

## 意味記憶システム I X

- 意味ネットマシン I X M 2 による word disambiguation の実験 -

## 4 F - 8

樋口哲也 高橋直人\* 半田剣一 古谷立美 国分明男

電子技術総合研究所 \*筑波大学

## 1. はじめに

自然言語解析における大きな問題点の一つに、ambiguityの解消がある。Ambiguityには、語義のambiguity、格関係のambiguity、統語的なambiguity等がある[1]。これらのうち語義のambiguityは、述語(主に動詞)とその格スロットのfiller(主に名詞)が満たすべき条件の組み合わせを考慮することで解消出来る。しかし同音異義語が多い場合や格スロットのfillerとなり得る語が数多く存在する場合には、組み合わせの爆発の発生により計算時間が急激に増大する可能性がある。

本稿では大容量連想メモリを活かして超並列処理を図る意味記憶システムIX上に単語間意味ネットワークのモデルを構築し、それによって語義のambiguityを解消するための手法について考察する。具体的には、多義性を持つ複数の単語から成る文において、正しい語義の組み合わせを求める。

## 2. 意味記憶システムIX

意味記憶システムIXは連想マルチプロセッサマシンIXM2と知識表現言語IXLからなる。

IXM2は、意味ネットワークやフレームなど、ネットワーク型の知識ベースを高並列処理する連想マルチプロセッサである。図1に示すように64台の連想プロセッサと9台の通信プロセッサから成るシステムが完成している。各連想プロセッサはT800トランジスタ(10MIPS)と4K語の連想メモリを中心に構成される。

連想メモリは最大256K語まで実装可能であり、一つの単語に連想メモリの2語を割り当てると全体で約13万個の単語ネットワークに対する同時検索や集合演算を行うことが可能である(現在作成中の処理プログラムでは実験範囲の拡張を考慮し、1単語あたり4語を割り当てている)。

各単語に対応する連想メモリの領域には、その単語に接続するリンクの情報、接続相手先のほかにマークビットと呼ぶ1ビットのフラグが複数個用意され、検索や集合演算の結果が格納される。

IXM2はワークステーションのバックエンドプロセッサであり、現在はSUN-3/260をホストとする。IXM2のプログラミングは、知識表現言語IXLを用いて行う。IXLはネットワーク型のデータ処理を

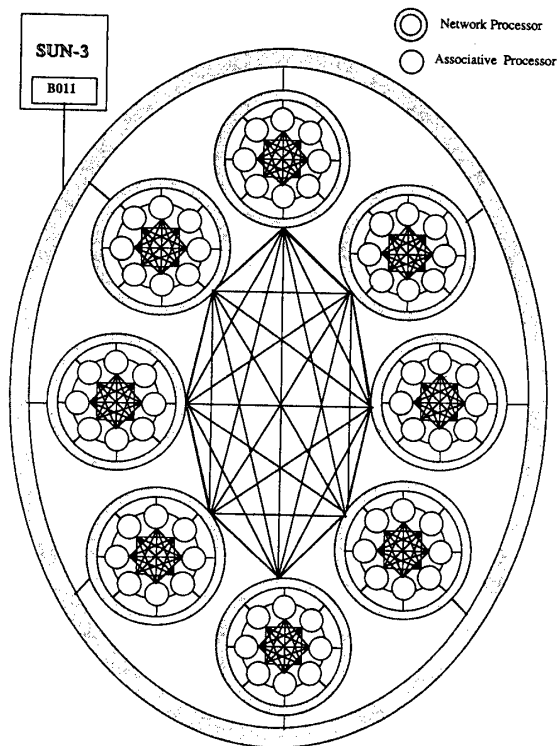


図1 IXM2の構成

柔軟に記述することができる。IXLはPrologのスーパーセットであり、通常のPrologの述語のほかに意味ネット処理用の述語を加えた。それら意味ネット処理用の述語はホスト上のIXL処理系からIXM2に渡され、そこで解が全探索で求まる。

今回の実験では、ambiguity解消用の述語を定義し、またその述語をIXM2上で実行するための、IXM2機械語ルーチンを記述する。

## 3. 単語ネットワークのモデル

実際の用例から大規模な単語間のネットワークを作るには大きな労力が必要とされる。今回は以下のような単語間ネットワークのモデルを定め、大規模単語間ネットワークをプログラムで生成してIXM2にロードする。

- ・ネットワークのユニットには、名詞ユニット(ex. 図2の飛行機)、動詞ユニット(ex. 滑走する)、名詞綴りユニット(ex. "terminal")、動詞綴りユニット(ex. "taxied")の4種類がある。またユニット間のリンクには、spelling、isa、subject、objectの4種類がある。

