

格充足原理に基づく係受け解析手法

— 構文原理とそのアルゴリズム —

立岡章* 鈴木勝男* 中村澄江* 林達也** *FDL

**富士通研究所

我が国における自然言語処理の研究はここ数年来非常に活発に行われてきているが、とりわけ機械翻訳システムの製品化がこの流れに拍車をかけてきたと云えよう。しかしながら、いわゆる“実用的”と称されている機械翻訳システムの研究開発も、ここ数年の間著しい技術的進展が見られたとは云い難い情況にあり、あらためて自然言語処理の難しさを思い知らされているのが正直なところではないかと思われる。

とはいへ、この間における努力の結果、今や基本的な問題点は何であるのかが明らかになって来つつあるように思われる。

筆者等は、これまで、大量の日本語文章の事例に基づき従来の解析手法の問題点を調査分析し日本語の論理を把握することに努めてきたが、一応の結論らしきものを得たのでここに報告する。本稿の主旨は、日本語をそのマクロな構造から把握しなおすことにより日本語解析手法に新たな展開の可能性がひらけてくることを示すにある。このことを以下の項目に沿って説明する。

- (1) 係受け解析の基本的な問題点
- (2) 構文原理 …… 日本語の構造を規定する基本的枠組み
- (3) 構文原理に基づく係受け解析手法 (アルゴリズム)

Dependency Analysis Using the Case Satisfaction Principle

Akira Tateoka* Katsuo Suzuki* Sumie Nakamura* Tatsuya Hayashi**

* Fujitsu Documents Service Ltd.

** Fujitsu Laboratories Ltd.

1015 Kamikodanaka, Nakahara-ku Kawasaki 211, JAPAN

Through analytical research on the problems in conventional methodologies for analyzing the Japanese language and an intensive study of the logical foundation of the structure of Japanese, the authors have gained an understanding of the structural principles of Japanese and developed an innovative analysis methodology based on those principles. The new methodology provides us a clear perspective on natural-language processing. This paper summarizes the authors' theory and methodology with a discussion of

- 1) the fundamental problems in the existing analysis methodologies,
- 2) the structural principles of Japanese, and
- 3) an analysis methodology for Japanese based on those principles.

1. 係受け解析の基本的な問題点

これまで機械翻訳システム等で実施されている解析手法の多くは、文をいわゆる木構造として解析しようとするものが一般的であるように思われる。

ここでは、その最も原始的なモデルを示し、そこでの基本的な問題を指摘する。

1.1 解析手法の一般モデル

後の議論のために、最も一般的であると思われる解析手法の1モデルを簡単に説明する(図1)。

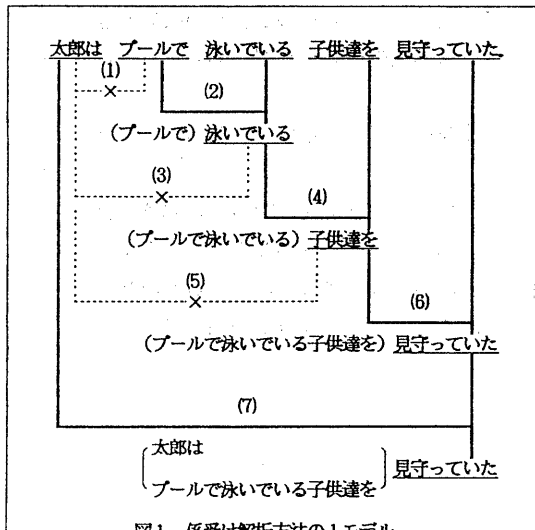


図1 係受け解析方法の1モデル

このモデルにおける係受け解析は文頭から文末方向へと進められ、隣り合う二つの文節同士の係受け関係を判定しつつ行われる。ここに、二つの文節の係受け関係が成り立つならば、この二つを木構造の1ノードとして縮退し、以降はこのうちの右側の文節(係り先文節)でもって代表させ、これと直前の文節との係受け判定に進む。一方、二つの文節の間に係受け関係が成立しない場合には、文節1ヶ分だけ文末方向にずらし、新たに二つの文節についての係受け判定処理を行う。このように、二つの隣り合う文節同士を順次縮退してゆくことにより、最後には文全体を1ヶのノードに縮退せしめ、これをもって係受け解析を完了する。

