

(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2004-355204

(P2004-355204A)

(43) 公開日 平成16年12月16日(2004.12.16)

(51) Int. Cl.⁷

G06F 17/27
G06N 5/04

F I

G06F 17/27 J
G06N 5/04 550N

テーマコード(参考)

5B091

審査請求 有 請求項の数 19 O L (全 21 頁)

(21) 出願番号 特願2003-150598(P2003-150598)
(22) 出願日 平成15年5月28日(2003.5.28)

(出願人による申告) 国等の委託研究の成果に係る特許出願(平成14年度通信・放送機構「多言語標準文書処理システムの研究開発」委託研究、産業活力再生特別措置法第30条の適用を受けるもの)

(71) 出願人 000000295
沖電気工業株式会社
東京都港区虎ノ門1丁目7番12号
(74) 代理人 100090620
弁理士 工藤 宣幸
(72) 発明者 北村 美穂子
東京都港区虎ノ門1丁目7番12号 沖電気工業株式会社内
Fターム(参考) 5B091 AA15 CA02 CA05 CC03 EA17 EA25

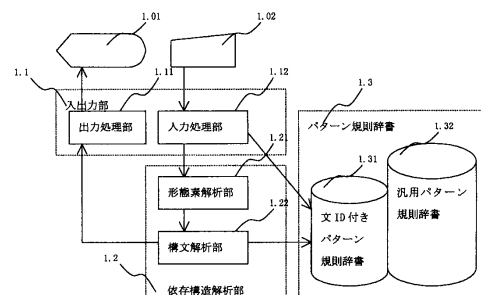
(54) 【発明の名称】 自然言語処理装置、自然言語処理方法及び自然言語処理プログラム

(57) 【要約】

【課題】 パターン規則を利用する構文解析での解析結果の精度を高めることができる自然言語処理装置、自然言語処理方法及び自然言語処理プログラムを提供する。

【解決手段】 本発明の自然言語処理装置、自然言語処理方法及び自然言語処理プログラムにおいては、予め、同一文に同時に適用する可能性の高さを示す文IDが付与された文ID付パターン規則を用意しておく。そして、入力文を、用意されているパターン規則を参照しながら構文解析する際に、その解析結果に、同一の文IDが付与されたパターン規則が多くなるようにする。

【選択図】 図1



【特許請求の範囲】**【請求項 1】**

少なくともパターン名及びパターン構成要素を有するパターン規則を利用して、少なくとも入力文の構文解析結果を得る処理を伴う自然言語処理装置において、同一文に同時に適用する可能性の高さを示す文 ID が付与されたパターン規則を格納している文 ID 付パターン規則辞書と、解析対象の入力文を形態素解析する形態素解析手段と、形態素解析結果に対し、上記文 ID 付パターン規則辞書を参照しながら、複数のパターン規則の木構造でなる構文解析結果を得るものであって、同一の文 ID が付与されたパターン規則が多くなる、パターン規則間の木構造を採用する構文解析手段とを有することを特徴とする自然言語処理装置。

10

【請求項 2】

上記構文解析手段は、上記文 ID を考慮しないで複数のパターン規則の木構造でなる構文解析結果を得た後、その結果に含まれているパターン規則に付与されている文 ID に基づき、同一の文 ID が付与されたパターン規則が多くなるように、複数候補の解消を行って最終的な構文解析結果を得ることを特徴とする請求項 1 に記載の自然言語処理装置。

【請求項 3】

上記構文解析手段は、同一の文 ID が付与されたパターン規則の数を計数する際に、選言的な構造内に同一パターン規則が複数存在する場合には、それらをまとめて 1 つとして数えることを特徴とする請求項 2 に記載の自然言語処理装置。

20

【請求項 4】

上記構文解析手段は、上記形態素解析結果における各形態素で上記文 ID 付パターン規則辞書を参照して木構造の下位のパターン規則を定め、上記各下位パターン規則に対する上位パターン規則の探索を、当該下位パターン規則に付与されている文 ID と同一のものを優先させて行いながら、最終的な構文解析結果を得ることを特徴とする請求項 1 に記載の自然言語処理装置。

【請求項 5】

上記構文解析手段は、上記形態素解析結果における各形態素で上記文 ID 付パターン規則辞書を参照して木構造の下位のパターン規則を定め、定められた複数の下位パターン規則に多く付与されている文 ID を検出し、上記各下位パターン規則に対する上位パターン規則の探索を、検出された文 ID と同一のものを優先させて行いながら、最終的な構文解析結果を得ることを特徴とする請求項 1 に記載の自然言語処理装置。

30

【請求項 6】

パターン規則の検索対象の辞書として、上記文 ID 付パターン規則辞書に加え、文 ID が付与されていない汎用的なパターン規則を格納している汎用パターン規則辞書を有することを特徴とする請求項 1 ~ 5 のいずれかに記載の自然言語処理装置。

【請求項 7】

上記文 ID 付パターン規則辞書は、文 ID 付パターン規則を追加して登録できるものであることを特徴とする請求項 1 ~ 6 のいずれかに記載の自然言語処理装置。

【請求項 8】

上記文 ID 付パターン規則辞書として、文書や分野などに応じて区別される複数の文 ID 付パターン規則辞書を有することを特徴とする請求項 1 ~ 7 のいずれかに記載の自然言語処理装置。

40

【請求項 9】

当該自然言語処理装置が機械翻訳装置であり、上記構文解析手段が、原言語文に対する構文解析を行うものであることを特徴とする請求項 1 ~ 8 のいずれかに記載の自然言語処理装置。

【請求項 10】

少なくともパターン名及びパターン構成要素を有するパターン規則を利用して、少なくとも入力文の構文解析結果を得る処理を伴う自然言語処理方法において、

50

同一文に同時に適用する可能性の高さを示す文IDが付与されたパターン規則を格納している文ID付パターン規則辞書を予め用意しておくと共に、
解析対象の入力文を形態素解析する形態素解析工程と、
形態素解析結果に対し、上記文ID付パターン規則辞書を参照しながら、複数のパターン規則の木構造でなる構文解析結果を得るものであって、同一の文IDが付与されたパターン規則が多くなる、パターン規則間の木構造を採用する構文解析工程と
を有することを特徴とする自然言語処理方法。

【請求項11】

上記構文解析工程は、上記文IDを考慮しないで複数のパターン規則の木構造でなる構文解析結果を得た後、その結果に含まれているパターン規則に付与されている文IDに基づき、同一の文IDが付与されたパターン規則が多くなるように、複数候補の解消を行って最終的な構文解析結果を得ることを特徴とする請求項10に記載の自然言語処理方法。

10

【請求項12】

上記構文解析工程は、同一の文IDが付与されたパターン規則の数を計数する際に、選言的な構造内に同一パターン規則が複数存在する場合には、それらをまとめて1つとして数えることを特徴とする請求項11に記載の自然言語処理方法。

【請求項13】

上記構文解析工程は、上記形態素解析結果における各形態素で上記文ID付パターン規則辞書を参照して木構造の下位のパターン規則を定め、上記各下位パターン規則に対する上位パターン規則の探索を、当該下位パターン規則に付与されている文IDと同一のものを優先させて行いながら、最終的な構文解析結果を得ることを特徴とする請求項10に記載の自然言語処理方法。

20

【請求項14】

上記構文解析工程は、上記形態素解析結果における各形態素で上記文ID付パターン規則辞書を参照して木構造の下位のパターン規則を定め、定められた複数の下位パターン規則に多く付与されている文IDを検出し、上記各下位パターン規則に対する上位パターン規則の探索を、検出された文IDと同一のものを優先させて行いながら、最終的な構文解析結果を得ることを特徴とする請求項10に記載の自然言語処理方法。

【請求項15】

パターン規則の検索対象の辞書として、上記文ID付パターン規則辞書に加え、文IDが付与されていない汎用的なパターン規則を格納している汎用パターン規則辞書も予め用意されていることを特徴とする請求項10～14のいずれかに記載の自然言語処理方法。

30

【請求項16】

上記文ID付パターン規則辞書は、文ID付パターン規則を追加して登録できるものであることを特徴とする請求項10～15のいずれかに記載の自然言語処理方法。

【請求項17】

上記文ID付パターン規則辞書として、文書や分野などに応じて区別される複数の文ID付パターン規則辞書が容易されていることを特徴とする請求項10～16のいずれかに記載の自然言語処理方法。

