

(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2015-219861

(P2015-219861A)

(43) 公開日 平成27年12月7日(2015.12.7)

(51) Int.Cl. F 1 テーマコード (参考)
G 0 6 F 1 7 / 2 7 (2006.01) G 0 6 F 1 7 / 2 7 Z 5 B 0 9 1

審査請求 未請求 請求項の数 7 O L (全 22 頁)

(21) 出願番号	特願2014-105221 (P2014-105221)	(71) 出願人	000005223 富士通株式会社 神奈川県川崎市中原区上小田中4丁目1番1号
(22) 出願日	平成26年5月21日 (2014.5.21)	(74) 代理人	100087480 弁理士 片山 修平
		(72) 発明者	鄭 育昌 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内
		(72) 発明者	長瀬 友樹 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内
		Fターム(参考)	5B091 AA15 CA05 CA12

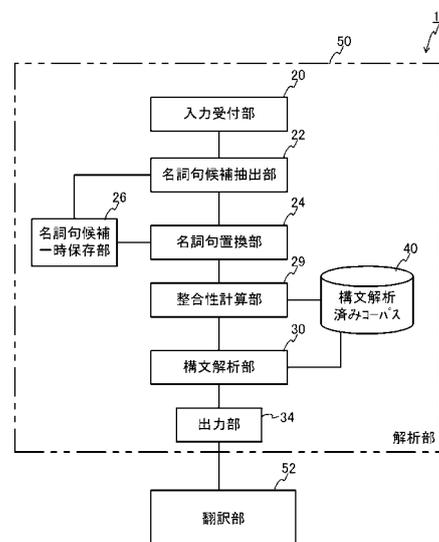
(54) 【発明の名称】 文書解析装置、文書解析プログラム及び文書解析方法

(57) 【要約】

【課題】 文における名詞句の範囲を精度よく解析する。

【解決手段】 解析部50では、名詞句候補抽出部22が、文書内に複数回出現する単語群を名詞句候補文字列として抽出し、整合性計算部29が、構文解析済みコーパス40に含まれる複数の構文事例のうち、名詞句候補文字列を含む複数の文と類似する構文事例をそれぞれ抽出し、複数の文それぞれについて抽出された構文事例に基づいて、名詞句候補文字列の抽出が正しいか否かを判断する。

【選択図】 図2



【特許請求の範囲】

【請求項 1】

文書内に複数回出現する単語群を名詞句候補として仮定する仮定部と、
 構文リストに含まれる複数の正しい構文のうち、前記名詞句候補を含む複数の文と類似する構文をそれぞれ抽出する抽出部と、
 前記抽出部による抽出結果に基づいて、前記名詞句候補の仮定が正しいか否かを判断する判断部と、を備える文書解析装置。

【請求項 2】

前記名詞句候補が複数存在する場合、
 前記抽出部は、文字数の多い名詞句候補を優先的に選択し、選択した前記名詞句候補を含む複数の文と類似する構文をそれぞれ抽出することを特徴とする請求項 1 に記載の文書解析装置。

10

【請求項 3】

前記抽出部が抽出した構文に基づいて、前記名詞句候補の意味を特定する特定部、を更に備える請求項 1 又は 2 に記載の文書解析装置。

【請求項 4】

前記文書内に、複数種類の名詞句候補を含む文が存在する場合に、
 前記抽出部は、前記複数種類の名詞句候補の少なくとも 1 つを含む複数の文を前記文書の中から特定して、特定した前記複数の文それぞれと類似する構文を前記構文リストから抽出し、

20

前記判断部は、前記抽出部による抽出結果に基づいて、前記複数種類の名詞句候補の仮定が正しいか否かを判断する、ことを特徴とする請求項 1 ~ 3 のいずれか一項に記載の文書解析装置。

【請求項 5】

前記判断部の判断結果に基づいて、前記文書の翻訳を実行する翻訳部を更に備える請求項 1 ~ 4 のいずれか一項に記載の文書解析装置。

【請求項 6】

文書内に複数回出現する単語群を名詞句候補として仮定し、
 構文リストに含まれる複数の正しい構文のうち、前記名詞句候補を含む複数の文と類似する構文をそれぞれ抽出し、
 抽出した結果に基づいて、前記名詞句候補の仮定が正しいか否かを判断する、
 処理をコンピュータに実行させる文書解析プログラム。

30

【請求項 7】

文書内に複数回出現する単語群を名詞句候補として仮定し、
 構文リストに含まれる複数の正しい構文のうち、前記名詞句候補を含む複数の文と類似する構文をそれぞれ抽出し、
 抽出した結果に基づいて、前記名詞句候補の仮定が正しいか否かを判断する、
 処理をコンピュータが実行する文書解析方法。

【発明の詳細な説明】

【技術分野】

40

【0001】

本発明は、文書解析装置、文書解析プログラム及び文書解析方法に関する。

【背景技術】

【0002】

従来、中国語などの品詞による語形変換がない言語を機械翻訳する場合、品詞の曖昧性により、名詞句を一つの単位として認識できず、誤った翻訳がなされる場合があった。これに対し、最近では、機械翻訳の際に、解析済コーパスデータベースを用いた構文解析を行う技術が知られている（例えば、特許文献 1 参照）。

【0003】

特許文献 1 においては、文法規則や統計手法に基づいて文を単語ごとに切って形態素に

50

分解し、形態素に基づいて文の構造を解析する。この解析の結果、正しい構文解析結果となりうる候補が複数存在する場合には、解析済コーパスデータベースに含まれるコーパスとの類似度に基づいて、複数の候補から正しい構文解析結果を決定する。

【先行技術文献】

【特許文献】

【0004】

【特許文献1】特開2007-241764号公報

【発明の概要】

【発明が解決しようとする課題】

【0005】

10

しかしながら、翻訳対象の文には様々な構造があるため、上述したようにコーパスとの類似度に基づいて構文解析を行ったとしても、正しい構文解析結果を得られない場合があった。

【0006】

1つの側面では、本発明は、文における名詞句の範囲を精度よく解析することが可能な文書解析装置、文書解析プログラム及び文書解析方法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】

【0007】

一つの態様では、文書解析装置は、文書内に複数回出現する単語群を名詞句候補として仮定する仮定部と、構文リストに含まれる複数の正しい構文のうち、前記名詞句候補を含む複数の文と類似する構文をそれぞれ抽出する抽出部と、前記抽出部による抽出結果に基づいて、前記名詞句候補の仮定が正しいか否かを判断する判断部と、を備えている。

20

【0008】

一つの態様では、文書解析プログラムは、文書内に複数回出現する単語群を名詞句候補として仮定し、構文リストに含まれる複数の正しい構文のうち、前記名詞句候補を含む複数の文と類似する構文をそれぞれ抽出し、抽出した結果に基づいて、前記名詞句候補の仮定が正しいか否かを判断する、処理をコンピュータに実行させる文書解析プログラムである。

【0009】

一つの態様では、文書解析方法は、文書内に複数回出現する単語群を名詞句候補として仮定し、構文リストに含まれる複数の正しい構文のうち、前記名詞句候補を含む複数の文と類似する構文をそれぞれ抽出し、抽出した結果に基づいて、前記名詞句候補の仮定が正しいか否かを判断する、処理をコンピュータが実行する文書解析方法である。

30

【発明の効果】

【0010】

文における名詞句の範囲を精度よく解析することができる。

【図面の簡単な説明】

【0011】

【図1】第1の実施形態に係る翻訳端末のハードウェア構成を概略的に示す図である。

【図2】翻訳端末の機能ブロック図である。

40

【図3】構文解析済みコーパスの一例を示す図である。

【図4】第1の実施形態における解析部の処理を示すフローチャートである。

【図5】解析対象の文書の一例を示す図である。

【図6】図4のステップS12の処理を示すフローチャートである。

【図7】図7(a)～図7(c)は、名詞句候補一時保存部に一時保存されるデータの一例を示す図である。

【図8】図8(a)～図8(d)は、名詞句候補ID(CANDI-1)の名詞句候補文字列の処理例を説明するための図である。

【図9】図4のステップS18の処理を示すフローチャートである。

【図10】図10(a)～図10(c)は、名詞句候補ID(CANDI-2)の名詞句

50

候補文字列の処理例を説明するための図（その１）である。

【図１１】図１１（ａ）～図１１（ｃ）は、名詞句候補ＩＤ（ＣＡＮＤＩ－２）の名詞句候補文字列の処理例を説明するための図（その２）である。

【図１２】図４のステップＳ２０の処理を示すフローチャートである。

【図１３】図１３（ａ）、図１３（ｂ）は、名詞句リストの一例を示す図である。

【図１４】図１４（ａ）、図１４（ｂ）は、名詞句候補ＩＤ（ＣＡＮＤＩ－３）の名詞句候補文字列の処理例を説明するための図である。

【図１５】文書の名詞句を意味記号で置き換えた状態を示す図である。

【図１６】第２の実施形態に係る翻訳端末の機能ブロック図である。

【図１７】第２の実施形態における解析部の処理を示すフローチャートである。

【図１８】第２の実施形態における解析対象の文書の一例を示す図である。

【図１９】図１９（ａ）～図１９（ｄ）は、図１７のステップＳ１２、Ｓ１３'を説明するための図である。

【図２０】図１７のステップＳ１３'の処理を示すフローチャートである。

【図２１】図２１（ａ）、図２１（ｂ）は、図１７のステップＳ２０'を説明するための図（その１）である。

【図２２】図２２（ａ）～図２２（ｄ）は、図１７のステップＳ２０'を説明するための図（その２）である。

【図２３】図２３（ａ）～図２３（ｄ）は、図１７のステップＳ２０'を説明するための図（その３）である。

【図２４】図１７のステップＳ２４に対応する名詞句リストの一例を示す図である。

【図２５】変形例を示す図である。

【発明を実施するための形態】

【００１２】

《第１の実施形態》

以下、文書解析装置としての翻訳端末の第１の実施形態について、図１～図１５に基づいて詳細に説明する。なお、本第１の実施形態の翻訳端末１０は、中国語の文書を日本語に翻訳する（中日翻訳する）端末である。

