自然言語処理 153－4
（2003．1．20）

意味解析システムSAGEのWeb化と連体・使役・受身における解析精度向上

井村裕 † 沓掛俊樹 †
佐藤直美 † 原田実 ††

我々はこれまで，EDR電子化辞書を元に，日本語文を意味解析し格フレーム群を生成するシステムSAGEを開発して，語意精度80.1%，深層格精度80.6%を達成している．

本研究では連体・使役・受身表現に対する解析手法を新たに開発し解析精度を向上する．連体表現においては，用言から体言への修飾方向の逆向きの補足関係を表す係り受け関係を仮定し，これを元に辞書検索し語意と格の確率計算を行うことによって解析精度を向上した．また，動詞に接辞'(r)areru' '(s)aseru' が付くと表層格が変化する受動表現と使役表現については，適切な表層格を仮定し，これを能動態表現に変換してから，辞書検索することによって解析精度を向上した．この結果，語意精度89.2%，深層格精度89.0%を達成した．また，SAGE Web Serverを開発し，インターネット上で利用できるようにした．

The accuracy improvement of Japanese semantic analysis system SAGE for the adnominal, causative, and passive expression.

YUTAKA IMURA ,† TOSHIKI KUTSUKAKE ,† NAOMI SATOU † and MINORU HARADA ††

We have developed the Japanese semantic analysis system SAGE which decides the meaning of a word and the deep case between the words by using the EDR electronic dictionary and generates the result as the case frames. This system achieved 80.1% for word meaning accuracy and 80.6% for deep case accuracy. Here, we will improve both analytical accuracy by developing the new analytical technique for adnominal, causative and passive expressions. In an adnominal expression, analytical accuracy has been improved by assuming the dependency relation which shows the supplementation relation of the opposite direction in the direction of the modification from the declinable word to the indeclinable word, and by retrieving the dictionary based on this relation and calculating the probability of the meaning of the words and the case. Moreover, for the causative and passive expressions where surface case changes when affix '(r) areru' and '(s) aseru' adhere to the verb, analytical accuracy has been improved by converting these expressions into the active voice expressions by using an appropriate surface case and retrieving the dictionary based on those expressions. As a result, word meaning accuracy was improved to 89.2% and deep case accuracy to 89.0%. Moreover, SAGE Web Server was developed, and it is now available on the Internet.

1. はじめに

原田研究室ではこれまで，EDR電子化辞書に記載された情報を元に，日本語文を意味解析し格フレーム群に変換するシステムSAGE[2)4)9)10)]を開発し，オブジェクト指向分析システムCAMEO[1)11)12)]の自然語要求仕様の意味解析や内容の類似による判例検索システムJCare[3)]などに応用してきた．しかし，実利用す

† 青山学院大学理工学部経営工学科
Department of Industrial and System Engineering, Aoyama Gakuin University
†† 青山学院大学理工学部情報テクノロジー学科
Department of Integrated Information Technology, Aoyama Gakuin University

るにはさらなる精度向上が必要なので，SAGE の評価実験を繰り返し行い，不正解の原因を次の４つに分類することができた．①連体節における補足関係を示す深層格の正解率が低い，②使役・受身を含む文章の正解率が低い，③昨年の研究であるシステム辞書からオンメモリ化による辞書構築の移行が不完全である，④その他ソースレベルにおけるシステムの不具合などがある．そこで，これらの原因に焦点を絞りより精度の高い意味解析を行うシステム SAGE2002 の開発を行った．具体的には以下の３つを行う．

1) 連体表現における解析精度向上:用言が体言を修飾する連体節において，従来の SAGE では語意と補足関係を表す深層格の精度が低かった．これに対し，連体表現を，被修飾名詞が述語に対して修飾方向の逆向きの補足語としての関係をもつ表現に変換し，語意-格総合評価値を求める手法を開発した．

2) 使役表現・受動表現における解析精度向上:従来の Sage では，接辞'(r)areru' や'(s)aseru' が付くと表層格が変わるが深層格はかわらない使役表現や受動表現に対応していなかった．そこで，使役表現・受動表現を含む文は適切な表層格を補って能動表現に変換して，辞書検索を行うことで解析精度を向上した．

3) SAGE の Web 化:現在 Windows 上のみで動作している SAGE を Unix 上でも動作するようにし，インターネット上で公開利用できるよう Web Server 版を開発した．

なお，従来の SAGE と本研究で開発した SAGE2002 における意味解析の処理の流れに大きな差はない．その差は，後で述べるように，連体節と使役・受身における解析精度向上の仕掛けを組み込んだことである．

