(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2017-208049
(P2017-208049A)
(43)公開日　平成29年11月24日(2017.11.24)

(51)Int.Cl.		F I			テーマコード（参考）
G06F 17/27	(2006.01)	G06F	17/27	675	5B091
G06F 17/30	(2006.01)	G06F	17/30	330C	

審査請求 未請求　請求項の数 7　ＯＬ　（全 19 頁）

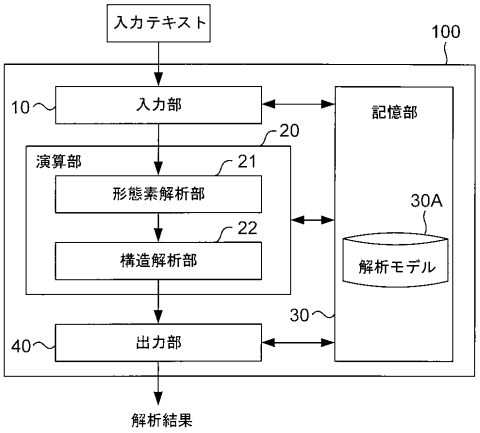
(21)出願番号　特願2016-101930 (P2016-101930)
(22)出願日　平成28年5月20日 (2016.5.20)

(71)出願人　000004226
日本電信電話株式会社
東京都千代田区大手町一丁目５番１号
(74)代理人　110001519
特許業務法人太陽国際特許事務所
(72)発明者　田中　貴秋
東京都千代田区大手町一丁目５番１号　日本電信電話株式会社内
(72)発明者　永田　昌明
東京都千代田区大手町一丁目５番１号　日本電信電話株式会社内
(72)発明者　林　克彦
東京都千代田区大手町一丁目５番１号　日本電信電話株式会社内

Ｆターム(参考)　5B091 AA15　CA02　CA07

(54)【発明の名称】言語解析装置、言語解析方法、及びプログラム

(57)【要約】
【課題】構文構造解析の精度を向上させる。
【解決手段】言語解析装置１００は、自然言語で記述された入力テキストを単語に分割して品詞の情報を付与し、単語の分割結果及び品詞の付与結果から得られる単語の素性と、１つ以上の単語の固まりであるチャンクの素性と、分割された単語をチャンクに統合すると共に依存関係を表わすラベルを付与するための予め定められた解析モデルとに基づいて、分割された単語をチャンクに統合すると共に、単語間、又はチャンク間の各々に依存関係を表すラベルを付与する。
【選択図】図１



【特許請求の範囲】
【請求項１】
自然言語で記述された入力テキストを単語に分割し、品詞の情報を付与する形態素解析部と、
前記形態素解析部による前記単語の分割結果及び品詞の付与結果から得られる単語の素性と、１つ以上の単語の固まりであるチャンクの素性と、前記分割された単語を前記チャンクに統合すると共に依存関係を表わすラベルを付与するための予め定められた解析モデルとに基づいて、前記分割された単語を前記チャンクに統合すると共に、単語間、又はチャンク間の各々に前記依存関係を表すラベルを付与する構造解析部と、
を含む言語解析装置。
【請求項２】
前記構造解析部は、前記分割された単語を前記チャンクとして長単位に統合すると共に依存関係を表わすラベルを付与するための前記解析モデルに基づいて、前記分割された単語を前記長単位に統合すると共に、単語間、又は長単位間の各々に前記依存関係を表すラベルを付与する
請求項１記載の言語解析装置。
【請求項３】
前記構造解析部は、前記分割された単語を前記チャンクとして単名詞句に統合し固有表現クラス又は一般名詞クラスの意味クラスを付与すると共に依存関係を表すラベルを付与するための前記解析モデルに基づいて、前記分割された単語を前記単名詞句に統合し、固有表現クラス又は一般名詞クラスの意味クラスを付与すると共に、単語間、又は単名詞句間の各々に前記依存関係を表すラベルを付与する
請求項１又は請求項２記載の言語解析装置。
【請求項４】
自然言語で記述された入力テキストを単語に分割して品詞の情報を付与し、
前記単語の分割結果及び品詞の付与結果から得られる単語の素性と、１つ以上の単語の固まりであるチャンクの素性と、前記分割された単語を前記チャンクに統合すると共に依存関係を表わすラベルを付与するための予め定められた解析モデルとに基づいて、前記分割された単語を前記チャンクに統合すると共に、単語間、又はチャンク間の各々に前記依存関係を表すラベルを付与する
言語解析方法。
【請求項５】
前記分割された単語を前記チャンクとして長単位に統合すると共に依存関係を表わすラベルを付与するための前記解析モデルに基づいて、前記分割された単語を前記長単位に統合すると共に、単語間、又は長単位間の各々に前記依存関係を表すラベルを付与する
請求項４記載の言語解析方法。
【請求項６】
前記分割された単語を前記チャンクとして単名詞句に統合し固有表現クラス又は一般名詞クラスの意味クラスを付与すると共に依存関係を表すラベルを付与するための前記解析モデルに基づいて、前記分割された単語を前記単名詞句に統合し、固有表現クラス又は一般名詞クラスの意味クラスを付与すると共に、単語間、又は単名詞句間の各々に前記依存関係を表すラベルを付与する
請求項４又は請求項５記載の言語解析方法。
【請求項７】
請求項１～請求項３の何れか１項に記載の言語解析装置の各部としてコンピュータを機能させるためのプログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、自然言語処理で使用される言語解析装置、言語解析方法、及びプログラムに

関する。
【背景技術】
【０００２】
自然言語処理においては、処理の対象である自然言語に対して、文の構造を同定する構文解析が行われる。日本語の場合、文を例えば文節等の１語以上の単語を結合した単語の固まり（チャンク）に分割した後に、文節間の依存関係として定義される文節係り受け解析による構文解析が標準的に行われている（非特許文献１参照）。
【０００３】
また、非特許文献２に開示される長単位解析手段等を用いて、文を文節とは異なる長単位のチャンクに分割した後に、チャンク間の依存構造として定義される依存構造解析が行われることもある（非特許文献３参照）。
【０００４】
上述した何れの手法においても、文の構造をある文法的な機能を有するチャンクに分割し、チャンク内の不要な内部構造を無視することで、同定する構文構造を単純化し、構文解析の精度の向上を図ろうとしている。
【先行技術文献】
【非特許文献】
【０００５】
【非特許文献１】Taku Kudo and Yuji Matsumoto: "Japanese Dependency Analysis using Cascaded Chunking", The 6th Conference on Natural Language Learning 2002 (CoNLL-2002), Vol.20, pp.1-7.
【非特許文献２】小澤俊介、内元清貴、伝康晴著、「BCCWJに基づく長単位解析ツール Comainu」第20回年次大会予稿集、言語処理学会、2014年、pp.582-585.
【非特許文献３】Takaaki Tanaka and Masaaki Nagata: "Word-based Japanese typed dependency parsing with grammatical function analysis": In Proceedings of the Association for Computational Linguistics 53rd Annual Meeting(ACL-2015), Vol.2, pp.237-242.
【発明の概要】
【発明が解決しようとする課題】
【０００６】
しかしながら、ある文法的な機能を有するチャンクの分割は、入力された単語列の１次元的な情報から決定されるものではなく、より上位の構文構造と整合するように決定することが好ましい。
【０００７】
例えば、「予備／調査／結果／に／つい／て／報告／し／た」という短単位列からなる例文の構文解析について検討する。なお、ここで“／”は短単位の区切りを示す記号である。
【０００８】
１次元的な処理の場合、例えば例文を、例文の先頭から順に「予備調査結果」、「について」、「報告し」、「た」という長単位のチャンクに分割した後、分割したチャンク間の依存構造を同定する。
【０００９】
一方、上述した例文と類似した別の例文、「猫／の行動／に／関する／予備／調査／結果／に／つい／て／報告／し／た」を構文解析する場合、本来「猫の行動に関する」という句の依存先は、「予備調査」になるべきである。しかし、上述したように、１次元的に例文を処理する場合、例文の分割の段階で「予備調査結果」を１つのチャンクとして分割することがある。この場合、チャンク間で文の依存構造を解析しようとするため、チャンクの内部構造が隠蔽されてしまい、適切な構文構造を構築し難いという状況が発生することがある。
【００１０】

すなわち、文をチャンクに分割する処理（チャンキング）が全て終了してから、チャンク間で文の依存構造を解析しようとした場合、構文構造解析の精度が低下する状況が発生することがある。