- ・名詞ユニットと名詞綴りユニットとの間、およ

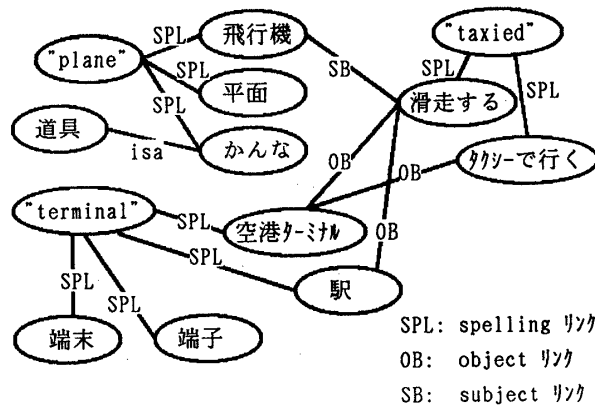


図2 単語ネットワークの例

び動詞ユニットと動詞綴りユニットとの間は spelling リンクで結ばれている。spelling リンクは多対1の関係である。これは同一品詞内での同音異義語の存在を表している。

・名詞ユニット相互間は isa リンクで結ばれている。一般に1つの名詞ユニットは複数の名詞ユニットを上位概念として持つ。ただし isa リンクはループを持たないものとする。これは名詞間のソーラスに対応するものである。

・動詞ユニットと名詞ユニットとの間は subject リンクあるいは object リンクで結ばれている。subject リンクおよび object リンクは動詞ユニットと名詞ユニットを多対多の関係でむすぶ。ある動詞ユニットと subject (object) リンクで結ばれた名詞ユニットは、その動詞の主語(目的語)となり得ることを意味する。また、動詞ユニット1と名詞ユニット1とが subject (object) リンクで結ばれており、かつ名詞ユニット1が名詞ユニット2の isa の意味での上位語であるならば、名詞ユニット2もまた動詞ユニット1の主語(目的語)になり得ることを意味する。

#### 4. Ambiguity 解消のアルゴリズム

この実験では、たとえば "Nadia's plane taxied to the terminal." という文があり、かつ SVO という構文解析結果があるときに、plane を「飛行機」、taxied を「滑走する」、terminal を「空港ターミナル」の意味に解釈したい。その際の word disambiguation を行なうための、IXM2 機械語レベルでのアルゴリズムを以下に示す。

基本的な手段は marker passing と集合演算であり、各ユニットが自分の持っている情報をリンクを通じて伝搬させることで正しい主語-動詞-目的語の組み合わせを決定する。

(1) 動詞の綴り (ex. "taxied") をホストから受け取り、その綴りに対応する動詞綴りユニットと

spelling リンクで結ばれる動詞ユニット全て (ex. "滑走する") に V というマークをつける (マーク付けは連想メモリ内の特定のマークレットのセットに対応する)。

(2) 主語 (又は目的語) の綴りをホストから受け取り、その綴りに対応する名詞綴りユニット (ex. "plane") と spelling リンクで結ばれる名詞ユニット全て (ex. 平面) に S (又は O) というマークをつける。また、マーク S (又は O) が付いている名詞ユニットの (isa の意味での) 下位ユニットにもマーク S (又は O) をつける。

(3) マーク S (又は O) のついた名詞ユニット、あるいはマーク S (又は O) のついた名詞ユニットの上位ユニットと subject リンク (又は object リンク) で結合されている動詞ユニットすべてにマーク S (又は O) をつける (subject マーク、あるいは object マークの伝搬)。

(4) SVO すべてのマークが付いている動詞ユニット (図2では "滑走する") の ID と、マーク S が付いている名詞ユニットで、この動詞ユニットの主語となり得るもの (飛行機) の ID、およびマーク O が付いている名詞ユニットでこの動詞ユニットの目的語となり得るもの (空港ターミナル) の ID をまとめてホストに返す。

要するに与えられた綴りを持つ動詞のうち、S、V、O すべてのマークが付けられたものが正しい語義の動詞である。また与えられた綴りを持つ名詞のうち、この動詞の主語または目的語となり得るものが正しい語義の名詞となる。

#### 5. おわりに

意味記憶システム IX 上に単語間意味ネットワークのモデルを構築し、それを用いて word ambiguity を解消するための仕組みについて論じた。現在本稿で述べた方式に従った実験を進めており、C、Prolog で同様のプログラムを作成して計算速度の比較を行なう予定である。また今回は disambiguation を同一品詞内に限ったが、将来的には多品詞語を含む文をも disambiguation の対象としたい。

末筆ながら連想メモリでご援助を頂いている NTT・LSI 研究所の小倉武氏、本研究の機会を与えられた柏木寛電総研所長、日頃御指導頂く棟上昭男情報アーキテクチャ部長に感謝する。

#### 参考文献

- [1] 田中・辻井共編: 自然言語理解, 丸善社 (1988)
- [2] Hirst, G.: Semantic Interpretation and the Resolution of Ambiguity, Cambridge University Press (1987).
- [3] 樋口、古谷、半田、楠本、国分: 意味ネットワークマシン IXM 第2版の概要、情報処理学会計算機アーキテクチャ研究会、78-3、1989.