2. SAGE での意味解析の概要

SAGE における意味解析とは，係り受け関係にあるすべての２文節の語意とその主辞同士の深層格を決定し，語毎の格フレームからなる意味ネットワークを生成することである．ここでいう主辞とは，文節の主要な語を指す．ただし図１に示すように，複合語，活用語（例「発生する」），派生語の場合は，主辞を明確に判別するために，主辞自身（発生）には head 格を付与している．また，文の主要な概念として文中における最後の文節の主辞には main 格を付与する．これは，日本語では一般に最後の文節がその文の主要な主張を表していることが多いためである．格の向きは，係り先→係り元とする（例：「人は並ぶ」並ぶ→ object 格→人）．

SAGE で係り受け関係にある意味解析を行う前段階

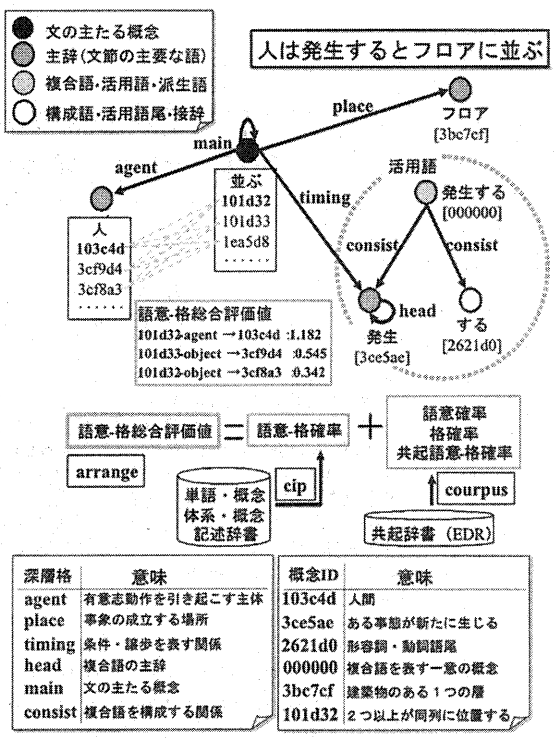


図１ SAGE による意味解析
Fig. 1 Semantic analysis by SAGE.

として形態素解析と係り受け解析を行う．本研究では，構文解析システムとして『茶筌・南瓜』[8] 引用文献と『JUMAN・KNP』[5)6)] を利用した．また preSAGE では，南瓜や KNP の出力ファイルを，prolog で扱いやすい形のリスト形式 (tree 述語形式) に変換する．この tree 述語を受けて SAGE は意味解析を行う．SAGE で行われる処理を大きく分けると，SAGE 本体，EDR 検索ルーチン，語意・深層格決定ルーチンとなる．図１示すように，この処理の流れを「人は並ぶ」という例に添って説明する．まず SAGE 本体が tree 述語形式ファイルを読み込み，そこから係り受け関係にある２文節（「並ぶ」と「人」）を取り出す．これら２文節を EDR 検索ルーチンに引き渡す．EDR 検索ルーチンでは，語意-格確率，格確率，語意確率，共起語意-格確率を求める．語意-格確率とは，与えられた２文節の主辞（「並」と「人」）の語意とその間の深層格の３つ組の尤もらしさを与える値で，概念体系辞書と概念記述辞書で検索し統計をとることで得られる．格確率，語意確率，共起語意-格確率は共起辞書を用いて検索し統計をとることで得られる．格確率とは，助詞（は）と係り先主辞（並）から，その助詞と単語が共

に出現した場合の2文節間における格の出現確率をいう．語意確率とは，係り先主辞と係り元主辞（「並」と「人」）から，この2つの語が共に出現した場合の，2語の語意組の出現確率をいう．共起語意-格確率とは，係り先主辞（並），助詞（は），係り元主辞（人）から，その助詞と2単語がともに出現した場合の，語意-格組（3つ組）の出現確率をいう．これらをすべての係り受け関係にある2文節に対して求め，個々の語意-格組毎に語意-格確率，格確率，語意確率，共起語意-格確率の和を語意-格総合評価値として算出する．このようにして求めた確率を基に，語意-格決定ルーチンで，文毎の係り受け木に語意と格を割り当てて，係り受け木全体で最も評価値が高くなる割り当てを持つ木（解釈木）を求める．多数の語意や格の候補から解釈木の準最適解を求める過程は，文節数に線形オーダーのJiri-Haradaアルゴリズムで高速化をする．この結果をSAGE本体に引き渡し，SAGEは，複合語における主辞以外の構成語の語意を個々に求めながら，解釈木を作成し，格フレームファイルを出力する．

