自 然 言 語 処 理 99－5
（1994. 1. 20）

数値データ処理のための自然言語インタフェース機能

宮部 隆夫　山口 智治　芝 温子　谷 幹也　市山 俊治

NEC 関西C&C研究所

金額や個数などの数量及び期日や期間等の時間に関する数値項目を含むデータベースを，自然言語インタフェースを介して利用する際に必要な機能について述べる。自然言語の文に現れる漠然とした大小比較表現や既知情報を省略した表現を対象データ中の数値項目と直接対応付けるのは困難である。ここでは，数値項目との対応付けが課題となる主要な言語表現として (1)省略を含む比較 (2)省略を含む統計表現 (3)数量相対表現 の３種表現を選択しその内容を分析する。各表現に対し数量や時間に関する問題事例を提示しつつ，その解決手法について述べる。

A Method of Natural Language Interface for Processing of Numerical Data

Takao Miyabe　Tomoharu Yamaguchi　Haruko Shiba
Mikiya Tani　Shunji Ichiyama

Kansai C&C Research Laboratory
NEC Corporation

This paper describes a method of natural language interface for numerical data, such as various accounts, dates and periods. To use numerical data base, expressions of input sentences should have concrete numerical informations. The following 3 expressions are typical ones as appear in input sentences and which lack the concrete numerical information. (1)Comparative form with ellipsis (2)Statistic form with ellipsis (3)Relative numerical expression. The analyses and solutions are shown to cover the information gap and to use these as input.

1　はじめに

情報を整理し格納するデータベースシステム等の情報処理システムでは，客観性のあるデータを管理しやすい形態で保持するという視点から，情報を数量化して数値データとして記憶する場合が多い。この情報を利用するために，特定属性について数値の指定や範囲の限定を行うとともに，属性値を四則演算や統計演算によって加工したり属性値同士を比較したりして，情報を特定する。このような実運用の情報検索を対象としたシステムのインタフェースには，強力な数値表現処理機能が必須である。本稿では省略の補完処理を中心に比較表現と統計表現の処理手法について説明する。また，相対表現に関する実体化処理手法も提示する。

我々は複数分野での適応性を考慮した自然言語インタフェースキットＩＦ－Ｋｉｔ[谷 他92][宮部 92]を開発してきた。図１に示すように，データベースから獲得した対象分野知識を利用して，日本語の入力文を解析し更に文脈処理を施した結果を使って意味解釈する。その解釈結果からデータベース検索言語ＳＱＬの検索コマンドを生成する。なお，ここで利用している対象分野知識は，ツール（知識ベース作成支援ツール）[谷 他93]によりデータベースから半自動的に構築したものである。

