

## 概念ベースを用いた常識場所判断システムの構築

杉本 二郎† 渡部 広一† 河岡 司†

人間はあいまいな情報を受け取った場合にも、適宜に解釈し、適切に会話を進めたり、適切な行動をとったりすることができる。これは、人間が長年の経験により、言語における知識を蓄積し、その基本となる概念に関する「常識」を確立しているからである。コンピュータをより人間に近づけるためには、常識を踏まえて物事を扱うことができる「常識的判断システム」の構築が必要不可欠であると考え、人間が日常的に用いている常識のうちには、「場所」に関する常識がある。本稿では、日常語の知識ベースを構築し、あらゆる名詞に対してそれが場所を表す語か判断し、その場所における存在物と何をやる場所なのかを連想する手法を提案する。また国語辞書等から自動的に構築された概念ベースを用いて、知識ベースに登録されていない言葉からも存在物と役割を想起する未知語処理方式を提案する。これによって、場所に関する意味理解の実現を目指す。

## Constructing a Commonsense Place Judgment System Which Uses a Concept Base

Jiro SUGIMOTO† Hirokazu WATABE† Tsukasa KAWAOKA†

When we humans receive uncertain information, we interpret it suitably, to understand what the speaker is trying to say. This is possible because we have "commonsense" concerning the word, which is built up from knowledge that is stored from long time experience. In order to understand the meaning, we think that the construction of a "Commonsense Judgment System" is necessary. Of the commonsense we have, one concerns the information of "place". This paper proposes a mechanism, which judges if a word expresses a certain place and associates existing objects and what people do at that place, with a knowledge base consisting of basic words. By using a concept-base, which is automatically constructed from dictionaries, this mechanism can understand a word that does not exist in the knowledge base. This study aims to retrieve the meaning concerning place, and deepen semantic understanding.

### 1. はじめに

コンピュータを人間にとってより使いやすいものにするには、人間同士が日常行っているコミュニケーション（意思疎通）の仕組みをモデル化し、コンピュータと人間とのインタフェースにもこれを取り入れることが望まれる。

人間はあいまいな情報を受け取った場合にも、適宜に解釈し、適切に会話を進めたり、適切な行動をとったりすることができる。これは、人間が長年の経験により、言語における知識を蓄積し、その基本となる概念に関する「常識」を確立しているからである。人間が日常的に用いている常識には、大きさ・長さ・重さ・速さなど比較的数値に直し易い量に関するものや、クリスマス・夕方といった時間に関するものや、神社・病院・目の前・向こう側といった場所や位置に関するもの、さらには、嬉しい・悲しいなどの感情に関するものなどがある。コンピュータをより人間に近づけるためには、このような常識を踏まえて物事を扱うことができる「常識的判断システム」の構築が必要不可欠であると思われる（図1）。

この「常識的判断システム」の要素技術には、「場所」に関する常識を判断する「場所判断システム」がある。

本稿では、ある名詞が場所に関する語であるか判断し、もし場所に関する語なら、そこには何が存在し、何をやる場所なのかを日常的な場所に関する語の知識ベースを構築することによって連想する手法を提案する。また、概念ベースを用いることによって、知識ベースにない語に対しても、同様の判断・想起する方法についても述べる。

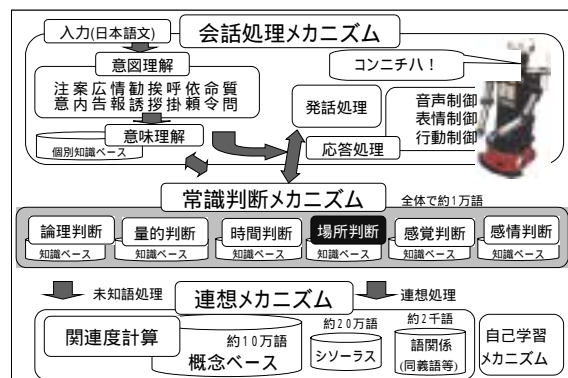


図1 常識判断システムの論理構成

†同志社大学 工学研究科  
Graduate School of Engineering, Doshisha University

## 2. 場所判断

人間は「近くに病院はありますか?」といえば「目の前にあります」や、「病院へ行ってきました」といえば「医者に診察してもらったのですか?」といったような会話を当然のように行っている。これは人間が「近く」や「目の前」といった位置的な概念を知識として持っており、また「病院」は場所であり、医者・患者などが存在し、診察・手術などをする所であると理解・判断し、相手のいわんとする物事を連想しているからである。

ここで、日常会話で用いられている場所に関する語について大別すると、以下のように2つに分けられる(表1)。

表1 場所を表す語

具体的な場所を表す語
神社, 病院, 警察署, 山, 川, オフィス etc.
論理的に場所を表す語
目の前, 向こう側, 端, 近く etc.

