

辞書情報を利用したコンピュータによる間接的照応関係解析の試み

An Indirect Anaphor Resolution on Computer by use of Vocabulary Information of English Dictionaries

良峯 徳和(Norikazu Yoshimine)

多摩大学グローバルスタディーズ学部(School of Global Studies, Tama University)

1. はじめに: 自然文を論理計算可能な形式に変換

筆者は近年、比較的平易な英語文を、SNePSと呼ばれる意味ネットワークシステム上で論理的、規則的に計算処理できる形式に変換するためのプログラム SNaLC(SNePS-Natural Language Converter)の開発を行ってきた。SNaLCは、通常の自然文を Stanford Parser によるパーズング結果を介して、SNePS 構文に変換し、論理処理の対象とすることができる。

これにより、人間が自然文を介して行っているさまざまな言語活動を、ある程度は計算機上で疑似的に実現できると期待される。とりわけ、人間が自然言語で表現されたテキストや発話をどのように処理し、理解しているかといった自然言語理解の認知プロセスのさまざまな側面について、分析したり、シミュレーションする有効な道具となりうるだろう。

そうした試みのひとつとして、人間のための書かれた一般の辞書(英語)を、自然言語処理のための知識ベースとして利用することで、意味理解なしには困難とされてきた間接的な照応関係を特定するという実験を行なった。

2. 結束性と照応関係

英文などで、先行する文の中で用いられている名詞や動詞などの単語や句が、後続の文中で指示代名詞や代動詞、あるいは指示代名詞+名詞などで置き換えられて文を構成している場合、置き換えられた語が先行文の何を置き換えたものかを特定することは、比較的簡単である。なぜなら、文の表層に現れる単語の語尾変化やジェンダー、近接性、語順などによって、先行詞の範囲が強く制約を受けるため、大概の場合、特定の候補者のみが選択対象となることが多いからである。実際、こうした表層的な手掛かりだけで、90パーセント以上の照応関係を特定することが可能とされる。こうした表層的な手掛かりだけでほぼ先行詞を特定することができる種類の照応関係は、直接照応(Direct Anapher Solution)と呼ばれる。

これに対し、そうした表層的な手掛かりだけでは、どれが先行詞なのかを特定することが困難な種類の照応関係も存在する。それらは「間接的照応関係(Indirect Anapher Solution)」もしくはそれを推論、推定する視点から「橋渡し推論(Bridging Inference)」とも呼ばれている。こうした間接的照応関係に分類される英文の例を以下にあげる。

間接的照応関係を含む英文例

Alphons Clein noticed a blue Mercedes. He had the impression that the driver was lying on the wheel. At that particular moment he noticed that the man was dead. His temples were shot through.

(アルフォンス・クラインは、1台の青のメルセデスに気付いた。その運転手はハンドルにもたれかかっているかに見えた。その瞬間、彼はその運転手が死んでいるのが分かった。彼のこめかみが打ち抜かれている)

【一重のアンダーラインは直接照応関係にある語句、二重のアンダーラインは間接的照応関係にある語句を示す】

この例文には、照応関係をなしている組み合わせが2種類存在している。2番目、3番目の文の"he"は、最初の文の人物名"Alphons Clein"を指示し、その代名詞として使用されている。これは直接的な照応関係である。また、2番目の文に登場する"the driver"は、最初の文の"a blue Mercedes"で表されている自動車の運転手であり、同じく"the wheel"は、その自動車のハンドルのことであろう。第3文の"the man"は、第2文の"the driver"を指し、同様に第4文の"His temple"は、その"driver"の"temple"のことになる。

上記の自然文の意味を理解できるのであれば、このような照応関係は当然理解されているだろう。とはいえ、文の意味を理解するアルゴリズムを実装していない直接照応関係推定プログラムであっても、最初の直接的な照応関係(第2文以降の"he"が"Alphons Clein"を指示していること)を、かなり高い確率で言い当てることができる。なぜなら、第2文のHeの先行詞となりうる語は、第1文には、人物の男性固有名詞に属する"Alphons Clein"しか存在せず、第1文の主語、第2文の主語がともに同一人物"Alphons Clein"を指示しているのならば、第3文の男性形代名詞"he"もまた、"Alphons Clein"を指示していると、かなり高い確率で予想されるからである。

