

人間の照応解決過程の計算モデル — “即時性” 仮説の妥当性の検討を中心に —

中島輝明*

阿部純一**

+北海道大学大学院文学研究科・日本学術振興会特別研究員

++北海道大学文学部心理システム科学講座

〒060-0810 札幌市北区北10条西7丁目

{tnaka, abe}@psych.let.hokudai.ac.jp

あらまし

文章の読みのいつの時点で、どの程度、代名詞の解釈が進められているのか、という問題に対して、人間の照応解決をシミュレートすることが可能な計算モデルを提案する。これまでの心理学的研究では、代名詞が即時的に解釈されていることを示唆するデータが報告されている一方で、即時的な解釈はつねに行われるものではなく、即時的な解釈を照応解決の一般的な原理とするには妥当性が低い、との主張もなされている。本研究では、代名詞の解釈に対して読み手が有する複数の選好性を仮定し、これらの選好性の相互作用によって、読み手の照応解決を説明するモデルを構築した。シミュレーションの結果から、本モデルは、ある特定の条件下において読み手が代名詞を即時的に解釈することを説明することができ、また同時に、文章の読みとともに代名詞の解釈を漸進的に進めていく人間の特性を説明することが可能であることが示された。

キーワード： 照応，代名詞，即時性の仮説

A computational model of human anaphoric resolution

Teruaki NAKAJIMA* and Jun-ich ABE*

+Research Fellow of the Japan Society for the Promotion of Science

*Department of Psychology, Hokkaido University

{tnaka, abe}@psych.let.hokudai.ac.jp

Abstract

This paper describes a computational model that can simulate human anaphoric resolution. One controversy in the study of human anaphoric resolution is when and to what extent an anaphor is interpreted in the process of reading a discourse. While several psychological studies suggest that a pronoun is immediately interpreted at the moment when the pronoun is encountered by readers, some researchers argue that a pronoun is not necessarily interpreted immediately. In the present study, we proposed a model in which several preferences regarding to pronoun interpretation interact with one another in the time course of reading a discourse. The results of the simulation suggested that our model can explain not only the immediacy of pronoun interpretation in a particular context but the incremental property of human pronoun interpretation.

Key words: anaphora, pronoun, immediacy hypothesis

1 問題と目的

照応 (anaphora) とは, “代名詞や名詞句などによって, 文脈あるいは環境の中に存在する ‘もの’ や ‘こと’ を参照する現象” である (阿部・桃内・金子・李, 1994)。通常, 我々は代名詞を認定したとき, その指示対象を容易に特定することができる。人間の照応解決過程を解明する上での重要な問題として, 文章を読むいつの時点で, どの程度, 代名詞の指示対象の特定が進められているのか, という問題がある。従来, 計算言語学の分野では, focusing theory (例えば, Sidner, 1983) や centering theory (例えば, Grosz, Joshi, & Weinstein, 1995) などの照応解決の理論が提案されている。近年では, これらの理論の心理学的妥当性について検証を試みた研究が報告されてきている (例えば, Chambers & Smyth, 1998; Gordon & Chan, 1995; Gordon, Grosz, & Gilliom, 1993; Kennison & Gordon, 1997)。計算言語学的な理論を人間の照応解決を説明するモデルとしてとらえようとした場合, 次のような問題点を指摘することができる。これらの理論では, 基本的に, 解析器の処理が文末まで達しない限り, 代名詞の解釈に関するアルゴリズムは起動されない。一方, これまでの心理学的知見から, 人間の読み手は必ずしも文末まで待たないうちに, 代名詞の解釈を進めていることが示唆されている。指示対象の特定がいつの時点で行われているのかについて, Just and Carpenter (1980) は, 読み手が照応表現を認定した時点で, たとえ仮にはあっても “即時的” にその指示対象を特定している, とする “即時性 (immediacy)” の仮説を提案した。以来, この仮説の妥当性をめぐって, 様々な議論がなされている (例えば, Sanford & Garrod, 1989)。

本研究は, 認知心理学的な立場から, 人間の照応解決の原理, および, 照応解決を支えるメカニズムを解明することを目指している。本報告の目的は, 照応解決における人間の多様な振る舞いを説明し, 心理学的により妥当性の高い計算モデルを構築することである。このために, 上記仮説の妥当性を検討する上で重要な先行研究の実験結果を踏まえて, ある特定の条件下では即時的に代名