本稿では以上2つのうち具体的な場所を表す語を対象とする。

場所判断は、ある名詞が場所に関する語であるか判断し、もし場所に関する語なら、そこには何が存在し、何をやる場所なのかを想起することを目指す。

このような判断ができることによって、以上のような人間が行っている柔軟な会話を可能にするだけでなく、音声認識の精度向上などにも役立つと考えられる。例えば、「ビョウイン ヘ インサツ ニ イキマシタ。」と誤認識したとしても、場所判断によって、「病院」から「診察」という言葉を想起でき、インサツ(印刷)をシンサツ(診察)に修正することができ精度向上が可能となる。

### 2.1 場所語

本稿における場所に関する語(以下、「場所語」と呼ぶ)とは、神社・病院・警察署・山・川・オフィスなどのようにその名詞単独で、そこには何が存在し、何をやる場所なのかを連想可能となる語と定義する。

更に詳しく定義付けすると、場所語は、シソーラス<sup>2)</sup>に存在する上記の条件を満たすノード「場所」・「家屋」以下に属する全ての語とする(図2)。

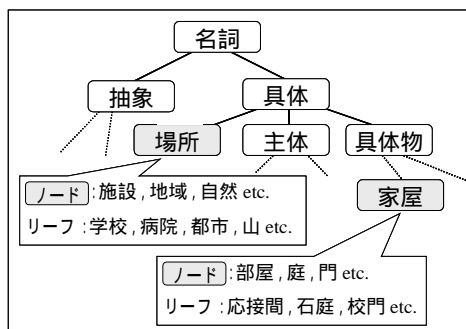


図2 場所語と定義された語の例

ここで、シソーラスとは一般名詞の意味的用法を表す2710個の意味属性(ノード)の上位-下位関係、全体-部分関係が木構造で示されたものであり、ノードに属する名詞として約13万語(リーフ)が登録されている。

### 2.2 主体語・目的語

前述した「場所語」において、何が存在するかを示した具体名詞を「主体語」と呼び、人間にとって何をやる場所なのかを示した語を「目的語」と呼ぶ(表2)。

表2 場所語と主体語・目的語の例

場所語	主体語	目的語
神社	神主, 鳥居 etc.	参拝, 祈願 etc.
病院	医者, 患者 etc.	診察, 入院 etc.

### 2.3 場所語知識ベース

すべての場所語に主体語と目的語を記憶させることは困難であり効率が悪い。そこで、よく使われる場所語を「代表語」(443語)として「場所語知識ベース」に持たせ、代表語から連想される主体語・目的語を上記の表2のように登録する。

この代表語は、2.1節で述べたように、シソーラスのノード「場所」・「家屋」以下に属するノード(120語)から選定されており、これらのノードを「分類語」と呼ぶ。

場所語知識ベースの「分類語」はシソーラスのノードの一部であり、この分類語自身にも主体語・目的語を付加している。ここで場所語知識ベースのイメージ図を図3に示す。

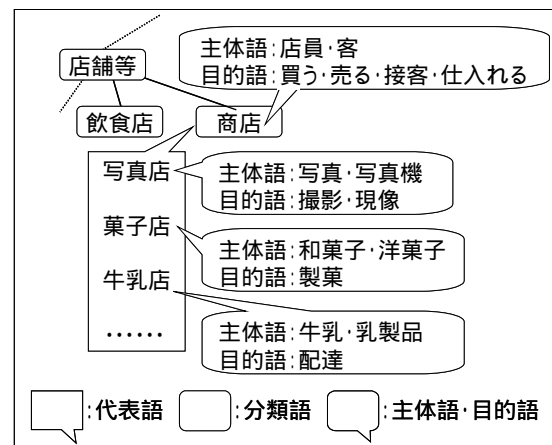


図3 場所語知識ベースのイメージ図

場所語知識ベースで、分類語を木構造化することによって、主体語・目的語の継承が可能となり、継承された情報を含めてその代表語の「主体語」・「目的語」とする。この場所語知識ベースによって代表語と主体語・目的語の間に論理関係を持たせることができ、代表語に関しては、場所判断、つまり、その特徴である主体語・目的語を想起することが可能となる。単語の固有の主体語・目的語と分類語とし

ての主体語・目的語が出力可能となる。例えば、代表語「写真店」の固有の主体語・目的語として「写真・フィルム・カメラ・カメラマン」、「撮影・現像」を取得し、分類語「商店」の主体語・目的語として「店員・客」、「買う・売る・接客・仕入れる」が取得され、最終的に「写真店」から「店員・客・写真・フィルム・カメラ・カメラマン」という主体語と、「買う・売る・接客・仕入れる・撮影・現像」という目的語が想起可能となる。

