

## マルチモーダル情報からの言語獲得

國吉 芳夫

yoshio@nak.math.keio.ac.jp

慶應義塾大学理工学部

### 1 はじめに

自然言語理解システムは小さく閉じた世界を対象として研究されてきた。そのような小さな世界から開いた実世界の問題への適用に際して生じる問題についてはあまり考えられてこなかった。

しかし、システムのサイズが大きくなるにつれて、必要とされる知識のサイズは大きくなり、知識間の関係が複雑になって一貫性を保った知識の開発が手に負えなくなってくる。また、実世界の問題は固定しておらず使用する間に変化することがよくあり、あらかじめ想定した範囲を超える状況に対処する必要に迫られることもある。したがって、サイズが大きく、また閉じてもない世界に応用するためのシステムを手作りの知識だけに頼って開発することは難しい。

このような問題に対処するためには、システムに学習能力を持たせることが重要になってくる。また、想定していなかった状況にも対処できるような学習を実現するためには、部分的に組み込まれた学習能力では限界があり、システムの各部との連携が欠かせない。

言語に関する知識を獲得する試みの一つとして考えられるものにコーパスベースのアプローチがある。コーパスは実際に使用されたテキストから収集されることから、ここから抽出された言語的情報は実用上有用な知識であることが期待されるからである。

しかしながら、コーパスからは言語情報しか得られない<sup>1</sup>ため、コーパスベースの知識獲得には限界がある。例えば、promise と persuade の2つの動詞を対比させる次の2つの文を考えてみる。

I promised her to go.

I persuaded him to go.

<sup>1</sup>タグ付けやシソーラスとの併用などにより補える点もあるが、それらに依存すればするほど、同じ問題の繰り返しになる。

この2つの文の「to go」という句の主語は各々の動詞が持つ構造の違いによって異なるが、どちらの文も同じ統語構造を持つので言語情報のみからこれらを区別することを学習させるのは困難である。

複数の領域からの情報を入力として領域間の関係を元に言語的知識の学習を行うシステムは少ないが[1]、上で示したような知識の獲得は、マルチモーダル情報からならば容易に学習できる可能性がある。我々は、マルチモーダル情報源からの学習に重点を置いた対話知識の学習モデルを提案する。

### 2 モデル

我々の提案する対話理解モデルは、視覚、聴覚、発声、文法、辞書、解釈、推論の7つのモジュールから構成され、次の3種類の領域の情報を扱う。

1. 言語領域
2. 視覚領域
3. 推論領域

ここで、推論領域とは、単に推論モジュールで使用される内部表現のことのみを指すのではなく、推論能力自体も含めたものである。推論部での処理過程を経た結果を学習対象の領域とすることによって、推論能力との相互作用による効果的な言語知識の学習を目指す。

人間は、言語をさまざまな環境から得られる情報と関連させて学習するが、このモデルでは、言語入力を聴覚から得るものとし、外界から得る情報としては簡単のために他に視覚情報のみを扱うようにしている。

本稿では、対話知識の獲得に焦点を当てるため、辞書と解釈モジュールにおける学習能力のみを考慮する。文法知識の学習能力については無視するが、

それはこのモデルの本質的な限界を意味するものではない。

次の2つの節において、それらの2つのモジュールとその学習能力についてさらに議論を行う。

### 3 辞書モジュール

我々のモデルでは、辞書モジュールは前節で述べた3つの領域の表現のペアを学習する。すなわち、言語-視覚と言語-推論の2種類のペアを獲得する。

辞書は双方向の検索が可能であり、言語情報から視覚や推論情報を検索したり、その逆の検索が可能である。

辞書項目の獲得は、ペアのどちらか一方、あるいは両方が未知のものであるときに行うもので、次に示す3つの状況において起きる。

1. 未知語 — 未知対象
2. 未知語 — 既知対象
3. 既知語 — 未知対象

#### 3.1 語彙の獲得における言語的制約

語の意味の獲得には、曖昧性の問題が避けられない。例えば、既知の対象とともに提示された未知語の意味としては、その対象の名前と、対象の属性の名前との、少なくとも2つの可能性が考えられる。語の意味の獲得に際して何らかの制約を仮定することによって効果的に学習を行わせることが可能になる。