【００１１】

本発明は、上記の事情を鑑みて成されたものであり、文をチャンクに分割し終えてから文の依存構造を解析する場合に比べて、構文構造解析の精度を向上させることができる言語解析装置、言語解析方法、及びプログラムを提供することを目的とする。

【課題を解決するための手段】

【００１２】

上記の目的を達成するために本発明に係る言語解析装置は、自然言語で記述された入力テキストを単語に分割し、品詞の情報を付与する形態素解析部と、前記形態素解析部による前記単語の分割結果及び品詞の付与結果から得られる単語の素性と、１つ以上の単語の固まりであるチャンクの素性と、前記分割された単語を前記チャンクに統合すると共に依存関係を表わすラベルを付与するための予め定められた解析モデルとに基づいて、前記分割された単語を前記チャンクに統合すると共に、単語間、又はチャンク間の各々に前記依存関係を表すラベルを付与する構造解析部と、を含む。

【００１３】

本発明に係る言語解析方法は、自然言語で記述された入力テキストを単語に分割して品詞の情報を付与し、前記単語の分割結果及び品詞の付与結果から得られる単語の素性と、１つ以上の単語の固まりであるチャンクの素性と、前記分割された単語を前記チャンクに統合すると共に依存関係を表わすラベルを付与するための予め定められた解析モデルとに基づいて、前記分割された単語を前記チャンクに統合すると共に、単語間、又はチャンク間の各々に前記依存関係を表すラベルを付与する。

【００１４】

本発明に係る言語解析装置のプログラムは、言語解析装置の各部としてコンピュータを機能させる。

【発明の効果】

【００１５】

以上説明したように、本発明の言語解析装置、言語解析方法、及びプログラムによれば、文をチャンクに分割し終えてから文の依存構造を解析する場合に比べて、構文構造解析の精度を向上させることができる、という効果が得られる。

【図面の簡単な説明】

【００１６】

【図１】言語解析装置の構成例を示す概略図である。

【図２】形態素解析結果の一例を示す図である。

【図３】文の構造解析に係るアクションの一例を示す図である。

【図４】素性テンプレートの一例を示す図である。

【図５】素性テンプレートの凡例を示す図である。

【図６】複合語辞書の一例を示す図である。

【図７】第１実施形態における構造解析ルーチンの処理の一例を示すフローチャートである。

【図８】言語解析装置における入力テキストの構造解析過程の一例を示す図である。

【図９】言語解析装置における入力テキストの構造解析結果の一例を示す図である。

【発明を実施するための形態】

【００１７】

以下、図面を参照して本発明に係る各実施の形態を詳細に説明する。なお、以下では、同じ働きを担う構成要素又は処理には全図面を通して同じ符号を付与し、重複する説明を適宜省略する。

【００１８】

＜第１実施形態＞

以下に示す実施形態では、一例として単語の最小単位を、非特許文献である「小椋秀樹、小磯花絵、冨士池優美、宮内佐夜香、小西光、原裕著『現代日本語書き言葉均衡コーパス(Balanced Corpus of Contemporary Written Japanese:BCCWJ)形態論情報規定集（上）』、国立国語研究所、2011年」において規定される短単位とする。すなわち、短単位とは、現代語において意味を持つ最小の単位を規定し、最小の単位を長単位の範囲内で短単位の認定規定に基づいて結合させるか、又は結合させないようにして認定した言語単位である。

【００１９】

また、以下に示す実施形態で扱うチャンクは、一例として、上記BCCWJ形態論情報規定集（上）において規定される長単位とする。すなわち、長単位とは、文節の認定を行った上で、各文節の内部を規定に従って自立語部分と付属語部分とに分割するか、又は分割しないようにして認定した言語単位である。したがって、長単位は文節を超えることはない。

【００２０】

しかしながら、チャンクの単位は長単位に限定されず、例えば１語以上の単語から構成されるものであれば、他の定義でチャンクを表してもよい。

【００２１】

また、以下に示す実施形態では、依存関係単位として長単位で表されるチャンクを用いる。したがって、文の依存構造は長単位間の依存関係で表される一方で、長単位、すなわち、チャンクに含まれる短単位に関する情報や、単名詞句或いは文節等の他のチャンクに関する情報も依存構造解析のための素性として利用することができる。

【００２２】

＜システム構成＞

図１は、第１実施形態に係る言語解析装置１００のシステム構成例を示す図である。言語解析装置１００は、ＣＰＵと、ＲＡＭと、後述する構造解析ルーチンを実行するためのプログラムを記憶したＲＯＭと、を備えたコンピュータで構成され、機能的には次に示すように構成されている。

【００２３】

言語解析装置１００は、入力部１０、演算部２０、記憶部３０、及び出力部４０を備える。また、演算部２０は、形態素解析部２１及び構造解析部２２を含み、記憶部３０は、解析モデル３０Ａを含む。

【００２４】

入力部１０は、自然言語で記述された入力テキストを受け付けると、受け付けた入力テキストを演算部２０に出力する。

【００２５】

演算部２０は、入力部１０から入力テキストを受け付けると、まず形態素解析部２１で受け付けた入力テキストに対して形態素解析を実行して、入力テキストを単語に分割すると共に、分割した単語の各々に対して、品詞及び出現形等の素性を対応付ける。

【００２６】

図２は、例文「昨日予備調査結果について報告した」を、形態素解析部２１で形態素解析した場合の結果の一例を示す図である。図２に示すように、例文「昨日予備調査結果について報告した」は、「昨日／予備／調査／結果／に／つい／て／報告／し／た」のように短単位で表される単語に分割される。

【００２７】

形態素解析部２１は、形態素解析によって分割された単語毎に、例えば短単位#s、出現形w、標準形m、及び品詞tが行方向に対応付けた形態素解析結果を得る。ここで、短単位#sは、入力テキストにおける単語の並び順に、入力テキストの先頭の単語から順に付与した番号であり、当該番号は互いに重複しないように各々の単語に付与される。また、出現形wは、入力テキスト中に現われた形（BCCWJにおける書字形）を示し、標準形mは、対応する単語が係り受け先や意味の違いによって語形変化を起こす活用語である場合の基本形

（BCCWJにおける語彙素）を示し、品詞tは、対応する単語の品詞情報を示す。

【００２８】

品詞t欄に記載した記号は品詞の種類であり、"ADV"は副詞、"NN"は名詞、"PCS"は格助詞、"VB"は動詞、"PCJ"は接続助詞、及び"AUX"は助動詞を表している。なお、図２に示した形態素解析結果は一例であり、例えば発音形等の他の素性を単語毎に対応付けるようにしてもよい。また、品詞tについても、BCCWJで規定された、より細分化された品詞を単語毎に対応付けるようにしてもよい。

【００２９】

構造解析部２２は、形態素解析部２１で得られた入力テキストの形態素解析結果からチャンク解析及び依存構造解析を行うための素性を抽出し、抽出した素性と、予め用意した解析モデルと、に基づいて、入力テキストの構造解析を行う。ここで、チャンク解析とは、形態素解析部２１で短単位に分割した単語を長単位のチャンクに統合する処理をいい、依存構造解析とは、統合したチャンク間又は単語間の係り受け関係を同定する処理をいう。

【００３０】

チャンク解析及び依存構造解析には、例えばMaltParser等の既存の構造解析器で使用されるshift-reduce法を用いることができる。

【００３１】

shift-reduce法とは、出現形wで表される入力テキストの短単位列$(w_0,\cdots,w_n)$を、短単位の並び順にキューQ内の配列$(Q_0,\cdots,Q_n)$にそれぞれ格納し、shiftアクション又はreduceアクションを行うステップを繰り返すことで、短単位列$(w_0,\cdots,w_n)$をスタックS上で長単位に統合してチャンクを生成すると共に、チャンク間の係り受け関係を同定する構造解析手法である。ここで"n"は入力テキストに含まれる短単位の数を表す。

【００３２】

以降では、キューQ内の配列$(Q_0,\cdots,Q_n)$に格納されるそれぞれの短単位を、短単位列$(q_0,\cdots,q_n)$として表すことにする。

【００３３】

なお、shiftアクションとは、キューQの先頭に格納される短単位q_0をスタックSにプッシュする動作をいい、reduceアクションとは、スタックSの最上位の要素s_0と、要素s_0に隣接する２番目の要素s_1と、の間で弧を張り、要素s_0の係り受け関係を決定する処理をいう。

【００３４】

この際、構造解析部２２は、チャンクの素性と、解析モデル３０Ａと、に基づいて、各ステップでshiftアクション及びreduceアクションの何れのアクションを行うかを決定する。