【００１３】

図１には、翻訳端末１０のハードウェア構成が示されている。翻訳端末１０は、例えば、ＰＣ（Personal Computer）等の端末であり、図１に示すように、ＣＰＵ（Central Processing Unit）９０、ＲＯＭ（Read Only Memory）９２、ＲＡＭ（Random Access Memory）９４、記憶部（ここではＨＤＤ（Hard Disk Drive））９６、ネットワークインタフェース９７、表示部９３、入力部９５、及び可搬型記憶媒体用ドライブ９９等を備えている。これら翻訳端末１０の構成各部は、バス９８に接続されている。表示部９３は、液晶ディスプレイ等を含み、入力部９５は、キーボードやマウス等を含む。翻訳端末１０では、ＲＯＭ９２あるいはＨＤＤ９６に格納されているプログラム（文書解析プログラムを含む）、或いは可搬型記憶媒体用ドライブ９９が可搬型記憶媒体９１から読み取ったプログラム（文書解析プログラムを含む）をＣＰＵ９０が実行することにより、図２に示す機能が実現される。なお、図２には、翻訳端末１０のＨＤＤ９６等に格納されている構文リストとしての構文解析済みコーパス４０も図示されている。

【００１４】

図２には、翻訳端末１０の機能ブロック図が示されている。翻訳端末１０では、ＣＰＵ９０がプログラムを実行することで、図２に示すように、解析部５０及び翻訳部５２としての機能が実現されている。解析部５０は、入力される翻訳対象の中国語文書（例えば図５に示すような特許文書）に含まれる各文の構文解析を行い、構文解析結果を翻訳部５２に対して出力する。翻訳部５２は、各文の構文解析結果に基づいて、各文を日本語に翻訳する。

【００１５】

解析部５０は、入力受付部２０、仮定部としての名詞句候補抽出部２２、名詞句置換部

10

20

30

40

50

24、名詞句候補一時保存部26、整合性計算部29、構文解析部30、及び出力部34を含む。

【0016】

入力受付部20は、翻訳対象、すなわち構文解析対象の中国語文書の入力を受け付け、名詞句候補抽出部22に対して送信する。名詞句候補抽出部22は、文書に複数回出現する複数の単語を含む単語群を名詞句候補文字列と仮定し、抽出する。なお、名詞句候補文字列として検出された単語群は、名詞句候補一時保存部26に一時保存される。ここで、名詞句候補一時保存部26は、一例として、図7(a)~図7(c)に示すようなデータを一時保存するものとする。

【0017】

名詞句置換部24は、名詞句候補一時保存部26に一時保存されたデータに基づいて、名詞句候補文字列を仮の意味記号で表される暫定名詞([**])で置換する。なお、暫定名詞[**]は、任意の意味を表す名詞であるものとする。

【0018】

整合性計算部29は、構文解析済みコーパス40に含まれる複数の正しい構文のうち、名詞句候補文字列を含む複数の文と類似する構文をそれぞれ抽出する。また、整合性計算部29は、該抽出結果に基づいて、名詞句候補文字列が正しい名詞句であるか否かを判断する。ここで、構文解析済みコーパス40は、複数の正しい(解析済みの)構文の例を格納するものであり、図3に示すようなデータ構造を有する。具体的には、構文解析済みコーパス40は、「事例ID」、及び「名詞句を意味記号で置換した構文構造」のフィールドを有する。「名詞句を意味記号で置換した構文構造」のフィールドには、((A)(B))(C)などの表現(A~Cは単語)により、解析済みの構文のツリー構造が格納されている。なお、事例ID=C1001の構文構造に含まれる[方法]や[土壌]という表現は、方法や土壌を意味する単語(同義/類義語)を統括して表す記号(意味記号ともいう)である。

【0019】

構文解析部30は、整合性計算部29の判断結果に基づいて、構文解析を実行する。構文解析結果は、出力部34を介して、翻訳部52に送信される。

【0020】

次に、本第1の実施形態の翻訳端末10において実行される処理について、図4、図6、図9、図12のフローチャートに沿って、その他図面を参照しつつ詳細に説明する。

【0021】

図4には、翻訳端末10の解析部50において実行される処理の概要がフローチャートにて示されている。

【0022】

図4の処理では、まず、ステップS10において、入力受付部20が、中国語文書が入力されるまで待機する。中国語文書が入力されると、ステップS12に移行し、名詞句候補抽出部22は、名詞句候補文字列抽出処理のサブルーチンを実行する。なお、本第1の実施形態では、図5に示すような特許文書が入力受付部20に入力されたものとする。図5では、説明の便宜上、特許文書に含まれる各文に対して、原文ID(Q301、Q302...)を付して示している。

【0023】

ステップS12においては、図6に示すフローチャートに沿った処理が実行される。図6の処理では、ステップS40において、名詞句候補抽出部22は、文書中のテキストを文単位で抽出する。この場合、図5に示す原文IDごとに文を抽出する。

【0024】

次いで、ステップS42では、名詞句候補抽出部22は、文ペアを抽出する。例えば、名詞句候補抽出部22は、原文ID=Q301、Q302の文ペアを抽出する。

【0025】

次いで、ステップS44では、名詞句候補抽出部22が、文ペアの部分一致文字列を検

10

20

30

40

50

出する。この場合、ステップ S 4 2 で抽出した文ペアを比較し、共通する文字列部分を検出する。

【 0 0 2 6 】

次いで、ステップ S 4 6 では、名詞句候補抽出部 2 2 が、部分一致文字列が検出された場合において、該部分一致文字列が既に検出された文字列か否かを判断する。ここでの判断が否定された場合には、ステップ S 4 8 に移行し、名詞句候補抽出部 2 2 は、名詞句候補一時保存部 2 6 に保存する。この場合、名詞句候補一時保存部 2 6 には、例えば、図 7 (a) ~ 図 7 (c) に示すような構造のデータが保存される。図 7 (a) ~ 図 7 (c) のデータは、「名詞句候補 I D」、「名詞句候補文字列」、及び「出現原文 I D」の各項目を有する。「名詞句候補 I D」の項目には、C A N D I - 1、C A N D I - 2、... のような名詞句候補の識別情報が格納される。「名詞句候補文字列」の項目には、ステップ S 4 4 で検出された部分一致文字列が格納される。「出現原文 I D」の項目には、部分一致文字列が検出された文の原文 I D が格納される。

10

【 0 0 2 7 】

ステップ S 4 8 の後は、名詞句候補抽出部 2 2 は、ステップ S 5 0 に移行する。一方、ステップ S 4 6 の判断が肯定された場合には、名詞句候補抽出部 2 2 は、ステップ S 4 9 において、ステップ S 4 2 で抽出された文ペアの原文 I D を名詞句候補一時保存部 2 6 の対応するデータの「出現原文 I D」の欄に格納する。その後は、ステップ S 5 0 に移行する。

【 0 0 2 8 】

ステップ S 5 0 に移行すると、名詞句候補抽出部 2 2 は、全ての文ペアを抽出したか否かを判断する。ここでの判断が否定された場合には、名詞句候補抽出部 2 2 は、ステップ S 4 2 に戻り、上述したステップ S 4 2 ~ S 5 0 の処理・判断を繰り返す。なお、例えば、文ペアとして原文 I D = Q 3 0 2、Q 3 0 4 の文ペアが抽出された場合には、図 7 (a) ~ 図 7 (c) に名詞句候補文字列として示されているような部分一致文字列が検出され、名詞句候補一時保存部 2 6 に格納される。また、例えば、文ペアとして原文 I D = Q 3 0 4、Q 3 0 5 の文ペアが抽出された場合には、図 7 (b)、図 7 (c) に名詞句候補文字列として示されているような部分一致文字列が検出される。また、例えば、文ペアとして原文 I D = Q 3 0 6、Q 3 0 7 の文ペアが抽出された場合には、図 7 (c) に名詞句候補文字列として示されているような部分一致文字列が検出される。

20

30

【 0 0 2 9 】

その後、すべての文ペアを抽出し、ステップ S 5 0 の判断が肯定されると、ステップ S 5 2 に移行する。ステップ S 5 2 に移行すると、名詞句候補抽出部 2 2 は、名詞句候補一時保存部 2 6 に保存されているデータを名詞句候補文字列の文字数が多い順に並べ替える（ソートする）。以上のようにして、図 6 の処理が終了すると、図 4 のステップ S 1 4 に移行する。

【 0 0 3 0 】

図 4 のステップ S 1 4 では、名詞句置換部 2 4 が、1 つの名詞句候補文字列を特定する。この場合、名詞句置換部 2 4 は、名詞句候補一時保存部 2 6 に一時保存されているデータのうち、候補文字列の文字数が最も多いものを特定する。ここでは、図 8 (a) に示すデータ（名詞句候補 I D = C A N D I - 1 の名詞句候補文字列）が特定されたものとする。

40

【 0 0 3 1 】

次いで、ステップ S 1 6 では、名詞句置換部 2 4 は、文書中の名詞句候補文字列を暫定名詞[*]に置換する。具体的には、名詞句置換部 2 4 は、図 8 (b) に示すように、文書の中から、ステップ S 1 4 で特定した名詞句候補文字列を含んでいる文を図 8 (a) のデータの出現原文 I D (= Q 3 0 2、Q 3 0 4) に基づいて抽出する。そして、名詞句置換部 2 4 は、図 8 (c) に示すように、抽出された文のうち、名詞句候補文字列の部分を暫定名詞（仮の意味記号）[*]で置換する。

【 0 0 3 2 】

50

次いで、ステップ S 1 8 では、整合性計算部 2 9 が、名詞句候補文字列に対する意味記号候補を抽出する処理のサブルーチンを実行する。このステップ S 1 8 の処理においては、図 9 のフローチャートに沿った処理が実行される。

【 0 0 3 3 】

図 9 の処理では、まず、ステップ S 6 0 において、整合性計算部 2 9 は、名詞句候補文字列を暫定名詞で置換した文集合 Z (文数 k_{max}) を取得する。ここでは、整合性計算部 2 9 は、図 8 (c) に示す 2 つの文 (文数 $k_{max} = 2$) を取得したとする。