しかしながら、"blue Mercedes"と"the driver"、"blue Mercedes"と"the wheel"との間の照応関係については、"this"、"that"といった指示句が用いられておらず、動詞の目的語であったり、主語であったり、修飾句の中の目的語であったりと、文法的な役割もバラバラである。それらがともに単数形の名詞であるということ以外、それらの関係を示す表層的な手掛かりはない。ゆえに、一般的な照応関係推定プログラムでは、こうした語句間の照応関係の推定は困難である。

とはいっても、この文をきちんと理解するには、これらの語の関係性を把握することが、重要かつ必須である。"blue Mercedes"、"driver"、"wheel"の間に相互の関係がなく、突如、同じ文脈に登場してきた無関係の個人、個物だとしたら、文脈としての結束性が希薄になり、文の理解に必要な背景状況の構築ができなくなってしまうからである。

3. 一般の辞書を使い、語句の情報を得る

では一般的な常識をもたないコンピュータが、どのようにして上記のような文の理解に必要な情報を得ることができるだろうか。比較的単純な自然文で書かれている常識的知識の情報源として、各種辞書(English Dictionaryなどのいわゆる語学用辞書)があげられる。辞書はいわば人間が歴史をへて積み重ねてきた言葉に関する常識の最良の知識ベースといえるだろう。とはいえ、それらの辞書は、常識を持った人間のみを読み手として想定し、作られているため、コンピュータがそのまま利用することには、不向きであった。しかし、SNaLCの機能は、そうした状況を和らげてくれるものと期待される。

今回の実験では、すでに電子化され、インターネット上で自由に利用できる2種類の英語辞書をサンプルとして用いる。ひ

とつは、WordNet であり、もうひとつは、Wiktionary である。Wiktionary については、WordNet に収録されていない固有な名、ブランド名などの意味を検索することを目的に、補助的に使用することにした。

4. 複数の定義候補をいかに扱うか

同じ言葉であっても、しばしば複数の定義や意味が付与されている。辞書でも同じ見出し語に対し、複数の定義や説明が記載されている。これをどのように扱うかが、コンピュータで人間用の辞書を利用する際にもっとも注意を要する点である。再び“driver”を例にとると、WordNet では、以下のように5つの意味が分類して掲載されている。

WordNetによる“driver”の検索結果

The noun driver has 5 senses (first 1 from tagged texts)

1. driver -- (the operator of a motor vehicle)
2. driver -- (someone who drives animals that pull a vehicle)
3. driver -- (a golfer who hits the golf ball with a driver)
4. driver, device driver -- ((computer science) a program that determines how a computer will communicate with a peripheral device)
5. driver, number one wood -- (a golf club (a wood) with a near vertical face that is used for hitting long shots from the tee)

例えばこの“driver”という語を、これらの5つの意味ごとに区別して扱うため、それぞれ“driver@wn\$1”、“driver@wn\$2”などとして表記することにする。“@wn”は「WordNet 辞書による」ことを意味し、“\$n”(n は自然数)は「何番目の定義であるか」を示す。また、WordNet の場合、語のすべての意味は、“synset”という概念で分類されており、固有の概念番号が付与されている。これを SNePS の命題表現で表すと、以下ようになる。

WordNetによる“driver”の検索結果をSNePS表現する

```
(assert superclass driver subclass driver@WN$1)
(assert source WordNet synset driver@wn$1 ID wn09885824)
(assert superclass driver subclass driver@WN$2)
(assert source WordNet synset driver@wn$1 ID wn09886348)
(assert superclass driver subclass driver@WN$3)
(assert source WordNet synset driver@wn$1 ID wn09886232)
(assert superclass driver subclass driver@WN$4)
(assert source WordNet synset driver@wn$1 ID wn06487211)
(assert superclass driver subclass driver@WN$5)
(assert source WordNet synset driver@wn$1 ID wn03210301)
```

Wiktionary (およびそれ以外の一般的な辞書) の場合には、WordNet の場合のような synset 番号が付与されていないので、その部分を省略して表現することにする。

