

語意に基づく深層レベルの指示代名詞照応解析システム Anasys/D

山岸 謙治[†] 村松 孝彦[†] 原田 実^{††}

文章では、同じ言葉の繰り返しが嫌われるため、指示代名詞が頻繁に使われる。指示代名詞の指示先が何であるかを決定することは、自然言語インターフェースを持つ応用システムにおいて、文章の内容を理解する上で重要である。我々は、指示代名詞の照応要素（先行詞）を決定する照応解析システム Anasys/D を開発した。ここでは、EDR 電子化辞書による指示代名詞の係り先や修飾先の語意と先行詞候補の語意の類似性の評価に基づく決定的手法とそれらの語間距離や指示代名詞の特性などの表層的な情報に基づくルールベース手法を融合した。

Development of demonstrative pronoun resolution system Anasys/D using similarity of word meaning

KENJI YAMAGISHI,[†] TAKAHICO MURAMATSU[†] and MINORU HARADA^{††}

In sentences, because the repetition of the same word is disliked, the demonstrative pronoun is frequently used. In the application system with the natural language interface, it is important for understanding the content of sentences to decide which word is indicated by each demonstrative pronoun. We developed the anaphoric analysis system Anasys/D which can decide the indicated element of each demonstrative pronoun. Here, a deterministic technique based on the evaluation of the similarity of the meaning of the antecedent candidate and that of the dependant or modificant of the demonstrative pronoun by using the EDR electronic dictionary and a rule base technique based on the shallow information such as the distance between those two words and the characteristic of the demonstrative pronoun were united.

1. はじめに

1.1 研究背景

日本語文章では、同じ言葉の繰り返しを嫌うために指示詞によって言葉が置き換えられる。これは自然言語処理において大きな問題となり、文章に表された内容を正しく解析し、適切な応答を返すためには、指示詞が指す対象が何であるかを解析する必要がある。これらの照応解析の研究では、村田^⑥のように表層的な情報を中心として用いる規則を多数用意して解析を行っている。しかし表層的な情報による解析では精度を上げていく際にその規則の量が膨大になる傾向がある。これに対して語意や深層格の情報を用いて、それぞれの単語が持つ概念の類似度により、指示先を決定できればこのような問題を解決できるのではないかと

考えた。

1.2 研究目的

本研究では公開されている EDR 電子化辞書から得られる語意や用言における深層格を用いて、先行詞を特定する照応解析システム Anasys/D を開発する。Anasys/D の解析では、指示詞の照応要素である先行詞特定において、EDR 電子化辞書^⑦による語意の類似性の評価に基づく決定的手法と語間距離や指示代名詞の特性などの表層的な情報に基づくルールベース推論を融合した手法を提案する。

2. EDR 電子化辞書

本論文では、先行詞の特定のため単語の語意と深層格を用いている。これらの深層格と語意に関する情報は、EDR 電子化辞書に記載されているものを用いる。EDR 電子化辞書は言語データを扱った電子化辞書であり、単語の語意を 16 進数 6 衔の概念 ID として与えている。本論文では EDR 電子化辞書の概念体系辞書と日本語共起辞書の 2 つを用いる。概念体系辞書は、単語の語意間の上下関係を木構造に体系化し、集約したものである。日本語共起辞書は、日本語コーパスに

† 青山学院大学理工学部経営工学科

Department of Industrial and System Engineering,
Aoyama Gakuin University

†† 青山学院大学理工学部情報テクノロジー学科

Department of Integrated Information Technology,
Aoyama Gakuin University

格納された実文の解析結果から係り受けを構成している部分を抽出し、句の表記の五十音順に並べたものである。

3. システム概要

3.1 入力仕様 ～格フレーム群～

Anasys/D は、原田研究室で開発された意味解析システム SAGE¹⁾²⁾³⁾⁴⁾ が日本語文章を意味解析した結果である格フレーム群を入力とする。SAGE における日本語意味解析とは、「文章中の単語の語意を決定し、係り受け関係にある 2 文節間の深層格を決定する」ことであるといえる。SAGE によって生成された格フレームには、単語の語意と係り受け関係にある全ての 2 文節間の格関係が記述されている。具体的には、格フレームは、図 1 に示すような形式をしており、フレーム番号、日本語文中に現れた形、品詞に関する情報、概念 ID、原文の文番号などの 13 の要素から構成される。