詞を解釈し, またある状況下では, 文章の読みとともに漸進的に解釈を進めていく人間の特性を説明することが可能なモデル化を試みる。

2 代名詞を含む文の読みに関する先行研究 — Garrod, Freudenthal, and Boyle (1994) からの知見

Garrod, Freudenthal, and Boyle (1994) は, 代名詞の解釈の即時性に関して, 示唆に富む実験結果を報告している。彼らは, 談話の “焦点 (focus)” が照応解決に及ぼす影響を調べるため, 図 1 に示すような材料を用いて, 被験者が照応文を読むときの眼球運動を測定した。この結果, (a) 文よりも (b) 文の方が, 代名詞直後の動詞領域における読み時間が有意に長いことがわかった。そして, この読み時間の差は, (c) 文と (d) 文の間では観察されなかった。

この結果から, 以下のことが推察される。(a) 文に比べて (b) 文の方が動詞領域の読み時間が増加したのは, 照応文を読み終わらないうちに, *she* = *Joan* の解釈がある程度進められており, この解釈に対して意味的, 語用論的に符合しない後続文脈 (*poured*) を読んだとき, 読み手はその解釈の変更を余儀なくされ, このことが処理負荷の増大となって読み時間に表れたものと考えられる。一方, (c) 文と (d) 文の間で, 上記のような読み時間の差が見られなかったのは, *he* の指示対象である *the steward* には, 先行文脈において読み手の焦点が形成されていなかったために, *he* = *the steward* の解釈が, (a) 文や (b) 文のときに比べて進められておらず, この解釈に対して意味的, 語用論的に符合しない後続文脈 (*ordered*) を読んでも, 解釈の変更にもなう処理負荷の増大が生じなかったことによると考えられる。

つまり, Garrod et al. (1994) の実験結果は, 代名詞の解釈過程にはいくつかの要因が複雑に相互作用しており, つねに即時的な解釈が行われるとは限らないことを示唆している。人間の照応解決過程をモデル化するためには, 代名詞が即時的に解釈されるか否か, という二項対立的な問題を設定するのではなく, 上で述べたような読み手の柔

タイトル： *Flying to America*

先行文：

Joan wasn't enjoying the flight at all. The dry air in the plane made her really thirsty. Just as she was about to call him, she noticed the steward coming down the aisle with the drinks trolley.

照応文：

- (a) Right away **she** ordered a large glass of coke.
- (b) Right away **she** poured a large glass of coke.
- (c) Right away **he** ordered a large glass of coke.
- (d) Right away **he** poured a large glass of coke.

図1. Garrod, Freudenthal, and Boyle (1994) で使用された材料例。

軟な振る舞いを説明することが求められる。

以下では、このような考察を踏まえて、人間の照応解決をより広範に説明することが可能なモデルの構築を試みる。

3 照応解決過程のモデル化

モデル構成

図2に示すような文章例に対して、代名詞の理解過程をシミュレートするモデルを構築した。本モデルでは、McRae, Spivey-Knowlton, and Tanenhaus (1998) が提案している文理解モデルを照応解決に応用した。モデル構成の概念図を図2に示す。本モデルは、“指示対象ノード”、“制約”、および、各制約から指示対象ノードへの“リンク”、から構成される。指示対象ノードは、代名詞の指示対象候補を表すものであり、ここでは花子とボーイという二つのノードを設定した。以下、指示対象ノードを表す添字を i とし、 $i=1$ を“花子”ノード、 $i=2$ を“ボーイ”ノードとする。これらのノードは0から1までの活性化レベルを有し、活性化レベルが高いほど、代名詞の指示対象として解釈される可能性が高いことを意味する。制約は、読み手の照応解決に関与する要因を表す。以下、制約を表す添字を j とする。先行文脈から照応解決に関与する要因として、読み手の焦点を仮定し、これを制約1 ($j=1$) としてモデルに組み込んだ。また、照応表現自体から規定される要因として、代名詞が有する性や数などの語彙的制約を仮定し、これを制約2 ($j=2$) とした。後続文脈から照応解決に関与する要因として、同文脈の情報に基づいた推論によって生じる、意味的・語用論的なバイアスを仮定し、これを制約

