

# 自然な訳文を生成する翻訳システム 開発中の全く新しい翻訳ソフトの「現在」

マーリンアームズ株式会社 代表取締役 武舎広幸さん

聞き手:MT研究会 小室誠一

まとめ:eとらんす編集部



武舎広幸さんは翻訳ソフトの研究・開発に携わってすでに20年のキャリアの持ち主だが、そのかたわら翻訳者としても数多くの書籍を手がけてこられた。武舎さんが開発に携わった翻訳ソフトにはノヴァ社（現クロスランゲージ）のPC-Transerシリーズ（直接開発に携わったのは1999年発売のPC-Transer 2000まで）などがあるが、翻訳者の視点で作られた使い勝手の良さには定評がある。武舎さんはこの2年間、情報処理推進機構（IPA）が行なっている「未踏ソフトウェア創造事業」の支援を受け、「翻訳者並みの自然な訳文」を生成する翻訳システムの開発を行なってきた。エンジニアとは違う翻訳者の視点から開発された、従来の翻訳ソフトを超える全く新しいシステムとはどのようなものなのだろうか。

## 新しい翻訳システムの実力

### 従来の翻訳ソフトとの比較

まず、なにはともあれ新しいシステム（以下新システムと略）がどのような訳文を出すのか、従来の翻訳ソフトの訳と比べて見てみよう。

【原文】 The object gives you direct control over the volume.

Mixer

#### 《訳例 1.》

【従来の翻訳ソフト訳】（加工していない一次出力による）

#### ● PC-Transer V11

「ミキサー物は、あなたにボリュームの直接の制御をする。」

#### ● LogoVista X PRO 3.0

「ミキサーオブジェクトはあなたにボリュームについての直接のコントロールを与える。」

#### ● エキサイトの自動翻訳サービスによる翻訳

<http://www.excite.co.jp/world/text>

「ミキサー・オブジェクトは、あなたにボリュームに対する直接の管理を与えます。」新しい翻訳システムの実力

【新システム訳】

「Mixer オブジェクトにより、ボリュームを直接制御できる。」

いかがだろうか。従来の翻訳ソフト訳がいかにも機械翻訳調のまわりくどい表現になっているのとは比べると、新しいシステムはかなりこなれた自然な訳になっている。ここでの訳出のポイントは“direct”という単語をどう訳すか、だ。

この“direct”は“control”を修飾する形容詞。従来の翻訳ソフトでは原文で形容詞だった語句は訳文でも形容詞とされるため、“direct”は日本語で「直接の・直接的な」などとなる。一方新システムでは副詞として「直接」と訳している。人間が翻訳する時には、訳文を自然なものにするために頭の中で品詞の変換を無意識に行なっている。この場合は形容詞を副詞に変えて「直接制御できる」と訳すわけだが、新システムではそれと同じ作業を行なっているのだ。これは従来の翻訳ソフトのフレームワークでは全く不可能だった。

“give ~ control”で「制御する」という訳を出すことは、現在の翻訳ソフトでも可能だ。しかし、これらの翻訳ソフトでは品詞の転換ができないため、どうしても「直接的な制御できる」といった不自然な訳しか出せないわけだ。

また先述の訳例では、翻訳ソフトは“give you control”の“you”を訳出しているが新システムではこれを省略して日本語らしい表現にしている。いわば受験英語レベルの直訳から、自然な日本語への翻訳に変換している、と言えるだろう。このような翻訳のテクニックをバベルでは31の公式にして「翻訳英文法」として教えているが、こうしたレベルでの新システムの実力はどのようなのだろうか。そこで、さらにもう一例、トライアルを試してみることにした。

## 自然な訳文を生成する翻訳システム 開発中の全く新しい翻訳ソフトの「現在」

### 翻訳のテクニックを適用する新システム訳

「翻訳英文法」の公式6によると「of + 名詞」が目的語を表わす場合は以下の手順で翻訳する。

A of B という形が出てきたら――

- (1) まずこの of が主語を表わすか、目的語を表わすかを確認し、
- (2) 目的語であるとわかれば「～を…する」という関係を読みとり、
- (3) それに応じた文章の形に読みほどこいてから訳す。

#### 【原文】

His application of the rule to this case was quite natural.

