

自然言語インタフェースにおける省略の扱い

加 藤 恒 昭†

代名詞や省略などを利用した文脈の参照は、自然言語による対話を簡潔かつスムーズなものとしている大きな要因である。自然言語インタフェースにおいてもこのような文脈参照を可能とすることによって、人間どうしでの対話と同様な自然さを実現する必要がある。文脈参照方略のひとつである省略 (Ellipsis) は、談話中の、不完全かつ断片的な文法的単位 (主に名詞句) の利用を指し、この発話断片と統語的もしくは意味的に対応する断片 (置換対象) を先行する発話より同定し、先行発話の意味内容の内での置換対象に相当する部分を発話断片の意味内容で置換することによって解釈される。本稿では、(1) 置換対象の同定において、その候補が複数個存在する場合の優先順位について、アンケート形式で行った実験の結果に基づいて報告する。本実験では、例えば、置換対象候補が共に同じ文の格要素である場合に、候補の優先順位には格種別等が影響し、特に主題化されているものが置換対象として選好されること等を明らかにした。これらの選好から示唆される置換候補探索順序を提案する。(2) 省略解釈のために必要となる意味内容置換処理に適した意味表現手法を提案する。本手法は統語構造の各節点に意味内容を構成するための部分的な情報を記述するもので、統語構造と意味内容との対応が明確である点、置換後の再構成が容易である点で省略処理に適したものである。

Handling Ellipsis in Natural Language Interface Systems

TSUNEA KI KATO†

Natural language interface systems must understand ellipsis if we want to make dialogue with them more fluent. To realize this they must identify as the correspondent, the constituent of the previous utterance that semantically corresponds to the elliptical utterance, and re-construct the semantic content after replacing the correspondent with the elliptical utterance. In this paper, (1) Based on the results of experiments using questionnaires, we report which type of constituent tends to be chosen as the correspondent, where several candidates exist. The constituent conveying older information or the largest constituent is favored. (2) We propose a semantic representation method that enables ellipsis understanding. In this method, partial information for constructing the complete semantic content of an utterance is distributively attached to the nodes of the syntactic structure. The semantic content can hold an explicit relation to the corresponding syntactic structure, and, thus, can be easily re-constructed.

1. はじめに

自然言語による対話は、代名詞や省略などを利用した文脈参照によって、簡潔かつスムーズなものとなる。例えば、「神奈川営業所の40歳以上の社員を知りたい」という質問をし、その応答を得た後に続けて「神奈川営業所の30歳以上の社員を知りたい」という内容の質問を行うとき、「30歳以上は」等の表現によって文脈が参照され、発話が節約される。自然言語インタフェースシステムは対象世界の事物を指示する表現を理解できるだけでなく、ユーザが行うこのような文脈参照の方略をも理解できなければならない。

文脈参照方略のひとつである省略 (Ellipsis) は、談話中の、それ自体では不完全で断片的な文法的単位 (主に名詞句) の利用を指し、欠落している要素を文脈から補うことによって理解される¹⁾。特に発話全体の統語カテゴリが文でないという点によって、省略は0代名詞化による要素の欠落と区別される。なお、「省略」という語は0代名詞化を含んだ広い意味で用いられることもあるが本稿ではこの語を上記の意味で用いる。つまり、完全な文として必要な要素が欠落した断片的発話で、先行発話中の要素を補完することによって理解されるものを指すとして、0代名詞化と区別する。省略を理解するための統語的意味的アプローチ^{2), 3)}では、発話断片と統語的もしくは意味的に対応する断片 (置換対象と呼ぶ) を先行発話より同定し、その発話の意味内容の内での置換対象に相当する部

† NTT 情報通信網研究所メッセージシステム研究部
Message Processing Systems Laboratory, NTT
Network Information Systems Laboratories

分を発話断片の意味内容で置換して得られるものを解釈結果とする。ただし、「神奈川営業所の社員を知りたい」「40歳以上は」のように発話断片の意味内容を先行発話の意味内容に付け加えることによって解釈されるような省略もあるが、これについては本稿では論じない。

本稿では、以上の処理における二つの問題について議論する。第一の問題は、置換対象の候補、つまり発話断片と同じ意味カテゴリに属す断片が複数個存在したときに生じる曖昧性である。このような曖昧な省略において、どのような場合にどの解釈を優先すべきかは、人が実際に行っている省略解釈の傾向（選好）に基づいて決定されるべきである。本稿では、この解釈の選好を明らかにするために実施した、人が省略をどのように解釈しやすいか、さらに発話者が曖昧性を避けるためにどのような方略を利用しているかに関する実験調査について報告する。第二の問題は、省略解釈において必要となる意味内容の置換処理である。置換対象の探索は統語構造に対する処理として実現する必要があるため、統語構造上で同定された部分に対応する意味内容を抽出し、それを発話断片の意味内容で置換し、全体の意味内容を再構成するという処理が必要となる。本稿では、このような置換処理に適した意味表現手法を提案する。