場所語知識ベースのすべての主体語・目的語情報は、人手によって作成され、検証された。したがって、この情報は正確であり、代表語に関する場所判断は正確である。

### 3. 概念連想メカニズム

場所語知識ベースを用いることによって、代表語に関する主体語・目的語の想起が可能となる。しかし、会話に出現する場所語、つまり、常識的判断システムに入力される多くの語は場所語知識ベースに登録された代表語ではない。これらに対する主体語・目的語の想起も必要であり、すべての場所語に対する場所判断が必要である。場所語知識ベースに登録されていない単語を「未知語」と呼び、これらに関して場所判断を行うメカニズムも必要となる。

未知語に対して、場所語知識を利用するためには、概念連想メカニズムを構成する概念ベース(3.1節)や関連度計算(3.2節)を用いることにより、これらの未知語について、意味的な関連やその強さの度合いを評価し、最も関連の強い代表語を決定する必要がある。この処理を総称して「未知語処理」と呼ぶ(図4)。

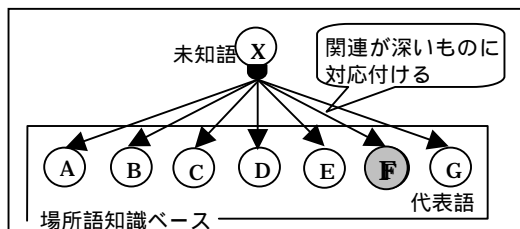


図4 未知語処理のイメージ図

#### 3.1 概念ベース

ある語  $A$  をその語と関連の強いと考えられる語  $a_i$  と重み  $w_i$  の対の集合として定義する。

$$A = \{(a_1, w_1), (a_2, w_2), \dots, (a_m, w_m)\} \quad (1)$$

ここで、 $a_i$  を1次属性と呼ぶ。また便宜上、 $A$  を概念表記と呼ぶ。このような属性の定義された語(概念)を大量に集めたものを概念ベースと呼ぶ。ただし、任意の1次属性  $a_i$  は、その概念ベース中の概念表記の集合に含まれているものとする。すなわち、属性を表す語もまた概念として定義されている。したがって、1次属性は必ずある概念表記に一致するので、さらにその1次属性を抽出することができる。これを2次属性と呼ぶ。概念ベースにおいて、

「概念」は  $n$  次までの属性の連鎖集合により定義されている。

本論文では、複数の国語辞書から、各見出し語を概念表記、その見出し語の説明文中の自立語を1次属性として抽出し、出現頻度に基づく重みを付加した約4万の概念からなる概念ベースを基に、新聞などから抽出した概念表記や属性を加え、質の向上を目的にした精練操作(属性の追加・修正など)を施し、更に、概念間に成り立つ一般的なルールに基づく適切な重みを付加した約10万の概念からなる概念ベースを構築し利用している。

#### 3.2 関連度計算

関連度<sup>4)</sup>とは、関連の種類しか特定できない意味ネットワークのようなものとは違い、概念と概念の関連の強さを定量的に評価するものである。具体的には概念連鎖により概念を2次属性まで展開したところで、最も対応の良い1次属性同士を対応付け、それらの一致する属性個数を評価することにより算出するものである。関連度は、概念間の関連の強さを0と1の間の数値で表す。表3に例を示す。

表3 関連度計算結果の例

概念 A	概念 B	概念 A と概念 B の関連度
病院	医者	0.42
病院	林檎	0.01

#### 4. 場所語であるかの判断

代表語または未知語の中でシソーラスに存在する語(以下、「未知語」と呼ぶ)に関しては、場所語知識ベースを利用し、分類語に属するか否かを参照することで場所語であるか判断ができるが、例えば「シアター」や「保健室」、「青林檎」などといったシソーラスに存在しない未知語(以下、「未知語」と呼ぶ)については、場所語であるかどうか分からない。そこで、未知語については、その単語が属する分類語を決定し、場所語であるか判断しなければならない。この処理を「分類語決定」と呼ぶ。

##### 4.1 分類語決定

「分類語決定」処理は、未知語が属する分類語を決定する。分類語を含めたシソーラスの全ノード(2710語)との意味的な関連の強さを関連度計算によって算出する。算出された関連度がある設定値を超えていれば、意味的に関連が強いと判断し、関連付ける。

この設定値は、概念ベース全体の中から無作為に抽出した概念  $X_i$  に関する関連の度合いを、「女性-婦人」、「山-丘」などの極めて密接な関係、「山-川」、「夕焼け-赤い」などの密な関係、「山-机」、「電車-眼鏡」などの疎な関係の3種類に分類する教師データを概念ベースから人手で無作為に抽出し、4人が4人共にその関係が正しいと判断したも

のを 200 セット (計 600 データ) 使用して得られた。それらの関連度の平均を求め、これを関連の強さの大まかな判断に利用する。

ある概念 X と極めて密接な関係、密な関係、疎な関係にある語の関連度の平均は、0.47, 0.16, 0.01 であり、各関係の間には関連度に十分有意な差が見られる。そこで、ある未知語 とシソーラスの全ノードとの関連度の内、極めて密接な関係と密な関係の中間値である関連度平均 0.32 を超えるものを、意味的に関連が強いと判断して対応付ける。

