

共起関係データの蓄積と利用のための基礎実験

An Experiment for Acquisition and Usage of Co-Occurrences

松本 一則

黒岩 真吾

鈴木 雅実

榊 博史

Kazunori MATUMOTO

Shingo KUROIWA

Masami SUZUKI

Hiroshi SAKAKI

KDD 上福岡研究所

KDD KAMIFUKUOKA R&D Laboratories

あらまし 「例文」と「例文の正しい統語解析木」から共起関係を蓄積し、解析への応用を試みている。今回、2つの手法で共起関係の蓄積・利用を行った。一つは、句の中心を成す語(ガバナ)に注目するもので、書換え規則の兄弟節点のガバナに共起関係があると仮定し、このガバナの並びを蓄積する。もう1つの方法は共起関係を人手によって書換え規則に記述する手法で、使用したい共起関係やその引数を自由に定義できる。どちらも正しい統語構造と解析結果を比較することで、出現した共起関係を正の事例・負の事例に分類する。そして、負の事例のみの蓄積例を解析のあい昧性解消に利用できる。蓄積例の増加に伴う解析性能の変化を両手法において測定した。蓄積した共起関係の利用により不適当な木の抑止や正しい木の選択に効果が見られた。

Abstract

This manuscript describes two methods which acquire lexical co-occurrences information and utilize it. These methods obtain co-occurrence relations from each example sentence and the corresponding right syntactic structure of the sentence. First of the two methods treats the governors appearing on sister nodes of the syntactic structure as a co-occurrence. On the other method, co-occurrence relationships are described manually in the rewriting-rules. Both methods discriminate between the proper appearances of co-occurrence and wrong ones, using the right syntactic structure affixed to the sentence treated. The experiment is conducted for these methods, to observe the performances of the analysis using the stored co-occurrences data.

1. はじめに

自然言語処理システムの解析性能を向上させるためには、文法及び実世界知識のデータを精密化することが重要である。

我々は、既存の規則に不足している実世界知識のデータ(特に単語間の共起関係)を実例から蓄積することを目標に、検討と実験を行っている。今回、解析処理部と共起関係蓄積機構を組み合わせることにより、「例文」と「例文の正しい統語構造」から共起関係の抽出と解析への応用を試みた。その手法と実験結果および問題点について報告する。

実験結果は使用したパーザの能力や文章の種類に影

響を受けるが、筆者らの実験では、不適当な解析木の抑止や正しい解析木の選択に効果がみられた。

本研究に関連し、参考にしたものに、自然言語の分析による知識データ獲得の研究[1]、文法と意味解析規則の帰納的学習を文と意味表現から行う研究[2]、また、解析処理後に曖昧性の無い共起関係をのみ抽出する研究[3]があげられる。

本稿では、まず、使用したパーザとその出力について説明する。

3章では蓄積・利用のための簡易な手法について説明する。この手法は、統語構造における句の中心となる語(ガバナ)に注目するもので、

(1) 解析結果と本来生成すべき統語構造との比較に

よって、書換え規則ごとに兄弟節点に出現したガバナの並びを正・負の事例に判定する。

(2) 負の事例のみであるガバナの並びを曖昧性解消に用いる。

ことを特徴とする。さらに、共起の蓄積に用いる文例の増加に伴う解析性能の変化に関する実験結果を示す。

4章ではこの簡易な手法の8つの問題点を述べる。

5章では先の問題点の内7つに対応可能な蓄積・利用の手法を説明する。その修正点は、

(a) 蓄積・利用に用いる共起関係について、出現する共起関係やその共起関係の引数となる属性を入手によって書換え規則に記述する。

(b) 解析木選択時に、正の事例のみの共起関係データが出現している解析木の優先度を高くする。

(c) 共起関係データの蓄積を2パスで実現する。

である。

6章では修正した手法で、蓄積用文例の増加に伴う解析性能の変化に関する実験結果を示す。さらに、解析対象に出現する共起データが蓄積済みの事例に含まれる割合が解析性能におよぼす影響についての実験結果を述べる。

2. 使用するパーザ

我々の手法では共起関係の蓄積にパーザを利用している。今回は機械翻訳システムKATEの解析処理部を流用した。このパーザは拡張LINGOLを改良したものであり、[4]