辞書で表現されている個々の概念をクラス(存在の集合)とすれば、それに属する任意のインスタンス(以下の例では d1 ~d8)は、辞書のそれぞれの記述をみたすことになる。例えば、WordNet の名詞“driver”の最初の意味、すなわち“driver is the operator of a motor vehicle”に対し、SNaLC を使って、SNePS 表現に変換したものを辞書情報に付加すると、以下になる。

“driver is the operator of a motor vehicle”を
SNePS命題表現¹に変換 (簡略化のため一部のコードを省略)

¹ SNePS 構文中に現れる d1 のような、アルファベット 1 文字と数字の組み合わせで表現されているのは、個々のインスタンス名である。インスタンス名は任意で付与することができる

```
(assert superclass driver subclass driver@WN$1)
(assert source WordNet concept driver@wn$1 ID wn09885824)
(assert instance d1 class driver@WN$1)
(assert instance d2 class be)
(assert instance d3 class operator)
(assert instance d4 class of)
(assert instance d5 class motor)
(assert instance d6 class vehicle)
(assert instance d7 class or)
(assert instance d8 class motor-vehicle)
(assert hasIDT (d1 d3) IDT a )
(assert has-prep-p d3 prep-p (build prep-obj (d6 d8) prep d4 ))
(assert adj-obj d6 adj d5)
(assert has-conj (d6 d8) conj-node d7)
(assert VP (build noun d3 verb d2) subject d1 )
```

辞書データを、このような SNePS の命題表現に変換することで、その情報をこの世界に存在するさまざまな物事や出来事に関する知識ベースとして利用可能な形に変換することができる。辞書から得られる情報は、個々の具体的な対象に関する情報とは異なり、あるクラスに属するインスタンスの一般的な特徴や属性に関する情報であるため、具体的な自然文を変換した場合とは違い、具体的な成立状況を設定する状況インスタンスや時制に関する属性は付加されない(ただし、この定義を実際の文に適用する際には、その文で指示されている状況インスタンスを付与することになる。)

5. Synset に関する情報(上位語・下位語)を付加

“driver”という見出し語で WordNet を検索した際、もっともこの文を理解するために適切と思われる定義は、上記5つの選択肢のうち、最初の “driver -- (the operator of a motor vehicle)” になる。一方“Mercedes”の語彙の方を Wiktionary で検索した際、最も適切と思われる意味は、“A car manufactured by Mercedes-Benz”となる。両者の語彙間に、何らかの論理的関係(同義やクラス-インスタンスの関係など)を見出すには、前者における“motor-vehicle”という表現と後者における“car”という関係にどんな関係があるかを見極める必要がある。この間に何の論理的関連性も見出せないならば、この文における間接的照応関係の同定は失敗することになる。

異なる二つの語の関係を見つけ出すためには、シソーラスからの情報を得て、比較を行う必要がある。WordNet の場合、一般のシソーラスと同様の情報を、synset 間の hypernym-hyponym (上位語-下位語) の関係として検索することができる。ここから見出し語のみを抽出して、同義語、上位語-下位語関係を列記した SNePS による命題表現に変換する。

SNePS表現による “motor-vehicle” の同義語、上位語-下位語関係の列記 (一部省略)

```
(assert synonym motor-vehicle@WN$1 synonym automotive-vehicle@WN$1)
(assert hypernym motor-vehicle@WN$1 hyponym (auto@WN$1 automobile@WN$1 machine@WN$1-4 motorcar@WN$1 four-wheel-drive@WN$1-1 4WD@WN$1-1))
(assert source WordNet concept (motor-vehicle@WN$1 automotive-vehicle@WN$1) ID wn03749282)
(assert source WordNet concept (car@WN$1 auto@WN$1 automobile@WN$1 machine@WN$4 motorcar@WN$1) ID
```

め、SNaLC では自動で振り当てるようにプログラムしてある。ここでは、d で始まるインスタンス名(d1 など)は、辞書の説明文に登場する存在、b(b1 など)で始まるインスタンス名は、もとの自然文に登場する存在を指し示すものとして、区別している。

wn02929975)
(assert source WordNet concept ambulance@WN\$1 ID
wn02676307)