3.2 Anasys/D の基本構成

Anasys/D システムにおける処理の流れを図 2 に示す。Anasys/D は、指示代名詞検出部と先行詞解析部から構成される。指示代名詞検出部は、意味解析システム SAGE が生成した格フレーム群を入力とし、指示代名詞が検出された場合、それを指示代名詞フレーム群として出力する。先行詞解析部では、格フレーム群と指示代名詞フレーム群を入力とし、先行詞を解析した結果を先行詞解析結果フレーム群として出力する。また格フレーム群中の、先行詞解析された指示代名詞に対する格フレームを、指示代名詞の照応要素として特定された先行詞の情報を付与した修正済み格フレームに入れ替える。

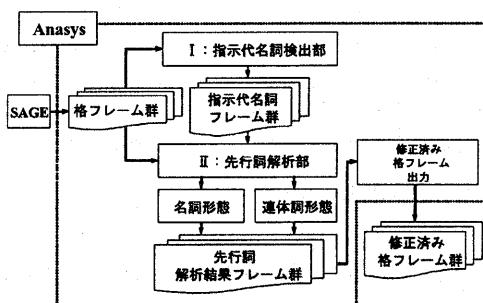


図 2 Anasys/D における処理の流れ
Fig. 2 The basic dataflow in Anasys/D

3.3 出力 1～指示代名詞解析結果フレーム群～

Anasys/D は、指示代名詞の照応要素を特定した結

果を先行詞解析結果フレームとして出力する。先行詞解析結果フレームは、指示代名詞と指示代名詞の照応要素である先行詞との関係を記述している。先行詞解析結果フレームは、図 3 に示すように指示代名詞 d の「frame 番号」、「先行詞であることを表す深層格 (antecedent と specify)」、「先行詞の frame 番号」の三つの要素から構成される。図 3 に示した例は、frame 番号 26 で表された指示代名詞「それ」の照応要素である先行詞が、frame 番号 17 で表された「桃」であることを示している。

Analysis_result(27,[[antecedent,18]]).

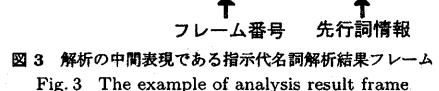


図 3 解析の中間表現である指示代名詞解析結果フレーム
Fig. 3 The example of analysis result frame

3.4 出力 2～修正済み格フレーム群～

Anasys/D は、SAGE が output した格フレームに Anasys によって指示代名詞の照応要素として特定された先行詞の情報を付与した格フレームを出力する。図 4 に示したように、指示代名詞の先行詞解析が行われたならば、% (prolog でコメントを表す) を用いてその指示代名詞フレームをコメントアウトし、新たに指示代名詞の照応要素である先行詞の情報を付与した修正済み格フレームを次行に挿入する。図 4 に示した例では、SAGE が output した frame 番号 26 で表された指示代名詞フレーム「それ」の先行詞解析が行われたので、このフレームを、% を用いてコメントアウトし、指示代名詞「それ」の照応要素として解析された先行詞として、frame 番号 17 で表された「桃」の antecedent 格として情報を付与した修正済み格フレームを新たに挿入している。

4. 指示代名詞検出部

SAGE の出力した格フレーム群から、KNP に対応する EDR 品詞コード、KNP に対応する日本語表記品詞をキーとして名詞を検索し、指示代名詞 d を抽出し、指示代名詞フレームとして出力する。SAGE の出力した格フレーム群に含まれる KNP による品詞を使用するのは、EDR 品詞による分類では指示代名詞は「指示詞-代名詞一般」と解析されるのに比べ、「指示詞-名詞形態指示詞」、「指示詞-連体詞形態指示詞」、「指示詞-副詞形態指示詞」のように指示代名詞の分類が細かいためである。指示代名詞フレームは、図 5 に示すように指示代名詞 d の「frame 番号」、「文中表記」な