- ・拡張CFGで記述された書換え規則を入力文に適用することで、各節点が属性情報を持つような解析木の集合を生成する。

- ・生成した木の各節点には、その節点が表す句のガバナが属性として与えてある。

- ・禁止したい部分木(禁止パターン)を登録しておく、その部分木を含むような解析木の生成が抑止できる。

ことを特徴としている。KATEが生成する解析木の例を以下に示す[例1, 2]。

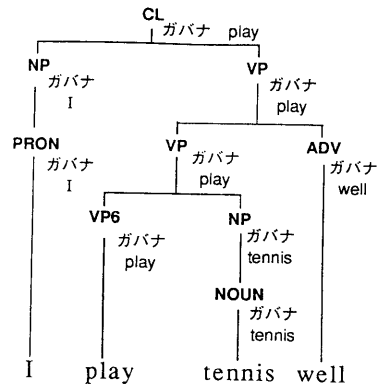
3. 簡易な手法によるガバナ間の共起関係の蓄積・利用

ここでは蓄積・利用のための簡易な手法について述べる。この手法とその実験結果は既に報告済み[5]であるが、説明のためここで詳しく述べる。

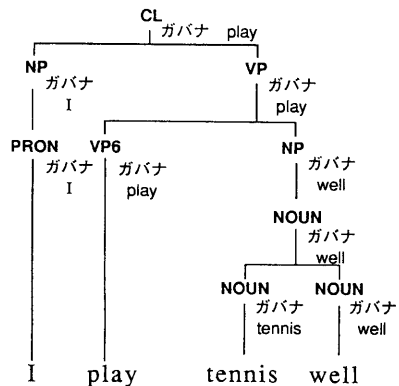
3.1. ガバナの共起関係の蓄積と分類手法

意味処理が不十分であると、生成する解析木には例

[例1 統語解析木の例]



[例2 統語解析木の例]



2のように、解釈として不適切な木が含まれている。しかし、例2の木でも、部分的にはガバナの正しい共起関係が存在する。そこで、既知の正しい統語構造と生成した解析木を比較して、使用した書換え規則ごとに下位節点に出現したガバナの並びを、「正しい局面で出現した(正の事例)」か、もしくは「誤った局面で出現した(負の事例)」かに自動判定している。この判定手法は次のようになっている。

まず、正しい統語構造と比較する時、

- (1) 同一の書換え規則が適用されており、
- (2) その書換え規則の各下位節点において、書き換えられた終端語列が一致している

ような規則の適用を「規則の正しい適用」もしくは「正しい適用」と呼ぶ。そして、それ以外の適用を「規則の誤った適用」もしくは「誤った適用」と呼ぶ。又、規則の正しい適用がなされている時、規則の下位節点のガバナを節点の順に左から並べたものを「正しい局面に出現したガバナの並び」、誤った適用によって出現したガバナの並びを「誤った局面に出現したガバナの並び」と呼ぶことにする。

例1の解析木が正しい解析木であるとする、例1、2から得られる「正しい局面で出現したガバナの並び」及び「誤った局面で出現したガバナの並び」は以下ようになる。[例3,4]

[例3 正しい局面に出現したガバナの並び]

・CL	→ NP VP	に対して	$\left. \begin{array}{l} I \text{ play} \\ I \\ \text{play well} \\ \text{play tennis} \\ \text{tennis} \end{array} \right\}$
・NP	→ PRON	に対して	
・VP	→ VP ADV	に対して	
・VP	→ VP6 NP	に対して	
・NP	→ NOUN	に対して	

[例4 誤った局面に出現したガバナの並び]

・VP	→ VP6 NP	に対して	$\left\{ \begin{array}{l} \text{play well} \\ \text{well} \\ \text{tennis well} \end{array} \right\}$
・NP	→ NOUN	に対して	
・NOUN	→ NOUN NOUN	に対して	

この分類基準を用いて、各書換え規則ごとにガバナの並びを蓄積していくと、正しい局面にも誤った局面にも出現するガバナの並びが生じる。そこで、我々は、ガバナの並びを

- (1) 「正しい局面にのみ出現する」(正の事例のみ)
- (2) 「誤った局面にのみ出現する」(負の事例のみ)
- (3) 「正しい局面にも誤った局面にも出現する」
(正にも負にもなる事例)