【請求項18】

当該自然言語処理方法が機械翻訳方法であり、上記構文解析工程が、原言語文に対する構文解析を行うことを特徴とする請求項10～17のいずれかに記載の自然言語処理方法。

40

【請求項19】

請求項10～18のいずれかの自然言語処理方法を、コンピュータが実行可能なコードで記述していることを特徴とする自然言語処理プログラム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、自然言語処理装置、自然言語処理方法及び自然言語処理プログラムに関し、例えば、事例ベース（過去の解析結果や翻訳結果を利用した）の構文解析処理や翻訳処理に

50

適用し得るものである。

【0002】

【従来の技術】

【0003】

【非特許文献1】山田 寛康、松本 裕治共著、「Support Vector Machineを用いた決定性上昇型構文解析」、研究報告「自然言語処理」、No. 149 - 009、2002年5月23日

【0004】

【特許文献1】特開平7 - 295991号公報

【0005】

【特許文献2】特開2002 - 41512号公報

機械翻訳等で利用される自然言語の構文解析技術の進歩は目覚ましい。従来型の構文解析は、構文的な情報を含んだ辞書と文法規則を予め人間が作成しておき、それをチャート法やア－リー法などのパーザを利用することにより、文の解析結果を求めていた。しかし、最近では、大量の文書の構文解析結果があれば、その構文解析結果（学習データ）を再現するための規則を自動的に作成し、以降は、その規則に基づいて構文解析結果を求める機械学習の手法を利用した構文解析システムの研究が進んでいる（非特許文献1）。

【0006】

また、大量の文書の構文解析結果を蓄積しておき、入力された文の構文解析結果と蓄積された構文解析結果とを比較し、その比較結果から正しい解析結果を求めるような方法も提案されている（特許文献1）。

【0007】

上述した過去の事例を利用する技術は、辞書や文法の手による作成が不要であり、また、構文解析の正解結果を多く用意すればするほど解析精度が良くなるという利点がある。

【0008】

さらに、事例を利用した技術は、検索や翻訳等の自然言語技術に応用しやすいという利点がある。特許文献1の記載方法においては、事例として対訳文書を利用することによって機械翻訳に応用している。この場合は、大量の対訳文書の構文解析結果を蓄積しておき、入力文と同言語の構文解析結果と入力文の構文解析結果とを比較し、最も類似する構文解析結果を選択し、その構文解析結果の相手言語側の構文解析結果を参照することにより、適切な翻訳結果を求めるという手法を採用している。

【0009】

しかし、非特許文献1の方法は、機械学習を利用しているため、予め作成しておく学習データ（規則）は、人間にとって理解不能であり、規則に手を加えることはできない。つまり、解析結果がより良くなるように規則を手で調整することはできない。また、規則が理解不能であるため、どんな解析結果が得られるかの推測が難しい。さらに、正解例が増えた場合には、学習し直し、規則を作り直す必要があるが、規則の再学習には膨大な時間がかかる。

【0010】

一方、特許文献1の方法は、入力文に最も類似する過去の構文解析結果から入力文中に含まれる語彙の用法を知ることにより構文解析支援を行うという提案であり、入力文の構文解析を全自動で行う方法ではない。また、利用する過去の構文解析結果も最も類似する1文のみである。

【0011】

さらに、特許文献1の提案では、比較の手法（照合手段）において、一文ごとに照合するという手法であるため、用例が何万文と大量になった場合、その比較において実用レベルの速度が得られないという課題も有する。

【0012】

上記の課題を解決するために、特許文献2では、既存対訳文書から翻訳パターン規則を作成し、それらを辞書として蓄積しておき、その辞書を利用して構文解析することにより、

10

20

30

40

50

既存文書を模倣した翻訳結果を得ることができる（同様の手法で構文解析処理のみを行うこともできる）方法を提案している。

【0013】

【発明が解決しようとする課題】

特許文献2の提案方法によって、既存対訳文書から作成された翻訳パターン規則は、入力文に応じて、適宜、構文解析結果に含まれるが、作成された翻訳パターン規則は、全て同列に扱われるものであった。

【0014】

このように、作成された翻訳パターン規則が全て同列に扱われるため、複数の構文解析結果候補間の順位付けに、翻訳パターン規則の作成に供した文の情報が反映されず、最適でない構文解析結果候補が最適と判断される恐れがある。

10

【0015】

仮に、翻訳パターン規則の作成に供した文が、構文解析対象の文として入力された場合において、作成された翻訳パターン規則を適用した構文解析結果候補以外の構文解析結果候補が生じて、必ずしも前者を有効とすることができなかつた。

【0016】

そのため、自然言語処理パターンの作成に供した文の情報も、入力文に対する構文解析などの自然言語処理に反映でき、最適な解析結果が得られる自然言語処理装置、自然言語処理方法及び自然言語処理プログラムが望まれている。

【0017】

20

【課題を解決するための手段】

かかる課題を解決するため、第1の本発明は、少なくともパターン名及びパターン構成要素を有するパターン規則を利用して、少なくとも入力文の構文解析結果を得る処理を伴う自然言語処理装置において、同一文に同時に適用する可能性の高さを示す文IDが付与されたパターン規則を格納している文ID付パターン規則辞書と、解析対象の入力文を形態素解析する形態素解析手段と、形態素解析結果に対し、上記文ID付パターン規則辞書を参照しながら、複数のパターン規則の木構造でなる構文解析結果を得るものであって、同一の文IDが付与されたパターン規則が多くなる、パターン規則間の木構造を採用する構文解析手段とを有することを特徴とする。

【0018】

30

また、第2の本発明は、少なくともパターン名及びパターン構成要素を有するパターン規則を利用して、少なくとも入力文の構文解析結果を得る処理を伴う自然言語処理方法において、同一文に同時に適用する可能性の高さを示す文IDが付与されたパターン規則を格納している文ID付パターン規則辞書を予め用意しておくと共に、解析対象の入力文を形態素解析する形態素解析工程と、形態素解析結果に対し、上記文ID付パターン規則辞書を参照しながら、複数のパターン規則の木構造でなる構文解析結果を得るものであって、同一の文IDが付与されたパターン規則が多くなる、パターン規則間の木構造を採用する構文解析工程とを有することを特徴とする。

【0019】

さらに、第3の本発明の自然言語処理プログラムは、第2の本発明の自然言語処理方法を、コンピュータが実行可能なコードで記述していることを特徴とする。

40

【0020】

【発明の実施の形態】

(A) 第1の実施形態

以下、本発明による自然言語処理装置、自然言語処理方法及び自然言語処理プログラムの第1の実施形態を図面を参照しながら説明する。第1の実施形態は、入力文に対する構文解析結果を得るものである。

【0021】

(A-1) 第1の実施形態の構成

図1は、第1の実施形態の自然言語処理装置（構文解析装置）の機能的構成を示すブロッ

50

ク図である。なお、実際上は、例えば、パソコンなどの情報処理装置上に、第1の実施形態の自然言語処理プログラム（固定データを含む）がローディングされて、第1の実施形態の自然言語処理装置が構築されるが（なお、専用装置として構築しても良い）、機能的には、図1に示すように表すことができる。

【0022】

図1において、第1の実施形態の自然言語処理装置は、大きくは、入出力部1.1、依存構造解析部1.2、パターン規則辞書1.3から構成されている。

【0023】

入出力部1.1は、キーボードやファイル読込装置等の入力装置1.02から、入力文を入力したり、入力文の構文解析結果から得られたパターン規則辞書の修正情報を入力したり、文ID付きパターン規則辞書1.31を登録入力したりする入力処理部1.12と、構文解析結果をディスプレイやプリンタやファイル格納装置等の出力装置1.01に出力する出力処理部1.11とから構成されている。

10

【0024】

依存構造解析部1.2は、入力文の構文解析結果を求めるための処理部である。依存構造解析部1.2は、単語区切り及び品詞推定を行う形態素解析部1.21、及び、区切られた単語の依存構造を求める構文解析部1.22から構成されている。

【0025】

パターン規則辞書1.3は、文ID付きパターン規則辞書1.31と汎用パターン規則1.32とから構成されている。

20

【0026】

文ID付きパターン規則辞書1.31は、ユーザが参考にしたい過去の文書の構文解析結果から作成されたパターン規則を格納しており、どの文書中のどの文に由来するかを示すための文の識別情報（以下、文IDと呼ぶ）を持っている（後述する図5参照）。なお、同一の文IDを有する複数のパターン規則は、同一の文をベースに形成されたものである。文ID付きパターン規則辞書1.31に格納されているパターン規則は、例えば、特許文献2に記載の作成方法で作成されたものであり、その際、ユーザによって、又は、当該装置が自動的に文IDを付与したものである。

