

着目語句周辺の文脈情報を利用して曖昧性を解消する 構文解析方式

1 B-8

青沢 秀憲 側島 康博 高木 朗

(株) CSK

1. はじめに

語と語の間の依存関係の判定は、一般に品詞や意味素性などの情報を、語に制約条件として付与しておき、一方の語の品詞や意味素性が、他方の語がもつ制約条件に合致するかどうかを調べることによって行なわれる。

解析すべき文が単純であり、係り合う候補となる語同士の間で直接的に両者の依存関係の可否の判定や多義・多品詞の曖昧性の解消を行ない得る場合は、これらの制約情報は、一定の効果を果たし得る。しかし、通常、2つの語が互いに特定の品詞・意味で依存関係を持ち得るかどうかをそれらの語のみで決定出来る場合は限定され、それらの語以外の周辺の語によって、依存関係の可否が左右されることが少なからず起こる。

そのような場合に対処するためには、周辺の文脈に依存した単語の構文的・意味的性質を辞書中に記述し、構文解析中にそれを読み取って周辺の文脈とつぎ合わせ、依存関係を判定するメカニズムが必要となる。本稿では、このような情報を記述するための制約式の導入と、それを利用した構文解析制御方式について述べる。

2. 従来の問題

語の制約情報を品詞や意味素性を用いて記述するということは、その語が持つ構文的意味的振舞いを、いわば、名詞1語相当の表現手段で表わす、ということの意味する。これによって、主として以下の2つの問題が生ずる。

(1) このような表現手段の情報表現能力は極めて限られたものであるにも拘らず、同一の品詞・意味カテゴリに属する語の間でさえ、個々の語の振舞いは多様に変動する。従って、品詞・意味素性等による制約だけでは、それら個別的振舞いを表現しきれず、解析精度を維持しようとする、勢い、解析ルール中に例外处理的な形でこれを記述せざるを得なくなる(プログラムの表現能力は充分高いので)。しかし、例外的な処理は出来るだけ排除し、個別情報は辞書中に記述する形が望ましい。

(2) 前述のように、語の振舞いは、周辺の文脈のあり方によって影響を受けることが少なくないが、そうした場合、周辺の状況によって変化する性質を制約情報として記述する必要が生ずる。これを行うには名詞1語相当の表現手段では不十分であり、必然的に文と同等の表現能力をもった表現手段が必要となる。

この様な問題を解決するためには、周辺の文脈のあり方を条件として記述し、その条件が満たされた場合に、語が示す振舞いを限定する表現手段を導入することが、まず必要となる。又、構文解析は、文中の各語の解析に際して、この制約記述を解釈し、評価する能力と、制約によって記述された周辺文脈のあり方を実際の語句の並びと照合するための“視野”をもち、文を読み進んだ後に初めて文頭側の語の性質が確定するような場合には、そこまで遡って解析をやり直す等の能力が必要となる。

A Natural Language Parsing Mechanism Using Constraint Formulas for the Contextual Disambiguation
Hidenori AOSAWA, Yasuhiro SOBASHIMA, Akira TAKAGI
CSK Corporation, e-mail:aosawa@mtmng.csk.co.jp

3. 制約式の概要

3.1 多義選択や係り先選択における制約

まず、いくつかの例を示す。

- | | | |
|-----|-------------------------|-----------|
| 例1) | 彼は今日のうちに担当者にあたる。 | (調べるの意) |
| | 彼は私のいとこにあたる。 | (あてはまるの意) |
| | 彼がドスンと私(の体)にあたる。 | (ぶつかるの意) |
| | 彼女にあたる。 | (?) |
| 例2) | 主力はやはり政策に関連する内需関連株である。 | |
| 例3) | 人はすべて、赤い糸で結ばれている。 | |
| | 人はみな、過去になしてきた過ちを反省すべきだ。 | |
| 例4) | そう言うて行かないのは馬鹿だ。 | |
| 例5) | 率直に言うて行かないのは馬鹿だ。 | |
| 例6) | 英語に限らず、日本語も、習うのは大変だ。 | |

例1の「あたる」はみな主格や与格に人の意味をとるため、主格や与格の制約だけでは候補を選択しきれないが、

- ・「至急」「今日のうちに」etc.が係るなら「調べる」を選択
- ・「勢いよく」「ドスンと」etc.が係るなら「ぶつかる」を選択
- ・主格と与格の意味が異なるなら「あてはまる」を棄却
- ・「ている」etc.がつくなら「あてはまる」を棄却