に分類し、蓄積に伴って分類を自動更新する機構を実装した。

3.2. ガバナの共起関係の蓄積実験

上記の分類法に従い、3,200文からガバナの並びを蓄積してみた。使用した文例は、英文法書「英語の型と用法」、辞書「Light House」の用例、英検4級の用例から抽出した。抽出基準は比較的易しく、かつ、状況によって文意に曖昧性が生じないような文例であるとした。

実験結果から、文例数の増加に伴って、正のみの事例及び負のみの事例が単調に増加するのが観察された。又、最終的に3,200文から、正のみの事例が約10,000種類、負のみの事例が約5,000種類、正にも負にもなる事例が約4,000種類収集できた。

しかし、収集したガバナの並びを観察した結果、各種の問題点[4章で後述]のため、無意味なデータやすぐに反例が見つかる負の事例が含まれることが分かった。

3.3. 負の事例に基づくフィルタリング

規則ごとに蓄積した負の事例を利用して、「解析時に出現したガバナの並びが収集済みの負の事例に含まれている時、その規則の適用を行わない」ことで、規則の過剰適用を抑止でき、曖昧な解析木の生成を抑えることができる。我々は、この負の事例を利用し

た規則の過剰適用の抑止を「負の事例に基づくフィルタリング」と呼んでいる。

負の事例によるフィルタリング機構をパーザに組み込んだ。そして、蓄積用文例の増加に伴う解析性能の変化を(1)生成された解析木の数(1文当たりの平均本数)、(2)正しい統語構造が生成できる確率、(3)正しい統語構造が選択できる確率、の3項目について測定した。

[実験1] 蓄積用文例の集合と解析用文例の集合が等しい時

先の事例収集に用いた3,200文をフィルタリングの効果測定に用いてみた。実際には自然言語システムの入力例をすべて蓄積用の文例にすることはできないので、実験の有用性については問題がある。しかし、先程の蓄積実験において、蓄積用文例の増加に伴い負の事例が十分に増加していることから、筆者らはフィルタリングの効果を予測するための基礎データとして利用できると考えている。

実験結果から以下のことが観察できた。[図1]

- (1) 蓄積文例の増加に伴って、ほぼ単調に平均解析木の数が減少し、最終的には2.5本から1.0本に減少した。図1のAが解析木の数の減少量を表している
- (2) 正しい統語構造を生成できる確率は、蓄積文例が少ない時、急落しており、そのから文例数の増加に伴って上昇をはじめた。最終的には蓄積開始前の値と等しくなった。
- (3) 正しい統語構造を選択できる確率は、蓄積用文例が少ない時、急落している。しかし、その後増加に転じ、蓄積が解析対象の3分の1程度を越えてからは、フィルタリングを行わない時の性能を上回ることができた。(図1のC参照)。

これは、解析木の生成を抑止することで、より少ない候補から解析木を選択することができたからである。

[実験2] 蓄積用文例の集合と解析用文例の集合が異なる時

実験1で事例収集に用いた文例を蓄積用の2,400文と解析用の800文に分けて、フィルタリングの効果を測定した。[図2]

- (1) 蓄積文例の増加に伴って、多少上下しながら平均解析木の数が減少し、最終的に、1文当たりの解析木数は3.5本から2.5本に減少した。(図2のA参照)
- (2) 正しい統語構造を生成する確率は実験1と同様、蓄積例が少ない時、急落した。その後上昇はするが、初期値までは回復しなかった。(図2のB参照)

正しい統語構造を生成する確率が初期値に達していないことから、この場合、負の事例によるフィルタリングは強すぎたといえる。

(3) 選択する確率も蓄積例が少ない時、悪化し、その後上昇が見られたが、実験1のようにフィルタリングを用いない時の性能を越えられなかった。(図2のC参照)

4. 簡易な手法の問題点

ここでは、簡易な手法における共起データ蓄積・利用に対する8つの問題点を列挙する。

問題点[1] 言語現象から見て、収集するのに不適切な並びを共起事例と扱ってしまう。

全ての書換え規則について一律にガバナの並びを収集しているため、共起関係の無いものまで収集してしまう。我々の場合では、TEXT → CL END という規則(節と終止符から文を生成する規則)等からガバナの並びを収集していた。このような並びを蓄積して利用するには無理があり、蓄積の対象から除かねばならない。