図1の例においては、二つの文節同士の係受け判定(および場合に応じて縮退処理)が(1)~(7)の順に進められる。

このような手法には以下のごとき特徴がある。

- 解析木の生成という目的によく適合している。
- 2項関係にほぼ全面的に依存している。
- 文中の部分部分ごとに順次係受け関係を決定してゆく(局所的・逐次的・決定論的に進める)。

従って、解析の精度を向上させるための手段としては“二つの文節同士の係受け関係の成否判定”をいかに精密化するかが中心的な課題となっている。特に、語と語の間の意味的な結合性(意味関係)を利用して精度の向上を図ろうとする方法が熱心に研究されている。

1.2 解析手法の基本的問題点

上の手法には次のような基本的な問題がある。

(1) 意味関係

意味関係を解析精度の向上に利用するには、各語への意味情報(意味カテゴリ等)の付与、語のもつ意味的な多面性、および“意味関係”規則の記述等の困難かつ膨大な整備作業を必要とするのみならず、“意味”の概念や分類法そのものもまだ明確に確立されていないのが現状であり、満足すべき成果が得られているとは云い難い。これは確かに重要な課題ではあるが、それ以前に日本語の構造について解決すべき課題がまだ残されているのではなからうか?

(意味関係は今後の研究に期待することとし、これ以降は専ら意味関係を抜きにして議論を進める)

(2) 解釈の多義性の認識

一般に文は人間が読んでも幾通りにも解釈されることが頻繁にある。たとえば、図1の例では“プールで”の係り先は“泳いでいる”であると決定されてしまうが、実際には“プールで泳いでいる”と“プールで見守っていた”の両方に解釈することができ、どちらも等しく正しい解釈である。これを機械が一方的に決めてしまうとすれば、そのこと自体が既に誤りを内包していることになる。理想的な解析法とは、このような場合にはこの二つの解釈を提示するものでなければならない。上のモデルにおいては、局所的・逐次的・決定論的に解析を進めるため、このような解釈の多義性を認識することが困難である。

(3) 文法記述の可能性

日本語文においては各文節ができるだけ近く of 文節に係るように配置されるという一般則がある⁴⁾

(卑近則と呼ぶことにする)。この通りに書かれている文であれば、上のモデルによって正しい解釈を得ることは可能である。しかし、全ての文章が卑近則を完全に満たしているわけではなく、近く of 文節を飛び越えて遠く of 文節に係るような例も頻繁に見られる。このような例に対しては、上の方法を単純に適用しただけでは正しい解釈が得られない。係受け判定や縮退といった処理そのものを制御することが必要となる。

たとえば次の例を見てみよう。

- (a) 図に示した方法でシステムを起動する。
(b) システムを図に示した方法で起動する。

(a)と(b)はともに同じ意味であるが、(a)では全ての係受けが卑近則を満たしており、(b)では“システムを”が卑近則に反して“起動する”に係っている。

(a)が上記方法によって正しく解釈されることは明らかであろう。

一方、(b)に対して、この方法を単純に適用すると“システムを図に示した”というように誤った解釈をしてしまうことになる。そこで“システムを→示した”の処理をひとまず保留して“示した→方法”の係受け処理を優先することが必要となる。つまり、係受けや縮退処理をどのような場合にどのように行うべきかということを規則として指示しなければならないことになる。(上の例(b)は、この方法において一体どのような規則を用いれば正しく解釈できるのだろうか?)これは実際には非常に困難な問題であり、見通しのよい規則が得られない。この種の問題は到るところに見られ、いきおい文法規則も複雑かつ大量になりがちである。

(4) 2項関係以外の構文規則

上の問題は、別の視点から見ると“文の構造は単に2項関係だけで規定されるものではない”ことを予測させる。たとえば上の例(b)において、2項関係から得られることは“システムを”の係り先として

“示した”と“起動する”の両者があり得るということだけであり、この両者のうちどちらに係り先とすべきかは2項関係からは出て来ない(意味関係からも出て来ない)。これは2項関係とは別の論理によって決定されるべき問題である。

別の例として、次の例文(c)~(f)における〈は〉の係り先を考えてみよう。

- (c) 彼は××の免許を持っている。
(d) 当社は××の免許を持っている。
(e) 彼は××の免許を持ち3年以上の実務経験がある。
(f) 当社は××の免許を持ち3年以上の実務経験がある人を求めている。

