

## 日本語の名詞句における照応計算のモデル

BLIN Raoul  
C. R. L. A. O. -E. H. E. S. S.

54 Bvd Raspail  
75006 Paris, France

s001077@ipe.tsukuba.ac.jp

あらまし

本稿では、日本語の名詞句における、照応の計算の形式的システムを紹介する。その目的は、照応の計算は短期記憶からのデータを用いて行われるのであろうという、広く言語学者に受け入れられてきた推論を、形式的に実施してみることにある。これらのデータの性質や、このデータを計算に用いるための方法について議論する。しかし、そこには多くの問題点が残されており、それらの解決は容易ではないことがわかる。よってこの推論は、今後一層詳しく検討される必要があると思われる。

キーワード 自然言語処理 意味論 照応 日本語 知識ベース 名詞句

## Model for the Anaphora Calculus in Japanese Noun Phrase

BLIN Raoul  
C. R. L. A. O. E. H. E. S. S.

54 Bvd Raspail  
75006 Paris, France

s001077@ipe.tsukuba.ac.jp

Abstract

This paper will present a formal system of anaphora calculus in the Japanese noun group. Our aim is to test operatively the idea, largely accepted among linguists of a calculus done from recorded data from a short term memory. We shall discuss the nature of the data and the way envisaged to use them in order to do the calculus. As we shall see, there remain a large number of questions which can not be resolved simply and therefore find that the theory needs to be refined further.

key words

NLP, Semantics, Anaphora, Japanese, Knowledge Base, Noun Phrase

## 1. はじめに

テキストを理解するには、テキストの異なる部分間の関係を理解する必要がある。その一つに、名詞句間の関係の理解がある。つまり、対象の名詞句と先行文中の名詞句との関係や、対象の名詞句に関係する知識との関係を理解することである。日本語では、この名詞句間の関係は、「こ、そ、あ、」などの指示詞の場合で、数多くの研究がなされてきた。これらの研究では、情報を短期記憶から抽出するという結論が多く出されてきた。しかし、具体的にこの情報抽出をどのように行うかは詳細に検討されてこなかったように思われる。そこで本稿では、この短期記憶からの情報抽出という理論を評価するための形式的なモデルを提案する。これにあたり、F. Renaud (ルノー) (2) がフランス語での定冠詞の研究に用いた言語モデルを、日本語の名詞句に応用してみる。その形式的なモデルを提示する前に、まず、具体例の中で、言語学的現象を観察し、言語構造の中で、どのような情報が記憶され、照応プロセスにおいて利用されるのかを考察する。次に、形式的なモデルを紹介し、その機能を説明する。そして、短期記憶からの情報抽出理論を、このモデルの使用によってより具体的に実施し、この理論の利点や問題点を明らかにすることができる。

名詞句には、様々な言語学的構造があるが、ここではその全てを扱うことはできないため、次の名詞句をとりあげ、分析対象とする：対象の名詞句は具体的なものを示す名詞句（つまり名詞化された動詞を省く）で、数詞・助数詞により数えられるものである。また、「お母さん」のような関数名詞句を除く。分析対象の名詞句は、主語やテーマとして機能する。

## 2. 言語学的分析

テキスト・会話中の名詞句を解釈するには、次の三点を区別する必要がある：1) まず、名詞句がある物を初めて提示しているのか、2) あるいは、既に知られている物を再度示しているのかである。さらに、もし名詞句の指す物が既に知られている物であるなら、それは、2-a) テキスト・会話内で既に出てきたため知られているのか、2-b) または全く別の方法により知られているのか、を調べる必要がある。

また、本研究の分析対象は、数えることのできる名詞句であるため、集合の概念を用いて名詞句の意味的表象を解釈することは、有益であると思われる。この集合の概念により、名詞句のより詳しい解釈が可能となる。上に述べた1) 2-a) 2-b) に加え、分析対象の名詞句によって示される集合と、既出の集合との関係を、同一の集合であるか、または部分集合であるかに分類できる。ここでは、分析中の名詞句が、以上述べた場合のどれであるかを知るため、名詞句中の言語学的要素を分析する。