問題点[2] 蓄積例を増やした時、負のみの事例に反例が出現しやすい規則がある。

例4では NP → NOUN (名詞から名詞句を生成する規則)において ≪ well ≫ を負の事例として蓄積する。もし、蓄積例の中に正の事例がなければ、NP → NOUN の規則適用時に ≪ well ≫ がフィルタリングに用いられる。しかし、明らかに反例が存在する。この問題を避けるために、蓄積するデータによってはフィルタリングを用いないようにしなければならない。

問題点[3] 書換え規則によっては、共起関係に無関係なガバナがガバナの並びの中に現れる場合がある。

たとえば、CL → NP ADV VP という規則(名詞句と副詞句と動詞句から節を生成する規則)の場合、NPとVPのガバナの並びには主語と動詞句との共起関係があるが、主語・副詞句・動詞句の3つのガバナの並びでは共起関係は無い。単に全下位節点のガバナを共起関係として収集するのは問題である。

問題点[4] 異なる書換え規則において、同一種類の共起関係が存在しても、同一の共起関係としては蓄積されない。

たとえば、CL → NP VP (名詞句と動詞句から節を生成する規則)と、問題点[3]で例示した CL → NP ADV VP の両規則とも、NPとVPのガバナの

組が、主語と動詞句との共起関係にある。同一の関係を別の関係として蓄積・利用することになり、共起関係蓄積の効率が悪い。異なる規則に出現する同一の共起関係を扱う必要がある。

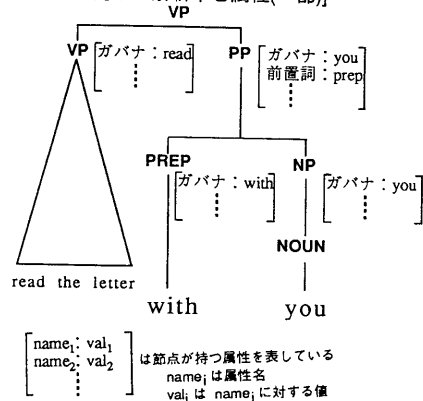
問題点[5] 下位節点のガバナの並びだけでは表現できない共起関係が存在する。

例5の場合、VP → VP PP (動詞句と前置詞句から動詞句を生成する規則)における下位節点のガバナの並びは ≪ read you ≫ である。そして、例7が正しい統語構造と一致して、VP → VP PP に対し、≪ read you ≫ が正の事例として蓄積される。しかし、文例

I read the letter from you.

では、前置詞句が直前の名詞句に係るべきであるが、VP → VP PP が誤って適用された場合、≪ read you ≫ が負の事例として蓄積される。動詞句と前置詞句の関係では、動詞、前置詞、前置詞句の目的語の3つが共起関係を成している。そこで、例5の VP → VP PP の場合、(1) VP のガバナの値、(2) PP の前置詞の値、(3) PP のガバナの値、の3つの属性の並びで共起関係を蓄積する必要がある。

【例5 I read the letter with you
に対する解析木と属性(一部)】



問題点[6] 正の事例を用いて統語解析木選択の成功率を上げていない。

負の事例を用いて解析木の生成を抑制することによって生成する解析木を減らし、選択の性能の向上計っているが、正の事例を用いることでも向上が計れる可能性がある。

ただし、正の事例を用いる場合、さらに以下の問題点[7][8]が生じる。これら2つに共通することは、言語現象としては「正のみの事例」に分類すべき事例が、3.1章の分類手法では「正にも負にもなる事例」に分類してしまう可能性があるからである。

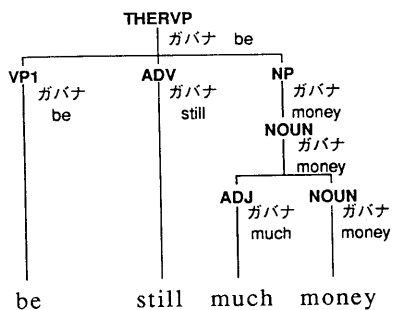
問題点[7] 不適当な部分木の上にさらに規則を適用して木を生成した時、そこで収集する負の事例のために不適当な分類が起こる。

例6では例文に対して、正しい統語構造を持つものを含んで2つの解析木が生成されている。この時、NOUN → ADJ NOUN の規則において、正しい統語解析木から、≪ much money ≫ を正の事例として収集する。そして誤った統語解析木から ≪ much money ≫ を負の事例として収集する。