以下、まず2章で、省略処理の問題を概観し、統語構造と意味内容とを同時に処理の対象とすることによって初めて適切な処理が実現できることを示す。3章で省略の曖昧さについて議論し、4章で省略処理に適した意味表現手法を論じる。最後に5章でまとめと考察を行う。

2. 省略解釈とその問題

前述のように、省略を理解するためには、第一に発話断片に対する置換対象を先行発話から同定すること、第二に同定された置換対象に対応する意味内容を発話断片の意味内容で置き換えることにより、先行発話の意味内容から新たな意味内容を再構成することが必要である。自然言語インタフェースにおいては、このようにして得られた文脈に依存しない意味内容が発話断片の解釈として後段の問題解決部（例えば、データベース検索システム）に渡される。本章では、第一の段階である置換対象の同定においては先行発話がどのような統語構造を有していたかが重要であることと、一方で、最終的に必要となる意味内容を得るため

に、前半の統語構造上の処理と後半の意味内容に関する処理との適切な融合が必要となることを示す。

置換対象の同定において、先行発話の統語構造が重要であることは、例えば以下の例からわかる。(2-1)に続いて(2-2)が発話された場合、もっともらしい置換対象は「〇〇社と△△社」であり、その結果「□□社と取引のある営業所は」が解釈となる。一方(2-2)が(2-1')に続いて発話された場合に、「□□社と取引のある営業所は」という解釈は少なくとも一般的ではない。(2-1)と(2-1')は、一階述語論理で意味内容を表した場合、共に(2-3)で表現される。ここで、 $X \vdash P(X)$ は質問文の意味内容の表現であり、式 $P(X)$ を満たす X を求めることを意味するとする。このように同じ意味内容を持つ先行発話の間で置換対象が異なるのは、その同定が統語構造に依存するためと考えられる。

「〇〇社と△△社と取引のある営業所は」
(2-1)

「□□社は」
(2-2)

「〇〇社と取引のある営業所で
△△社とも取引のあるのは」
(2-1')

$X \vdash (\text{営業所}(X) \wedge \text{取引}(X, \text{〇〇社})$
 $\wedge \text{取引}(X, \text{△△社}))$
(2-3)

比較構文においても同様である。以下に示す(2-4)と(2-4')とはほぼ同義であり、(2-4')の下線部は、(2-4)の表層には現れていないが、意味内容として補完すべき部分である。(2-4)に続いて「既婚男性では」という入力があった場合、望まれる解釈は(2-5)であって(2-5') (2-5'')ではない。一方、直接(2-4')が入力となった場合は、これらの解釈も可能である。(2-5')の解釈がありえないことより、表層に現れていない部分は置換対象とならないことがわかる。これも先行発話の統語（表層）構造の影響である。これらのことから、置換対象の同定は統語構造上での探索処理として実現すべきであるといえる。

「営業部より、独身女性が多い部」
(2-4)

「営業部の独身女性の数より、独身女性の数が多い部」
(2-4')

「営業部の既婚男性の数より、既婚男性の数が多い部」
(2-5)

「営業部の既婚男性の数より、独身女性の数が多い部」
(2-5')

「営業部の独身女性の数より、既婚男性の数が多い部」
(2-5'')

前半の置換対象同定が統語構造上の処理であるとする、これを後半の意味内容に関する処理と適切に融合する必要が生じる。例えば、置換対象は統語構造の部分として同定されるので、この部分に対応する意味内容を容易に得られることが必要である。しかし、先の等位接続構造の例では、意味内容(2-3)に「〇〇社と△△社」に直接対応する部分が存在せず、このため、それと発話断片「□□社」の意味内容との置換は困難となる。このような困難を避けるためには任意の統語的部分に対応する意味内容が存在し、それが容易に抽出できることが必要である。また、比較構文の場合は、上例で(2-5^{*)}の解釈がありえないことより、単純に置換対象と発話断片の意味内容を置換するだけでは不十分で、表層に現われない部分の意味内容を補完する処理をその後に改めて行う必要があることがわかる。

以上より、統語構造上の処理である置換対象の同定と後半の意味内容に関する処理を融合させるためには、置換対象として同定された統語構造の部分と対応する意味内容が常に同定でき、その部分の置換や意味内容の補完などを含む新しい意味内容の再構成処理を容易に実現できるような意味表現手法を利用する必要があることがわかる。本稿では、置換対象の同定に影響する統語的な特徴について述べると共に、上記の条件を満たす意味表現手法を提案する。

