

# 文脈理解のための拡張シソーラス知識表現法

4L-3

井佐原 均 石崎 俊 半田 剣一 橋田 浩一  
電子技術総合研究所

## 1: はじめに

筆者らは、文脈理解システムの研究を行う中で、自然言語理解に有効な知識表現法を検討してきた。本稿では、筆者らの開発している文脈理解システムで用いられている概念表現法の概要を説明した後、概念間の照合の基本方式としてアクティブマッチング法を提案し、さらに筆者らの概念表現法に視点の表示を導入する手法について述べる。

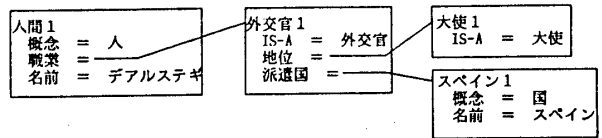


図1 ス페인大使デアルステギ氏の拡張シソーラス知識表現

## 2: 拡張シソーラス知識表現法

自然言語処理のための知識表現法としては数多くのものが提案されているが、筆者らは、概念の詳細記述と、(通常シソーラスなどで表現される)概念間の相互関係とを同一の枠組で表現できる手法を提案してきた。その特徴は、動詞概念と名詞概念とが同じ形式で記述でき、また、それらの示す情報をダイナミックに詳細化できることである。図1に「スペイン大使デアルステギ氏」を示す拡張シソーラス知識表現法の記述を示す。図における各箱はユニットと呼ばれる。また、文字列の最後に数字の付いているものはユニットを特定するためのラベルである。

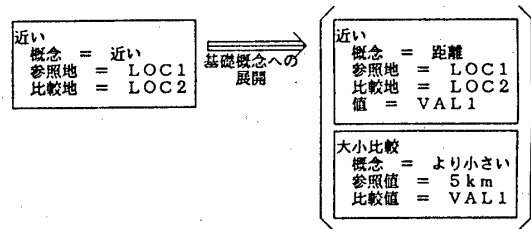


図2 「近い」の概念表現とその「基礎概念」への展開

さて、このような知識表現法を用いる場合には概念をどの程度に詳細化しておくかということが、実際のインプリメントの上で重要な問題となる。ここでは、距離概念について考察した例を示す。

「駅の近くのアパート」という表現の示すものをどのように表現するかを考えてみよう。「近い」という概念を何らかの形で表現するわけであるが、ひとつの手法として考えられるのは「基礎概念」といった非常に細かな概念を考え、取り扱う概念は全てこの「基礎概念」の組合せで表現しようとするものである。しかしながら、このように常に全ての概念を「基礎概念」にまで細分化すると、細分あるいは照合に不必要に手順がかかることが予想される。

ここでは、「近い」という言葉の示すものそれ自体を概念としておき、必要に応じて、より詳細な情報を検索できるようにしておく。ここで用いる「近い」の表現と、必要な場合に参照される「距離概念」の枠組を図2に示す。

(図では「近い」ということは、数字で表わすと5 kmより小さいことだと定義してある。)日本語の「近い」という単語は「近い」を示す一つの概念を指す。その概念が、より詳細に言及されたり(たとえば、「5 m」とか)、別の表現で言及されたり(たとえば、「付近」とか)した場合には、この「近い」という概念を付属する「『近い』と距離概念一般とを対応付ける情報」と関連付けて処理することにより、「近い」と「5 m」あるいは「付近」とが、同一の概念を示しうるかどうかを判定できることになる。

このような枠組を翻訳システムにおいて利用することを考えた場合には、たとえば、日本語の「近い」と英語の“near”が同じ概念を示しているとする、これらの対応付けにおいては、直接の照合が可能であり、この表現を展開する必要はなく効率が良い。また、概念が微妙に異なる言語間でも必要に応じて展開することによって、必要な情報は得ることができる。

## 3: アクティブマッチング法

一般に自然言語による文章では同じ対象を指示するために、様々な表現を用いる場合が多い。人間はこのような相異なる表現が同じ物を指すのかどうかを、容易に判断できるが、計算機にこのような作業をさせることは決して簡単な処理ではない。

H. P. Griceの会話の公理は、我々人間が協動的に会話を進めるうえでの従来からの慣習を明文化したもので、Quantity, Quality, Relation, Mannerの4種の公理からなる。これらの公理は人間の会話に関するもので、自然言語処理のレベルとの間には大きなギャップがあることは言うまでもない。

ここで筆者らは「アクティブマッチング」という考え方を提案する。これは、我々人間が文章を読んだり、人の話を聞いて理解する時に、大変効率的に理解が行われることの一つの基準を示すものである。

意味的にひとまとまりの文章において、表層上でいくつかの表現があり、それらの指示物(外延)を同定する時、明確な矛盾がない限り、アクティブに同一物としてマッチングを取ることを「アクティブマッチング」と呼ぶ。

