

3B-1 オブジェクト指向型概念辞書COOLにおける共起関係の表現

横山 晶一 ・ 羽中田 賢治 ・ 鈴木 雅実

(電子技術総合研究所) (シュトゥットガルト大学) (KDD上福岡研究所)

□はじめに 我々は、自然言語処理のための知識表現の体系を構築することを目指して、国語辞典の語義文から語と語の関係を抽出し、それを、大規模処理に向けたオブジェクト指向型言語であるCTALK上にインプリメントすることによって、概念辞書の構築を試みてきた¹。また、外国語の辞書の語義文と比較することによって、概念のグループ同士が多言語間で多対多の関係で対応することも示唆した²。本稿では、国語辞典の語義文に含まれる例文を利用して、語の共起関係を抽出し、それをオブジェクト指向型概念辞書の上を実現する手法について述べる。これまでに述べてきた手法を拡張すれば、ほぼそのままの形で例文に示された共起関係を概念辞書の上にインプリメントすることができる。なお、このオブジェクト指向型概念辞書をCOOL (Conceptual, Object-Oriented Lexicon)と名付ける。

□背景 一般に、辞書的な意味を語の意味と定めた場合には、次のような三層にわたる意味を定義することができる。

意味の三層

1. 静的な意味関係：国語辞典の語釈文（例文を除く）から抽出される上位-下位、全体一部分、同値関係を、その語の意味とする。また、前記の関係以外でも、語義文から得られる語と語の関係を、その語の意味とする。

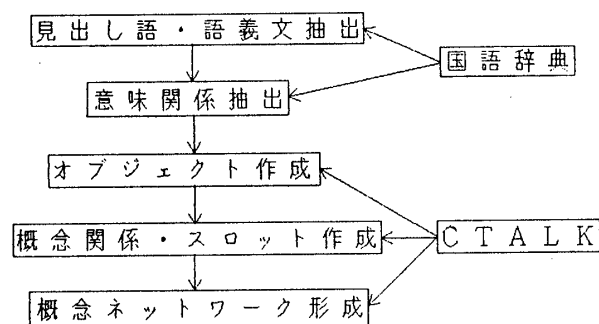


図1 概念ネットワークの作成手順

2. 半動的な意味関係：国語辞典の語義文中に含まれる例文は、1に比べると、動的だが、半ば固定した決まった関係である。

3. 動的な意味関係：実際の文を構文解析している際にその場の状況から得られる情報である。これは、語用論的な意味を取り出すことに相当する。

国語辞典で取り扱える範囲は1と2である。このうち1についての手順を図1に示す。

この図は、すでに示したもの^{1,2}とほぼ同じであるが、簡単に説明する。まず、国語辞典の語義文から、上下、全体一部分などの関係（抽象的な表現の場合には、特に語と語の関係）を抽出して、CTALK上にインプリメントする。各々の語に対して、その語の名前を持つ world（CTALKで、オブジェクトをグループ化したもの）と、語義文に関係した語を反映する複数のオブジェクトによって、上下関係をオブジェクトの親子関係とし、そのほかの関係を主としてスロット関係として取り扱うネットワークを作成する。2の部分で、今回述べるところであるが、この基本部分を変更せずに、システムを拡張する形で共起関係をCOOL上に実現する。

□手順 辞書の例文は、図2に示すような形で辞書に記載されている（-が見出し語を示す）。

「-を・果たす（達する・貫く・見失う）：
-に・かなう（沿う・反する）：-のため
には手段を選ばない：-地④」

図2 「目的」の例文表現³

これを、次のような手順でCOOL上にインプリメントする。

まず第一に、図2を見ても分かるように、辞書の例文は、おおむね助詞を基準にして並置されている。そこで、助詞を中心とした簡単な解析を行う。図2の例文では、たとえば、助詞「を」を中心にする、「目的」に対して4つの動詞が共起

する(図2の"~(…)"という表現は、OR表示で、~が…以下でそれぞれ置き換え可能であることを示す)。「目的のためには手段を選ばない」のような、文の形をとるものについては、語義文の取り扱いとほとんど同じことをする。つまり、この文を(手で)解析して格関係に直し、文中の各語の関係を意味関係として捉える。