人間の子供の言語獲得に関する研究から、語の意味の獲得に関するいくつかの制約が提案されている。われわれのモデルでは、そのような制約のうち、対比の原則、相互排除の仮定、分類の仮定の3つの制約に注目している。以下で、これらの制約について述べる。

##### ● 対比の原則

Clark [2] が提案したもので、複数の言語の同時獲得を行う子供は、ある特定の意味を持つ両方の言語の語を獲得するのが困難であるという観察から導かれた制約である。例えば、英語

とフランス語を同時に学習する子供は、ある時期になるまで‘chien’か‘dog’のどちらかしか覚えられない。このように、未知の語の意味を獲得する際に同義語になるようなものを制限するための制約である。

##### ● 相互排除の仮定

Markman[3] によって提唱されたもので、この制約によると、異なる語の外延は重複しない。これは、対比の原則を強めた形になっている。

##### ● 分類の仮定

子供における語意の獲得において、未知の対象とともに提示された語はその対象の名前と解釈される傾向があり、既知の対象とともに提示された場合にはその対象の属性と解釈される傾向があることが観察されている。分類の仮定は、この現象を説明する制約であり、Markman[4] によって提案された。

### 4 解釈モジュール

解釈モジュールは、文法モジュールから得られた統語構造を付与された言語の表現を、推論モジュールで使用される表現に変換する処理を担当する。例えば、

Put the block in the box.

という文に対して、文法モジュールの働きにより、  
directive(vp(put, np(the, block)), np(the, box))

のように表層構造が得られているものとする。解釈モジュールはこの構造を変換し、推論モジュールの入力として使用できるような次の構造を作り出す。

put(block5, box2)

解釈モジュールにおける学習は、このように、ある構造から別の構造に変換するための写像を獲得することを目的として行われる。

#### 4.1 構造写像理論

上で述べたような変換は、一種のアナログカルな写像として捉えることができる。Gentner[5] は、

アナロジーのための構造写像の理論を提案している。彼は、*literal similarity, analogy, abstraction* といったいくつかの異なるタイプの領域比較を考察し、アナロジーにおける比較では、関係を表わす述語を写像する傾向があるのに対して、属性を表わす述語は写像しない傾向があることを明らかにしている。さらに、彼は、写像される述語は高階の述語に基づいた概念である *systemicity* によって特徴付けられると主張している。

Falkenhainer[6] は Gentner の結果を一般化し、SME(Structure Mapping Engine) と呼ばれる構造写像を行うエンジンとなるプログラムの実装を行った。SME では写像規則はエンジンから分離されており、写像規則の定義を交換することによって Gentner の理論を含む、非常に柔軟な構造変換を可能にしている。

しかしながら、構造写像理論は知識表現の統語的な性質のみに依存している。以下で議論するように、対話を理解するためにはターゲットとなる領域上における推論過程の利用が重要になってくるので、そのままでは使えない。

## 4.2 関連性理論

関連性理論 [7, 8] は Sperber と Wilson によって提案された談話および伝達 (communication) の理論である。これは意味論や語用論を含むさまざまな言語学および認知科学の領域に関わる理論であり、最近大変注目されている。この理論によって、*coherence* や *cohesion* といった概念に基づく伝統的な談話の理論が説明できなかつたさまざまな事実が説明可能であることが示されている。関連性の原理とは、伝達の基本的な性質であり、意図明示的伝達においては、聞き手が最適の関連性を期待できるということを述べている。最適の関連性とは、適当な処理努力によって得られる十分な (ただし必ずしも最大であることを意味しない) 文脈効果のことである。文脈効果とは、新しい情報が既存の文脈を変化させる効果のことである。ただし、文脈に対する単なる付加は文脈効果を生まない。

この理論によれば、話し手の意図は十分な文脈効果を与える意図を推論する聞き手によって解釈可能である。このように、関連性理論では、対話の理

解における推論過程の役割を大変重視している。

## 4.3 解釈と学習

解釈モジュールでは、構造写像と関連性理論に基づく複合的な処理過程によって言語情報の理解を試みる。まず、写像の候補が構造写像エンジンによって最初に計算される。次に、得られた候補の上位にあるものに対してさらに関連性を計算し、より詳細なランク付けを行う。関連性の計算を行っているとき、必要に応じて推論部との相互作用による新たな変換が加えられる。これらの処理の結果、最適の関連性を持つ候補が、解釈モジュールの最終的な出力となる。