対応付けられたノードの中で最高関連度の語が分類語であれば、その語は場所語であると決定し、それ以外ならば場所語でないとする。

#### 4.2 場所語であるかの評価

ここで、未知語 の分類語決定処理により、場所であるかどうか決定した結果を踏まえて、ある名詞が場所語であるかの評価を行う。実験データとして用いた単語は、日常的に使用している語をランダムに 300 語選出したものである。

これらに対して、まず人の判断で場所であるか・場所でないかを決定する。決定方法を表 4 に示す。4 人に各語を場所であるか・場所でないか判断してもらい、3 人以上が場所であると判断すれば、その語は「場所である」とし、逆に 3 人以上が場所でないか判断すれば、その語は「場所でない」とする。また、場所であると判断する人と場所でないか判断する人が同じ数であれば、その語は「人それぞれ」とする。

表 4 人の場所語であるか・ないかの決定法

完全未知語	人 A	人 B	人 C	人 D	人の判断
シアター					場所である
お皿		x	x	x	場所でない
運転席		x		x	人それぞれ

このようにして人の判断が決定され、次に実際のシステムの評価を行う。但し、人の判断が「人それぞれ」の場合は実験データから除外した(296 語)。また、このときの代表語と未知語の個数はそれぞれ 98 語と 198 語であった。評価方法は、システムが人の判断と同じ判断をすると「正解 (○)」、人の判断と違った判断をすると「不正解 (×)」という 2 段階とし、未知語 198 語の評価を図 5 に示す。

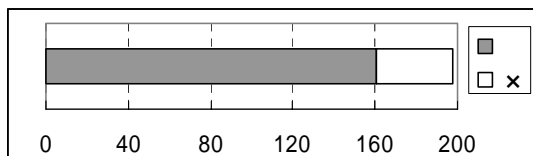


図 5 未知語についての場所であるかの評価

未知語における正解・不正解は、それぞれ 81.30%、18.60% となり、全体の正解・不正解は、それぞれ 87.50%、12.50% となった。この結果は充分であるといえ、本稿で定義した「場所語」が我々人間の常識的に思う場所と十分な確率で一致したといえる。

#### 5. 未知語の主体語・目的語の想起

代表語と分類語に属する未知語、分類語決定によって場所語であるとされた未知語における主体語・目的語の想起手法を述べる。

代表語に関しては、場所語知識ベースを利用することによって、主体語・目的語を出力することができる。しかし、未知語に関しては、シソーラスや分類語決定処理を行うことで分類語が特定され、分類語の主体語・目的語が取得できるだけで、その未知語固有の主体語・目的語が取得できない。このような問題を解決するために、本章と 6 章において様々な場所判断特有の未知語処理手法を提案する。

##### 5.1 単一代替処理

###### 5.1.1 処理方法

未知語が属する分類語から取得された主体語・目的語以外にも、更に未知語固有の主体語・目的語を取得する必要がある。

そこで、未知語が属する分類語に属する代表語と関連度計算を行い、最も関連性が高い語に代替しその代表語の主体語・目的語を未知語の主体語・目的語として更に追加する。この処理を「単一代替処理」と呼ぶ(図 6)。

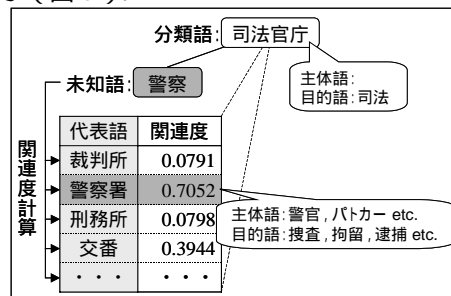


図 6 単一代替処理の例

###### 5.1.2 単一代替処理の精度評価

この処理によって得られた未知語の主体語・目的語の評価を行う。入力された語に対して、意味的に関連が強い主体語・目的語が正しく出力されると「常識的(正解)[○]」、誤ったものを出力すると「非常識(不正解)[×]」とする。また、意味的に関連が強くはないが出力してもおかしくないものは「非常識ではない[□]」とする。

例えば、「警察」の場合、「警官・捜査」などは常識的、「医者・キャンプ」などは非常識、「教官・教える」などは非常識でない解である。この評価は人手で行う。

実験データとして用いた単語は、未知語かつ、日常的に使用している語をランダムに 150 語選出したものであり、出力された主体語・目的語に対して 3 段階評価を行う。結果を図 7 に示す。図 7 に示している は「分類語から得られた主体語・目的語」であり、 は「代替された語から得られた主体語・目的語」である。