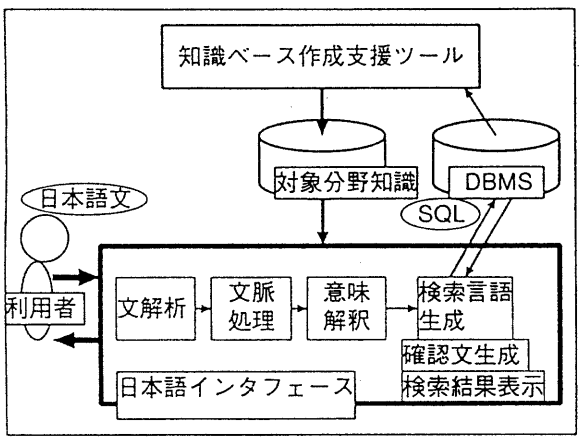

図１　IF-Kitの構成

図２には，このＩＦ－Ｋｉｔから日本語を用いてデータベース中の数値項目を検索する例を示す。

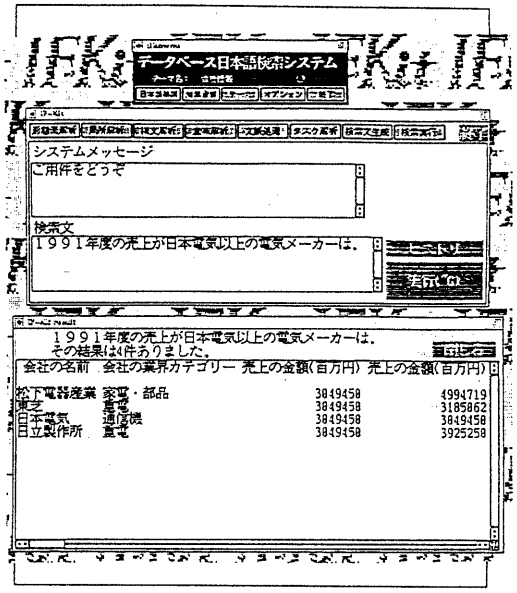

図２　ＩＦ－Ｋｉｔ　利用例

以下，２章では自然言語インタフェースにおける数量表現処理の特徴と問題点について述べ，問題の分類と代表例を提示する。３章では２章の分類と代表例をもとに，数量表現処理の枠組みを示す。４章では３章の枠組み中で対処困難な現象について，事例を列挙しながら指摘しつつ解消するための方式について説明する。

2　自然言語インタフェースの処理の特徴

自然言語インタフェースの特徴として，解釈目標の意味表現が対象分野に強く依存している点がある。前章で記したような数値項目を有するデータベースの場合には，助数詞や数量詞移動[井上 編89]の問題の他に，省略や相対表現に対する情報補完や変換による言語表現と対象知識中の項目との対応付け問題がある。自然言語インタフェースで利用される数量表現の中で，特に数値項目条件を含む問い合わせ文中に現れるものを，省略と相対という観点で３種類の表現として提示する。

(1) 省略を含む比較表現
(2) 省略を含む統計表現
(3) 数量相対表現

2．1 省略を含む比較表現

比較表現は，「比較関係」「比較対象」「被比較対象」「比較差分」の４項の関係ととらえることができる。「比較関係」は，大小関係を記述する関係表現（例．多い）に相当する。「比較対象」は，較べる相手（例．９１年度の売上が）であり，「被比較対象」は，較べられる相手（例．９０年度の売上より）のことである。また，「比較差分」は，比較対象と被比較対象の差の値（例．１兆円以上）である。これを定式化すると，各項目はそれぞれ複数の属性から成り立ち，全体として図３の構造として記述できる。

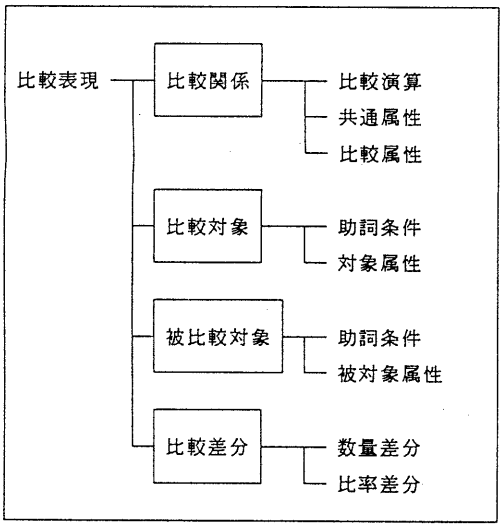

図３　比較表現の構造

各項目は，以下の内容である。

比較演算：比較演算子の値，大小関係(図７参照)
共通属性：比較対象と被比較対象に共通する性質
比較属性：比較の際に基準にする属性

(比較)助詞条件：　比較項目に現れる助詞の条件
「が」，「方が」，・・
対象属性　　：比較対象の性質

(被比較)助詞条件：被比較項目の助詞の条件
「より」，「に比べて」，・・
被対象属性　　：被比較対象の性質

数量差分：比較属性に関する比較・被比較対象間の数値量の差の値
比率差分：比較属性に関する比較・被比較対象間の数量差分の被比較対象に対する比率

次に比較表現の例を例１に示す。

（例１）　比較表現例
(1-1) ９１年度の売上が９０年度の売上より１兆円以上多い会社を教えて。
(1-2) ９１年度の売上が９０年度より多い会社は？
(1-3) ９０年度の売上より９１年度の方が大きくなった会社は？
(1-4) Ｎ社より多い会社は？
(1-5) 売上が９０年度の２倍くらいの会社は？
(1-6) 売上が９０年度より１０％以上多い会社は？
(1-7) ９１年度について，Ｎ社より売上が伸びた会社は？