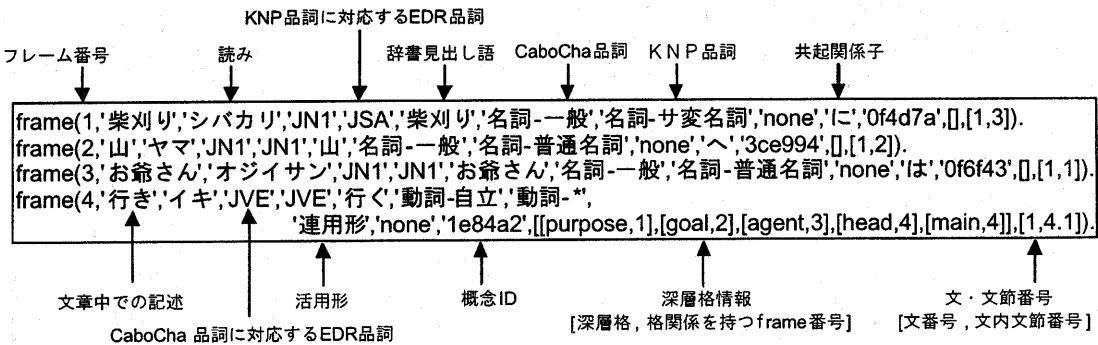
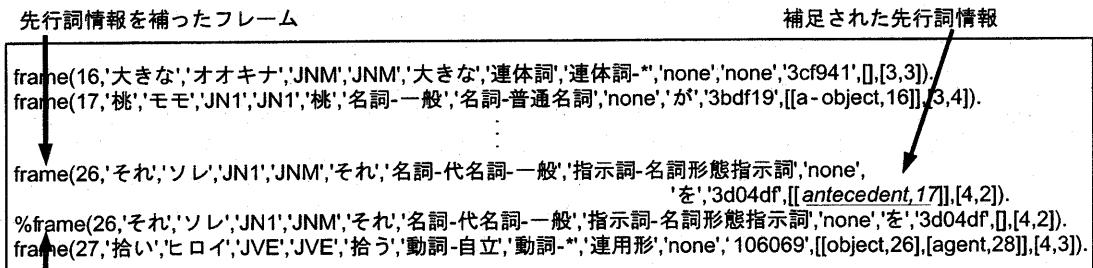


図 1 Anasys/D の入力である格フレーム群

Fig. 1 Case frames input to Anasys/D

図 4 Anasys/D の出力である修正済み格フレーム
Fig. 4 Corrected case frames generated from Anasys/D

どの 7 の要素から構成される。

表 1 解析対象となる指示代名詞
Table 1 Demonstrative pronoun図 5 解析の中間表現である指示代名詞フレーム
Fig. 5 The depro frame used as an intermediate expression

5. 先行詞解析部

先行詞解析部は、指示代名詞検出部で生成された指示代名詞フレームと SAGE が output した格フレーム群を入力とし、検出された指示代名詞の照応要素の解析を行う。

表 1 に示すように、指示代名詞は「名詞形態指示詞」と「連体詞形態指示詞」と「副詞形態指示詞」の 3 つに分類される。本研究では、「名詞形態指示詞」と「連体詞形態指示詞」の先行詞解析を行う。

名詞形態 指示詞	これ ここ こちら こっち	それ そこ そちら そっち	あれ あそこ あちら あっち
連体詞形態 指示詞	この こんな こういう こうした こういった このような	その そんな そういう そうした そういった そのような	あの あんな ああいう ああした ああいった あのような
副詞形態 指示詞	こう このように こんなに こんなふうに	そう そのように そんなに そんなふうに	ああ あのように あんなに あんなふうに

5.1 名詞形態指示詞の解析

指示代名詞 d を深層格として持つ用言 v が存在する場合、先行詞解析を行う。図 6 に名詞形態指示詞の解析における処理の流れを示す。

5.1.1 指示代名詞と用言の関係の調査

SAGE が output した格フレーム群から、指示代名詞 d と同じ「文番号」を持つ用言フレーム v でその深層格

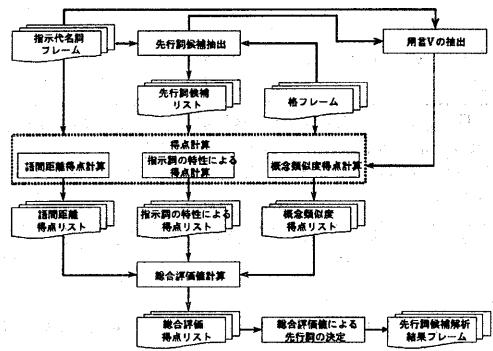


図 6 名詞形態指示詞解析部の処理の流れ

Fig. 6 The data flow of nounform pronouns detection part

c による接続先が d のものを求める。

例として、図 7 に示した指示代名詞「それ」の格フレームを用いて説明する。指示代名詞「それ」の「文番号」「4」と等しい「文番号」を持つ格フレームから「EDR 品詞コード」が“JVE(動詞)”又は“JSA(サ変動詞)”のものを探し、そのフレームが「それ」に深層格で接続しているかを調べると、frame 番号 27 「拾い」の object 格に「それ」が含まれることがわかる。

