリーを他方のツリーの内部に「割り込ませる」演算（adjoining）が定義されているのが特徴です。具体的には、通常のツリー（initial tree）に加えて、adjoin演算用のauxiliary treeが構成要素に含まれます。Auxiliary treeは、ルートと葉のノードの一つが同じシンボルを持ち、これと同じシンボルを持つinitial treeの内部のノードに「割り込む」ことができます。

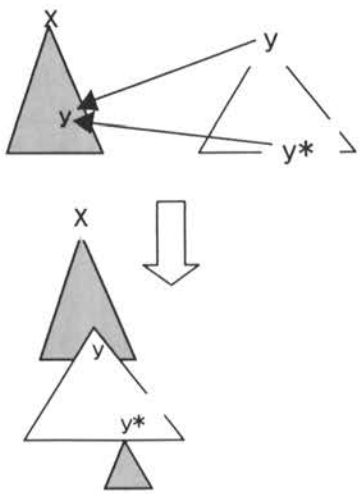


図4　TAGにおけるagjoinの操作

図4はこの「割り込む」adjoin操作を説明する図です。xをルートとする左のツリーが割り込まれるinitial tree、ルートと葉っぱにシンボルyを持つ右のツリーがauxiliary treeです。左のツリーの内部にはシンボルyを持つノードがあります。adjoin演算は、左のツリーをyのノードで上下に分割し、auxiliary treeを割り込ませます。もともとyの下にあった部分は、auxiliary treeのy＊の下にぶら下が

ります。TAGでは、このようなadjoin演算と、TSG式の結合演算を繰り返しながらツリーを成長させていきます。

この演算が導入された主要な動機は、長距離依存現象の扱いにあります。例えば英語のWH疑問文では、

What do you think he bought

のように、本来は動詞の目的語であるはずのwhatが動詞buyから遠く離れた位置にあらわれます。この距離を超えてwhatがbuyの目的語であることを認定できないと、正しい解釈を得ることができません。さらに、一般に

What do you think she said he bought

のように、WH語と動詞の間には任意個の動詞が入り得るので、WHが移動した全ての可能なパターンをあらかじめ文法規則として用意しておくという対処方法はとれません。

長距離依存は、例えばGB理論では、whatがmove-αによりbuyの目的語位置から繰り返し文を超えて文頭位置に移動する、という説明を行います。あるいは、HPSGでは、buyの目的語位置の単語の欠落が素性伝播の原則に従ってツリーの上を伝達され、最終的にwhatと出会って空所が埋まる、という説明をします。どちらの説明も、単純な句構造文法の枠には納まらず、movementや素性の伝播といった別の道具を必要とします。

TAGでは、adjoining操作が導入されているおかげで、句構造文法の枠内でこの現象を説明することができます。そのためには、動詞が主語と目的語を取り込むことを示す通常のツリーの他に、

what does he buy

のように目的語がWH化されてその文の先頭に現れることを記述するツリーをinitial treeとして用意しておきます。そして、do you thinkやshe saidのように間に挿入されるツリーをauxiliary treeとして用意しておきます。遠距離依存構造は、これらのauxiliary treeをWH語と残りの部分の間にadjoin演算で挿入していくことにより生成することができるのです。このように、ヘッドの単語が周囲の語を支配する局所的構造をあらわすツリーと、実際の文で支配の局所性を壊す原因となるツリーを別々に記述し管理することができるのがTAGの特徴です。

2.3. 依存文法

句構造文法としばしば対比される文法形式に依存文法があります。依存文法は、単語と単語を直接依存関係で結んで構造化を行う文法形式です。図5に依存文法による解析の一例を示します。句構造文法との最大の違いは非終端記号に相当する中間ノードがないことです。そのため、依存関係を結ぶ順序を可変にするのが容易であるという特徴があり、日本語のように語順に自由性が高い言語の文法記述にしばしば用いられます。

依存文法は、単語の内部にすべての文法情報が格納されている点で後述の語彙化文法の枠組みにそっており、語彙と文法を統合した確率モデルを作るのが容易であるなどの利点があるため、最近では、英語など従来句構造文法が主流であった言語の文法記述にも使われることが多くなってきています。

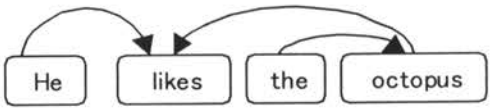


図5　依存文法による解析例

3．文法の語彙化

実際にあらわれる言葉は実に多様なので、文法開発を進めれば進めるほど規則の数は増えていきます。その一方で、文法規則を増やせば増やすほど、解析結果の曖昧性が増えていき、正解を選択するのが困難になります。文法が実用的であるためには、構造記述能力が高いだけでは不充分で、適切な曖昧性解消が導入できる枠組みになっていなければなりません。

一般に、曖昧性解消には具体的な語彙項目の情報が非常に重要です。例えば、動詞がどのような補語をとるか、どのような名詞が主語になりやすいかなどは、抽象的な構造規則だけからは決めることができません。

文法の語彙化は、文法と辞書を分離せずに、一体の形式として扱おうというアプローチです。構造の記述能力が増えるわけではありませんが、曖昧性解消や文法開発の観点から利点が多いので、最近では多くの文法開発プロジェクトで採用されるようになりつつあります。

語彙化した文法では、文法はすべて辞書の中に語

彙と関連付けて格納されます。正確には、語彙化文法は以下の要素からなる文法として定義されます。

1）各々が語彙項目と関連付けられた有限個の構造

2）構造を結合する演算

ここで構造といっているのは、TSG や TAG のツリーに相当し、演算はツリーの結合のことです。構造は、通常、関連付けられた語彙項目の辞書の中に格納され、その語彙項目と一緒に辞書からロードされます。

この構成から明らかにわかることですが、語彙化文法には、入力単語と無関係な文法規則は最初からシステムにロードされない、という利点があります。そのため、文法が大きくなってもパフォーマンスが落ちにくくなります。文法開発が進むと、あるところから個別語に特有の詳細な文法・表現を記述していく必要がありますが、これら個別語専用の文法規則をいくら増やしても他の語の処理に影響を与えないわけで、これは重要なポイントです。また、熟語も、複数の語彙項目と関連付けられた文法規則として、この枠組みで自然に記述できます。

別の利点として、文法開発の見とおしがよくなるという利点があります。文法規則というのは、様々な単語の振る舞いを観察している中から一般化して作られるものです。文法開発者が苦労する点の一つは、その規則がどこまで一般化できるのか、その境界を正確に見極めることにあります。これが狭すぎれば、ある単語がその文法規則から取り残されることになりますし、広すぎれば過剰に文法規則が適用されてしまい、誤った結果を出力することになります。語彙化されていない文法では、適用されるべき単語にマーク（一般には属性－属性値の組）を付与しておき、文法規則中でこのマークをチェックすることにより規則適用を制御しますが、このマークが増えるとだんだん管理し切れなくなってきます。それに対して、語彙化文法では、マークを介した間接的な制御ではなく、直接「この単語にはこの文法を適用する」と直接対応させてしまうわけですから、管理がずっと簡単になります。その代償として、同じ文法規則をたくさんの単語に繰り返し書かなければならない、という欠点がありますが、これはソフトウェア的に別の方法で対処可能です。

別の重要な利点として、曖昧性解消の際に、語彙の情報と構造の情報の双方を容易に利用できることがあります。特に、近年盛んに研究されている、確率モデルによる曖昧性解消において、語彙項目に関わる確率と構造に関わる確率の双方が自然な形で融合できるという利点があります。

曖昧性解消は、自然言語処理の最大の問題の一つですが、近年、電子化されたコーパス（文例集）が大量に手に入るようになったのにともない、大量のデータを統計処理して曖昧性解消に利用する研究が盛んに行われています。基本的には、曖昧性解消をしたい対象物、例えば統語構造の集合に確率構造を導入し、実際のデータから具体的な確率分布のパラメタを推定します。例えば、確率的文脈自由文法（PCFG）では、文法規則を単位に確率を付与し、木構造の確率は、その構造の構成要素である文法規則の確率の積として計算されます。ところが、構造としては同じであっても、周囲の環境で構造の現れやすさは異なります。例えば、that 節補語をとる動詞の格パターン規則

VP→V thatC

があったとして、これがどの動詞に属する規則かでそのあらわれやすさはまったく異なります。つまり、V 以下にどのような語彙項目が展開されるかに依存するのですが、上記の文法規則を単位に確率を付与するのでは、このような語彙的な傾向をとらえることができません。一方、この規則の動詞部分が語彙化されて、

VP→” think” thatC

VP→” believe” thatC

のように別々の規則として書かれていれば、これらにお互いに異なった確率を与えてこの語彙的傾向をとらえることができるわけです。

第一章で説明したように、理論言語学でも句構造の大枠は語彙が支配する枠組みになっていることを考えれば、適切な句構造の選択のために語彙の情報が必須であるのは当然と言えるでしょう。語彙化文法は、こうした文法と語彙の関連を自然に捉えることのできる枠組みとなっています。