ここには、WordNet 上の"motor-vehicle"という見出し語の第1番目の概念と、上位-下位概念の関係にある WordNet 上の見出し語とが、列挙されている。しかしながら、通常、言葉は、辞書に挙げられている意味番号を付して使用されることはない。例えば、ある場面で、"motor vehicle"という言葉に、"motor-vehicle@wn\$2"ではなく、"motor-vehicle@wn\$1"に限定した意味を持たせたいとしても、話し手は"motor vehicle"という言葉を使って、それを言い表すことになる。ようするに、"motor-vehicle"という言葉は、そのうちの一部の意味を表す一種の換喩(metonymy)として使用されているのである。

したがって「ある名前を持った具体的対象(インスタンス)の属性が、辞書に載っている複数の意味のどれかに該当する場合、その対象をその名前と呼ぶことは、ある程度の蓋然性を伴うが、適切な呼び方でありうる(なぜならば、その言葉が意味しうるいくつかの概念のうち、少なくともひとつについては、満たしているからである)」という規則を定式化する。ここでは、このような蓋然性を伴う推定状況を SNePS の命題表現を使って表すために、"(assert modal possible state (build ~))"という構文(自然文で表すと、"It is possible that ~")を導入した。

同様に「辞書のふたつの見出し語(例えば、"motor-vehicle"と"car")の間に、上位-下位関係が見出されるとき、前者の言葉で呼ばれる対象は、ある程度の蓋然性は伴うが、適切な呼び方でありうる」を SNePS 構文で定式化する。

辞書定義に蓋然性を付与するルールのSNePS表現

```
(forall ($A $B $C)
  &ant ((build instance *A class *B)
        (build hypernym *B hyponym *C))
  cq (build modal possible state (build instance *A class *C)))
[AがBのインスタンスで、辞書の見出し語BとCの間に上位-下位関係(hypernym-hyponym)が成り立っているならば、AがCのインスタンスである可能性がある]
```

一方ですでに見たように、"driver"には WordNet によって5つの定義が与えられている。上記の議論から"driver"と呼ばれる対象には、これら5つの意味のどれかが蓋然的に当てはまることになる。まず、上記の1番目の辞書記述により、"It is possible that the driver is the operator of a motor-vehicle (1)"を表す SNePS 表現形式の命題が生成される。さらに、2番目の記述により、"It is possible that the driver drives animal that pull a vehicle (2)"という SNePS 表現形式の命題が生成される。(3番目以降も同様に変換・生成される。)

(1)の場合、すでに例文および"motor-vehicle"についての WordNet および Wikitionary の記述から、"Mercedes"が"car"であり、さらに"car"は"motor-vehicle"である蓋然性が導出されている。ゆえに、(1)は、"It is possible that the driver is the operator of car"と書き換えられる。さらに、WordNet の"driver"に関する記述("driver" ⇔ "a operator of motor-vehicle")によって、"operator"の対象が"motor-vehicle"の場合には、"operator"は"driver"と置き換えることができるため、(1)は、"It is possible that a driver of car exists"を含意²していることになる。

²先に見た"driver"に関する WordNet の定義 1 ("driver" ⇔ "a operator of motor-vehicle")は、以下のように解釈、定式化される。
motor-vehicle(x) ∧ operator(x) ⇔ motor-vehicle(x) ∧ driver(x)。
(もし motor-vehicle とその operator が存在すれば、その operator

しかし(2)の場合、"animal that pull a vehicle"に対応する語や句が、辞書の記述にも、元の文にも存在しないため、置き換えが成立しない。3番目以降の辞書の意味についても同様に置き換えは成立しない。こうして、(1)の辞書記述のみが、もとの自然文に適用されることになる。その結果、元の自然文の第1文と第2文に、意味の共通項が見出されることになる。

6. 間接的照応関係の蓋然的導出

以上のように、辞書から得られる情報を元の文に適用する際には、蓋然的な性格が伴わざるをえない。ある語彙の意味を知るために辞書引きをする際、複数の定義が挙げられていれば、どの定義が当の語彙に当てはまるのかは、実際にその定義をあてはめてみて得られた自然文の解釈の適切性をもって判断するしかない。場合によって、その文章だけで判断ができないときには、前後の文の内容をも含めて、文脈としての適切性を考慮する必要がある。

今回の実験で用いた SNePS 表現では、例えば、以下のように辞書から得られた具体的存在に関する情報は、すべて(assert modal possible state (build [個々の具体的記述]))という記述形式で表現される。