翻訳英文法・公式6に基づいた訳例を挙げておこう（人間による翻訳）。

「彼がこの規則を今回のケースに当てはめたのは、きわめて自然なことだった。」

これを新システム訳と従来の翻訳ソフト訳と比較してみよう。参考までに例文の rule を theory に置き換えた文章も翻訳してみた。

#### 《訳例 2.》

[新システム訳]

- ① His application of the rule to this case was quite natural.

彼が今回のケースにこの規則を当てはめるのはきわめて自然なことであった。

- ② Her application of the theory to this case was quite natural.

彼女が今回のケースにこの理論を当てはめるのはきわめて自然なことであった。

[従来の翻訳ソフト訳]

#### ● PC-Transer V11

- ① このケースへの規則の彼の適用は、全く自然だった。
- ② このケースへの理論の彼女の応用は、全く自然だった。  
(★注目点 of に続く単語の意味素性によって application の訳語を切り替えている。)

#### ● LogoVista X PRO [英×日]Ver.3.0

- ① このケースへの規則の彼のアプリケーションは非常に自然だった。
- ② このケースへの理論の彼女のアプリケーションは非常に自然だった。

#### ● エキサイト翻訳

<http://www.excite.co.jp/world/text/>

- ① この場合への規則の彼の適用は全く自然でした。
- ② この場合への理論の彼女の適用は全く自然でした。

この例でも新システムの有効性は明らかだろう。唯一新システムの訳文の難点は「当てはめるのは」という部分で時制の一致が適用されていないことぐらい（「当てはめたのは」となっていれば完璧）。このルールが必要なのは今回トライアルにかけて分かった点だが、ここに時制の一致を適用するルールを書き込むのは難しくないそうだ。それにしてもどのようにしてここまで自然な翻訳ができるようになったのだろうか。ここで新システムの構造を見てみよう。

## 新システムの構造

### 辞書の役割を切り分け翻訳を効率化

武舎さんによれば、新システムは図1のような概念構造になっている。実は原文から訳文に至る根幹部分のプロセスは従来の翻訳ソフトとほとんど変わらないとのこと。ここで一般的な翻訳ソフトの作業プロセスを簡単に見ておこう。図1の根幹（中心）部分を参照していただきたい。

辞書引き：原文を最初から見ていって、各単語を辞書のデータと付け合せていくプロセスだ。各単語の品詞の照合を行ない、単数・複数形、時制などの情報をリストアップする。品詞などが複数あるものについてはそれらもリストに挙げられる。

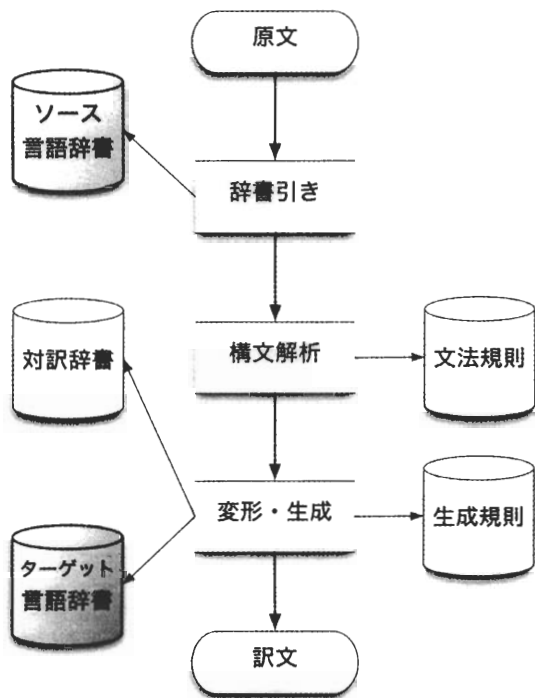


図1

**構文解析：**ここで各単語を文法規則にあてはめていく。冠詞と名詞の組み合わせが名詞句になり、それと他動詞が組み合わさって動詞句になり…という具合にすこずつ大きなまとまりを作って、最終的に文の構造を明らかにしていく。また、文法規則に当てはめて、品詞や構造の絞込みを行なう。