5.1.2 係り元の共起事例の検索

抽出された用言 v の「語幹」と「概念 ID」をキーとして共起辞書を検索し、用言 v に関する共起事例から深層格 c による係り元の語を抽出する。

例えば、用言「拾う」の「語幹」「拾」と「概念 ID」「3ce7f0」の二つをキーとして共起辞書検索を行い、共起事例を抽出する。さらに抽出した共起事例から深層格が“object”的ものをとり出し、その係り元の語を列挙すると表 2 のようになる。

5.1.3 先行詞候補の抽出

指示代名詞 d の前 2 文と後 1 文の範囲から、「EDR 品詞コード」をキーとして、その範囲中の全ての名詞格フレームを先行詞候補 a として抽出する。

図 8 に frame 番号 26 で表された指示代名詞「それ」に対する先行詞候補群を示す。

5.1.4 総合評価値による先行詞の決定

各々の先行詞候補について、「概念類似度得点」と「語間距離得点」と「指示詞の特性による得点」を求め、図 9 のようにして総合評価値を計算し、最も高得点のものを先行詞として確定する。

$$\text{総合評価値} = \text{概念距離得点} + \text{語間距離得点} + \text{特性得点} \\ \times 0.2 \quad \times 0.2$$

図 9 総合評価得点算出式

Fig. 9 Total score

表 2 用言「拾（う）」の共起事例
Table 2 The collocation example of a verbal "hirou"

受け側概念ID	深層格	係り側概念ID
拾(3ce7f0)	object	えんぴつ(3c1164)
拾(3ce7f0)	object	ごみ(3ce810)
拾(3ce7f0)	object	球(0fc580)
拾(3ce7f0)	object	リンゴ(3bd8db)
拾(3ce7f0)	object	落とし物(0e7d0b)
拾(3ce7f0)	object	磁束(3c6209)
拾(3ce7f0)	object	財布(1e8f92)
拾(3ce7f0)	object	風船(1066f8)

5.1.5 概念類似度得点の計算

抽出された先行詞候補それぞれについて、用言 v の共起事例 e 中の係り元語との概念類似度計算を行う。概念類似度計算は、図 10 に示すように 1 個の先行詞候補 a と用言 v の各共起事例中の係り元語 m との語意類似度を計算し、その平均値を求めるものである。語意類似度計算とは先行詞特定のために頻繁に用いられる二つの単語の語意の近さを計算するものである。先行詞候補の語 a の概念 ID を用いて、概念体系辞書検索を行い、上位概念束を生成する。さらに生成した上位概念束を上位概念リスト UCLa に展開する。同様にして係り元語 m の上位概念リスト UCLm も作成する。展開された語 a の上位概念リスト UCLa と語 m の上位概念リスト UCLm を用いて、図 12 のように両概念間の類似度を評価する。実際の評価に当たっては、個々のリストの比較になる。

以下に語意類似度の算出処理ステップを述べる。①語 a と語 m の共通上位語 c を両上位概念リストを比較することによって求め、最上位概念である「概念」から共通上位語 c までの距離を Dc とする。②語 a から語 c までの距離 Dar を項数 ar, 初項 1, 公比 0.7 の等比数列の和として求める。同様に比較対象の語 m から共通上位語 c までの距離 Dmr を項数 mr, 初項 1, 公比 0.7 として求める。③①と②で求めた値を図 11 に示すように語意類似度計算式に代入する。

