

キーボード会話における名詞句の同一性の理解

野垣内 出 飯田 仁

ATR自動翻訳電話研究所

本報告では、対話における名詞と名詞の照応を理解する手法を示す。対話において話者と聞き手は、共通する知識や推論によって照応を理解していると仮定し、ドメインに依存する知識および推論を用いて、名詞と名詞の照応を理解する手法を示す。特に、照応関係の先行詞の候補であるが照応関係とならない例などから、制限的な条件を集め、実験システム上で検証をした。また、照応関係には発話上に明示的に現れないが、照応の先行詞となっていると考えられるものがあり、これらについて推論によって理解できることを示す。また、照応関係の理解の応用について考察する。

A Method of Semantic Identification for Noun Phrases in Inter-Keyboard Dialogues

Izuru NOGAI, Hitoshi IIDA

ATR Interpreting Telephony Research Laboratories

Sanpeidani Inuidani Seika-cho Soraku-gun Kyoto 619-02 Japan

A nominal anaphora is often occurred on telephone & inter-keyboard dialogues. The identity conditions for anaphora is given, but these rules give a possibilities of anaphora. So it is difficult to filter out many possibilities. This paper presents a method for identification of simple nominal phrase by restriction rules from domain knowledge and utterance. This method has been implemented on expert system for verifying the rules.

[1] はじめに

対話などで省略(ゼロ代名詞)の補完などを考慮する場合、主題の連続を条件とすることがある。しかし、連続した主題でも、主題となる名詞は、照応関係にある別の名詞で現れることがある。形態の異なる名詞では、単純には、同じ主題であることが判明しないため、照応の理解が必要となる。また、名詞の訳語選択などにも照応関係の理解は重要である。

本報告では、協調性のある対話において話者は聞き手に誤解のないように何らの情報を与えているか、話者と聞き手は共通する知識や推論によって、照応を理解していると仮定し、ドメインに依存する知識や推論などを与えることによって、名詞と名詞の照応を理解する手法を示す。特に、対話における照応に関しての調査を行い、照応の先行詞の候補となる条件、また、この条件があるが照応関係とならない例などから、制限的な条件を集め、実験システム上でこれらの条件を検証した。照応関係には発話上に明示的に現れないが、照応の先行詞となっていると考えられるものがあり、これらについて推論によって理解できることを示す。また、名詞に関する照応関係の理解結果の応用について考察する。

[2] 名詞と名詞の照応の種類

具体的な照応の例を調べるために、日本語と英語のそれぞれを母国語とする話者2名と通訳者によるキーボード間対話^[Arita 87](英日各26対話、同一話者が続けて話した発話を1発話として各発話20前後)について照応関係の調査を行なった。この調査から照応関係にあった名詞の組の形態上、意味上の分類を以下に示す。

[a] 同一名詞の繰り返し

会議 - 会議
レディスプログラム - レディスプログラム
conference - conference
lady's program - lady's program

[b] 複合名詞の一部

通訳電話国際会議 - 会議
登録用紙 - 用紙
大阪空港 - 空港
京都プリンスホテル - 京都プリンス

registration form - registration form
Kyoto Station - station
sightseeing tour - tour
credit card - card

複合名詞の一部による照応の場合には、複合名詞の一番最後の語が使われることが多い。しかし、固有名詞の場合には、この例によらない。

[c] 上位の概念による

参加料 - 料金
登録費 - 費用
参加申込書 - 書類
registration fee - cost
breakfast - meals

[d] 同じものを指示する別の語による

申込用紙 - 登録用紙、参加申込書、...
登録費 - 登録料、参加費、参加料、申込料、...
費用 - 料金
期限 - 締切
住所 - 連絡先
妻 - 家内
application form - registration form
attendance fee - application fee
refund - cancellation
hotel - accommodation
paper - draft

[e] パラフレーズによる

使用言語 - 使われる言葉
ご質問 - おたずねしたいこと
研究内容 - 研究している内容
ホテル - 宿泊の場所
substitute - someone in my place
observer - those wishing only to observe