**変形・生成：**まず構文解析で作った構造を、日本語の生成に便利な構造に変形する。これは単純に言えば日本語の語順に並べ替える作業、と言える。変形では語順を並び替えただけなので、そのまま訳すと「てにをは」が無いカタコトの日本語のような状態になる。これを生成規則にしたがって書き換え、日本語らしい訳文にするのだ。

ここで従来のソフトと大きく違うのが、辞書を「ソース言語辞書」「対訳辞書」「ターゲット言語辞書」の3つに分けた点だ。従来のほとんどのシステムではこの「対訳辞書」と呼ばれている部分しか用意されていない。それぞれの辞書は翻訳の各段階で役割が違っている。

翻訳作業の流れから言うと、一番最初に参照するのはソース言語の辞書になる。「ソース言語辞書」の中には訳語との対応はなく、例えば direct という語句なら、その品詞と活用、副詞化すると directly になる、といった言葉の関係が記述される。ここではソースの語句全ての解を出す。

訳語との対応は「対訳辞書」で行なう。さらに「ターゲット言語辞書」の中にある品詞変換の情報を利用して、「対訳辞書」の中にある生成規則によって自然な文章になるように変換される(図2参照)。《訳例1》では direct を副詞化するとどうなるか、という情報をたどって「直接」という訳語にする、という判断をしたわけだ。

従来のソフトでは翻訳のそれぞれ違う段階で必要な情報を一つの辞書から引いてくるため、役割や情報が重複していた。それを効率良く切り分けた形だ。このため品詞の変換のような従来のソフトでは不可能だったこともできるようになったのだ。

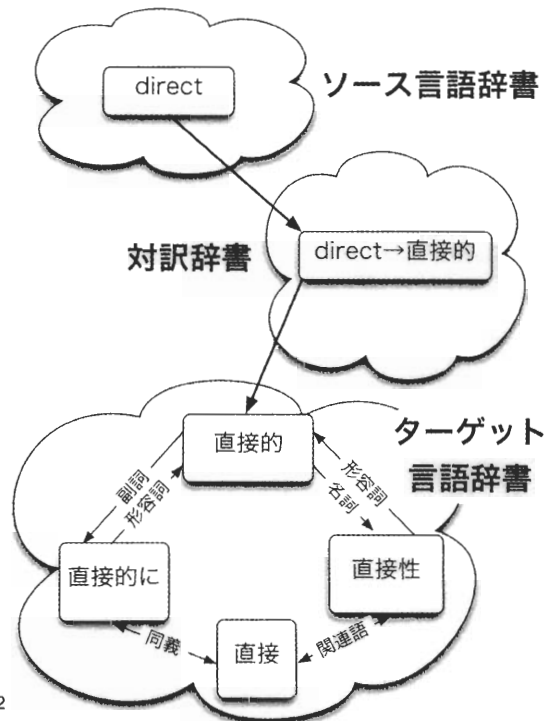


図2

## 多言語に対応可能な画期的システムの誕生

### 言語に依存しない翻訳プログラム

武舎さんによればソース辞書とターゲット辞書は入れ替え可能だという。また、新システムでは一連の翻訳プロセスから辞書と文法データを切り離したため、各部が完全にコンポーネント化された。このことによって、原文から訳文を生成する根幹部分は全く言語から独立したプログラムとなり、ソースとターゲットの入れ替えが可能だけでなく、

## 自然な訳文を生成する翻訳システム 開発中の全く新しい翻訳ソフトの「現在」

別の言語対の辞書に入れ替えるだけで英語から中国語にしたり韓国語にしたりということが自由にできる、という。従来のほとんどの翻訳エンジンはこの根幹のプロセスが対象言語に依存しているので、他言語に自由に対応することは難しい。まさに画期的なシステムが実現可能になったのだ。

ただ、現段階で翻訳の根幹部分はほぼ完成したが、辞書についてはほとんど手が着いていない状況なので、日本語⇄英語と別の言語に対応させるためには、まだ遭遇していない現象や、別の言語にしか表れない現象に対処するために根幹部分にも調整が必要になってくるかもしれない、とのことだった。