【 0 0 3 4 】

次いで、ステップ S 6 2 では、整合性計算部 2 9 は、文の処理数を示すパラメータ k を 1 に設定する。次いで、ステップ S 6 4 では、整合性計算部 2 9 は、文集合 Z に含まれる文 z_k ($= z_1$) に対し、構文解析を行い、構文解析結果を獲得する。ここでは、例えば、図 8 (c) の原文 ID = Q 3 0 2 の文についての構文解析を行い、図 8 (d) に示す構文解析結果を獲得したものとす。

10

【 0 0 3 5 】

次いで、ステップ S 6 6 では、整合性計算部 2 9 は、構文解析済みコーパス 4 0 において、構文解析結果と類似する構文解析事例を検索する。なお、図 8 (d) の原文 ID = Q 3 0 2 については、類似する構文解析事例が存在していなかったものとする。

【 0 0 3 6 】

次いで、ステップ S 6 8 では、整合性計算部 2 9 は、ステップ S 6 6 の結果、類似する構文解析事例が存在したか否かを判断する。ここでの判断が否定された場合には、ステップ S 7 4 に移行し、整合性計算部 2 9 は、k が k_{max} ($= 2$) であるか否かを判断する。ここでの判断が否定された場合には、ステップ S 7 6 に移行し、整合性計算部 2 9 は、k を 1 インクリメント ($k \rightarrow k + 1$) し、ステップ S 6 4 に戻る。

20

【 0 0 3 7 】

ステップ S 6 4 に戻ると、整合性計算部 2 9 は、次の文 z_2 として、図 8 (c) の原文 ID = Q 3 0 4 の構文解析を行い、構文解析結果を獲得する。ここでは、図 8 (d) の原文 ID = Q 3 0 4 の構文解析結果を得ることができたとする。次いで、ステップ S 6 6 では、整合性計算部 2 9 は、構文解析済みコーパス 4 0 において構文解析結果と類似する構文解析事例を検索する。ここでは、類似する構文解析事例が存在しなかったものとする。したがって、次のステップ S 6 8 の判断は否定され、ステップ S 7 4 に移行する。ステップ S 7 4 に移行すると、整合性計算部 2 9 は、k が k_{max} ($= 2$) であるか否かを判断する。ここでの判断が肯定されると、図 9 の全処理を終了し、図 4 のステップ S 1 9 に移行する。

30

【 0 0 3 8 】

図 4 のステップ S 1 9 に移行すると、整合性計算部 2 9 は、ステップ S 1 8 の処理において意味記号候補を抽出できたか否かを判断する。すなわち、整合性計算部 2 9 は、直前に行われたステップ S 1 8 の処理において、ステップ S 7 2 の処理が実行されたか否かを判断する。このステップ S 1 9 の判断が否定されると、ステップ S 2 6 に移行する。なお、ステップ S 1 9 の判断が否定される場合とは、名詞句候補文字列として抽出した部分が、正しい名詞句でなかったことを意味する。

40

【 0 0 3 9 】

ステップ S 2 6 に移行すると、整合性計算部 2 9 は、全ての名詞句候補を特定したか否かを判断する。このステップ S 2 6 の判断が否定されると、ステップ S 1 4 に戻り、名詞句置換部 2 4 は、2 番目に文字数の多い名詞句候補文字列を特定する。ここでは、図 1 0 (a) に示すように、名詞句候補 ID = C A N D I - 2 の名詞句候補文字列が特定されたものとする。

【 0 0 4 0 】

次いで、ステップ S 1 6 では、名詞句置換部 2 4 は、文書中の名詞句候補文字列を暫定名詞 [* *] に置換する。具体的には、名詞句置換部 2 4 は、図 1 0 (b) に示すように、文書の中から、名詞句候補文字列を含んでいる文を図 1 0 (a) のデータの出現原文 ID (

50

= Q 3 0 2 , Q 3 0 4 , Q 3 0 5 , Q 3 0 6) に基づいて抽出する。そして、名詞句置換部 2 4 は、抽出された文のうち、名詞句候補文字列の部分を暫定名詞 [**] で置換する。

【 0 0 4 1 】

次いで、ステップ S 1 8 では、整合性計算部 2 9 は、名詞句候補文字列に対する意味記号候補を抽出する処理を実行する (図 9) 。

【 0 0 4 2 】

図 9 の処理では、ステップ S 6 0 において、整合性計算部 2 9 は、名詞句候補文字列を暫定名詞で置換した文集合 Z (文数 $k_{\max} = 4$) を取得する。

【 0 0 4 3 】

次いで、ステップ S 6 2 では、整合性計算部 2 9 は、文を表すパラメータ k を 1 に設定する。次いで、ステップ S 6 4 では、整合性計算部 2 9 は、文集合 Z に含まれる文 z k (= z 1) に対し、構文解析を行い、構文解析結果を獲得する。ここでは、例えば、図 1 0 (b) の原文 ID = Q 3 0 2 の下線部を暫定名詞で置換したものについての構文解析を行い、図 1 0 (c) に示す構文解析結果を獲得したものとする。

【 0 0 4 4 】

次いで、ステップ S 6 6 では、整合性計算部 2 9 は、構文解析済みコーパス 4 0 において、構文解析結果と類似する構文解析事例を検索する。なお、図 1 0 (c) の原文 ID = Q 3 0 2 については、類似する構文解析事例として、図 3 に示す、事例 ID = C 1 0 0 1 , C 1 0 0 2 の 2 つの事例が検索されたものとする。

【 0 0 4 5 】

次いで、ステップ S 6 8 では、整合性計算部 2 9 は、ステップ S 6 6 の結果、類似する構文解析事例が存在したか否かを判断する。ここでの判断が肯定されると、ステップ S 7 2 に移行し、整合性計算部 2 9 は、検索された構文解析事例から、名詞句候補文字列の意味記号候補を特定し、意味候補リストに保存する。この場合、図 1 1 (a) に示すように、検索された構文解析事例 (C 1 0 0 1) のうち、文の暫定名詞 [**] と対応する意味記号が、 [中子] であり、検索された構文解析事例 (C 1 0 0 2) のうち、文の暫定名詞 [**] と対応する意味記号が、 [細菌] であるので、整合性計算部 2 9 は、これらの意味記号を意味記号候補と特定し、意味候補リストに保存する。

【 0 0 4 6 】

次いで、ステップ S 7 4 に移行すると、整合性計算部 2 9 は、k が $k_{\max} (= 4)$ であるか否かを判断する。ここでの判断が否定された場合には、ステップ S 7 6 に移行し、整合性計算部 2 9 は、k を 1 インクリメント ($k \rightarrow k + 1$) し、ステップ S 6 4 に戻る。

【 0 0 4 7 】

ステップ S 6 4 に戻ると、整合性計算部 2 9 は、次の文 z 2 として、図 1 0 (b) の原文 ID = Q 3 0 4 の文を暫定名詞 [**] で置換したものを構文解析し、構文解析結果を獲得する。ここでは、図 1 0 (c) の原文 ID = Q 3 0 4 の構文解析結果を得ることができたとする。次いで、ステップ S 6 6 では、整合性計算部 2 9 は、構文解析済みコーパス 4 0 において構文解析結果と類似する構文解析事例を検索する。ここでは、類似する構文解析事例として、事例 ID = C 1 0 0 4 の事例が検索されたものとする。この場合、ステップ S 6 8 の判断が肯定され、ステップ S 7 2 において、整合性計算部 2 9 は、意味記号 [細菌] を意味記号候補と特定し、意味候補リストに保存する。その後、ステップ S 7 4 の判断が否定されると、ステップ S 7 6 に移行し、整合性計算部 2 9 は、k を 1 インクリメント ($k \rightarrow k + 1$) し、ステップ S 6 4 に戻る。

【 0 0 4 8 】

以降、ステップ S 6 4 ~ S 7 6 の処理、判断を繰り返し、整合性計算部 2 9 は、図 1 0 (b) の残りの 2 つの文についても構文解析を行う。なお、本第 1 の実施形態では、原文 ID = Q 3 0 5 の文に関しては、意味記号 [細菌] が意味記号候補と特定され、原文 ID = Q 3 0 6 の文に関しては、意味記号候補は特定されなかったものとする。

【 0 0 4 9 】

その後、ステップ S 7 4 の判断が肯定されると、図 4 のステップ S 1 9 に移行する。

10

20

30

40

50

【 0 0 5 0 】

ステップ S 1 9 に移行すると、整合性計算部 2 9 は、ステップ S 1 8 の処理において意味記号候補を抽出できたか否かを判断する。すなわち、整合性計算部 2 9 は、直前に行われたステップ S 1 8 の処理において、ステップ S 7 2 の処理が実行されたか否かを判断する。このステップ S 1 9 の判断が肯定されると、整合性計算部 2 9 は、ステップ S 2 0 に移行する。

【 0 0 5 1 】

ステップ S 2 0 では、整合性計算部 2 9 は、名詞句候補文字列が意味記号候補の場合の整合性評価のサブルーチンを実行する。本ステップ S 2 0 では、具体的には、図 1 2 のフローチャートに沿った処理が実行される。

10

【 0 0 5 2 】

図 1 2 の処理では、まず、ステップ S 8 0 において、整合性計算部 2 9 は、文の処理数を示すパラメータ k を 1、意味記号候補の特定数を示すパラメータ c を 1、整合性評価に用いるパラメータ n を 0 に設定する。なお、ここでは、パラメータ k の最大値 k_{max} は 4 であり、パラメータ c の最大値 c_{max} は 2 である。

【 0 0 5 3 】

次いで、ステップ S 8 2 では、整合性計算部 2 9 は、意味記号候補 $T_c (= T_1)$ を特定する。ここでは、意味記号候補 T_c として、[中子] が特定されたものとする。次いで、整合性計算部 2 9 は、文 $z_k (= z_1)$ の意味候補リストに意味記号候補 $T_c (= T_1)$ が存在するか否かを判断する。図 1 1 (b) に示すよう原文 $ID = Q 3 0 2$ の文は、意味記号候補 T_c が [中子] である場合に、一致事例 ($C 1 0 0 1$) が存在しているので、ステップ S 8 4 の判断は肯定され、ステップ S 8 6 に移行する。