[3] 名詞の同一性の理解

[3-1] 概要

対話での照応に関する知識および判定基準となるルールの検証を目的として、名詞句の照応の理解のための実験システムを作成した。この実験システムは、エキスパートシステム^[Art 87]上にインプリメントされている。ここでは、この実験システムおよび、使用されたルールの概要について述べる。

最初に、ここで扱う照応関係について述べる。照応関係は単純名詞句の照応とし、対話の中にある名詞句を原則的に先行詞として解析するものとする。また、扱う照応関係は発話に現れた名詞Aと名詞BがどちらもCと呼べるものを指示している考えらる関係とする。そして、この関係を集合としての同一と呼ぶことにする。例えば、申込み書と用紙がどちらも参加申込み書と呼べるものを指示しているとき、集合としての同一と呼ぶことにする。また、これらが1枚の参加申込み書を指示しているとき、個としての同一であるとする。これらの個としての同一性の理解には、集合としての同一の理解が、前提となる。

この実験の入力は、意味解析された発話[Kogure 88]である。発話中の名詞は、対応する知識ベースの要素とリンクされて名詞句ごとに1つのエントリーを作成する。このエントリーには、この名詞が発話での位置の情報が含まれる。以後の解析はこのエントリーについて行われる。第一段階では、2つの名詞に対応する知識ベース(ここではドメイン知識と呼ぶことにする。)上の要素の関係を照合することによって、照応となるための条件を満たすかを判定する。これらは、すべての名詞の組に適用される。ここで与えられる条件は、ある名詞の先行詞の候補であるかを判定する。この条件は、例えば、英語の代名詞“He”の先行詞となるための条件の1つは、男性であり、単数であること

であるが、これらの条件の相当する。以後、候補の中から、さらに同定を行う。

[3-1-1] ドメイン知識

ドメインに依存した知識ベースとしてドメイン知識を作成した。会議登録に関するドメイン知識である。フレーム型の知識として記述し、スロットとして、格名に相当するもの、IS-A(上下関係)、HAS-A(部分/全体関係)、CAUSAL(因果関係)などを含む。

[3-1-2] 先行詞の可能性

ある名詞が他の名詞と照応できるための条件を可能性のための条件と呼ぶことにする。特に文脈によらない可能性の条件を前述した調査から分類しルールを作成した[Nogaito 87]。ここでのルールは、[fig-1]に示すドメイン知識の階層によるルールである。主に知識ベース上の概念の2項間の関係から可能性を与えるルールとなっている。[fig-2]に「2つの名詞に対応する概念がIS-A(上位/下位)の関係にあれば、先行詞の可能性がある。」に対応するルールを示す。この他、HAS-A(部分/全体)関係、CAUSAL(因果)関係などが検査される。また、2項以上のノードにおよぶルールも含まれている。これらのルールは、照応関係となる概念が知識ベース上で直接リンクを持たないことがあり、このため、知識ベースのいくつかのノードを検査するためのものである。例えば、登録料金と登録費用の双方は、料金とも費用と照応関係となるが、