辞書については最終的に役割をどう分担させるかについても現在手探りの開発が続いている。それを物語る本誌とのやり取りを一つ収録しよう。

**武舎:**「まずソース言語の分析をする場合に、そこで別の言語を絡ませるのはそもそも違っていると思います。まず必要なのはソース言語の品詞と活用と言葉の関係ですね。ただ、現在はある単語と他の単語との共起関係（組み合わせ）みたいなものは対訳辞書の中に書き込むようになっています。その理由は、その言語の中でこの語とこの語が結びついたときの意味は、他の言語ではこうなる、というふうに記述した方がいいと考えたからです。」

**本誌:**「そういう語句の組み合わせに関しては、やはりソース言語のところで見てしまうのがいいのではないのでしょうか。今コーパス言語学が盛んになっていますが、言語はチョムスキーが唱えた生成文法ではなくて、実は共起関係で成り立っているんじゃないか、という説が有力になってきています。今の翻訳ソフトでは要するにパターン辞書みたいな形でそれに対応していますが、こういう形のシステムであれば、それを原文の解析時にある程度拾っておくと面白いと思います。」

**武舎:**「確かにその通りですね。逆にそうでないといけなかもかもしれません。ここで書いた方が美しいかもかもしれません。考えてみます（笑）。」

### 広い範囲の文を見渡して共起関係を認識するソフト

ここでもうひとつ新システムの訳例を紹介したい。英語の原文はアップル社の製品 iPod に関するニュースリ

リースをウェブサイトから引用したものだ。(http://maccentral.macworld.com/news/2002/10/21/ipod/ より)

**【原文】** This problem should be remedied with a new point release. MusicMatch spokesperson Jennifer Roberts said that MusicMatch Jukebox 7.5 – due out sometime in the next several weeks – would have iPod support that should be maintained by future releases as well.

### 《訳例 3.》

この問題は次のマイナーバージョンアップで修正されるはずである。MusicMatch 担当者の Jennifer Roberts は（今後数週間のうちに出る予定の）MusicMatch Jukebox 7.5 が iPod をサポートするだろう、そしてそれは今後のリリースによっても変わらないはずであると語った。

この訳文でまず注目されるのは、ハイフンで括られた挿入節があるにもかかわらず、システムが動詞 have の主語を正しく MusicMatch Jukebox 7.5 と認識して訳している点だろう。現在の翻訳ソフトではこのような挿入節がある場合、文の構造をうまく認識できなくなってしまうことが多い。これはこの新システムのもう一つの特徴である「広い範囲を見渡して文の共起関係を認識する」という機能によるものだ。

さらに驚かされるのは “would have iPod support that” 以下を「…iPod をサポートするだろう。そしてそれは今後のリリースによっても変わらないはずであると語った。」と訳している点だ。

この原文の構造では主語は Jennifer Roberts で、said の目的語が that MusicMatch Jukebox 7.5... 以下の名詞節になっており、さらにその節の主語 MusicMatch Jukebox 7.5 の動詞 have の目的語は support で、それを修飾する形で名詞の iPod が support にかかっており、support に対して後ろから関係詞が修飾しているという複雑な形になっている。従来の翻訳ソフトではこういう関係詞による複合的な構造の文を正しく訳出することはなかなかできなかった。

特に、この新システム訳では “would have iPod support that” の関係詞のところを切ってしまう、別の節にして「サポートするだろう。そして」という訳を出している。これは先の「翻訳英文法」では公式1の「原文の流れを生かし、

語順を変えずに訳す」という方法にあてはまる。文の構造の広い範囲を自由に見渡して必要な変形を行える例と言えるだろう。

従来のソフトでは文の構造のうち恐らく動詞と目的語や、名詞とその形容詞の関係という、小さな範囲しか見ることができなかつたが、このシステムでは関係する語句がかなり離れた場所にあっても参照することができ、前の文の主語が何かということまで見ることができる、ということだ。