問題は(f)において“当社は”→“持ち”という係受け関係の成否をどのように判定したらよいかである。詳細は省略するが、この問題も2項関係だけでは解決が困難であり、やはり別の論理(“〈は〉は連体修飾節の一員にならない”など)が必要になってくる。

このように、文の構造には、2項関係を基礎としながらも、なおそれ以外のいわばマクロな構文論理が働いているように考えられる。前記したモデルにおいては、このような論理を吸収することが困難であり、文法規則が一段と複雑に(非論理的に)なりがちである。

1.3 構文解析の課題

以上の議論から、少なくとも次のような課題が解決されていなければならないことがわかる。

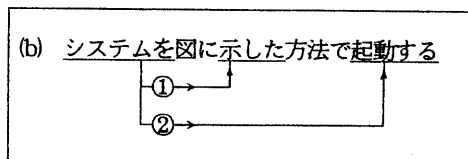
- ー 解釈の多義性の認識：現実の文は人間が読んでもいろいろに解釈されることが多い。従って、このような解釈の多様性をも明解に説明できる(複数の解として提示できる)ものでなければならない。
- ー マクロな構文論理の反映：前述したごとく、文の構造は、2項関係を基本としつつも、なおマクロな論理によって規制されていると考えられる。このような規制を具体的な規則や機構として実現する手段が必要である。

2. 日本語の構造を規定する基本的枠組み

しからば、このマクロな論理とは具体的にどのようなものなのであろうか？ まず、懸案の上記(b)の問題の解を示し、次いで、日本語の構造がいかなる枠組みによって規定されているのかについてあらためて見なおすことにする。

2.1 例(b)の問題の解

例文(b)において“システムを”の係り先は下のごとく二通り考えられるのであった。



問題は“この文では②の方が正しい係受けである”ことを決定する論理は何かである。

①②の係受け解釈それぞれに応じて文を節ごとにまとめてみよう：

①：[システムを図に示した方法]で起動する。

②：システムを[図に示した方法]で起動する。

両者を比べると、解釈①は何か欠けて未完の文の感じがするのに対し、解釈②ではそのような欠落感がなく自然な文章として受けとめられる。

これを論理的に詰めれば、次のように云えよう。

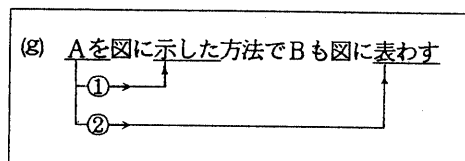
解釈①では述語“起動する”に“～ヲ”という必須の情報が欠けており、このままでは陳述として全うされていない。一方、解釈②では述語“起動する”が“～ヲ”という情報を満たしており陳述としての完成度が高い。

さらに、もう一つの述語“示す”についても同様な観点から見ると、解釈①②ともに“～ヲ・～ニ”という必須の情報が満たされている。（[]の中は①②ともにごく普通の正しい解釈である）

即ち、解釈①と解釈②のいずれが正しいかは“陳述としての完成度”によって判定されるのであり、具体的には「2ヶの述語‘示す’‘起動する’それぞれにおける必須格の充足性によって決まる」ものと考えられる。

これが上記(b)の問題の解を与える論理である。

この裏付けとして、次の例文(g)を考えてみる。



今度は係受け解釈①の方が正解となる。これもまた上と同じ論理によって説明されることは明らかである。

2.2 格の充足性

以上の考察から次の如き仮説が得られる。

格充足の原則

文章においては、それぞれの述語がその必須格を - それが自明でないかぎり - 充足しようとする傾向にある。

以下では、この‘格の充足性’や‘自明な格’の意味するところをもう少し具体的に把握することを試みる（図2の事例を中心に検討する）。

3.2 間接アドレスアクセス法

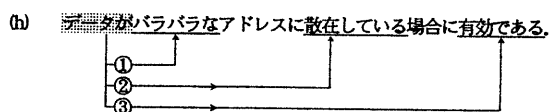


図2 事例（文書の一部）

図2は、ある文書の一部であり、項目のタイトル“3.2 間接アドレスアクセス法”の直後に本文(h)がはじまっている。この文(h)において“データが”の係り先は①②③の三つが考えられるが、正解は②である。これを具体的な方法でもって示そう。