【0027】

一方、汎用パターン規則辞書1.32は、特定の文に依存しない汎用的なパターン規則（汎用パターン規則）を格納しており、人手によって作成される（後述する図6参照）。なお、汎用パターン規則には、文IDは付与されていない。

30

【0028】

（A-2）第1の実施形態の動作

次に、第1の実施形態の自然言語処理装置の動作（第1の実施形態の自然言語処理方法）を説明する。以下では、適宜、入力文書に“work a 40 hour week”という文が含まれ（図3の5.1参照）、この文の構文解析を行うとして具体的な説明も加える。

【0029】

図2は、第1の実施形態の自然言語処理装置の動作（構文解析処理）を示すフローチャートである。

40

【0030】

まず、ユーザは、キーボード等の入力装置1.02を用いて、入力処理部1.12より入力文を入力する（S31）。入力処理部1.12は、その入力文を形態素解析部1.21に渡す。形態素解析部1.21は、その文を形態素解析し（S32）、形態素解析結果を構文解析部1.22に渡す。次に、構文解析処理部1.22は、形態素解析結果を構文解析する（S33）。なお、ここでの形態素解析処理及び構文解析処理は、以下の通りである。

【0031】

形態素解析部1.21では、文を単語単位に区切り、品詞や変化形の情報を付与する（特

50

許文献2記載のものと同様である)。形態素解析結果は、ルートノードを " N o d e " とした木構造で表現される。複数候補がない形態素の場合には、ルートノードの直下に各形態素の標準形と品詞や変化形などの形態素情報とが付与される。一方、複数候補がある形態素の場合には、orノードの子ノードとして各形態素候補の情報が付与される。図4は、上述した入力文 " w o r k a 4 0 h o u r w e e k " に対する形態素解析結果を示している。なお、形態素解析結果に複数候補が存在する場合には、図4のように全ての候補を求める(符号4.1参照)。なお、図4などにおける " p o s = " は品詞情報を表しており、" n " は名詞、" v " は動詞、" a r t " は冠詞である。

【0032】

構文解析部1.22は、パターン規則辞書1.3に格納されているパターン規則を、形態素解析結果にボトムアップに適用させて、入力文を構成するパターン規則の集合(木構造)を求めることによって構文解析する。これは、上記特許文献2のものと同様である。但し、上記特許文献2のものでは、「パターンの評価処理」を行っているが、第1の実施形態では、後述するように、構文解析結果候補の競合を解消しているため、上記特許文献2のような「パターンの評価処理」は実行しない。

10

【0033】

図5は、文ID付きパターン規則辞書1.31の格納例を示す説明図であり、上述した入力文例に係るパターン規則6.1を示している。パターン規則6.1には、上述したように、文ID6.2が対応付けられている。図6は、汎用パターン規則辞書1.32の格納例を示す説明図であり、上述した入力文例に係るパターン規則7.1を示している。

20

【0034】

両パターン規則6.1及び7.1は、同様な表記方法で表記されており、構文解析では区別することなく、適用される。パターン規則は、[言語名:パターン名 パターン構成要素]からなる。言語名は、そのパターン規則に係る言語名を規定するものであり、図5及び図6では英語(en)を規定している。言語名は、所定言語の構文解析専用であれば省略されていても良い。言語名に続くパターン名は、例えば、VP(動詞句)、NP(名詞句)、N(名詞)等の句構造規則での標識が適用される。パターン構成要素は、単語、変数、又は、単語と変数の2以上の並び、からなる。変数は[任意の数字:パターン名(木構造の下位ノードに対応する)]で記述される。任意の数字部分は、翻訳処理用の対となっている原言語及び目的言語パターン間での対応関係を示すものである(第2の実施形態参照)。構文解析においては、変数に、別のパターンが適用されることにより、パターンは入れ子構造をとることができる(変数が解消される)。また、単語及びパターン名は、意味情報などの詳細な情報(素性情報)を持つことができる。さらに、単語及びパターン名は、詳細情報を変数化して、情報の参照をすることもできる。

30

【0035】

構文解析部1.22は、構文解析が終了していないことを確認しつつ、パターン辞書引き処理、パターン検査処理及びパターン適用処理の3つの処理を繰り返し行うことにより、構文解析結果(候補)の木構造を形成する。

【0036】

パターン辞書引き処理は、形態素解析結果及びそれまでのパターン適用処理の結果から、次に適用の可能性のあるパターン規則をパターン規則辞書1.3から引く処理である。パターン検査処理は、辞書引きの結果得られたパターン規則が現在構築中の木構造に適合するか否かを、各木構造毎に検査する処理である。パターン適用処理は、検査の結果、適合すると判定された木構造とパターン規則とに基づいて、木構造にそのパターン規則を実際に適用する処理である。

40

【0037】

図7は、図4に示した形態素解析結果に対し、図5及び図6に示したようなパターン規則を適用して得られた構文解析結果(候補)を示すものである。多くの場合、構文解析結果は一意に定まらず、複数の候補を含むものとなる。図7の例では、" o r " ノード9.1

50

、 9 . 2 によって複数の構文解析結果候補を有している。ここで、図 7 に示すような構文解析結果（候補）において、適用されたパターン規則が文 I D 付きのパターン規則であれば、その文 I D も、木構造の該当するノードの情報として付与される。また、文 I D 付きパターン規則辞書 1 . 3 1 と汎用パターン規則辞書 1 . 3 2 とに、付与されている文 I D を除けば同じパターン規則が格納されている場合には、文 I D 付きパターン規則辞書 1 . 3 1 に格納されているものを優先する。

【 0 0 3 8 】

構文解析結果（候補）を得た後は、文 I D を利用した複数候補の解消（1 個への絞り込み）を行う（S 3 4 ~ S 3 6 ）。

【 0 0 3 9 】

まず、図 7 に示すような構文解析結果（候補）の木構造全体から、解析結果を構成しているパターン規則の文 I D の個数を、例えば、構文解析部 1 . 2 2 が内蔵する図 8 に示すような文 I D 数え上げ表（バッファメモリの一種）を利用して数え上げる（S 3 4 ）。

【 0 0 4 0 】

なお、図 9 に示すような、“ o r ” ノード直下の 2 個のパターン規則が同じパターン名かつ同じ文番号の場合には（言い換えると、選言的な解析結果にまたがって存在する複数の文 I D の場合には）、1 つとして数えることにより、数え上げの重複を避ける。

【 0 0 4 1 】

図 7 の構文解析結果（候補）の場合、文 I D が “ 1 2 0 ” をもつパターン規則はア、イ、ウ、工、オの 5 つであるので、図 8 の文 I D 数え上げ表の “ 1 2 0 ” の結果の欄には「 5 」がセットされ、一方、文 I D が “ 9 2 ” をもつパターン規則はカ、キの 2 つなので、“ 9 2 ” の結果には「 2 」がセットされる。なお、図 9 の “ < - > ” は汎用パターン規則であるため文 I D を持たないことを表している。従って、数え上げの対象から外されている。以上のようにして、図 8 の文 I D 数え上げ表の結果を得る。

【 0 0 4 2 】

次に、その表中で最も数え上げ数の多い文 I D を選択し、その文 I D のパターン規則を最も多く有する構文解析結果候補を（最終的な）構文解析結果として選択する（S 3 5 ）。図 8 の例では、文 I D が “ 1 2 0 ” の数え上げ数が最大であるので、図 7 の構文解析結果候補（解析木）の中からア - オのパターン規則を有する構文解析結果候補が選択される。

【 0 0 4 3 】

その後、選択された構文解析結果の中に、複数候補（選言的な部分）があるか否かを判別し、選択された構文解析結果の中に複数候補（選言的な部分）がなければ、一連の解消処理を終了する（S 3 6 ）。

【 0 0 4 4 】

図 7 の例では、ア - オのパターン規則を有するが構文解析結果候補が選択された段階で、複数候補がなくなるため、解消処理を終了する。

【 0 0 4 5 】

一方、ステップ S 3 5 の処理により、選択された構文解析結果の中に、また複数候補が存在すれば、先に決定したパターン規則の文 I D を除き、再度、文 I D の数え上げ処理を行い（S 3 4 ）、複数候補の解消処理を繰り返す（S 3 5 ）。例えば、“ o r ” ノードが多段に存在するような場合においては、ステップ S 3 4 ~ S 3 6 でなる処理ループが繰り返されることも生じる。

【 0 0 4 6 】

そして、全ての候補が確定して複数候補が解消すれば（S 3 6 ）、依存構造解析部 1 . 2 は、構文解析結果を出力処理部 1 . 1 1 に渡し、C R T ディスプレイなどの出力装置 1 . 0 1 から出力させ（S 3 7 ）、構文解析処理を終了する。