ただ現在のところは、その範囲をセンテンス内に設定しであるとのこと。というのはパラグラフの範囲にすると、処理するデータ量が大きくなり過ぎて現実的でないのだそうだ。

## 新しいシステムの開発意図と今後の課題

以上見てきたようにこの新たな翻訳システムは実に大きな可能性をもっているが、今後解決しなければならぬ課題も数多く残っている。特に重要なのは現在ほとんど未着手の辞書の充実をどうしていくか、という点だ。武舎さんにこの新たなシステムの開発意図と今後の課題についてもお話をうかがった。

こういうソフトを開発しようとお考えになったのは、実際の翻訳作業に利用できるソフトを作ろう、ということだったのですか。

そうです。もちろん現在の翻訳ソフトも実際に利用していますが、現在の翻訳ソフトは極端に言うともまだ「辞書引きソフト」のレベルなのです。自動的に辞書に登録している語に置き換わっていて、訳文をつくるのにカット・ペーストする材料を作っているだけなのですが、翻訳者としてはもう一歩、ああこういう訳もありだね、というぐらいにはしたい。かつてワープロで仮名漢字変換した時に「ああ、こういう字なんだ」って思いましたよね。それと同じように、「ああ、こういう訳もいいね」って思わせるぐらいの訳も時々は出したいな、と思うんですね。

そのぐらいのレベルになれば、ある程度の人たちがまた使ってくださるのではないかと、という気持ちがあります。この新しいシステムでは従来に無い機能をいろいろ持たせていますが、そういう機能が無いと私が自分で翻訳しているレベルの訳文は出せないのです、これは絶対必要だ、ということを実現したものです。

新システムでは、要するに2回翻訳にかけているみたいなイメージでいいんでしょうか。《訳例1》でいえば、従来の翻訳ソフトだと「直接的な」という訳が出て終わりになるところを、もう一段階翻訳がかかる、という感じでしょうか。

そうですね、自然な日本語にするためにもう一段階かける、ということです。

荒い訳をつくっている時は頭の中で従来の翻訳ソフト訳みたいなものを出していることもあるんです。「制御を与える」はいかにも日本語では変だから「制御する」にしよう、などの作業は、翻訳者の頭の中でやっていると思うんですね。それを規則として書けるようにした、ということです。

その時に大きかったのは、ある翻訳学校のテキストを作った経験ですね。そこで自分が翻訳をやっている時のテクニックをまとめて客観的に見ることができた。それによって、ああ、これはルール化できるじゃないか、と思ったのです。

つまり、システムの中に、より人間がやっている翻訳作業に近いルールを書いているということですね。

ええ、自分がつくりたい訳文を出力できるような仕組みを新たに加えた、ということですね。まだ本当に出したい訳のどの辺までいっているかわかりませんが、従来の翻訳ソフトの訳が10段階のうち3段階くらいだとすれば、2つくらい仕組みが加わってもう2段階くらい進んで、少し近い訳が出るようになった、とは思っています。

プロセスとしては直訳は1度出しているわけですね。

ええ、というか正確には直訳が出るための構造は作っています。

つまり翻訳者が頭の中でやっていることと同じ作業をやっているわけですね。

そうですね、翻訳者は読むときには英語の構造で理解しているわけですよね。それを完全に日本語には直してはいないけれども、日本語に直す段階ではこれは日本語では不自然なので、こういう形に変えましょう、ということをやっているわけです。

英語のネイティブであればそのまま理解しているものを、日本語に変換する段階の作業を加えているわけですね。ここにさらにもう一段階入ることで「翻訳」になるわけですね。それに比べると従来のは「和訳」だった。

## 自然な訳文を生成する翻訳システム 開発中の全く新しい翻訳ソフトの「現在」

その「翻訳」のレベルを実現するためにはどうしても自然な日本語にするプロセスが必要だということですよ。今までの翻訳ソフトは、あまりそういうことを気にしてつくられてはなかったと思います。エンジニアの方はそういう感覚はあまりお持ちでなかった。エンジニアでなおかつ翻訳をやっている方はそんなにはないと思いますから。