3 ($j=3$)、および制約4 ($j=4$) とした。

これらの制約は、指示対象ノードに対してリンクを有する。このリンクは、各制約が“バイアス値”と“ウェイト値”という2種類のパラメータを有し、これらの値によって指示対象ノードの活性化レベルが計算されることを示すものである。以下、バイアス値、ウェイト値をそれぞれ b , w で表す。バイアス値 b_{ij} は、ある制約 j に関して、いずれのノードがより代名詞の指示対象として解釈されやすいのか、という読み手の選好性 (preference) を表すパラメータである。例えば、焦点化の制約 ($j=1$) を考えた場合、一般的に、焦点化されている人物の方が、焦点化されていない人物よりも代名詞の指示対象として解釈されやすい。この場合、 $b_{1i} > b_{1l}$ となる。選好性が大きいほど、 b_{1i} と b_{1l} の差は大きくなる。ウェイト値 w_j は、制約 j がその他の制約と比較して相対的にどの程度代名詞の解釈に寄与するのかを表すパラメータである。

動作原理

以下では、各制約が有するバイアス値とウェイト値に基づいて、指示対象ノードの活性化レベルを計算する三つのステップについて述べる。

第一ステップでは、おのおのの制約に関して、バイアス値の総和が1となるように正規化を行う。これは次式で表される。

$$b_{ij_norm} = b_{ij} / \sum_i b_{ij}$$

ここで、 b_{ij} は、制約 j から指示対象ノード i への正規化前のバイアス値を表し、 b_{ij_norm} は、正規化後のバイアス値を表す。同様に、ウェイト値に関しても、ウェイト値の総和が1となるように正

タイトル：「披露宴での出来事」

先行文： 花子は結婚式が無事に終わってほっとした。うきうきしながらパーティー会場へ入ったとき、一人のボーイがいきなりぶつかってきて、花子のお気に入りのドレスにワインをこぼした。

照応文： (a)すると彼女は怒りで震え出した。(Focus-referring 条件)
 (b)すると彼は丁重に謝った。(Nonfocus-referring 条件)

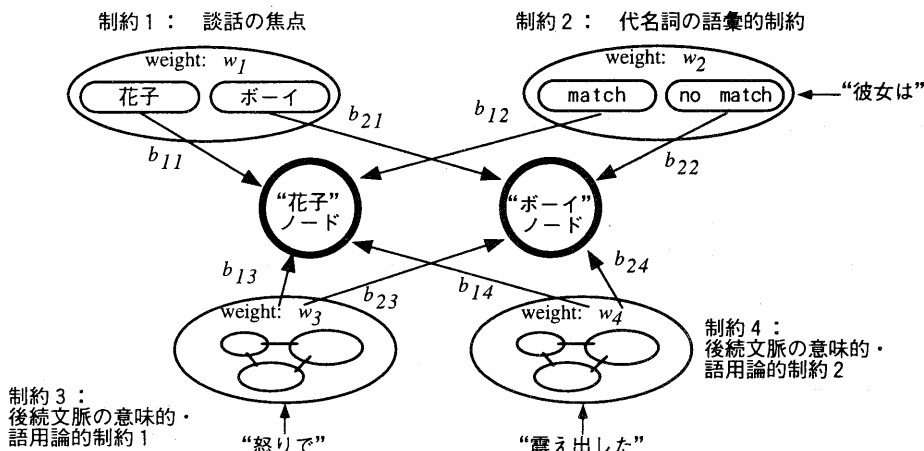


図2. モデル化の対象とした文章例とモデル構成の概念図. 上の図は、照応文(a)に対するモデル化の例であり、照応文(b)に対しては、制約2の楕円中の“match”と“no match”が入れ替わる。

規化を行う。これは次式で表される。

$$w_j_norm = w_j / \sum_j w_j$$

ここで、 w_j , w_j_norm はそれぞれ、正規化前、正規化後の制約 j のウェイト値を表す。

第二ステップでは、第一ステップで正規化されたバイアス値とウェイト値に基づいて、各ノードの活性化レベルを次式により計算する。

$$A_i = \sum_j w_j_norm * b_{ij_norm}$$

ここで、 A_i はノード i の活性化レベルを表す。

第三ステップでは、指示対象ノードの活性化レベルに比例した値が各制約のバイアス値にフィードバックされる。

$$b_{ij} = b_{ij_norm} + A_i * w_j_norm * b_{ij_norm}$$

以上の三つのステップを実行し、これを1サイクルとした。サイクル数の増大にともなって、二つの指示対象ノード間の活性化レベルの差が増大する。二つのノード間での活性化レベルの差が増大するほど、代名詞の解釈が進められていることを意味する。