### 2-1. 照応対象の有無

ここでは「照応対象」は、「既に知られているもの」という意味で用い、分析対象の名詞句に先行するテキスト・会話文に見られるもののみならず、一般知識として既に知られているものも含む。

ここではまず、分析対象の名詞に照応対象があるかどうか、ある場合はその照応対象が先行の文章で現れたかどうかを観察する。また、この照応対象の有無を分析対象の名詞句A上でどのように理解することができるかを調べる。つまり、名詞句A上の言語的な構造により照応対象の有無の識別が可能かどうか、またはその意味により可能かどうかを調べる。

照応対象の存在を明確に理解させる名詞句の中には二つの構造がある。一つは無論、指示詞が前置している場合で、必ず先行の文章・会話に先行詞がある。今までに指示詞に関して多数の研究がなされたので、ここでは特に例をあげない。もう一つの名詞句構造は省略された名詞句である。省略されたからこそ、照応対象の存在が不可欠になる。

(1)-(3)は数量詞を核にした省略された名詞句で、(4)(5)は形容詞を核にした省略された名詞句の場合である：

- |                       |                     |
|-----------------------|---------------------|
| (1) 隣のおばさんが犬を3匹飼っていた。 | (4) あそこにチューリップがあるね。 |
| (2) 2匹がしょっちゅう病気になった。  | (5) うん、黄色いのがいっぱい。   |
| (3) 2匹が雌だった。          |                     |

ただし数量詞の場合は、あるテキスト・対話では注意すべき場合がある。この点に関しては後述する。

上の2つの構造では照応対象の有無が明確にわかるが、この他に、その有無がわからない場合もある。この場合、一般的知識やその名詞句と前後の文章構造との意味論的關係など、いわゆる文脈が重要になる。この例として、名詞のみからなる名詞句の場合を検討する。この場合、同一の構造に三つの異なる照応機能が見られる。以下に、その具体例を示しておく。

まず、(7)の場合、指示対象は先行の文章から抽出する。

(6) バスが時速30 kmで出発してから5分後に、同じ地点から自動車が  
時速40 kmで出発し、バスを追いかけた。

(7) 自動車がバスに追いつくのはいつですか。 (「中学校、数学1」、大日本図書、平成5年2月1日より)

ここでは、先行文のバスと自動車はそれぞれ1台しかないことが分かる。すると、(7)の「自動車」は(6)の「自動車」にしか照応しえない。(実際、元のテキストには「この自動車」とあるが、日本人のインフォーマントによると、「この」は余計であるとのことだった。)

次に(8)の場合は、指示対象が先行の文章にないので、知識ベースから抽出することになる。

Knowledge Base (家に2匹の犬がいる)

(8) 母：あら！テーブルの上にあったクッキーはどうしたの？

(9) 父：犬が食べたんだ。

また(10)の場合、指示対象は先行の文章で現れたが名詞句は初めて登場するものを示す可能性もある。

KB (この辺りは野犬がたくさんいる)

(10) 家の犬はポチとパチという。昨日この2匹が庭に出たきり  
いなくなってしまった。それで、あちらこちら探しまわったけれども、  
どうしても見つからず、あきらめた。

(11) 今朝、家を出ると1匹の犬が庭で寝ていた。

(11') 今朝、家を出ると1匹の犬が帰っていて外で待っていた。

(11)の犬は必ずしも家の犬の1匹ではない。よその犬である可能性がある。しかし、この文章の動詞で、その曖昧さをなくすことができる：(11')では家の犬であることがわかる。

つまり指示詞のない、数量詞+名詞からなる名詞句には、指示対象があるのかどうか不明なのである。また、先行詞があっても、それをどこで検索するのか(先行文中かあるいは知識ベースの中からか)も不明である。

また、上に示した省略した名詞句にも例外がある。特に前に文章はないが数詞量を核にした名詞句が現れる場合である：

(12) モスクワの地下鉄で爆発、4人死亡 (朝日新聞の記事の見出し)

このような場合は新しく登場する4人の集合は全く知られていないため、知識データベースにはそれに関する知識はないはずである。従って、規則として、知識ベースにも先行の文章にも先行詞がない場合、数量詞を核にした名詞句は新しい集合を表し、しかもその集合は、他の集合の部分集合ではない。