$$\frac{\sum \text{語意類似度}}{\text{全共起事例}} = \frac{\text{概念類似度}}{\text{共起事例の総数}}$$

図 10 概念類似度算出式

Fig. 10 Conceptual similarity score

5.1.6 語間距離得点の計算

指示代名詞の先行詞は、指示代名詞の前方に存在す

```

frame(1,'柴刈り','シバカリ','JN1','JSA','柴刈り','名詞-一般','名詞-サ変名詞','none','に','0f4d7a',[[],[1,3]]).
frame(2,'山','ヤマ','JN1','JN1','山','名詞-一般','名詞-普通名詞','none','へ','3ce994',[[],[1,2]]).
frame(3,'お爺さん','オジイサン','JN1','JN1','お爺さん','名詞-一般','名詞-普通名詞','none','は','0f6f43',[[],[1,1]]).

frame(16,'大きな','オオキナ','JNM','JNM','大きな','連体詞','連体詞-*','none','none','3cf941',[[],[3,3]]).
frame(17,'桃','モモ','JN1','JN1','桃','名詞-一般','名詞-普通名詞','none','が','3bc503',[[],[a-object,16],[3,4]]).
frame(18,'上流','ジョウリュウ','JN1','JN1','上流','名詞-一般','名詞-普通名詞','none','から','0f811a',[[],[3,2]]).

frame(25,'家','イエ','JN1','JN1','家','名詞-一般','名詞-普通名詞','none','へ','0e5cdb',[[],[4,4]]).
frame(26,'それ','ソレ','JN1','JNM','それ','名詞-代名詞-一般','指示詞-名詞形態指示詞','none','を','3d04df',[[],[4,2]]).
frame(27,'拾い','ヒロイ','JVE','JVE','拾う','動詞-自立','動詞-*','連用形','none','106069',[[],[object,26],[agent,28],[4,3]]).
frame(28,'お婆さん','オバアサン','JN1','JN1','お婆さん','名詞-一般','名詞-普通名詞','none','は','0e7edf',[[],[4,1]]).

frame(33,'お爺さん','オジイサン','JN1','JN1','お爺さん','名詞-一般','名詞-普通名詞','none','と','0f6f43',[[],[5,3]]).
frame(34,'一緒','イツショ','JSA','*','一緒','名詞-サ変接続','*','none','に','0e4b56',[[],[and,33],[5,4]]).
frame(35,'早速','サッソク','JD1','JD1','早速','副詞-助詞類接続','副詞-*','none','none','3ced5f',[[],[5,2]]).
frame(36,'食べる','タベル','JVE','JVE','食べる','動詞-自立','動詞-*','基本形','none','3bc6f0',[[],[manner,34],[manner,35],[5,5]]).

```

図 7 例フレーム群

Fig. 7 The case frames

```

frame(8,'洗濯','センタク','JSA','JSA','洗濯','名詞-サ変接続','名詞-サ変名詞','none','を','0fa20f',[[],[2,3]]).
frame(9,'川','カワ','JN1','JN1','川','名詞-一般','名詞-普通名詞','none','で','0ea9dd',[[],[2,2]]).
frame(10,'お婆さん','オバアサン','JN1','JN1','お婆さん','名詞-一般','名詞-普通名詞','none','は','0e7edf',[[],[2,1]]).
frame(17,'桃','モモ','JN1','JN1','桃','名詞-一般','名詞-普通名詞','none','が','3bc503',[[],[cause,16],[3,4]]).
frame(18,'上流','ジョウリュウ','JN1','JN1','上流','名詞-一般','名詞-普通名詞','none','から','0f811a',[[],[3,2]]).
frame(25,'家','イエ','JN1','JN1','家','名詞-一般','名詞-普通名詞','none','へ','0e5cdb',[[],[4,4]]).
frame(28,'お婆さん','オバアサン','JN1','JN1','お婆さん','名詞-一般','名詞-普通名詞','none','は','0e7edf',[[],[4,1]]).
frame(33,'お爺さん','オジイサン','JN1','JN1','お爺さん','名詞-一般','名詞-普通名詞','none','と','0f6f43',[[],[5,3]]).

```

図 8 先行詞候補

Fig. 8 Noun case frames as the antecedent candidates

$$\text{語意類似度} = \left(\frac{0.6dc}{0.6dc + (1 - 0.7^{ar}) + (1 - 0.7^{mr})} \right) \times 100$$

図 11 語意類似度算出式

Fig. 11 Word meaning similarity score

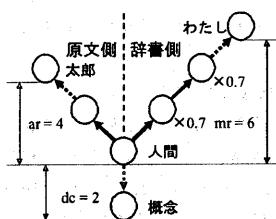


図 12 上位概念リストの比較

Fig. 12 The comparison of upper concepts lists

る(前方照応)可能性が高く、先行詞が指示代名詞の後方に存在する(後方照応)可能性は低い。また、先

行詞は指示代名詞の近くに存在する可能性が高い。以上の2点の性質から、図14に示す計算式を用いて語間距離得点を求める。

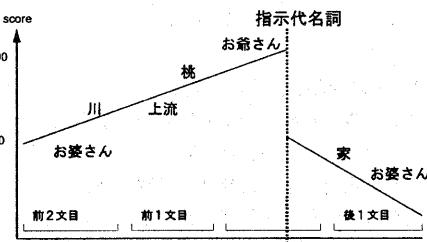


図 13 語間距離得点の例

Fig. 13 The example of word distance score

$$\text{語間距離得点} = \frac{\text{指示代名詞から数えて何番目の単語か}}{\substack{\text{(先行詞候補が指示代名詞の} \\ \text{前方にある場合)}}} \times 100$$

$$\text{語間距離得点} = \frac{\text{指示代名詞から数えて何番目の単語か}}{\substack{\text{(先行詞候補が指示代名詞の} \\ \text{後方にある場合)}}} \times 50$$