【 0 0 4 7 】

図 1 0 は、図 7 の構文解析結果候補に対し、複数候補の解消処理を行って得た最終的な構文解析結果を示している。

【 0 0 4 8 】

10

20

30

40

50

なお、ステップ S 3 3 での構文解析処理による構文解析結果において、文 I D 付きのパターン規則が適用されず、全て汎用パターン規則であって、複数候補が存在する場合には、他の複数候補の解消処理を行う。例えば、特許文献 2 に記載のものを適用することができる。また、文 I D の数え上げによって、数え上げ数が最大となった文 I D が複数生じた場合にも、例えば、特許文献 2 に記載の複数候補の解消処理を適用することができる。

【 0 0 4 9 】

(A - 3) 第 1 の実施形態の効果

上記第 1 の実施形態によれば、以下の効果を奏することができる。

【 0 0 5 0 】

正解の構文解析結果が得られた後でそれに基づいて作成された文 I D 付きパターン規則を利用して、構文解析の精度を向上させることができる。すなわち、文 I D に基づいて、同一の文の解析結果から得られた複数のパターン規則を新たな文の解析結果に含めることができ、構文解析の精度を向上させることができる。

【 0 0 5 1 】

例えば、図 3 における文 “ w o r k a 4 0 h o u r w e e k ” の前にある同種の文 “ w o r k a 5 d a y w e e k ” の解析結果が提示されたときに、ユーザがその解析結果に満足せず、パターン規則 (文 I D 付きパターン規則) を作成したとすると、文 “ w o r k a 4 0 h o u r w e e k ” の構文解析では、“ w o r k a 5 d a y w e e k ” の解析結果が反映された文 I D 付きパターン規則が適用されて、“ w o r k a 4 0 h o u r w e e k ” の構文解析結果として良好なものが得られる。

【 0 0 5 2 】

また、上述したステップ S 3 4 ~ S 3 6 でなる処理ループの繰り返し処理により、複数の文 I D を有するパターン規則を適用することもでき、過去の解析結果を反映させる場合において、過去の 2 文以上の解析結果を、今回の入力文に対する解析結果に反映させることができる。

【 0 0 5 3 】

さらに、過去の事例から作成された文 I D 付パターン規則と、当初より人手によって作成された汎用パターン規則の両方を利用しているため、適用できる事例が少ない場合でも構文解析処理を実行することができる。

【 0 0 5 4 】

(B) 第 2 の実施形態

次に、本発明による自然言語処理装置、自然言語処理方法及び自然言語処理プログラムの第 2 の実施形態を図面を参照しながら説明する。第 2 の実施形態は、入力文 (原言語文) を他の言語文 (目的言語文) に変換する機械翻訳に、第 1 の実施形態と同様な技術思想を適用したものである。

【 0 0 5 5 】

(B - 1) 第 2 の実施形態の構成

図 1 1 は、第 2 の実施形態の自然言語処理装置 (機械翻訳装置) の機能的構成を示すブロック図である。なお、実際には、例えば、パソコンなどの情報処理装置上に、第 2 の実施形態の自然言語処理プログラム (固定データを含む) がローディングされて、第 2 の実施形態の自然言語処理装置が構築されるが (なお、専用装置として構築しても良い) 、機能的には、図 1 1 に示すように表すことができる。

【 0 0 5 6 】

図 1 1 において、第 2 の実施形態の自然言語処理装置は、大きくは、入出力部 2 . 1、翻訳処理部 2 . 2、及び、翻訳パターン規則辞書 2 . 3 から構成されている。

【 0 0 5 7 】

入出力部 2 . 1 や翻訳パターン規則辞書 2 . 3 は、第 1 の実施形態のものと同様である。なお、第 2 の実施形態の翻訳パターン規則辞書 2 . 3 は、第 1 の実施形態のパターン規則辞書に準じているが、格納されている規則は、二言語対からなるパターン規則 (翻訳パターン規則) となっている。図 1 3 は、翻訳パターン規則辞書 2 . 3 における文 I D 付

翻訳パターン規則 2.3.1 の格納例を示し、図 1.4 は、翻訳パターン規則辞書 2.3 における汎用翻訳パターン規則 2.3.2 の格納例を示している。文 ID 付翻訳パターン規則 2.3.1 では、二言語対からなる各対の翻訳パターン規則に文 ID が付与されている。

【0058】

翻訳処理部 2.2 は、形態素解析部 2.2.1、構文解析・生成部 2.2.2、及び、形態素生成部 2.2.3 から構成されている。

【0059】

形態素解析部 2.2.1 は、第 1 の実施形態のものと同様なものである。構文解析・生成部 2.2.2 における構文解析機能は、第 1 の実施形態の構文解析部の機能と同様である。構文解析・生成部 2.2.2 における構文生成機能は、対となっている目的言語のパターン規則に基づいた生成処理を行う機能である。形態素生成部 2.2.3 は、目的言語の各単語の変化形や活用形の整形を行うものである。なお、翻訳処理部 2.2 は、原言語の構文解析結果の複数候補の解消処理を除き、特許文献 2 に記載の翻訳処理部とほぼ同様なものである。

10

【0060】

(B-2) 第 2 の実施形態の動作

次に、第 2 の実施形態の自然言語処理装置の動作（第 2 の実施形態の自然言語処理方法）を説明する。以下では、適宜、入力文書に “work a 40 hour week” という文が含まれ（図 3 の 5.1 参照）、この文を機械翻訳するとして具体的な説明も加える。

20

【0061】

図 1.2 は、第 2 の実施形態の自然言語処理装置の動作（機械翻訳処理）を示すフローチャートである。

【0062】

第 2 の実施形態における入力処理（S121）及び形態素解析処理（S122）は、第 1 の実施形態と同様であるので、その詳細説明は省略する。

【0063】

構文解析処理（S123）も、第 1 の実施形態とほぼ同様であるが、以下の点が異なっている。まず、構文解析処理に利用するパターン規則が、図 1.3 及び図 1.4 に示したような英語パターン規則と日本語パターン規則の対の翻訳パターン規則であるという点である。入力文を原言語側のパターン規則で構文解析することにより、目的言語（翻訳側）の構文解析結果も同時に得る（特許文献 2 参照）。上述した入力文の形態素解析結果（図 4）を図 1.3 及び図 1.4 に示す翻訳パターン規則で構文解析した結果を図 1.5 に示している。図 1.5 と第 1 の実施形態における図 7 の違いは、構文に関する複数候補に加え、符号 1.5.1 に示すような訳語に関する複数候補も現われる点である。すなわち、ステップ S123 での構文解析処理では、原言語側でのパターン規則が同じであっても、訳語でのパターン規則が異なる場合には、訳語でのパターン規則が異なることを明らかにして、構文解析木に、原言語でのパターン規則をそれぞれ含める。

30

【0064】

しかし、構文の複数候補も訳語の複数候補も、第 1 の実施形態と同様に、文 ID の数え上げ表を用いることにより解消する。

40

【0065】

形態素解析結果に対する構文解析処理が終了すると、次に、文 ID の数え上げ処理を行う（S124）。図 1.5 に示す構文解析結果に対しては、図 1.6 に示すような文 ID の数え上げ表が作成される。文 ID が “120” の結果が 5 つとなって最大であるので、文 ID “120” の翻訳パターン規則が採用され（S125）、この結果、図 1.7 に示すような文 ID “120” の翻訳パターン規則を最も含む構文解析結果候補を得る。

【0066】

図 1.7 には複数候補が存在しないので（S126）、次の処理に進む。なお、第 1 の実施形態と同様に、ステップ S124 から S126 となる処理ループが、複数候補がなくなる

50

まで繰り返し実行される。

【0067】

ステップS124からS126でなる処理ループの繰り返し処理により、複数候補が存在しなくなると、原言語の構文解析結果が得られると同時に、図18に示すような目的言語の構文解析結果も得られる。図12では、構文生成処理を別個のステップで記載しているが、原言語の構文解析結果を得る処理とほぼ並行して、目的言語の構文解析結果を生成する構文生成処理を実行する(S127)。

【0068】

構文生成処理では、翻訳パターン規則辞書2.3を参照し、原言語(英語)のパターンと対をなしている目的言語(日本語)のパターンを利用して、構文解析結果に対応する日本語の木構造を得る(特許文献2参照)。なお、翻訳パターンは、原言語パターンと翻訳パターンとの対になっており、その対応が一意であるので、実際上は、構文解析処理と構文生成処理とがほぼ並行して実行される。

