

### 永田 昌明

上席特別研究員 NTTコミュニケーション科学基礎研究所



## 専門用語に頼らず、分かりやすいパフォーマンスを心掛けよ

いろいろな国・地域の人たちとスムーズなコミュニケーションをリアルタイムで図ることができる翻訳機械があったら…現実になる日はそう遠くはないようです。最先端を走る日本の言語解析の研究を支える永田昌明上席特別研究員に、機械翻訳技術の潮流と課題について、また、研究者たちはどうあるべきかも伺いました。



### 世界をリードする統計的機械翻訳技術

#### ●今手掛けている研究について教えてください。

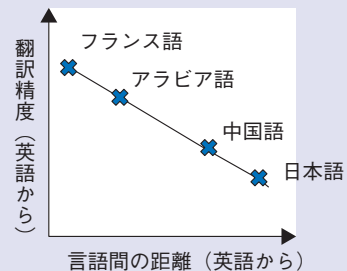
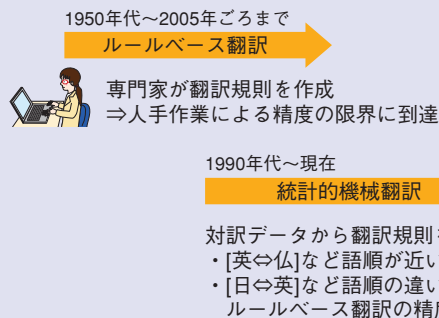
現在は統計的機械翻訳の研究を担当しています。これは、すでに翻訳されている膨大な対訳データから、翻訳規則・対訳辞書に相当する統計モデルを学習して機械翻訳を実現する技術です。

従来の機械翻訳では、翻訳元の文を一度分解して構造を調べ、翻訳先の言語に単語を置き換えてから再び文を組み立てます。具体的にいうと、名詞や形容詞など単語の品詞を特定し、翻訳元の言語の文法を使って主語と述語や主節

と従属節の関係など構文を調べます。構文が把握できたら、日英翻訳なら和英辞典を使い、日本語の単語を英語に置き換えます。そして主語、述語、目的語などの語順を翻訳先の言語の文法に合うように変更します。

この方法は文法などのルールに沿って行われるため、ルールベース翻訳といわれる方式で、専門家が自分の知識とノウハウに基づき翻訳規則を作成する方法です。しかし、この人手による作業は複雑化し過ぎて精度の限界に達し、現在は既存の膨大な対訳データから自動的に学習する「統計的機械翻訳」が主流です(図1)。

人手で作成したルールから自動学習した統計モデルへという変化は、音声認識やコンピュータビジョンなど人間の機能をコンピュータで真似る技術に近年共通してみられる



典型的な誤り例 (翻訳元の語順のまま直訳)

Language is a means of communication ⇒ 言語は道具の通信です

図1 機械翻訳技術の潮流と課題

パターンです。人が経験に頼って作成する規則ではすべてを網羅するのは難しいので、このような方式へと移っていきます。

また、翻訳の精度に影響するのが言語間の距離です。例えば、日本語から韓国語は文法も語順もよく似ているので、対訳辞書を使って単語を置き換えるだけで意味が通じます。ところが、日本人の多くが翻訳を必要としているであろう英語は、日本語からもっとも遠いところに位置します。フランス語は英語に近いので、比較的早い段階から精度の高い翻訳はできましたが、英語と日本語ではそう簡単にはいきませんでした。

しかし、最近では主辞後置性という日本語の特徴を使って、あらかじめ英語を日本語の語順に並べ替えてから翻訳するという手法をNTTが提案し、翻訳の精度を上げることができました。

#### ●具体的な例で説明をお願いしますか。

それでは英文和訳の例で説明しましょう。

主辞とは文の部品である句において句の文法的な役割を決める単語のことです。例えば前置詞句では前置詞が主辞です。別の言い方をすると句の中で修飾先になる単語が主辞です。日本語には、前の言葉が必ず後ろの言葉を修飾する、すなわち、修飾先の単語が必ず文の後ろにある主辞後置性という性質があります。これは世界の言語の中ではまれな性質です。