ある二つの概念について、それらの概念が持つ属性の値(実際にはスロット名とそこに埋め込まれる概念の構造の組合せで表現される)が互いに矛盾しなければ、その二つの概念が同一のものであると仮定する。これは、unificationを行っていることになり、たとえば、一方の属性値(として埋め込まれた構造)が他方の部分集合(空集合も含む)となる場合も同一のものであると仮定される。このような照合規則に基づいて、入力文章に出現する概念同士の同一性を判断してゆく。

4: 概念表現への視点の導入

ここまで述べてきた知識表現法は元来、事実関係のみを記述するものであった。すなわち、「AがBを殺した」といったような事実を表現するものであり、「Aが殺人者だ」とか「Bを殺したのはAである。」とかいった表現の違いを記述することは出来ない。このような記述が必要となるのは、その記述自体を参照する場合が考えられる。たとえば、「『Bを殺したのがAであった』ために、Cは・・・」といった場合である。このような記述を可能とするために、ここでは、図3に示すような特別なユニットを導入する。図3において“ROLE”というラベルを持つユニットがこのような視点を表現しており、これを日本語で読み下すと、「関係 スロットの指す概念(ここでは、“kill” )における、概念 スロットが指定する名前のスロット (“actor”) の中身が、“role” スロットの指す概念 (“A”) であるということ」をこの箱は示している。「AがBを殺した。」「Bの殺人者A」「Aが殺人者である」は各々図において「殺す1」「A1」「ROLE1」を指すことによって示されることになる。したがって、図3は全体として「『Bを殺したのがAであった』ために、Cは破滅した。」という事実を表現している。

さて、このような表現を導入した場合に“role” スロットの指す概念が、既に出現している特定のインスタンスとなる場合ばかりではなく、そのインスタンスの上位のジェネリックな対象を指す場合がある。次の例を参照されたい。

性格が良く、賢い男がいた。  
 Bを殺したのが「性格が良く、賢い男」だったために・・・  
 Bを殺したのが「賢い男」だったために・・・  
 Bを殺したのが「男」だったために・・・  
 このような場合には表層の文から抽出されて“role” スロットに埋め込まれた概念は特定のインスタンスを指すのではなく、すでに言及された、あるインスタンスの上位となるジェネリックな概念を示すことになる。ここで、これらのインスタンスとジェネリックな概念との間を結びリンクを適切に繋げることが必要となる。(図4)

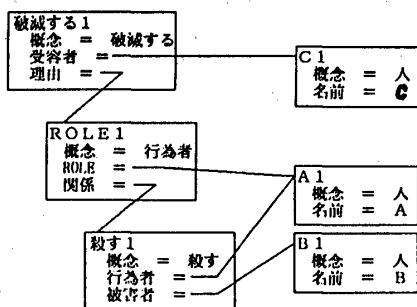


図3 視点の表現

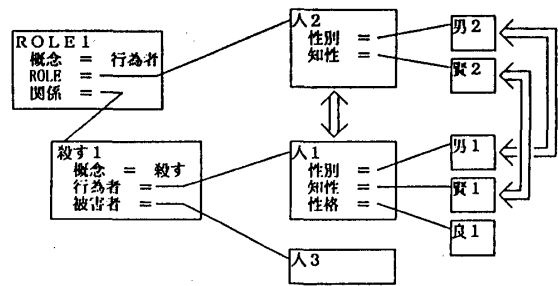


図4 ジェネリックな知識の表現

(太い矢印はジェネリックとインスタンスとの対応を示す)

5: おわりに

本稿で述べた手法により、新聞記事等に出現する情報を適切に表現し、照合することが可能であるが、一般の文章に対しても有効であることは十分に期待できると思われる。今後は、さらに多くの種類の記事を参照することにより、そこに出現する様々な情報をも表現できるように概念表現法を拡張して行く予定である。

参考文献

- [1] 石崎俊・井佐原均, 文脈と言語理解, 電子通信学会言語処理とコミュニケーション研究会, 1986
- [2] 井佐原均・石崎俊, Context Analysis System for Japanese Text, Proc. of COLING'86
- [3] 石崎俊・井佐原均・半田剣一, Natural Language Processing System with Deductive Learning Mechanism, International Symposium on Language and Artificial Intelligence, 1986
- [4] 井佐原均・石崎俊・半田剣一, 文脈解析システムにおける概念表現とその照合法, 情報処理学会第32回全国大会, 1986
- [5] 石崎俊・井佐原均・内田ユリ子・橋田浩一・元吉文男, 機械翻訳システムCONTRASTにおける文脈情報の利用, 情報処理学会第33回全国大会, 1986
- [6] H.P.Grice, Logic and Conversation, Syntax and Semantics, Vol.3, Academic Press, 1975