10

【0069】

次に、目的言語の木構造(構文生成結果)に基づいて形態素生成処理を行い(S128)、最終的な翻訳結果を得て、この翻訳結果を、CRTディスプレイなどの出力装置2.01により出力する(S129)。ここで、形態素生成処理では、構文生成結果を、端末ノードに位置する日本語単語を左から順に並べ、目的言語形態素辞書(図示せず)を用いて、動詞の活用形を整えたりなど、各単語の整形を行う。

【0070】

例えば、“work a 40 hour week”という原文に対して、「週40時間の仕事」という翻訳結果が得られる。

20

【0071】

(B-3)第2の実施形態の効果

第2の実施形態によれば、第1の実施形態と同様な効果に加えて、以下の効果を奏することができる。

【0072】

過去の翻訳事例から作成した文ID付翻訳パターン規則を適用して構文解析すると共に、一旦得られた構文解析結果に対して、文IDを利用した複数候補の解消を行うことにより、構文の複数候補の解消と訳語の複数候補の解消を同時に行うことができる。

30

【0073】

文単位の既存対訳文書の利用ではなく、過去の翻訳文を翻訳パターン規則として部分的に分解して利用することにより、既存対訳文書の利用機会を高めることができる。部分的に分解して利用すると、部分同士の関係情報がないため間違っ組み合わされるという問題が生じるが、文ID情報を利用することにより、組み合わせ時に、過去の翻訳文を再現しようという仕組みが働くため、より適切な組み合わせが選択される。

【0074】

事例に基づいた翻訳方式として一般的な用例主導型翻訳では、過去の翻訳例文から最も類似している文を見つけ出し、その差分(異なっている箇所)を抽出して、その差分を機械翻訳し、もとの翻訳例文に差分を置き換えるという、処理工程の多い手法をとるが、この第2の実施形態の手法では、採用された文IDが付与されていない部分が上記の差分に該当し、構文解析処理だけで用例主導型翻訳と類似する結果が得られる。

40

【0075】

(C)第3の実施形態

以下、本発明による自然言語処理装置、自然言語処理方法及び自然言語処理プログラムの第3の実施形態を図面を参照しながら説明する。第3の実施形態は、入力文に対する構文解析結果を得るものである。

【0076】

第3の実施形態の自然言語処理装置(構文解析装置)も、例えば、パソコンなどの情報処理装置上に、第3の実施形態の自然言語処理プログラム(固定データを含む)がローディ

50

ングされて構築されるが（なお、専用装置として構築しても良い）、機能的には、第1の実施形態に係る図1で表すことができる。

【0077】

第3の実施形態の自然言語処理装置は、第1の実施形態と比較すると、構文解析部1.22での処理が異なっている。

【0078】

上述した第1の実施形態では、図7に示すような構文解析結果（構文解析木）を得る際には文IDを利用せず、構文解析木での複数候補の解消時に、文IDを利用するものであったが、この第3の実施形態は、構文解析木を形成していく処理中でも文IDを利用して、構文解析を高速に実行できること、構文解析木が得られた際に、できるだけ複数候補が生じないようにすることを達成しようとしたものである。

10

【0079】

第3の実施形態は、ボトムアップ的な手法では、下位パターン規則の条件を満たす上位パターン規則の適用によって、構文解析木が構築されていくが、新しいパターン規則を適用する際に、そのパターン規則が持つ文IDと同じ文IDを持つパターン規則（上位パターン規則）を優先的に選択するように解析することにより、適用するパターン規則の探索空間を狭くして、高速化と複数候補の排除とを達成しようとしたものである。

【0080】

図19は、第3の実施形態における構文解析処理（図2のS33～S36に対応）を示すフローチャートである。なお、図19は、文IDの利用面に重点を置いて、処理の流れを示している。また、図19でのバッファ1及びバッファ2は、構文解析部1.22が内蔵するものである。

20

【0081】

まず、形態素解析結果から未処理の1個の形態素を選択し（S191）、パターン規則辞書1.3から、その形態素に適用するパターン規則を検索し、検索結果をバッファ1に格納する（S192）。かかる処理を、形態素解析結果の全ての形態素に対して繰り返す（S193）。ここでも、文ID付きパターン規則辞書1.31と汎用パターン規則辞書1.32とに、付与されている文IDを除けば同じパターン規則が格納されている場合には、文ID付きパターン規則辞書1.31に格納されているものを優先してバッファ1に格納する。

30

【0082】

例えば、図4における形態素「work, pos = n」、「work, pos = v」、...毎にステップS191～S193の処理が繰り返される。形態素「work, pos = n」については、図5の符号6.3のパターン規則がバッファ1に格納され、形態素「work, pos = v」については、図6の符号7.2のパターン規則がバッファ1に格納される。

【0083】

全ての形態素に対するパターン規則の検索が終了すると、ステップS194以降の関連するパターン規則（主として上位パターン規則）の検索に移行する。

【0084】

関連パターン規則の検索ではまず、バッファ1内の1個の未処理のパターン規則を処理対象とし、そのパターン規則が有する文IDをバッファ2に格納し（S194）、その未処理のパターン規則の関連パターン規則として、バッファ2に格納されている文IDを有するものの中から検索する（S195）。なお、処理対象の未処理のパターン規則に文IDが付与されていない場合には、文IDのバッファ2への格納は省略され、又は、無意味な値がバッファ2に格納される（S194）。また、ステップS194での処理対象となる未処理のパターン規則には、上述したステップS192で格納されたものだけでなく、後述するステップS197やS198で格納されたものもなり得る。

40

【0085】

例えば、図5の符号6.3が付されたパターン規則が処理対象となったときには、文ID

50

が120を有するパターン規則が検索対象となる。

【0086】

その後、バッファ2に格納されている文IDを有する関連パターン規則が検索できたか否かを判別する(S196)。そして、検索できた場合には、検索された関連パターン規則をバッファ1に追加する(S197)。この追加時には、パターン規則間の上下関係などの関係情報も格納される。一方、該当する文IDを有する関連パターン規則が検索できなかった場合には、その文IDを持たないパターン規則から検索を行い、検索された関連パターン規則をバッファ1に追加する(S198)。なお、この際の検索で関連パターン規則が検索できなかった場合には、そのことを無視して次の処理に移行する。また、ステップS197又はS198で検索結果をバッファ1に格納した際には、処理対象となっているパターン規則以外のバッファ1に格納されているパターン規則の中に、今回、検索された関連パターン規則と連結されて自動的に処理済みになるものも生じる。

10

【0087】

次に、今回検索された関連パターン規則が、終了カテゴリ(S(センテンス)に関するパターン規則)か否かを判別する(S199)。

【0088】

終了カテゴリに達していなければ、バッファ1に関連パターン規則の検索を行っていない未処理のパターン規則が残っているか否かを判別し(S200)、残っていれば、上述したステップS194に戻り、残っていなければ、構文解析失敗として一連の処理を終了する。

20

【0089】

関連パターン規則の検索により、終了カテゴリに達した場合には、第1の実施形態と同様に、構文解析木に含まれている文IDの多少に応じて、複数候補の解消などを行い、構文解析結果を一通りに絞り込んで一連の処理を終了する(S201、S202)。

【0090】

第3の実施形態によれば、第1の実施形態と同様な効果に加え、構文解析木の構築時において、下位パターン規則が持つ文IDと同じ文IDを持つ関連パターン規則(上位パターン規則)を優先的に選択するようにしたことにより、適用するパターン規則の探索空間が狭くなり、構文解析処理の高速化や複数候補の排除とを達成できるという効果をも奏する。

30

【0091】

(D)第4の実施形態

以下、本発明による自然言語処理装置、自然言語処理方法及び自然言語処理プログラムの第4の実施形態を図面を参照しながら説明する。第4の実施形態も、入力文に対する構文解析結果を得るものである。

【0092】

第4の実施形態の自然言語処理装置(構文解析装置)も、例えば、パソコンなどの情報処理装置上に、第4の実施形態の自然言語処理プログラム(固定データを含む)がローディングされて構築されるが(なお、専用装置として構築しても良い)、機能的には、第1の実施形態に係る図1で表すことができる。

40

【0093】

第4の実施形態の自然言語処理装置は、第1の実施形態と比較すると、構文解析部1.2での処理が異なっている。

【0094】

この第4の実施形態も、第3の実施形態と同様に、構文解析木を形成していく処理中でも文IDを利用して、構文解析を高速に実行できること、構文解析木が得られた際に、できるだけ複数候補が生じないようにすることを達成しようとしたものである。