加えて"Mercedes"と"the driver"の照応関係を導出するための規則を設定する。すなわち、"Mercedes"が「自動車(car)」のことを意味し、"driver"が「自動車の操作者(operator of car)」のことを意味すると仮定したうえで、与えられた自然文の情報の範囲内から、両者が同一の自動車インスタンスを指示していると、矛盾なく解釈されるならば、先の仮定は(たんに可能というだけでなく)もっともらしい(plausible)仮定であるといえる、という命題を導出する規則を設定する。

辞書から得られた"Mercedes"に関する情報では、それが女の子の名前であるとか、会社名であるといった可能性もあるとされるが、このルールを設定することで、そうした可能性は棄却される。また、"driver"に関する辞書情報においても、"driver"がゴルフクラブの一種であるとか、台車を引く家畜の誘導者であるといった可能性が述べられているが、そうした可能性も、このルールによって棄却される。その結果、唯一、"Mercedes"が自動車のインスタンスであり、"driver"がその自動車の操作者インスタンスであるという仮定のみが、この文を解釈する際の蓋然性の高い仮定として保持されることになる。

7. 論理処理プロセスの実装の試み

以上の論理処理プロセスを SNePS システム上で実装する際、その要となるのは、自然文の SNePS 表現における状況インスタンスの値をすべて同一(以下に示す実装例では C1)に設定する点にある。というのも、同じ状況インスタンス値をもつ命題間では、相互に論理的な制約が働くからである。

たとえまったく同じ出来事が記述してあっても、それらが生じた状況インスタンスが異なれば、違った時空間で生じた事象とみなされ、共存が可能となる。逆に、矛盾した出来事であっても、状況インスタンス値が異なればやはり共存が可能と

は driver と呼ばれる)

いま、"motor-vehicle"とその"operator"の存在 x が前提されているため、operator(x)を driver(x)で置き換えることができ、driver(x)の存在が推定される。

これを辞書の定義から導かれる"driver"に関する一般規則として、SNePS 表現に変換し、元の自然文に適用することで、この推論結果が導き出される。

なる。しかし、状況インスタンス値を等しく設定すると、同じ状況インスタンス値をもった記述の間に論理的な矛盾が許容されなくなる。さらに、相互に何の関連もない複数の出来事や事象の記述に関して、同じ状況インスタンス値を付加すれば、たまたま同じ時空間で生起していると解釈されることになる。

一般に人は、いくつかの偶然が重なりあって生起することを認めたがらない。むしろ、偶然が同時に重なりあって生起しているようにみえるのは、それらの間に表に現れていない何らかの関連性があるからだと考えたがる。今回の例文では、"a blue Mercedes"と"the driver"が、連続した文脈で登場する。一般に読み手は連続した文脈には、何らかの結束性、関連性が存在しているはずだと推測する。加えて、指示詞を伴った"the driver"という語句の使用には、必ず先行する照応対象がなければならないという、文理解に際しての強い要請が伴う。

しかしながら、常識を欠いているコンピュータの場合、与えられた自然文だけでは、その要請を満たす媒介項を見つけ出すことができない。そこで今回の実験では、人間における常識の代替として、辞書から得られさまざまな情報を、要請を満たす可能性をもった情報として付加したのである。その結果、それらの情報の中に、"Mercedes"と"driver"とを媒介し、結びつけることのできる情報が見出されることになった。

これらの情報は、文脈の要請する関連性、結束性の要求に応えてくれるがゆえに、他の情報よりもより確からしい情報として受け入れられる。これが、今回の実装実験した推論処理プロセスの概要である。このような推論処理プロセスを導入したSNePSコードを実行した結果、以下の結果が得られた。

SNePSコードの実行結果：間接的照応関係の導出

(M38! (MODAL PLAUSIBLE)
(STATE
(M37 (HAS-PREP-P B13) (PREP-P (M36 (PREP D5) (PREP-OBJ B6))) (SITUATION C1))))

この SNePS 表現は、SNePS の論理演算の結果、38 番目の命題 (M138)として、「状況インスタンス値 C1 のもとでは、対象 B13("driver"のインスタンス)と対象 B6("Mercedes"のインスタンス)とが、前置詞 D5("of")によって結合した状態が、もっともらしい」ことを示している。