一方、英語は動詞があったときに、主語は前から修飾、目的語は後ろから修飾します。そして形容詞は名詞に対して前から、前置詞は名詞に対して後ろから修飾するという日本語とは異なる性質を持っています。

この日本語の性質ののっとして、翻訳元の英語の単語を修飾先が必ず後ろになるように順番を置き換えます。こう

すると日本語と同じ語順になるので、あとは逐語訳するだけで自然な日本語ができあがるのです(図2)。

まとめると、私たちが提案した事前並べ替え翻訳方式は、漢文の訓読のような規則に従い、英文を和文に訓読します。中間段階として英語を日本語の文法で話すルー大柴さんのような言葉に並べ替え、最終的に正しい日本語に直しているということです(図3)。

問題は、英語から日本語への翻訳はほぼ実用レベルに達していると思いますが、日本語から英語については、まだ取り組まなければならない課題がたくさんあることです。

現在は、実際に使える製品レベルまで英語から日本語への自動翻訳の精度を高めながら、日本語から英語への新しい翻訳方式を研究しているところです。今後の統計翻訳・言語解析の研究の方向性ですが、1つは日本語にありがちな主語が省略されている文から、主語が何かを特定する技術を確認することですね。図4で示すところの省略照応解析という部分です。話し言葉のような「くだけた文」を翻訳する機能は難しいのですが、旅行で使う定型的な会話などは比較的早い段階から提供できると思います。その先にはリアルタイム、つまり同時通訳型の翻訳装置を通して違う言語の人たちが会話をスムーズにできるシステムも構想にあります。一方、特許文のような技術文書は、近い将来、実用化に手が届くでしょう。



### 自称、語学オタクがコンピュータ技術を学ぶと翻訳器ができる?!

#### ●言語解析に携わるきっかけは、

実は私は語学オタクでした。今でもNHKで新しい語学講座が始まるとつい手を出してしまいます。英語はしゃべ

2010年にNTTが提案

- ・日本語は主辞後置 (head final) 言語である  
⇔日本語の係り受けは必ず前から後ろになる
- ・前から後ろに係るように翻訳元言語の単語を並べ替えてから逐語訳する
- ・英語 (他言語) から日本語への翻訳では、とても有効  
—日本語から英語 (他言語) への翻訳には使えない
- ・構文解析の精度が翻訳精度を決める

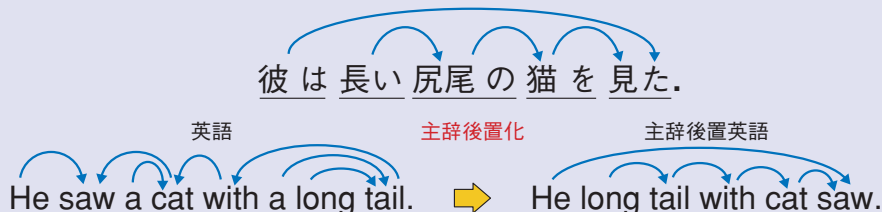


図2 日本語の主辞後置性に基づく事前並べ替え



最先端の統計翻訳がたどりついたところは千年の歴史を誇る漢文訓読

### ・漢文訓読

— 「有朋自遠方来」

→ 朋-有-遠-方-自-来 日本語語順の漢字列

→ とも-あり-遠方-よ<sup>リ</sup>-きた<sup>ル</sup> 訓読み

1. 白文を構文解析する
2. レ点・返り点を打つ
3. 送り仮名を付ける

### ・英文訓読（事前並べ替え翻訳）

— “I love you”