3. 連体節解析の精度向上

3.1 旧SAGEの解析誤りの原因

連体節とは，名詞を修飾する働きをするものであり，被修飾名詞に対する修飾の仕方の違いにより，補足語修飾節・相対名詞修飾節・内容節に分かれる．今回問題とされるのは補足語修飾節である．

補足語修飾節（例「私が借りた本」）は，被修飾名詞「本」が連体節中の述語「借りた」に対する補足語の関係にあるものをいう．まず，修飾関係を表すwhich格を「本→ which →借りた」の向きに割り当てる．次にこの補足関係を表す格を「借りた→ object →本」の向きに割り当てる．しかし，これまでのシステムによる解析では，この補足関係を表す格「object」が誤っていることが多くあった．

これまでのシステムでは，ＥＤＲ電子化辞書では連体節に対する事例が非常に少ないので連体節を構成する2単語については全ての語意-格組の総合評価値は同値としていた．そして，Jiri-Haradaアルゴリズムによってこれらの語意-格組の総合評価値から最適解釈木「本→ which →借りた」を求めた後に，その補足関係を表す格「借りた→ object →本」を共起辞書のみから求めていた．従って，連体修飾関係にある2単語についてはそれらの語意は単に辞書に該当する語意が無作為に選ばれているにすぎなかった．

3.2 連体節の新ルールに関する流れ

被修飾名詞「本」は述語「借りた」に対して補足語の関係（「本を借りた」という関係）にある．この格関係「借りた→ object →本」は，十分にＥＤＲ電子化辞書に存在する情報であると考えられる．

そこで「本を借りた」（借りた→ object →本）の語意と格関係を調べた方が，「借りた本」よりははるかに尤もらしい深層格をみつけることができる．

まず，係り受け解析から得た2文節から連体節の関係（名詞を修飾する関係）があるかどうかを判定する．そのような関係が得られなければ通常通りの計算を行う．連体節の関係があると判定されたらこの2文節に対して連体節における総合評価値計算システムadnominalを用いて，係り受け関係を逆向きに仮定し（「借りた本」→「本を借りた」）総合評価値を得てから，Jiri-Haradaアルゴリズムによって最適解釈木を得る．

3.3 連体節の語意-格総合評価値の算出

①連体ＣＩＰ：語意-格確率の計算

通常「借りた本」であれば，係り元は「借りた」，係り先は「本」であるが，これを「本を借りた」というように係り受け関係を逆向きに仮定して辞書引きを行う．まず，2単語「本」と「借りた」を単語辞書検索し，それぞれの概念ＩＤを得る．これらの概念ＩＤを用いて概念体系辞書検索し，それぞれの上位概念を最上位まで求める．そして係り元を「本」の概念ＩＤを含む全ての上位概念，係り先を「借りた」の概念ＩＤを含む全ての上位概念として，概念記述検索し，語意-格組候補を得る．

次に，それぞれの語意-格組候補における語意が，概念構造の末端からどれだけ離れているかを調べる．これを概念距離と呼ぶ．各語意-格組候補における2つの語意の概念距離の和を求め，さらにその逆数を求める．これを優先率と呼ぶ．すべての語意-格組候補について優先率を求め，すべての優先率の中での各候補の割合を語意-格確率とする．

② 連体 Courpus：共起語意-格確率の計算
（A:共起辞書に連体節での情報が存在する場合）

ＥＤＲ電子化辞書にはごくわずかではあるが連体節としての情報が存在する場合がある．例えば「読んだ本」という連体節の関係をもつ文において通常通り係り元を「読んだ」，係り先を「本」として検索すると「1e867e(文章を読む) ― object → 3ce30(書籍)」という情報を得ることができる．このように連体節としての情報がある場合は通常通りの共起語意-格確率の計算を行う．