8. もうひとつのケースと今後の課題

例文中には"Mercedes"と間接的照応関係にあるもうひとつの語"the wheel"が存在している。WordNet によると、"wheel" (名詞)には、7個の定義が存在している。この文を適切に理解するためには、どの定義が"wheel"の意味として使用されているかを確定しなければならない。先の例では car という概念が媒介となっており、"Mercedes"と driver という単語が間接的に照応していた。これと同じように、"car"という概念を媒介としただけでは、"wheel"定義の候補として「車輪(wheel)」と「ハンドル(steering wheel)」の二つの候補が残ってしまう。

しかし、例文中には「その運転手はハンドルにもたれかかっていた("driver was lying on the wheel")」という表現が含まれており、これにしたがって"driver"と"wheel"の間に成り立つ空間的・物理的条件を考慮に入れることで、常識ある読み手には、"wheel"とは"steering wheel"のことだという推測が成り立つ。

したがって、今回の例文内容から"wheel"の意味を確定させるためには、通常の辞書には書いていない以下のような常識的知識を付加することが必要となる。

"the driver was lying on the wheel"を

可能にする条件としての常識的知識：

- (a) "driver"と"wheel"間に物理的障壁がない、距離が近い。
- (b) "wheel"は"driver"よりも下方、もしくは同じ高さにある。

加えて、こうした条件が満たされるかどうかを判定するため、次のような常識的知識も必要となる。

「車輪」および「ハンドル」としての"wheel"、および運転手に関する常識的知識記述例

- (I) 「車輪」としての"wheel"は、一般に車のボディの外側の前側と後側の左右に各 1 個、全部で 4 個存在する。
- (II) 「ハンドル」としての"wheel"は、一般に車のボディの内側の車室の前方に位置する運転席と車室の最前面に位置するフロントガラスの中間に、1 個存在する。
- (III) 「運転手」は、一般に車のボディの内側の、前方に位置する運転席に座る姿勢で、1 名存在する。

上記(a)(b)、および(I)と(III)から、"driver"がハンドルとしての"wheel"にもたれかかる(be lying on)ための 条件を満たしていることが導かれる。他方、(II)と(III)からは、車輪としての"wheel"は、"driver"がもたれかかるための条件を満たさないことが導かれる。その結果、この例文を適切に理解するための"wheel"の意味とは、自動車のハンドルとしてのそれではないことが判明する。

しかしながら、このような形で、間接的照応関係の同定、および複数の辞書定義から適切な意味情報を選択する精度を、さらに高めるためには、何らかの方法で、ここで例示したような常識的知識をシステムに付加してやる必要がある。前者については、前置詞や動詞、副詞などが実際の文脈で果たしている役割を分析し、整理してデータベース化することによって、ある程度の実装化を進めていくことができると期待される。また、後者の情報については、図や画像を使った関係性の解析など、マルチモーダルな手段をも含めた総合的な情報抽出の方法が、今後必要になってくると思われる。

参考文献

- Christiane Fellaum(ed.) (1998): *WordNet: An Electronic Lexical Database*, The MIT Press, Cambridge, Massachusetts.
- Dmitriy Dligach. SNePS and WordNet: Using WordNet as a Source of Background Knowledge in Contextual Vocabulary Acquisition. (<http://www.cse.buffalo.edu/~rapaport/CVA/dligach-wordnet-final-report.pdf>)
- Monika Schwarz-Friesel (2007): Indirect anaphora in text A cognitive account. In Schwarz-Friesel, Monika. Consten, Manfred. Knees Mareille (Eds). *Anaphors in Text : Cognitive, Formal and Applied Approaches to Anaphoric Reference*. Amsterdam,NLD: John Benjamins Publishing Company.
- Ruslan Mitkov (1999): Anaphora resolution: the state of the art, Working paper (Based on the COLING'98/ACL'98 tutorial on anaphora resolution). University of Wolverhampton, Wolverhampton.
- Stuart Shapiro, The SNePS Implementation Group (2008): SNePS 2.7 User's Manual, University of Buffalo, The State University of New York.
- Tomoko Matsui (2000): *Bridging and Relevance*. Philadelphia, PA, USA: John Benjamins Publishing Company.