【0095】

パターン規則を用いた構文解析では、ボトムアップ的な手法をとっており、最初は、語彙(形態素)が含まれるパターン規則の適用から始まる。第4の実施形態は、同一の文ID

50

をもつパターン規則を優先的に適用するものであり、語彙が含まれるパターン規則の適用時に、優先すべき文IDを予め決定しておき、それ以降の関連パターン規則（主として上位パターン規則）の検索時に、その文IDを有するパターン規則の適用を優先するという方法である。これは、語彙に関するパターン規則をチェックするだけでも、優先すべき文IDの予想がつくためである。

【0096】

第4の実施形態では、まず、全ての語彙のいずれかを含むパターン規則の適用を最初に決め、最も適用数が多い文IDを選択する（数個でも構わない）。以降は、その選択された文IDを持つパターン規則を優先的に適用させる。予め語彙に関するパターン規則で探索すべき文IDを制限することによって、探索空間を狭めることができるため、高速化が見込めると共に、構文解析木が形成された際に複数候補もほとんど生じさせなくすることができる。

10

【0097】

図20は、第4の実施形態における構文解析処理（図2のS33～S36に対応）を示すフローチャートである。なお、図20は、文IDの利用面に重点を置いて、処理の流れを示している。また、図20でのバッファ1～バッファ3は、構文解析部1.22が内蔵するものである。

【0098】

まず、形態素解析結果の全ての形態素のそれぞれに対し、パターン規則辞書1.3から、その形態素に適用するパターン規則を検索し、検索結果をバッファ1に格納する（S211～S213）。かかる処理は、上述した第3の実施形態と同様である。

20

【0099】

次に、バッファ1に格納されている形態素（語彙）に適用するパターン規則に付与されている文IDを、文ID毎に数え上げ、最も適用数が多い文IDをバッファ2に格納する（S214、S215）。

【0100】

例えば、上述した入力文“work a 40 hour week”の場合であれば、図5の符号6.3や6.4などに係るパターン規則が、形態素（語彙）に適用するパターン規則となり、その文ID“120”のパターン規則が最も多く適用されていることになり、120がバッファ2に格納される。

30

【0101】

文IDのバッファ2への格納が終了すると、ステップS216以降の関連するパターン規則（主として上位パターン規則）の検索に移行する。

【0102】

関連パターン規則の検索ではまず、バッファ1内の1個の未処理のパターン規則を処理対象とし、その未処理のパターン規則の関連パターン規則を、バッファ2に格納されている文IDを有するものの中から検索し、検索できたか否かを判別する（S216、S217）。すなわち、処理対象の未処理のパターン規則に文IDが付与されていない場合や異なる文IDが付与されていても、バッファ2に格納されている文IDを利用した検索が実行される。なお、ステップS216での処理対象となる未処理のパターン規則には、上述したステップS212で格納されたものだけでなく、後述するステップS218やS223で格納されたものもなり得る。

40

【0103】

例えば、バッファ2に格納されている文IDが“120”の場合において、仮に、図5の符号6.5を付したパターン規則（文ID92）や、図6の符号7.3を付したパターン規則が処理対象となった場合でも、ステップS216での検索では、文IDが“120”のパターン規則を探索範囲として実行される。

【0104】

バッファ2に格納されている文IDを有する関連パターン規則が検索できた場合には、検索された関連パターン規則をバッファ1に追加する（S218）。この追加時には、パタ

50

ーン規則間の上下関係などの関係情報も格納される。また、検索結果をバッファ 1 に追加格納した際には、処理対象となっているパターン規則以外のバッファ 1 に格納されているパターン規則の中に、今回、検索された関連パターン規則と連結されて自動的に処理済みになるものも生じる。一方、該当する文 ID を有する関連パターン規則が検索できなかった場合には、検索できなかった旨の情報と共に、処理対象となっているパターン規則をバッファ 3 に格納する (S 2 1 9)。

【 0 1 0 5 】

次に、今回検索された関連パターン規則 (S 2 1 8 による) により、終了カテゴリ (S (センテンス) に関するパターン規則) に達したか否かを判別する (S 2 2 0)。

【 0 1 0 6 】

終了カテゴリに達していなければ、バッファ 1 に関連パターン規則の検索を行っていない未処理のパターン規則が残っているか否かを判別し (S 2 2 1)、残っていれば、上述したステップ S 2 1 6 に戻る。

【 0 1 0 7 】

終了カテゴリに達しておらず、しかも、バッファ 1 に未処理のパターン規則が残っていなければ、バッファ 3 に格納されているパターン規則があるか否かを判別する (S 2 2 2)。この場合において、バッファ 3 に格納されているパターン規則がなければ、構文解析失敗として一連の処理を終了する。

【 0 1 0 8 】

バッファ 3 に格納されているパターン規則があれば、その中の未処理 (S 2 2 3 について未処理) のパターン規則を 1 個取り出し、取り出したパターン規則に関連するパターン規則 (上位パターン規則) を、バッファ 2 に格納されている文 ID を有するパターン規則以外のパターン規則の中から検索し、検索されたパターン規則をバッファ 1 に追加する (S 2 2 3)。なお、この際の実検索で関連パターン規則が検索できなかった場合には、そのことを無視して次の処理 (S 2 2 4) に移行する。

【 0 1 0 9 】

このような処理を、バッファ 3 に格納されている全てのパターン規則について繰り返す (S 2 2 4)。そして、バッファ 3 に格納されている全てのパターン規則について、それぞれ、バッファ 2 に格納されている文 ID に関係しないパターン規則からの検索を終了すると、上述したステップ S 2 2 3 での検索でバッファ 1 にパターン規則が追加されたか否かを判別する (S 2 2 5)。

【 0 1 1 0 】

バッファ 1 に追加されたパターン規則がなければ、構文解析失敗として一連の処理を終了する。一方、バッファ 1 に追加されたパターン規則があれば、バッファ 3 をクリアして、上述したステップ S 2 1 6 に戻る。

【 0 1 1 1 】

上述したようなボトムアップの検索を繰り返し、終了カテゴリに達した場合には、構文解析成功として一連の処理を終了する。

【 0 1 1 2 】

なお、上記では、ステップ S 2 1 5 の処理によってバッファ 2 に格納する文 ID が 1 個の場合を説明したが、形態素 (語彙) に適用するパターン規則の文 ID の多い方の複数個を格納するようにしても良い。この場合にも、バッファ 2 に格納された複数の文 ID のいずれかを有するパターン規則の集合が関連パターン規則 (上位パターン規則) の探索範囲となる。この場合には、終了カテゴリに達し、構文解析成功とした後に、上述した第 3 の実施形態に係る図 1 9 のステップ S 2 0 1 及び S 2 0 2 でなるような、複数候補の解消処理を実行することを要する。

【 0 1 1 3 】

第 4 の実施形態によれば、第 1 の実施形態と同様な効果に加え、構文解析木の構築時において、全ての語彙のいずれかを含むパターン規則の適用を最初に決め、適用数が多い文 ID を選択し、以降は、その選択された文 ID を持つパターン規則を優先的に適用させるよ

10

20

30

40

50

うにしたので、探索空間を狭めることができ、高速化が見込めると共に、構文解析木が形成された際に複数候補もほとんど生じさせなくすることができる。

【0114】

(E)他の実施形態

上記各実施形態の説明においても、種々変形実施形態に言及したが、さらに、以下に例示するような変形実施形態を挙げることができる。

【0115】

第1の実施形態で説明した文ID付きパターン規則の作成方法に代え、参考にしたい文書が既に存在し、そこからパターン規則を作成したい場合には、<http://cl.aist-nara.ac.jp/lab/nlt/NLT.html>のような統計的な手法を利用した構文解析ツールを用いて構文解析し、その構文解析結果から、名詞句、動詞句、形容詞句、副詞句などの句単位のパターン規則に分割し、パターン規則を作成する方法を適用することができる。

10

【0116】

文ID付き翻訳パターン規則の作成方法(第2の実施形態参照)として、以下の方法を適用できる。参考にしたい翻訳文書が既に存在し、そこから翻訳パターン規則を作成したい場合には、特願2002-367553号明細書及び図面の記載方法を用いることによって翻訳パターン規則を作成することができる。

【0117】

文ID付(翻訳)パターン規則辞書は複数存在していても構わない。文ID付(翻訳)パターン規則辞書を分野や文書毎に複数用意し、参考にしたい分野や文書に合わせて、文ID付(翻訳)パターン規則辞書を使い分けることにより、参考となる分野や文書における結果を模倣する構文解析結果や翻訳結果を得ることができる。