まず，係り元「読んだ」，共起関係子「@rentai」（連体情報がある場合ＥＤＲ電子化辞書には全て共起関係

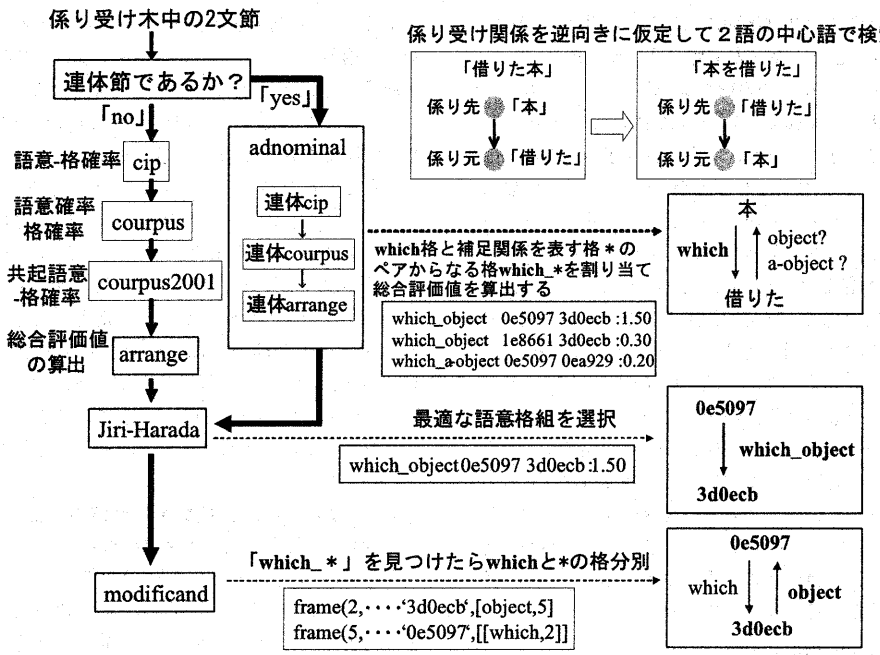


図 2 連体節解析の処理の流れ

Fig. 2 Process flow of adnominal clause analysis.

子が「@rentai」として登録されている)，係り先「本」の３つ組（助詞付き２単語）に注目する．この助詞付き２単語が出現した場合に係り元文節と係り先文節がどのような語意-格組をとることが多いのかを共起辞書を検索することで調査する．そして，語意-格組毎にその出現確率を求める．共起語意-格確率の算出には，表層共起頻度，項目共起頻度を用いる．表層共起頻度は助詞付き２単語すべての語意-格組の総出現数であり，項目共起頻度は，ある語意-格組が日本語コーパス内で登場している回数で，その語意-格組がどの程度実際に使われているかを示す指標となる．つまり，該当する共起事例の項目共起頻度を表層共起頻度で割ることで，その語意-格組の妥当性を示すことができる．

(B:共起辞書に連体節での情報が存在しない場合)

「読んだ本」のように共起辞書に連体節としての情報が存在しない場合,「本を読んだ」というように係り受け関係を逆向きに仮定して考えて，係り元を「本」，係り先を「読んだ」として２単語で共起辞書検索を行う．この場合共起関係子を「を」として検索することも考えられるが，本来の文「読んだ本」からこの共起関係子を特定することは困難であると考え，あえて２単語のみでの検索にした．

これによって共起事例を得るが，ここからの共起語意-格確率の計算方法は上で述べた方法と全く同じである．

③連体 Arrange：語意-格総合評価値の算出

連体 Arrange では，連体 CIP で求められた語意-格確率，連体 Corpus で求めた共起語意-格確率からそれぞれの語意-格組候補の尤もらしさを表す語意-格総合評価値を求める．

この語意-格総合評価値はそれぞれの同じ語意と深層格の組み合わせの語意-格確率と共起語意-格確率を足し合わせることで求める．

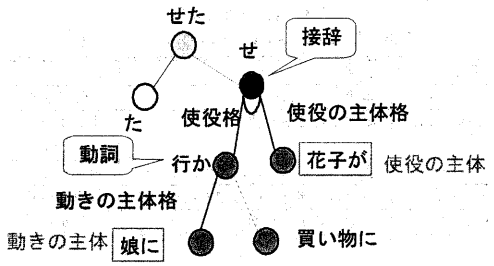
ここでの深層格の表現は連体節である「which」とその補足関係を示す深層格を全て表現するため「which_object」のように表し，

0e5097 - which_object → 3d0ecb :1.50

という形式で Jiri-Harada アルゴリズムに渡す．

3.4 modificand：補足関係の格の割り当て

Jiri-Harada アルゴリズムによって最適解釈木が生成された後の当該２単語間の語意-格組は 0e5097 (書籍) - which_object → 3d0ecb (返す約束で他人のものを使う) という情報を持っているに過ぎない．システム modificand はこの情報を得て，連体関係を表す「0e5097 - which → 3d0ecb」と当該２単語間の補足関係を表す「3d0ecb - object → 0e5097」という二つの



「花子が娘に買い物に行かせた」

図 3 使役の係り木

Fig. 3 The dependency tree of the causative.

格フレームを生成する.