(1-1)は４項成分をもつ標準例で，先の説明例中で利用した。

(1-2)-(1-4)は省略例である。それぞれ(1-2)は「被比較項目」から『売上』が，(1-3)は「比較項目」から『売上』が省略されている。(1-4)では，「比較属性」の『売上』が省略されている。

(1-5)(1-6)は比率差分の例である。(1-5)は『２倍くらい（－１倍）』が，(1-6)は『１０％以上』が「比較

差分」の値である。

(1-7) は『売上の伸び』という派生概念を「比較属性」として利用する例である。

２．２　省略を含む統計表現

「最大」「平均」等の統計表現については，「統計関係」「統計対象」「統計範囲」の３項関係ととらえる。

「統計関係」は，最大や平均等の統計関係（例．売上が最大）である。「統計対象」は，統計処理をする相手（例．メーカー）である。「統計範囲」は統計処理をする領域・範囲（例．電気業界）のことである。統計表現を比較表現と同様に定式化して図４に示す。

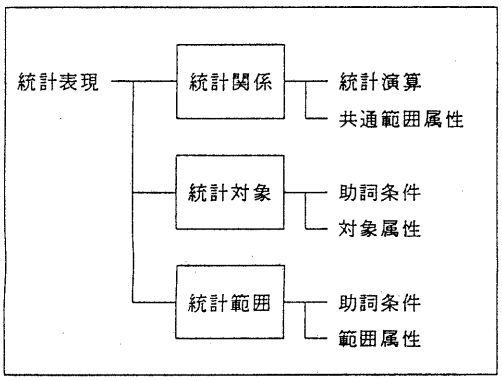

図４　統計表現の構造

以下が項目の説明である。

統計演算：統計演算子の値，最大等(図７参照) 共通範囲属性：対象の属性中で，統計の範囲に関係する性質
（対象）助詞条件：対象項目に現れる助詞の条件「が」，・・ 対象属性：　統計対象の性質
（範囲）助詞条件：範囲項目に現れる助詞の条件「で」，・・ 範囲属性：　統計範囲の性質

次に例を示す。

（例２）　統計表現例
(2-1) 電気業界で９１年度の売上が最大のメーカーを教えて。
(2-2) 売上が最大の電気会社は。
(2-3) ９１年度の売上が業界平均より１兆円以上多いメーカーは。

(2-1)は３項目をもつ標準例であり，先の説明でも利用した。

(2-2)は，「範囲属性」である『電気業界』を省略した例である。この場合は「共通範囲属性」の『電気（会社）』が効力を持つ。

(2-3)は　２．１節の比較表現との複合例である。ここでは，範囲は『業界』全てになり複数になる。

２．３　数量相対表現

数量相対表現は，時刻や順序などの順序関係を有する表現に付随して現れる。文脈に依存して定まる基準値が確定した後，はじめて実体のある項目データと対応可能となる。（例３参照）

(3-1)のように，処理時点の時刻に依存して決定される場合と，(3-2)(3-3)の様に前文脈に依存して決定する場合とがある。

（例３）　数量相対表現例
(3-1) 一昨年のＮ社の売上は。
(3-2) その前年の売上は。
(3-3) 売上が１０番目の会社は。その次は。

３　数量表現処理の枠組み

この章では，自然言語インタフェースでの数量表現処理に必要な知識と処理過程について説明する。前章では数量表現の特徴について(1)-(3)の３項目に分けて説明した。本章もこの分類項目に従い，各節毎に目標となる意味表現及び処理に利用する知識を提示し，次

いでこの知識を利用して言語表現から目標意味表現への解釈手順を示す。省略補完や相対表現の実体化処理は，構文解析結果と目標意味表現の中間のフレーム表現上で行う。ここでは，比較や統計や相対表現毎にその形式的な意味を反映したフレームで記述するので，省略箇所の指定とその後の補完情報の推定処理が形式的に平易に提示可能となる。