図 14 語間距離得点算出式

Fig. 14 Word distance score

5.1.7 指示代名詞の特性による得点の計算

指示代名詞 d の特性による得点計算を行う。各々の指示代名詞には、「どのような事物を指示対象としやすいか」という性質がある。例えば、指示代名詞「ここ」、「そこ」、「あそこ」の場合には、場所を表す語を指示対象としやすいであろうと推測できる。EDR 電子化辞書のコーパス辞書等の事例検索から各々の指示代名詞がどのような事物を指示対象としやすいかを判断し、指示代名詞個別に作成した規則に従い得点を計算し、「指示詞の特性による得点」とする。

指示代名詞「それ」にあてはまる EDR 電子化辞書の概念 ID は 4 種である。それぞれの概念 ID の EDR コーパスにおける共起事例件数を図 15 に示す。この共起事例件数から、指示代名詞「それ」は「時」や「場所」を示すことが極端に少ないと言える。よって、「指示詞の特性による得点」の満点を 100 とし、先行詞候補が上位概念に「時」"30f776" や「場所」"3aa938" を含む場合は 0.8 を掛け、得点を減点する。

「それ」にあてはまる概念ID		EDRコーパス 事例件数
概念ID:3d04df	事物を示す中称の指示代名詞	1326
概念ID:0fb3bd	すぐ前に述べた事柄を繰り返して示す時に用いる語	573
概念ID:0faed0	時を示す中称の指示代名詞	3
概念ID:0fb3bc	場所を示す中称の指示代名詞	5

↓
減点対象となる概念

概念ID : 30f776 時	/ time
概念ID : 3aa938 場所	/ location, position

図 15 指示代名詞「それ」の特性

Fig. 15 Property of "sore"

5.2 連体詞形態指示詞の解析

連体詞形態指示詞は指示の仕方から限定指示と代行指示という二つに大別される。図 16 に示すように、限定指示の場合、指示詞自体は意味を持たず、代行指示の場合は指示詞が先行詞を代行するというように指示詞の使われ方が違う。そこで解析も限定指示と代行指示でわけて行うこととする。限定指示は先行詞と被修飾名詞 A は概念が非常に近いので、限定指示で先行詞が存在する場合、その先行詞の評価得点は非常に高くなる。また代行指示で先行詞が存在する場合、被修

限定指示例文

ex. 天狗たちはおじいさんが来たことを大喜びました。
しかし、このおじいさんは踊りが下手でした。

このおじいさん= おじいさん
「この」自体は意味を持たない。

代行指示例文

ex. 鳥のような顔をした天狗もいました。
その口は鋭く尖っているのでした。

その口 = 天狗の口
「その」は「天狗の」を代行している。

図 16 連体詞形態指示詞の分類

Fig. 16 The adminal demonstrative pronoun

飾名詞 A と先行詞候補の概念が近くて限定指示であると誤解しそうな文章というの日本語としておかしなものになってしまふ。そこでこの性質を利用し、限定指示解析で高めのしきい値を設定して、しきい値に達したらそれを先行詞とする。基準点に達しなかった場合は代行指示の可能性があるとして代行指示解析を行う。なお照應解析の基本となる名詞を先行詞に持つ場合の解析に特化するために本研究では指示詞が用言を指示する場合の解析と文章を指示する場合の解析は行わない。

5.2.1 連体詞形態指示詞が修飾する名詞 A の抽出

連体詞形態指示詞の解析において重要な手がかりとなるのがその指示詞に修飾される名詞 A なので、これをまず SAGE が抽出した格フレームから深層格情報をキーに抽出する。被修飾名詞 A とは図 17 が示すように、指示詞 d を modifier 格として持つ名詞である。

5.2.2 先行詞候補の名詞の抽出

連体詞形態指示詞が存在する文番号からその 3 文前まで、そして指示詞が存在する文中では指示詞の文節番号以前を先行詞候補が存在する範囲とする。この範囲内から EDR の品詞番号をキーとして “JN1(普通名詞)”, “JN2(固有名詞)”, “JB1(接尾語)”, “JN6(名詞-接尾-助数詞)”, “JSA(名詞-サ変接続)” を持つ名詞を抽出し、これらを先行詞候補として図 18 に示す先行詞候補リストとして作成する。

5.2.3 限定指示解析部

被修飾名詞 A のフレームと先行詞候補の抽出で生成された先行詞候補リストを入力として解析を行う。図 19 に限定指示の解析の流れを示す。被修飾名詞 A と先行詞の関係は基本的に被修飾名詞 A と先行詞が等しいか、名詞 A は先行詞の上位概念として存在する場合の二通りがある。ここで EDR 電子化辞書の上位概念検索を用いる。5.1.5 中で論じた語意類似度計算により被修飾名詞 A と先行詞候補との概念の近さを得点化

frame(4, 'この', 'コノ', 'JNM', 'JAM', 'この', '連体詞', '指示詞-連体詞形態指示詞', 'none', 'none', '3d1839', [], [1, 1]).
 frame(5, '歌', 'ウタ', 'JN1', 'JN1', '歌', '名詞-一般', '名詞-普通名詞', 'none', 'は', '3cf9d9', [[modifier, 4]], [1, 2]).