4．語彙化文法のいろいろ

本節では具体的な語彙化文法についていくつか紹介します。

4.1. 文脈自由文法と語彙化

単純な文脈自由文法に対しては、対応する語彙化が簡単に作れます。例えば、図6は図1の文法を語彙化したTSGです。CFGとTSGが等価であったことを思い出してください。

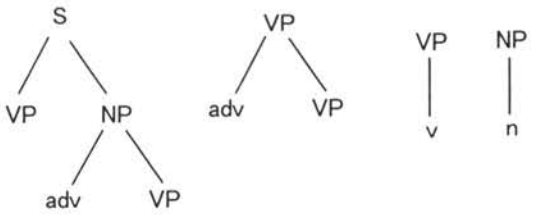


図6　図2のTSGの語彙化の例

一般のTSG（CFG）に対しては、対応する語彙化（LTSG）が存在するとは限りません。これは、LTSGでは、任意のツリーのルートから語彙項目までの距離がある一定値（基本ツリーでの距離の最大値）以下であるのに対し、一般のTSGでこれを満たさないものが作れることからもわかります。一般のCFG（TSG）の語彙化は、LTAGと同様なadjoin演算を追加してTree Insertion Grammar（TIG）に拡張して考えることにより可能になります。

また、自然言語の句構造規則には、通常、ヘッドの概念を導入することができますが、このようなCFGでは、各規則をヘッドの単語と関連付けることによっても語彙化が可能です。

4.2. Lexicalized Tree Adjoining Grammar

Lexicalized Tree Adjoining Grammar（LTAG）はTAGの語彙化です。すなわち、すべての基本ツリーが葉のノードのどこかに語彙項目を持つようなTAGとして定義されます。

各々の単語に付随するツリーはその単語の"extended domain of locality"を表現しています。すなわち、各語彙項目は、自分の支配する（通常は自分をヘッドとする）ツリーを持ち、自分をヘッドとしてどのような構造が生成されるかを各語彙項目が知っています。

図7（a）に単独の直接目的語をとる格フレームを持つ動詞が持つツリーの例を、図7（b）に二重目的語を持つ動詞のツリーの例を示します。さらに、これらのツリーに受身変形を行ったり、目的語をWH移動させたりしたツリーの一群が必須格パターンごとにあらかじめ用意されており、tree familyとして動詞と関連付けられています。

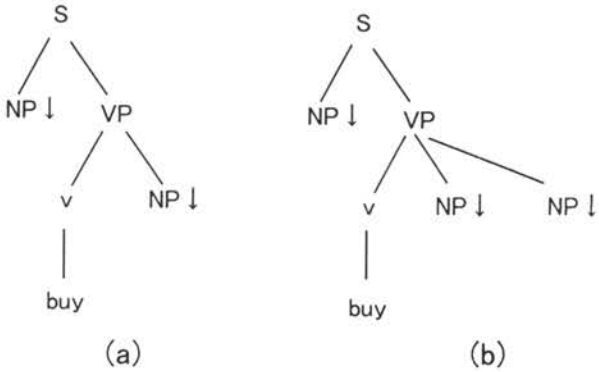


図7　LTAG動詞に付随するツリーの例

現在も、ペンシルベニア大学で、ユニフィケーションベースのLTAGをベースとした英語文法開発が進んでおり（XTAGプロジェクト）、辞書、文法、開発環境等がフリーで配布されています。最近では、今年（98年）の8月末に最新の文法がリリースされたようです。

XTAG関連ではいろいろ面白いアクティビティがあるのですが、その中からSuperTaggingについて紹介しておきましょう。従来はタギングといえば品詞タギングのことであり、もっぱら構文解析前に統計モデル等を使って品詞の曖昧性解消を行うために使われていました。ところが、LTAGでは、単語の統語的性質は、品詞ではなく単語に付随するツリーにより直接表現されています。そこで、品詞タギングと同様なテクニックで、ツリーの曖昧性解消をやってしまおう、というのがSuperTaggingの基本的なアイディアです。SuperTaggingの後では、使用すべきツリーが一意に決まっているので、例えば構文解析は単にこれらを結合して構造化するだけの処理となります。LTAGにおける品詞とSuperTag（ツリー）のパラレリズムというのは面白い着眼点だと思います。

4.3. 依存文法

依存文法は、単語と単語を直接依存関係で結んで統語構造をつくる文法形式です。句構造文法のように非終端記号による中間ノードがないため、文法要素は自然な意味で語彙項目を含み、その意味で語彙化の枠組みに合致しています。

4.4. Link Grammar

カーネギメロン大学のSleatorらによるLink

Grammarは、依存文法に類似した形式を持ちます。基本構成要素は単語と単語を結ぶ「リンク」です。リンクは方向と種類を持ち、同じ種類のリンクが出会うことにより単語どうしを結んで構造をつくります。単語は単なるリンクの束であり、すべての統語的性質はこの束によって完全に表現されています。純粋な依存文法とは、ヘッドとディペンダントの区別がない等の点で異なります。

Link Grammarでは、単語にすべての構造化規則(リンク)が書いてあり、それが文法記述のすべてであるという意味で、完全に語彙化された文法のよい例です。辞書(文法)とパーサーがフリーで公開されています。最近、パーサーを外部のアプリケーションで利用できるようにするためのAPI(Application Program Interface)が公開されたようです。

5．おわりに

機械翻訳をはじめとする自然言語処理で用いられる具体的な文法形式について、形式の間の関係と語彙化を中心に駆け足で説明しました。この分野は、自然言語処理技術の中でも最も形式整備の進んだ分野だと思いますが、新しい概念やテクニックが次々と導入されているアクティブな分野でもあります。その紹介として少しでもお役に立てば幸いです。

なお、煩雑になるのを避けるために、詳細な文献情報は挙げませんでしたが、関心のある方のために一般的なポインタを示しておきます。

まず、TAG、LTAG、XTAG関連の情報は、
http : //www.cis.upenn.edu/
http : //www.cis.upenn.edu/~ircs/homepage.html
http : //www.cis.upenn.edu/~linc/home.html
から手に入ります。レポート類もフリーでftpできるものが多いです。

Link Grammarに関する詳細情報は、
http : //www.link.cs.cmu.edu/link/
からアクセスできます。

また、計算言語学関連の論文を集めたオンラインアーカイブがhttp : //xxx.lanl.gov/cmp-lg/
にあります。最新の結果が盛んに投稿されているので、関心のある方はぜひ眺めてみてください。

事務局だより

'98年3月～9月にご入会された方々

—敬称略—

（個人会員）

信田　恵壱		'98年3月
池田　　昭		'98年6月
蔡　　東風		'98年6月
佐藤　幸浩		'98年6月
宮澤信一郎		'98年6月
加藤　秀和		'98年7月
小島　順一		'98年7月
久保　泰朗		'98年7月
西垣　　通		'98年7月
Carl Kay		'98年8月
小島　照章		'98年8月
高野　　治		'98年8月

（法人会員）

バーンジャパン株式会社	業種：　ソフトウェア開発 所在地：　東京都新宿区 担当窓口：翻訳チーム　野村　真弓	'98年9月

業界動向

産業翻訳業界概観

技術翻訳者　佐藤幸浩

はじめに

一般的に「翻訳」というと、どちらかというと文芸翻訳の方面から論じられる場合が多い。教育現場やマスメディアでも翻訳といえば、やはり取り上げられるのは文芸翻訳の場合が多い。しかし、市場規模が大きく、また社会の現実をより的確に反映しているのは、文芸翻訳ではなく産業翻訳である。「産業翻訳」は、別名「ビジネス翻訳」「技術翻訳」ともいわれる。また、文芸、映画、ビデオ、芸能などの翻訳を除いたすべての分野を「実務翻訳」という場合もあるが、市場と翻訳の方法で大部分は「産業翻訳」と共通している。

1　産業翻訳の特徴

1.1　大量で納期と翻訳均質性が重要

産業翻訳では、取り扱い説明書や仕様書の翻訳が多く、翻訳ドキュメントの制作を製品の発売にあわせている。また、文芸翻訳とは違い「翻訳」は「作品」ではなく、「生産物」としてとらえている。そのため納期と均質性がきわめて重要である。

1.2　新概念や新語が次々に登場

翻訳の対象が文字どおり「産業」であるため、翻訳の対象は常に技術革新が進んでいる。翻訳者も産業にかかわっている以上技術革新のスピードにあわせて知識を習得する努力が必要となる。翻訳するドキュメントは、新製品や新技術であるため、そのなかで使用されている用語は、まだ世間一般に普及していない場合が多い。そのため翻訳者は、業界の最新資料やインターネットなどを駆使して、不明な言葉を調査しながら翻訳を進めている。