しかし、先行文または同文中に集合名詞句(collective noun)がある場合は注意が必要である：

(13) このクラスはそろったが、3人はだいぶ遅れてきた。

「クラス」が「生徒」の集合という集合名詞句として扱われる場合、「3人」の集合はその「クラス」の部分集合である。

## 2-2. 先行詞との関係

以下、「先行」という語は、分析中の名詞が現れる文章の直前文から5文章前くらいまでを指す。

ここでは、先行詞を確実に持った名詞句に限り、この名詞句が先行詞の示す集合と同一の集合であるのか部分集合であるのかを、名詞句の表層的な構造を分析することにより調べる。このとき、名詞句や先行詞の示す集合に関しての数量データが深く関わっている。

「その内」や「そのなか」が前置された名詞句は先行詞との部分集合の関係を持つ。

- (14) 今日、犬をいっぱい見かけた。
- (15) その中の何匹かは、首輪をしてなかった。

また、数量詞を核にした省略名詞句と先行詞の関係はそれぞれの数量で決まる。ただし、「とも」がつくと、いつも先行詞と同一の集合となる(18)。

- (16) 隣のおばさんが犬を3匹飼っていた。
- (17) 2匹が雌だった。
- (18) 3匹とも年をとってた。

ところが、「その中」が部分集合の関係を表すのに対して、「その」という指示詞だと、指示対象との関係が曖昧で場合によって変わる。

- (19) このあたりに野良犬が1匹いる。
- (20) この辺りに野良犬が2、3匹いる。
- (21) この辺りに野良犬がたくさんいる。
- (22) その野良犬がよく家の庭でねている。

この場合、(22)の”その野良犬”はその前文が(19)であれば、”1匹の野良犬”と同一の集合であることが明確にわかるが、前文が(20)、(21)の場合は、その部分集合であるか、同一集合であるかは曖昧である。また、(20)の場合、2、3匹が全て庭にねに来ることは十分あり得るため、同一集合である可能性があるのに対し、(21)の場合、たくさんいる野良犬が全て家の庭にねに来るのは想像しにくい。従って、部分集合である可能性が高くなる。しかしこのような場合では、個人に固有の価値観などによって微妙な違いが現れてくる。

## 2-3. 既出情報の記憶

以上の一般考察で分かるように、テキスト・対話中の既出情報を扱うため、言語の形式モデルではその情報を何らかの形で保存しなければ先行詞の抽出は不可能になる。そこで、フランス語の場合、F. Renaud (ルノー) (2) は、その情報を二種類に分別して保存することを提案した。一方で、先行文を分析し、その意味論的表現をデータベースで保存し、もう一方では、前出の名詞句に関する言語的なデータを短期記憶ベースで保存することにした。筆者はこの考えを日本語に応用してみた。そのモデルをここで紹介するため、具体的な例でこのシステムがどのように働くのかを示したいと思う。

## 2-4. 照応対象に関する形式モデルの提案：「その中の3匹の犬」とその変形の照応機能

ここでは、次の名詞句構造：((その(中の))(3匹の)犬(達/とも))