英語のオリジナルな構文をそのまま日本語にすると、非常に不自然になるところを、日本語らしい構造に変えて出してあげましょう、ということです。

確かに現在、翻訳者は辞書引きソフトに近い形で、出力された文章は翻訳者が直すからいいんだ、という前提で使っています。その一方で一般的な使い方としてブラウザ対応のソフトが随分出てきて、それは訳文に一切手を加えられずただ翻訳されたものを利用しなければならない。このシステムが実用化されればそういう分野での応用も考えられますね。

そうですね、現在は特定の原文に対してのデータを辞書に入れているだけですが、こういう例をいっぱい貯めていけば、どんどんヒットする例が多くなってくると思います。本当は出来るだけ有能な翻訳者が100人くらい集まって作業すればある程度いい訳が出せるようになると思うんですが。

このシステムのように登録するのはソース辞書なりターゲット辞書だけでいい、ということになれば、皆で力を合わせてやることのできるかもしれませんね。

ええ、でも対訳辞書もやらないと、正しい訳が出せないので。こういう場合はこう訳す、という記述は対訳辞書で行なうんですが、この記述がなかなか面倒なのです。

その辺を簡単にしてみんなで共同でできるようなシステムをつくれればいいでしょうね。

それは是非やりたいですね。

そういうことは個人だとか普通の企業単位では無理だと思います。皆が自分の専門分野のデータを共同で登録していけるようなシステムがあれば、本当に面白いものが出来ると思います。

まあ、かなり資金力がある企業が、すでにかかなりのベースとなるデータを持っていて、それを改良していく、という手順なら可能かな、とも思うのですが。ベストなのは世界中の人が皆共同で一緒にやっつけてしまおうよ、という形ではないかと思いますが。なかなかそこまでののが大変なんですけれども。私がこれだけに掛りきりでやれば15年とか20年やればできるかもしれませんが。私も生活していかなければなりませんから（笑）。

今現在はシステムの中心部分はほぼ出来上がっているんですか。

一応は出来ています。今後はソース辞書、ターゲット辞書、対訳辞書にどれくらい実例を貯めていけるか、ということなのですね。しかし現状はほとんどゼロに等しいくらいで、従来の翻訳ソフトのシステムが持っている対訳辞書の基本的なものすらまだ、無い状況です。一つの方法は、どちらかの会社から買うとか協力していただくなどで辞書を揃えて、その上にさらに知識を加えるような方法でしょうか。従来の対訳辞書と互換性を保って、その上にこういう記述が出来るようになる、というものを加えることは可能ですから、そういう形が現実的かな、とは思うんですが。

やっぱり人間の頭の中に入っているのは膨大な知識ですね。それをひとつひとつ入れていくのは一人や二人でできるものではなくて、何年、何十年とかかる作業ですから。

翻訳の仕事をしている人が簡単に登録できる方法があるといいですね。

いい表現が浮かんだ時に取りあえず簡単に登録しておける形があれば、そこに登録しておくことによってまた、自分でも恩恵を受けることが出来ると思います。そういうものを是非形にできればいいなと思っています。

むしゃ ひろゆき

1980年国際基督教大学教養学部卒業。83年山梨大学大学院工学研究科修士課程修了。同年から1年間リソースシェアリング株式会社勤務。機会翻訳システムSTARの開発に従事。89年東京工業大学大学院博士課程満期退学。在学中、米国オハイオ州立大学大学院、カーネギーメロン大学機械翻訳センター留学。89年有限会社ワードハウス設立。93年マーリンアームズ株式会社設立。現在同社代表取締役。2002年6月から、独立行政法人 情報処理推進機構 (IPA) の未踏ソフトウェア創造プロジェクトに採択され、次世代の機械翻訳システムの開発を行っている。著書に『プログラミングは難しい!』(チューリング)、『理工系大学生のための英語ハンドブック』(共著、三省堂)、『パソコン英日翻訳ソフト活用法。(ピアソンエデュケーション)』など。訳書に『マッキントッシュ物語 僕らを変えたコンピュータ』、『暴走する帝国インターネットをめぐるマイクロソフトの終わりなき闘い』(以上翔泳社)、『HTML入門』『Java言語入門』(以上ピアソンエデュケーション)など多数 <http://www.musha.com/> にホームページがある。