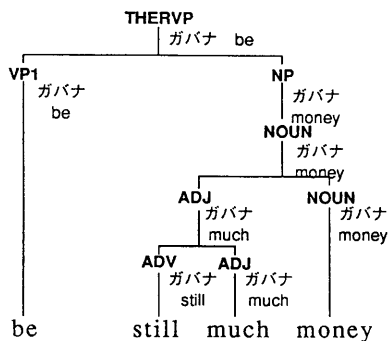
このため、≪ much money ≫ を「正にも負にもなる事例」に分類することになる。これは、誤った木において、still と much が結びついたことが原因である。この場合、still と much が ADJ → ADV ADJ において結び付かないことが既知となれば、この誤った解析木を生成すべきでない。そして、この木から事例を収集すべきでない。

【例6 There is still much money に対する統語解析木の例(一部)】

a) 正しい統語解析木



b) 誤った統語解析木



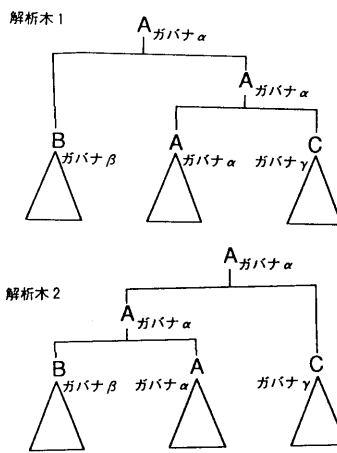
問題点[8] 書換え規則 $A \rightarrow B A$ と書換え規則 $A \rightarrow A C$ とで生じる曖昧性によって不適当な分類が起こる可能性がある。

たとえば、 $B A C$ の順に3つのカテゴリが並ぶ時、適用順序によって、例7の2つの解析木が生成で

きる。もし、例7のように、共起する属性が各節点に与えられている場合、以下の不都合が起こる。

3.1章の分類基準によると、「解析木1」が正しい場合にせよ、「解析木2」が正しい場合にせよ、いずれの場合でも、 $A \rightarrow B A$ に対して ≪ $\beta \alpha$ ≫ は「正にも負にもなる事例」になり、 $A \rightarrow A C$ に対して ≪ $\alpha \gamma$ ≫ が「正にも負にもなる事例」となってしまう。

【例7 2つのあいまいな解析木】



5. 蓄積・利用方法の修正

4章の問題点[1]~[7]に対応するために、簡易な蓄積・利用手法を修正した。しかし、問題点[8]については、共起関係を用いて曖昧な解析木を抑止することは困難である。そこで、我々はKATEパーザの禁止パターンによる解析木の生成禁止機能[2章]を利用することを検討中である。

以下、各修正点について述べる。

【修正1】 人手による共起関係の定義

共起関係自体をアルゴリズム的に自動的に決定するには大量の入力例が必要であることが予想される。そこで、我々は、人間が想定する有益そうな共起関係について、計算機を用いて個々の事例を蓄積し、この蓄積したデータを解析性能の向上に用いるべきと判断した。

この判断に基づき、蓄積・利用に用いる共起関係について、出現する共起関係やその共起関係の引数となる属性を人手によって書換え規則に記述する。そのため書換え規則は、

- (1) 下位節点の任意の属性を参照することで使用する共起関係を決定できる

(2)下位節点の任意の属性を共起関係の引数に指定できる

ような記述を可能にした。

この修正に対し、解析処理時に実際に出現する共起関係データは以下のようにして求める。

まず、適用された書換え規則の共起関係定義の記述に従い、使用する共起関係の名前とその引数の値が決定する。この共起関係の名前とその引数の値を実際に出現した共起関係データとする。