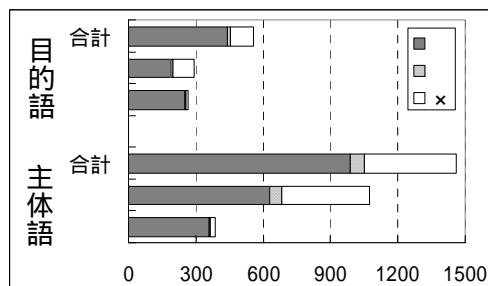


図7 単一代替処理の精度結果

出力された主体語・目的語の数はそれぞれ合計1460語・558語であり、そのうち常識的な解・非常識でない解・非常識な解は、主体語では67.60%、4.32%、28.08%、目的語では78.67%、2.69%、18.64%のようになった。

その内訳を見ると、得られた出力された主体語・目的語の数はそれぞれ合計386語・265語であり、そのうち常識的な解・非常識でない解・非常識な解は、主体語では92.75%、2.07%、5.18%、目的語では94.72%、1.89%、3.40%のようになり非常に高い確率で常識的な解が出力されているといえる。

一方、得られる出力された主体語・目的語の数はそれぞれ合計1074語・293語であり、そのうち常識的な解・非常識でない解・非常識な解は、主体語では58.57%、5.12%、36.31%、目的語では64.16%、3.41%、32.42%のようになり、非常に正答率が悪いといえる。

全体的には、常識的な解と非常識でない解を合わせると、主体語は71.92%、目的語は81.36%の正答率があり、すでにある程度の精度で未知語からも主体語・目的語を想起することが可能といえるが、より正答率を上げるため、によって得られる主体語・目的語の改善を行い、非常識な主体語・目的語を除去することを試みる。

## 5.2 閾値打ち切り処理

### 5.2.1 処理方法

まず5.1節で述べた処理の問題点について考える。例えば、未知語「小学校」が入力されると、最も関連性が高い代表語として「学校」に代替され、「小学校」の主体語・目的語には「中学生・教授・研究etc.」といった語を出力してしまう。

つまり、入力語と代替された代表語とは、あくまで関連性が高い語であって、同義・類義であるとは言えない。そのため、その代表語の主体語・目的語の全てを入力語の主体語・目的語として追加したのでは、不適切な単語が多々含まれる。

そこで今回は、代替された語から取得された主体語・目的語と未知語について関連度計算を行い、算出された関連度がある閾値を超えていれば、意味的に関連があるとして出力し、逆に閾値を超えなければ、不適切な単語として除去する。

この閾値は、場所語知識ベースの各代表語 443

語とその各主体語・目的語とで関連度計算を行い求められた、主体語・目的語のそれぞれの平均値であり、それぞれの平均値は0.18、0.13となる(図8)。

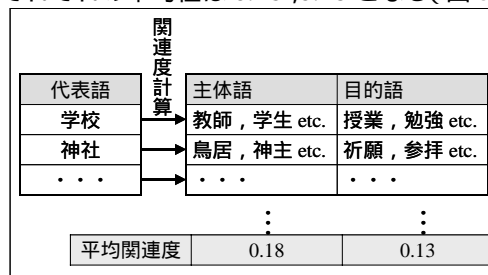


図8 主体語・目的語の平均関連度

この閾値以上の主体語・目的語だけを未知語の主体語・目的語として出力する。この処理を「閾値打ち切り処理」と呼ぶ。

### 5.2.1 閾値打ち切り処理の精度評価

この「閾値打ち切り処理」によって得られた主体語・目的語、および得られた主体語・目的語全体について評価する。

先程と同じように、入力された語に対して、出力される主体語・目的語が「常識的(正解)[ ]」、「非常識(不正解)[ × ]」、「非常識ではない[ ]」であるかを調査する。

実験データとして用いた単語は、5.1.2節で用いた語と同じであり、取得された主体語・目的語に対して3段階評価を行う。結果を図9に示す。ここで示すは5.1.2節と同じ「分類語から得られた主体語・目的語」であり、'は「閾値打ち切り処理より得られた主体語・目的語」である。

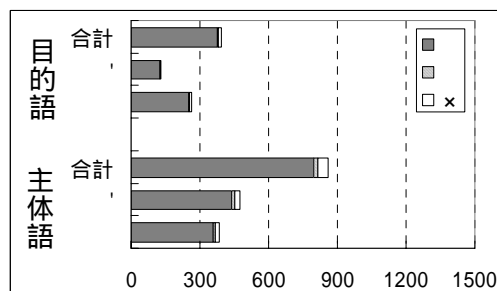


図9 閾値打ち切り処理の精度結果

今回の閾値打ち切り処理によって得られた'について見ると、主体語・目的語の数はそれぞれ合計474語・130語であり、そのうち常識的な解・非常識でない解・非常識な解は、主体語で92.83%、2.11%、5.06%、目的語では95.38%、0.77%、3.85%のようになり、単一代替処理で得られたに比べて格段に精度が高くなったといえる。

全体的には、出力された主体語・目的語の数はそれぞれ合計860語・395語であり、そのうち常識的な解・非常識でない解・非常識な解は、主体語では92.79%、2.09%、5.12%、目的語では94.94%、1.52%、3.54%のようになり、常識的な解と非常識でない解