1.3　新製品発売前後に大きな翻訳需要

技術革新のスピードが速い分野では大きな翻訳の需要がある。例として、パソコン用のOSの変遷で、MS-DOS→Windows 3.1→Windows 95→Windows 98という変遷では、各段階でマニュアルや解説書の書き換え需要があり追加翻訳がでてくる。

1.4　日本人でも和文英訳をする

翻訳業界には、「母国語をターゲット言語とする」という原則がある。文芸翻訳ではこの原則がよくあてはまり、日本人が和文英訳することはほとんどない。しかし、産業翻訳では、和文英訳に大きな需要があり、加えて英語圏出身（英語ネイティブ）の翻訳者が慢性的に不足している。そのため日本人も日常的に和文英訳をしている。

2　各分野の動向

表1に、翻訳会社・翻訳事務所が取り扱う分野の

表1　翻訳会社取扱い分野の比率

分野	全体に占める割合%			増減
	1982年	1992年	1997年	
コンピュータ	8	21	17	増
電気・電子	25	19	16	減
機械・プラント	32	20	18	減
通信・精密	5	14	16	急増
鉄鋼・エネルギー	4	3	1	減
医学・薬学	4	4	9	微増
土木・建築	12	9	13	横ばい
化学・環境	10	10	10	横ばい

比率を示す。

表1からわかるとおり、ここ15年間でコンピュータと通信・精密の占める割合が急増している。電気・電子、機械・プラントは、82年には大きな比率を占めていたが、徐々に落ち込んできている。鉄鋼・エネルギーも減少、土木・建築、化学・環境はほぼ横ばい、医学・薬学は微増であった。

コンピュータと通信・精密の増加は、明らかにパソコンとインターネットの急速な普及に呼応している。それから、筆者の推測であるが、機械や電気製品では内部に必ずマイコンやデジタル回路を使用している。そのため、機械や電気製品とはいってもマニュアルや解説書の内容がコンピュータ的内容である場合がある。このことから、コンピュータの占める割合は、表に示す割合よりも大きいものと考えられる。

以上から、これから産業翻訳に携わりたい人や、現在携わっていて業務を拡大していこうと考えている人は、コンピュータ、通信・精密の分野に力を注ぐことをお勧めする。

3　業界の特徴

3.1　クライアント獲得にはコネが重要

翻訳業界に限った話ではないが、仕事を獲得するにはコネが重要である。一個人が見ず知らずの翻訳会社の求人広告に応募しても、トライアルでたいてい落ちるか、あるいは仮にその会社に登録できても、仕事をもらうことは容易ではない。産業翻訳の業界では、翻訳者として独立後、以前勤めていた会社から仕事を請け負うといったケースや、知人から請け負うといったケースが非常に多い。また、パートナーや弟子が、先生から客先を融通してもらうというケースも多い。

3.2　下克上の業界

翻訳業界は、事務所や会社の設立が比較的容易である。経営者は労務管理を誤った場合、社員が謀反を起こしていとも簡単に独立する。その際に経営ノウハウ、顧客、登録翻訳者を奪う場合が多い。

4　仕事の流れと業務

図1に仕事の発注元から、フリーランス翻訳者までの仕事の流れの経路を示す。ここでいう発注元とは、大手メーカーや商社、公官庁である。ドキュメント製作会社とは、印刷会社や出版会社である。従業員も50人以上で、印刷、編集、翻訳、営業などの

図1　仕事の流れ

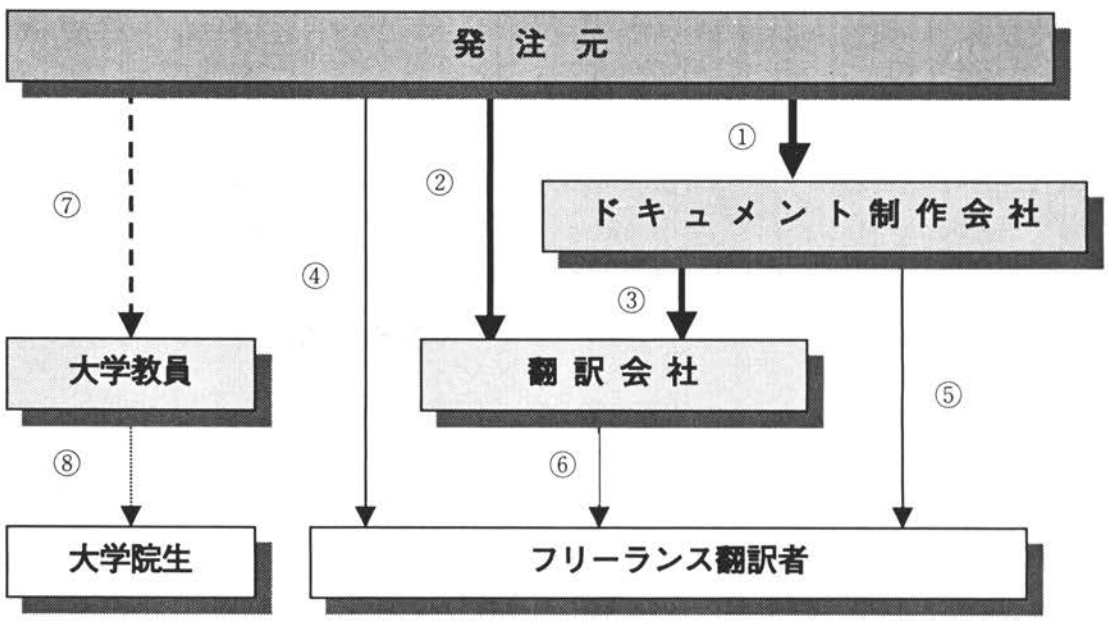
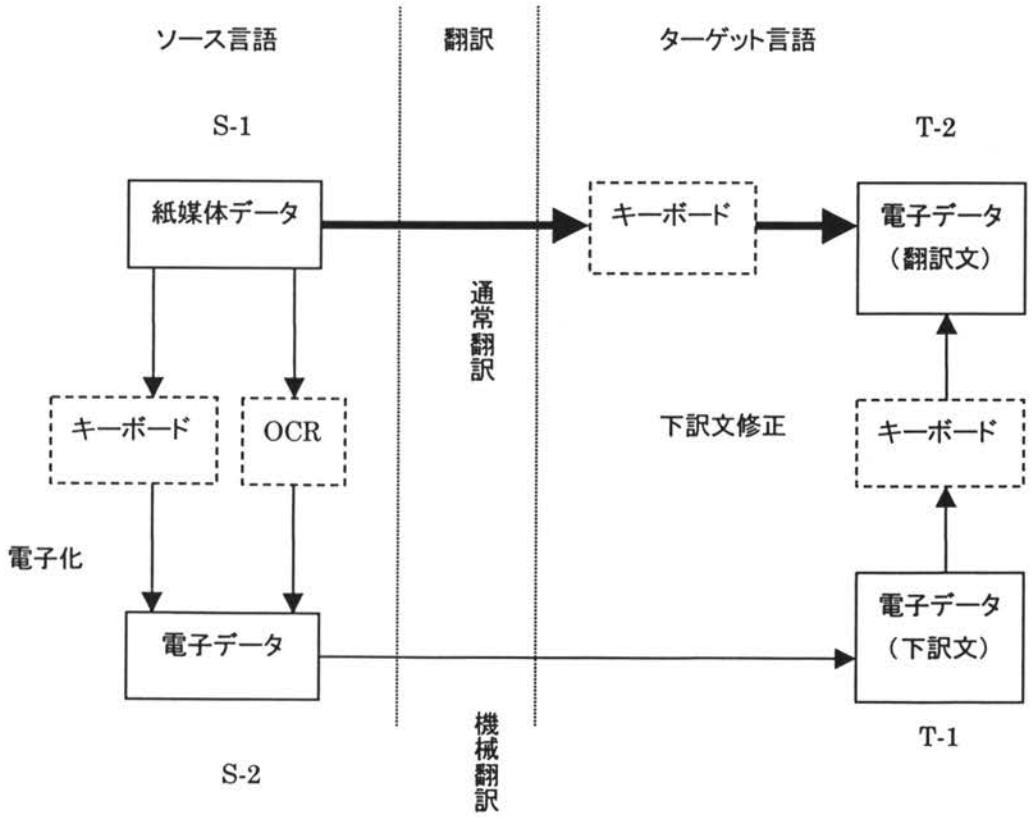