→ I you love 日本語語順の英単語列（主辞後置英語）

→ は youを loveする ル-大柴風日本語

→ 私は あなたを 愛する 訳語選択

1. 英文を構文解析する
2. 単語を並べ替える
3. 逐語訳する

有<sup>レ</sup>朋<sup>リ</sup>自<sup>ニ</sup>遠<sup>ク</sup>方<sup>ヨ</sup>来<sup>ル</sup>。不<sup>ニ</sup>亦<sup>レ</sup>楽<sup>シ</sup>乎<sup>カ</sup>。

図3 漢文訓読と事前並べ替え翻訳

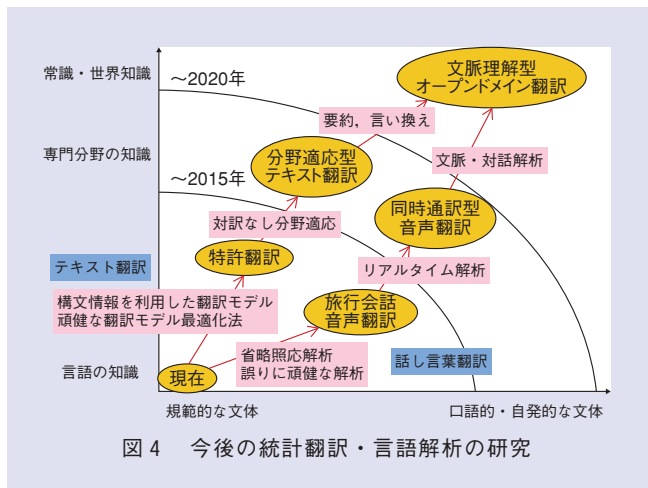


図4 今後の統計翻訳・言語解析の研究

るのは得意ではないですが、聞くのは大丈夫です。中国語は簡単なコミュニケーションができますし、韓国語を集中して勉強していた時期は韓流ドラマを字幕なしで見ました。

若いころから海外旅行が好きで20カ国以上回りました。フランス語が役に立ったのはモロッコの旧市街で迷って道を尋ねたとき、ドイツ語が役に立ったのはプラハへ深夜に列車で着いて駅から電話でホテルを予約したときですね。見知らぬ国でありえない体験をして常識をくつがえされるのが楽しいんです。

一方、大学では情報工学を専攻していました。私が学んでいた約30年前はハードウェア（電子回路）とソフトウェアの両方をつくれる人材を育成していた時代ですから、半田ゴテなど見たこともないだろう今の情報系の学生さんとはだいぶ異なります。

大学在学中には音声に関する研究を主に行っていました。パターン認識、言葉ではなく、人間の目と耳に相当する機能をコンピュータで扱う研究室でした。そして、NTT入社後、言語処理を研究している部署に配属されたことが、この研究を始めるきっかけといえますね。

## 研究成果を実感してもらいたいという 願いから生まれたデモンストレーションスタイル

言語処理の研究がなぜ必要なのか。今日では、検索サイト、翻訳サイトなど多く活用されていることもあり、その必要性は理解されています。しかし、私が研究を始めた当時は、NTTが民営化されてまだ2、3年目のころです。電話、すなわち音声通信や、FAXなどの画像通信が事業の中心である時代。社内には言葉、言語処理を扱うセクションは電報と番号案内しかありませんでした。言語処理は将来を見据えた研究であり、その当時は主流ではなかったのです。

なぜ研究していたかということ、必要というよりも自分が好きだったというほうが大きいかもしれません。1980年代後半、入社してすぐにATR（国際電気通信基礎技術研究所）に出向し、革新的な研究がしたいと思い、今でいう翻訳電話をつくるための研究を4年間していました。

ここでは、世界初の翻訳電話と称して音声認識と機械翻訳と音声合成を組み合わせた装置を作成し、ドイツや米国の大学と共同実験をしました。そのとき、非常に苦勞してつくった翻訳システムのデモンストレーションが、世間の注目を集めました。