以下では、それぞれの解釈①②③に対して、各述語がその必須格を備えているか否か（格充足度）を数量的に評価することにする。評価は減点法によるものとし、必須格はガ、ヲ、ニの範囲に限るものとし、ある述語がその必須格を一つ欠くごとに20点減点するものとする。

図3に、それぞれの解釈に対する必須格の充足性の評価結果（減点値）を示す。

①の解釈	{データがバラバラなアドレス}に散在している場合}に有効である
評価: -40	0 20 (ガ) 20 (ガ)
②の解釈	{データが [バラバラなアドレス] に散在している場合}に有効である
評価: -20	0 0 20 (ガ)
③の解釈	データが { [バラバラなアドレス] に散在している場合} に有効である
評価: -20	0 20 (ガ) 0
必須格: バラバラだ:ガ 散在する:ガ、ニ 有効だ:ガ	

図3 事例(h)における解釈①②③の格充足性評価

この結果、格充足度という観点からすれば②or③の解釈の方が妥当性が高く、①は妥当性が低いと判定される。このように、本例では、格充足の原則により三つの解釈が二つに絞られることになる。

では②と③のいずれかを選ぶべきか？（正解は②なのだが）……その前に‘格の充足性’の意味を理解するために、もう一つ、図2を变形した例を見てみよう（図4）。

図4は、図2でのタイトル“～アクセス法”を本文に取り込んだ文について、図3と同様に格の充足性を評価したものである。

この場合には、評価値により確実に正解②が出てくる。

②の解釈では全ての述語が格を満たしていること（減点0）に注目された。即ち、この解釈ならば文が単独で必要な情報を全て満たすことになるのである。実際この一文だけで内容が過不足なく完結しており‘前後の文章を見ないとわからない’といった部分が一つもない。

これに比べると、前の図2.3の事例では②③いずれの解釈においても、格の不足が生じる。即ち、この文だけでは完結しておらず、前後の文章を見ないとわからないことになる。

実際、直前にタイトル“～アクセス法”がある。これが“自明な格”として述語“有効である”を暗黙裏に満たしているのである。

（このような議論は、文脈処理への一つのきっかけとなる）

以上のように、文においては、各述語が必須格を（明示・暗黙いずれの形にせよ）引きつけようとする強い潜在力を持っていると考えられる

さて先刻の問題‘図2における解釈②と③のうちいずれを選ぶべきか’の問題に移ろう。

答えは簡単である。ここに卑近則が働くのである。つまり「解釈②と③が格の充足性という点で同等な評価値をとるならば、このうち卑近の係受けを優先させる」ことにすればよい。

これにより正解②が選ばれる。

××法はデータがバラバラなアドレスに散在している場合に有効である。	
① ② ③	
①の解釈	××法は { [データがバラバラなアドレス] に散在している場合} に有効である
評価: -20	0 20 (ガ) 0
②の解釈	××法は {データが [バラバラなアドレス] に散在している場合} に有効である
評価: 0	0 0 0
③の解釈	××法はデータが { [バラバラなアドレス] に散在している場合} に有効である
評価: -20	0 20 (ガ) 0

図4 例文(h)の变形例の格充足性評価

2.3 構文原理

以上に考察してきたごとく卑近則および格充足の原則は文の構造を規定する基本的な枠組みを与えるものであると考えられる。この二つに“非交叉の原則”を加えて以下のごとき構文原理が得られる。

構文原理

文は、基本的に以下の三つの条件を満足するように構成される：

- 非交叉則：係受けは交叉しない。
- 格充足則：文中の各述語は必須格を満たす。
- 卑近則：各文節は卑近の語に係ろうとする。

言葉（文）は、人と人の中でやりとりされる情報伝達の形式であり、この意味で‘陳述’とは本来必ず何らかの情報を担っているはずのものである。

言葉によって伝達される情報を一種の‘データ’として見るならば、そのデータの構造には自ずから規約（人と人との間の普遍的なインタフェース）がなければならず、しかもその規約は人間の言語機能という生理に依拠している以上、できるかぎり単純な構造を持つていなければならない。

この意味で、非交叉の原則や卑近則は言葉に対するごく自然な要請であると理解されよう（言葉は線条的であるから）。

次に陳述が情報を担うということの意味を考えてみよう。

たとえば“読んだ”という一語はそれ自身単独では（前後の脈絡なしには）陳述として受けとめられることがない。それは単に“読んだ”という語の静的概念としてのみ受けられ、“……?(それが?)”という反問を誘発するだけのことである。人はそれを情報として理解することはない。