を徹底的に分析し、その変形の照応機能を考察し、この名詞句構造の各要素の機能を考察する。同時に、照応の機能に必要なデータをまとめておきたい。

その変形の照応機能を要約すると、次の図のようになる：

| 可能性 | 名詞句構造      | 照応対象 |       |    | 照応対象との関係 |    | 探索パラメーター |
|-----|------------|------|-------|----|----------|----|----------|
|     |            | 短期記憶 | 知識ベース | 有無 | 部分       | 同一 |          |
|     | 犬          | +    | +     | -  | ?        | ?  | ?        |
|     | 犬達         | +    | +     | -  | -        | +  | ?        |
| *   | 犬とも        | *    | *     | *  | *        | *  | *        |
|     | その犬        | +    | -     | +  | ?        | ?  | (犬,匹,>=) |
| *   | その犬とも      | *    | *     | *  | *        | *  | *        |
|     | 内の犬        | *    | *     | *  | *        | *  | *        |
| *   | 内の犬とも      | *    | *     | *  | *        | *  | *        |
|     | 3匹の犬       | -    | +     | +  | +        | +  |          |
|     | 3匹の犬       | +    | ?     | -  | +        | +  | ?        |
| *   | 3匹の犬とも     | *    | *     | *  | *        | *  | *        |
|     | その3匹の犬     | +    | -     | +  | +        | +  | (犬,匹,=)  |
| *   | その3匹の犬とも   | *    | *     | *  | *        | *  | *        |
|     | その中の3匹の犬   | +    | -     | +  | +        | +  | (犬,匹,>)  |
| ?   | その中の3匹の犬とも | *    | *     | *  | *        | *  | *        |
|     | 匹          | *    | *     | *  | *        | *  | *        |
|     | 3匹         | +    | -     | +  | +        | +  | (_,匹,=)  |
|     | 3匹とも       | +    | -     | +  | +        | +  | (_,匹,=)  |
|     | その3匹       | +    | -     | +  | +        | +  | (_,匹,=)  |
|     | その3匹とも     | +    | -     | +  | +        | +  | (_,匹,=)  |
|     | その中の3匹     | +    | -     | +  | +        | +  | (_,匹,=)  |
| *   | その中の3匹とも   | *    | *     | *  | *        | *  | *        |

注1) 「可能性」の欄は名詞句は文法的に不可(\*)、容認可(A)、判断不可(?)である。

注2) 各名詞句の照応対象が短期記憶に有り得る場合、知識ベースに有り得る場合にはその欄に「+」を、可能性がなければ「-」を記入した。照応対象が短期記憶、知識ベースの両方の記憶に保存された可能性がある場合、両方の欄に「+」を記入した。

注3) 先行詞の有無の可能性については「有無」の欄に、可能性があれば「+」を、無ければ「-」を記入した。

注4) 先行詞がある場合、同一集合の関係であれば「同一」欄に「+」を、部分集合の関係であれば「部分」欄に「+」を記入している。それがはっきりわからない名詞句構造は、「同一」「部分」の両欄に「+」を記入した。

注5) 「探索パラメーター」の欄は、名詞句を分析した結果から、どのようなデータを用いて、短期記憶で照応対象の検索を行うのかを記入した。以下、どのように読むのか、「(A,Class,Test)」の場合で示す：

A: 名詞句の頭という文法的なfeature(3章を参照)である。

Class: 分析中の名詞句に用いられた助数詞

Test: 分析中の名詞句の集合と照応対象の集合との要素数における関係

この欄中で「\_」とは、データが決まっていないという意味である。

この図から、名詞句の要素に次の照応機能があると思われる：

- 「その」や省略がある場合、先行詞があり、それは先行のテキスト内にある
- 「内」や「その内」があると、先行詞との関係は、部分集合である (inclusion)
- 「その」があると、先行詞との関係は同一集合、または、部分集合である。
- 「とも」がある場合、必ず複数であり、かつ同一集合である。

本章では、この結果を形式的な言語モデルで表現する。

### 3- 形式理論

#### 3-1- 言語形式モデル

ここで使用する形式モデルは、F. Renaud (1) により詳細に紹介されているため、ここでは詳しい説明は省略する。このモデルはfeature constraint logicに基づいたもので、それに知識ベースが加わったものである (Smolka 92も参考)。

このモデルでは文章の分析が次のような手順で行われる：まず文章がパーザーに分析され、その意義論表現が述語の形式言語で書かれる。文章が断言であれば、その式をスコレム化(skolemization)して知識ベースに書き込み、疑問であればそれを否定化した式をスコレン化し、refutationのプロセスにおいて知識ベースの中で適当な情報の検索を行う (Chang & Leeの例も参照)。短期記憶は筆者がpileの形でモデルにした。以下のその説明である。

以下、変数は大文字のローマ字で示す。

### 3-2. 名詞句

まず、本言語分析システムにおける名詞句の表記法を以下に示す (応用例は2-3) 章を参照)：

(部分集合の場合)

名詞句 $\leftarrow$  ...

U0:: (f助数詞: Class & f数: N & f頭: A) ...