私はこのときに2つのことを学びました。

1つは、実際に手掛けていた研究の内容はルールベースといわれる、専門家が人力で規則をつくり上げるもので、私はATRに在籍していた4年間、ひたすらこれに取り組んでいました。しかし、なかなか良いシステムを構築することができなかつたため、言語データ等から自動的に規則を学習する研究の必要性に気付きました。

そして、もう1つは、研究成果をデモンストレーションのかたちで行うことは、他者の反応をつぶさに感じることができ、非常に面白いと感じました。研究者は、論文を書いて終わりではなく、「取りあえず動かしてみせる」というスタンスです。大学の恩師が常々話していた「例が必要」ではなく、「例がすべてだ」というのも、このときに再認識して、私のスタイルが確立しました。



## 専門分野以外の人に分かりやすく伝えることが大切

### ●永田さんの1日というのはどのようなものなのでしょうか。研究スタイルを教えてください。

グループリーダーも務めていますので、連日、メールや会議などに時間を割かれるために、なかなか十分な時間を研究に費やすことはできません。

今はグループの研究員の報告を聞いて議論をしながら、研究の方向性を決めていくことが多いですね。私自身は論文や研究報告を読みながら、次に何をすべきかを考えています。ひらめきを得るために何か特別にすることはありませんが、論文を書くときは、良い例を交えて分かりやすく説明することを常に考えていますね。説明したい原理をもっとも端的に表すことはとても重要だと思います。あまり意識したことはありませんが、分かりやすいか否かの判断基準は、研究所内の専門分野が異なる人にも理解できるかどうかだと思います。

NTTの研究所はさまざまなテーマを手掛けていますから、専門分野も多岐にわたりその数だけ研究者がいます。専門分野が違えば、専門用語は通じません。研究の意義を理解してもらえないということは、研究を継続する意義も認められないということですから、実用化にこぎつけることはできません。

自分が研究したことを世の中に伝えたいという欲求は、研究者として常にありますね。今の時代であればブログなどを使えば良いのしょうけれど、私の時代は教科書を書くことでその欲求を満たしていました。

ちなみに、最近はチームメンバーのプレゼンテーションの作成の中でも、どうやったら他人に分かりやすく伝わるだろうかという観点で参加しています。

また、ここには知的好奇心の旺盛な研究者が集まってい

ると思います。立場、年齢的にも上になっているので、若い世代の研究者が研究をしやすい環境をつくるか、彼らの視野を広げる手伝いをしたり、他者に研究内容を理解してもらえそうな表現方法を一緒に考えたりしていますね。



## 技術を先に進めたいという情熱を持って

### ●永田さんは、研究者としての転機がもう1つあったそうですね。

はい。40歳前後のときに股関節の病気をしました。医者から65歳まで安静にしているしかないと言われて、研究者としての生命は終わったと実感しました。

しかし、他に何か手立てはないのかと医学論文をあさり、最先端の手術によって治療できるということを発見し、その手術ができる医師を自分で探して手術を受けました。そのときに思ったのが、医学も研究の世界であるということ。ある問題に対してさまざまなアプローチが存在し、それに対していろいろな学派も意見もあるということでした。

また、最先端の医療技術に支えられて現場に復帰できたことから、私も自分の研究分野で、社会に貢献したいと思いました。病気をする前とは違う情熱が宿り、自動機械翻訳を現実のものとするために邁進しようと決意しました。

### ●後進の研究者の皆さんにアドバイスをお願いします。

自分でやりたいこと、正しいと思うことを周りの声に影響されることなく突き進むことです。

今は、社内の評価で短期的に成果を求められるなど、昔のような牧歌的な雰囲気は確かになくなりました。世の中全体の評価は厳しくなってきたという実感は私にもあります。情報のスピードも速いため、ほかの研究者が何をしているかがあつという間に分かってしまいます。インターネットが普及していなかった時代とは環境が全く違います。唯我独尊はあり得ない中で、信念を貫くのは難しいでしょうが、自分は何でこの研究をするのか、方向性をどうするかなど、軸をしっかり持ってください。私もできるだけサポートしますので、一緒に頑張りましょう。