

3. 実験結果

3.1. 文例数の増加に対するガバナ列の数の変化

比較的易しく、かつ、状況によって曖昧性が生じない文例[表1]を50文ずつ増やしながらガバナ列の収集・分類を行った。その結果、文例数の増加に伴い正しい局面にのみ出現したガバナ列も、誤った局面のみのガバナ列もほぼ単調に増加することが判明した。[図2]

以前正しい局面でのみ用いられていたガバナ列でも、増加した文例の中に誤った例が出現したり、又その逆が生じる現象がある。この現象を50文ごとに計測してみた。その結果、新たに獲得されていくガバナ列の増加数に対し、このような現象の数は4%程度であった。つまり、この実験の範囲内では、さほど信念の翻意も無く、学習文例の増加に伴って、ガバナ間の共起に関する知識が獲得されていたことが判明した。

3.2. ガバナ列によるフィルタリング

誤った局面でのみ出現したガバナ列を規則ごとに前もって求めておき、このガバナ列が構文解析時にその規則で出現したら、規則の適用を行わないフィルタリング機能を持つパーザを用意した。そして、ガバナ列収集用文例数の増加と解析性能の変化を以下の3項目について測定した。(1) パーズに成功する入力文の数。(2) 正しい統語構造を生成する入力文の数。(3) 木選択部によって正しい統語木が選択された入力文の数。

まず、解析対象とガバナ列収集用文例を共に全文例として測定した[図3]。この時、最終的には統語的に曖昧な木は、ほぼ生成されることが確認できた。又、解析対象の約3分の1程度のガバナ列収集用文例で、正しい木を選択する確率が高まること判明した。次に、入力文とガバナ列収集用文を別にして同様の測定を行った[図4]。その結果、大抵の場合、フィルタリングを用いない時に比べて、生成・選択の確率が0.3%から3%ほど劣化した。元々、正しい木を生成する確率が平均97.5%と高かったのでフィルタリングが悪く働いたようであった。しかし、各実験を通じて生成された木の数は15%から20%減少した。

3.3. フィルタリングによるパーズ時間の変化

フィルタリングを用いた時と用いない時の、構文解析の時間と、生成された木の集合から最も確

からしい木を選択する時間を測定した[表2]。オーバヘッドにもかかわらずフィルタリングによって解析・選択の実行速度が向上することが判明した。

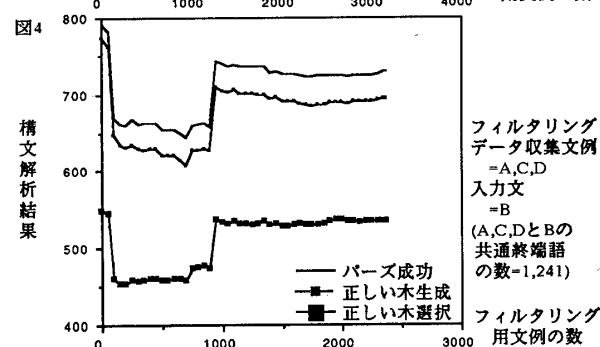
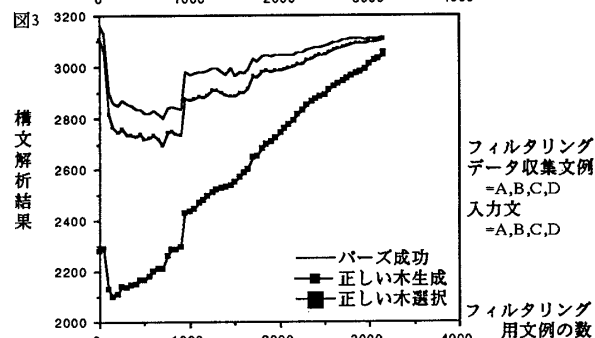
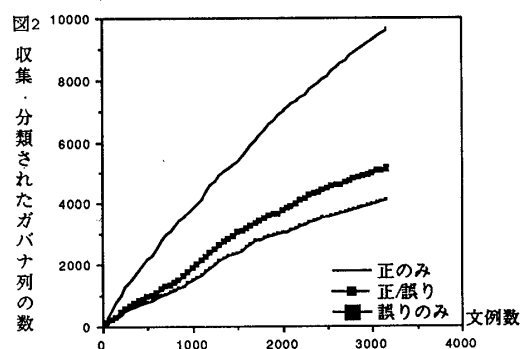


表1

文例	出典	文数	出現語数	終端語の種類
A	英語の型と語法	900	8,065	1,500
B	英語の型と語法	810	9,011	1,582
C	Light House 用例	834	7,328	1,125
D	英検4級用例	669	5,741	1,031

表2

	解析時間 (S)	選択時間 (S)
フィルタリング無し	3.56	0.297
ガバナ3,499例	2.623	0.181
ガバナ5,134例	2.617	0.177

謝辞 日頃御指導頂くKDD上福岡研究所小野所長、浦野次長、浅見担当主任研究員並びに知能処理研究室各位に感謝します。

参考文献 [1] 松本, 他: 辞書及び拡張CFGに対する知識調節機能の実現について, 情報学会36回(前期)全国大会, 7U-6