この共起関係データを「簡易な手法の時のガバナの並び」と同じように分類して蓄積し、利用を計る。

この修正によって問題点[1]~[5]が解決できることを以下、簡単に例で示す。

問題[1] TEXT → CL END 等では、共起関係を使用しない

問題点[2] NP → NOUN で共起関係を使用しなければ良い。

問題点[3] CL → NP ADV VP では、NPとVPのガバナを表す属性の2項関係で共起関係を定義する。

問題点[4] CL → NP ADV VP と CL → NP VP で同一の共起関係を定義する。

問題点[5] VP → VP PP で共起関係を定義する際、VPのガバナ、PPの前置詞、PPのガバナを共起関係の引数にする。

【修正2】 「正の事例のみ」の共起関係データを解析木選択のデータに利用

共起関係データの利用時に、出現した共起関係データが「正の事例のみ」として蓄積されていたら、書換え規則に与える木選択処理用の優先度のデータを最高値にする。出現した共起関係データが「負の事例のみ」として蓄積されていたら、規則の適用を禁止するように共起関係データの利用を定めた。

この修正によって問題点[6]にある程度対応できる。

【修正3】 共起関係データの蓄積を2パスにする

1パス目では通常の方法で蓄積を行う。1パス終了後、「正のみの事例」と「正にも負にもなる事例」をクリアする。そして蓄積した「負の事例のみ」をフィルタリングに用いて2パス目の蓄積を行う。

ADJ → ADV ADJ の ≪ still much ≫ が1パス終了時に、「負の事例のみ」に分類されていたら、2パス目ではフィルタリングによって例6の誤った解析木が生成されず、≪ much time ≫ が「正のみのデータ」となる。

6. 修正版による共起関係データの蓄積と利用

使用した解析用規則が簡易な版と修正版とで多少異なってしまった。しかし、変更点は僅かであり、解析能力にも差はないことから、以降の実験で、簡易な版と修正版の実験結果を比較する上で解析規則の差異は無視できると筆者らは信じている。

6.1. 修正版による共起関係データの蓄積実験

まず、簡易な手法の時と同様の3,200文から共起関係の事例を2パスで蓄積した[図3]。簡易な版と同様、「正のみの事例」「負のみの事例」は文例数の増加に伴ってほぼ単調に増加した。

簡易な版より正のみの事例が大量に蓄積できたのは、共起関係の種類を多く用意したからである。たとえば、単に主語と主格補語の関係を収集するだけでなく、同時に肯定文の時の主語と主格補語の関係をなどを収集したからである。

また、簡易な版で4,000以上あった正にも負にもなる事例が、修正版の1パス目で1,080になり、2パス目で970となった。1パス目で正にも負にもなる事例が大幅に減ったのは、人手による共起関係の定義を用いることにより無意味なデータやすぐに反例の見つかる負の事例が減ったからである。1パス目で正にも負にもなる事例の内、110種類が2パス目で正のみの事例として収集された。

6.2. 修正版による共起関係データの利用実験

簡易な版の実験1、2と同様に解析性能を(1)1文当たりの平均生成解析木数、(2)正しい統語構造が生成できる確率、(3)正しい統語木が選択できる確率、の3項目について測定した。

【実験3】 蓄積用文例の集合と解析用文例の集合が等しいとき

実験1と同じ蓄積用例、同じ解析対象であるが、実験結果[図4]は、実験1の結果[図1]とかなり異なった。

(1)実験1と同様、蓄積用文例数の増加に伴い、生成する解析木の数が減ってはいるが、最終的な減少量は修正版の方が小さい。(図1,4のA参照)

(2)文例が少ない時における正しい統語木生成の確率の悪化は実験1に比べて小さい。また、全体を通じての変化量も小さい。

(3)文例が少ない時にも、選択の確率がほとんど悪化しなかった。そのため、全体を通して選択率の向上が観察できる。しかし最終的な選択率の値は実験1を下回った。(図1,4のC参照)

[実験4] 蓄積用例の集合と解析用例の集合が異なる時

実験2と同じ蓄積用例、同じ解析対象を用いて修正版で実験した。[図5]

- (1) 文例数の増加に伴って平均解析木数は実験2同様減る傾向にあるが、実験2と比べて、より単調な減少が観察できる。しかし、実験2と比べて減少の幅は小さかった。(図2,5のA参照)
- (2) 実験2のような文例数が少ない時の急激な悪化はない。また全体的に変化量も小さい。最終的に正しい木を生成できる確率は実験2より良い。(図2,5のB参照)
- (3) 文例数の少ないとき僅かに選択が落ちているが、実験2のような急落はない。また、全体を通しての変化量は小さい。最終的には選択率は実験2より僅かに良い。(図2,5のC参照)

6.3. 蓄積済み共起関係データの出現率と解析性能

[実験5]