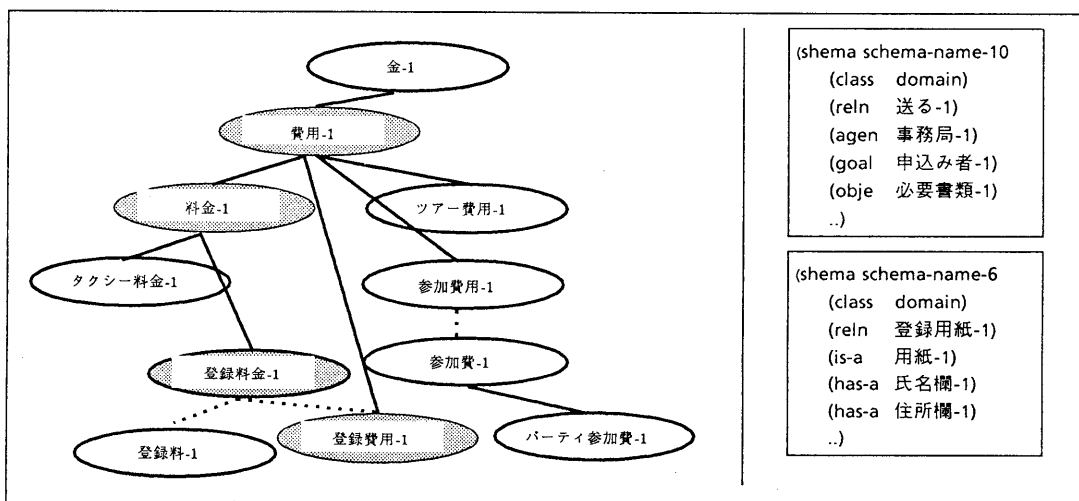


fig.1 ドメイン知識

費用に関連するドメイン知識の一部を左に示した。実線はIS-A関係を示し、破線はEQ関係を示している。左に具体的なドメイン知識の例を示した。

```

(defrule identify100
  "setting is-a-weak-ident"
  (schema 名詞1)           ;発話中の名詞1
  (schema 名詞2)           ;発話中の名詞2
  (schema 名詞1 (is-a 名詞2))
  =>
  (bind ?rel 'is-a)
  (assert
    (schema =(gentemp "weak-ident-")
      (class weak-ident)
      (rel ?rel)
      (arg1 名詞1のIDナンバー)
      (arg2 名詞2のIDナンバー))))
  ;IDナンバーは対話全体での識別ナンバー

```

fig. 2 ルール化した先行詞の可能性の条件

この関係は知識ベース上では、登録料金は、料金、また登録費用は、費用をそれぞれIS-Aの上位の概念として持ち、登録費用と登録料は、EQ(同値)関係で結ばれた関係に対応する。すなわち、登録費用と料金を直接結ぶリンクは、知識ベース上にない。このような例のために、「ある名詞に対して、そのEQで結ばれた概念のIS-A上位のものも、可能性がある。」というルールが与えられている。このように、いくつかのノードを検索するルールも含まれる。

[3-1-3] 候補の同定

先行詞の候補について、同定を行う。同定のための条件の内、代表的なものを以下に示す。

・名詞句の意味構造からの制限

・別のものを指示するマークが存在する。

最初の

前者の

・共起しえない属性がある。

宿泊のための費用

参加のための費用

・一貫性 ある語があるものを指示していれば、その語は、別のものを指示できない。

・ドメイン知識による

「(事務局)が(質問者)に(登録用紙)を送る」などの知識により推論および制限を行う。

・発話のタイプによる制限

否定文や疑問文の内、いくつかのタイプに関しては制限を与えている。

[3-1-4] 推論による照応

これまで、照応関係について、発話中の名詞句を中心に述べてきたが、ここでは、推論により発話の中に名詞として現れていないものを推論によって扱う。

照応における推論の深さに関しては次の4分類がある [Ishizaki 86]。

直接照応 先行詞が文脈中存在する照応
(太郎は、昨日大阪にいった。彼は阪大で開かれる学会へ出席する。)

1次推論型 先行詞が文脈中に明示的に表現されていないが文章中の名詞句や動詞句に含まれる標準的なスロットを指示する場合
(殺人事件が発生したが、まだ犯人は捕まっていない。)

2次推論型 文脈中の事象と因果関係にある事象や状態を指示する場合
(塩を水に溶かし、できた溶液を凍らした。)

高次推論型 2次推論型以上の高度な推論を要する場合
(太郎は山登りが大好きである。それが自分に挑む姿に引き付けられるのである。)

直接照応に関しては、より詳細化した分類し、その解釈の方法についても述べた。ここではその他の推論について述べる。

事務局: 会議で使われる言語は...。公式言語は...

この例は、調査の分類では、パラフレーズによる照応であり、1次推論型に含まれると考えられる。同様な例では、

申込者: 研究生が発表をしたいと言ってます。

事務局: わかりました。発表者は、...