“私はこの本を読んだ”となってはじめて情報として（即ち陳述として）受け付けられる。

つまり、述語はそれ自身単独では文たり得ず、必要最低限の格を伴うことによって、その述語たる意味・役目を全うするのである。

このように、格充足の原則とは“言葉が伝達情報として成り立つための要請である”ということができる。

- 非交叉の原則はほとんど例外がなく、まず絶対的な規制として働く。
- 格充足の原則もまた（自明な格を考慮に入れるならば）例外はまれであり、強い規制力を持つ。
- 卑近則には例外が頻繁に見られ、強制的なものではない。これは、いわば文における‘場の力’として働くものと考えられる。（例外があるとはいえ、それは格充足というフィルターを経た上で見るならば必ずしも例外とは云いきれない）

3. 構文原理に基づく解析手法

文は文節から構成され、この文節の間の係受け関係で構文が定まる。この文節の間の係受け関係を規制するのが構文原理なのである。

従って、日本語解析の処理論理は、基本的に、“この原理を最も良く満たすような解釈”を選びだすための論理でなければならない。

非交叉の原則は文節の相対位置の簡単な算術演算で判定できる。問題は、格充足の原則と卑近則をどのように調和させるかである。

‘格充足性’は前に示したような方法で定量的に表わされる。もう一つの‘卑近則’も同じく数量化し、これらに基づいて‘構文の評価値’を算定することを試みよう。

つまり‘卑近則による力’と‘格充足性による力’の二つから構文全体の力学的安定度（妥当性）を求めようというわけである。

この考えから、以下に述べるような解析手法（アルゴリズム）が出てくる。

以下では、まず、この手法の基本的な処理の流れを説明し、次いで、構文解釈の妥当性の評価方法を説明し、最後に、この手法の効果や課題について述べる。

ここに、先に取り上げた事例（図2の文(h)）に適用した例を図6～図8に示しつつ説明することとする。（図における①～⑥は文節の通番を示す）

3.1 処理の流れ

図5に処理の流れを示す。

(1) 係先候補の洗い出し

係受け文法規則（2項関係規則）に基づいて、処理対象文中の文節それぞれに対し、その文節が係り得る文節（係先候補）を全て洗い出し、当該係元文節とともに係先候補リストとして記録する。図6に係先候補リストの事例を示す。

(2) 係受解釈ケースの洗い出し

文中の文節それぞれについて、その係先文節を候補リストから一つずつ取り出すことにより1ケの係受解釈が得られる。但し係受けが交叉するようなものは破棄する。（以後解釈ケースと呼ぶ）

この処理に係先候補リスト中の候補の組合せ全てについて行い、可能な解釈ケースを洗い出す。この中に必ず正解があるはずである。

これらの処理は文節番号に関する簡単な算術的演算で行なえる。

図7.8に解釈ケース群の事例を示す。

(3) 構文評価による絞り込み

(2)により解釈ケース1ケを得る都度、その解釈ケースの構文的な妥当性を評価し、妥当性の高いケースを保存する。（構文評価の詳細は次項）

即ち、まず当該解釈ケース（現ケース）の構文評価値を求め、これと最適解保存部にある解釈ケース（仮正解と呼ぶ）の構文評価値とを比較し、この結果により以下の処理を行う。

- ・もし最適解保存部に仮正解が存在しない（最初の1回目）ならば、現ケースを保存する。
- ・もし現ケースの評価の方が良いならば、仮正解を消去し、代わりに現ケースを保存する。
- ・もし現ケースと仮正解の評価が同等ならば、現ケースもまた最適解保存部に保存する（複数のケースが保存されることになる）。
- ・もし現ケースの評価値が仮正解よりも悪いならば、現ケースを破棄する。