解析対象テキストを解析した時に出現する共起事例が蓄積済みの共起例に含まれる割合と解析性能との関係を知るため、以下の実験を行った。

まず、実験2,4で用いた解析対象を5つのテキストに分割した。分割前のテキストとあわせて6テキストについて、個別に共起関係データを蓄積した。そして、各々で蓄積した共起関係データが実験2,4の蓄積用例から蓄積できる共起関係データに含まれる割合を求めた。この割合を、蓄積用例に対する共起関係の出現率とする。蓄積用例は実験2,4と同じにし、6つのテキストを順次解析対象に替えて解析性能の変化を測定した。

各テキストの文例数が少ないため、信頼性についての検討は必要であるが以下の結果を得た。

平均生成解析木数は、共起関係の出現率が高い方が大きく減少する。[図6]

正しい統語構造を生成できる確率はすべてのテキストで減少した。しかし共起関係の出現率が高いほうが生成できる確率の減少度合いが小さい。[図7]

正しい解析木を選択する確率は出現率の高いほうで上昇し、出現率の低いほうで下がった。[図8]

7. まとめ

実験3,4,5の結果は共起のデータが不足する状況でも人手によって記述した共起は悪影響を及ぼしにくいことが観察できた。本稿で用いた例文が基本的で少

数あること割り引いても、共起関係データの処理において人間の持つ知識を利用することの重要性を示したと言える。今後は各種の例文に渡って、解析に有用な情報の蓄積とその利用の検討を続けたい。

謝辞 日頃御指導頂くKDD上福岡研究所小野所長、浦野次長並びに浅見主任研究員、橋本主査、知能処理研究室各位に感謝します。

参考文献:

- [1] 田中ほか: 自然言語の知識獲得、情報処理学会 自然言語処理研究会、88-NL-69-3、1988
- [2] 中川ほか: 自然言語の文法と意味解析規則の機能学習システム、情報処理学会論文誌、Vol.30, No.1, 1989
- [3] 中島ほか: テキストからの共起関係自動抽出の試み、情報処理学会38回全国大会、1989
- [4] H.Sakaki et al.、A Parsing method of Natural Language by Filtering Procedure, Transaction of the IECE of Japan, No.10, '86
- [5] 松本ほか: ガバナ間の共起関係学習のための基礎実験、情報処理学会38回全国大会、1989

図6 共起関係の出現率と1文当たりの生成される解析木の数の変化

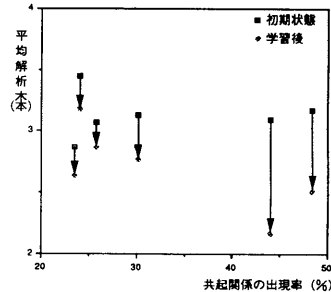


図7 共起関係の出現率と正しい解析木を生成できる確率の変化

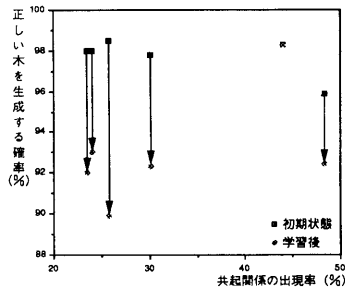


図8 共起関係の出現率と正しい解析木を選択できる確率の変化

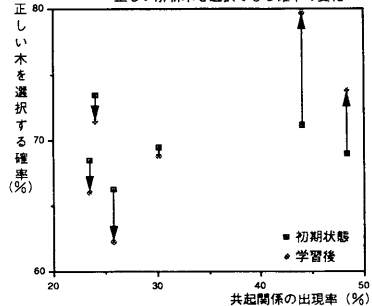


図1 負の事例に基づくフィルタリングの効果
(蓄積用文例=解析用文例)

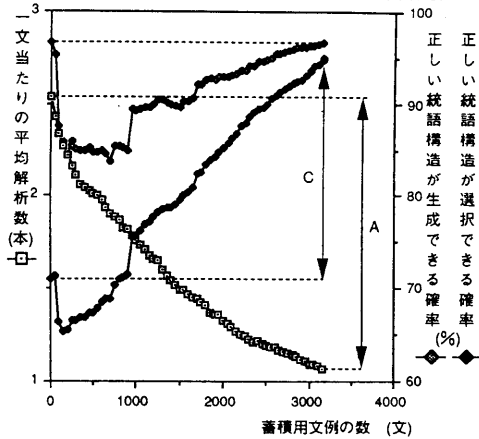


図4 修正版による効果
(蓄積用文例≠解析用文例)