20

【 0 0 5 4 】

ステップ S 8 6 に移行すると、整合性計算部 2 9 は、 n を 1 インクリメント ($n \rightarrow n + 1$) し、ステップ S 8 8 に移行する。ステップ S 8 8 では、整合性計算部 2 9 は、 k が最大値 (k_{max}) であるか否かを判断し、判断が否定されると、ステップ S 9 0 において、 k を 1 インクリメント ($k \rightarrow k + 1$) し、ステップ S 8 2 に戻る。その後は、整合性計算部 2 9 は、全ての文の意味候補リストに意味記号候補 T_1 が存在するか否かを判断し、 k が最大値 (k_{max}) となった段階で、ステップ S 9 2 に移行する。

【 0 0 5 5 】

ステップ S 9 2 に移行すると、整合性計算部 2 9 は、暫定名詞 [**] が意味記号候補 $T_c (= T_1)$ である場合の整合性評価値を算出する。具体的には、次式 (1) より、整合性評価値を算出する。

30

$$\text{整合性評価値} = n / k_{max} \quad \dots (1)$$

【 0 0 5 6 】

図 1 1 (b) の場合、 $n = 1$ となるため、整合性評価値は、 $1 / 4 = 0 . 2 5$ となる。

【 0 0 5 7 】

次いで、ステップ S 9 4 では、整合性計算部 2 9 は、 c が c の最大値 ($c_{max} = 2$) であるか否かを判断する。ここでの判断が否定されると、ステップ S 9 6 に移行し、整合性計算部 2 9 は、 k を 1 に戻すとともに、 c を 1 インクリメント ($c \rightarrow c + 1$) し、ステップ S 8 2 に戻る。

40

【 0 0 5 8 】

ステップ S 8 2 に戻った後は、上述したように、ステップ S 8 2 ~ S 9 2 の処理・判断を実行する。ここで、意味記号候補 T_2 として、[細菌] が抽出された場合、図 1 1 (c) に示すように、意味候補リストに意味記号候補 T_2 が存在する文の数 n は、3 である。したがって、ステップ S 9 2 では、整合性計算部 2 9 は、整合性評価値として、 $3 / 4 = 0 . 7 5$ を算出する。

【 0 0 5 9 】

その後、ステップ S 9 4 の判断が肯定されると、整合性計算部 2 9 は、図 1 2 の全処理を終了し、図 4 のステップ S 2 2 に移行する。ステップ S 2 2 に移行すると、整合性計算

50

部 29 は、整合性評価値が最大、かつ閾値を超えた意味記号候補があるか否かを判断する。例えば、閾値が 0.5 であるとする、図 11 (b)、図 11 (c) の例では、整合性評価値の最大値 0.75 が閾値よりも大きいので、ステップ S 22 の判断が肯定され、ステップ S 24 に移行する。ステップ S 24 では、整合性計算部 29 は、検出した名詞句と、意味記号候補を名詞句リストに登録する。図 13 (a) には、名詞句リストの一例が示されている。名詞句リストにおいては、ステップ S 14 で特定された名詞句候補文字列が、「抽出された名詞句」の欄に格納され、ステップ S 22 で特定された意味記号候補が、「意味記号」の欄に格納される。その後は、ステップ S 26 に移行する。なお、ステップ S 22 が肯定される場合とは、名詞句候補文字列として抽出した部分が正しい名詞句であったことを意味する。一方、ステップ S 22 の判断が否定された場合には、ステップ S 24 を経ずに、ステップ S 26 に移行する。なお、ステップ S 22 が否定される場合とは、名詞句候補文字列と仮定して抽出した部分が正しい名詞句ではなかったことを意味する。

10

【0060】

ステップ S 26 に移行すると、整合性計算部 29 は、全ての名詞句候補文字列を特定したか否かを判断する。このステップ S 26 の判断が否定されると、ステップ S 14 に戻り、次の名詞句候補文字列について、ステップ S 14 ~ S 26 の処理・判断が実行される。例えば、図 14 (a) に示すように、名詞句候補 ID = CANDI - 3 の名詞句候補文字列が特定されたとする (S 14)。この場合、図 14 (b) に示すように、名詞句候補 ID = CANDI - 3 の出現原文 ID のうち、名詞句候補 ID = CANDI - 2 の出現原文 ID に含まれていない ID の文 (原文 ID = Q307、Q308) の名詞句候補文字列を暫定名詞 [**] に変換するなどして、該名詞句候補文字列の意味記号として、[細菌] を特定する。これにより、名詞句リストには、図 13 (b) に示す 2 つ目のデータが追加される。

20

【0061】

その後、ステップ S 26 の判断が肯定されると、ステップ S 28 に移行し、構文解析部 30 は、名詞句リストに基づいて名詞句を意味記号に置換し、構文解析済みコーパス 40 を用いて文書を解析する。これにより、構文解析部 30 は、名詞句を適切に区切り、適切な意味記号で置換した文を解析することができるため、高精度な構文解析結果を得ることができる。その後、構文解析部 30 は、ステップ S 30 に移行し、出力部 34 を介して、解析結果を翻訳部 52 に対して出力する。

30

【0062】

なお、翻訳部 52 では、高精度な構文解析結果を用いて、文書を翻訳することができる。これにより、高精度な翻訳結果を得ることが可能となる。

【0063】

なお、上記説明から分かるように、図 4 のステップ S 14 ~ S 22 の処理においては、構文解析済みコーパス 40 に含まれる構文事例のうち、名詞句候補文字列を含む複数の文と類似する構文事例をそれぞれ抽出し (S 66、図 11 (a) 参照)、ステップ S 18、S 20、S 22 において、抽出結果に基づいて、名詞句候補文字列の抽出が正しかったか判断し、名詞句候補文字列の意味を特定しているといえる。すなわち、本実施形態の整合性計算部 29 により、構文解析済みコーパス 40 に含まれる複数の正しい構文のうち、名詞句候補文字列を含む複数の文と類似する構文をそれぞれ抽出する抽出部、及び抽出部による抽出結果に基づいて、名詞句候補文字列が正しい名詞句であるか否かを判断する判断部としての機能が実現されている。また、本実施形態の整合性計算部 29 により、抽出部が抽出した構文に基づいて、名詞句候補文字列の意味を特定する特定部としての機能が実現されている。

40

【0064】

以上、詳細に説明したように、本第 1 の実施形態によると、解析部 50 では、名詞句候補抽出部 22 が、文書内に複数回出現する単語群を名詞句候補文字列として抽出し (S 12)、整合性計算部 29 が、構文解析済みコーパス 40 に含まれる複数の構文事例のうち、名詞句候補文字列を含む複数の文と類似する構文事例をそれぞれ抽出し、複数の文それ

50

それぞれについて抽出された構文事例に基づいて、名詞句候補文字列の抽出が正しいか否かを判断する（S 1 4 ~ S 2 2）。これにより、複数の文を用いて名詞句候補文字列を仮定し、該仮定が正しいか否かを判断することで、文全体において名詞句候補文字列が正しい名詞句であるか否かを精度よく判断することができる。この場合、例えば、1つの文において名詞句候補文字列を仮定し、構文解析済みコーパス40の構文事例と比較する方法よりも、精度よく名詞句候補文字列が正しい名詞句か否かを判断することができる。これにより、品詞による語形変換がないため品詞の曖昧性により誤訳が生じやすい中国語の翻訳において、高精度な翻訳結果を得ることができる。

【0065】

また、本第1の実施形態では、名詞句候補文字列のうち文字数の多い名詞句候補を優先して、処理することとしている（S 1 4）。これにより、文字数の少ない名詞句候補文字列を先に処理した場合に生じる、文字数の多い名詞句が分断される事態の発生を回避し、高精度な構文解析を実現することができる。

10

【0066】

また、本第1の実施形態では、整合性計算部29は、複数の文それぞれについて抽出された構文に基づいて、名詞句候補文字列の意味を特定する（意味記号を決定し、名詞句リストに登録する）。これにより、名詞句の意味を考慮した構文解析を行うことができるため、高精度な構文解析が可能となる。

【0067】

また、本第1の実施形態では、解析部50の解析結果に基づいて、翻訳部52が文書の翻訳を実行するので、高精度な構文解析に基づく翻訳により、高精度な翻訳結果を得ることができる。

20

【0068】

《第2の実施形態》

以下、第2の実施形態について、図16～図24に基づいて、詳細に説明する。なお、本第2の実施形態では、1つの文に複数の名詞句候補文字列が含まれる場合の例について説明する。なお、翻訳端末10の装置構成は、第1の実施形態と同様であるが、図16に示すように、解析部50が名詞句集合一時保存部25としての機能を有している点が異なる。

【0069】

図17は、本第2の実施形態における、解析部50の処理の概要を示すフローチャートである。図17において、第1の実施形態と異なる処理については、ステップ番号に「'」を付して示している。

30

【0070】

図17の処理では、翻訳対象（解析対象）の文書が入力されると、ステップS12において、第1の実施形態と同様にして、名詞句候補抽出部22が、名詞句候補抽出処理を実行する（図6参照）。なお、本第2の実施形態では、図18に示すような特許文書が入力されたものとし、ステップS12では、図19（a）において下線を付して示す部分が名詞句候補文字列として抽出されたものとする。この場合、名詞句候補一時保存部26には、図19（b）、図19（c）のデータが一時保存されたものとする。

40

【0071】

次いで、図17のステップS13'では、名詞句候補抽出部22が、名詞句集合候補抽出処理のサブルーチンを実行する。具体的には、ステップS13'においては、図20に示す処理が実行される。

【0072】

図20の処理では、ステップS102において、名詞句候補抽出部22が、名詞句候補のパラメータ*i*を1に設定するとともに、文のパラメータ*k*を1に設定する。次いで、ステップS104では、名詞句候補抽出部22は、名詞句候補*N_i*（=*N*1）を特定する。ここでは、名詞句候補ID=CANDI-1の名詞句候補文字列が特定されたものとする。そして、名詞句候補抽出部22は、特定した名詞句候補が文*z_k*（=*z*1）の中に存在