のように文脈次第で絞り込めることがわかる。「彼女にあたる。」のように手掛かりに乏しい場合もあるが、我々の分析では、何らかの手掛かりが同一文に出現する場合が多く認められた。例2の副詞「やはり」は通常、直後の述語に係りやすいが、この例のように直前、または、直後に「は」句があれば、「やはり」の係り先は、「は」句と同一になる。例3の「すべて」「みな」も同様である。また、例4の「言うて」のように、通常、単独では係り先が曖昧であっても、例5の「率直+に+言うて」のように、一定の纏まりができる、係り先の種類が限定されてくる場合がある。例6は「限らず」がまとまってはじめて直後の「も」句と同じ係り先になる例である。

このように考えていくと、単語単位の制約としては持ちようがなかった周辺文脈の情報をキーとなる語の情報として登録し、解析に利用できるようになれば、解析精度の向上に寄与すると考えられる。

3.2 制約式の記述形式

以上の検討に基づき、表1に示すようなシンタクスを持つ制約式を導入する。各語の振舞いはこの制約式によって記述され、辞書に登録される。表2に簡単な例と内容を示す。「着目句」は係り受け判定のために着目している語句ではなく、文頭から着目位置までの語句のうち、制約式を評価するために着目している語句を表し、「直前句」等はその相対位置にある語句を示す。

表1 制約式の記述形式(抜粋)

制約式	: 《条件部⇒返却値》
条件部	: [対象. 条件[{ &}対象. 条件]...]
返却値	: [選択指定 係先指定]...
選択指定	: 対象={選択 選択・他棄却 棄却}...
係先指定	: 対象. 係り先[={≠}] [対象. 係り先 対象. 条件]
対象	: {着目句 直前句 直後句 前方句 後方句 上位句 下位句}... [対象]
下位句	: {主格 対象格}... [主体属性 強弱属性]... [任意下位]
条件	: 要素[={≠}]値群
要素	: {見 意 品 活 係り状態 背景}...
	(以下略)

表2 制約式例とその内容

<p>A) 「あたる」(調べるの意)の持つ制約式 内容 : 「至急」に係るなら自身を選択し他の多義を棄却 制約式 : 《 [直前句. 見 = (“至急”) &直前句. 品 = (副詞) &直前句. 上位句 = 着目句] ⇒ [着目句 = 選択・他棄却] 》</p>
<p>B) 「やはり」の持つ制約式 内容 : 直前句が「は」、「も」なら直前句と同一係り先 制約式 : 《 [直前句. 見 = (“は” “も”) &直前句. 品 = (副助詞 格助詞)] ⇒ [着目句. 係り先 = 直前句. 係り先] 》</p>
<p>C) 「限る」の持つ制約式 内容 : 「限らず」は、後方の「も」句と同一係り先 制約式 : 《 [着目句. 上位句. 見 = (“ぬ”) &着目句. 上位句. 品 = (助動詞) &着目句. 上位句. 活 = (連用形) &後方句. 見 = (“も”) &後方句. 品 = (副助詞)] ⇒ [着目句. 係り先 = 後方句. 係り先] 》</p>

4. 構文解析概要

4.1 構文解析方式

次に各単語毎に辞書中に記述されている制約式を利用して、依存関係の判定を行なう構文解析方式について述べる。但し、制約式を解釈するインタプリタはまだ実装されておらず、単なる文字列のマッチングによる制約内容の同定を行なうに留まっている。

我々は、これまで、依存構造解析に基づく構文解析システムを開発してきているが[1][2]、本方式では縦型的な探索と横型的な探索を合わせ持たせている。その縦型的な部分では、従来の方式を踏襲し、多品詞選択ルールを用いて多品詞語の曖昧性を解消し、大域的係り先選択ルールを用いて周辺の構造を参照しながら接続助詞などの係り先を補正していく。

横型的な部分では、文頭までの視野を確保しつつ、着目する語位置の1つの品詞における全ての多義語に対して、文頭側に隣接する部分木(語)と係り受け判定を行い、同時に、文頭から着目する位置までに候補として残っている全ての語の組み合わせについて、各語の持つ制約式を評価して候補を絞り込み、可能な候補とその組み合わせを、随時修正し、保持していく。従って、係り受け判定のために着目された時点で正しいかどうか不明だった(ので候補として残された)語でも、後方の適当な文脈がそろった時点で、前に遡って棄却される。例えば、例7、例8においては、「あたる」が着目された時点では多義を棄却する根拠がないが、後方の語句(「ている」「事業」「山田太郎」等)が着目された時点で、「あたる」の多義が絞り込まれる。