3. 省略の曖昧性と選好

統語構造上での探索処理として実現される置換対象の決定は、基本的には先行発話の統語構造から発話断片と置換可能な意味カテゴリを有する部分を発見する処理である。条件を満たす意味カテゴリを持つ部分が複数個存在するとき、そこに曖昧性が生じるが、統語的な情報に基づいていずれかが選好されると考えられる。また、話し手が積極的に特定の解釈を示唆する方法として Clue の利用がある。省略の最も単純な表現である「'名詞句'+は」から外れる言い回し、もしくはこれに付加される助詞や接続詞は、特定の解釈を示唆するための Clue であると考えられる。また、人は曖昧さのない範囲で最も簡単な表現を行うと仮定すれば、これらの Clue は無意味に利用されたのではなく曖昧さを減ずるために必要であると判断されたと考えられる。

本章では曖昧性を含んだ省略に対して、人がどのような解釈をしやすいかという省略解釈の選好と、曖昧

さを避けるために話し手が用いる方略について、アンケート形式で行った実験に基づいて述べる。さらにそれに基づき、曖昧性がある場合でも、選好される置換対象を先に発見するような探索順序を提案する。

3.1 実験方法

実験は成人男女 40 名の被験者を対象に以下の 2 種類を行った。

【解釈実験】 続けて発話されたという設定で二つの質問文(その間には適当な応答があったと仮定させた)を提示し、第一の発話なしに第二の質問だけを尋ねる場合に、どのような表現を用いるかを記入する。

問題例 「神奈川営業所の 30 歳以下の社員を知りたい」

「横浜営業所は」

回答例 「横浜営業所の 30 歳以下の社員を知りたい」

【生成実験】 類似した条件を持つ二つの質問文を提示し、第一の発話を行った後に(それに対する適当な応答を仮定して)、続けて第二の質問を尋ねる場合、どのような表現を用いるかを記入する。

問題例 「神奈川営業所の 30 歳以下の社員を知りたい」

「横浜営業所の 30 歳以下の社員を知りたい」

回答例 「横浜営業所は」

問題数は各実験約 75 題であり、これを 3 回に分けて実施した。解釈実験からは、Clue のない単純な形式の表現を第二の発話とすることで、曖昧性を含む省略の用法について人が行う解釈の選好を直接抽出した。生成実験では、回答に含まれる Clue 等、ある解釈を示唆するために利用される方略を抽出すると同時に、Clue のない単純な表現が回答に多い発話対は解釈実験の裏返しであることに着目して、省略解釈の選好を間接的に抽出した^{4),5)}。

置換対象に曖昧性が生じる場合として、基本的には以下の二つが考えられる。

(1) 格要素間の曖昧性: 同じ意味カテゴリを持つ二つの名詞句が共に同じ文の格要素である場合

例 「太郎が次郎を殴りましたか」「三郎は」

…置換対象が「太郎」か「次郎」かの曖昧性

(2) 統語的包含関係にある名詞句間の曖昧性: 名詞句とその部分や、並列句のように、同じ意味カテゴリを持つ二つ以上名詞句が統語的な包含関係にある場合

例 1 「太郎の上司はどこに住んでいますか」

「次郎は」

…置換対象が「太郎」か「太郎の上司」かの曖昧性
例2 「ピアノとバイオリンを習っている生徒は」

「フルートは」

…置換対象が「ピアノとバイオリン」か「ピアノ」
か「バイオリン」かの曖昧性

以下、これらに関する実験の結果について述べる。
なお、実験に用いた文例は特に断らない限り、すべて
表に示した。主に解釈実験の結果について論じるが、
全体の75題のうち、上記の問題に関連する設問は40
題である。

3.2 実験結果

(1) 格要素間の曖昧性において、選好に影響を与
える可能性のある要因として、格の種別、副助詞ハに
よる主題化、出現順序などが考えられる。文の自然性
などの理由により、網羅的に実験を行えたわけではな
いが、三つの基本文(「太郎が次郎を殴りましたか」
「太郎が次郎にケーキを贈りましたか」「神奈川営業所
は横浜営業所よりトラックをたくさん販売しています
か」)をもとにこれらの要因を変化させた32文につい
て、解釈実験を行った。その結果の一部を表1、表2
に示す。表中の×印は自然な文とならなかったために
実験を行わなかったことを示す。数値はそこに示され
た項目を置換対象とした被験者数を示す。例えば、表
1の1行1列のデータは、「太郎が次郎を殴りました
か」「三郎は」という発話列における「三郎は」の解
釈として、「太郎」を置換対象として「三郎が次郎を
殴りましたか」という解釈を行ったものが32名、逆

に「次郎」を置換対象として「太郎が三郎を殴まし
たか」という解釈を行ったものが4名であることを示
している。総和が40にならないのは、これ以外の回
答をした被験者がいるためである。