50

しているかどうかを確認する。なお、文 z 1 は、一例として、図 19 (a) の原文 ID = Q 3 0 2 であるものとする。

【 0 0 7 3 】

次いで、ステップ S 1 0 6 では、名詞句候補抽出部 2 2 が、i が i の最大値 (i max = 2) であるか否かを判断する。ここでの判断が否定されると、ステップ S 1 0 8 において、名詞句候補抽出部 2 2 は、i を 1 インクリメント (i = i + 1) し、ステップ S 1 0 4 に戻る。

【 0 0 7 4 】

ステップ S 1 0 4 に戻ると、名詞句候補抽出部 2 2 は、名詞句候補 N 2 として名詞句候補 ID = C A N D I - 2 の名詞句候補文字列を特定し、文 z 1 に該名詞句候補文字列が存在するか否かを確認する。その後、ステップ S 1 0 6 における判断が肯定されると、ステップ S 1 1 0 に移行し、名詞句候補抽出部 2 2 は、文 z k (= z 1) に名詞句候補文字列が 1 以上存在していたか否かを判断する。

10

【 0 0 7 5 】

このステップ S 1 1 0 の判断が肯定されると、ステップ S 1 1 2 に移行し、名詞句候補抽出部 2 2 は、名詞句集合一時保存部 2 5 に文 z k に存在していた名詞句候補文字列の情報を名詞句集合の情報として格納する。図 19 (d) には、名詞句集合一時保存部 2 5 が一時保存するデータの一例が示されている。図 19 (d) に示すように、名詞句集合一時保存部 2 5 に一時保存されるデータは、「名詞句集合 ID」と、「名詞句集合」の項目を含んでいる。文 z k (ID = Q 3 0 2) の場合、名詞句候補 ID = C A N D I - 1、C A N D I - 2 の両方が存在しているので、ステップ S 1 1 2 では、名詞句集合として、名詞句集合 ID = 1 に示すような情報が格納されることになる。なお、名詞句候補抽出部 2 2 は、名詞句集合一時保存部 2 5 に既に保存されている情報については、重複して保存しないようにする。

20

【 0 0 7 6 】

ステップ S 1 1 2 の後、又はステップ S 1 1 0 の判断が否定された場合には、ステップ S 1 1 4 に移行する。ステップ S 1 1 4 では、名詞句候補抽出部 2 2 は、k が k の最大値 (k max) と同一であるか否かを判断する。ここでの判断が否定された場合には、名詞句候補抽出部 2 2 は、ステップ S 1 1 6 において、i を 1 に戻すとともに、k を 1 インクリメントした後、ステップ S 1 0 2 に戻る。その後は、ステップ S 1 0 4 以降の処理をステップ S 1 1 4 の判断が肯定されるまで実行する。そして、図 19 (d) に示すようなデータが名詞句集合一時保存部 2 5 に格納され、ステップ S 1 1 4 の判断が肯定された段階で、図 20 の処理を終了する。その後は、図 17 のステップ S 1 4 ' に移行する。

30

【 0 0 7 7 】

図 17 のステップ S 1 4 ' に移行すると、名詞句置換部 2 4 は、図 19 (d) の中から、1 つの名詞句集合候補を特定する。例えば、名詞句置換部 2 4 は、図 19 (d) の中から、名詞句集合 ID = 1 の名詞句集合候補を特定したものとする。

【 0 0 7 8 】

次いで、ステップ S 1 6 ' では、名詞句置換部 2 4 は、文書中の名詞句候補文字列を暫定名詞に置換する。この場合、ステップ S 1 4 ' で特定した名詞句集合候補に含まれる名詞句候補文字列を、図 21 (a) に示すように、暫定名詞 [**1]、[**2] を用いて置換する。

40

【 0 0 7 9 】

次いで、ステップ S 1 8 ' では、整合性計算部 2 9 は、名詞句候補文字列に対する意味記号候補の抽出処理を実行する。このステップ S 1 8 ' では、図 21 (a) の各文を図 21 (b) のように構文解析し、各構文構造と類似する構文構造事例が構文解析済みコーパス 40 に含まれているかどうかを検索する。例えば、図 22 (a) に示す文 (Q 3 0 2) は、図 22 (b) に示す構文構造事例 (C 1 0 0 1、C 1 0 0 2) と類似していたとする。この場合、図 22 (c) に示すように、[**1] の意味候補リストに意味記号候補 [中子] と [細菌] が保存され、[**2] の意味候補リストに意味記号候補 [土壌] と「金属」が保存さ

50

れる。以下、同様に、他の文についても構文解析が行われるが、意味候補リストは、図 2 2 (c) のままであったとする。

【 0 0 8 0 】

次いで、ステップ S 1 9 では、整合性計算部 2 9 は、意味記号候補を抽出できたか否かを判断する。ここでの判断が否定された場合には、ステップ S 2 6 ' に移行するが、肯定された場合には、ステップ S 2 0 ' に移行する。

【 0 0 8 1 】

ステップ S 2 0 ' では、整合性計算部 2 9 は、名詞句集合候補が意味記号候補の場合の整合性を評価する処理を実行する。ここでは、整合性計算部 2 9 は、図 2 2 (d) に示すように、図 2 2 (c) の意味候補リストに基づいて、各暫定名詞の仮説 (仮説 1 ~ 4) を立て、第 1 の実施形態と同様に、整合性評価値を算出する。例えば、仮説 1 を採用した場合の構文解析の結果、図 2 3 (a) に示すように、原文 ID = Q 3 0 8 の文のみ、構文解析済みコーパス 4 0 に一致事例が存在していたとする。この場合、仮説 1 の整合性評価値は、 $1 / 4 = 0 . 2 5$ となる。同様に、仮説 2 を採用した場合の構文解析の結果、図 2 3 (b) に示すように、原文 ID = Q 3 0 2 の文のみ、構文解析済みコーパス 4 0 に一致事例が存在していたとする。この場合、仮説 2 の整合性評価値は、 $1 / 4 = 0 . 2 5$ となる。また、仮説 3 を採用した場合の構文解析の結果、図 2 3 (c) に示すように、原文 ID = Q 3 0 2、Q 3 0 4、Q 3 0 8 の文に一致事例が存在していたとする。この場合、仮説 3 の整合性評価値は、 $3 / 4 = 0 . 7 5$ となる。更に、仮説 4 を採用した場合の構文解析の結果、図 2 3 (d) に示すように、原文 ID = Q 3 0 4 の文のみ、一致事例が存在していたとする。この場合、仮説 4 の整合性評価値は、 $1 / 4 = 0 . 2 5$ となる。

10

20

【 0 0 8 2 】

以上のように、ステップ S 2 0 ' において、図 2 3 (a) ~ 図 2 3 (d) の整合性評価値を得ると、整合性計算部 2 9 は、次のステップ S 2 2 ' に移行する。

【 0 0 8 3 】

ステップ S 2 2 ' に移行すると、整合性評価値が最大、かつ閾値 (例えば、 $0 . 5$) を超えた仮説が存在するか否かを判断する。ここでの判断が否定された場合には、ステップ S 2 6 ' に移行するが、肯定された場合には、ステップ S 2 4 に移行する。ステップ S 2 4 に移行すると、整合性計算部 2 9 は、図 2 4 に示すように、名詞句リストに意味記号を登録した後、ステップ S 2 6 ' に移行する。

30

【 0 0 8 4 】

ステップ S 2 6 ' に移行すると、整合性計算部 2 9 は、全ての名詞句集合候補を特定したか否かを判断する。このステップ S 2 6 ' の判断が否定された場合には、ステップ S 1 4 ' に戻り、上述した処理を繰り返す。なお、図 1 9 (d) の名詞句集合 ID = 2、3 の名詞句集合のように、既に前の処理 (名詞句集合 ID = 1 の処理) において意味記号が確定している名詞句候補文字列のみを含む集合については、ステップ S 1 4 ' 以降の処理を実行しなくてもよい。一方、ステップ S 2 6 ' の判断が肯定された場合には、ステップ S 2 8 に移行する。

【 0 0 8 5 】

ステップ S 2 8 では、第 1 の実施形態と同様に、構文解析部 3 0 は、名詞句リスト (図 2 4) に基づいて名詞句を意味記号に置換して、文書の解析を行う。そして、ステップ S 3 0 では、構文解析部 3 0 は、出力部 3 4 を介してステップ S 2 8 の解析結果を翻訳部 5 2 に対して出力する。

40

【 0 0 8 6 】

以上、詳細に説明したように、本第 2 の実施形態によると、第 1 の実施形態と同様の効果が得られるほか、複数種類の名詞句候補文字列を含む文が存在する場合に、名詞句候補抽出部 2 2 は、複数種類の名詞句候補文字列の少なくとも 1 つを含む文を前記文書の中から特定し、整合性計算部 2 9 は、特定した文それぞれと類似する構文事例を構文解析済みコーパス 4 0 から抽出する。そして、整合性計算部 2 9 は抽出した構文事例に基づいて、複数の名詞句候補文字列が正しい名詞句であるか否かを判断する。これにより、複数種類

50

の名詞句候補文字列が正しい名詞句であるか否かを一度に判断することができるので、複数種類の名詞句候補文字列を用いた総合的な判断を行うことができる。これにより、1つずつ名詞句候補文字列が正しい名詞句であるか否かを判断する場合よりも、高精度な判断が可能となる。

【0087】

なお、上記第1、第2の実施形態では、解析部50及び翻訳部52を翻訳端末10が有する場合について説明したが、これに限られるものではない。例えば、図25に示すように、ネットワーク180に接続されたサーバ110が解析部50や翻訳部52を有していても良い。この場合、クライアント120から翻訳対象の文書を入力することで、該文書がサーバ110において高精度に翻訳され、翻訳文がサーバ110からクライアント120に対して出力されるようになる。なお、図25の場合、解析部50及び翻訳部52のいずれかをクライアント120が有していてもよい。

10

【0088】

なお、上記第1、第2の実施形態では、中国語から日本語への翻訳を例にとり説明したが、これに限られるものではない。中国語以外の、品詞による語形変化がない言語の翻訳において、上記第1、第2の実施形態の装置や方法を用いることとしてもよい。また、中国語等から日本語以外の言語に翻訳する場合に、上記第1、第2の実施形態の装置や方法を用いることとしてもよい。