○ (2X3)の処理を全ての解釈ケースについて行うことにより、最終的には、最も妥当と考えられる解釈ケースが保存部に残されることになる。

これが求める解である（図7ではケース4）。

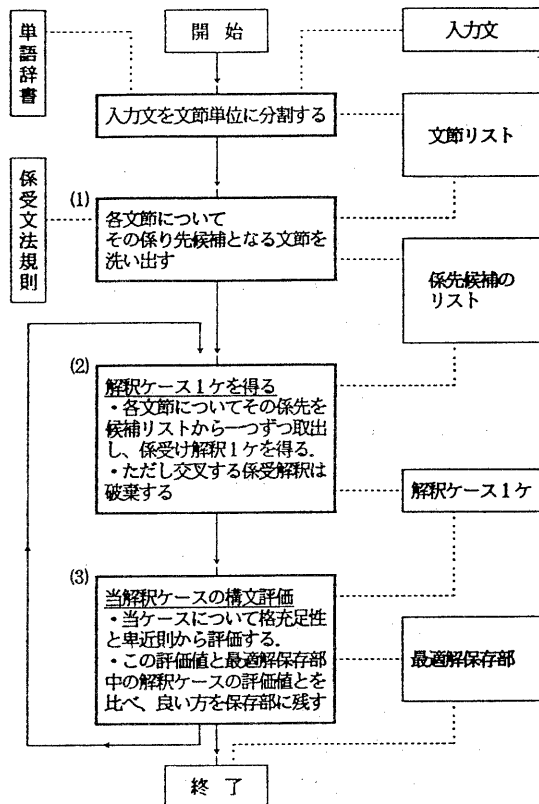


図5 処理の流れ

① データが ② バラバラな ③ アドレスに ④ 散在している ⑤ 場合に ⑥ 有効である					
係元文節	係先候補文節	係受評価 (減点値)			
		確度	準近則		
① “データが”	② “バラバラな”	0	0		
	④ “散在している”	0	10		
	⑥ “有効である”	0	20		
② “バラバラな”	③ “アドレスに”	0	0		
	⑤ “場合に”	0	10		
③ “アドレスに”	④ “散在している”	0	0		
	⑥ “有効である”	0	10		
④ “散在している”	⑤ “場合に”	0	0		
⑤ “場合に”	⑥ “有効である”	0	0		

図6 係先候補リストの事例

3.2 構文評価

ここでいう構文評価とは、既に或る係受け解釈が与えられているものとし（解釈ケース）、その解釈の構文的な妥当性を評価するものである。

この‘構文解釈の妥当性’の評価には、種々の方法が考えられるが、ここでは最も基本的・普遍的である構文原理のみに準拠することとする。

即ち、その解釈が「どの程度構文原理を満たしているか」を評価するものとする。

構文原理のうち、非交叉の原則は既に解釈ケースの洗い出しで折り込み済みである。故に、格充足の原則と卑近則の両者をどの程度満たしているかを評価することになる。

ここに、評価は減点法で行うものとする。

- ・格充足性の評価値は「2.2 格の充足性」で述べた方法で算定する。

図7の‘格充足評価’欄に各解釈ケースの格充足評価値の事例を示す。

- ・卑近則については、解釈ケースにおける係受けそれぞれについて卑近則評価値を求め、それらの合計を以て当解釈ケースの卑近則評価値とする。

ここに、1ケの係受けに関する卑近則評価値は、次のごとく設定されるものとする：

ある文節の係先候補の洗い出しにおいて、係元に最も近い係り先候補に対しては評価値（減点値）を0とし、次に近い候補に対しては評価値（減点値）を10とし、以下、遠くなるに従って評価値（減点値）を10ずつ増加してゆく。

図6の‘卑近則’欄に各係受け候補の卑近則評価値の事例を示す

図7の‘係受け評価’欄に各解釈ケースの卑近則評価値の事例を示す。

- ・以上二つの評価値（格充足性と卑近則）の合計を以て当該解釈ケースの構文評価値とする。

図7の‘総合評価’欄に各解釈ケースの構文評価値の事例を示す。この事例で評価値の最も良いものはケース4であり、これが求める解となる。

確かにケース4が正解である（図8参照）。

のみならず、次点となったケース1,5 も構文的には大いにあり得る解釈である（従来の手法ではケース1のごとく解析してしまいがちである）。

このように、この構文評価法は、構文原理のみに基づく単純な計算であるにもかかわらず、構文解釈の妥当性を良く反映しており、正しい解釈から構文的にあり得る解釈へ、さらに全くあり得ない解釈へと順に評価が下がってゆく（減点値が増えてゆく）。

