

文の生成に必要な言語知識と辞書構成

3B-2

鈴木 雅実

国際電信電話株式会社上福岡研究所

1. はじめに

概念的意味表現を基に文を生成する際に、類義語の中から適切な語を選択するための知識に焦点を当て、辞書の持つべき機能について考察する。

2. 問題提起 意味表現から文を生成すると言っても、どのような文脈や前提情報が与えられるかによって事態は変化する。さらに、その生成処理が目標とするタスクや伝達すべき意図によっても、参照される語彙情報の種類は異なる。実際いずれの場合も、似通った意味を持つ複数の語群の中から、然るべき語を選択して生成に用いるという手順を避けられないことが多い。

一例として、次のような状況を仮定する。

課題: 与えられた意味表現を用いて特許出願用の文章を作成する。

全体の構成: 従来技術の概括を行なった後、著者独自の新技术が述べられる。

状況設定: 従来技術Aについての説明

部分意味表現: USE(A) → ◇OCCUR(B)

注) USE, OCCURは概念記号で、()内はその引数
ここでは様相論理的な表現を用いた

上記の可能演算子◇の部分の語彙表現は幾つも考えられる。たとえば、～する可能性がある、～し得る、～となることがある、etc. ここでもし、著者が従来技術Aを利用した場合の副作用Bについて論じていて、後でその対比として新技术の特長を主張したいのであれば、この部分は、～とになってしまうこともある、～となるおそれがある...のような表現も考えられる。このように、文脈や意図によって、選択すべき語彙が変わり得る。

3. 生成に必要な言語知識と辞書 従来の概念辞書は、概念と語彙表現の対応関係が多対多のものも存在するが、類義語間の関係を記述したものや、文脈によって動的に語彙を選択するような機能を備えた辞書は極めて少ない。また、概念表現の側に求められるのは、明確な意図の表示とタスクの目標設定である。さらに読者への効果を考慮した、レトリックの活用も重要である。^[1]

筆者が先に提案した、オブジェクト指向の多言語辞書は、類義語グループを中心とした構成をとり、単語間の示差的特徴と階層関係を利用して、逐次的に適切な語彙を選んで行く辞書モデルの実現を目指したものである。^[2] 現在はこれをさらに一歩進めて、意図や背景の設定に応じた語彙選択方略の規則化を検討している。

各語彙オブジェクトに記述されている情報は大きく分けると次のようになる。

- 1) 語の使用に関する知識: 統語・意味情報など
- 2) 語の選択に関する知識: ある状況において自身を選ぶか、他の語を選ぶかを判断する規則

この2の規則に必要な情報は、外部の実世界知識ベースや文脈情報を参照しなければ得られないものも含まれる。ところで、ある視点(語彙の意味に影響を及ぼす特定の立場) P_i から見た場合の特徴 f_i を一時的に語彙オブジェクトに与え、語の選択にそれを利用することが考えられる。これを実現するために、外部知識ベースとリンクしたスロットを語彙オブジェクトに設け、状況に応じてその内容を書き換えるような枠組みを考案した。これにより、語の意味のうち状況に大きく依存する部分を外的知識と関連づけて陽に扱うことができる。

4. 検討 以上で述べた内容は、どちらかという to say に近いが、もちろん what to say は重要である。ここでは意味表現を与えられたものとして扱ったが、ある時点の文脈・意図の下で生成すべき内容を決定するための戦略的な知識についても考察したい。また、ある戦略に従って文を生成して行くと、それ以降の概念構造を変更せざるを得ない場合もあることが知られている。^[3] これを避けるためには、上記の1と2の知識の関連を辞書に反映させる工夫が必要となろう。

参考文献

- [1] 田淵・辻井・長尾: 文脈を考慮したテキスト生成システム, 情処NL研65-6, 1988
- [2] 鈴木: CTALKを用いたオブジェクトベース辞書の構成について, 情処全大36, 4U-2, 1988
- [3] L. Danlos: Conceptual and Linguistic Decisions in Generation, COLING'84

Lexical Knowledge for Sentence Generation

Masami SUZUKI

Kamifukuoka R & D Laboratories, Kokusai Denshin Denwa CO., LTD.