ここから、例えば表1で、1行1列、2行1列の比
較から、語順よりも格の種別の方が大きく選好に影響
しており、助詞ガを伴うものの方が助詞ヲを伴うもの
より選好されることがわかる。また、表2の二つの列
の間に大きな差が見られないことから、先行発話全体
がWH疑問文であるか、YN疑問文であるかは影響
がないこともわかる。これらをまとめると次のように
なる。ここで、AがBより置換対象になりやすい傾向
にある(選好される)ことを $A > B$ で表すものとする。

格の種別

主題のハ > 主格のガ > その他の格 影響大
名詞の種類

固有名詞などの名詞 > 疑問代名詞 ↑

出現順序

前方に現れた名詞 > 後方の名詞 影響小
全体がWH疑問文であるか、YN疑問文であるか 影響なし

主題は旧情報を表現し、疑問代名詞は新情報を表現
する。また、語順も情報の新旧を反映している⁶⁾ので、
ここで観察された傾向は、旧情報が新情報よりも選好
されるという観点でまとめられる。またこのうち、格
の種別の優先順序は0代名詞の処理に関連するセンタ
の優先順位⁷⁾に類似している。

同様の傾向は生成実験でも確認された。また、選好
の低い解釈を示唆するためのClueとして、以下のよう
なものが見られた。

「太郎が次郎にケーキを贈りましたか」「三郎には」
「太郎が次郎を殴りましたか」「三郎を殴りました
か」
「太郎が誰を殴りましたか」「三郎ですか」

最初の2例は格を明示することにより置換対象を限
定する。最後の例は、Copulaを使用することで疑問
代名詞が置換対象であると示唆している。この例は、
「太郎が誰を殴りましたか」「三郎です」「三郎です
か」というように驚きを表現しているという印象を受
ける。これはCarberryのいうExpress-Surprise-
Obtain-Corroboationに近い発話⁸⁾と考えられるが、
日本語の場合、統語的特徴だけでもこのタイプの省略
が区別されることがわかる。

表1 格要素間の曖昧性における選好(1)

Table 1 Preferences in case elements pairs (1).

問題パタン「～が～を殴りましたか」/「三郎は」

	A=太郎, B=次郎	A=誰, B=次郎
AがBを～	太郎が-32 次郎を-4	誰 が-22 次郎を-14
BをAが～	太郎が-25 次郎を-9	次郎を-25 誰 が-9
AはBを～	太郎は-34 次郎を-3	×
BをAは～	太郎は-29 次郎を-1	×
BはAが～	次郎は-24 太郎が-6	次郎は-28 誰 が-6

表2 格要素間の曖昧性における選好(2)

Table 2 Preferences in case elements pairs (2).

問題パタン「～が～に～を贈りましたか」/「三郎は」

	ケーキを	何 を
太郎が次郎に～	太郎が-35 次郎に-2	太郎が-33 次郎に-4
次郎に太郎が～	太郎が-32 次郎に-4	太郎が-30 次郎に-4
太郎は次郎に～	太郎は-35 次郎に-3	太郎は-35 次郎に-4
次郎に太郎は～	太郎は-33 次郎に-2	太郎は-35 次郎に-1
次郎は太郎が～	次郎は-14 太郎が-6	×

表 3 統語的包含関係にある名詞句間の曖昧性における選好 (解釈実験)
Table 3 Preferences in syntactically embedded NP pairs (The result of the interpretation experiment).

問	題	外側名詞句	内側名詞句
“名詞”の“名詞”	太郎の弟はどこに住んでいますか／次郎は	26	13
	太郎の上司はどこに住んでいますか／次郎は	22	18
	太郎の弟は何歳ですか／次郎は	21	16
	太郎の上司は何歳ですか／次郎は	18	21
“名詞”+“機能的接尾辞”	～部の30歳以上の社員は／40歳は	15	22
	年収400万円以上の～出身の社員は／600万円は	21	14

(2) 統語的包含関係には「名詞」の「名詞」, 「名詞」+「機能的接尾辞」, 「文」+「名詞」などが考えられる。前者2種類について解釈実験を行い、表3に示す結果を得た。なお、統語的にこれらの構造を持っているものがすべて曖昧となるわけではない。全体とその中に含まれた部分とが同じ意味カテゴリを持つことがさらに必要である。例えば「名詞」の「名詞」は多様な意味関係を表現しうるが、曖昧さを生じうる可能性を持つ関係は非常に少ない。

「太郎の弟」のように、この名詞句全体の意味カテゴリ、例えば男性、とそこに含まれる「太郎」の意味カテゴリとが同一である場合だけが問題となる。本実験はそれらを対象に行った。表3において、「外側名詞句」は「太郎の弟はどこに住んでいますか」「次郎は」という問題で、統語的な包含関係に関して外側、つまり包含する側である「太郎の弟」を置換対象として「次郎はどこに…」という回答をしたことを示す。「内側名詞句」は同じ問題に包含される側である「太郎」を置換対象として「次郎の弟はどこに…」と回答したことを示す。表が示すように解釈実験では明確な傾向は認められない。一方、生成実験の結果から、それぞれの解釈を示唆するために単純な表現で十分と考えている被験者数を求めると表4のようになる。この表で、「外側名詞句」は与えられた発話、例えば、「太郎の弟はどこに…」において、外側の名詞句を置換対象とした解釈（「次郎はどこに…」）を得るために「次郎は」という表現を用いた被験者の数であり、「内側名詞句」は、内側の名詞句を置換対象とした解釈（「次郎の弟はどこに…」）を得るために「次郎は」を用いた被験者の数である。ここからは、以下の傾向が見られる。

表 4 統語的包含関係にある名詞句間の曖昧性における選好 (生成実験)

Table 4 Preferences in syntactically embedded NP pairs (The result of the generation experiment).