図2　翻訳のプロセス



セクションに分れている会社である。翻訳会社は、翻訳のみ、あるいは翻訳と版下作成を請け負う会社で、3人から5人ぐらいの規模の会社である。
通常、ドキュメント製作会社は、翻訳・版下作成・印刷をパッケージで請け負う（①）。その中から翻訳の部分を切り分け、翻訳会社やフリーランスの翻訳者に発注する（③、⑤）。翻訳会社やフリーランスの翻訳者が発注元クライアントと直接の取り引きをおこなう場合もある（②、④）。

5　翻訳のプロセス

図2に翻訳のプロセスを示す。最も一般的なプロセスは、S－1→T－2のプロセスで、紙媒体のソース言語（英文和訳では英語原稿）から、ターゲット言語（日本語）の訳文をワープロに入力していく方法である。現在、翻訳者の多くは原稿を紙媒体で受け取ることが多いため、多くの翻訳者はこのプロセスで翻訳をおこなっている。

一方、一部の翻訳者の間で機械翻訳が試みられており、その場合のプロセスは、S－1→S－2→T－1→T－2（原稿を電子データで受け取った場合には、S－2→T－1→T－2）となる。今のところ、T－2に至るまでの時間は従来の翻訳プロセスの方が速く、また簡便である。しかし今後、翻訳ソフトの改善が進み、スピードの面でメリットがでてくるか、あるいはデータ保存における有効性が高まれば、機械翻訳を取り入れた翻訳プロセスが普及していくに違いない。

6　英文和訳に役立つテクニカルライティング

本項では、日本語テクニカルライティングを取り上げてみたい。テクニカルライティングの手法は、英文和訳の場合に役に立つ。また、和文英訳で、原稿の日本文がわかりにくい場合に、テクニカルライティングの手法で、原文の日本語を訳する前にわかりやすく書きかえるというテクニックにも応用でき

る。

6.1　翻訳臭さの少ない自然な日本語表現

和文英訳においては、読み手は日本人なので、極力違和感のない日本語に訳す。翻訳者でも熟練者と初級者の違いは、自然な日本語表現ができるかできないかという点に明瞭に現われてくる。ただし、例外もあり、ソフトウェアのマニュアルでは直訳調が好まれる。

6.2　文中の漢字比率30%前後が読みやすい

文中の漢字比率は30%程度が読みやすく、漢字が多すぎると、見ためが黒々として心理的に抵抗がでてくる。逆に漢字が少なすぎると、意味を把握するのに手まどる。しかし、ここでも例外はソフトウェアのマニュアルである。この種のマニュアルでは、カタカナ語を多く使用しなければならないため、漢字の比率が相対的に下がる。

6.3　一文には一つの事柄を記述

一つの文にたくさんの事柄が書いてあるとわかりづらくなる。そのため一文には極力ひとつの事柄を記述するようにする。例えば、原文を下訳して、「ディスプレイに電源を入れ、次に外付けディスクユニットの電源を入れ、いちばん最後に本体の電源を入れます。」と訳したとすると、このとおりにはせず、「最初にディスプレイに電源を入れます。」「次に外付けディスクユニットの電源を入れます。」「最後に本体の電源を入れます。」というように分けて書くとわかりやすくなる。

6.4　必ず一義的となること

2通り以上に解釈できる文を書かない。例えば、「青い端子が光る部品」とせず、読点をつけて「青い、端子が光る部品」か「青い端子が、光る部品」と書いて、意味が明瞭になるようにする。

6.5　一文の文字数が50字以内を目安

ひとつの文章が相当長い英語の原文は、翻訳する際に二つの文に分ける工夫が必要である。分け方の基準としては、一文の字数が50字以内とする。ただテクニカルライティングでは50字以内としているが、筆者の英文和訳の経験上では、50字以内に収めることはほとんど不可能であった。そのため、筆者は文字数の基準を60字以内においている。

6.6　業界の習慣に従った表現

一般的に原文の意味する内容を自然な日本語に正確に置き換えるのが質のいい翻訳である。しかし、業界の日本語表記の習慣に従わない場合、いくら質がよくても品位を落とす場合がある。放送業界に放送禁止用語があるように、業界によって使用すべきではない言葉があるので留意が必要である。一例として家電業界では、根拠や数字を明確に表示しない場合、「世界一」とか「最高級」といった表現は、使用しないという習慣になっている。

6.7　不快な印象を与えるような表現は避ける

たとえ原文に書かれているとしても、特別な場合を除いて、特定に人に不愉快な印象をあたえる表現は、避けるか婉曲な表現に書き換える。例えば原文を忠実に訳して、「当社は、インドやベトナムといった後進国から、日本やアメリカといった先進国にまで、幅広く輸出しています。」とするよりは、「インドやベトナムといった国々から、日本やアメリカといった工業諸国にまで、幅広く輸出しています。」というように書き換える。

7　翻訳者が頻繁に直面する問題

7.1　和文英訳で原稿の日本語の質が良くない

和文英訳において、原稿の日本語の質が良くないということは、以前より多くの翻訳者から指摘されてきている。質が良くない文章とは、何をいいたいのかが明確でない文章、流れが論理的でない文章のことをいう。小説や随想であれば、論理はそれほど問題ではない。しかし、技術文や仕事文であれば、論旨が明確で論理的であることに越したことはない。明確で論理的な日本語を書こうとすれば書くことはできる。しかし現実には翻訳者に回ってくる日本語原稿は、不明確で論理的性に欠ける場合が多い。では、なぜこのような日本語の文章がでまわるのか。その理由として次の3点をあげてみたい。

1．学校で「人に説明する文章」や「論理的な文章」を書くための教育を受けてきていない。
2．マニュアルや仕様書作成といった業務は、企

業内で補助的作業ととらえられている。そのため、経験の浅い社員が担当する場合が多く、経験豊富な人が書いてはいない。

3．企業内で質の高い文章を書ける人が、ドキュメント作成部門に配属されているとは限らない。質の高い文章を書ける人は、たいてい自己表現が上手な人で、このような人たちは社内で開発、企画、広報（花形３K 職）といった部門で働いているケースが多い。

原稿の日本語の質がよくないということに関しては、書く人の個人的表現力ばかりではなく、社会の構造的な問題も絡んでいる。

次に、翻訳者・翻訳コーディネータがどように対応すればよいかを検討してみたい。根本を解決するような特効薬はない。しかし、ある程度の工夫によって対処可能である。一つ目として、英語に翻訳する前に、日本語の原文をテクニカルライティングの手法で書き直すこと。二つ目は、能力の高い翻訳者に訳してもらうこと。明確さや論理性に欠ける日本語をそのまま英語に訳したり、あるいは能力の低い翻訳者に訳してもらうと、後工程の翻訳チェックで必ず余計な手間がかかる。

7.2　訳語統一の問題

翻訳者が苦労するもう一つの大きな問題に訳語統一がある。ドキュメントを通しての訳語の一貫性は、産業翻訳、特にコンピュータ関連の翻訳では重要課題である。この課題に対しても特効薬はない。対応策として考えられる次の方法について検討してみたい。

・翻訳の段階で翻訳者が注意
（メリット：確実　デメリット：手間がかかる）

・翻訳支援ソフトの活用
（メリット：ドキュメントが少品種・大量の場合に効果あり　デメリット：カスタマイズが難しい）

・ワープロの検索機能の有効利用
（メリット：操作が容易　デメリット：この機能だけでは限界がある）

7.3　商取引上の問題

発注側の問題として、翻訳者に契約額を支払わないとか、経理上の都合で納期や支払日を頻繁に変えるといったことが指摘されている。翻訳者側の問題として、請け負った仕事を途中で放り出すといったケースがある。

7.4　翻訳者の健康上・職業上の問題

肉体的・精神的問題として、ほとんどの翻訳者は OA 病（目の疲れ、慢性的肩こり、指の不調）、運動不足、慢性的なストレス（締め切りと継続的精神労働に対するストレス）といった問題を抱えている。現実問題、30代で突然死に至った翻訳者もいる。

社会的問題として、翻訳者には社会性の欠如している人が多いということが指摘されている。独立開業しやすいため、唯我独尊の人、会社や組織で従業員として向かない人が翻訳者になっているので、うなずける話である。

参考文献

アルク社「翻訳辞典」1983
アルク社「翻訳辞典」1993
アルク社「実務翻訳ガイド」1998
日経 PB テクニカルコミュニケーション協会編「文章・用事用語スタイルブック」
The ITI Japanese Network「IJET–97 Eighth International Conference on Japanese–English Translation and Interpreting」

