自然言語インターフェースにおける解析誤りの一回復手法

2F−5

大北 和弘　李 庸碩　青江 順一
徳島大学

1．はじめに

自然言語インターフェースは，システムの対象世界に関する辞書と文法構築者によって予め定義された文法規則を保有する．システムは，その辞書と文法規則に基づいて解析を進める．現段階では，定義された範囲内の表現に対しては有効なものとなっている．しかし，実際にはその範囲を越える入力が多々ある．そこで，これらの表現を全て網羅する文法規則を構築すべきであるが，これは構築に費やされる時間の莫大なことやメモリ容量の限界などから実現は難しい．よって実用的な自然言語インターフェースでは，不特定多数の利用者による様々な言語表現を既存の文法規則で受理する能力が必要になってくる．

本稿では，キーボードから日本語を入力し，計算機との対話を行う自然言語インターフェースを想定し，単一化文法の枠組み内で，検出された統語的誤りと意味的誤りを融合的に訂正していく手法を提案する．

松本による分類[1]のように，誤りの種類は各解析レベルに広く分布し，しかも各レベルにおいても多種多様であるが，本稿での統語的誤りとは，既存の文法規則に合致しない構文構造をもった入力によって生じる誤りを指し，意味的誤りとは，キーボード入力で生じたタイプミス（変換ミス）による同音異義語の誤り等のように語彙レベルで回復可能な誤りとしておく．

2．誤り回復の概要

誤り回復の大略は，構文レベルでは，既存の文法規則に照らして，削除，または挿入のための文法要素のみを決定し，解析を進める．なお，挿入候補語の決定が保留されていることに注意されたい．また意味レベルの誤りは，文全体を眺めて誤りを訂正しようとするものである．つまり誤りが検出されると，文末まで解析を続行し，誤り位置以後の解析で得られた情報で誤りを訂正する．このように，文全体を考慮した意味情報は，構文レベルの回復において保留された文法要素の訂正候補語の決定にも利用される．

2．1　誤り回復の具体例

例として入力文「太郎だけが長い神を接着剤で壁に張った」の訂正を考えてみる．

ここで既存の文法規則では名詞句の後には1つの助詞しか続かないような規則しか用意されてないと仮定すると，図1の(1)の位置で誤りを検出する．構文解析器は，解析を停止せずに後述する統語的な誤り回復によって，限定を表す助詞"だけ"を削除し，"太郎が"に訂正し解析は進む．次に形容詞"長い"は con または mat のいずれかの素性を持った名詞としか結合しないとすれば，誤り位置(2)が検出される．ここで辞書内から置換可能な候補語を探索し，解析を進める．これら候補語からの訂正語の決定は，一時保留し，後の解析で得られた情報を基に決定する．動詞"張った"の選択制限を名詞"接着剤"および"壁"は満足しない．よって解析器は(3)の誤り位置を検出する．そこで意味誤り訂正部に処理が移り，置換可能な候補語を辞書内から検索する．次に意味的制約条件を満たす動詞"貼った"が選ばれ，更に"貼った"の制約条件から先ほど保留されていた候補が"紙"に同定される．結果として「太郎が長い紙を接着剤で壁に貼った」に訂正される．

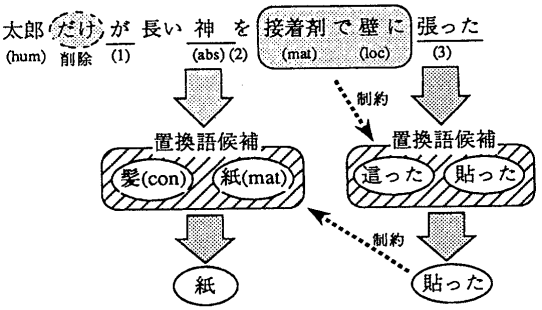

図1　誤り訂正の概念図

表層格	格マーカ(格助詞)	要求する意味素性		
		張る	這う	貼る
Suj	が	hum	hum	hum
Obj	を	mat	loc	mat
Obj2	に	—	—	loc
Obj3	で	—	—	mat

注) Suj は主語を表し,Objは直接目的語,
Obj2は間接目的語,Obj3はその他の目的語とする

図2　表層格に対応する動詞の選択制限

2．2　統語的誤り回復

統語的な誤り回復の手法としてM.D. Mickunas[2]らによる回復手法を用いる．この手法は，基本的に，解析表を参照しながら，誤り位置に対して何らかの終端語の挿入を行うものである．

2．2．1　誤り回復の概要

誤り回復ルーチンは，現在の構文解析器のスタックの先頭の状態からシフト動作も還元動作も起こさせないような入力記号に対して起動される．誤り回復ルーチンには，誤り発生時のLR構文解析器の配置(configuration)のコピーが渡され，結果として成功したか否か成功した場合には修正された配置が返される．

誤り回復処理は，圧縮フェーズ(condensation phase)及び訂正フェーズ(correction phase)の二つのフェーズから成る．

圧縮フェーズでは，誤り発生前の結果は無視して，誤り記号を最初の入力として改めて解析を続ける．つまり誤り記号をシフト可能な入力とする状態集合 P の要素 p に対して

① 再び誤りが発生する
② 誤り位置を越えるような還元が生じる（図3参照）

まで解析を続行させる．

解析の結果，①になる状態 p を保留候補(holding candidate)，②になる状態 p を訂正候補(correction c

---

A Method of Recovering Syntactic and Semantic Errors for Natural Language Interfaces
Kazuhiro OOKITA, LEE Yong-seok, Jun-ichi AOE
Tokushima University

andidate)と呼ぶ．各訂正候補に対して訂正フェーズは適用される．

訂正フェーズは，誤り発生時のスタックと誤り記号以後を解析した状況を示すスタックとを結合させる終端語を挿入する．挿入できなかったときは誤り位置を1つ後退させ，再度，挿入を試みる．誤り位置を後退させても挿入が失敗した場合，誤り位置の次にくる入力記号を削除する．過度の削除，あるいは誤り位置が初期状態まで後退した場合，訂正フェーズは失敗したとし，別の訂正候補を適用し，回復を試みる．