S0 $\leftarrow$  lbd([P], ex([E], 集合.E.N & const(E, kn) & 部分集合(E, km) & prtt(X, 依存(X, E) => (S2\*X & P.X)))

(同一の場合)

名詞句 $\leftarrow$  ...

U0:: (f助数詞: Class & f数: N & f頭: A) ...

S0 $\leftarrow$  lbd([P], prtt(X, app(X, kn) => (S2\*X & P.X)))

このconstraintを説明すると、名詞句を分析するとき、まずどんなf助数詞が用いられているのかを示す。次にf数は、例えば「その内の3匹の犬」という名詞句においてのf数は3である。f頭は、中心の名詞のf頭に相当するので、この場合、「3匹の犬」の中心にある「犬」がそのf頭の値となる。これが省略された名詞句の場合、値は0となる。Lambda Term中の[集合(E, N)]で示す内容は、名詞句が指示するものが集合であり、その集合の要素の数量がNである。knは下に説明する形式的なスタック中にあるように名詞句の指示対象になっているものに定数としてつける。この際、初めて現れた集合それぞれに対して違う定数knをつける。ここではconst(E, kn)における定数Eは定数knと同一のものである。

本稿第一部で示したように名詞句の指示詞を抽出するには、直前の文中の名詞句とそれに関するいくつかのデータを記憶する必要がある。そのデータには言語的な情報と意味に関する情報の両方があるため、意味に関する情報のみを記憶する知識ベースには記録ができない。従って、先行詞に関する情報は知識ベースとは別のペースに記憶することになる。これは、本言語分析システムでは形式的なスタックである。スタック使用の利点として、述語論理式でデータを扱うのに比較して、データの扱いが簡潔で速いことがあげられる。また、名詞句の現れる順序を記憶する必要があるが、述語論理式ではそれが記憶されないのに対し、スタックではそれが可能である。このスタックにどのような情報を記録するのか、またそのスタックのデータをどのように扱うのかを以下説明する。

名詞句の分析を行うとその分析結果には、言語的、意義的な情報が含まれている。この情報の中で、他の名詞句分析での照応の計算に必要な情報のみをスタック (短期記憶) に記憶する。この照応に必要な情報とは、図1を参照すると、次のようなものである：文法的な情報としては、名詞句のf頭、f助数詞、f数。意義論的な情報としては、定数kn。よって、スタックは次のような形になる：

name\_stack([(f頭1, K1, f助数詞1, f数量1), ..., (f頭n, fKn, f助数詞n, f数量n)])

スタックで指示対象の検索を行う際、次のパラメーターを使用する：名詞句のf頭、f助数詞、f数の値、その名詞句の数と指示対象の数との関係のテスト（同一か否か）、そして名詞句の特性である。これらのパラメーターに対応するデータが検出されれば、数のテストを行い、これがうまくいった場合、意味のテストも行う。これもうまくいけば、そのデータの定数knを出し、分析中の名詞句の指示する集合との関係を表す式を出す。例えば、名詞句が「そのうちの3匹の黒い犬」であれば、スタックで(犬, K, 匹, N)を検索するが、それに相当するデータ(犬, ki, 匹, ni)があれば、そのniは必ずn>3である。これに対応するデータが1つあれば、その定数kiは意義論テストで確認される。この例では、色(黒, K)が、知識ベースのrefutationで確認される。この検索は[検索1]関数において行われる。

数量詞に省略された名詞句の場合も同様に働く。ただし調べる際、頭の値は決まっていない(頭=0)。従って、次のようなデータを調べることになる：( \_, K, Numerical, N)。

また、調べられた指示対象が複数ある場合もある：

(23) 家には、猫1匹と犬1匹がいる。 (24) 2匹とも、雌だ。

この場合、文章(23)の分析が終わると、スタックは次のような状態になる：

```
name stack([(猫, k1, 匹, 1)]).
```

(24)の「2匹とも」という名詞句の場合、検索1関数は次のように働く：( \_, k, 匹, 2)を検索し、相当するデータは無いが、数をのぞけば可能なデータがあるため、検索2を起こす。検索2は数を無視して、助数詞に応じる全てのデータを新しい集合としてまとめる。この新しい集合の数は、部分集合の数の和である。その和はテストでNに比べられ、うまくいく場合はこの新しい集合の要素、つまりその部分集合の要素を意味テストで確認する。うまくいく場合、この新しい集合に定数knをつけ、「2匹とも」の分析結果としてその定数を出す。この場合、数量詞に省略された名詞句であるため、意味のテストは無いが、次の例ではそれがある。スタック上でデータを調べる際、複数の候補が現れる場合があるため、正しいものを選択するためには知識ベースの知識を必要とする。次の形容詞に省略された名詞句には頭の値がない：