単純な表現を利用した被験者数

問	題	外側名詞句	内側名詞句
“名詞”の“名詞”	太郎の弟はどこに住んでいますか	30	4
	太郎の上司は何歳ですか	26	6
“名詞”+“機能的接尾辞”	～部の30歳以上の社員は	20	0
	年収400万円以上の～出身の社員は	25	3

表 5 並列句内の曖昧性

Table 5 Preferences in coordinate structures.

問	題	並列句全体	前方要素	後方要素
ピアノとバイオリンを習っている生徒は ○○社と△△社と取引のある営業所は		37	1	2
		25	2	11

外側の名詞句 > 内側の名詞句

並列句に起因する曖昧性の場合でも、表5に示す解釈実験の結果から同様の選好が明らかである。表5では「ピアノとバイオリンを習っている生徒は」「フルートは」という発話列で、「フルートを習っている…」「フルートとバイオリンを習っている…」「ピアノとフルートを習っている…」という解釈を行った被験者数が、それぞれ「並列句全体」「前方要素」「後方要素」として示されている。ここから、

並列句全体（外側の名詞句）>

並列句内の要素（内側の名詞句）

この傾向は生成実験でも明らかで、並列句全体との置換を指定した場合にはほとんど Clue は見られなかった。

内側の名詞句との置換を示唆するための適切な Clue はほとんど観察されず、以下のように外側の名詞句もしくは並列句全体との置換として扱えるような発話が利用されることが生成実験で観察された。

「次郎の弟は」

「ピアノとバイオリンを習っている生徒は」

「ピアノとフルートでは」

3.3 統語構造探索の指針

以上の実験から、選好される解釈を優先して発見するための統語構造の探索順序は以下のようになる。

- 先行発話全体から始めて、その部分構造へというトップダウンの探索を行う。
- 文の節点ではその文を構成している格要素を列挙し、主題のハ、主格のガ、その他の格の順に探索する。
- ある部分構造が条件を満たした場合はそこで停止し、その部分構造は探索しない。

Clue としての助詞や Copula の使用は、ある部分構造が置換対象として適切であるかの条件を付加する形で働く。これらの Clue が無い場合は発話断片の中心をなす名詞句の意味カテゴリだけが条件となり、それと置換可能な意味カテゴリを有する部分構造が探索されることになる。

4. 省略処理のための意味表現手法

置換対象が同定された後に行うべき処理は、その置換対象部分に相当する意味内容を発話断片の意味内容で置換し、先行発話全体の意味内容を再構成することである。前述のように、このためには置換対象として同定された統語構造の部分に対応する意味内容が常に同定できるとともに、容易に新しい意味内容が再構成できるような意味表現手法が必要である。

論理に基づく意味表現言語と、それを利用した、部分式の連言結合と部分式に含まれる項（変数および定数）の単一化とに基づく意味合成の枠組みをすでに提案している^{9),10)}。この意味表現言語（以下、Lと呼ぶ）は、一階述語論理式における基本式の連言結合を拡張したものであり、平板な構造を維持したまま選言や集約演算等を扱うことができる。本章では、このLを利用した意味表現手法を提案し、これが上述の条件を満たしており適切な省略処理を可能とすることを述べる。

4.1 意味表現手法と評価処理

意味表現言語 L は、類間に汎化専化 (IS-A) 階層関係を導入して拡張した多類論理に基づくもので、一般には述語として扱われる個体の順序組を個体として扱うことを特徴とする。L の個体は $E:C$ と記述される。ここで、E は個体記号であり、英大文字の場合は変数

である。C は類名である。基本式は、 $E0:C0 (E1:C1, \dots, En:Cn)$ であり、これは個体 $E1:C1, \dots, En:Cn$ の組であるような個体 $E0:C0$ が存在するときに、真である。ある類の個体がどの類の個体の組であるかは、分野知識として定義される。これが多類論理における述語定義に対応する。L の整式はこの基本式の連言結合であり、特に指定がない場合、量記号としては存在量記号が仮定される。

「東京にある会社が計算機を生産する」に対応する L の表現は、

$A: \text{ある} (B: \text{会社}, \text{東京}: \text{場所}) \wedge C: \text{生産} (B: \text{会社}, D: \text{計算機})$