4. 使役表現における解析精度向上

従来の Sage では使役文において，語の区切り・深層格・係り先の解析精度は低かった．これは接辞'(s)aseru' を付けると深層格は変わらないが表層格は変わる使役表現に対する事例が EDR 辞書に少ないからである．そこで，以下のような解決方法を採った．

4.1 使役の係り受け解析と格表現

使役表現は接辞'(s)aseru' がついた文を指すが，語尾が「〜てもらう」の文もその語尾が付くことによって深層格は変わらないが表層格は変わるので，共に使役表現とみなすことにする [7]．

使役文「花子が娘に買い物に行かせた」において従来の Sage では 2 つの主体「花子」と「娘」が 1 つの動詞「行かせる」に係っていた．また,「花子」の深層格が agent 格,「娘」の深層格も agent 格となってしまい，主体の区別がなされていなかった．そこで図 3 のように使役文の動詞 1 文節を動詞と接辞に分けて 2 文節にし，使役の主体 (使役を行わせる側) を接辞に，動きの主体 (使役によって，特定の動作を行う側) を動詞に係るようにした．また，動きの主体と動詞間の格を「動きの主体格」，使役の主体と使役の接辞間の格を「使役の主体格」，動詞と使役の接辞との格を「使役格」とし，それぞれの深層格を決定するようにした．

4.2 使役格の決定

使役格は，動きの主体と使役の主体の係り先を分ける為に，使役動詞を動詞と接辞'(s)aseru' に分ける際に発生した格である．従来使用している EDR の深層格の中に当てはまる格が存在しなかったので，使役格として causative 格を作った．

4.3 使役の主体格の決定

「使役の主体格」も「使役格」と同様に，独自に係り先を決めた際に発生した格なので，従来使用している共起辞書で検索しても事例が存在しない．そこで，係り先単語と係り元単語を元に独自に深層格を割り当てることにした．具体的には語尾が「〜てもらう」の文では相手の意志を尊重する表現だが，接辞'(s)aseru' の文では人やものがある事態を何らかの形で引き起こす表現なので [7]，使役の主体格は動詞の語尾が接辞'(s)aseru' か「〜てもらう」かで候補となる深層格が異なると考えた．そこで，語尾が「〜てもらう」の文での使役の主体格は利益・不利益の移動先である beneficiary 格とした．また，語尾が接辞'(s)aseru' の文での使役の主体格は，意志を持つもの，使役の主体の上位概念が「生物」なら agent 格とし，意志を持たないもの，使役の主体の上位概念が「生物」でないなら cause 格とした．

4.4 動きの主体格の決定

動きの主体格を決定するにあたって，まず，動きの主体を決定しなければならない．動きの主体は表層格が「二格」もしくは「ヲ格」であることに着目し,「B に（を）C させる」が与えられた時，以下のようにして求める．

(1) 動きの主体を見つける

まず，表層格がニ格，ヲ格を含む文節「B に (を)」を見つけ，B を使役の主体の候補とする．次に B の上位概念を調べ，上位概念に「人間」(30f6b0) があれば，動きの主体とみなす.「人間」がない場合でも,「位置」(30f751),「時」(30f776),「物事」(3998076) がなければ，また動きの主体とする．

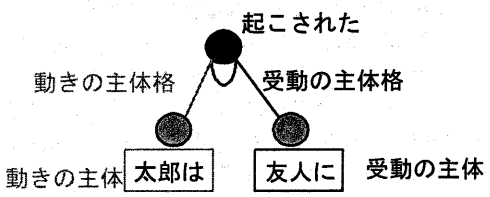
(2) 動きの主体 B の語意と深層格を決定する

B の表層格ニ格（ヲ格）をガ格に変更する．つまり受動表現「B に（を）C させる」に対し能動表現「B が C する」を仮定して，この係り受け及び表層格を用いて辞書を引き語意と深層格の確率を求める．

ここで，(1) で動きの主体の候補である B の上位概念に「時」や「位置」が含まれているなら B を動きの主体でないとする理由は,「8 時に起きる」,「9 時から働く」,「京都から来る」,「東京に行く」のように，時間や場所を表している語は動きの主体にはならないからである．また，B の上位概念に「物事」が含まれているなら B を動きの主体でないとする理由は,「花子が娘に買い物に行かせた」の「買い物」のように上位概念に「物事」が含まれている語は動きの主体にならないからである．

4.5 「文内の主たる概念 (main 格)」の決定

従来の Sage では，文内の主たる概念 (main 格) は，文末の位置で文を支えている述語とし，主たる概念が



「太郎は友人に起こされた」

図 4 受動の係り木
Fig. 4 A passive charge tree.