発表者の例で具体的な推論を示す。「(人)が発表する」ならば「(人)は発表者である」というドメ

イン知識により、発表をするのが(人)の下位にあるかを照合して、発表者が対話上のエントリーとして導入される。導入されたエントリーは、照応の可能性を調べられ、後の発話の発表者と同定される。

次の2例は、ホテルの宿泊料金の話題に続く会話の例である。同じ話題に続く対話であるが、料金は、それぞれタクシー料金、ホテルの宿泊料金となる例である。ここでの推論は、2次推論型に含まれると考えられる。

事務局: 1泊シングルで、6000円から10000円です。

申込者: わかりました。ところで、京都駅からタクシーを使った場合、料金はいくらぐらいかかりますか？

この例では、料金は、直前のホテルの宿泊料金ではないが、この発話の代わりに以下の発話があると料金はホテルの宿泊料金となる。

申込者: わかりました。ところで、割引を使った場合、料金はいくらぐらいかかりますか？

前者の場合、ドメイン知識の先天的な知識「(交通手段)を使えば(費用)がかかる」「(費用)は(交通手段)に対応する」が与えられているとする。「(交通手段)を使えば(費用)がかかる」を発話の意味構造が満足させるかを判定する。発話の意味構造からタクシーを得て、ドメイン知識の階層性からこれが交通手段であることを得る。また、タクシーに対応する費用は、タクシー料金であることから、これによりタクシー料金が、対話のエントリーとして導入される。導入されたタクシー料金は、可能性などの照合の後、料金の先行詞となる。後者では、推論により、料金の先行詞となるものは、導入されない。