シミュレーション

図2の照応文(a)のように、代名詞が、先行文脈において焦点化されている人物を指示する場合(以下、focus-referring 条件)と、照応文(b)のように、代名詞が、先行文脈において焦点化されていない人物を指示する場合(nonfocus-referring 条件)を想定し、シミュレーションを行った。照応文の読みの領域を、代名詞領域(彼女は/彼は)、代名詞の直後領域(怒りで/丁重に)、動詞領域(震え出した/謝った)、の三つに分け、それぞれの領域において、先に述べた三つのステップを繰り返して、指示対象ノードの活性化レベルを計算した。読み手が代名詞を認定したとき、焦点化の制約と代名詞の語彙的制約のもとで、指示対象の特定が進められると考え、代名詞領域では、これら二つの制約が有するバイアス値とウェイト値に基づいて、シミュレーションを行った。代名詞の直後領域では、上記二つの制約の他に、後続文脈の意味的・語用論的制約1を加え、三つの制約のもとで計算を行った。動詞領域ではさらに、後続文脈の意味的・語用論的制約2を加え、四つの制約のもとで計算を行った。

以下に、設定したパラメータの値について述べ

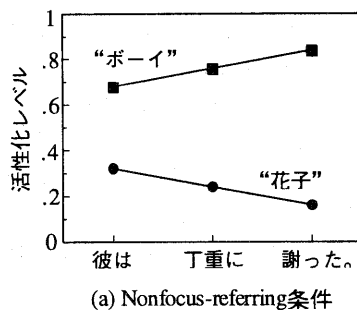
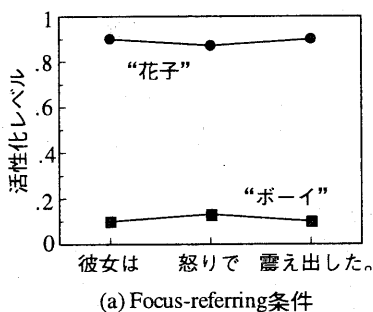


図3. 照応文の読みの各領域において計算された指示対象ノードの活性化レベル。

る。バイアス値は、焦点化の制約については、 $b_{11} = .7$, $b_{21} = .3$ とした。代名詞の語彙的制約については、focus-referring条件のとき、 $b_{12} = 1$, $b_{22} = 0$ に、nonfocus-referring条件のとき、 $b_{12} = 0$, $b_{22} = 1$ とした。後続文脈の意味的・語用論的制約については、focus-referring条件のとき、 $b_{13} = .8$, $b_{23} = .2$, $b_{14} = .9$, $b_{24} = .1$ に、nonfocus-referring条件のとき、 $b_{13} = .2$, $b_{23} = .8$, $b_{14} = .1$, $b_{24} = .9$ に設定した。ウェイト値は、代名詞領域では、焦点化の制約と代名詞の語彙的制約の二つが等しいウェイト値を有すると仮定し、 $w_1 = .5$, $w_2 = .5$ に設定した。以降、代名詞の直後領域と動詞領域では、直前の読みの領域における制約間のウェイト値のバランスを変化させない条件下で、すでに加えられていたウェイト値の総和と新たに加えられる制約のウェイト値とが等しくなるように設定した。このウェイト設定法は、McRae et al. (1998)の方法に準じたものである。サイクル数は5とし、1サイクルから5サイクルまでに計算された活性化レベルを平均化し、これをモデルの出力とした。

結果と考察

照応文の読みの各領域において計算された指示対象ノードの活性化レベルを図3に示す。同図において、focus-referring条件の場合、代名詞領域(彼女は)では、“花子”ノードの活性化レベルが.90であり、“ボーイ”ノードの活性化レベルが.10であった。“ボーイ”ノードに比べて“花子”ノードの方が活性化レベルが極めて高く、この傾向は、照応文の文末まで大きく変化することはなかった。一方、nonfocus-referring条件では、代名詞領域(彼は)において、“花子”ノードの

活性化レベルが.32、“ボーイ”ノードの活性化レベルが.68であった。この傾向は、照応文の読みとともに変化した。文の読みが文末へ進むにつれて、“ボーイ”ノードの活性化レベルが増加し、“花子”ノードの活性化レベルは減少した。上記二つの条件間で結果を比較してみると、代名詞領域では、nonfocus-referring条件に比べてfocus-referring条件の方が、指示対象ノード間での活性化レベルの差が大きいがわかる。この結果は、nonfocus-referring条件よりもfocus-referring条件の方が、代名詞の解釈がより即時的に進められていることを示すものと考えられる。そして、nonfocus-referring条件では、focus-referring条件の場合ほどには代名詞の解釈が進められていないが、代名詞の認定後、照応文を読むにつれて、漸進的に代名詞の解釈が行われていることを示している。