【0089】

なお、上記の処理機能は、コンピュータによって実現することができる。その場合、処理装置が有すべき機能の処理内容を記述したプログラムが提供される。そのプログラムをコンピュータで実行することにより、上記処理機能がコンピュータ上で実現される。処理内容を記述したプログラムは、コンピュータで読み取り可能な記録媒体（ただし、搬送波は除く）に記録しておくことができる。

20

【0090】

プログラムを流通させる場合には、例えば、そのプログラムが記録されたDVD（Digital Versatile Disc）、CD-ROM（Compact Disc Read Only Memory）などの可搬型記録媒体の形態で販売される。また、プログラムをサーバコンピュータの記憶装置に格納しておき、ネットワークを介して、サーバコンピュータから他のコンピュータにそのプログラムを転送することもできる。

30

【0091】

プログラムを実行するコンピュータは、例えば、可搬型記録媒体に記録されたプログラムもしくはサーバコンピュータから転送されたプログラムを、自己の記憶装置に格納する。そして、コンピュータは、自己の記憶装置からプログラムを読み取り、プログラムに従った処理を実行する。なお、コンピュータは、可搬型記録媒体から直接プログラムを読み取り、そのプログラムに従った処理を実行することもできる。また、コンピュータは、サーバコンピュータからプログラムが転送されるごとに、逐次、受け取ったプログラムに従った処理を実行することもできる。

【0092】

上述した実施形態は本発明の好適な実施の例である。但し、これに限定されるものではなく、本発明の要旨を逸脱しない範囲内において種々変形実施可能である。

40

【0093】

なお、以上の第1、第2の実施形態の説明に関して、更に以下の付記を開示する。

（付記1） 文書内に複数回出現する単語群を名詞句候補として仮定する仮定部と、構文リストに含まれる複数の正しい構文のうち、前記名詞句候補を含む複数の文と類似する構文をそれぞれ抽出する抽出部と、

前記抽出部による抽出結果に基づいて、前記名詞句候補の仮定が正しいか否かを判断する判断部と、を備える文書解析装置。

（付記2） 前記名詞句候補が複数存在する場合、

前記抽出部は、文字数の多い名詞句候補を優先的に選択し、選択した前記名詞句候補を

50

含む複数の文と類似する構文をそれぞれ抽出することを特徴とする付記 1 に記載の文書解析装置。

(付記 3) 前記抽出部が抽出した構文に基づいて、前記名詞句候補の意味を特定する特定部、を更に備える付記 1 又は 2 に記載の文書解析装置。

(付記 4) 前記文書内に、複数種類の名詞句候補を含む文が存在する場合に、

前記抽出部は、前記複数種類の名詞句候補の少なくとも 1 つを含む複数の文を前記文書の中から特定して、特定した前記複数の文それぞれと類似する構文を前記構文リストから抽出し、

前記判断部は、前記抽出部による抽出結果に基づいて、前記複数種類の名詞句候補の仮定が正しいか否かを判断する、ことを特徴とする付記 1 ~ 3 のいずれかに記載の文書解析装置。

(付記 5) 前記判断部の判断結果に基づいて、前記文書の翻訳を実行する翻訳部を更に備える付記 1 ~ 4 のいずれかに記載の文書解析装置。

(付記 6) 文書内に複数回出現する単語群を名詞句候補として仮定し、

構文リストに含まれる複数の正しい構文のうち、前記名詞句候補を含む複数の文と類似する構文をそれぞれ抽出し、

抽出した結果に基づいて、前記名詞句候補の仮定が正しいか否かを判断する、

処理をコンピュータに実行させる文書解析プログラム。

(付記 7) 前記名詞句候補が複数存在する場合、

前記抽出する処理では、文字数の多い名詞句候補を優先的に選択し、選択した前記名詞句候補を含む複数の文と類似する構文をそれぞれ抽出することを特徴とする付記 6 に記載の文書解析プログラム。

(付記 8) 前記抽出する処理において抽出された構文に基づいて、前記名詞句候補の意味を特定する、処理を前記コンピュータに更に実行させる付記 6 又は 7 に記載の文書解析プログラム。

(付記 9) 前記文書内に、複数種類の名詞句候補を含む文が存在する場合に、

前記抽出する処理では、前記複数種類の名詞句候補の少なくとも 1 つを含む複数の文を前記文書の中から特定して、特定した前記複数の文それぞれと類似する構文を前記構文リストから抽出し、

前記判断する処理では、前記抽出する処理における抽出結果に基づいて、前記複数種類の名詞句候補の仮定が正しいか否かを判断する、ことを特徴とする付記 6 ~ 8 のいずれかに記載の文書解析プログラム。

(付記 10) 前記判断する処理における判断結果に基づいて、前記文書を翻訳する、処理を前記コンピュータに実行させる付記 6 ~ 9 のいずれかに記載の文書解析プログラム。

(付記 11) 文書内に複数回出現する単語群を名詞句候補として仮定し、

構文リストに含まれる複数の正しい構文のうち、前記名詞句候補を含む複数の文と類似する構文をそれぞれ抽出し、

抽出した結果に基づいて、前記名詞句候補の仮定が正しいか否かを判断する、

処理をコンピュータが実行する文書解析方法。

(付記 12) 前記名詞句候補が複数存在する場合、

前記抽出する処理では、文字数の多い名詞句候補を優先的に選択し、選択した前記名詞句候補を含む複数の文と類似する構文をそれぞれ抽出することを特徴とする付記 11 に記載の文書解析方法。

(付記 13) 前記抽出する処理において抽出された構文に基づいて、前記名詞句候補の意味を特定する、処理を前記コンピュータが更に実行する付記 11 又は 12 に記載の文書解析方法。

(付記 14) 前記文書内に、複数種類の名詞句候補を含む文が存在する場合に、

前記抽出する処理では、前記複数種類の名詞句候補の少なくとも 1 つを含む複数の文を前記文書の中から特定して、特定した前記複数の文それぞれと類似する構文を前記構文リストから抽出し、

10

20

30

40

50

前記判断する処理では、前記抽出する処理における抽出結果に基づいて、前記複数種類の名詞句候補の仮定が正しいか否かを判断する、ことを特徴とする付記 1 1 ~ 1 3 のいずれかに記載の文書解析方法。

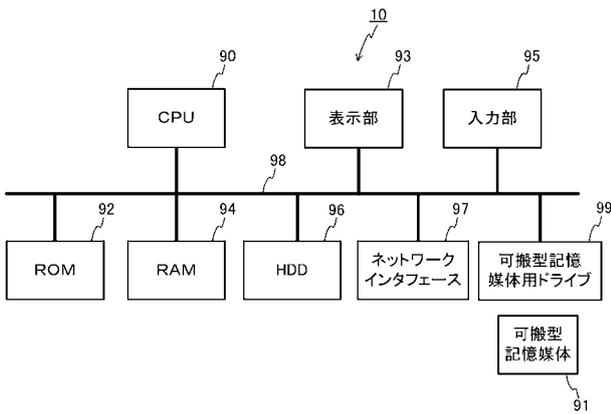
(付記 1 5) 前記判断する処理における判断結果に基づいて、前記文書を翻訳する、処理を前記コンピュータが更に実行する付記 1 1 ~ 1 4 のいずれかに記載の文書解析方法。

【符号の説明】

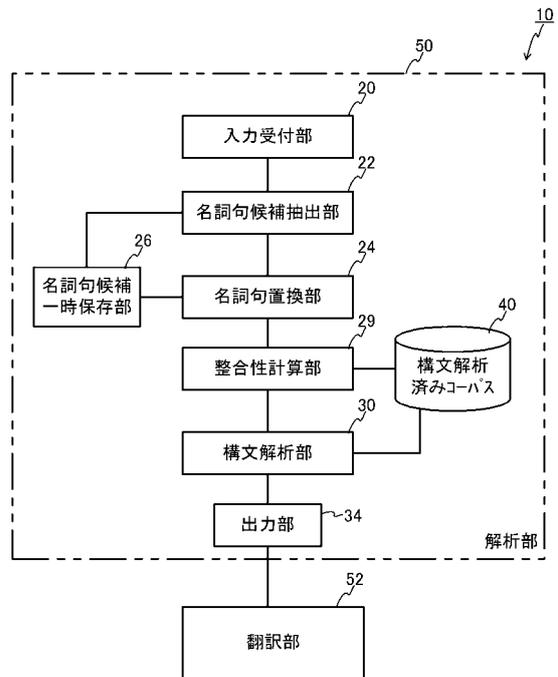
【0094】

- 1 0 翻訳端末（文書解析装置）
- 2 2 名詞句候補抽出部（仮定部）
- 2 9 整合性計算部（抽出部、判断部、特定部）
- 4 0 構文解析済みコーパス（構文リスト）
- 5 2 翻訳部

【図 1】



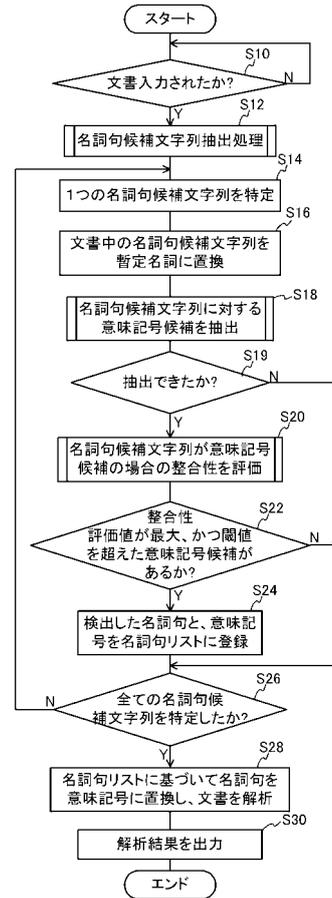
【図 2】



【 図 3 】

＜構文解析済みコーパス＞	
事例ID	名詞句を意味記号で置換した構文構造
C:1001	((一)種)((採用)((中))((修)復)((土)壤))((的))((方)法))
C:1002	((一)種)((採用)((細)菌))((修)復)((土)壤))((的))((方)法))
C:1003	((一)種)((採用)((特)性)菌))((修)復)((土)壤))((的))((方)法))
C:1004	((本)発)明)((採)用)((細)菌))((製)造)((土)壤))((進)行)((修)復))
C:1005	((本)発)明)((採)用)((特)性)菌))((製)造)((土)壤))((的))((方)法))
C:1006	((通)過)((本)方)法)((採)用)((特)性)菌))((培)養)出)産))((可)有)效))((用)于)((汚)染)((土)壤))((的))((生)物))((修)復))
C:1007	((而)且)((特)性)菌))((中)子))((的))((製)造)((方)法))
C:1008	((而)且)((特)性)菌))((的))((製)造)((方)法))
C:1009	((經)過)((特)性)菌))((之)後)((再)次)培)養))((一)次))((以))((保)持)((細)菌))((含)量))
...	...