を合わせると、主体語は 94.88%，目的語は 96.46% の正答率があり、精度的には何ら問題がないといえる。

しかし、閾値打ち切り処理によって単一代替処理によって出力されていた の正しい主体語・目的語も幾つか除去されてしまうため、全体的にも主体語・目的語の想起数が多少減少してしまった。そこで、精度を保ちつつ、さらに常識的な主体語・目的語を想起することを試みる。

## 6. 未知語の主体語・目的語の想起拡張

これまで実験データとして用いている未知語 150 語について、我々人間ならどのくらいの主体語・目的語を想起できるのか調査した結果、主体語・目的語それぞれ 1709 語、863 語想起した(図 10)。1つの未知語について見ると、それぞれ 11.39 語 5.75 語を想起するという結果が得られた。

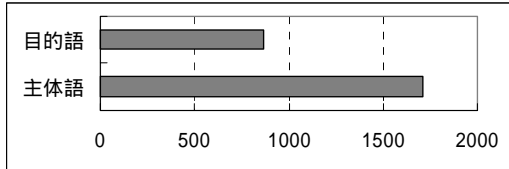


図 10 人間が想起する主体語・目的語の総数

ここで「人間が想起する主体語・目的語の総数」を  $T$  とし、「閾値打ち切り処理で得られた常識的な解と非常識でない解の主体語・目的語の総数」を  $S$  とし、 $S/T$  の出力個数を見る(図 11)。

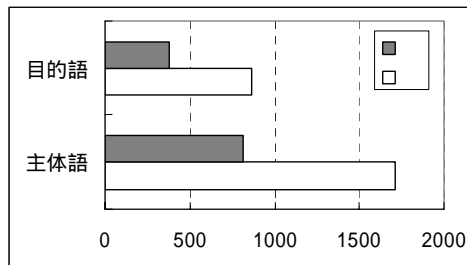


図 11 の出力個数

その結果、主体語・目的語のそれぞれにおいて、 $S/T$  を  $T$  で割った値、つまり人間が想起する個数に対してシステムが想起する個数の比率(以下、想起率と呼ぶ)は 47.75%、44.15%となり、さらに常識的な主体語・目的語を想起することが必要であることがわかる。

### 6.1 複数代替閾値打ち切り処理

#### 6.1.1 処理方法

更なる常識的な主体語・目的語の想起を目指し、これまで未知語の代替語を 1 つに限定していた処理(単一代替処理)を、新たに意味的に関連が強いとされる代表語を全て取得し、それを代替語とし全ての代替語の主体語・目的語について、5.3 節の「閾値打ち切り処理」により、常識的な主体語・目的語だ

けを想起する。

意味的に関連が強いとされる代表語の選定の仕方は、4.1 節でも述べた、ある概念 X と極めて密接な関係、密な関係の中間値である関連度平均 0.32 を超えるものを、意味的に関連が強いと判断して代替語とする。

この処理を複数代替閾値打ち切り処理と呼び、この処理を施すことで、例えば、これまでは未知語「農地」の代替語が「農場」だけであったのに対して、新たに「農園」を代替語に、そして、「耕地・牧場」を分類語にすることができ、新たな常識的な主体語・目的語である「果樹園・菜園・茶畑 etc.」、「耕作・摘む etc.」などが出力できた。

#### 6.1.2 複数代替閾値打ち切り処理の精度評価

この「複数代替処理」によって得られた主体語・目的語について評価する。

先程と同じように、入力された語に対して、出力される主体語・目的語が「常識的(正解)[○]」、「非常識(不正解)[×]」、「非常識ではない[□]」であるかを調査する。

実験データとして用いた単語は、5.1.2 節で用いた語と同じであり、取得された主体語・目的語に対して 3 段階評価を行う。結果を図 12 に示す。ここで示す「○」は「複数代替閾値打ち切り処理により分類語から得られた主体語・目的語」であり、「□」は「複数代替閾値打ち切り処理により代替語から得られた主体語・目的語」である。

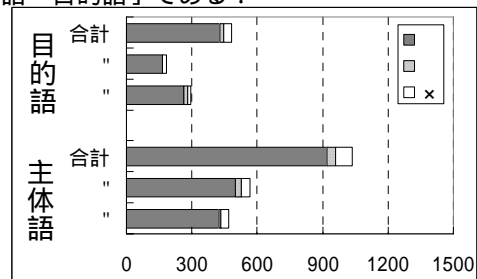


図 12 複数代替閾値打ち切り処理による精度結果

また、出力個数を図 13 に示す。ここで示す「○」は、図 11 で示したもので、「□」は今回の「複数代替語処理で得られた常識的な解と非常識でない解の主体語・目的語の総数」である。