（技術翻訳者　AAMT 会員　佐藤幸浩）

国際会議報告

COLING–ACL'98参加報告

NEC C&C メディア研究所　土井伸一
NEC C&C メディア研究所　池田崇博
通信・放送機構渋谷上原リサーチセンター　若尾孝博

1．はじめに

COLING–ACL'98は、本年8月9日～16日にカナダ・ケベック州のモントリオール大学にて、

・COLING'98：17th International Conference on Computational Linguistics

・ACL'98：36th Annual Meeting of the Association for Computational Linguistics

の両会議の共催として、ICCL（the International Committee on Computational Linguistics）とACL（the Association for Computational Linguistics）の主催で開催された。9日は8件のチュートリアル、15、16日は11件のワークショップが開催され、本大会は10日～14日の間行われた。機械翻訳を始めとする自然言語処理の分野では史上最大規模の国際会議の一つであり、総数で300件以上の発表、800人近い参加者があった。以下、始めにCOLINGとACLについて解説し、続いて本大会の概要を述べ、最後に機械翻訳に関連するワークショップとして、Translingual Information Managementをテーマとしたワークショップの内容を報告する。

2．COLING/ACLについて

COLINGは、毎回200件近い発表がある自然言語処理の分野では最大規模の国際会議で、2年に1度開催されている。最近の開催地は、90年がフィンランドのヘルシンキ、92年がフランスのナント、94年が京都、96年がデンマークのコペンハーゲンであった。今後は、北米、ヨーロッパ、アジアの3地域で順次大会を開催することとしており、次回2000年がルクセンブルグを中心にドイツ、フランスを含めた3カ国で、その次2002年は、台北で行うことになっている。

一方ACLはアメリカの研究者が中心の学会であり、年大会の開催地も、94年ニューメキシコ州ラスクルーセス、95年マサチューセッツ州ケンブリッジ、96年カリフォルニア州サンタクルスと、アメリカ近辺に限られていた。しかし最近では学会の国際化を進めており、昨年の大会は、EACL（ACLのヨーロピアンチャプター）との共催でスペインのマドリッドで開催された。今回のCOLINGとの共催もおそらくこの流れの中で行われたものであり、両大会の共催は、1984年にスタンフォード大学でCOLING–ACL'84が開催されて以来14年ぶり2回目である。

さらに、今大会中に開催されたACLのビジネスミーティングでは、この国際化のための計画が中心議題となっていた。ACLは、今後2000年までにヨーロピアンチャプターに加えて北米チャプター、アジアチャプターを作り、各チャプター地域毎の年次大会を開く計画である。アジアチャプター設立のための担当は、香港科学技術大学のデカイ・ウー先生が指名されていた。また、この地域別チャプターの大会とは別にACL Internationalの大会を、北米、ヨーロッパ、アジアの3地域で順次回して開催する予定でもある。ACLの北米チャプターを作るにあったては、ビジネスミーティングの場で賛否両論の意見が出された。同時に地域会員とInternational会員の関係、会費の問題などが討論されていた。この国際化計画の推進については、まだまだ全体の同意を得るには時間がかかりそうである。なお、次回のACL–99は、来年6月にメリーランド大学で開催される。

3．COLING-ACL'98

今大会は、自然言語処理の主要な二つの国際会議の合同だったことと、やはり北米での開催ということもあって、多くのメンバが参加して活気のある大会となった。総参加者数は772名で、日本からもアジア太平洋機械翻訳協会前会長の長尾真京大総長、現会長の田中穂積東工大教授を始めとして、多くの参加者があった。日本からの参加者数は78名で、アメリカからの346名は別格として、地元カナダの76名を抑え第2位であった。その他主なところでは、英国とフランスから52名、ドイツから42名が参加した。

今大会の論文発表件数は以下の通りである。

論文種類	件　数
Invited paper	2
Regular paper	135
Project note	100
Demonstration	35
Student paper	12
Workshop paper	120

最近のCOLINGでの発表が200件弱、ACLでの発表が50件程だったのに対して、今回は全体としてはそれらの合計を上回る発表件数となっている。しかし、レギュラーペーパーの数は絞り込まれて、代わりに今回から新しく登場したプロジェクトノート&デモセッションで多くの発表がなされた。

このプロジェクトノートは、50名程が入る中規模の部屋で1コマ30分の発表を行うもので、さらにデモを行う場合は多くの場合2コマが割り当てられ、この1時間を自由に使う（30分のデモ付き発表を2回行っても良いし、30分発表の後に30分デモを行っても良い）ことができるものであった。単に口頭発表だけでなくデモを見たり、小人数で活発な質疑応答を行ったりすることができた点は良かったのだが、初めての試みということもあって、

- 各発表はテーマ別に分けられておらず、また予稿集も著者名順に並んでいて論文のタイトルからは検索できない形式だったため、自分が聴講したい発表を見つけるのが大変だった。
- しかも、小人数を対象にした発表でありながら、ポスターセッションのように各部屋を自由に行き来できる形ではなかった。
- 予定の部屋が直前まで授業で使われていて、予約していたデモ用の機器が準備されていないことがあった。
- 昼休みが20分しかなく、プロジェクトノートの最初のセッションには遅れる人がほとんどだった。

等の問題点や事務的な不備も一部で目立っていた。

自然言語処理の基本分野での論文の傾向としては、シンボリックな言語処理に統計処理を統合させたものが、一般的になっているようであった。例えば、シェフィールド大学Y.Wilks教授の論文"Word Sense Disambiguation using Optimised Combinations of Knowledge Sources"がそうである。単語の意味の曖昧性を解消するのにレキシコン、統語、意味などの種々の情報を組み合わせて高精度のDisambiguationを実現させているが、この組み合わせの決定には、標準的な機械学習の手法が使用されていた。多量の機械可読の言語データが入手可能になるにつれて、特に80年代後半から盛んになった統計的処理であるが、種々の言語処理に適用してみる試みが一段落した感があり、今大会では、従来のシンボリックな手法と統計処理を上手く組み合わせることが中心になって来ていると思われた。電子協のテキスト処理技術専門委員会が昨年度、COLING、ACL等の最近の自然言語処理に関する主要国際会議における技術項目ごとの発表数の推移を分析する技術調査を行っている［1］が、そこでも、現在活発に研究が行われているコーパス使用や統計処理に関係する論文の比率はあまり増加していないことが報告されている。やはり、コーパスや統計に基づく自然言語処理の研究は、基本的には定着してきているが、今後は実用化に向けて伝統的な手法といかに融合させるか等、新たな展開が求められていると言えるだろう。

その他の機械翻訳関連では、やはり最近の傾向として、純粋な解析、生成、翻訳といった発表はあまり多くはなかったが、談話／対話処理、音声翻訳等に関して、Regular paperのDiscourse、Anaphora、Dialogue、Speech Translation、Multimodal NLPの5項目のセッションの合計で23件にのぼる発表があった。プロジェクトの中間報告的な発表や小規模の実験に基づくもの等、今回は内容的には目新しいものは少かった。しかし今後、文章や音声言語、マルチモーダル情報の処理は必ず必要となる技術であり、機械翻訳の応用分野としても発展していくことは必須である。また、セッション以外では、L&H社がディクテーション製品のデモを行っていた。

情報検索関連では、テキストの要約結果を検索に利用するという発表がいくつかあった。例えば、クエリに適合するテキストのうち、クエリに関連する部分を要約して、次回のクエリに追加するリレバンスフィードバック方法や、ニュース記事のリーディングテキスト部を抜き出して、その中に限定したキーワードのマッチングを行うことで適合率を上げる方法などである。また、文章中のキーワードの5W1Hに基づく役割を指定して情報を検索することで、キーワードだけによる検索よりも検索ノイズを削減

する情報分類ナビゲーションの提案もあった。さらに、情報抽出・要約関連では、辞書にある固有名詞の認定が現状でも高精度できていることから、コーパスから固有名詞を抽出し、そこから前後の文法構造などによるルールを獲得することで、未定義の固有名詞の種類を特定する手法、語彙連鎖に着目してテキスト中の主題部分を判定するテキスト要約手法などの発表があった。

4．COLING-ACL'98 post-conference workshop : "Translingual Information Management : Current Levels and Future Abilities"

以下では、16日に行われた標記のワークショップについて報告する。このワークショップは、今後、マルチリンガル・マルチモーダルな情報を対象として、機械翻訳、情報検索、音声認識、情報要約等、自然言語を対象とする各分野の壁を乗り越えた研究を進めていくために開催されたもので、Eduard Hovy（ISI/USC）、Nancy Ide（Vassar College）、奥村（NEC）らがコミッティーを務めている。今回のワークショップは、今年の6月に、スペインで開かれた同ワークショップの続きという位置づけで、前回の内容を踏まえ、各分野で現在どのくらいのレベルのことが可能になっているか、各々の機能を組み合わせることで、近い将来どんなシステムが実現できるか、マルチリンガル・マルチモーダルな情報を扱えるようにするために考えなければいけないことは何か、という点についてまとめたものを代表者が発表するという形式で行われた。以下では、その中の情報抽出・テキスト要約、機械翻訳のセッションの内容について述べる。