【００３５】

ここで、解析モデル３０Ａは、様々な例文に関する、チャンクの依存構造を示す構造解析結果を予め記憶した構造解析モデルであり、例えば記憶部３０に予め構築されている。解析モデル３０Ａに、例文毎の構造解析結果（正解データ）を学習データとして大量に蓄積しておくことで、構造解析部２２は、正解データからチャンクの素性と、実行されたアクションと、の組み合わせを抽出し、例えばSupport Vector Machine(SVM)、対数線形モデル、又はニューラルネットワークを用いたパターン認識等の公知の分類モデルに基づいて、入力テキストに含まれるチャンクの素性に対応する最適なアクションをステップ毎に決定し、入力テキストの構造解析を行う。

【００３６】

なお、チャンクの素性とは、構造解析の対象となっているチャンク自身の素性の他、チャンクに含まれる単語に関する素性や、単名詞句或いは文節等の他のチャンクに関する素性も含まれるものとする。

【００３７】

図３は、構造解析部２２で用いるアクションを示した図である。図３に示すように、構

造解析部２２は、"$Shift_{LUW}$(POS)"、"$Reduce_{LUWL}$(d1)"、"$Reduce_{LUWR}$(d1)"、"Pop_{LUW}"、"$Shift_{SUW}$"、"$Reduce_{SUWL}$(d2)"、"$Reduce_{SUWR}$(d2)"の７つのアクションを用いて、入力テキストの構造解析を行う。なお、構造解析部２２で用いるアクションは図３に示したアクションに限られず、例えば形態素解析部２１での形態素解析内容に応じて、他のアクションを規定してもよい。

【００３８】

構造解析部２２は、$Shift_{LUW}$(POS)、又は$Shift_{SUW}$で、キューQに格納された短単位列(q_0,・・・,q_n)の先頭から順に短単位をスタックSにプッシュしていき、$Reduce_{SUWL}$(d2)、又は$Reduce_{SUWR}$(d2)でスタックSの短単位間に弧を張り、チャンク内部の依存構造を決定する。また、構造解析部２２は、Pop_{LUW}でステップＳの最上位の要素s_0をチャンクとして確定させ、$Reduce_{LUWL}$(d1)、又は$Reduce_{LUWR}$(d1)でチャンク間に弧を張り、入力テキストにおけるチャンク単位の依存構造で表す。

【００３９】

具体的には、$Shift_{LUW}$(POS)は、キューQの先頭から短単位q_0を取り出し、品詞がPOSで表される長単位のチャンクを構成する先頭の短単位として、スタックSにプッシュする。

【００４０】

すなわち、$Shift_{LUW}$(POS)は、短単位q_0をスタックSにプッシュするpush(S,q_0)、短単位q_0をスタックSにプッシュした後、キューQの各短単位をポップするpop(Q)、要素s_0のチャンクの品詞をPOSに設定するSetPos(s_0,POS)、及び要素s_0がまだチャンクとして確定していないことを示すSetLuw(s_0,f)の各動作によって構成される。ここで、SetLuw関数の引数fは"false"を表しており、要素s_0がまだチャンクとして確定していないことを示している。すなわち、SetLuw関数はチャンクの確定状態を示すチャンクフラグの値を設定する。

【００４１】

$Shift_{SUW}$は、キューQの先頭要素から短単位q_0を取り出し、スタックSにプッシュする。すなわち、$Shift_{SUW}$は、push(S,q_0)、pop(Q)、SetLuw(s_0,f)の各動作によって構成される。

【００４２】

$Reduce_{SUWL}$(d2)は、それぞれ短単位であるスタックSの最上位の要素s_0と、要素s_0に隣接する２番目の要素s_1との間に、短単位間の関係ラベルd2を持つ左向きの弧を張って部分木を生成し、生成した部分木をスタックSの要素s_0とする。この際、スタックSから要素s_1を削除する。なお、左向きの弧とは、要素s_0から要素s_1に向かう方向の係り受け関係をいう。

【００４３】

すなわち、$Reduce_{SUWL}$(d2)は、スタックSの要素s_0と、要素s_1との間に、関係ラベルd2を持つ左向きの弧を張るLeftSArc(s_1,s_0,d2)、及びスタックSから要素s_1を削除するremove(S,s_1)の各動作によって構成される。

【００４４】

$Reduce_{SUWR}$(d2)は、それぞれ短単位であるスタックSの最上位の要素s_0と、要素s_1との間に、短単位間の関係ラベルd2を持つ右向きの弧を張って部分木を生成し、生成した部分木をスタックSの要素s_0とする。この際、スタックSから要素s_1を削除する。なお、右向きの弧とは、要素s_1から要素s_0に向かう方向の係り受け関係をいう。

【００４５】

すなわち、$Reduce_{SUWR}$(d2)は、スタックSの要素s_0と、要素s_1との間に、関係ラベルd2を持つ右向きの弧を張るRightSArc(s_1,s_0,d2)、及びremove(S,s_1)の各動作によって構成される。

【００４６】

Pop_{LUW}は、スタックSの最上位の要素s_0を、直前に行われた$Shift_{LUW}$(POS)のPOSで表される品詞を有するチャンクとして確定させる。

【００４７】

すなわち、Pop_{LUW}は、要素s_0がチャンクとして確定したことを示すSetLuw(s_0,t)の動作

によって構成される。ここで、SetLuw関数の引数tは"true"を表しており、要素s_0がチャンクとして確定したことを示している。また、チャンクの確定に伴い、当該チャンクに直前の$Shift_{LUW}$(POS)で設定された品詞が割り当てられる。

【００４８】

$Reduce_{LUWL}$(d1)は、スタックSの最上位の要素s_0と、要素s_1との間に、チャンク間の関係ラベルd1を持つ左向きの弧を張って部分木を生成し、生成した部分木をスタックSの要素s_0とする。この際、スタックSから要素s_1を削除する。

【００４９】

すなわち、$Reduce_{LUWL}$(d1)は、スタックSの要素s_0と、要素s_1との間に、関係ラベルd1を持つ左向きの弧を張るLeftLArc(s_1,s_0,d1)、及びremove(S,s_1)の各動作によって構成される。

【００５０】

$Reduce_{LUWR}$(d1)は、スタックSの最上位の要素s_0と、要素s_1との間に、チャンク間の関係ラベルd1を持つ右向きの弧を張って部分木を生成し、生成した部分木をスタックSの要素s_0とする。この際、スタックSから要素s_1を削除する。

【００５１】

すなわち、$Reduce_{LUWR}$(d1)は、スタックSの要素s_0と、要素s_1との間に、関係ラベルd1を持つ右向きの弧を張るRightLArc(s_1,s_0,d1)、及びremove(S,s_1)の各動作によって構成される。

【００５２】

ここで、図４は、構造解析部２２が抽出するチャンクの素性の一部を示した素性テンプレートであり、構造解析部２２は、チャンクに含まれる短単位の素性の他、例えばアクションの決定等に長単位素性及び複合語素性を用いることができる。また、図５は、図４に示した素性テンプレートの凡例を示す図である。

【００５３】

例えば、図４において、長単位素性の一例である

$$s_0.h_{Lw} \circ s_1.h_{Lw}$$

は、スタックSの要素s_0の主辞（部分木の根ノード）の長単位出現形の素性と、スタックSの要素s_1の主辞の長単位出現形の素性と、の組み合わせ素性を表す。

【００５４】

構造解析部２２は、shift-reduce法を用いて、スタックS中に入力テキストの部分的な依存構造である部分木を生成しながら入力テキストの構造解析を行うため、チャンクの素性として、部分木のノードの情報を用いることができる。更に、構造解析部２２は、チャンクを依存構造とは区別して扱うことができるため、部分木のノードの情報の他、品詞や出現形等のチャンク自体の素性も用いることができる。

【００５５】

また、図４に示すように、構造解析部２２は、固定的な表現を有するチャンクの分割精度を向上させるため、複合語辞書との照合結果を素性として取り入れることができる。ここで、複合語素性$q_n.f_{comp}(c_0,c_m)$とは、キューQのn番目の短単位q_nを先頭とする短単位列と、短単位数がc_0以上c_m以下の複合語辞書に含まれる複合語との照合フラグであり、照合OK、すなわち、短単位q_nを先頭とする短単位列が複合語辞書に含まれる場合は"1"となり、照合NG、すなわち、短単位q_nを先頭とする短単位列が複合語辞書に含まれない場合は"0"となる。