[modifier (修飾), frame 番号 4 と 格関係]

図 17 被修飾名詞 A の格フレーム

Fig. 17 The case frame of the embellished noun A

```
frame(1, '柴刈り', 'シバカリ', 'JN1', 'JSA', '柴刈り', '名詞-一般', '名詞-サ変名詞', 'none', 'に', '0f4d77', [], [1, 3]).  

frame(2, '山', 'ヤマ', 'JN1', 'JN1', '山', '名詞-一般', '名詞-普通名詞', 'none', 'に', '3ce994', [], [1, 2]).  

frame(3, 'お爺さん', 'オジイサン', 'JN1', 'JN1', 'お爺さん', '名詞-一般', '名詞-普通名詞', 'none', 'は', '0f6f43', [], [1, 1]).  

frame(9, '川', 'カワ', 'JN1', 'JN1', '川', '名詞-一般', '名詞-普通名詞', 'none', 'で', '0ea9dd', [], [2, 2]).  

frame(10, 'お婆さん', 'オバアサン', 'JN1', 'JN1', 'お婆さん', '名詞-一般', '名詞-普通名詞', 'none', 'は', '0e7edf', [], [2, 1]).  

frame(17, '桃', 'モモ', 'JN1', 'JN1', '桃', '名詞-一般', '名詞-普通名詞', 'none', 'が', '3bc503', [[cause, 16]], [3, 4]).  

frame(18, '上流', 'ジョウリュウ', 'JN1', 'JN1', '上流', '名詞-一般', '名詞-普通名詞', 'none', 'から', '0f811c', [], [3, 2]).  

frame(25, '家', 'イエ', 'JN1', 'JN1', '家', '名詞-一般', '名詞-普通名詞', 'none', 'へ', '0e5cdb', [], [4, 4]).  

frame(26, 'それ', 'ソレ', 'JN1', 'JNM', 'それ', '名詞-代名詞-一般', '指示詞-名詞形態指示詞', 'none', 'を', '3d04df', [], [4, 2]).  

frame(28, 'お婆さん', 'オバアサン', 'JN1', 'JN1', 'お婆さん', '名詞-一般', '名詞-普通名詞', 'none', 'は', '0e7edf', [], [4, 1]).  

frame(33, 'お爺さん', 'オジイサン', 'JN1', 'JN1', 'お爺さん', '名詞-一般', '名詞-普通名詞', 'none', 'ど', '0f6f43', [], [5, 3]).  

frame(37, 'こと', 'コト', 'JN1', 'JN1', 'こと', '名詞-非自立-一般', '名詞-形式名詞', 'none', 'に', '0e84ad', [[which, 28]], [5, 6]).
```

図 18 先行詞候補としての名詞格フレーム

Fig. 18 Noun case frames as the first antecedent candidates

する。得られた得点をその先行詞候補の先行詞確信度とし、これを先行詞候補全てに対して行う。さらに名詞形態指示詞の解析と同様にして、指示詞と各先行詞候補との語間の距離によって先行詞確信度に加点処理を施す。これまでの各先行詞に対する加点の最終結果を評価値として、しきい値 110 を超えている先行詞候補のうちで最高得点のものを正しい先行詞であると判断し、先行詞解析結果ファイルに出力する。