複合語の場合は，複合語の主辞としていた．今回，使役動詞1文節を動詞と接辞の2文節に分けたことにより，主辞も2つになり，文内の主たる概念はこれまで動詞であったが，使役文では使役の接辞'(s)areru' の主辞を文内の主たる概念とした．

5. 受動表現における精度向上

従来の Sage では受動表現においても深層格の精度が低かった．能動表現「B が (は)A に (を・から)C する」の動詞 C に接辞'(r)areru' を付けると受動表現「A が (は)B に (から・によって)C られる」となり，「B が (は)」は「B に (から・によって)」と表層格がガ格又はハ格からニ格又はカラ格又はニヨッテ格になる．同様に「A に (を・から)」は「A が (は)」となり表層格が二格又はヲ格又はカラ格からガ格又はハ格になる．

Sage では係り元文節と係り先文節の2文節ごとに語意と深層格の確率を求めているので，「A が (は)B に (から・によって)C られる」は「A が (は)C られる」と「B に (から・によって)C られる」に分けてそれぞれ，語意と深層格の確率を求めている．動詞の辞書引きは不変化部分で引かなければならないので「A が (は)C られる」は「A が (は)C」という形式に，「B に (から・によって)C られる」は「B に (から・によって)C」という見出しで辞書を引いており，動詞の不変化部分で辞書を引く段階では原文が受動表現であることが効いておらず，深層格として間違ってたものが出てきていた．また受動表現の事例は辞書に少ないので語意も誤ることが多かった．従って，辞書引きする際，能動表現「A に (を・から)C する」や「B が C する」を仮定して，「A に (を・から)C」や「B が (は)C」という見出しで辞書を引けば，深層格の精度が向上すると考えた．

具体的には動詞に接辞'(r)areru' があれば受動文と判定し，以下のような解決策を採った．

5.1 受動文における格表現

受動表現「A が (は)B に (から・によって)C られる」では，使役と同様主体が2つあるので，A を動きの主体 (受動によって特定の動作を行う側)，B を受動の主体 (受動を引き起こす側) とした．図 4 のように動きの主体と動詞間の格を「動きの主体格」，受動の主体と動詞間の格関係を「受動の主体格」とし，それぞれの深層格を以下のようにして決定する．

表 1 深層格とその意味
Table 1 Depths case and that meaning.

EDR 格	意味	例
time	事象の起こる時間	8 時に起きる
material	素材または構成要素	牛乳から バターを作る
source	事象の主体または対象の最初の位	京都から来る
goal	事象の主体または対象の最後の位	東京に行く
time-from	事象の始まる時間	9 時から働く

5.2 動きの主体格決定

受動表現「A は C られる」は能動表現にすると「A に C する」または「A から C する」または「A を C する」の3通りあり，例えば「太郎は愛されている」を能動表現にすると「太郎を愛す」であり，「鈴木さんは書類を渡された」は「鈴木さんに書類を渡した」となり，「花子はバックを奪われた」は「花子からバックを奪った」となる．このように，動きの主体の表層格は受動表現では「ハ格」又は「ガ格」であるが，能動表現では「ニ格」，「カラ格」，「ヲ格」のいづれかである．この能動表現の3通りの可能性に対する語意格組の尤もらしらを Jiri-Harada で一意に絞り込む．具体的には，以下のようにした．

(1) 動きの主体を見つける

表層格がガ格（ハ格）を含む文節「A が（は)」を見つけ A を動きの主体とする

(2) 動きの主体の語意と深層格を決定する

「A が (は)C られる」に対して，「A に C する」と「A から C する」と「A を C する」の3通りの能動表現を考え，これを係り元単語，係り先単語，表層格（ニ，カラ，ヲ）として，辞書引きを行い2章における通常の Sage の語意-格組の総合評価値を求める手法でそれぞれ計算し，これら全ての総合評価値を Jiri-Harada に渡して最適な語意格組を決定する．

6. SAGE の Web 化

SAGE の応用において生じてきた様々な要求に沿った改良を行った．具体的には次のようなことを行った．①形態素解析・係り受け解析との一貫した連動，②文節番号の追加，③KNP による解析結果の品詞の追加，④生成される格フレーム要素の選択，⑤SAGE システムの Web 化．

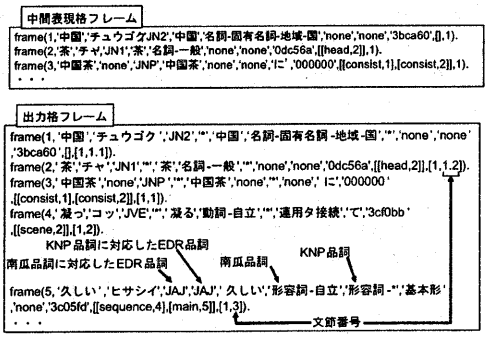


図 5　中間表現格フレームと出力格フレームの違い
Fig. 5　A difference between the middle expression and final expression frame.