【００５６】

複合語辞書は、例えば記憶部３０の予め定めた領域に格納される。

【００５７】

図６は、複合語辞書の一例を示す図である。複合語辞書は、例えば複合語の表記及び複合語の品詞を表す「複合語／複合語品詞」、複合語に含まれる短単位の数を表す「短単位

数」、並びに、複合語に含まれる短単位の出現形、品詞、及び標準形等を表す「短単位列」を複合語毎に対応付けた情報を含む。

【００５８】

例えば複合語／複合語品詞欄の「について／PCS」は複合語「について」の品詞が格助詞で、且つ、短単位数欄の"3"は複合語「について」に含まれる短単位の数が3であることを表している。更に、短単位列欄には、複合語「について」が、「に／つい／て」のように短単位に分割でき、短単位「に」の品詞は格助詞で標準形は「に」、短単位「つい」の品詞は動詞で標準形は「つく」、及び短単位「て」の品詞は接続助詞で標準形は「て」であることが示されている。

【００５９】

複合語辞書の短単位数は、極端に短い複合語、或いは極端に長い複合語への照合を制限し、照合時間を短縮するために設けられた情報である。例えば、図４の複合語素性のテンプレートに含まれる$q_0.f_{comp}(3,5)$は短単位数が3以上5以下の短単位で構成される複合語を照合対象としているため、図６の複合語辞書において、複合語「だが」は照合対象とならない。

【００６０】

なお、図６に示した複合語辞書は一例であり、上述した項目以外の情報を複合語辞書に含めてもよいことは言うまでもない。

【００６１】

構造解析部２２は、短単位に分割された入力テキストの各単語に対して、図３に示したshift-reduce法に伴うアクションを実行して生成した、入力テキストに対応するチャンクの依存構造を示す部分木を、入力テキストの構造解析結果として出力部４０に出力する。

【００６２】

出力部４０は、構造解析部２２から入力テキストの構造解析結果を受け付けると、受け付けた構造解析結果を、例えば入力テキストの構造解析結果に基づいて入力テキストと類似する文を検索する検索装置、或いは液晶ディスプレイ等の表示装置等に出力する。

【００６３】

＜言語解析装置の作用１＞

次に、第１実施形態に係る言語解析装置１００の作用について説明する。

【００６４】

言語解析装置１００は、入力テキストを入力部１０で受け付けると、受け付けた入力テキストを形態素解析部２１に出力する。

【００６５】

形態素解析部２１は、既存の形態素解析器を用いて、受け付けた入力テキストを短単位である単語に分割して形態素解析を行い、例えば図２に示した形態素解析結果を構造解析部２２に通知する。

【００６６】

そして、構造解析部２２は、形態素解析部２１から形態素解析結果を受け付けると、ＣＰＵで図７に示す構造解析ルーチンを実行する。

【００６７】

まず、ステップＳ１００において、記憶部３０にスタックS及びキューQを生成し、スタックSの最上位の要素s_0に、スタックSが空であることを示す文字列"ROOT"を設定する。また、キューQの各配列の値を"NULL"に初期化した上で、形態素解析結果における出現形欄の短単位列$(w_0,・・・,w_n)$を、短単位の並び順にキューQの配列$(Q_0,・・・,Q_n)$にそれぞれ格納して短単位列$(q_0,・・・,q_n)$を生成する。

【００６８】

ステップＳ１０５において、素性テンプレートに基づいて、スタックS及びキューQの状態に対応した素性を抽出する。

【００６９】

ステップＳ１１０において、ステップＳ１０５で抽出した素性と、記憶部３０の解析モ

デル３０Ａと、に基づいて、図３に示したアクションの中から、抽出した素性に対応する最適なアクションを決定する。

【００７０】

ステップＳ１１５において、ステップＳ１１０で決定されたアクションが構造解析ルーチンの終了を指示するものであるか否かを判定し、肯定判定の場合、図７に示す構造解析ルーチンを終了する。

【００７１】

なお、スタックSの要素s_0の主辞が"ROOT"で、且つ、キューQに何れの短単位も格納されていない状態、すなわち、キューQの短単位q_0が"NULL"である場合に、ステップＳ１１０で終了を指示するアクションが決定される。

【００７２】

一方、ステップＳ１１５の判定処理が否定判定の場合には、ステップＳ１２０に移行する。

【００７３】

ステップＳ１２０において、ステップＳ１１０で決定されたアクションが"$Shift_{LUW}$(POS)"であるか否かを判定し、肯定判定の場合、ステップＳ１５５に移行する。

【００７４】

ステップＳ１５５では、キューQの短単位q_0に対して$Shift_{LUW}$(POS)を実行する。この際、キューQから短単位q_0が取り出されるため、短単位q_1を短単位q_0、短単位q_2を短単位q_1、・・・というように、キューQに格納されている短単位q_1以降の各短単位をpopさせるが、これに伴い、短単位q_nが格納されなくなったキューQの配列Q_nには"NULL"を設定する。

【００７５】

一方、ステップＳ１２０の判定処理が否定判定の場合には、ステップＳ１２５に移行する。

【００７６】

ステップＳ１２５において、ステップＳ１１０で決定されたアクションが"$Shift_{SUW}$"であるか否かを判定し、肯定判定の場合、ステップＳ１６０に移行する。

【００７７】

ステップＳ１６０では、キューQの短単位q_0に対して$Shift_{SUW}$を実行する。この際、$Shift_{LUW}$(POS)と同様に、キューQから短単位q_0が取り出されるため、短単位q_1以降の各短単位のpop処理によって短単位q_nが格納されなくなったキューQの配列Q_nには"NULL"を設定する。

【００７８】

一方、ステップＳ１２５の判定処理が否定判定の場合には、ステップＳ１３０に移行する。

【００７９】

ステップＳ１３０において、ステップＳ１１０で決定されたアクションが"Pop_{LUW}"であるか否かを判定し、肯定判定の場合、ステップＳ１６５に移行する。

【００８０】

ステップＳ１６５ではPop_{LUW}を実行し、スタックSの要素s_0をチャンクとして確定する。

【００８１】

一方、ステップＳ１３０の判定処理が否定判定の場合には、ステップＳ１３５に移行する。

【００８２】

ステップＳ１３５において、ステップＳ１１０で決定されたアクションが"$Reduce_{LUWL}$(d1)"であるか否かを判定し、肯定判定の場合、ステップＳ１７０に移行する。

【００８３】

ステップＳ１７０では、スタックSの要素s_0と、要素s_1に対して$Reduce_{LUWL}$(d1)を実行し、チャンク間の依存構造を設定する。

【００８４】
一方、ステップＳ１３５の判定処理が否定判定の場合には、ステップＳ１４０に移行する。
【００８５】
ステップＳ１４０において、ステップＳ１１０で決定されたアクションが"$Reduce_{LUWR}$(d1)"であるか否かを判定し、肯定判定の場合、ステップＳ１７５に移行する。
【００８６】
ステップＳ１７５では、スタックSの要素s_0と、要素s_1に対して$Reduce_{LUWR}$(d1)を実行し、チャンク間の依存構造を設定する。
【００８７】
一方、ステップＳ１４０の判定処理が否定判定の場合には、ステップＳ１４５に移行する。
【００８８】
ステップＳ１４５において、ステップＳ１１０で決定されたアクションが"$Reduce_{SUWL}$(d2)"であるか否かを判定し、肯定判定の場合、ステップＳ１８０に移行する。
【００８９】
ステップＳ１８０では、スタックSの要素s_0と、要素s_1に対して$Reduce_{SUWL}$(d2)を実行し、チャンク内部の依存構造を設定する。
【００９０】
一方、ステップＳ１４５の判定処理が否定判定の場合には、ステップＳ１５０に移行する。
【００９１】
ステップＳ１５０において、スタックSの要素s_0と、要素s_1に対して$Reduce_{SUWR}$(d2)を実行し、チャンク内部の依存構造を設定する。
【００９２】
そして、ステップＳ１５０、Ｓ１５５、Ｓ１６０、Ｓ１６５、Ｓ１７０、Ｓ１７５、及びＳ１８０の各ステップを実行した後、ステップＳ１０５に移行する。そして、ステップＳ１１０で終了を指示するアクションが決定されるまで、ステップＳ１０５～Ｓ１８０の各ステップを繰り返し実行して、スタックS及びキューQの状態に対応する素性と、解析モデル３０Ａと、に基づいて、アクションを決定して実行し、入力テキストに含まれる単語をチャンクに統合すると共に、チャンク間の依存構造を示す部分木を生成する。
【００９３】
＜動作例＞
図８は、例えば入力テキストとして「昨日予備調査結果について報告した」を言語解析装置１００に入力した場合に、図７に示した構造解析ルーチンで実行される構造解析過程の一例を具体的に示す図である。
【００９４】
なお、図８では、スタックSは要素s_0から要素s_3まで、キューQは短単位q_0から短単位q_2までを記載している。Step欄は処理を識別するIDであり、各StepでAction欄に記載されたアクションを実行した後のスタックS及びキューQの状況が、行方向に対応付けられて記載されている。
【００９５】
ステップ０は、図７のステップＳ１００に対応する処理であり、スタックSの要素s_0として"ROOT"を設定し、キューQの短単位列(q_0,q_1,q_2)を（昨日,予備,調査）に設定する。
【００９６】
そして、ステップ０の状態で抽出されたチャンクの素性と解析モデル３０Ａとに基づいて、キューQの短単位q_0を先頭にして構成されるチャンクの品詞と、実行するアクションが決定される。
【００９７】
チャンクの品詞が"ADV"、及びアクションが$Shift_{LUW}$(ADV)に決定されると、ステップ１