4.1. Information Extraction and Automated Text Summarization

情報抽出とは、大量の情報から予め与えられるテンプレートに合致するものを抽出する操作をいう。テンプレートは通常、ある種のイベントに対応しており、誰が何をどうしたというスロットを埋める形で情報が抽出される。テキスト要約は、抽出の基準を与えるのではなく、テキストで述べられている主題を抽出するものである。要約の方法には、文字通り要約文を生成するものと、テキストから必要な文だけを抜粋するものとがある。

現在、情報抽出では、決定木を利用して、名詞部分や動詞部分、節のパターンの認定、抽出すべき部分かどうかの判断を行っている。現状で、60から70％の正解率を達成しているが、これで十分かどうかは問題である（人間では65％から80％との分析がある）。また、テンプレートをよりフレキシブルなものにしたり、適用ドメインを拡張する際のコストをいかに小さくしていくかという課題がある。テキスト要約では、適切な評価法がないため正確なことが言えないが、新聞記事では、先頭の2割を取ってくる方法で十分なこと、元のテキストの2割以下に圧縮した場合には意味がとれなくなること、どんなテキストでも、ぜひ抽出すべき文と絶対にいらない文が必ず入っているが、6割くらいの文は、そのどちらとも言えないことなどが分かってきている。

将来の研究課題として、特に、音声認識結果やOCR済みテキストなどの誤り含みのテキストへの適用、また、複数の文書を対象とするイベントの抽出や要約手法の確立などが残されている。

4.2. Machine Translation of Text and Speech

これまでに、原言語から目的言語に直接変換するものと中間言語を通すもの、言語学的な見地からルールを書いたものと統計的にルールを集めたもの、ルールベースの翻訳と用例ベースの翻訳、複数の翻訳手法を組み合わせるマルチエンジンの手法など、さまざまな翻訳手法が提案されている。汎用ドメインの翻訳システムでは、性能は必ずしもよいと言えないが、安価なPC用ソフトやWebでのサービス等、広く普及してきている。一方、限定ドメインのシステムであれば、例えば天気予報の翻訳など、98％の精度を上げているものもある。しかし、ドメインを限定したシステムでは、そのドメインが十分大きくないとコストが割に合わないこと、多くのドメインを管理するのは事実上不可能なことが問題となっている。

現在、語彙的・文法的知識、対訳などのリソースをどう収集し、どう活用するかということが大きな問題である。統計的な手法で知識の獲得が容易になったとしても、深層の概念の獲得までできる訳ではなく、限界がある。言語学的なアプローチや統計的な手法の結合を考慮していかなければならない。一方で、マルチエンジンのアプローチも重要である。

ユーザは時間のかかる高価な翻訳システムを求めているわけではない。機械翻訳の本質的な問題ではないが、真の汎用システムの実現のためには複数エンジンの効果的な組み合わせを研究する必要がある。また、他のシステムに機械翻訳エンジンを統合していくためには、各言語処理モジュールが共通のリソースを共通のフォーマットで扱えるようにしていかなければならないだろう。

[参考文献]
[1] 日本電子工業振興協会：平成10年3月：98－情－9
「自然言語処理システムに関する調査報告書」

事務局だより

協会活動報告

（1998年7月～9月）

運営委員会	7月24日	①第8回通常総会および関連行事の報告 ②シンガポールサミットの企画　③委員会・研究会の活動状況報告 ④会員および委員の増員と活性化対策　⑤AAMTの収支見通し ⑥事務局の体制強化策
	9月18日	①MT白書作成・シンガポールサミット参画の方針 ②MT市場調査の企画　③シンガポールサミットの企画
市場動向調査委員会	9月25日	①映像翻訳についてのヒアリング　講師：水谷美津夫氏 ②MT市場調査の企画と取り組み
技術動向調査委員会	8月31日	①シンガポールサミットの準備状況報告 ②キャノンにおける機械翻訳システムの事例報告 ③UPFの活用策　④MT白書の作成
編集委員会	9月4日	①AAMT Journal No. 23の反省　②AAMT Journal No. 24の企画 ③シンガポールサミットの準備状況報告
インターネットWG	7月15日	①UPFダウンロードサービスの進め方 ②ホームページのメンテナンス
	9月8日	①UPFサービスの進め方　②メーリングリストの検討 ③カウンター設置に関する報告
シンガポールサミットPC	7月10日	①メインテーマの検討　②Call for Paper（論文募集）の作成 ③プログラムの検討　④準備スケジュール
	9月18日	①組織委員会・実行委員会の準備状況　②プログラムの内容 ③展示会　④チュートリアル・ワークショップ・レセプション

MT Summit Ⅶ
September13—17, 1999, Singapore

テーマ： MT in the Great Translation Era
期　間： 1999年９月13日（月）～17日（金）
会　場： シンガポール国立大学内 KRDL
参加予定人数：200名（日本50、シンガポール50、アジア各国50、欧米50）

MTサミットの準備が日本とシンガポールとで協同しつつ進められていますが、シンガポールから写真が届きましたのでご紹介いたします。

シンガポールの象徴になっている
マーライオン像と高層ビル

KRDL の本館。椰子の木や熱帯樹に囲まれた静かな環境にある。

広大なシンガポール国立大学構内の小高い丘の上にある KRDL

新製品紹介

チャットもできる翻訳ソフト

The翻訳インターネット

株式会社　東　芝

1．はじめに

東芝は、「The 翻訳」シリーズ第２段としてインターネット上のWeb情報や、チャット翻訳機能やインターネットニュース／メールの翻訳機能を搭載した「The 翻訳インターネット V3.0」を発売しました。

The 翻訳インターネットは、24万語の英日辞書と12万語の日英辞書を搭載し、翻訳精度を向上しています。また、HTML3.2への対応に伴いスタイルシート（CSS１）や、Java など、原文のレイアウトをより忠実に維持しながら翻訳することが可能です。さらに、ユーザ認証が必要なホームページでも指定した時刻に自動的にダウンロードして自動翻訳する「ユーザ認証付きオートパイロット機能」や、現在表示しているホームページの画像データとテキストデータをパソコンにダウンロードできる「キャッシュ機能」を装備しています。

2．主な機能

（１）チャット翻訳

日本語が分からない人と英語が分からない人との間のチャットをサポートするための機能を有しています（注１）。

この機能は、チャットツール（Microsoft®Net Meeting および Netscape　Conference）のインターフェースそのままで翻訳を行えるため、新たな操作を覚える必要がありません。また、入力した日本語を英語へ翻訳してから送信、受信した英語を日本語に翻訳してから表示することができるため、チャット相手側は従来のチャットツールを変更する必要がないという特長を持っています。さらに、口語翻訳用のパターンを提供し、翻訳をスムーズに行えるようにしています。

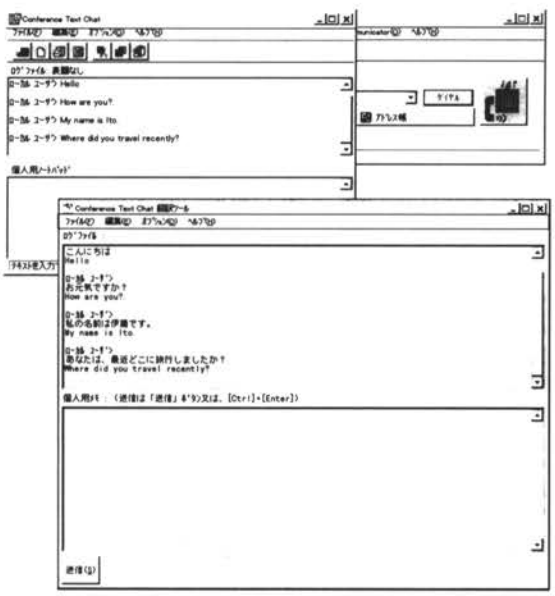


図１　チャット翻訳

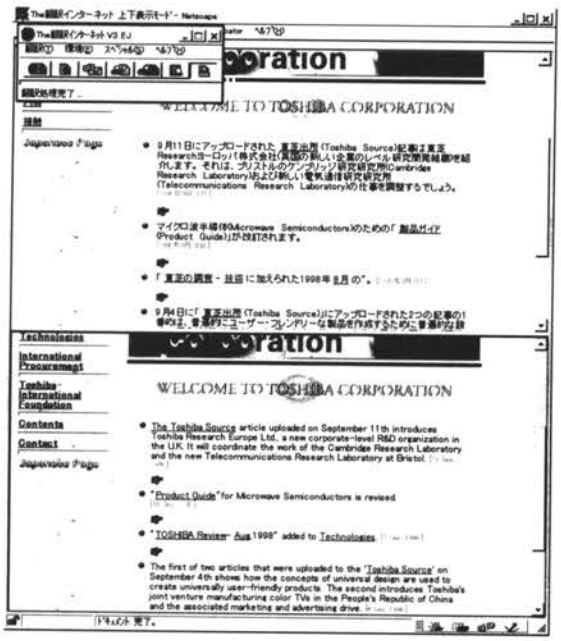


図２　ブラウザ翻訳

（2）ニュース／メール翻訳

受信した英文のインターネットメールやニュースを日本語に翻訳することが可能です。

（3）スタイルシート対応の Web ページ翻訳

HTML3.2に対応し、Web ページの翻訳後のレイアウトとオリジナルの画面の差異が少なくなり忠実性がアップしました。また、スタイルシート（CSS 1）の情報も翻訳結果に反映が可能となり、従来より対応していた JAVA 等と合わせ、最新の Web ページの翻訳も可能となりました。

