

言語学的比例類推空間相似

イヴ・ルパージュ & 白井 諭

{yves.lepage, satoshi.shirai}@slt.atr.co.jp

エイ・ティ・アール - 音声言語通信研究所

要約

「類推は、モデルとその規則的模倣を予想する (ソシュール、小林英夫訳)」。

即ち、類推は事例を一定の規則に基づいて模倣することにより成立する。例えば、「食べる: 食べます = 決める: x \Rightarrow x = 決めます」。本論文では、機械翻訳にその原則の適用可能性を示す。類推関係は、記号レベルで適用することとし、この関係に基づく言語処理手法は普遍的な手法であるため、いろいろな言語に適用可能であると考えられる。

1 言語学的背景

音韻学—形態学 本研究の着目は、ソシュールによる「一般言語学講義 (小林英夫訳 岩波書店)」、第 I I I 部第 4 章の中の次の指摘である。

[古典語形 *honor* と競争する新語形 *honor*,] この二重性は、*ōrātor* : *ōrātōrem*, etc. をモデルにして作られた新語形 *honor* によって除かれた; その手順, 以下に研究するが, これからさき, 比例四個式の計算に引きもどしてみようと思う:

$\text{ōrātōrem} : \text{ōrātor} = \text{honōrem} : x \Rightarrow x = \text{honor}$

ソシュールは、アリストテレスの類推の定義を単語に応用した。即ち、類推とは3つの形を与え、欠けている4つ目の形を見い出すことであり、類推方程式として規定される。

歴史的には、類推とは古典文法学者によるラテン語ないしはギリシャ語の形態的記述、および、あらゆるアラビア語の伝統文法を意

味する。これらは、先ずモデルを提供し、次に類推による曲用、活用、派生等を再構築するのに必要で十分な情報という形で、語彙的要素を提供する。類推は言語の比較研究、より正確にはインドヨーロッパ語研究にも適用されている。(もしギリシア語 /h/ がラテン語の /s/ に対応すると言えるなら、 $\acute{\upsilon} \pi \epsilon \rho$ が *super*(最高) で、 $\acute{\epsilon} \mu \iota$ が *semi*(半分) である、等)。類推関係は Hjelmslev と Jakobson の音韻研究の骨格をなし (/p/ は /t/ となり、/b/ は /d/ に当たる、等) 関与の音韻対立の概念を伴っている。

統辞論 共時的観点から、類推だけであらゆる言語現象を説明できないし、例外はそこかしこにみられる。しかし、ある意味ではヘルマン・パウエルとブルームフィールドは統辞論が類推の適用対象になりうると考えている。

柿が彼に食べられた。: 柿を彼が食べた。=
打ち合わせ場所が彼に決められた。: x
 $\Rightarrow x$ = 打ち合わせ場所を彼が決めた。

統辞論における類推の効果に関する議論は、チョムスキーによって否定された。類推によって非文法的文章が簡単に生成されることから、彼は類推が判断材料とならないであろうと結論づけた。

柿が彼に食べられた。: 柿を彼が食べた。=
先生が学院に來られた。: x
 $\Rightarrow x$ = *先生を学院が來た。

しかし、チョムスキーの言語生得説及び文脈自由文法の仮説は、心理学の実験でも、言語学的データ (形態論 [バンバラの例]・統辞論 [スイスのドイツ語圏の例]) でも否定された。こうして類推は表舞台に帰り咲くことができる。

tu es une fille.	あなたは少女だ。
tu es un garçon.	あなたは少年だ。
il est un garçon.	彼は少年だ。
elle est une fille.	彼女は少女だ。
elle est professeur.	彼女は先生だ。
il est professeur.	彼は先生だ。
elle est infirmière.	彼女は看護婦だ。
tu n'es pas infirmière.	あなたは看護婦ではない。
il n'est pas mon professeur.	彼は私の先生ではない。
professeur	先生
professeur	教授
fille	少女
garçon	少年

図 1: 2 言語の文章－翻訳

2 目的：類推空間相似に基づく翻訳

次のような 2 言語の文章－翻訳の列を使い、類推のみに基づく機械翻訳方式の実験システムを作成した。

システムにある文章、例えば tu es infirmière を入力し、elle est une fille, tu es une fille, elle est infirmière という 3 つのフランス語文章を類推関係で見つけようとするものである。実際には、以下ようになる。

elle est une fille : tu es une fille =
 elle est infirmière : x
 $\Rightarrow x = \text{tu es infirmière}$