で$Shift_{LUW}$(ADV)を実行して、短単位q_0(＝昨日)をキューQから取り出してスタックSの要素s_0としてpushする。この際、要素s_0はチャンクを構成していないため、SetLuw(s_0,f)によってチャンクフラグを"false"に設定すると共に、SetPos(s_0,ADV)によってチャンクの品詞を"ADV"に設定する。また、pop(Q)によってキューQに格納されている各短単位を移動させる。

【００９８】

そして、ステップ１の状態で抽出されたチャンクの素性と解析モデル３０Ａとに基づいて、次に実行するアクションが決定される。

【００９９】

実行するアクションがPop_{LUW}に決定されると、ステップ２でPop_{LUW}を実行して、「昨日」をチャンクとして確定し、SetLuw(s_0,t)によってチャンクフラグを"true"に設定すると共に、「昨日」の品詞を"ADV"に設定する。

【０１００】

そして、ステップ２の状態で抽出されたチャンクの素性と解析モデル３０Ａとに基づいて、キューQの短単位q_0を先頭にして構成されるチャンクの品詞と、次に実行するアクションが決定される。

【０１０１】

チャンクの品詞が"NN"、及び実行するアクションが$Shift_{LUW}$(NN)に決定されると、ステップ３で$Shift_{LUW}$(NN)を実行して、短単位q_0(＝予備)をキューQから取り出してスタックSの要素s_0としてpushする。この際、要素s_0はチャンクを構成していないため、SetLuw(s_0,f)によってチャンクフラグを"false"に設定すると共に、SetPos(s_0,NN)によってチャンクの品詞をNNに設定する。また、pop(Q)によってキューQに格納されている各短単位を移動させる。

【０１０２】

そして、ステップ３の状態で抽出されたチャンクの素性と解析モデル３０Ａとに基づいて、次に実行するアクションが決定される。

【０１０３】

実行するアクションが$Shift_{SUW}$に決定されると、ステップ４で$Shift_{SUW}$を実行して、短単位q_0(＝調査)をキューQから取り出してスタックSの要素s_0としてpushする。この際、要素s_0はチャンクを構成していないため、SetLuw(s_0,f)によってチャンクフラグを"false"に設定すると共に、pop(Q)によってキューQに格納されている各短単位を移動させる。

【０１０４】

そして、ステップ４の状態で抽出されたチャンクの素性と解析モデル３０Ａとに基づいて、次に実行するアクションが決定される。

【０１０５】

実行するアクションが$Reduce_{SUWL}$(cn)に決定されると、ステップ５で$Reduce_{SUWL}$(cn)を実行して、LeftSArc(s_1,s_0,cn)によってスタックSの要素s_0「調査」と、要素s_1「予備」との間に短単位間の関係ラベルcnを持つ左向きの弧を張った部分木「予備←(cn)←調査」を生成し、当該部分木を新たな要素s_0とすると共に、要素s_1「予備」をスタックSから削除する。なお、関係ラベルcnは、複合名詞の内部構造を構成する関係を表すラベルである。したがって、要素s_0は、２つの短単位「予備」と「調査」とが関係ラベルcnを持つ部分木で構成されていることを示している。

【０１０６】

そして、ステップ５の状態で抽出されたチャンクの素性と解析モデル３０Ａとに基づいて、次に実行するアクションが決定される。

【０１０７】

実行するアクションが$Shift_{SUW}$に決定されると、ステップ６で$Shift_{SUW}$を実行して、短単位q_0(＝結果)をキューQから取り出してスタックSの要素s_0としてpushする。この際、要素s_0はチャンクを構成していないため、SetLuw(s_0,f)によってチャンクフラグを"false"に設定すると共に、pop(Q)によってキューQに格納されている各短単位を移動させる。

【０１０８】
そして、ステップ６の状態で抽出されたチャンクの素性と解析モデル３０Ａとに基づいて、次に実行するアクションが決定される。
【０１０９】
実行するアクションが$Reduce_{SUWL}$(cn)に決定されると、ステップ７で$Reduce_{SUWL}$(cn)を実行して、LeftSArc(s_1,s_0,cn)によって要素s_0の短単位「結果」と、要素s_1の短単位の部分木「予備←(cn)←調査」と、の間に短単位間の関係ラベルcnを持つ左向きの弧を張った部分木「予備←(cn)←調査←(cn)←結果」を生成し、当該部分木を新たな要素s_0とすると共に、要素s_1「予備←(cn)←調査」をスタックSから削除する。
【０１１０】
そして、ステップ７の状態で抽出されたチャンクの素性と解析モデル３０Ａとに基づいて、次に実行するアクションが決定される。
【０１１１】
実行するアクションがPop_{LUW}に決定されると、ステップ８でPop_{LUW}を実行して、「予備←(cn)←調査←(cn)←結果」をチャンクとして確定し、SetLuw(s_0,t)によってチャンクフラグを"true"に設定すると共に、ステップ３のアクションに基づいて、チャンク「予備←(cn)←調査←(cn)←結果」の品詞を"NN"に設定する。以降では、チャンクの後ろに"/POS"（POSは品詞）を付加してチャンクに設定した品詞を表すことにする。
【０１１２】
そして、ステップ８の状態で抽出されたチャンクの素性と解析モデル３０Ａとに基づいて、キューQの短単位q_0を先頭にして構成されるチャンクの品詞と、次に実行するアクションが決定される。
【０１１３】
チャンクの品詞が"PCS"、及び実行するアクションが$Shift_{LUW}$(PCS)に決定されると、ステップ９で$Shift_{LUW}$(PCS)を実行して、短単位q_0(＝に)をキューQから取り出してスタックSの要素s_0としてpushする。この際、要素s_0はチャンクを構成していないため、SetLuw(s_0,f)によってチャンクフラグを"false"に設定すると共に、SetPos(s_0,PCS)によってチャンクの品詞を"PCS"に設定する。また、pop(Q)によってキューQに格納されている各短単位を移動させる。
【０１１４】
そして、ステップ９の状態で抽出されたチャンクの素性と解析モデル３０Ａとに基づいて、次に実行するアクションが決定される。
【０１１５】
実行するアクションが$Shift_{SUW}$に決定されると、ステップ１０で$Shift_{SUW}$を実行して、短単位q_0(＝つい)をキューQから取り出してスタックSの要素s_0としてpushする。この際、要素s_0はチャンクを構成していないため、SetLuw(s_0,f)によってチャンクフラグを"false"に設定すると共に、pop(Q)によってキューQに格納されている各短単位を移動させる。
【０１１６】
そして、ステップ１０の状態で抽出されたチャンクの素性と解析モデル３０Ａとに基づいて、次に実行するアクションが決定される。
【０１１７】
実行するアクションが$Reduce_{SUWL}$(mwe)に決定されると、ステップ１１で$Reduce_{SUWL}$(mwe)を実行して、LeftSArc(s_1,s_0,mwe)によって要素s_0の短単位「つい」と、要素s_1の短単位「に」と、の間に短単位間の関係ラベルmweを持つ左向きの弧を張った部分木「に←(mwe)←つい」を生成し、当該部分木を新たな要素s_0とすると共に、要素s_1「に」をスタックSから削除する。なお、関係ラベルmweは、複合語表現の内部構造を構成する関係を表すラベルである。
【０１１８】
そして、ステップ１１の状態で抽出されたチャンクの素性と解析モデル３０Ａとに基づいて、次に実行するアクションが決定される。