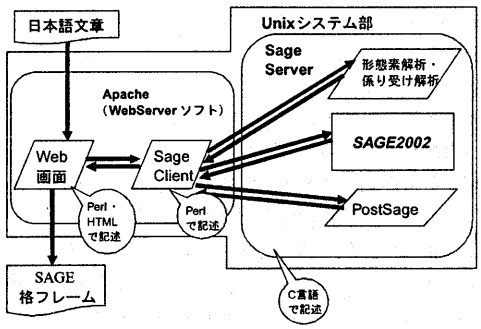


図 6　SAGE Web Server の流れ
Fig. 6　The flow of SAGE Web Server.

6.1　形態素解析・係り受け解析との一貫した連動

以前より要求の多かった形態素解析・係り受け解析との一貫した連動を行った．この統合は，Web 化を行う上で必要であり，これによって解析を行いたい日本語の文章を入力するだけで，格フレームの出力までを一気に行えるようになった．

6.2　postSAGE

上記目的の②〜④を実現するために，新たに postSAGE の開発を行った．これにより，従来の格フレームを中間表現とし，図 5 に示す本研究で新しく定めた格フレームを SAGE からの最終出力とする．このような格フレーム情報の充実により，SAGE が生成する格フレームを用いた応用システムの開発が進むものと考えられる．実際現在，この格フレーム情報の追加により本研究室において新たに文章要約システムと指示詞の照応解析システムについての研究が行われている．この postSAGE による各機能の仕組みは次のようになっている．

6.2.1　文節番号の追加

従来の SAGE 出力での格フレームでは，図 5 のように文番号は出力していたが，その文内での各文節の順番に関する情報は全く存在していなかった．

新 SAGE では，新たに図 5 の m や m.n（ドット区切り付）のような文節番号を追加する．ここで，単純に数字 1 つだけの m は，この文節が「複合語・活用語・派生語」の「構成語・活用語尾・接辞」もしくはその主辞（いずれの格フレームにおいても consist 格のあて先として指定されていない）ではなく，そのようなものの中で，文中で m 番目に現れていることを表す．次に m.n は，当該文節が m という文節番号が振られた文節の構成要素（文節 m の consist 格のあて先）であり，その中で n 番目の要素であることを表わしている．

6.2.2　品詞の追加

中間表現格フレームでは図 5 のように南瓜による解析結果の品詞とそれに対応する EDR 品詞が出力されていた．本研究では，さらに図 5 の格フレームのように KNP による解析結果の品詞とそれに対応する EDR 品詞を追加した．このことにより，応用システムで使われる品詞として二系統の品詞に対応できるようになった．

6.3　SAGE Web Server

3 章と 4 章と 5 章で改良された SAGE と 6.1 節と 6.2 節で論じた改良を行ったシステム全体を Unix 上に移植し，さらに Web 化を行った．

システム構成は，図 6 に示すようになっている．フロントエンドである Web Server 上の Web ページは，HTML と Perl で記述されており，Sage Server システムは C 言語で記述されている．また，複数ユーザーの同時利用への対策として，SAGE 利用中には利用中を示すファイルが作られ，このファイルが存在してる時には，ほかのユーザが利用できないようにしている．これにより，固定名である中間生成ファイルへの複数同時のアクセス等が防げるようになっている．

7.　おわりに

SAGE の解析精度を自動的に評価するシステム EvalSAGE を構築し，実際に 100 文に対して評価を行った．我々は評価対象文として EDR 電子化辞書のコーパス辞書に記述されている例文をランダムに選ぶことにした．EvalSAGE は，SAGE が生成した格フレームと，コーパス辞書からの意味解析済みデータを格フレーム形式に変換したコーパスフレームを入力し，語意・深層格・係り受け関係をそれぞれ自動的に照合

する．この照合の結果，SAGEが生成した格フレームとコーパスフレームの結果が一致しているものを"1"，一致していないものを"0"として，フレームごとに語意の正誤，深層格の正誤，あて先の正誤をExcel表の各セルに出力する．