[3-2] 実験

先の調査と同じ対話より5対話、130発話(文に対応するものを1発話として、232発話)、他に修正ルールの再検証のための対話(発話数2から4ぐらい

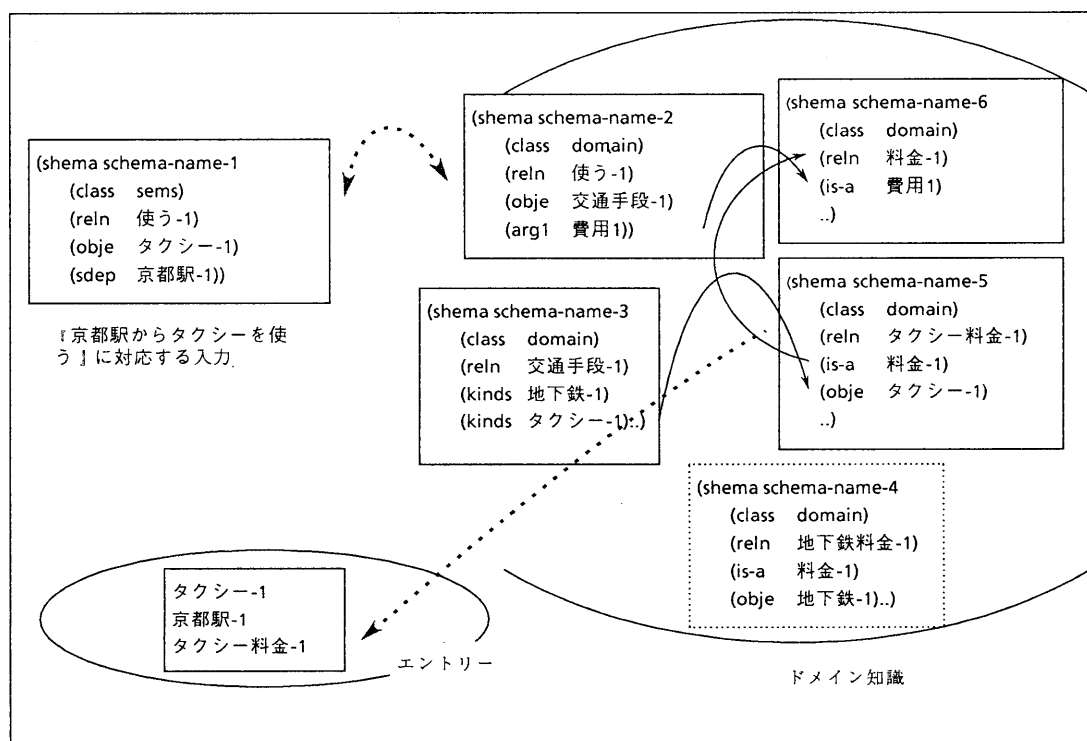


fig. 3 推論による同定

「京都駅からタクシーを使う」を意味解析結果より得たドメイン知識上の知識から、推論を行う。ドメイン知識上の知識については、省略部分、略記がある。

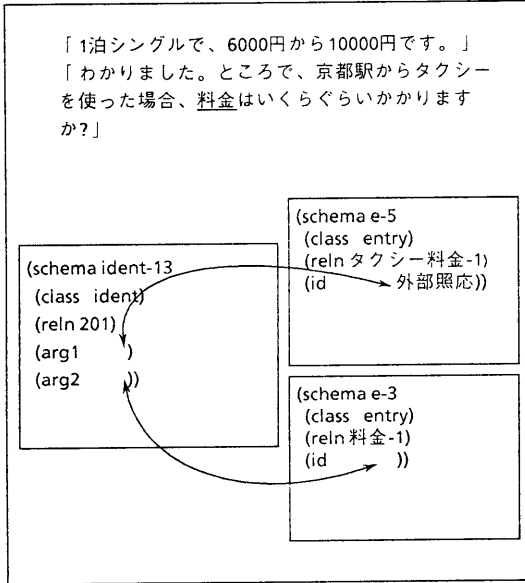


fig. 4 名詞句の同一性の解析の出力

料金の先行詞がタクシー料金であることを示している。左の2つの構造は、発話中の名詞に1対1に対応するエントリーである。右の構造は、この2つが照応関係であることを示している。

のもの)を10対話についてルールの検証を行った。指示的な名詞(両方、他など)を除く名詞について、次の発話を除いて照応関係を理解できた。

事務局: 今回の会議は国際会議ですので、使用言語は英語です。

前者の会議は特定の会議を示すが、後者の国際会議は、国際会議一般を示している。しかし、解析結果は、特定の会議を指示していた。これについては、「今回の会議」が「今回」により、特定の会議を指示すると考えてよいことと、「Aは、Bだ。」の形の発話について、AとBが同じものを指示している場合が、会議の登録に関する問合せにおいて考えにくいことなどから、知識ベースの記述により対応できると考えている。

しかし、総称的な名詞の用法は、「Aは、Bだ。」の形だけではなく、広範に現れ、名詞の用法に関しては、形態上もしくは意味的に総称的な用法であることがわからなければ、常に総称的な場合とそうでない場合の可能性を持つことになる。

[4] これまでの研究

照応の研究は多いが、次の2つについて考察する。これらは、いずれも本報告で扱った「集合としての同一」より詳しい「個としても同一」を扱うための手法である。

名詞の解釈について、プランナーを起動し、会話の対象を検索することで、指示および照応を理解する手法がある [Winograd 72]。この手法は、古典的な手法ではあるが、会話の対象すべてについて具体的な記述が可能である場合には、実際の、有効な方法と考えられる。しかし、本報告の対象とする対話では、会話の対象の具体的な記述は、できない。

力学の問題を入力し、それを問題解決するためのデータ構造に変換するための自然言語インターフェイスの中で照応を扱っており、物理的な知識などを記述して候補を制限している手法 [Mellish 1985] がある。

本報告の手法は、自然言語として発話や文に現れない情報をあらかじめ知識として用意しておき、これらを利用して、候補を制限するという意味で後者の方法に近い。後者では、題材とする力学の問題では、1つの名詞に対して、文として入力される情報が、多く、これらと物理的知識による制約条件により、この条件をより満たすものを推論により求めて、結果的に候補を制限をしている。また、定冠詞“the”をマーカとして、既知を判定している。本報告の手法は、発話内に照応とならない条件があるか、推論による新たな先行詞の候補の導入に重点がおかれている。