【０１１９】
実行するアクションが$Shift_{SUW}$に決定されると、ステップ１２で$Shift_{SUW}$を実行して、短単位q_0(＝て)をキューQから取り出してスタックSの要素s_0としてpushする。この際、要素s_0はチャンクを構成していないため、SetLuw(s_0,f)によってチャンクフラグを"false"に設定すると共に、pop(Q)によってキューQに格納されている各短単位を移動させる。
【０１２０】
そして、ステップ１２の状態で抽出されたチャンクの素性と解析モデル３０Ａとに基づいて、次に実行するアクションが決定される。
【０１２１】
実行するアクションが$Reduce_{SUWL}$(mwe)に決定されると、ステップ１３で$Reduce_{SUWL}$(mwe)を実行して、LeftSArc(s_1,s_0,mwe)によって要素s_0の短単位「て」と、要素s_1の短単位の部分木「に←(mwe)←つい」と、の間に短単位間の関係ラベルmweを持つ左向きの弧を張った部分木「に←(mwe)←つい←(mwe)←て」を生成し、当該部分木を新たな要素s_0とすると共に、要素s_1「に←(mwe)←つい」をスタックSから削除する。
【０１２２】
そして、ステップ１３の状態で抽出されたチャンクの素性と解析モデル３０Ａとに基づいて、次に実行するアクションが決定される。
【０１２３】
実行するアクションがPop_{LUW}に決定されると、ステップ１４でPop_{LUW}を実行して、「に←(mwe)←つい←(mwe)←て」をチャンクとして確定し、SetLuw(s_0,t)によってチャンクフラグを"true"に設定すると共に、ステップ９のアクションに基づいて、チャンク「に←(mwe)←つい←(mwe)←て」の品詞を"PCS"に設定する。
【０１２４】
そして、ステップ１４の状態で抽出されたチャンクの素性と解析モデル３０Ａとに基づいて、次に実行するアクションが決定される。
【０１２５】
この場合、素性テンプレートにより様々なチャンクの素性を抽出することができるが、そのうちの１つである長単位素性

$$s_0.h_{Lw} \circ s_0.h_{Lt}$$

により「に←(mwe)←つい←(mwe)←て/PCS」という、品詞が確定したチャンクの素性が抽出される。解析モデル３０Ａは、抽出したチャンクの素性に基づいて、チャンク間の依存関係を設定するアクション$Reduce_{LUWL}$(pobj)の実行を決定する。
【０１２６】
実行するアクションが$Reduce_{LUWL}$(pobj)に決定されると、ステップ１５で$Reduce_{LUWL}$(pobj)を実行して、LeftLArc(s_1,s_0,pobj)によってスタックSの要素s_0「に←(mwe)←つい←(mwe)←て/PCS」と、要素s_1「予備←(cn)←調査←(cn)←結果/NN」との間に長単位間の関係ラベルpobjを持つ左向きの弧を張った部分木「[予備←(cn)←調査←(cn)←結果/NN]←(pobj)←[に←(mwe)←つい←(mwe)←て/PCS]」を生成し、当該部分木を新たな要素s_0とすると共に、要素s_1「予備←(cn)←調査←(cn)←結果/NN」をスタックSから削除する。なお、関係ラベルpobjは、チャンク間の依存構造を示すラベルである。すなわち、要素s_0は、２つの長単位のチャンク「予備←(cn)←調査←(cn)←結果/NN」と「に←(mwe)←つい←(mwe)←て/PCS」とが、関係ラベルpobjで関係付けられた部分木を示す。
【０１２７】
以下、上述した処理と同様の処理が行われ、ステップ１６～２４が実行されると、スタックSの要素s_0に、ROOTを主辞とする部分木が構成される。
【０１２８】
なお、ステップ１８の関係ラベルnvは、サ変動詞の内部構造を構成する関係を表すラベルである。また、ステップ１９の関係ラベルiobj、ステップ２２の関係ラベルaux、ステ

ップ２３の関係ラベルadvmod、及びステップ２４の関係ラベルrootは、関係ラベルpobjと同様に、チャンク間の依存構造を示すラベルである。更に、図８の短単位q_0～短単位q_2において、"-"は"NULL"が設定されていることを示す。

【０１２９】

ステップ２４によって、スタックSの要素s_0の主辞が"ROOT"となり、且つ、キューQの短単位q_0が"NULL"となるため、図７に示したステップＳ１１５の判定処理が肯定判定となり、構造解析ルーチンが終了する。

【０１３０】

図９は、図８に示した例文「昨日予備調査結果について報告した」の構造解析によって得られた結果の一例を示す図である。

【０１３１】

図９では、例文が５つの長単位のチャンク「昨日」、「予備調査結果」、「について」、「報告し」、及び「た」に分割され、チャンクの品詞は、それぞれ"ADV(副詞)"、"NN(名詞)"、"PCJ(格助詞)"、"VB(動詞)"、及び"AUX(助動詞)"であることを示している。チャンクの中には、例えば「昨日」のように１つの短単位でチャンクを構成するものも存在する。

【０１３２】

ここで、背景色が黒色で表される関係ラベルが付加された弧からなる構造は、チャンク内部の依存構造を表している。

【０１３３】

図９の構造解析結果によれば、チャンク「予備調査結果」は、短単位「予備」と短単位「調査」との間、及び、短単位「調査」と短単位「結果」との間に、それぞれ関係ラベルcnで表される依存構造を含み、チャンク「については」は、短単位「に」と短単位「つい」との間、及び、短単位「つい」と短単位「て」との間に、それぞれ関係ラベルmweで表される依存構造を含む。チャンク「報告し」は、短単位「報告」と短単位「し」との間に、関係ラベルnvで表される依存構造を含む。

【０１３４】

また、背景色が白色で表される関係ラベルが付加された弧からなる構造は、チャンク間の依存構造を表している。図９の構造解析結果によれば、チャンク「昨日」とチャンク「報告し」との間に、関係ラベルadvmodで表される依存構造が存在し、チャンク「予備調査結果」とチャンク「について」との間に、関係ラベルpobjで表される依存構造が存在し、チャンク「について」とチャンク「報告し」との間に、関係ラベルiobjで表される依存構造が存在し、チャンク「た」とチャンク「報告し」との間に、関係ラベルauxで表される依存構造が存在していることを表している。

【０１３５】

このように、第１実施形態に係る言語解析装置１００は、チャンク処理と依存構造解析を分離させることなく、チャンク解析を行いながら、同時に依存構造解析を行うことで入力テキストの構造を解析する。

【０１３６】

まず入力テキストをチャンクに分割し終えてから、分割したチャンク間の依存構造解析を行うという手順を踏んで、チャンク処理と依存構造解析とを分離して入力テキストの構造解析を行う場合には、チャンク処理でのチャンクへの分割精度が依存構造解析にまで影響を及ぼす場合がある。

【０１３７】

しかしながら、第１実施形態に係る言語解析装置１００の場合、チャンク処理と依存構造解析とを同時に行うため、チャンク処理を行う際には依存構造解析の結果を参照し、依存構造解析を行う際にはチャンク処理の結果を参照するというように、互いの処理結果を考慮しながら自身の処理を決定することができる。すなわち、第１実施形態に係る言語解析装置１００は、分割されたチャンクの尤もらしい依存構造を同定しながら、当該依存構造に整合する適切なチャンクへの分割を行うことができるため、入力テキストの構造解析

の精度を向上させることができる。

【０１３８】

＜第２実施形態＞

第１実施形態に係る言語解析装置１００では、素性テンプレートに基づいて、入力テキストから単語の素性及びチャンクの素性を抽出し、単語の素性と、チャンクの素性と、解析モデル３０Ａと、に基づいて、チャンクに統合すると共にチャンクの依存構造を解析するためのアクションを決定した。

【０１３９】

第２実施形態では、チャンクとして単名詞で構成される単名詞句に統合すると共に、当該単名詞句がどのような内容を表す単名詞句であるのかを表す情報（意味クラス）を単名詞句に付与する言語解析装置１００について説明する。

【０１４０】

なお、第２実施形態に係る言語解析装置１００のシステム構成は、図１に示した第１実施形態に係る言語解析装置１００のシステム構成例と同様である。

【０１４１】

入力テキストの構造解析を行う上で、単名詞句の意味クラスを認識することは構文構造解析の精度を向上するための重要な要素技術となる。したがって、第２実施形態に係る言語解析装置１００では、チャンクとして統合された単名詞句のクラスとして、固有表現クラス又は一般名詞の意味クラスを付与する。