第二に、このような関係を、各語を表わすオブジェクト(図1のオブジェクト作成のところで、各語に対応したオブジェクトと、各語の名前を持つworldとを作成し、語義文の解析結果を反映したネットワーク関係がすでに成立している)のスロットの中に埋めこむ。スロット名は、語義文ではなく、共起関係から得られたものであることを明示するために、語の前後に"co"をつける。たとえば、オブジェクト"mokuteki/mokuteki"(/の後ろの"mokuteki"は、world名を示し、前の"mokuteki"は、そのworld名の中にあるオブジェクト名を示す)には、"co-wo"というスロットの中に、
(hatasu tassuru turanuku miusinaw)

という値が共起関係を表わすものとして入れられる。また、このworldのトップレベルオブジェクトとして、"TOP/mokuteki"が存在するが、ここには、生の語義文とともに、"cooc"というスロット名の中に例文自身を入れてある。これは、人間とのインタフェースを行う部分である。これらの記述は、methodに対してメッセージ送信を行うことによって出力される。

第三に、否定的な表現がある。辞書には、共起関係とともに、決して共起しない表現というものがある。国語辞典にはこの記載例はあまり多くないが、「~とは言わない」という説明があった場合には、「neg-cooc」という非共起関係を記述するスロットを設ける。たとえば、上の例で、「目的を果たす」とはいえるが、「目標を果たす」とはいえない。辞書に明示されていない非共起関係であっても、類似の語に対して人間が非共起関係を見付けた場合には、このスロットに値を入れる。実際の知識表現では、これが抑制的なスロットとして働くことを目指している。

第四に、共起する単語が、現在のworldに存在しない場合には、その単語の名前と、対称のスロット値を持ったオブジェクトがそのworld内に生成される。たとえば、「目的を果たす」の例では、「hatasu/mokuteki」というオブジェクトが、新たに生成され、その中には、「wo-co」("co-wo"の逆の名前)というスロットが存在して、「mokuteki」というスロット値が入っている。これによって、

「目的を果たす」という共起関係に対して語の順序関係を保存するようなスロット関係が生成される。「hatasu/mokuteki」というオブジェクトは、上位概念として、(そのworldのTOPというオブジェクトはもちろんであるが)「hatasu/hatasu」というオブジェクトも同時に有する。この親オブジェクトが存在しない場合には、新たに生成する。

第五に、国語辞典では、しばしば語義の部分ではなく、例文を示すことによって特殊な意味を説明することが多い(むしろ例文でオーソドックスな意味が出てくることは少ない)ので、これに対処するために特別な意味の取り扱い規則を設けておく。たとえば、図2の例文中には含まれていないが、類似の語「目標」の例文中には、

「目標 [=ねらい] を決める」

という形で、[]の中にこの場合の「目標」の意味が含まれている。この場合には、「mokuhyou」と「kimeru」の両方のスロットに、「nerai」という意味があることを明示する。たとえば、「mokuhyou/mokuhyou」の「co-wo」というスロットには、

(utidasu kakageru (kimeru nerai) …)

という形で、「kimeru」と共起する場合には、「nerai」という特殊な意味になることが明示される。

□考察 以上のように、辞書の例文からの共起関係の抽出に焦点を絞って概念辞書COOLの構築の一端を紹介した。これによって、語義文だけでなく、やや動的な意味を持ちながら、しかも慣用的には固定化された表現をとる、共起関係の記述を概念辞書の中に取り入れることができた。しかしながら、上の小規模な例からも分かるように、ネットワークの形が次第に複雑になって来る。構文解析、意味解析も含めて、多くの部分をCTALKの種々の組み込み関数やそれらを組み合わせたmethodを用いて自動化するとともに、現在の平面的な記述方法から、ある程度メタなオブジェクト構造を導入した階層的な構造に概念辞書を書き換えるべく、現在検討を進めている。また、いくつかの例で示したように、スロット名やスロット値は、現在のところローマ字で記載されている。これを日本語化することも今後の課題である。国語辞典の情報は、語義文、例文にとどまらず、派生語、用法の説明など多岐にわたる。したがって本稿で述べた以外にも未使用の情報が数多く存在する。今後はこれらの情報の利用も検討していきたい。

□参考文献 1. 横山・羽中田: 情処全大34, 6L-3 (1987.3). 2. 横山・羽中田: 情処全大35, 1P-7 (1987.9). 3. 金田一他編: 新明解国語辞典(第3版), 三省堂(1981)