であり、これは、以下の一階述語論理式に対応する。

$\exists B \exists D (\text{会社}(B) \wedge \text{計算機}(D) \wedge \text{ある} (B, \text{東京}) \wedge \text{生産} (B, D))$

なお、表記を簡単にするために、特に問題がない場合は、 $E0:C0 (E1:C1, \dots, En:Cn)$ を $C0 (E1:C1, \dots, En:Cn)$ と記述する。

発話の意味内容は三つ組 (Y, P, X) で表現される。

図1にその例を示す。統語構造の各節点の上に記述されているものがその節点を根とする部分構造に対応する意味内容である。ここで、第二項 P が L の整式であり、第一項 Y、第三項 X はそこに現れる項の並びである。第一項 Y に含まれる項はすべて変数で、直感的にはラムダ計算に基づく意味合成においてラムダ抽象された変数に相当するものである。第三項 X に含まれる項は、それ以外の変数もしくは定数で、変数はすでに量化されたものとして意味づけられる。また、第三項 X において、その先頭には常にその句の主なる意味を表現するもの、つまり名詞句では主名詞に対応する項、文では主用言に対応する変数がおかれる。意味合成は、左右部分構造の意味内容の第二項を連言結合し、そこに含まれる項を適当に単一化することにより進められる。例えば、図1の根の節点では、全体の意味内容を得るためにその左右部分構造の $B: \text{会社}$ と $E: \text{会社}$ とが単一化されている。

等位接続構造や比較構文においてはその部分構造の意味内容を複数回、別様に具体化して利用しなければならない場合があり、これはダブルインスタンスションと呼ばれる¹¹⁾。例えば、

「〇〇社と△△社と取引のある営業所は」

(4-1)

$([], C: \text{取引} (D: \text{営業所}, \text{〇〇}: \text{会社}) \wedge$

$C1: \text{取引} (D: \text{営業所}, \triangle\triangle: \text{会社}), [D: \text{営業$

所, …]) (4-2)
 において取引に対応する意味内容 $_$: 取引 ($_$, $_$) が 2 回現れている。部分の意味内容を連言結合していく上記の手法だけでは、このような複数回の具体化を実現できないことが問題となる。そこで、意味内容を構文意味解析処理時に逐次的に直接生成するのではなく、解析がすべて終了し完全な統語構造が得られた後で、評価処理と呼ぶ後処理によって合成することとし、この評価処理の中でダブルインスタンスーションの問題に対処する。まず、解析時には、その節点を根とする部分の意味内容とその部分構造の意味内容とを関係付ける関係情報を、各節点に記述する。これは以下のような四つ組である。

((YN, PN, XN), (YL, PL, XL),
 (YR, PR, XR), (Y, PL $\wedge$ PN $\wedge$ PR, X))

ここで、第二項と左部分木の意味内容、第三項と右部分木の意味内容とをそれぞれ単一化すると、第四項にその節点を根とする部分に対応する意味内容が得られるようになっている。第一項はその節点で付加される意味内容である。例えば、「東京の会社」という表現の解釈として「東京にある会社」を導く場合、「ある」に相当する意味内容は左右どちらの部分木にも含まれない。これを節点の時点で加えるのが第一項の役割である。ここで第四項の第二要素である L の整式は常に $PL\wedge PN\wedge PR$ であるので、このことを利用すれば、上の関係は以下で表現できる。図 1 の口内にその例を示す。

((YN, PN, XN), (YL, XL), (YR, XR), (Y, X))

評価処理では、統語構造を再帰的に処理してゆき、

各節点に記述された上記の関係情報から順次、ボトムアップ的に意味内容の具体化と合成を行っていく。等位接続や比較構文のヨリ格の節点等、特別な処理が必要となる可能性のある節点には、構文意味解析時にそれを明記しておき、その節点を処理する際に、その節点が含まれる統語構造の特徴に照らして必要な処理を起動する。処理の起動が必要であるかを決定するための、上位の統語構造の情報や、下位の節点で起動されている処理種別とそれに付随する情報は評価処理のパラメータとして伝播される。

等位接続構造を例に、本枠組みでダブルインスタンスーションの問題が回避できることを述べる。図 2 に等位接続構造を含んだ発話の例を示す。日本語の等位接続構造「～と～」は or 並列であるか and 並列であるかが曖昧な表現であり、どちらの解釈が適切であるかはそれが含まれる統語構造などに依存する。構文意味解析時にはこれに or 並列の解釈を付与しておく。or 並列にとらえた場合、ダブルインスタンスーションの問題は生じないので、構文意味解析によって各節点に記述された関係 (図中口内) を用いて単純な合成を行っていけばよく、この例では以下が得られる。