20

【0118】

上記各実施形態では、英語構文解析装置や、英日機械翻訳装置の場合を例に示したが、処理対象文の言語はいずれの言語であっても構わない。

【0119】

第3の実施形態や第4の実施形態の特徴的な技術思想は、機械翻訳装置における構文解析処理(第2の実施形態参照)に適用することができる。

【0120】

上記各実施形態における解析結果や翻訳結果をユーザに表示し、その結果をユーザに確認させ、正解ならば、その際、使用された(翻訳)パターン規則の全て、又は、文IDが付与されていないものを、文IDを付与して、文ID付(翻訳)パターン規則辞書に格納することによって、利用すればするほど、規則が蓄積され、処理の精度を向上させることもできる。すなわち、パターン規則学習部やユーザ登録部を設けるようにしても良い。また、ある文章に対して得られた構文解析結果を構成する、全てのパターン規則、又は、文IDが付与されていないものを、ユーザに確認させることなく、自動的に、文IDを付与して、文ID付パターン規則辞書に格納するようにしても良い。

30

【0121】

第1の実施形態で説明した文IDが付与されたパターン規則が存在しない場合だけでなく、文IDを利用した複数候補の解消と、特許文献2記載のコスト計算を利用した複数候補解消とを組み合わせることもできる。例えば、最も多い数が出現した文IDでも、その出現数が所定数以下であれば、文IDを利用した複数候補の解消方法ではなく、特許文献2記載のコスト計算を利用した複数候補の解消方法を利用する。また例えば、特許文献2記載のコスト計算式に、文IDの数上げた数をパラメータとする項などを設け、文IDの数が多ければ多いほどコストが低くなるようなコストを定義し、そのコストと他で定義される構文解析結果のコストを合わせて計算し、最小のコストでなるパターン規則を選択することによって複数の構文解析結果候補から最適な構文解析結果を求めるようにしても良い。

40

【0122】

50

第 1 の実施形態や第 4 の実施形態で、数え上げ数が閾値数より少ない文 I D を無視するようにしても良い。

【 0 1 2 3 】

また、文 I D と構文要素のカテゴリの両方を同時に評価するようにしても良い。例えば、一部の特別なカテゴリ (N P (名詞句) や V P (動詞句) などの自立語系のカテゴリ) をもつパターン規則の文 I D だけを数えるようにしても良い。つまり、構文要素のカテゴリを考慮して文 I D を利用するようにしても良い。

【 0 1 2 4 】

上記各実施形態では、同一の文 I D は、同一の文から形成されたパターン規則に付与するものを示したが、文 I D をパターン規則の同時適用度として付与するようにしても良い。

10

【 0 1 2 5 】

例えば、同時適用されやすいパターン規則に共通の文 I D を与えることによって、同時適用され易いパターンの組み合わせからなる解析結果が優先的に選ばれるようになる。同一の文 I D の付与は、過去の文書で 1 文中に同時に出現する場合だけでなく、他の手段によって付与することもできる。例えば、関連分野別にパターン規則を分類しておき、関連分野毎に同一の文 I D を付与すると、同一の関連分野のパターン規則の組み合わせからなる解析結果が優先されるようになる。パターン規則の関連分野別の分類は、文章を分野に振分け、その文章から得られたパターン規則に文 I D を付与することにより行うことができる。

【 0 1 2 6 】

また例えば、“ work a 40 hour week ” に基づいて、パターン規則を作成して文 I D を付与した場合において、その文の類似文 “ work a 5 day week ” を考慮してパターン規則を作成し、その作成したパターン規則にも同一の文 I D を付与するようにしても良い。

20

【 0 1 2 7 】

【 発明の効果 】

以上のように、本発明によれば、同一文に同時に適用する可能性の高さを示す文 I D が付与された文 I D 付パターン規則を用意して、同一の文 I D が付与されたパターン規則が多くなっている構文解析結果を採用するようにしたので、構文解析結果の精度を高めることができる。

30

【 図面の簡単な説明 】

【 図 1 】 第 1 の実施形態の自然言語処理装置の機能的構成を示すブロック図である。

【 図 2 】 第 1 の実施形態の自然言語処理装置の動作を示すフローチャートである。

【 図 3 】 第 1 の実施形態の処理の具体的な説明のための入力文の例を示す説明図である。

【 図 4 】 図 3 の入力文に対する第 1 の実施形態での形態素解析例を示す説明図である。

【 図 5 】 第 1 の実施形態の文 I D 付きパターン規則辞書の格納例を示す説明図である。

【 図 6 】 第 1 の実施形態の汎用パターン規則辞書の格納例を示す説明図である。

【 図 7 】 第 1 の実施形態の複数候補の解消前の構文解析結果例を示す説明図である。

【 図 8 】 第 1 の実施形態の文 I D 数え上げ表の一例を示す説明図である。

【 図 9 】 第 1 の実施形態の文 I D の数え上げ方法の例外の説明図である。

40

【 図 10 】 第 1 の実施形態の複数候補の解消後の構文解析結果例を示す説明図である。

【 図 11 】 第 2 の実施形態の自然言語処理装置の機能的構成を示すブロック図である。

【 図 12 】 第 2 の実施形態の自然言語処理装置の動作を示すフローチャートである。

【 図 13 】 第 2 の実施形態の文 I D 付き翻訳パターン規則辞書の格納例を示す説明図である。

【 図 14 】 第 2 の実施形態の汎用翻訳パターン規則辞書の格納例を示す説明図である。

【 図 15 】 第 2 の実施形態の複数候補の解消前の構文解析結果例を示す説明図である。

【 図 16 】 第 2 の実施形態の文 I D 数え上げ表の一例を示す説明図である。

【 図 17 】 第 2 の実施形態の複数候補の解消後の構文解析結果例を示す説明図である。

【 図 18 】 第 2 の実施形態の構文生成結果例を示す説明図である。

50

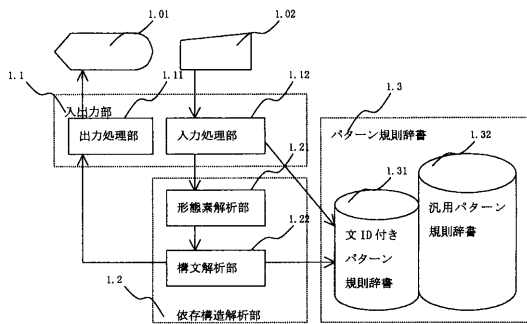
【図19】第3の実施形態の構文解析処理を示すフローチャートである。

【図20】第4の実施形態の構文解析処理を示すフローチャートである。

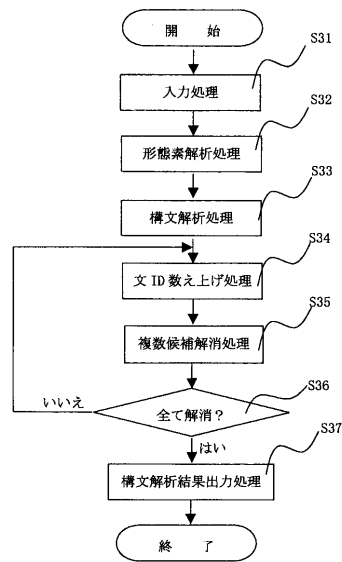
【符号の説明】

1.1...入出力部、1.1.1...出力処理部、1.1.2...入力処理部、1.2...依存構造解析部、1.2.1...形態素解析部、1.2.2...構文解析部、1.3...パターン規則辞書、1.3.1...文ID付きパターン規則辞書、1.3.2...汎用パターン規則、2.1...入出力部、2.1.1...出力処理部、2.1.2...入力処理部、2.2...翻訳処理部、2.2.1...形態素解析部、2.2.2...構文解析・生成部、2.2.3...形態素生成部、2.3...翻訳パターン規則辞書、2.3.1...文ID付翻訳パターン規則、2.3.2...汎用翻訳パターン規則。