[5] 応用

ここでは、集合としての同一性の理解の応用について述べる。

[5-1] 変換および生成

言語間で名詞が示す概念上の差があるとき、訳語選択のために照応関係が有効である [Nogaito 88]。

例えば、用紙と対訳の候補である **form** の語義を比べると用紙は、何かの目的のための紙であり、**form** は、線などで区切られた答えなどを記入する紙である。つまり、**form** は記入するための線を持つことが必要であるが、用紙は特に、必要としない。したがって、用紙が何の用紙であるか、それが線を持つかを調べなければならない。

用紙が照応関係にあり先行詞が、登録用紙である

とすれば、登録用紙の訳語からregistration formを得る。用紙の対訳候補の中からこのregistration formと照応関係を持つ語を知識ベースから得てこれを対訳とする。このように対訳を選べば、対象言語でも照応関係が成立し、対象言語側で先行詞を見失わない。

[5-2] 内部データとの照合

機械翻訳システムには多くの内部データが存在する。照応関係による名詞の変化により、これらの内部データと解析された発話のデータ構造とは、単純な照合が、できないことがある。ドメイン依存の理解を行う場合、これら詳しく特定の名詞が理解のためのキーワードとなることが多い。例えばホテル宿泊料金からホテルの予約に関する何かの発話であることが理解できるが、発話中では、これに照応する料金が使われていると、料金だけでは、ホテルの予約、観光、交通手段のどの話題に関する発話なのか理解できない。これは、その名詞と照応されたエントリから内部データと照合できる。

[5-3] 定名詞句の理解

日本語のように、定/不定のマーカをもたない言語からマーカを持つ言語への翻訳を考慮すると、定/不定の解釈は必要である。「定(definite)」は、話者と聞き手にとって、同定できると定義できるが、既出であることであるときには、この定義を満たす。この既出であることとある名詞と同じ名詞が、対話中に存在することは、異なる。同じ用紙が、たびたび使用されても、一方は参加申込み用紙であり、他方は予稿集の購入申込書であるかもしれない。

逆に、同じものであっても、名詞が言い替えられるために、単純な照合はできないことが多いので、照応の解析が前提条件となる。照応されたものがある場合には、既出のものがあるので定である。なお、“the”の生成に関しては、より詳しい考察が必要である。

[6] さいごに

名詞句の同一性の理解の手法を示した。および、この手法をエキスパートシステム上で実現し、いくつかの対話について処理実験を行った。

英語における照応理解の手法は、定冠詞や不定冠詞の有無に処理の手順が依存している。これらの

手法をそのまま、無定冠詞で名詞を表現する日本語には適用できないことが多い。表層のマーカなしにその名詞が既知であるか未知であるか決定しなければならない。しかし、これらは、決定できないことが多く、結局、候補数は多くなり、誤った結果を得ることになりやすい、このため適切な方法でできるだけ候補を少なくする必要がある。

ここでは、会議の申込みに関するドメイン知識と推論により制限をおこなった。物理の力学の問題などに比べ、かならずしも十分な情報を持たないと考えられる対話においてもドメインに応じて照応の候補の同定のために制限的に働くルールや推論規則や知識を記述することは可能で、協調的な対話においては、名詞の同定に有効であることが、予測できた。

謝辞

本研究の機会を与えて下さるとともに、適切な助言を述べられたATR自動翻訳電話研究所 樽松 明 社長、同言語処理研究室 相沢輝昭 室長、また議論に参加してくれた言語処理研究室の方々に感謝します。