３．１　比較・統計表現に関する知識と処理

(1) 目標意味表現

目標となる意味表現は，ＳＱＬ記述文と同等の情報を有する。利用している記号「：」は左辺の「対象」「条件」項目と右辺の式とを対応づける演算子である。「．」は図５に示すモデル（会社のモデル）での上／下位関係を現し，上位概念の指定の下で存在する下位概念を規定する。「＝」などの関係式や「＋」などの演算は，数値に関する等号や四則演算に相当する。

同じ桁の記述は AND条件 であり，右へずらした「条件」は，その上の条件の前提条件となる。

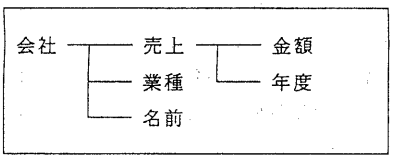

図５　会社のモデル

なお，同じ属性を別の観点から利用するときは，属性にインデクス（１，２，・・）を付与する。

「９１年度の売上が１兆円以上の会社は？」という文に対する意味表現は以下のようになる。

```
対象：会社．名前
条件：会社．売上．金額　≧　１兆円
  条件：売上．年度　　＝　９１年度
```

例１について意味表現を図６に記載する。(1-2)(1-3)の意味表現は一致する。「売上１」は「９１年度の売上」に相当し，「売上２」は「９０年度の売上」に対応する。「会社２の名前」は「Ｎ社」である。

```
(1-1)
  対象：会社．名前
  条件：会社．売上１．金額
            ≧会社．売上２．金額＋１兆円
    条件：売上１．年度　＝９１年度
          売上２．年度　＝９０年度
(1-2)(1-3)
  条件：会社．売上１．金額
            ≧　会社．売上２．金額
    条件：売上１．年度　＝９１年度
          売上２．年度　＝９０年度
(1-4)
  条件：会社１．売上．金額
              ≧　会社２．売上．金額
        会社２．名前　＝　Ｎ社
    条件：売上．年度　＝９１年度
(1-5)
  条件：会社．売上１．金額
              ≒　会社．売上２．金額＊２
    条件：売上１．年度　＝９１年度
          売上２．年度　＝９０年度
(1-6)
  条件：会社．売上１≧　会社．売上２
                    ＋会社．売上２＊０．１
    条件：売上１．年度　＝９１年度
          売上２．年度　＝９０年度
(1-7)
  条件：会社１．売上の伸び．金額
              ≧　会社２．売上の伸び．金額
    条件：売上の伸び．金額
              ＝　売上１．金額－売上２．金額
      条件：売上１．年度　＝９１年度
            売上２．年度　＝９０年度
```

図６　目標意味表現（例１）

算術演算	＋，－，×，÷
比較演算	＝，≠，≒，≧，≦，＞，＜
統計演算	avg 平均　count 数　sum 合計 max 最大　min　最小
包含関係	in　含む

図７　意味記号（比較・統計で利用）

また，算術演算や比較演算は通常の記号と同義で用いている。後述する統計演算の記号も含めて図７に一括記載する。これらの記号は原則としてＳＱＬ中の各種演算記号に相当する。

(2) 知識と処理

比較処理に用いるフレーム表現を示す。２章の２．１節でも示したように，比較表現は特定の比較軸に従って対象と被対象とを量的に比べる表現であり，同時に差や比についても表現可能である。図３の構造の知識を利用した枠組みとしてフレーム表現を提示する。

```
「比較関係」比較演算：共通属性：比較属性：
「比較対象」助詞条件：明示部：省略部：
「被比較対象」助詞条件：明示部：省略部：
「比較差分」数量差分：比率差分
```

図３の構造との差の部分を中心に以下に説明する。

「比較対象」　図３の対象属性は，明示部と省略部とに分かれる。明示部は入力表現から定まる。省略部については，「被比較対象」の明示部と「共通属性」とから，その内容を推定する。

[被比較対象]　比較対象に準ずる。

次に処理過程を説明する。例１(1-1)(1-2)を参考にして以下を入力文として説明する。

```
入力文：９１年度の売上が９０年度より
　　１兆円多い会社を教えて。
```