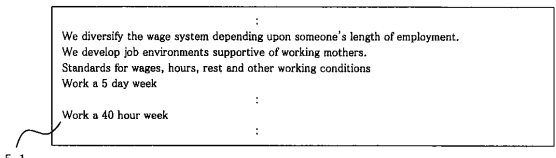
【図1】



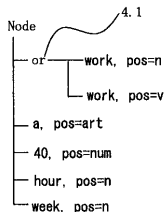
【図2】



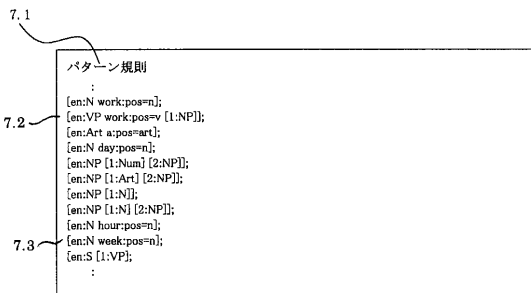
【図3】



【 図 4 】



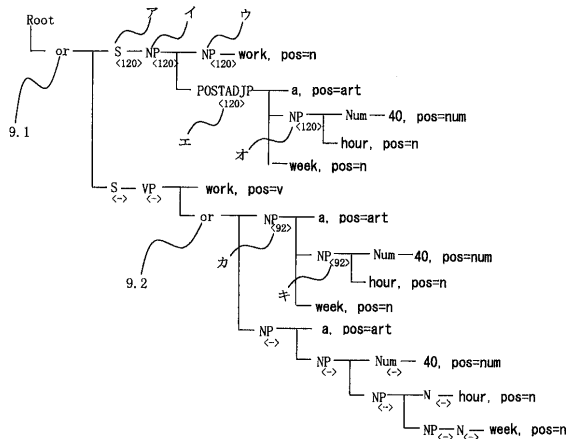
【 図 6 】



【 図 5 】

パターン規則	文 ID
[en:NP a:pos=art [1:NP] week:pos=n];	92
[en:NP [1:Num] day:pos=n];	92
[en:NP standard:pos=n for:pos=prep [1:NP]];	199
[en:NP [1:Num] day:pos=n];	120
[en:POSTADJP a:pos=art [1:NP] week:pos=n];	120
[en:NP [1:NP] [2:POSTADJP]];	120
[en:S [1:NP]];	120
[en:NP [1:N]];	120
[en:N work:pos=n];	120

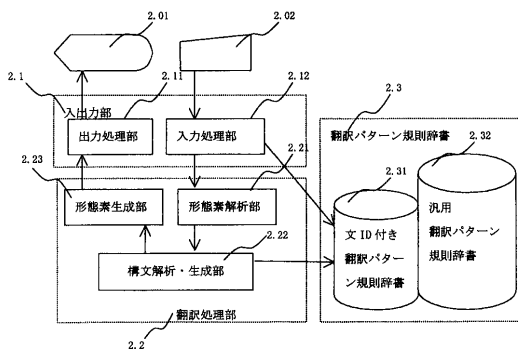
【 図 7 】



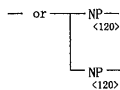
【 図 8 】

文 ID 番号	集計リスト	結果
120	1 1 1 1 1	5
92	1 1	2

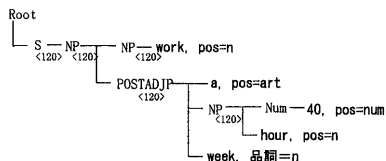
【 図 1 1 】



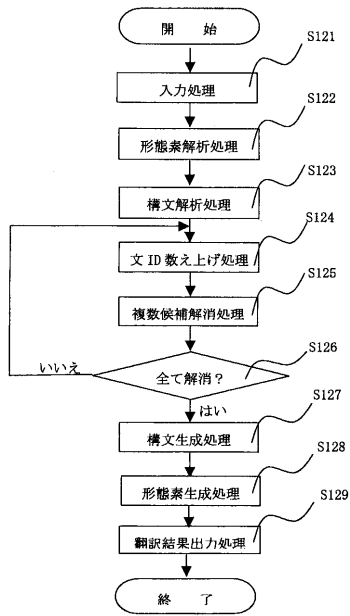
【 図 9 】



【 図 1 0 】



【 図 1 2 】



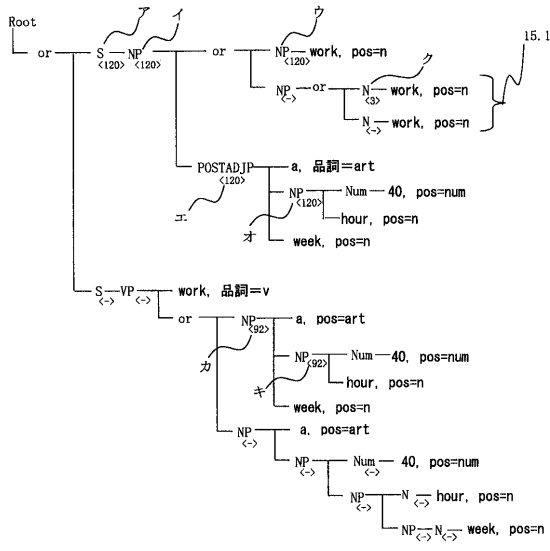
【 図 1 3 】

パターン規則	文 ID
[en:N work:pos=n]	:
[ja:Ms 仕事:pos=ms];	3
:	:
[en:NP a:pos=art [1:NP] week:pos=n]	:
[ja:NP 1 週に付き:pos=rt [1:NP]];	92
:	:
[en:NP [1:Num] day:pos=n]	:
[ja:NP [1:Num] 日:pos=ms];	92
:	:
[en:NP standard:pos=n for:pos=prep [1:NP]]	:
[ja:NP [1:NP] に関する:pos=kj 標準:pos=ms];	199
[en:NP [1:Num] day:pos=n]	:
[ja:NP [1:Num] 日:pos=ms];	120
:	:
[en:POSTADJP a:pos=art [1:NP] week:pos=n]	:
[ja:POSTADJP 1 週に付き:pos=rt [1:NP] の:pos=kj];	120
[en:NP [1:NP] [2:POSTADJP]]	:
[ja:NP [2:POSTADJP] [1:NP]];	120
[en:S [1:NP]]	:
[ja:S [1:NP]];	120
[en:NP work:pos=n]	:
[ja:NP 仕事:pos=ms];	120
:	:

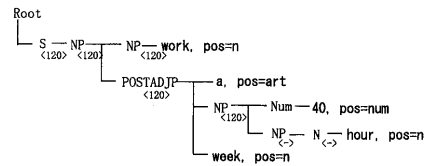
【 図 1 4 】

パターン規則
[en:N work:pos=n]
[ja:Ms 仕事:pos=ms];
[en:VP work:pos=v]
[ja:VP 働<:pos=ds];
[en:VP work:pos=v [1:NP]]
[ja:VP [1:NP] を:pos=kj 動か<:pos=ds];
[en:NP [1:Num] [2:NP]]
[ja:NP [1:Num] [2:NP]];
[en:NP a:pos=art [2:NP]]
[ja:NP [2:NP];
[en:NP [1:N]]
[ja:NP [1:N]]
[en:NP [1:Ms]]
[ja:NP [1:N] [2:NP]]
[en:NP [1:Ms] [2:NP]];
[en:N hour:pos=n]
[ja:Ms 時間:pos=ms];
[en:N week:pos=n]
[ja:Ms 週:pos=ms];
[en:S [1:VP]]
[ja:S [1:VP]];
:

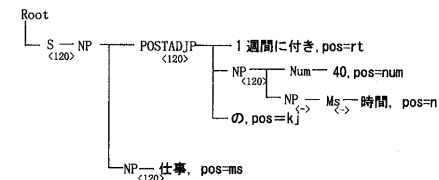
【 図 1 5 】



【 図 1 7 】



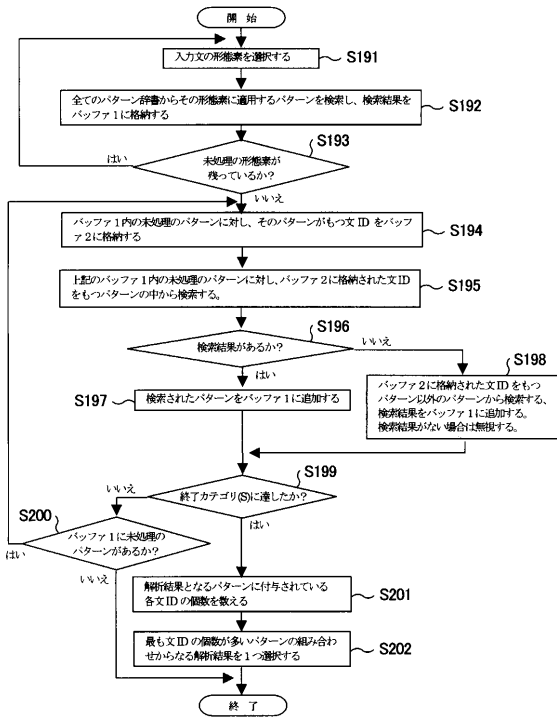
【 図 1 8 】



【 図 1 6 】

文 ID 番号	集計リスト	結果
120	1 1 1 1 1	5
92	1 1	2
3	1	1

【図 19】



【図 20】