表示ページの翻訳をはじめ、ダウンロードしながらの翻訳、キャッシュ登録を行った後の翻訳、ドラッグ範囲の翻訳など用途に応じた様々な翻訳形態を提供しています。

（4）用途に応じた翻訳結果表示

画面の表示方法は ①画面を２分割し、原文・訳文を同時表示、②１文ごとに原文と訳文を表示、③訳文のみを表示、という３つの表示方法から選択することができます。また、上下分割の場合は自由に比率を変えることができ、使いやすさと見やすさが向上しました。

さらに、訳文の頭にあるマークをクリックするだけで、対応する原文を呼び出すことができる「翻訳ハイパーリンク」を採用しており、効率よく英文のホームページの閲覧ができるほか、英語学習などにも役立ちます。

また、目標のページに素早く到達するために、ページ間のリンク関係のみを階層的に表示することができます。

（5）ユーザ認証付きオートパイロット

指定した時刻に指定したホームページにアクセスし、自動的にダウンロードして自動的に翻訳をおこなう「オートパイロット機能」を引き続き採用しています。今回はさらにユーザ認証機能を付加しているので、パスワードの入力が必要なホームページでもパスワードが指定してあれば、自動ダウンロードが可能です。

（6）ファイル翻訳

テキスト形式で保存されたファイルだけでなく、HTML 形式のファイルに対してもタグを保存したまま翻訳をおこなうことができます。

（7）その他の機能

本ソフトでは、既存のインターネット環境に合わせて自動的に環境設定をおこなうので、インストール後すぐに使いはじめることができます。

また、英日基本辞書24万語と日英基本辞書12万語を標準装備し、他に最大約20万語が登録できるユーザ辞書機能を採用しているので精度の高い翻訳が可能です。さらに、オプションの専門用語辞書を使用すれば、より高度な翻訳結果が得られます。

3．動作環境（全て日本語版）

OS：　Windows®95/98,
　　Windows NT®Workstation4.0

HW：PC/AT 互換機、PC98シリーズ
　　Pentium®以上推奨／MEM２４M 以上
　　ディスク55MB

ブラウザ翻訳対象ツール
　　Netscape Navigator3.01～4.05、
　　Internet Exproler3.0～4.01

ニュース/メール翻訳対象ツール
　　Netscape Collabra4.0～4.05、
　　Netscape Messenger4.0～4.05、
　　Microsoft®Outlook Express4.0

チャット翻訳対応ツール
　　Netscape Conference Text Chat4.0～4.05、
　　Microsoft®NetMeeting2.0～2.1

お問い合わせ先

株式会社東芝　コンピュータ・ネットワークプロダクト事業部（03－3457－2725）

製品情報 URL：http：//eiplaza.toshiba.co.jp

カタログご請求先：http：//eiplaza.toshiba.co.jp/pcserver/astransacctlg.htm

（注１）チャットができるのは、英日／日英双方向版のみです。

・Microsoft, Windows, Windows NT は米国マイクロソフト社の米国ならびにその他の国における商標です。
・Pentium は Intel Corporation の登録商標です。
・その他の名称は、各社の商標、登録商標の場合があります。

新製品紹介

英日・日英翻訳支援ソフト「PENSEE V3.0」

沖ソフトウェア株式会社

1　はじめに

沖ソフトウェアでは、沖電気工業株式会社での長年にわたる研究開発成果を1994年よりPCパッケージの「PENSEE」シリーズとして製品化を行っております。このたび、パーソナルユースからプロユースまでの幅広い範囲でご活用いただける本格的な英日・日英翻訳支援ソフトウェアであるPENSEE Vの最新版 Ver3.0のリリースを行いました。Ver3.0では、最新の翻訳エンジンを採用し、更に翻訳精度を向上、翻訳編集機能などのパワーアップによりより使いやすくなっております。

2　機能のご紹介

2.1　WORD連携翻訳

他社に先駆けいち早くWordの最新版 Word98からのダイレクト翻訳を可能にしています。Word98のツールバー上にPENSEEの翻訳アイコンが表示され通常のWordのオプション機能と同様な感覚で、ボタンをクリックすることで翻訳を実行することができます。もちろんWord97,95との連携も可能です。

2.2　スマート辞書機能

今までの翻訳ソフトでは、翻訳品質を上げるために単語登録を行っていましたが、PENSEEでは、2つの離れた単語すなわち、慣用表現（イディオム）が登録可能となっています。これにより、翻訳時の文書解析力が向上し、より文法に沿った的確な訳を得ることができます。

例文：I take train at Tokyo.

"[take]＋[train]＋[at]＋[場所]"という並びを[場所]で電車に乗ると解釈するイデオムを登録する。

登録前：“Tokyoで列車で行く。”

登録後：”私は、Tokyoで電車に乗る。”

このイデオム登録で、Tokyoの部分は、抽象的に「場所」として登録しますので、TokyoがOsakaに変わっても、また、主語が変わった場合でも正しい翻訳を実行します。

スマート辞書への登録も、図に示す簡単な登録画面よりガイダンスにしたがって登録することができます。日英版でも同様の効果を得ることができます。

2.3　PENSEEエンジン98搭載

最新版、PENSEE翻訳エンジン98を搭載、97年度版に比べて単語数大幅アップ日英20万語、英日15万語をシステム辞書に搭載、また最新の文書にも対応できるように各種新語登録および構文解析力強化を行っています。

例）舟木が金メダルを獲得した。

PENSEE97エンジン

A ship tree acquired a gold medal.

PENSEE98エンジン

Funaki acquired a gold medal.

例）He is all of a twitter.

PENSEE97エンジン　彼は、さえずりの全員である。

PENSEE98エンジン　彼は、早口でしゃべる。

また、別売の専門用語辞書「PENSEE辞書PRO」を組み合わせることでさらに翻訳品質を向上させることが出来ます。

2.4　エディタ機能の大幅強化

●マルチドキュメントサポート･文書サイズ制限無し

文書を翻訳している時に、さらに別の文書ファイルを翻訳してみたくなることがありませんか？PENSEE Vでは、同時に複数の文書ファイルを開い

て翻訳作業を効率よく行うことが可能です。また、開ける文書ファイルのファイルサイズに制限はありません。（搭載メモリ容量など、システムリソースの制約は受けます。）

●マルチユーザサポート

PENSEEは、完全なマルチユーザ環境をサポートしています。翻訳環境や辞書の優先順位などユーザ毎のPENSEE環境を保持します。会社などで同一マシンを複数人で利用する場合に個人個人の環境を分けることが出来ます。

●未知語検索機能

翻訳する文書の中にシステム辞書、ユーザ辞書に含まれていない単語（未知語）がないか検索、表示することができます。これらの未知語を辞書に登録することにより翻訳時には快適な翻訳を実行します。

●ワープロ感覚の編集機能

翻訳文書としてテキストファイル以外にRTF形式ファイルをサポート。原文の文書のフォーマット変換なしにそのまま翻訳できます。また翻訳結果をワープロ感覚のエディタで文書整形でき、完成した文書に仕上げることが出来ます。

●電子メール送信機能

完成した文書、対訳文書をそのまま、電子メールで送信することが出来ます。翻訳した文書は、そのまま世界へ電子メールで送信することができます。(注意　MAPI対応のメール送信ソフトが必要です。)

2.5　「らくらく英文メール」付加（日英のみ）

「らくらく英文メール」は、豊富な例文テンプレートを選択しながら、簡単に英文の手紙を作成します。テンプレートは、株式会社創元社　「パターン活用ビジネス英文手紙」を利用し、ビジネスのいろいろな場面での英文レター作成を強力に支援します。

●レターテンプレートを利用した簡単レター作成

「らくらく英文メール」は、あらかじめビジネスで使う英文レターのテンプレートを用意しています。この中から使うテンプレートを選択すると、あなたの名前、住所などは自動的に置換します。また、相手先の名前なども、ガイダンスにしたがって入力するだけで英文レターを作ることが出来ます。

●多くの例文を用意

らくらく英文メールは、テンプレート以外にもよく使う英語の例文を用意しています。

例）カテゴリ　ありがとう

日本語　　ご連絡ありがとうございました

英語　　Thank you for letting us know.