また、本シミュレーション結果は、先に述べたGarrod et al. (1994)の結果に符合するものである。彼らの実験において、代名詞が、先行文脈において焦点化されている人物を指示する場合に、後続文脈の意味的、語用論的なバイアスの違いによって、動詞領域の読み時間が変化したのは(図1の(a)文 vs. (b)文)、読み手の心内において、focus-referring条件のシミュレーション結果に示されるような、即時的な代名詞の解釈が進められていたことによるのではないだろうか。そして、この解釈に対して、意味的、語用論的に符合しない後続文脈を処理したために、代名詞の解釈の変更が引き起こされ、このことによって処理負荷の増大が生じたことが推察される。この処理負荷の増大は、代名詞の認定時において、代名詞の解釈がより進められているほど読み時間に顕著に表れ、解釈が

進められていないときには、読み時間に表れる可能性が小さいと考えられる。代名詞が、先行文脈において焦点化されていない人物を指示するときには(図1の(c)文と(d)文),本シミュレーションの nonfocus-referring 条件のように,代名詞の解釈がそれほど進められておらず,読み時間に表れるほどの処理負荷の増大が生じなかったことが推察される。

以上のことから,本モデルは,代名詞の解釈がいつの時点で,どの程度進められているのか,という問題に関して,代名詞の認定時における即時的な解釈はつねに行われるものではなく,“即時性”の仮説は,ある特定の条件下における人間の照応解決の一側面を説明するものであること,また同時に,人間の読み手は漸進的に代名詞の解釈を進めていること,をシミュレートすることが可能であると考えられる。

4 まとめ

本報告では,人間の照応解決をシミュレートすることが可能な,心理学的により妥当性の高い計算モデルの構築を試みた。“即時性”の仮説の妥当性について検討するために,示唆に富む先行研究を取り上げ,これまでの心理学的知見と合わせて,即時的な代名詞の解釈はつねに行われる一般性の高い原理でなく,ある条件下で見られる人間の照応解決の一側面であることを考察した。この考察を踏まえて,照応解決における人間の振る舞いをより広範に説明するモデルを提案し,このモデル化によって,人間の照応解決の心内プロセスを解明する上での示唆を得た。

本モデルは,代名詞の解釈について読み手が有する複数の選好性を数値的に表し,これらの選好性の相互作用によって,読み手の振る舞いを説明するものである。このモデルは,代名詞の解釈に対する制約が読み手の心内でどのように生じるのか,すなわち,代名詞の解釈に対する選好性が生じるメカニズムを説明するものではない。今後は,この心的メカニズムを明らかにするためのモデル化を行う必要がある。

引用文献

- 阿部純一・桃内佳雄・金子康朗・李光五(1994). 人間の言語情報処理 言語理解の認知科学. サイエンス社.
- Chambers, C. G., & Smyth, R. (1998). Structural parallelism and discourse coherence: A test of centering theory. *Journal of Memory and Language*, **39**, 593-608.
- Garrod, S., Freudenthal, D., & Boyle, E. (1994). The role of different types of anaphor in the on-line resolution of sentences in a discourse. *Journal of Memory and Language*, **33**, 39-68.
- Gordon, P. C., & Chan, D. (1995). Pronouns, passives, and discourse coherence. *Journal of Memory and Language*, **34**, 216-231.
- Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science*, **17**, 311-347.
- Grosz, B. J., Joshi, A. K., & Weinstein, S. (1995). Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, **21**, 203-225.
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, **87**, 329-354.
- Kennison, S. M., & Gordon, P. C. (1997). Comprehending referential expressions during reading: Evidence from eye tracking. *Discourse Processes*, **24**, 229-252.
- McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, **38**, 283-312.
- Sanford, A. J., & Garrod, S. C. (1989). What, when, and how?: Questions of immediacy in anaphoric reference resolution. *Language and Cognitive Processes*, **4**(3/4), S1235-262.
- Sidner, C. L. (1983). Focusing in the comprehension of definite anaphora. In M. Brady & R. C. Berwick (Eds.), *Computational models of discourse* (pp. 267-330). Cambridge, Massachusetts: The MIT Press.