(25) 箱に白い瓶1本と青い瓶1本を入れた。 (26) 青いのが割れている。

「青いのが」を分析する際に、スタックの状態は次のようになっている：

```
name stack([(箱, k0, _, 1), (瓶, k1, 本, 1), (瓶, k2, 本, 1)]).
```

検索1が( \_, K, N)を検索すると、候補のデータは3つある。正しいデータを選択するには、その内のどの名詞句が「白い」のかを調べる必要がある。そのために知識データベースを使用する。

検索1も検索2も失敗すると、名詞句の構造により検索を知識ベースで行うかどうかが決まる。

### 3-4. 「そのうちの3匹の犬」

ここでは図の全ての名詞句を分析する余地はないので、「そのうちの3匹の犬」に絞って形式的構造と意味を示す。

#### 辞書

犬 (名詞, (f助数詞:匹, f数:0, f頭:犬), lbd([X], 犬(X))

3 (数詞, (f値:3), 3)

匹 (助数詞, (f助数詞:匹), vrai)

その中 (前置詞, (f連体:否),

lbd([A, Class, N, Rs, P], ex([E], 集合(E, N) & 検索1(A, Class, lbd(X, N<X), Rs, name\_stack))

& prtt(Y, app(Y,E)=>Rs\*Y & P\*E)))

規則

R1 数量詞 ← 数詞 助数詞

U0:: (f助数詞:Class & f数:N & f連体:否) & U1:: (f値:N) & U2:: (f助数詞::Class)

S0<= S1

R2 前置詞 ← 前置詞 の 数量詞

U0:: (f助数詞:Class & f数:N & f連体:否) & U1:: (f連体:否) & U2 (f助数詞:Class & f数:N & f連体:否)

S0<= S1\*S2

R3 名詞句 ← 前置詞 の 名詞

U0:: (f助数詞:Class & f数:N & f頭:A) & U1:: (f助数詞:Class & f数:N f連体:否) &

U1:: (f助数詞:Class & f数:0 & f頭:A).

S0<=S1\*A\*Class\*N\*S2

結語

以上本稿では、短期記憶に基づく照応の計算の形式モデルを紹介した。このモデルの実施により、照応の計算が短期記憶のデータを用いて行われるという推論の、いくつかの問題点が明らかになった。その問題点の中で特に、短期記憶のデータ量が不明であったことがあげられる。計算の成功の条件として、短期記憶スタックでの記録可能なデータの最大量を定めることが必要である。このように問題点があるとしても、モデルの使用は、すべての名詞句構造とその機能を体系的に分析することを可能とするため、有益であると思われる。これにより、短期記憶上か、あるいは知識データベース上のどちらで検索が行われるべきか分かりにくい、難しいケースを解決する方法を見いだせるかもしれない。そのためにも、モデルが開発される価値があると思われる。また、その価値は、既にフランス語で用いられたことが示すように、一言語の研究に限定されないことにある。

-----  
文献

RENAUD F. (1), (1996年秋予定), Semantique du temps et lambda-calcul,  
(テンスの意義論とラムダ計算), PUF. (フランス語)

RENAUD F., 予定, L'article defini, in Journal of Semantics, Vol 13 Num 1,

KAMPS H. & REYLE U., 1993, From discourse to logic, Introduction to Model Semantics  
of Natural Language, Formal Logic and Discourse Representation Theory, Kluwer.

SMOLKA G., 1992, Feature-constraint Logics for Unification Grammars, in J. Logic Programming 12.

CHANG CH-L., LEE R. CH-T. (1973), Symbolic Logic and Mechanical Theorem Proving, Academic Press.