ただし，この照合に際して，語意に対しては，コーパス語意が補足付概念説明であるもの（手作業で確認してあっているものを正解とした）はEDRにおいても概念IDが与えられていないので評価対象外とした．また語意が完全に一致していなくても評価実験者が常識から判断して正しいとしたものは正解とした．さらに活用語尾の語意は応用システム上必要ないとして評価対象外とした．また，深層格に対しては，係り受けが一致していないもの（構文解析結果は南瓜とKNPに委ねているため）を評価対象外とした．また，使役や受身のようにEDRに存在しないでSAGEが独自に出力した深層格は手作業で確認した．

コーパス辞書の例文100例文において，SAGE2002が生成した格フレームを，この精度評価システムを用いて評価した．なお，この100例文中の照合に用いられたデータ数は語意で1131件，深層格で1361件であった．この実験結果では，語意正解率は89.2%，深層格正解率は89.0%であった．

SAGEは南瓜とKNPの出力である係り受け木を入力としているため，係り受け解析の誤りはSAGEの意味解析に影響する．また，共起辞書に補足つき概念説明が記述されている（共起事例が存在するのに語意が存在しない）場合間違った語意と深層格の確率が高くなることがある．これらのことは特にあて先の語意の誤りとその前後の2文節の深層格の誤りを誘発し，その文中のそれらの精度を下げていると考えられる．これらの事は本システムの対象外の所与条件であるので，これらに起因する誤りは現段階では改善しようがない．

本研究により，SAGEの解析精度は実利用可能なレベルに近づいたといえる．今後の課題としては，『（，[，「』などの記号，箇条書き，数式を含む表現や対話表現の文に対する適用範囲の拡張や，さらなる高速化に向けて処理の並列化などがある．

謝辞

本研究を進めるにあたり，『茶筌』と『南瓜』を提供してくださった奈良先端科学技術大学院大学の松本裕治教授に深く感謝いたします．また，『JUMAN』と『KNP』を提供してくださった京都大学の言語メディア研究室，そして，Windows版『JUMAN』と『KNP』を開発・提供してくださった東京大学の黒橋禎夫助教授に深く感謝いたします．なお，この研究の一部は，文部科学省科学研究費基盤研究C『日本語文章の常識を用いた意味理解・文脈理解システムの開発研究』の補助金を用いて行われました．

参考文献

1) 原田実，野村佳秀，山本幸二，大野雅志，田村浩樹，高橋史郎：自然語要求仕様からオブジェクト指向設計図を自動生成するシステムCAMEO，情報処理学会論文誌，Vol. 38, No. 10, pp. 2031–2039(1997).
2) 原田実，水野高宏：EDRを用いた日本語意味解析システムSAGE，人工知能学会論文誌，Vol. 16, No. 1, pp. 85–93(2001).
3) 原田実，鈴木亮，南児瑞：意味グラフのマッチングによる事故問い合わせ文からの判例検索システムJCare，言語処理学会論文誌，Vol.9, No.4, pp.3-22 (2002.4).
4) 原田実，田淵和幸，大野博之：日本語意味解析システムSAGEの高速化・高精度化とコーパスによる精度評価，情報処理学会論文誌，Vol.43, No.9, pp.2894-2902, (2002.9).
5) 黒橋禎夫，長尾真：日本語形態素解析システムJUMAN version 3.61使用説明書，京都大学大学院情報学研究科 (1998).
6) 黒橋禎夫，長尾眞：並列構造の検出に基づく長い日本語文の構文解析，自然言語処理，Vol.1, No.1, pp.35-57(1994).
7) 益岡隆志，田窪幸則：基礎日本語文法-改訂版-，くろしお出版 (1992).
8) 松本裕治，北内啓，山下達雄，平野善隆，今一修，今村友明：日本語形態素解析システム『茶筌』version 2.0使用説明書，奈良先端科学技術大学院大学松本研究室 (1999).
9) 水野高宏，原田実：日本語意味解析システムSAGEの高速化と精度向上，人工知能学会第14回全国大会論文集，pp. 149–152(2000).
10) 尾見孝一郎，原田実，岩田隆志，水野高宏：日本語文章からの意味フレーム自動生成システムSAGE(Semantic frame Automatic GEnerator)の開発研究，人工知能学会第13回全国大会論文集，pp. 213–216(1999).
11) 矢後友和，原田実：日本語要求仕様文章からオブジェクト指向による動的モデルを生成するシステムCAMEO/Dの開発，情報処理学会第62回全国大会論文集，pp. 95–98(2001).
12) 矢後友和，原田実：日本語要求仕様文章からシーケンス図を自動生成するシステムCAMEO/Dの開発と販売管理システム問題への適用，情報処理学会オブジェクト指向2001シンポジウム論文集，pp. 9–16(2001).