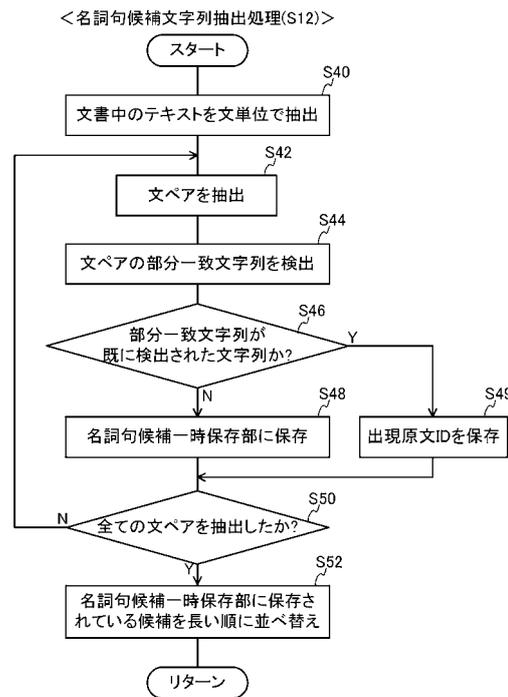
【 図 4 】



【 図 5 】

＜解析対象の特許文書＞	
原文ID	原文
Q301	农民为了获得高产而过量施用氮肥, 导致硝酸盐积累, 蔬菜品质下降。
Q302	一种采用棉秆高得率浆液生物转化菌种N4修复土壤的方法。
Q303	由于普通细菌培养基中含有铵盐, 在好氧培养过程中一部分铵态氮被合成为细菌自身细胞物质。
Q304	本发明采用棉秆高得率浆液生物转化菌种N4对土壤进行修复。
Q305	本发明所使用的棉秆高得率浆液生物转化菌种N4的制备方法为
Q306	每隔4h用血球板做菌的计数, 即可获得棉秆高得率浆液生物转化菌种N4
Q307	用废液培养的转化菌种N4可明显降低土壤中的硝态氮
Q308	通过本方法所培养出的转化菌种N4可有效用于污染土壤的生物修复。
Q309	要求的溶解氧浓度较低, 在一定范围内反硝化率不受DO值的影响。
...	...

【 図 6 】



【 図 7 】

(a)

名詞句候補ID	CANDI-1
名詞句候補文字列	采用棉秆高得率浆废液生物转化菌种N4
出現原文ID	Q302、Q304

(b)

名詞句候補ID	CANDI-2
名詞句候補文字列	棉秆高得率浆废液生物转化菌种N4
出現原文ID	Q302、Q304、Q305、Q306

(c)

名詞句候補ID	CANDI-3
名詞句候補文字列	转化菌种N4
出現原文ID	Q302、Q304、Q305、Q306、Q307、Q308

【 図 8 】

(a)

名詞句候補ID	CANDI-1
名詞句候補文字列	采用棉秆高得率浆废液生物转化菌种N4
出現原文ID	Q302、Q304

① 文書から名詞句候補文字列を含む文を取り出す

(b)

原文ID	原文
Q302	一种采用棉秆高得率浆废液生物转化菌种N4修复土壤的方法
Q304	本发明采用棉秆高得率浆废液生物转化菌种N4对土壤进行修复

② 名詞句候補文字列を暫定名詞(仮の意味記号) [**]で置換

(c)

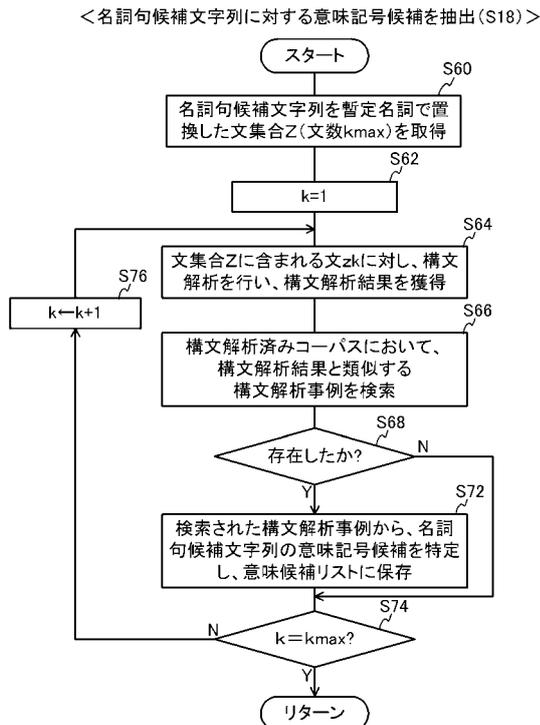
原文ID	原文
Q302	一种[**]修复土壤的方法
Q304	本发明[**]对土壤进行修复

③ 構文解析

(d)

原文ID	原文
Q302	(((一种)([**])(修复)(土壤))的)(方法))
Q304	(((本)(发明)([**])(对)(土壤))(进行)(修复))

【 図 9 】



【 図 10 】

(a)

名詞句候補ID	CANDI-2
名詞句候補文字列	棉秆高得率浆废液生物转化菌种N4
出現原文ID	Q302、Q304、Q305、Q306

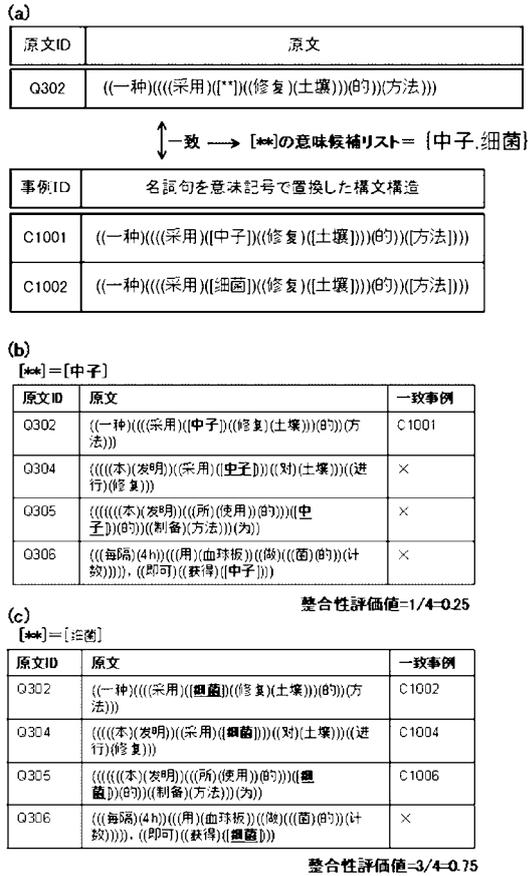
(b)

原文ID	原文
Q302	一种采用棉秆高得率浆废液生物转化菌种N4修复土壤的方法
Q304	本发明采用棉秆高得率浆废液生物转化菌种N4对土壤进行修复
Q305	本发明所使用的棉秆高得率浆废液生物转化菌种N4的制备方法为
Q306	每隔4h用血球板做菌的计数,即可获得棉秆高得率浆废液生物转化菌种N4

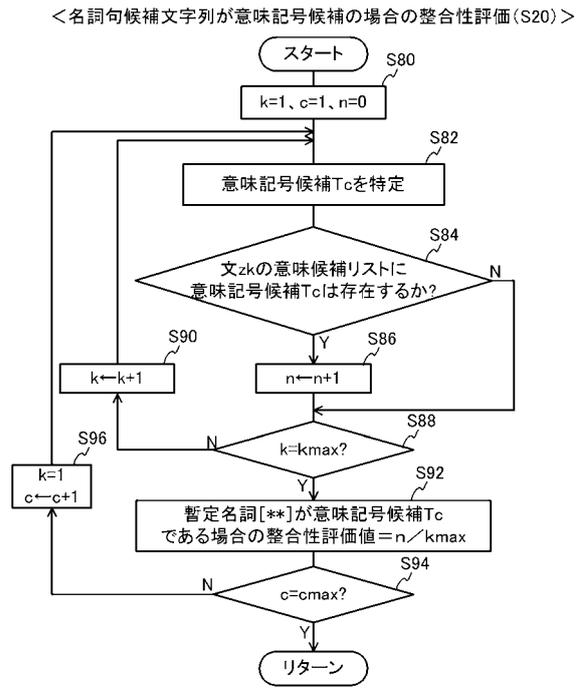
(c)

原文ID	原文
Q302	((一种)((采用)([**])(修复)(土壤))的)(方法))
Q304	(((本)(发明)((采用)([**])(对)(土壤))(进行)(修复))
Q305	((((本)(发明)((所)(使用))的))([**])的)(制备)(方法))(为))
Q306	(((每隔)(4h)((用)(血球板))((做)(菌的)(计数))))((即可)(获得)([**]))