この構文解析結果は，

```
教える {obj 会社{mod
  ge{obj 売上
   {mod ９１年度{ako 時};scm が;ako 金};
     bas ９０年度{scm より; ako 時刻} ;
     val １兆円{ako 金}}}}
```

となる。なお，格ラベル obj,bas,val,mod,... ako: 意味属性　scm: 助詞　とする[谷 他92]。

ここで，次に示す "ge(以上)" と比較表現との対応規則を用いて，

```
     ge              <--> 比較関係「≧」
case:obj 助詞:が     <--> 比較対象
     bas      より  <--> 被比較対象
     val             <--> 比較差分
```

以下のフレーム表現を作成する。

```
「比較関係」　比較演算：≧
「比較対象」　明示部：９１年度の売上
「被比較対象」明示部：９０年度
「比較差分」　数量差分：１兆円
```

ここで，省略の判断と省略箇所の補完を行う

1) 省略の判断　「比較対象」「被比較対象」「比較差分」及び「比較属性」について，その属性の一致（包含関係）を検証する。不一致時（省略可能性）に，「対象」間の構造比較と「差分」や「属性」に対する対象の概念関係の検証（図５参照）を行う。結果として「対象」中の省略箇所や「比較属性」の省略などが判明する。

　例では，「比較属性」と「被比較対象」に「年度」情報が省略されている。

2)省略補完　省略があると判定した時点で，４項目の一致（包含）内容が「比較属性」として補完される。この属性に対して対象が情報不足の時に，概念関係に準拠して情報を補完する。

　例では「比較属性」が「売上」になり，また「被比較対象」の省略部に「売上」が補完される。

「比較関係」　比較演算：≧
　　比較属性：売上（金）
「比較対象」　明示部：９１年度の売上
「被比較対象」明示部：９０年度
　　省略部：（９０年度の）売上
「比較差分」　数量差分：１兆円

比較構文の目標意味表現(条件部)は以下の通りである。

条件:比較対象 比較演算 被比較対象+比較差分
　条件:比較対象の条件 & 被比較対象の条件

これに上記のフレーム表現中の情報を代入して

条件：売上1 ≧ 売上２＋１兆円
　条件:売上1.年度=91年度 & 売上2.年度=90年度

後は対象に関する一般的な整形規則により，目標意味表現が生成可能となる。

他の例に関しても同様である。なお，(1-5)の「比較演算」の『≒』については±１０％の誤差と解釈する。また(1-5)(1-6)の比率差分については，『比率×被比較対象』を計算して意味表現に反映する。(1-7)で「売上の伸び」は予め定義しておく。

３．２　統計表現に関する知識と処理

(1) 目標意味表現　例２の意味表現を図８に記す。

対象：会社.名前　---------　(2-1)(2-2)
条件：会社.売上.金額=max(会社１.売上.金額)
　条件：売上．年度　＝　９１年度
　　会社１．業種　＝　電気
　　会社１．業種　＝　会社．業種

条件：会社．売上．金額　----　(2-3)
　　≧avg(会社１．売上．金額)＋１兆円
　条件：会社１．業種　in　＊(全て)
　　会社１．業種　＝　会社．業種

図８　目標意味表現（例２）

(2) 知識と処理

統計表現のフレーム表現について図４と比較する。

「統計対象」比較表現と同様に，対象属性が明示部と省略部とに分かれる。

「統計範囲」同様に，明示部と省略部とに分かれる。

フレーム表現例として，例２(2-1)を示す。

「統計関係」演算：max 共通範囲属性：メーカー
「統計対象」明示部：９１年度の売上
「統計範囲」明示部：電気業界

(2-2)のような省略例については推論結果を例示する。

「統計関係」演算：max 共通範囲属性：電気会社
「統計対象」明示部：売上　省略部：９１年度
「統計範囲」省略部：電気（会社）

(2-3)は比較との複合表現であり，以下のようになる。

「比較関係」比較演算：≧　共通属性：メーカー
「比較対象」明示部：９１年度の売上
「被比較対象」明示部：　《
「比較差分」数量差分：１兆円

「統計表現」統計演算:avg 共通範囲属性:メーカ
「統計対象」省略部:９１年度の売上
「統計範囲」明示部:業界

処理手順は３．１同様なので，本節では割愛する。

３．３　数量相対表現

(1) 目標意味表現　数量相対表現の目標は，相対的な表現に対して基準値を与えて実体化することである。例３の(3-1)(3-2)の「一昨年」や「その前年」は，「今年」や「その年」を基準点として「１９？？年」と実体化できる。また(3-3)は「１０番目の次」として「１１番目となる。」