フランス語の類推関係を日本語の類推関係に移植すると、次のようになる。

彼女は少女だ : あなたは少女だ =
 彼女は看護婦だ : x
 $\Rightarrow x = \text{あなたは看護婦だ}$

そうすると、結果的に tu es infirmière のフランス語の文章に対応する「あなたは看護婦だ」の日本語訳の文章が得られる。

3 方法：基本演算としての類推解決

類推空間相似 上記の方法では、フランス語と日本語の二つの空間があり、第 2 空間の類推的構成は、第 1 空間の要素との照応によって移植または保存されると仮定している。数学では構成を保存する機能が相似と呼称される。

先に挙げた例では、入力した文章 tu es infirmière の訳として「あなたは看護婦だ」を即座に得ることができる。実際には常に 3 つの文章を見つけることができるとは限らないので、2 つの文章を探し、3 つ目の位置に出現すべき文章を類推解決アルゴリズムを用いて獲得し、その 3 つ目の文章をフランス語と日本語の翻訳情報の中に存在するかどうかを確かめる。こうして獲得したその 3 つ目の文章を用いることにより、手法は帰納的に適用される。

記号列の基礎演算 上記の方法をコンピュータで実行するため、次のような記号列の間の類推方程式を解決する機械的な手順（アルゴリズム）を提案した。

食べます : 食べる = 決めます : x
 $\Rightarrow x = \text{決める}$

提案されたアルゴリズムは普遍的であり、入力した 3 語の記号列以外の知識に影響されてはいけない。即ち言語知識や世界知識を一切用いない。

$\square\bigcirc : \triangle\bigcirc = \square\bullet : x \Rightarrow x = \triangle\bullet$

実際には、上記のような接頭辞と接尾辞の入れ替え操作だけでは不十分である場合もある。具体的には、セム語については数限りない接中辞の平行挿入を扱わなければならないので、問題はもっと複雑である。¹

¹アラビア語 : aslama ‘彼は改宗した’ は ‘イ

tu es une infirmière.	あなたは看護婦だ。	(× 1)
tu n'es pas mon infirmière.	あなたは私の看護婦ではない。	(× 3)
	あなたは私の看護婦ではない。	(× 2)
elle n'est pas mon infirmière.	彼女は私の看護婦ではない。	(× 22)
	彼女は私の看護婦ではない。	(× 18)
	彼女は私の看護婦ではない。	(× 1)
tu n'es pas mon professeur.	あなたは私の先生ではない。	(× 21)
	あなたは私の教授ではない。	(× 1)

図 2: 文字レベル類推空間相似によって翻訳結果

aslama : muslim = arsala : x
 $\Rightarrow x = \text{mursil}$

4 結果

ここに前述の手法と 2 章の例文 (図 1) によって得られた幾つかの翻訳がある (図 2)。括弧内の数字は一つの翻訳が出現する頻度を示している。入力と同じ場合、最も頻度が高い翻訳が最尤であると推測できる。

5 考察

この方式は十分に改良された機械翻訳システムに比べると品質では劣るかもしれないが、方法論としては大きな利点がある。

第 1 に、文書を異なる 2 言語について並べるだけで事足りる。これはある程度自動的に行うことができる。

第 2 に、このタイプのシステムでは文法規則を明示する必要はない。文法記述は一般の機械翻訳システムの構築には必須である。文法は、前もって一つの言語のデータを分析し、明示的に分類することである。ここでは類推が暗黙のうちにこの分析を行う役を担っている。翻訳結果の精度の高さは、収集したデータの妥当性に依存する。

第 3 に、普通の語彙・文法の分割は不要である。この点は認識論の観点から重要である。実際にいかなる言語学者でも、単語や文章等に明確な定義を与えるのは困難である。

スラム教徒に' とほのめかされている。語源的には muslim は '改宗者'、つまり 'イスラム教徒'。arsala: '彼は送った'、mursil: '送付者'。

6 結論

心理学で人間活動の特徴の一つと考えられている 4 項類推式に基づき、言語に依存しない新しい言語処理方式について述べた。類推関係を翻訳に適用し、2ヶ国語の対応を用いて、翻訳が可能であることを示した。

また、この類推関係は、画像や音波の空間にも適用可能と考えられるので、今後は音響処理や画像処理に対する本技術の適用性を検討する予定である。

参考文献

Esa ITKONEN

Iconicity, analogy, and universal grammar
Journal of Pragmatics, 1994, vol. 22, pp. 37-53.

Yves LEPAGE & 飯田仁

言語に依存しない早期終了型類推解決手法
 言語処理学会第 4 回年次大会, 九州大学, 1998 年 3 月, pp. 266-269.

Yves LEPAGE

Open Set Experiments with Direct Analysis by Analogy
Proceedings of NLPRS-99, Beijing, November 1999, pp 363-368.

Yves LEPAGE

Languages of Analogical Strings
Proceedings of COLING 2000, vol 1, Saarbrücken, July-August 2000, pp. 488-494.

フェルディナン・ド・ソシュール

一般言語学講義

小林英夫訳 岩波書店