例7) 彼があたっている事業は難題だらけだ。
例8) 彼のおじにあたる山田太郎です。

勿論、適当な文脈がそろわないなら、候補はそのまま残される。一般に、長文になれば候補は増えやすいが、逆に手掛かりも多くなるため、この方式では候補も増えにくい。図1に解析アルゴリズムの概要を示す。

4.2 制約式適用の具体例

図2は「至急彼にあたりたい」の形態素並び例である。本方式では、実線単位の多品詞選択は縦型で、実線内の多義選択は横型で行うが、これを(便宜上、品詞選択は除外し、「あたる」の多義も3個として)解析したものを表3に示す。この場合、4で「毒にあたる」の多義は依存関係判定で棄却され、他の多義が候補として残るが、

5で「調べる」の多義が持つ表2Aの制約式が適用され、「ぶつかる」が棄却されて「調べる」に絞り込まれる。

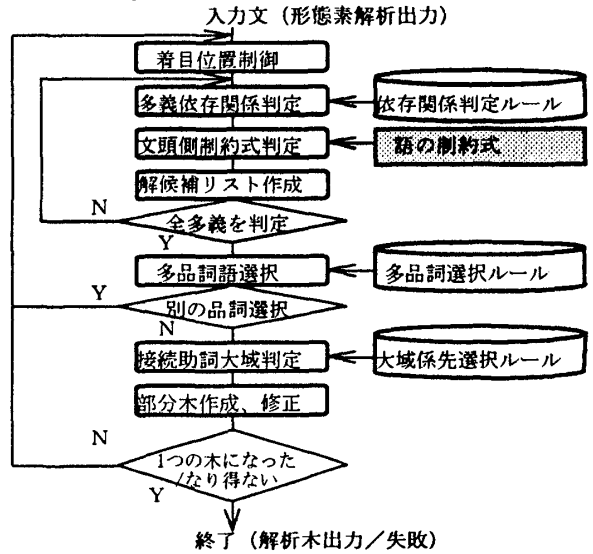


図1 解析アルゴリズムの概要

至急: 副	彼: 代名	に: 助	あたり: 名	たい: 名
		に: 格助	あたる: 動	たい: 助動

図2 形態素並び例

表3 「至急彼にあたりたい」の解析例

1) 至急: 彼	→係らない
2) 彼: に	→係る
3) 至急: 彼に	→係らない
4a) 彼に: あたる(ぶつかる)	→係る
4b) 彼に: あたる(毒にあたる)	→係らない(棄却)
4c) 彼に: あたる(調べる)	→係る
5a) 至急: 彼にあたる(ぶつかる)	→係る→制約式で棄却
5b) 至急: 彼にあたる(調べる)	→係る→制約式で選択

「主力はやはり政策に関連する銘柄である」を(多義多品詞抜きで)解析したものを表4に示す。7で「やはり」は「関連する」に係るが、8で「主力は」が「関連する」に係らないと判定されると表2Bの「やはり」の制約式が適用され、「関連する」から切り離される。そして14で「ある」に係っても15で「主力は」も「ある」に係るので切り離されない。

表4 「主力はやはり政策に関連する銘柄である」の解析例

1) 主力: は	→係る
2) 主力は: やはり	→係らない
7) やはり: 政策に関連する	→係る
8) 主力は: やはり政策に関連する	→係らない→“やはり”切離し
9) 政策に関連する: 銘柄	→係る
14) やはり: 政策に関連する銘柄である	→係る
15) 主力は: やはり政策に関連する銘柄である	→係る

5. おわりに

周辺の文脈に視野を拡大した上で制約式を取り入れ、多義や係り先の曖昧性を解消する手法について述べた。しかし、制約式の解釈方式は試作段階であり、制約式自体もまだ少ない。今後は、より多くの現象に対して評価実験を重ねながら改良していく予定である。

[参考文献]

- [1]青沢、石井、笹野、高木:多義多品詞選択ルールを採用した依存構造解析, 情報処理学会第50回全国大会, 1995.3
- [2]青沢、笹野、高木:曖昧性解消ルールを用いた依存構造解析, 人工知能学会全国大会(第9回), 1995.7