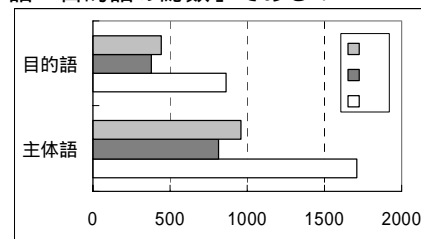


図 13 と の出力個数

出力された主体語・目的語の数はそれぞれ合計

1034 語・480 語であり、そのうち常識的な解・非常識でない解・非常識な解は、主体語では 89.07%、3.58%、7.35%、目的語では 88.96%、4.17%、6.87% のようになった。常識的な解と非常識でない解を合わせると、主体語は 92.65%、目的語は 93.13% の正答率がある。また、想起率も主体語・目的語それぞれ 56.06%、51.80% と向上した。

以上のことから、正答率をほとんど下げることなく、想起率を上げることができたといえる。しかし、まだ想起率をあげる余地があり、さらに常識的な主体語・目的語を想起することを試みる。

## 6.2 属性参照処理

ここでは、複数代替処理で取得されている主体語・目的語と概念ベースを参照することで得られる未知語の属性を利用して、常識的な主体語・目的語を想起することを試みる。

### 6.2.1 処理方法

複数代替処理で取得されている主体語・目的語と未知語の属性を利用して、常識的な主体語・目的語を想起する処理を 2 通り行う。

1 つ目の処理について述べる。まず、これまでに取得されている全ての主体語・目的語についてソーラスによってノードを特定する。次に、未知語の属性を参照し、それらと同じノードに属する属性があるかどうかを探る。そして、同じノードに属する属性があれば、5.2.1 節で述べた閾値以上（主体語の場合は関連度 0.18、目的語の場合は 0.13 以上）を示した語を常識的な主体語・目的語とみなし、新たに取得する。

2 つ目の処理は、これまでに取得されている主体語・目的語と同義となる属性を取得し、常識的な主体語・目的語を拡張しようとする処理である。同義となる属性の判断方法は、各属性とこれまでの主体語・目的語の全ての組合せとを関連度計算をし、関連度 0.62 以上の語が存在するかどうかで決定する。この 0.62 という値は、同義語辞書<sup>5)</sup>に存在する、例えば、「今 - 現在」、「医師 - 医者」、「学習 - 勉強」などの同義となる組を無作為に 1800 組抽出し、それぞれに関連度計算を行った平均値であり、その値以上を示すと、同義とみなしてもよいとされた値である。

このようにして同義とみなされた属性を新たな主体語・目的語として取得する。

これまでに述べた「複数代替閾値切り処理」と、この処理を利用した「属性参照処理」を組み合わせると主体語・目的語を想起する処理を「組合せ想起処理」と呼ぶ。

### 6.3 組合せ想起処理の精度評価

組合せ想起処理によって得られた主体語・目的語、および、得られた主体語・目的語全体について評価する。

先程と同じように、入力された語に対して、出力される主体語・目的語が「常識的（正解）[ ]」、「非常識（不正解）[ × ]」、「非常識ではない[ ]」であるかを調査する。

実験データとして用いた単語は、5.1.2 節で用いた語と同じであり、取得された主体語・目的語に対して 3 段階評価を行う。結果を図 1 4 に示す。ここで示す ” と ” はそれぞれ 6.1 節で示した複数代替処理によって分類語、代替語から得られた主体語・目的語であり、” は属性参照処理によって得られた主体語・目的語である。また、出力個数を図 1 5 に示す。ここで示す は、図 1 1 で示したもので、

は今回の「組合せ想起処理で得られた常識的な解と非常識でない解の主体語・目的語の総数」である。

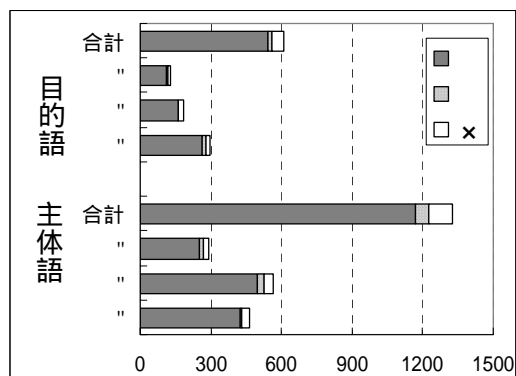


図 1 4 組合せ想起処理の精度結果

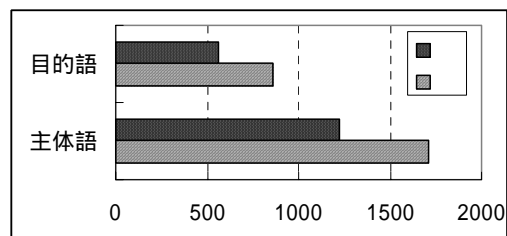


図 1 5 の出力個数

属性参照処理によって新たに出力された主体語・目的語の数はそれぞれ 233 語・103 語であり、そのうち常識的な解・非常識でない解・非常識な解は、主体語では 85.62%、6.16%、8.22%、目的語では 88.37%、0.78%、10.85% のようになり、常識的な解と非常識でない解を合わせると、主体語は 91.78%、目的語は 89.15% の正答率があり、新たに常識的な主体語・目的語が取得できたといえる。