【０１４２】

そのため、例えば単名詞句毎に、単名詞句と、単名詞句のクラス（固有表現クラス又は一般名詞の意味クラス）と、を対応付けたクラス対応表を予め記憶部３０に記憶しておく。そして、構造解析部２２は、チャンクとして単名詞句に統合する際に当該単名詞句がクラス対応表に含まれる場合、当該単名詞句に対応するクラスを付与する。

【０１４３】

具体的には、解析モデルによって決定されるアクションが、スタックSの最上位の要素s_0を、クラス対応表から得られるクラスが付与された単名詞句として確定させるアクションを含むように設計すればよい。

【０１４４】

なお、第２実施形態に係る言語解析装置１００の他の構成及び作用については、第１の実施の形態と同様であるため、説明を省略する。

【０１４５】

このように、第２実施形態に係る言語解析装置１００は、チャンクとして単名詞句に統合する場合、単名詞句に固有表現クラス又は一般名詞の意味クラスを付加する。言語解析装置１００は、例えばチャンクに付加したクラスもチャンクの素性として扱ってもよい。この場合、図３に示したアクションの中からチャンクの素性に対応するアクションを精度よく決定することができる。

【０１４６】

また、チャンクに付与したクラスは、例えば言語解析装置１００での構造解析結果を入力として様々な処理を行う別装置で利用することができる。例えば言語解析装置１００での入力テキストの構造解析結果を入力として、入力テキストと類似する文を検索する検索装置では、チャンクにクラスが付与されていない場合と比較して、検索精度を向上させることができる。

【０１４７】

なお、本発明は、上述した各実施形態に限定されるものではなく、この発明の要旨を逸脱しない範囲内で様々な変形や応用が可能である。

【０１４８】

例えば、各実施形態に係る言語解析装置１００は、キューQの短単位列$(q_0,\cdots,q_n)$の先頭から順に短単位を取得し、最尤の解析候補のみをスタックに保持して決定的に解析を行っている。しかし、例えばスタックを複数用意し、複数の解析候補をスタックに保持

しながら、動的計画法等を用いて最尤の解析候補を探索する方法と組み合わせるようにしてもよい。

【０１４９】

また、上述の言語解析装置１００は、内部にコンピュータシステムを有しているが、「コンピュータシステム」は、ＷＷＷシステムを利用している場合であれば、ホームページ提供環境（あるいは表示環境）も含むものとする。

【０１５０】

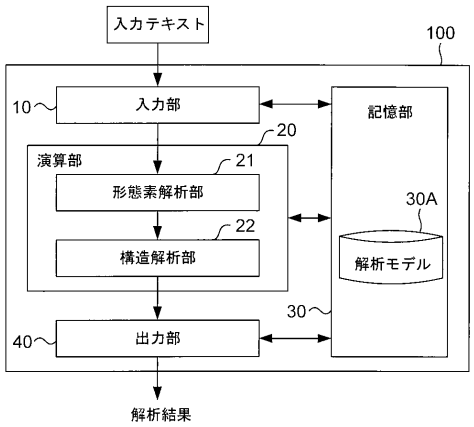
また、本願明細書中では、プログラムが予めＲＯＭにインストールされている実施形態を説明したが、当該プログラムを、コンピュータ読み取り可能な記録媒体に格納して提供することも可能である。

【符号の説明】

【０１５１】

１０・・・入力部
２０・・・演算部
２１・・・形態素解析部
２２・・・構造解析部
３０・・・記憶部
３０Ａ・・・解析モデル
４０・・・出力部
１００・・・言語解析装置
Ｑ・・・キュー
Ｓ・・・スタック

【図１】



【図２】

短単位 ♯s	出現形 w	標準形 m	品詞 t
1	昨日	昨日	ADV
2	予備	予備	NN
3	調査	調査	NN
4	結果	結果	NN
5	に	に	PCS
6	つい	つく	VB
7	て	て	PCJ
8	報告	報告	NN
9	し	する	VB
10	た	た	AUX

【図３】

Shift LUW (POS)	キュー Q の先頭から 1 短単位 q_0 を取り出し，長単位（品詞=POS）を構成する先頭の短単位として，スタック S に push する．
Reduce LUW_L (dep)	スタック S の最上位の 2 つの長単位の部分木 s_0, s_1 の間に長単位間の関係ラベル dep を持つ左向きの弧を張る．
Reduce LUW_R (dep)	スタック S の最上位の 2 つの長単位の部分木 s_0, s_1 の間に長単位間の関係ラベル dep を持つ右向きの弧を張る．
Pop LUW	スタック S の最上位の部分木 s_0 を長単位（品詞は $Shift_{LUW}$ (POS) の POS）として確定させる．
Shift SUW	キュー Q から先頭の 1 短単位 q_0 を取り出し，スタック S に push する．
Reduce SUW_L (dep)	スタック S の最上位の 2 つの短単位の部分木 s_0, s_1 の間に短単位間の関係ラベル dep を持つ左向きの弧を張る．
Reduce SUW_R (dep)	スタック S の最上位の 2 つの短単位の部分木 s_0, s_1 の間に短単位間の関係ラベル dep を持つ右向きの弧を張る．

【図４】

長単位素性	
$s_0.h_{Lw} \circ s_1.h_{Lw}$	$s_0.h_{Lt} \circ s_1.h_{Lw}$
$s_0.h_{Lm} \circ s_1.h_{Lm}$	
$s_0.h_{Lt} \circ s_1.h_{Lt} \circ s_0.l_{Lt}$	$s_0.h_{Lt} \circ s_1.h_{Lt} \circ s_0.r_{Lt}$
$s_0.h_{Lt} \circ s_1.h_{Lt} \circ s_0.l_{Lw}$	$s_0.h_{Lt} \circ s_1.h_{Lt} \circ s_0.r_{Lw}$
$s_0.h_{Lt} \circ s_1.h_{Lt} \circ s_0.l_{Lm}$	$s_0.h_{Lt} \circ s_1.h_{Lt} \circ s_0.r_{Lm}$
$s_0.h_{Lw} \circ s_1.h_{Lw} \circ s_2.h_{Lw}$	$s_0.h_{Lt} \circ s_1.h_{Lt} \circ s_2.h_{Lt}$
$s_0.h_{Lt} \circ s_1.h_{Lt} \circ q_0.t$	$s_0.h_{Lw} \circ s_1.h_{Lt} \circ q_0.t$
$s_0.h_{Lt} \circ s_1.h_{Lw} \circ q_0.t$	$s_0.h_{Lw} \circ s_1.h_{Lw} \circ q_0.w$
$s_0.h_{Lw} \circ s_1.h_{Lw} \circ q_0.m$	
複合語素性	
$q_0.f_{comp}(3,5) \circ s_0.h_{Lt}$	$q_0.f_{comp}(3,5) \circ s_0.h_{Lt} \circ q_0.t$
$q_0.f_{comp}(3,5) \circ s_0.h_{Lt} \circ q_0.t \circ q_1.t$	$q_0.f_{comp}(3,5) \circ s_0.h_{Lw}$
$q_0.f_{comp}(3,5) \circ s_0.h_{Lw} \circ q_0.w$	$q_0.f_{comp}(3,5) \circ s_0.h_{Lw} \circ q_0.w \circ q_1.w$

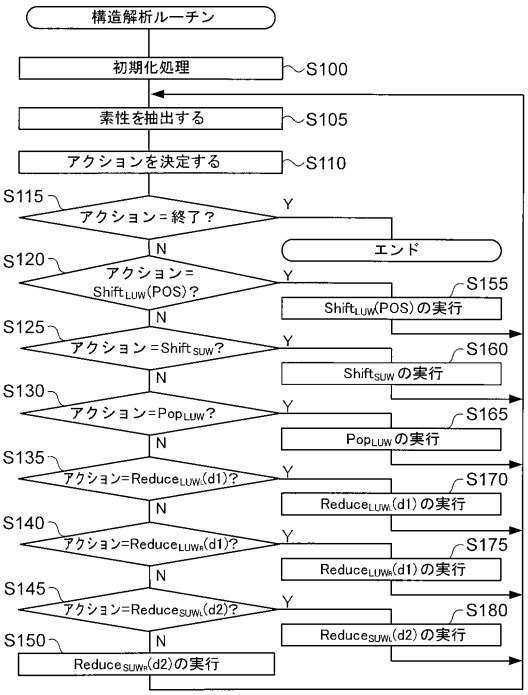
【図５】

長単位素性	
$s_n.h_{Lw}$	s_n の主辞（部分木の根ノード）の長単位出現形
$s_n.h_{Lm}$	s_n の主辞（部分木の根ノード）の長単位標準形
$s_n.h_{Lt}$	s_n の主辞（部分木の根ノード）の長単位品詞
$s_n.l_{Lw}$	s_n の最左の子ノードの長単位出現形
$s_n.l_{Lm}$	s_n の最左の子ノードの長単位標準形
$s_n.l_{Lt}$	s_n の最左の子ノードの長単位品詞
$s_n.r_{Lw}$	s_n の最右の子ノードの長単位出現形
$s_n.r_{Lm}$	s_n の最右の子ノードの長単位標準形
$s_n.r_{Lt}$	s_n の最右の子ノードの長単位品詞
短単位素性	
$s_n.h_w$	s_n の主辞（部分木の根ノード）の短単位出現形
$s_n.h_m$	s_n の主辞（部分木の根ノード）の短単位標準形
$s_n.h_t$	s_n の主辞（部分木の根ノード）の短単位品詞
$q_n.w$	q_n の短単位出現形
$q_n.m$	q_n の短単位標準形
$q_n.t$	q_n の短単位品詞
複合語素性	
$q_n.f_{comp}(c_0, c_m)$	q_n を先頭とする短単位列と、複合語辞書に含まれる複合語エントリと、の照合フラグ 照合 OK:1,NG:0（照合対象の短単位数，最小 c_0，最大 c_m）