ケース No	各文節の係り先					構文評価（減点値）			備 考
	①	②	③	④	⑤	係受け 評価	格充足 評価	総合評価 (合計)	
1	②	③	④	⑤	⑥	0	40	40	△ (次点)
2			⑤			10	60	70	
3		⑤	④			10	40	50	
4	④	③	④			10	20	30	◎ (正解)
5	⑥	③	④			20	20	40	△ (次点)
6			⑥			30	40	70	
7		⑤	④			30	40	70	

図7 解釈ケースのリスト (事例)

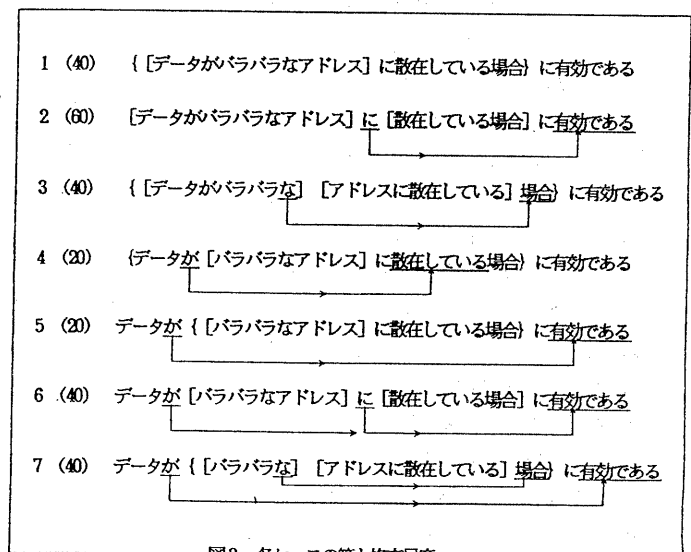


図8 各ケースの節と格充足度

3.3 手法の効果

この手法の要点は、「2項関係に基づいてあらゆる可能な係受け解釈を洗い出してしまう、それらの解釈群から妥当なものを絞り込む」という戦略をとっている点、および、この絞り込みの手段（フィルター）として「構文原理といった文全体からする構文評価法を用いる」点にある。

これにより、以下のごとく、先に指摘した‘解析上の基本的な問題点’の殆ど全てが、かなり明解に - 論理的に - 解決されることになる。

(1) 解釈の多様性の認識

ここで示した構文評価手法は、素朴ながら構文解釈の妥当性の程度を良く反映している。

従って、評価値が上位の複数の解釈ケース（たとえば評価値の差が10以内の上位の解釈ケース群）を保存するようにすれば、人間が読んでも多様に解釈されるような文は、それなりの多様性を以て複数の解として出力されることになる。

(2) マクロな構文論理の吸収

文の構造は、単に2項関係のみによって規定されるものではなく、文全体からの規制等2項関係以外の規制があるはずであり、本手法は、これを具体的に取り込む手段を提供している。

このような、文全体からの構文的規制という観点を明確に意識した研究ならびに具体的方法論は、従来にはあまり多くは見られなかったものであり、今後の課題として大いに発展が期待できる。

(3) 文法記述の可能性

従来の解析手法では文法規則が複雑化しつつ大量の規則が必要になるという傾向にあり、文法開発作業に大きな負荷がかかっていた。本手法においては、言語現象が2項関係とマクロな関係とに分離して把握されるので、見通しの良い規則が記述できることになる。

(4) 構文評価の効用

2項関係によってあらゆる解釈が洗い出されること、および構文評価によってその妥当性が評価されることから、以下のことが可能となる。

- ・ 構文評価（減点）の要因を保持しておくことによ

り、文章推敲情報を出力したり自動的に文章を推敲することが可能となる。

- ・ 対話処理においては、指摘すべき情報や問合わせるべき情報が的確に抽出できるばかりでなく、対話により得た情報を迅速に反映することも可能となる。
- ・ この手法においては、個々の解釈ケースに対してその評価要因が明確に意識されている。従って文法規則の整備やデバッグは、（アドホックでなく）より論理的に進めることが可能となる。

(5) 並列処理

2項関係による係受け解釈ケースの洗い出し以降、解釈ケースそれぞれの構文評価は互いに独立して無関係に進められる。従って、この解析手法を並列処理コンピュータにおいて実現することにより、超高速の処理が可能となろう。