([], or (〇〇: 会社, $\Delta\Delta$: 会社, G: 会社) $\wedge$
 C: 取引 (A: 営業所, G: 会社), [A: 営業所, …])

これは、「〇〇社か $\Delta\Delta$ 社かのどちらかであるような会社と取引している営業所」の意味である。一方、本例のように等位接続構造が埋め込み文の格要素であるとき、それは and 並列と解釈されることが多い¹²⁾が、

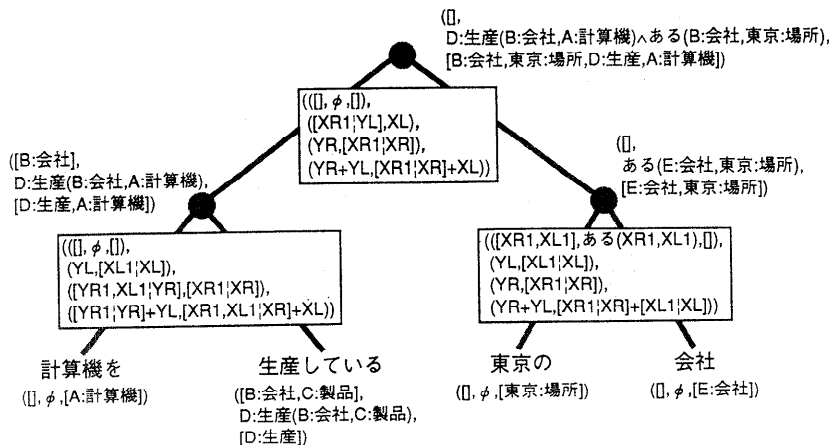


図 1 統語構造上での意味内容表現

Fig. 1 A semantic content representation on a syntactic structure.

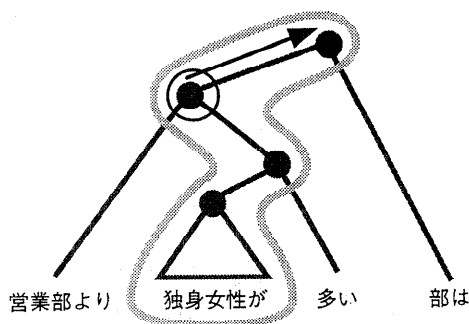


図3 比較構文における評価処理
Fig. 3 Handling comparative sentences.

図3は比較構文について示したものである。2章で述べたようにこのような発話では「～の独身女性」に相当する意味内容を補完して「営業部」と関係付けなければならない。等位接続構造の場合と同様、これは評価処理において行われる。この処理は通常の意味制約を満たさないようなヨリ格の節点によって起動され、図中、破線で囲んだ部分の意味内容を2回具体化し、一方を営業部と関係付けるものである。この発話に続いて「既婚男性は」という発話がなされた場合、統語構造上で「独身女性」が置換対象と同定され、それとの置換が行われる。評価処理はまったく同様に行われるが、このときに複数回具体化されるのは置換後の「既婚男性」である。このように評価処理に必要な補完が行われるために2章で述べた問題は回避でき、不適切な解釈が得られることはない。

5. おわりに

自然言語インタフェースにおける省略処理の実現に関して、二つの問題を議論した。

第一は、置換対象候補の優先順位についてである。日本語の0代名詞に関しては本研究と同じ観点の研究が比較的多くなされているが^{7), 13), 14)}、省略に関しては、冠詞等に関するいくつかの制約が示されている^{1), 2)}だけで、ここで扱ったような意味での曖昧性がある場合についての知見は得られていない。なお、本実験では視点や共感等の要因に関しては調査していないが、自然言語インタフェースという観点ではこれらを考慮しなければならない状況はほとんどないので実用上問題にならないと考えられる。

第二は、統語構造と意味内容とを対応付けることによる省略処理の実現である。省略処理を実現するためには統語構造と意味内容とのこのような対応付けが必

須であり、分野文法を利用して意味規則と統語規則を対応させておき、統語構造を探索して置換対象を発見する手法¹⁵⁾や、省略に近い現象である do-anaphora に対処するために、意味内容をλ抽象して維持しておく手法¹⁶⁾などが提案されている。本手法は、対象分野に依存しない一般的な文法で意味規則と統語規則を対応させ、さらに置換処理が任意の部分について行えるようにするために、抽象化に相当する処理を統語構造のすべての部分で行っている。この点でより一般的な手法である。

構文意味解析において意味内容を直接合成せず、何らかの中間表現を設けることで、文脈に関連する問題、量記号のスコープ決定の問題、等位接続構造等でのダブルインスタンス化の問題を解決しようという考えやそれに基づく手法はいくつか提案されている^{17)~21)}。本手法はこのような流れに属するものであり、自然言語インタフェースという観点から、データベース検索システム等での利用を考慮して一階述語論理に近いものを意味表現言語として採用し、式の連言結合と項の単一化による意味合成を実現している。特に本稿では、本手法が省略解釈を容易にし、自然な解釈のみを生成することを述べることで、その有効性を示した。