【図６】

複合語 / 複合語品詞	短単位数	短単位列
...	...	...
について/PCS	3	に/PCS/に, つい/VB/つく, て/PCJ/て
に関して/PCS	3	に/PCS/に, 関し/VB/関する, て/PCJ/て
かもしれない/AUX	5	か/P/ か, も/PBD/も, しれ/VB/しる ない/AUX/ない
だが/CONJ	2	だ/AUX/だ, が/PCJ/が
...	...	...

【図７】



【図８】

Step	Action	Stack s_3	s_2	s_1	s_0	Queue q_0	q_1	q_2
0	--				Root	昨日	予備	調査
1	Shift $_{LUW}$ (ADV)			Root	昨日	予備	調査	結果
2	Pop $_{LUW}$			Root	昨日/ADV	予備	調査	結果
3	Shift $_{LUW}$ (NN)		Root	昨日/ADV	予備	調査	結果	に
4	Shift $_{SUW}$	Root	昨日/ADV	予備	調査	結果	に	つい
5	Reduce $_{SUW_L}$ (cn)		Root	昨日/ADV	予備 $\overset{cn}{-}$ 調査	結果	に	つい
6	Shift $_{SUW}$	Root	昨日/ADV	予備 $\overset{cn}{-}$ 調査	結果	に	つい	て
7	Reduce $_{SUW_L}$ (cn)		Root	昨日/ADV	予備 $\overset{cn}{-}$ 調査 $\overset{cn}{-}$ 結果	に	つい	て
8	Pop $_{LUW}$		Root	昨日/ADV	予備 $\overset{cn}{-}$ 調査 $\overset{cn}{-}$ 結果/NN	に	つい	て
9	Shift $_{LUW}$ (PCS)	Root	昨日/ADV	予備 $\overset{cn}{-}$ 調査 $\overset{cn}{-}$ 結果/NN	に	つい	て	報告
10	Shift $_{SUW}$	昨日/ADV	予備 $\overset{cn}{-}$ 調査 $\overset{cn}{-}$ 結果/NN	に	つい	て	報告	し
11	Reduce $_{SUW_L}$ (mwe)	Root	昨日/ADV	予備 $\overset{cn}{-}$ 調査 $\overset{cn}{-}$ 結果/NN	に $\overset{mwe}{-}$ つい	て	報告	し
12	Shift $_{SUW}$	昨日/ADV	予備 $\overset{cn}{-}$ 調査 $\overset{cn}{-}$ 結果/NN	に $\overset{mwe}{-}$ つい	て	報告	し	た
13	Reduce $_{SUW_L}$ (mwe)	Root	昨日/ADV	予備 $\overset{cn}{-}$ 調査 $\overset{cn}{-}$ 結果/NN	に $\overset{mwe}{-}$ つい $\overset{mwe}{-}$ て	報告	し	た
14	Pop $_{LUW}$	Root	昨日/ADV	予備 $\overset{cn}{-}$ 調査 $\overset{cn}{-}$ 結果/NN	に $\overset{mwe}{-}$ つい $\overset{mwe}{-}$ て/PCS	報告	し	た
15	Reduce $_{LUW_L}$ (pobj)		Root	昨日/ADV	に $\overset{mwe}{-}$ つい $\overset{mwe}{-}$ て/PCS	報告	し	た
					予備 $\overset{cn}{-}$ 調査 $\overset{cn}{-}$ 結果/NN $\overset{pobj}{\swarrow}$			
16	Shift $_{LUW}$ (VB)	Root	昨日/ADV	に $\overset{mwe}{-}$ つい $\overset{mwe}{-}$ て/PCS	報告	し	た	-
			予備 $\overset{cn}{-}$ 調査 $\overset{cn}{-}$ 結果/NN $\overset{pobj}{\swarrow}$					
17	Shift $_{SUW}$	昨日/ADV に $\overset{mwe}{-}$ つい $\overset{mwe}{-}$ て/PCS		報告	し	た	-	-
		予備 $\overset{cn}{-}$ 調査 $\overset{cn}{-}$ 結果/NN $\overset{pobj}{\swarrow}$						
18	Reduce $_{SUW_L}$ (nv)	Root	昨日/ADV	に $\overset{mwe}{-}$ つい $\overset{mwe}{-}$ て/PCS	報告 $\overset{nv}{-}$ し/VB	た	-	-
			予備 $\overset{cn}{-}$ 調査 $\overset{cn}{-}$ 結果/NN $\overset{pobj}{\swarrow}$					
19	Reduce $_{LUW_L}$ (iobj)		Root	昨日/ADV	報告 $\overset{nv}{-}$ し/VB	た	-	-
					に $\overset{mwe}{-}$ つい $\overset{mwe}{-}$ て/PCS $\overset{iobj}{\swarrow}$			
					予備 $\overset{cn}{-}$ 調査 $\overset{cn}{-}$ 結果/NN $\overset{pobj}{\swarrow}$			
20	Shift $_{LUW}$ (AUX)	Root	昨日/ADV	報告 $\overset{nv}{-}$ し/VB	た	-	-	-
				に $\overset{mwe}{-}$ つい $\overset{mwe}{-}$ て/PCS $\overset{iobj}{\swarrow}$				
				予備 $\overset{cn}{-}$ 調査 $\overset{cn}{-}$ 結果/NN $\overset{pobj}{\swarrow}$				
21	Pop $_{LUW}$	Root	昨日/ADV	報告 $\overset{nv}{-}$ し/VB	た/AUX	-	-	-
				に $\overset{mwe}{-}$ つい $\overset{mwe}{-}$ て/PCS $\overset{iobj}{\swarrow}$				
				予備 $\overset{cn}{-}$ 調査 $\overset{cn}{-}$ 結果/NN $\overset{pobj}{\swarrow}$				
22	Reduce $_{LUW_R}$ (aux)		Root	昨日/ADV	報告 $\overset{nv}{-}$ し/VB	-	-	-
					に $\overset{mwe}{-}$ つい $\overset{mwe}{-}$ て/PCS $\overset{iobj}{\swarrow}$ $\overset{aux}{\searrow}$ た			
					予備 $\overset{cn}{-}$ 調査 $\overset{cn}{-}$ 結果/NN $\overset{pobj}{\swarrow}$			
23	Reduce $_{LUW_L}$ (advmod)			Root	報告 $\overset{nv}{-}$ し/VB	-	-	-
				昨日/ADV $\overset{advmod}{\swarrow}$	に $\overset{mwe}{-}$ つい $\overset{mwe}{-}$ て/PCS $\overset{iobj}{\swarrow}$ $\overset{aux}{\searrow}$ た			
					予備 $\overset{cn}{-}$ 調査 $\overset{cn}{-}$ 結果/NN $\overset{pobj}{\swarrow}$			
24	Reduce $_{LUW_L}$ (root)				Root	-	-	-
					報告 $\overset{nv}{-}$ し/VB $\overset{root}{\swarrow}$	-	-	-
				昨日/ADV $\overset{advmod}{\swarrow}$	に $\overset{mwe}{-}$ つい $\overset{mwe}{-}$ て/PCS $\overset{iobj}{\swarrow}$ $\overset{aux}{\searrow}$ た			
					予備 $\overset{cn}{-}$ 調査 $\overset{cn}{-}$ 結果/NN $\overset{pobj}{\swarrow}$			

【図９】