いろいろなカテゴリ別に整理された例文を利用することで英文レター作成を強力にバックアップします。

●使いやすい日本文・英文同時表示

らくらく英文メールは、テンプレートで作成した文書の日本語、英語の両方を表示、英文レター作成中に日本語を見ながら作業を行うことができます。

●翻訳機能連携

翻訳ソフトウェアのPENSEE翻訳エンジン98を搭載、らくらく英文メール内から、日英翻訳を行うことで、テンプレートで足りない部分の作成が簡単に行えます。

●お気に入り登録機能

よくお使いになる表現を登録することが出来ます。例えば電子メールでよく使われる顔マーク（＾ｏ＾）などをワンタッチで英文レターに貼り付けることが可能です。

3　動作環境

OS：	Windows95 Windows98 WindowsNT4.0
必要メモリ	16Mbyte以上（Windows95Windows98）
	32Mbyte以上（WindowsNT）
HDD容量	英日　32Mbyte　日英　45Mbyte
	カップリング版　65Mbyte

4　お問い合わせ先

沖ソフトウェア株式会社
東京都板橋区舟渡１－12－11　ヘリオスⅡビル
営業部 TEL03－5970－9952
E-Mail : pkg@okisoft.co.jp
製品情報 URL http : //www.okisoft.co.jp/

新製品紹介

国際契約書システム V1.0

株式会社富士通ラーニングメディア

1．はじめに

国際ビジネスの進展とともに、企業は、国際的な取引の増大を余儀なくされています。このような国際取引においては、契約書が基本となります。

従来の英文契約書作成においては、和文の原案を作成し、翻訳し、さらに専門家のチェックを受ける等、多大な時間（平均１ケ月）と費用がかかっていました。本システムでは、ユーザが必要とするほとんどの書式（600編５万文）が和英対訳で用意されているので、それを参照し、一部を変更するだけで原案が完成します。さらに、自動翻訳ソフトATLASと豊富な文書編集・辞書引き機能を使って簡単に翻訳できるので、短期間（平均３日）かつ若干の費用（従来の約1/10）で１つの契約書を完成させることができます。

2．特長および主要な機能

2.1　大規模契約書データベース

本契約書データベースは、国際事業開発株式会社が、10余年の歳月をかけて作成したものです。

データベースは、５編に分かれており、それぞれ標準書式集と差替条項集が用意されています。

（１）物品取引編（輸出入、代理店、販売店等）
（２）技術取引編（技術輸出、技術導入等）
（３）資本取引編（国際合弁、会社定款等）
（４）大型取引編（大量売買、機械設備売買等）
（５）特殊取引編（業務委託、共同開発、雇用等）

これらのデータは、通常のファイルとしてオープンできる他、後述するATLASの文例参照翻訳用にパターン化されています。また、一つの書式をオープンした状態で、他の書式や差替条項を参照することもできます。

2.2　ATLAS文例参照翻訳

契約書の文は、全般的に長く、特殊な表現が多いことから、通常の機械翻訳では、まともな翻訳結果を得ることが困難です。本システムでは、契約書の文が比較的定型化されていることに着目し、文例参照翻訳（コーパス翻訳）を採用しています。

契約書集のすべての文は、パターンデータとして準備されています。名詞（句）の部分を変数に指定してあるので、ユーザが標準書式や差替条項の一部（変数部分）を変更しても、元の文型を維持したまま訳文を生成します。以下、実際の例を示します。

標準書式

和文	販売店は、年間を通じて最低（　）回の購入を行うことに同意する。
英文	Distributor agrees to make at least（　） purchase of Products throughout the year.

ユーザ更新和文に対する文例参照翻訳結果

変更和文	会社は、年間を通じて最低１回の購入を行うことに同意する。
文例参照翻訳	Company agrees to make at least one purchase of Products throughout the year.

2.3　用例参照機能

契約書データベースのすべての標準書式と差替条項は、用例集として参照することもできます。また、検索キーで、そのキーが含まれている文（和文・英文のペア）を簡単に検索、表示できます。（図１）

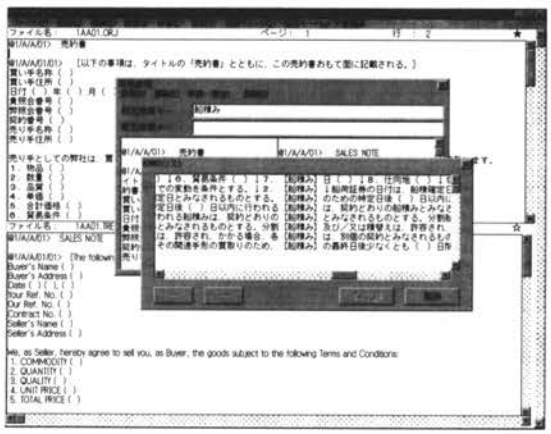


図１　用例参照画面

2.4 辞書引き機能

ATLASで使う辞書とは別に、文書編集中に簡単に画面上で引ける辞書（基本辞書：日英／英日各14万語、専門用語辞書：日英／英日各52万語［オプション］）も準備されています。（図2）

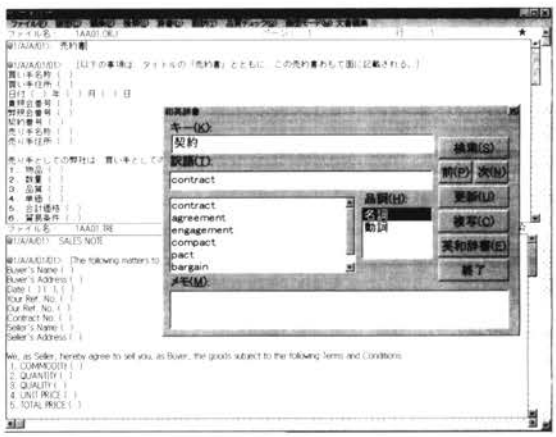


図2　辞書引き画面

2.5 ATLASの他の機能

システムには、ATLAS for Windowsの機能が装備されていますので、辞書の選択、優先順位の設定も簡単に行うことができます。また、ATLAS for Windows用の専門用語辞書も利用できます。

2.6 文書チェック機能

和文チェッカでは、文の長さのリスト、見出し語リスト、さらに日本語の文書中に含まれるさまざまな文字列リストを自動的に出力します。「漢字・カタカナ」リストで出力される単語は、機械翻訳用の辞書項目として登録すると便利です。

オプションで提供する英文チェッカを使えば、スペルはもちろん、英文法、表現等もチェックすることができます。

2.7 ユーザ書式、ユーザ辞書登録

ユーザ書式は、翻訳メモリに登録することができます。登録されたユーザ書式は、システムが提供する書式と同様に、ファイルとしてオープンでき、また用例として参照できます。さらに、文置換機能では、別の文書の翻訳時に、同一文が翻訳メモリにあれば、その訳文を自動的に置換します。

ユーザ辞書は、ATLAS用の利用者辞書と、辞書引き用のユーザ辞書の両方に登録できます。

2.8 用語埋込と用語抽出

用語埋込は、準備しておいた対訳用語リストを使い、原稿文書に訳語を埋め込む機能です。

用語抽出は、文書内のマークを付与した用語を対訳リストとして出力する機能です。

■関連製品

・翻訳テキストエディタ TranText
（富士通ラーニングメディアより販売）
・英日／日英翻訳ソフト ATLAS
（富士通より販売）

■　動作環境

日本語Windows95/98，NT3.51/4.0が動作するパソコン
必要メモリ：　32MB以上
必要ディスク：　最大300MB（OS部分を除く）

■製品構成

個別製品：物品取引編、技術取引編、資本取引編、大型取引編、特殊取引編：各180,000円
オプション製品：TranText　DXC／専門用語辞書、TranText　DXC／英文チェッカ、ATLAS for Windows用専門用語辞書

■開発と販売

以下の3社による共同開発・販売製品です。

・（株）富士通ラーニングメディア
・富士通（株）
・国際事業開発（株）

■お問合せ先

東京都大田区新蒲田1－17－25（情報処理システムラボラトリ）
株式会社富士通ラーニングメディア
翻訳事業部）第一翻訳部
TEL：03－3733－9341
FAX：03－3730－3624
URL：http://mall.ec.infoweb.ne.jp/flm/

AAMT
ジャーナル

No. 24
(Oct. 1998)

発　　行　アジア太平洋機械翻訳協会
所 在 地　〒105-0011　東京都港区芝公園3-5-12 芝公園真田ビル
TEL：03-5473-7135　FAX：03-5473-0569
E-mail : aamt 0001@infotokyo.or.jp
ホームページ：http : //www.jeida.or.jp/aamt/
編集委員会　野村浩郷(委員長)　大倉清司　奥西稔幸　山端　潔
事 務 局　中島泰雄　木谷　恵
印 刷 所　伸光写植印刷株式会社