本手法に基づく構文意味解析処理は Prolog を利用して実現されている²²⁾。

謝辞 日頃よりご指導いただく、坂間保雄主席研究員、中村稔メッセージシステム研究部部長に感謝いたします。また、本研究を進めるにあたって熱心に議論いただき、本論文執筆に関しても貴重なコメントをいただいた東田正信主幹研究員、大山芳史主幹研究員、林良彦氏、石崎雅人氏、石川有紀子氏に感謝いたします。特に、石川有紀子氏には本論文3章の実験計画、課題設定に関して有益なアドバイスを数多くいただきました。ここに感謝いたします。

参考文献

- 1) Walker, D.E. (ed.): *Understanding Spoken Language*, Elsevier, North-Holland, New York (1978).
- 2) Allen, J.F.: *Natural Language Understanding*, The Benjamin/Cummings Publishing, Menlo Park (1988).
- 3) Carbonell, J.G.: Discourse Pragmatics and Ellipsis Resolution in Task-Oriented Natural Language Interfaces, *Proceedings of 21st Annual Meeting of the ACL*, pp. 164-168 (1983).

- 4) 加藤ほか：省略における曖昧性とその解消，第43回情報処理学会全国大会論文集，5H-7 (1991).
- 5) Kato, T. et al.: Ellipsis in Japanese Discourse, *Natural Language Processing Pacific Rim Symposium*, pp. 217-224 (1991).
- 6) 久野：談話の文法，大修館書店 (1978).
- 7) Walker, M. et al.: Centering in Japanese Discourse, *Proceedings of COLING '90*, Vol. 1, pp. 5-10 (1990).
- 8) Carberry, S.: A Pragmatics-Based Approach To Ellipsis Resolution, *Computational Linguistics*, Vol. 15, No. 2, pp. 75-96 (1989).
- 9) 加藤：語用論処理を考慮した意味表現，情報処理学会「談話理解モデルとその応用」シンポジウム，pp. 101-112 (1989).
- 10) 加藤：式の連言結合と項の単一化に基づく意味合成，情報処理，Vol. 32, No. 8, pp. 971-980 (1991).
- 11) Moore, R.C.: Unification-Based Semantic Interpretation, *Proceedings of 27th Annual Meeting of the ACL*, pp. 33-41 (1989).
- 12) 藤崎ほか：データベース照会システム「ヤチマタ」と名詞句データ模型，情報処理学会論文誌，Vol. 20, No. 1, pp. 77-84 (1979).
- 13) 吉本：談話処理における日本語ゼロ代名詞の扱いについて，情報処理学会 NL 研究会，56-4 (1986).
- 14) Kameyama, M.: A Property-Sharing Constraint in Centering, *Proceedings of 24th Annual Meeting of the ACL*, pp. 200-206 (1986).
- 15) Hendrix, G.G. et al.: Developing a Natural Language Interface to Complex Data, *ACM Trans. Database Systems*, Vol. 3, No. 2, pp. 105-147 (1978).
- 16) Webber, B.L.: *A Formal Approach to Discourse Anaphora*, Garland Publishing, Inc., New York (1979).
- 17) Schubert, L.K. and Pelletier, F.J.: From English to Logic: Context-Free Computation of 'Conventional' Logical Translations, *Computational Linguistics*, Vol. 8, No. 1, pp. 26-44 (1982).
- 18) Woods, W.A.: Semantics and Quantification in Natural Language Question Answering, Yovits, M. (ed.), *Advances in Computers*, Vol. 17, pp. 2-64, Academic Press, New York (1978).
- 19) Alshaw, H. and Eijck, J.: Logical Forms in the Core Language Engine, *Proceedings of 27th Annual Meeting of the ACL*, pp. 25-32 (1989).
- 20) Pollack, M.E. et al.: An Integrated Framework for Semantic and Pragmatic Interpretation, *Proceedings of 26th Annual Meeting of the ACL*, pp. 75-86 (1988).
- 21) Pereira, F. et al.: Incremental Interpretation, *Artif. Intell.*, No. 50, pp. 37-82 (1991).
- 22) 加藤ほか：自然言語インタフェースにおける構文意味解析方式，NTT R&D, Vol. 39, No. 5, pp. 785-792 (1990).

(平成4年8月3日受付)

(平成5年6月17日採録)



加藤 恒昭 (正会員)

昭和34年生。昭和56年東京工業大学電気電子工学科卒業。昭和58年東京工業大学総合理工学研究科電子システム専攻修士課程修了。同年，日本電信電話公社横須賀電気通信研究所に入所。自然言語理解に関する研究に従事。現在，NTT 情報通信網研究所メッセージシステム研究部主任研究員。電子情報通信学会，人工知能学会各会員。