また、組合せ想起処理全体においても、常識的な解と非常識でない解を合わせると、主体語は 92.46%、目的語は 92.28% の正答率があり、想起率も主体語・目的語それぞれ 71.74%、65.12% となり、組合せ想起処理によって、正答率、想起率ともに十分な結果が得られた。

## 7. システム全体の評価

日常的に使用する場所語として、これまで使用し

てきた未知語 150 語とランダムに選出した代表語 150 語の計 300 語を実験データに用いて、想起する主体語・目的語の正答率と出力個数の評価を行う（図 1 6）。ここで示す A は実験データで、我々人間が想起した主体語・目的語の総数であり、B はシステムが想起した主体語・目的語の総数である。

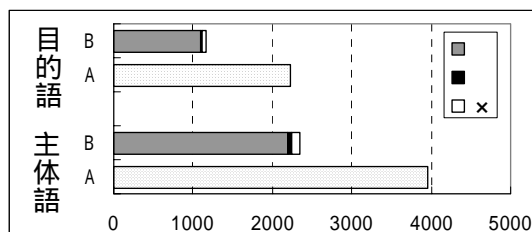


図 1 6 システム全体の正答率と出力個数

常識的な解と非常識でない解を合わせると、主体語は 95.73%、目的語は 95.97% の正答率があり、正答率に関しては全く問題ない。しかし、想起率を見ると、主体語・目的語それぞれ 59.28%、52.24% となり、未知語のときに比べ、相当減少してしまっている。これは、代表語のときに主体語・目的語の想起数が少ないために起こっていることは明らかである。

そこで、代表語に対しても、未知語に対して行った組合せ想起処理を施す。その結果を図 1 7 に示す。

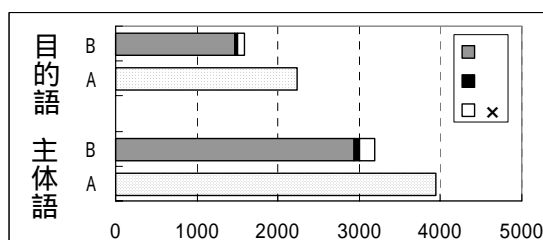


図 1 7 代表語にも組合せ想起処理を施したシステム全体の正答率と出力個数

代表語に対しても、組合せ想起処理を施すことで、想起率が、主体語・目的語それぞれ 76.10%、67.40% となり、十分な想起ができたといえる。また、正答率についても、常識的な解と非常識でない解を合わせると、主体語は 94.08%、目的語は 94.59% の正答率があり十分な精度が得られたといえる。

## 8. おわりに

本論文では、日常的な自然言語による会話中に現れる単語が場所に関する語であるか判断し、もし場所に関する語なら、そこには何が存在し、何をやる場所なのかを連想する「場所判断システム」を構築し、更に、電子化国語辞書などから作成した概念ベースや関連度計算により意味的な関連の強さを考慮して、未知語として入力された語を、場所語知識ベースに登録された語に対応付ける未知語処理方式を提案し、その有効性を実験により検証した。

提案した様々な主体語・目的語の想起拡張処理によって、精度が高く、想起率の高い主体語・目的語が取得できる。これによって、常識を踏まえて物事を扱うことができる「常識的判断システム」の構築に貢献でき、会話の意味・意図を理解できる人間にやさしいコンピュータの実現につながるであろう。

## 謝辞

本研究は文部科学省からの補助を受けた同志社大学の学術フロンティア研究プロジェクトにおける研究の一環として行ったものである。

## 参考文献

- 1) 広瀬 幹規, 渡部 広一, 河岡 司, “概念間ルールと属性としての出現頻度を考慮した概念ベースの自動精練手法”, 信学技報, NLC2001-93, pp.109-116, 2002.
- 2) NTT コミュニケーション科学研究所監修, 「日本語語彙体系」, (岩波書店, 東京, 1997)
- 3) 渡部 広一, 河岡 司, “常識的判断のための概念間の関連度評価モデル”, 自然言語処理, Vol. 8, No. 2, pp. 39-54, 2001.
- 4) 小島 一秀, 渡部 広一, 河岡 司, “常識判断のための概念ベース構成法: 概念間論理関係を用いた概念属性の重み決定法”, 信学技報, AI2000-80/KBSE88(2001-3), pp.57-64, 2001.
- 5) 杉本 二郎, 渡部 広一, 河岡 司, “概念ベースを用いた場所に関する連想メカニズムの構築”, 情報処理学会第 64 回全国大会講演論文集, 5M-3, 2002.