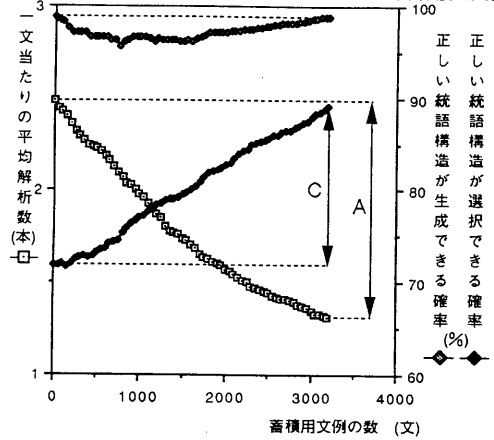


図2 負の事例に基づくフィルタリングの効果
(蓄積用文例≠解析用文例)

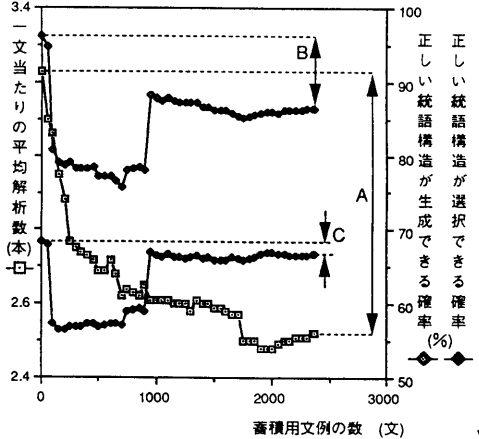


図5 修正版による効果
(蓄積用文例≠解析用文例)

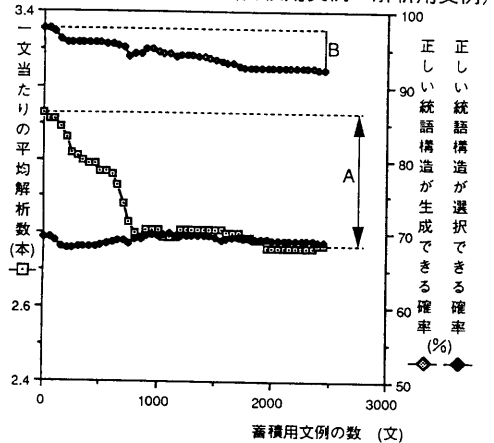
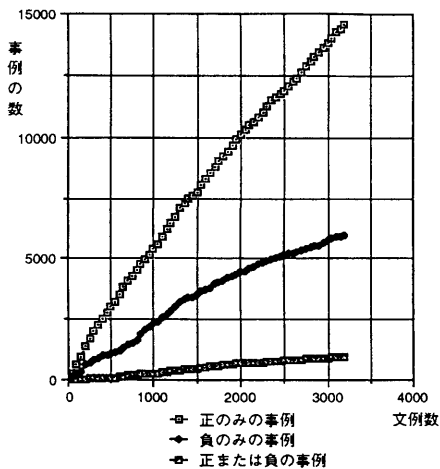


図3 蓄積できる事例の種類の変化



付1 統語カテゴリがVP23であるmakeの共起事例検索結果(部分)

VP23_SVOC[Jill,make,Jack,wife]

Correct examples: 1 ← 正の事例

SA-01:9270 ← 出典

Jill has made Jack an excellent wife.

Wrong examples: 0

VP23_SVOC[she,make,job,success] ← 出現した共起事例

Correct examples: 1

SA-04:9170

She's made the job a success.

Wrong examples: 0

VP23_SVOCは、主語のガバナと、ホーンビーコードでVP23AまたはVP23Bの他動詞と、直接目的語のガバナと、補語のガバナとの共起関係を表している。

VP23_SVOC[she,make,party,dress]

Correct examples: 0

Wrong examples: 1 ← 負の事例

SA-01:6670

She made a new party dress for her youngest daughter.

VP23_SVOC[you,make,such,mistake]

Correct examples: 0

Wrong examples: 1

SA-02:7320

It was silly of you to make such a mistake.