解釈モジュールにおける学習は2種類の関連性に基づいてなされる過程である。一つは言語的関連性であり、言語 - 推論の相互作用によって計算される。もう一つは視覚的関連性であり、視覚 - 推論の相互作用から計算される。視覚的関連性は、事象の同時性、事象の時系列の連続性、文脈に対する効果などから計算される。

言語的関連性が期待できなく、視覚的関連性が十分にある場合には、言語的関連性が低いことが言語知識の薄さにあると判断して学習過程に入る。まず、視覚 - 推論の関係を言語 - 推論の関係に変換するための写像が構造写像エンジンによって計算される。次に、こうして得られた写像を利用して、言語表現を推論部で使用される表現に変換するための写像が求められる。この結果得られた写像が学習の対象となる。

## 5 考察

### 5.1 コードモデルと推論モデル

Sperber と Wilson が指摘しているように [7]、伝達のモデルにはコードモデルと、推論モデルの2つがある。コードモデルでは、伝達はメッセージの符号化と複号化によってなされる。すなわち、伝達によって伝えられる情報はすべてコードとしてのメッセージに含まれると考える。一方、推論モデルでは、話し手は意図の証拠を提示し、聞き手が話し手の意図を証拠から推論することによって伝達が達

成される。したがって、伝達される内容は話し手のみならず、聞き手にも高度に依存することになる。

伝達はコードがなくても達成可能であることから、コードモデルだけでは伝達のモデルとして不十分である。もちろん、言語が恣意的な記号からなる人工的なシステムであることも明らかであるから、コードモデルをまったく無視することはできない。

ほとんどの伝統的な自然言語理解システムはコードモデルの立場を強調してきたが、これは、人間によってなされる推論が、その高度な内容にもかかわらず多くの場合無意識になされるためにコードとしての側面が特に注意を引いたことと、無意識になされる推論は定式化が難しかったためであると考えられる。

我々のモデルは、伝達における推論モデルの側面を重視し、解釈とその学習が推論過程との相互作用によって達成されるように考えられている。

## 5.2 モジュールに依存した学習能力

我々のモデルによると、学習の能力はシステムの情報処理能力による制限を受けることになる。このことはモデルの弱点のように見えるかもしれないが、決してそうなのではない。例えば、積み木の世界のシステムは航空券の予約に関して会話をする能力は獲得できないであろう。そのシステムが持っている知識は積み木に関することで、積み木に関する推論ができるだけであり、たとえ航空券に関する言語的知識があってもシステムには利用することのできない無駄な知識である。逆に、積み木のシステムはもっぱら積み木に関する言語知識の学習に努めることにより、学習過程の進行を加速し、効率良く進めることが可能になる。

## 6 むすび

マルチモーダル情報源からの言語知識の学習のための対話システムのモデルを提案し、対話解釈のための知識の獲得について論じた。

我々のモデルを実装するシステムは現在開発中であり、UNIX 上の SICStus Prolog を用いて記述されている。

## 参考文献

- [1] Mikiko Nishikimi, Hideyuki Nakashima, and Hitoshi Matsubara. 1992. Language Acquisition as Learning. In *Proceedings of the 16th International Conference on Computational Linguistics (COLING-92)*, pp. 707-713.
- [2] E. V. Clark. 1986. The principle of contrast: a constraint on language acquisition. In B. MacWhinney, editors, *Mechanisms of Language Acquisition*. Erlbaum.
- [3] E. M. Markman. 1987. How children constrain the possible meanings of words. In U. Neisser, editor, *Concepts and conceptual development*. Cambridge University Press.
- [4] E. M. Markman and J. E. Hutchinson. 1984. Children's sensitivity to constraints on word meaning: Taxonomic vs. thematic relations. *Cognitive Psychology*, **16**, pp. 1-27.
- [5] D. Gentner. 1983. Structure-Mapping: A Theoretical Framework for Analogy. *Cognitive Science*, **7**(2), pp. 155-170.
- [6] B. Falkenhainer, K. D. Forbus, and D. Gentner. 1986. The Structure-Mapping Engine. In *Proceedings of the Fifth National Conference on Artificial Intelligence*, pp. 272-277..
- [7] Dan Sperber and Deirdre Wilson. 1986. *Relevance: communication and cognition*. Blackwell, Oxford.
- [8] Diane Blakemore. 1992. *Understanding Utterances*. Blackwell, Oxford.