本手法は完成されたものではなく未だ検討の余地はあるものの、上に見る如くその基本的な考え方は日本語解析に新たな可能性をひらくものと思われる。

○ 補足説明 — これまでの説明を少し補足する。

－ 解釈ケースの個数

係先候補の組み合わせによって解釈ケースを洗い出すのであるが、この方法だと、一見極めて多数の解釈ケースが発生するよう思われがちである。

しかし、実際には - 係受け規則と非交叉の原則により - ケースの個数はそれほど多くはない。

－ 構文評価に関して

- ・ 本手法においては構文評価手段が重要な働きをするが、2項関係以外の構文的な法則や性質等に基づく評価手段は今後の研究にまっところが大きい。

構文原理は基本的なものではあるが、これのみで構文を完全に評価できるわけではない。

- ・ 配点方法についても、詳細は今後の課題である。ただ、先のごとき単純素朴な算定法でもかなり有効な結果を得るのは驚くべきことである。

- ・ 図6の‘確度’欄については触れなかったが、ここに表層文法や意味関係等に基づく2項関係の確度を設定し、これを構文評価に組み込むことにより、さらに精度の向上が図られよう。

- 必須格について

- ・ここでは必須格を助詞が、ヲ、二等としたが、これを深層格にした方がより妥当であろう。
- ・必須格の定義については、ここでは曖昧なままに常識に委ねているが、これも突き詰めれば語と語の結合度の一環として把握されるべき課題であろう。

4. 今後の課題

(1) 構文評価手法の追求

この解析法の決め手は構文評価手法の整備にある。基本は構文原理にあるとしても、これ以外にも構文にかかわる有効な論理がまだまだ存在すると思われる。

(2) 構文原理の詰め

この原理が本当に間違いのないものか否かは、なお今後の検証を必要とする（かなりの事例に基づいて検証済みではあるが）。特に格充足の原則については、必須格とか自明な格などという考えの妥当性や具体的な扱いについての詰めが必要と思われる。今後ともよりの確かな形での構文原理を求めてゆく必要がある。

(3) 未解決の課題

次のような問題は依然として残されている。

- 共有格（二つ以上の文節に同時に係る現象）
- 並立句の認識
- 述語と述語の間の係受け関係（連用中止等）
- 連体修飾節と被修飾語の間の深層の格関係
- 名詞と名詞の深層格関係（ノの問題など）

これらの問題は、本稿の手法に直接関与するものではないが、これも「陳述は情報を担うものである」という観点から見れば、解決の糸口が得られるのではないかと考えられる。

(4) 文脈処理

先に述べたように自明な格という考えは、文脈処理への一つのきっかけを与えるものと思われる。

これもやはり「陳述が情報を担う」という観点に立って把握することが必要となろう。というのも、自明な格といった現象は、個々の文章単独では本来あり得ないことであり、前後の脈絡情報と密接に絡むからである。

(5) 他言語への応用

ここで示した手法は、日本語における文節同士の係受け関係（2項関係）から出発しているが、この手法の基本的な考え方を形態素のレベルから適用することも考えられる（文字レベルからの適用も可）。

さらに、この基本戦略を他の言語の解析に応用することも可能であろう。

5. おわりに

本稿で提示した構文原理や解析手法は、その基本的な考えを理解されるならば、何れも、ごく当たり前のことと思われよう。筆者等もこの意見に同感であるばかりか、この“あたりまえ”であることこそまさに自負すべき点であると考えている。

ここで特に強調しておきたいことは、これらの理論や手法と同時に「陳述は情報を担うものである」という視座が得られたことの持つ意義である。

この観点から見れば、各種の言語現象は、かなり明確に位置づけられてくるように思われる。

この意味で、本稿が日本語処理 及び 自然言語処理における新たな展開に少しでも寄与できるとすれば幸いである。

参考文献

- (1) Hayashi, T. et al: ATLAS: Fujitsu's machine translation system, Proc. Asia and Pacific regional Conference on Translation, p20 (1986)
- (2) 首藤公昭：日本語文解析の難しさ、電子通信学会誌、Vol70-9, pp. 891-896 (1987)
- (3) 柴谷方良、影山太郎、田守育啓：言語の構造、- 意味統語篇 - くろしお出版 (1982)
- (4) 尾関和彦：最適文節列を選択するための多段決定アルゴリズム、電子通信学会技術研究報告、SP 86-32, pp. 41-48 (1986)