(2) 知識と処理　時刻に関する相対的表現の関係知識として，図９と図１０に示すような定義知識を有する。

図９の知識では，基準となる「今」に関する情報が

求まれば，対応して「年（，月，週，日，時刻，等）」が定まり，相対時刻の実体化が可能となる。

同様に，図１０のような知識は，前文脈の時刻が基準値となる。この前文脈の基準時刻を利用して，相対時刻の実体化が可能となる。

{昨年，去年} ＝ {今年，本年} －１ {来年} ＝ {今年，本年} ＋１　（月，日，..）

図９　時刻関係知識(1)

翌(N)　≡ 基準(N)＋１（N＝年，月，日，..）	
MN後 ≡ 基準(N)＋M	N＝年，カ月，日，.. M＝１，２，３，..
MN前 ≡ 基準(N)－M	

図１０　時刻関係知識(2)

更に，時刻以外の，「順序」などに関する一般的な相対表現[奥津74]に対しても図１１に示す相対表現に関する知識で実体化が可能である。

次，先(W) ≡ 基準＋W　　（W＝範囲幅）
前(W) ≡ 基準－（W）　　（W＝範囲幅）

図１１　相対表現関係知識

例えば，「１番から１０番まで１０個は」「その次は」では，「１１番から２０番までの１０個」となる。

４．課題

前章の知識と処理の枠組みにより，基本的な数値表現は処理可能と考える。しかしながら入力文の中にはこの枠に納まらない事例が存在する。

４．１　課題１　誤差範囲の問題

（例４） (4-1) １９９０年４月頃の売上と株価は？ ＝＞　１９９０年３月－５月の売上？ (4-2) 売上が１兆円くらいの会社は？ ＝＞　売上が９千億－１兆１千億円の会社は？

「頃」や「くらい」などに対する適切な誤差範囲指定の問題である。(4-2)の例では前述したような１０％をデフォールト誤差としているが，(4-1)の例のような離散数値に対しては適切でない。

［解決策］　「月」の様に１０％で小数になる離散値例や順序については±１を誤差基準とする。

４．２　課題２　範囲選択の問題

値を決定する範囲の決定に関する問題である。

（例５）　１９９０年４月の売上と株価は？ ＝＞１９９０年の売上？ ＝＞１９９０年の株価は？

１日単位に値が記述してある場合，例では「売上」は１ヶ月分の和であるが，「株価」は月平均をとるのが通常である。

［解決策］　売上のように加算値をとる対象と，株価のように変動する対象とを分別する情報を付与し，その値に従って数値決定範囲を変更する。

５　おわりに

自然言語インタフェースにおける数値処理について説明した。問題を (1)省略を含む比較表現 (2)省略を含む統計表現 (3)数量相対表現 の３種類の表現に分類し，それぞれについて必要な知識と処理方式について述べた。またこの枠外の問題についても言及した。本方式の中で(1)(2)の主要部分は既に実装済みである。今後は本方式を実用システム中で運用しながらその評価を進める予定である。

＜参考文献＞

［井上 編89］　井上和子編「日本文法小辞典」大修館

［奥津 74］　奥津敬一郎「生成日本文法論」

［谷 他92］　谷幹也，飯野香，山口智治，市山俊治「自然言語インタフェース構築キット：ＩＦ－Ｋｉｔ」信学研資　NLC91-62

［谷 他93］　谷幹也，市山俊治　「日本語による対象分野の知識獲得」情処　NL97-13

［宮部 92］　宮部隆夫　「日本語インタフェース文脈文法 - 解析手法 -」 45回情処全大2F-01