参考文献

- [Arita87] 有田,飯田,「日本語におけるタスクオリエンティッドな対話の構造」,電子情報通信学会, NLC-87-10,1987
- [ART87] ART Reference Manual, Inference Corporation, 1987
- [Brachman83] Brachman R: What IS-A Is and Isn't: Analysis of Taxonomic Links in Semantic Networks,"COMPUTER",Oct 1983, IEEE
- [Ishizaki 86] 石崎,井佐原,「文脈処理技術」,情報処理学会, Vol. 27 No.8,1986
- [Kogure88] Kogure K: A method of analyzing Japanese speech act types, The 2nd International Conference on theoretical and Methodological Issues in Machine Translation of Natural Languages,Jun. 1988
- [Mellish 85] Mellish. C: Computer Interpretation of Natural Language Descriptions, Ellis Horwood, 1985 (田中穂積訳、コンピュータのための自然言語理解の基礎、サイエンス社)
- [Nogaito87] 野垣内,「照応・指示関係の同一性の分類・解析」,人工知能学会研究会資料SIG-FAI-8701-1,1987
- [Nogaito88] NOGAI, "Noun Phrase Identification in Dialogue and its Application", 2nd International Conference on Machine Translation, CMT,CMU,1988
- [Winograd 72] Winograd. T: Understanding Natural Language, Academic Press, 1972

Dialogue	Identification
<p>1 お尋ねしますが通訳電話国際会議事務局ですか？ 2 はい 3 そうです 4 会議に参加したいのですがどうすればよいですか？ 5 先ず登録用紙にて手続をしていただかなくてははいけませんがお持ちでしょうか 6 いいえ 7 ではこちらから用紙をお送りいたします 8 ご住所とお名前をお願いします 9 大阪市東区徳井町の二階水次郎です 10 はい 11 分かりました 12 それでは早速送らせていただきます 13 参加費用はいくらですか？ 14 16000円です 15 発表したいと思いますがどうすればよいですか？ 16 発表申込み用紙を同封致しますのでそれに必要事項を記入してまず送って下さい 17 はい 18 分かりました 19 発表は日本語ですかそれとも英語ですか？ 20 両方とも可能です 21 当日は同時通訳が付きますのでどちらでも結構です 22 はい 23 分かりました 24 分科会はいくつぐらいありますか？ 25 八つの分科会がございます 26 分科会の発表で聞けないものについてはコピーをもらえますか？ 27 論文系を最後にお渡ししますのでそれでご勘弁願います 28 はい 29 どうもありがとう 30 差を連れて行きたいと思いますが参加費用はいくらですか？ 31 会議に参加されないのであれば無料で結構です 32 レディースプログラムはありますか？ 33 はい 34 ございます 35 同行の二夫人も多いでしょうか？ 36 そうです 37 多分たくさんのお夫人がお見えになると思います 38 ではレディースプログラムの方も参加申込みます 39 はい 40 大歓迎でございます 41 では発表申込み用紙を待っています 42 締切はいつですか？ 43 十二月二十五日までですのでお急ぎ下さい 44 はい 45 分かりました 46 どうもありがとう 47 宜しくお願いします 48 こちらこそ宜しくお願い致します 49 さようなら</p>	<p>レディースプログラム-1/3201 レディースプログラム-1/3801 202 夫人-1/3702 夫人-1/3502 202 参加費用-1/3802 参加費用-1/1301 202 分科会-1/2401 分科会-1/2502 202 分科会-1/2601 分科会-1/2502 202 発表-1/1501 発表-1/1901 202 発表-1/2602 発表-1/1901 202 発表申込み用紙-1/1601 発表申込み用紙-1/4103 202 用紙-1/701 登録用紙-1/501 202 会議-1/401 会議-1/3101 202 会議-1/401 会議-1/3101 202 会議-1/3101 会議-1/401 202</p>
<p>Show All Ident L2: Window menu.</p>	<p style="text-align: center;">RELATIONS</p> <p>Your options are:</p> <ul style="list-style-type: none"> Refresh Board Window Refresh Ident Window Show All Entry Show All Weak-Ident Show Weak-Ident Show All Ident Show Ident Show All Ident-in Show All Not-Ident

CL ACU: (no window)

fig. A 名詞句の同一性のための実験システム