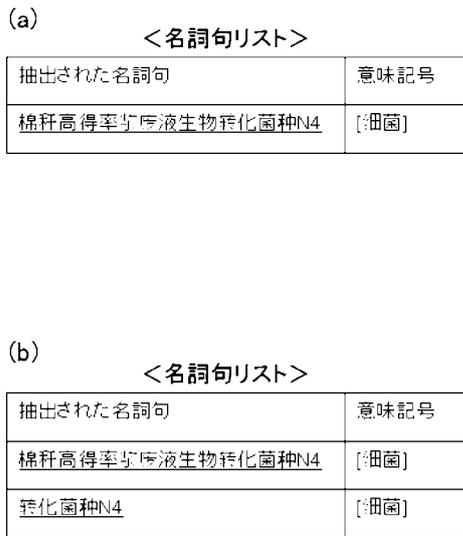
【 図 1 1 】



【 図 1 2 】



【 図 1 3 】



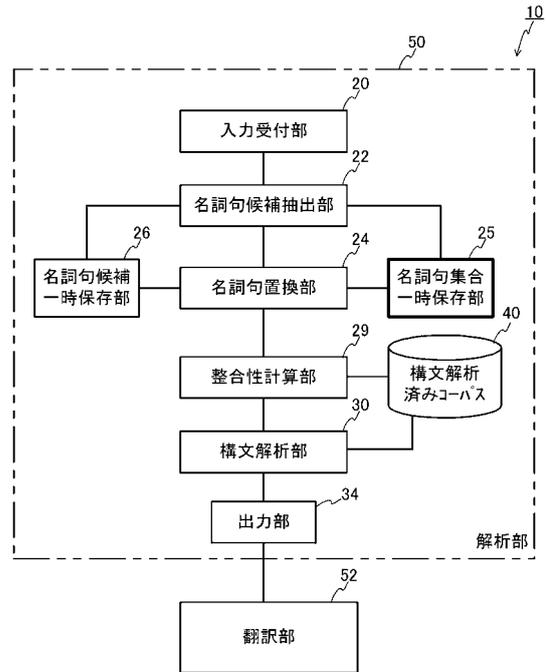
【 図 1 4 】



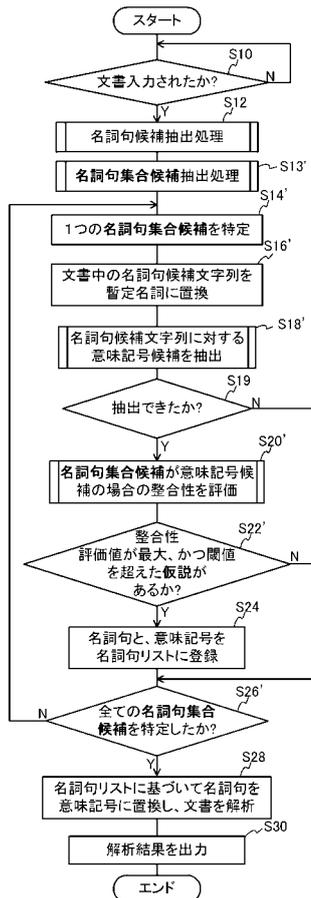
【 図 1 5 】

原文ID	名詞句を意味記号で置き換え済み原文
Q301	农民为了获得高产而过量施用氮肥, 导致硝酸盐积累, 蔬菜品质下降。
Q302	一种采用[细菌] 修复土壤的方法
Q303	由于普通细菌培养基中含有硝酸盐, 在好氧培养过程中一部分硝态氮被合成为细菌自身细胞物质,
Q304	本发明采用[细菌] 对土壤进行修复
Q305	本发明所使用的[细菌] 的制备方法为
Q306	每隔4h用血球板做菌的计数, 即可获得[细菌]
Q307	用原液培养的[细菌]可明显降低土壤中的硝态氮
Q308	通过本方法所培养出的[细菌]可有效用于污染土壤的生物修复。
Q309	要求的溶解氧浓度较低, 在一定范围内反硝化率不受DO值的影响。

【 図 1 6 】



【 図 1 7 】



【 図 1 8 】

<解析対象の特許文書>

原文ID	原文
Q302	一种经由棉秆高得率原液生物转化菌种N4修复遭硝酸盐严重污染的土壤的方法
Q304	本发明利用棉秆高得率原液生物转化菌种N4对土壤进行修复
Q307	用原液培养的菌种可明显降低遭硝酸盐严重污染的土壤中的硝态氮
Q308	通过本方法所培养出的菌种可有效用于遭硝酸盐严重污染的土壤的生物修复
⋮	⋮

【 図 1 9 】

(a)

原文ID	原文
Q302	一种经由棉秆高得率浆液生物转化菌种N4修复遭硝酸盐严重污染的土壤的方法
Q304	本发明利用棉秆高得率浆液生物转化菌种N4对土壤进行修复
Q307	用废液培养的菌种可明显降低遭硝酸盐严重污染的土壤中的硝态氮
Q308	通过本方法所培养出的菌种可有效用于遭硝酸盐严重污染的土壤的生物修复。
⋮	⋮

(b)

名詞句候補ID	CANDI-1
名詞句候補文字列	采用棉秆高得率浆液生物转化菌种N4
出現原文ID	Q302, Q304

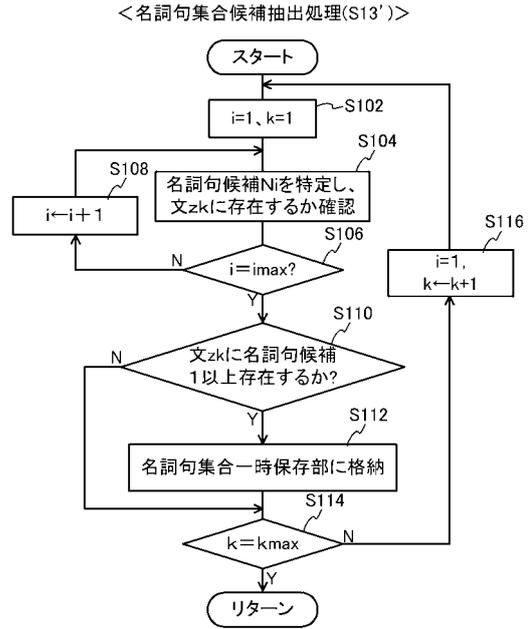
(c)

名詞句候補ID	CANDI-2
名詞句候補文字列	遭硝酸盐严重污染的土壤
出現原文ID	Q302, Q307, Q308

(d) <名詞句集合一時保存部>

名詞句集合ID	名詞句集合
1	CANDI-1, CANDI-2
2	CANDI-1
3	CANDI-2

【 図 2 0 】



【 図 2 1 】

(a)

原文ID	原文
Q302	一种经由[**1]修复[**2]的方法
Q304	本发明利用[**1]对土壤进行修复
Q307	用废液培养的菌种可明显降低[**2]中的硝态氮
Q308	通过本方法所培养出的菌种可有效用于[**2]的生物修复。

(b)

原文ID	原文
Q302	((一种)((理由)([**1])(修复)([**2]))的)(方法))
Q304	((本(发明)(利用)([**1]))对(土壤))(进行(修复))
Q307	((用(废液)(培养)的)(菌种)((可)(明显))降低)(([**2])中)的)(硝态氮))
Q308	((通过(本(方法))所培养出的)(菌种)((可)(有效))用于)(([**2])的)(生物(修复)))

【 図 2 2 】

(a)

原文ID	原文
Q302	((一种)((理由)([**1])(修复)([**2]))的)(方法))

(b)

事例ID	名詞句候補を意味記号で置換された構文構造
C1001	((一种)((理由)([中子])(修复)([金属]))的)(方法))
C1002	((一种)((理由)([細菌])(修复)([土壤]))的)(方法))

(c) [**1]の意味候補リスト(GSL) = {中子, 細菌}
 [**2]の意味候補リスト(GSL) = {土壤, 金属}

(d) 仮説1: [**1]は[中子]、[**2]は[土壤]
 仮説2: [**1]は[中子]、[**2]は[金属]
 仮説3: [**1]は[細菌]、[**2]は[土壤]
 仮説4: [**1]は[細菌]、[**2]は[金属]

【 図 2 3 】

(a) 仮説1の整合性評価値 = 1 / 4 = 0.25

原文ID	原文	一致事例
Q302	(((一种)((自由)([中子])([修复])([土壤]))的方法))	×
Q304	(((本)([明])([利用])([中子])([土](壤))([进行])([修复]))	×
Q307	(((用)([溶液])([培养])([菌种])([可](明)显)([降]低)([土](壤)([中])([的])([明](确)([气]))	×
Q308	(((通)过)([本](方)法)([所](培)养)出)([的])([菌]种)([可](有)效)([用]于)([土](壤)([的])([生]物)([修]复))	C1006

(b) 仮説2の整合性評価値 = 1 / 4 = 0.25

原文ID	原文	一致事例
Q302	(((一种)((自由)([中子])([修复])([金属]))的方法))	C1001
Q304	(((本)([明])([利用])([中子])([土](壤))([进行])([修复]))	×
Q307	(((用)([溶液])([培养])([菌种])([可](明)显)([降]低)([金](属)([中])([的])([明](确)([气]))	×
Q308	(((通)过)([本](方)法)([所](培)养)出)([的])([菌]种)([可](有)效)([用]于)([金](属)([的])([生]物)([修]复))	×

(c) 仮説3の整合性評価値 = 3 / 4 = 0.75

原文ID	原文	一致事例
Q302	(((一种)((自由)([细菌])([修复])([土壤]))的方法))	C1002
Q304	(((本)([明])([利用])([细菌])([土](壤))([进行])([修复]))	C1004
Q307	(((用)([溶液])([培养])([菌种])([可](明)显)([降]低)([土](壤)([中])([的])([明](确)([气]))	×
Q308	(((通)过)([本](方)法)([所](培)养)出)([的])([菌]种)([可](有)效)([用]于)([土](壤)([的])([生]物)([修]复))	C1006

(d) 仮説4の整合性評価値 = 1 / 4 = 0.25

原文ID	原文	一致事例
Q302	(((一种)((自由)([细菌])([修复])([金属]))的方法))	×
Q304	(((本)([明])([利用])([细菌])([土](壤))([进行])([修复]))	C1004
Q307	(((用)([溶液])([培养])([菌种])([可](明)显)([降]低)([金](属)([中])([的])([明](确)([气]))	×
Q308	(((通)过)([本](方)法)([所](培)养)出)([的])([菌]种)([可](有)效)([用]于)([金](属)([的])([生]物)([修]复))	×

【 図 2 4 】

< 名詞句リスト >

抽出された名詞句	意味記号
棉秆高得率及尿液生物转化菌种N4	[细菌]
遭硝酸盐严重污染的土壤	[土壤]

【 図 2 5 】