た場合には、限定指示解析部と同様に被修飾名詞 A のフレームと先行詞候補リストを入力として、図 20 に示すように代行指示解析を行う。図 15 の代行指示の例文を用いて代行指示解析について説明する。「その口」とは「天狗の口」を示しており、「その」は「天狗の」を代行するかたちとなっている。そこで「その口」を「その口」と捉えて、先行詞候補の中から「それ」に当てはめてもっとも相応しいものを先行詞であると判断する。しかし「その」のような連体詞形態指示詞は名詞形態指示詞と異なり用言の格ではないので、用言との意味関係からは先行詞候補をしほることができない。そこで共起辞書から「(名詞 α) の口」という用例を被修飾名詞 A である「口」を用いて抽出し、その用例から名詞 α リストを作成する。ここで限定指示解析部とは異なり、5.1.5 で論じた概念類似度を先行詞候補と名詞 α リストから求め、これを先行詞候補の先行詞確信度とする。また限定指示解析部と同様に指示詞と各先行詞との語間の距離により先行詞確信度に加点処理を施す。これまでの各先行詞に対する加点の最終結果を評価値として、このうちで一番得点の高いもので、その得点がしきい値 60 を超えていたら代行指示を行っていると判断し、先行詞解析結果ファイルに深層格情報 antecedent で出力する。しきい値を超えていない場合は文や用言を指示しているか、外界照応であると考えられるが、本研究ではこの先の解析は行っていない。これらの解析を行うことは今後の課題である。

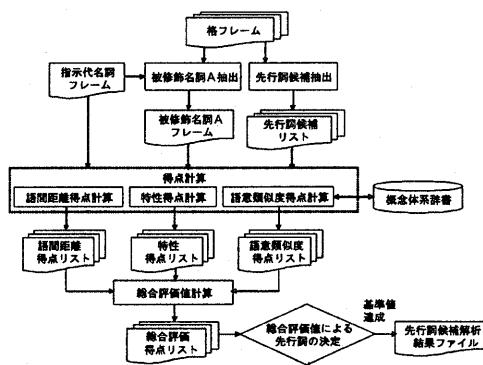


図 19 限定指示解析部の処理の流れ

Fig. 19 The dataflow of the specified demonstrative pronoun analysis part

5.2.4 代行指示解析部

限定指示解析で基準点に達するものが存在しなかつ

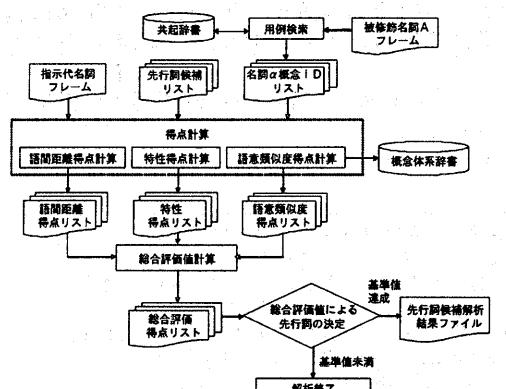


図 20 代行指示解析部の処理の流れ

Fig. 20 The dataflow of the possessive demonstrative pronoun analysis part

6. 結 論

GDA コーパスによる 1994 年毎日新聞の文章を用いて評価実験を行った。名詞形態指示詞において、54.8% の正答率を得た。連体詞形態指示詞の解析において限定指示解析は 85.3%、代行指示解析は 72.2% の正答率で、用言や文章を指示する例文を限定指示や代行指示であると誤って検出した例は 7.7% だった。

謝辞 本研究の一部は文部省科学研究費基盤研究 C『日本語文章の常識を用いた意味理解・文脈理解システムの開発研究』の補助金を用いて行われました。ここに感謝いたします。

参 考 文 献

- 1) 原田実、水野高宏: EDR を用いた日本語意味解析システム SAGE, 人工知能学会論文誌, Vol. 16, No. 1, pp. 85-93(2001).
- 2) 原田実、田淵和幸、大野博之: 日本語意味解析システム SAGE の高速化・高精度化とコーパスによる精度評価, 情報処理学会論文誌, Vol.43, No.9, pp.2894-2902, (2002.9).
- 3) 水野高宏、原田実: 日本語意味解析システム SAGE の高速化と精度向上, 人工知能学会第 14 回全国大会論文集, pp. 149-152(2000).
- 4) 尾見孝一郎、原田実、岩田隆志、水野高宏: 日本語文章からの意味フレーム自動生成システム SAGE(Semantic frame Automatic GENERator) の開発研究, 人工知能学会第 13 回全国大会論文集, pp. 213-216(1999).
- 5) 南旭瑞、原田実: 語意に基づく深層レベルの照應解析システム Anasys, 情報処理学会論文誌, Vol.41, No.6, pp.1234-1241 (2000.6).
- 6) 村田真樹、長尾真: 用例や表層表現を用いた日

本語文章中の指示詞・代名詞・ゼロ代名詞の指示対象の推定, 自然言語処理, Vol.2, No.3, pp.3-26 (1995.7).

- 7) (株) 日本語電子辞書研究所: EDR 電子化辞書使用説明書 (第 2 版), (株) 日本語電子辞書研究所 (1995).