保留候補については，2番目の誤り位置に対して再帰的に誤り回復が試みられる．

誤り回復の結果，修正された配置が複数存在した場合，つまり修正後の文に曖昧性があった場合は，それらのうちコストの低い配置を選ぶ．コストは削除，挿入の回数を基に計算される．

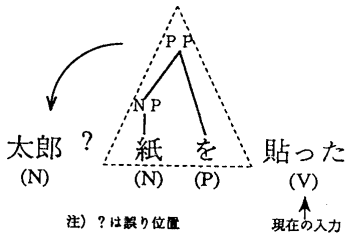

図3　誤り位置を越える還元

2．2．2　誤り訂正の具体例

例として入力文"太郎 紙 を 貼った"の回復を考えてみる．

構文解析器は，解析を進めていくうち"紙"の位置で誤りを検出する．回復処理部が実行され，処理は圧縮フェーズに移る．ここで図3に示すような誤り位置を越える還元が生じ，訂正フェーズにより，誤り位置にP(助詞)が挿入される．ここでは，挿入できる語の候補を保留しておき，後の解析で得られた共起情報を基に絞り込む．この例では"が"が挿入される．

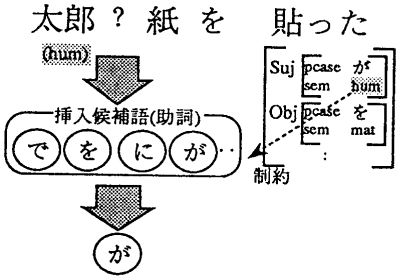

図4　統語誤りの回復図

2．3　意味的誤り訂正

文の各構成要素の属性構造の合成は，統語的な解析によって還元が成功した後に行われる．つまり構文木の節点において，構成要素の持つ属性構造の単一化が行われ両構成要素の属性の融合が行われる．ここで語句同士の整合性がチェックされ，単一化の失敗によって意味的な非文は排除されることになる．

従来の単一化文法による日本語解析では，名詞に続行する助詞(格助詞)によって名詞句の文法機能を定めていた．単一化文法で，意味的なチェックを行うには，辞書内の名詞に意味素性を付与し，動詞に対しても図2のように，文法機能を決定する条件として，名詞の意味素性を記述できるように拡張することが必要となってくる．つまり表層格の決定を助詞(格助詞)のみによって定めるのでなく，助詞(格助詞)と共起し得る名詞の意味素性の対により行う．これらの情報は動詞の語彙項目に予め記述されている．

図1の入力文「太郎だけが長い神を接着剤で壁に張った」の場合，(2)で最初の意味誤りが検出される．意味誤り訂正処理部は，誤り位置の語と同音の語を候補語として辞書より抽出する．例では素性conを持った"髪"，matを持った"紙"が選ばれる．動詞の場合も同様に同音の語句が候補として選ばれる．候補の絞り込みは名詞の場合は，動詞の属性構造に記述された共起情報を基に，また動詞は名詞に付与された意味素性と表層格マーカによって行われる．

候補語の曖昧性を保持させるには，属性構造の選言(disjunction)を取り入れる必要がある．選言によって候補語の決定を遅延させる．例えば図5の属性構造は，"長い紙"か"長い髪"のいずれかを表す属性構造である．

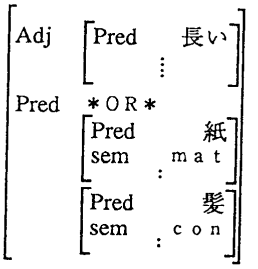

図5　属性構造の選言

3．まとめ

キーボード入力による自然言語インターフェースを例に挙げて，誤り回復の手法について述べてきた．

本稿で提案した手法の特徴としては，次の点が挙げられる．

(1)誤りが発生したら解析を停止するのではなく，とりあえず解析を続行し，文全体の解析結果から誤りを訂正する．

(2)誤り語を置換する候補語を列挙し，その絞り込みを文末を解析するまで遅延させる．
　候補語の選定はキーボード入力ならば同音語，OCRならば字形が類似しているものを辞書内から探索する．

(3)構文レベルでの逸脱に対しても語句を挿入，削除するなどして既存の文法規則に適合させて解析を行う．

4．今後の課題

統語的な誤り回復において，削除の対象となる語句が名詞や動詞のように重要な語である場合の処理，挿入語句の選定方法が問題となってくる．解決法として，利用者に問い合わせる等が考えられるが決定的な方法の発見は今後の課題といえよう．

また今回のように単一化文法の属性構造と意味素性によって訂正を行うのでなく，文の意味を重視した深層格の情報を誤り訂正に用いれば，より正確な訂正が可能になると思われる．そのためには，得られた属性構造を深層格フレームに変換，あるいは文献[3]のように直接生成することが必要となってくる．

［参考文献］

[1]松本：頑健な自然言語へのアプローチ，情報処理，Vol.33, No.7, pp.757-767, 1992

[2]M.D.Mickunas and J.A.Modry:Automatic Errorr Recovery for LR Parsers, Comm.ACM, Vol.21, No.6, 1978

[3]二口ほか：LFGに基づく並列型パージング法，情報処理学会自然言語処理研究会，72-6, 1989