

(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2004-334729

(P2004-334729A)

(43) 公開日 平成16年11月25日(2004.11.25)

(51) Int. Cl.⁷
G06F 17/27

F I
G06F 17/27

テーマコード(参考)
5B091

審査請求 有 請求項の数 6 O L (全 10 頁)

<p>(21) 出願番号 特願2003-132527(P2003-132527)</p> <p>(22) 出願日 平成15年5月12日(2003.5.12)</p> <p>特許法第30条第1項適用申請有り 2002年11月12日 社団法人情報処理学会発行の「情報処理学会研究報告 情処研報Vol. 2002, No. 104」に発表</p>	<p>(71) 出願人 301022471 独立行政法人情報通信研究機構 東京都小金井市貫井北町4-2-1</p> <p>(74) 代理人 100085338 弁理士 赤澤 一博</p> <p>(74) 代理人 100118245 弁理士 井上 敬子</p> <p>(72) 発明者 木田 敦子 東京都小金井市貫井北町4-2-1 独立行政法人通信総合研究所内</p> <p>(72) 発明者 山本 英子 東京都小金井市貫井北町4-2-1 独立行政法人通信総合研究所内</p>
--	---

最終頁に続く

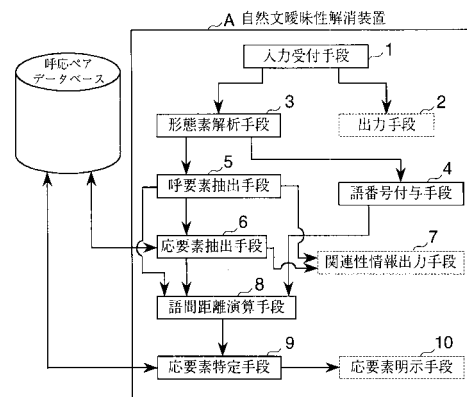
(54) 【発明の名称】 自然文曖昧性解消装置、及び自然文曖昧性解消プログラム

(57) 【要約】

【課題】 日本語文における「係り結び」を形成する呼応ペアを蓄積したデータベースを利用することで、入力済の日本語自然文からなるテキストデータにおいて、真に正しいと推測される呼応関係を見出し、当該自然文の曖昧性を解消する。

【解決手段】 入力済のテキストデータに基づく日本語自然文中において、複数の呼応関係が出現してその文章の意味が曖昧となった場合に、当該自然文中の呼要素と応要素との語間距離のうち、呼応ペアデータベースDB1に格納されている当該呼要素と当該応要素との平均的な語間距離に対応する呼応関係にある応要素を機械的に選出するようにした。

【選択図】 図1



【特許請求の範囲】

【請求項 1】

日本語の自然文中において共起する 2 つの語のうち係り結びを形成する呼要素と応要素とを対にした呼応ペアの集合を、当該呼応ペアを含む元の自然文中における呼要素と応要素との平均語間距離を付加した状態で格納してある呼応ペアデータベースを検索することによって、入力された日本語自然文テキストにおける呼応関係の曖昧性を解消するコンピュータからなるものであって、

日本語テキストデータの入力を受け付ける入力受付手段と、

前記入力受付手段で受け付けた日本語テキストデータについて形態素解析を実行する形態素解析手段と、

前記形態素解析手段で形態素解析された各形態素に先頭から順に語番号を付与する語番号付与手段と、

前記形態素解析手段で形態素解析された形態素から所定の呼要素を抽出する呼要素抽出手段と、

前記呼応ペアデータベースから、前記呼要素抽出手段で抽出した呼要素に対応する応要素を検索し、その検索結果に基づいて形態素解析手段で形態素解析された形態素のうち応要素に該当するものを抽出する応要素抽出手段と、

前記呼要素抽出手段で抽出した応要素と前記応要素抽出手段で抽出した応要素との語間距離を前記語番号付与手段で付与された語番号に基づいて算出する語間距離演算手段と、

前記語間距離演算手段による演算結果に対応する平均語間距離を有する一の応要素を前記呼応ペアデータベースから特定する応要素特定手段とを具備していることを特徴とする自然文曖昧性解消装置。

【請求項 2】

前記入力受付手段で受け付けた日本語テキストデータを出力する出力手段と、

前記呼要素抽出手段で抽出した呼要素と前記応要素抽出手段で抽出した応要素との関連性を示す情報を出力する関連性情報出力手段と、

前記応要素特定手段で特定した応要素を明示する応要素明示情報を出力する応要素明示手段とをさらに具備している請求項 1 記載の自然文曖昧性解消装置。

【請求項 3】

前記出力手段、関連性情報出力手段、応要素明示手段が、それぞれディスプレイ等の表示装置に日本語テキストデータ、関連性情報、応要素明示情報を表示出力するものである請求項 2 記載の自然文曖昧性解消装置。

【請求項 4】

日本語の自然文中において共起する 2 つの語のうち係り結びを形成する呼要素と応要素とを対にした呼応ペアの集合を、当該呼応ペアを含む元の自然文中における呼要素と応要素との平均語間距離を付加した状態で格納してある呼応ペアデータベースを検索することによって、入力された日本語自然文テキストにおける呼応関係の曖昧性を解消するコンピュータを、

日本語テキストデータの入力を受け付ける入力受付手段と、

前記入力受付手段で受け付けた日本語テキストデータについて形態素解析を実行する形態素解析手段と、

前記形態素解析手段で形態素解析された各形態素に先頭から順に語番号を付与する語番号付与手段と、

前記形態素解析手段で形態素解析された形態素から所定の呼要素を抽出する呼要素抽出手段と、

前記呼応ペアデータベースから、前記呼要素抽出手段で抽出した呼要素に対応する応要素を検索し、その検索結果に基づいて形態素解析手段で形態素解析された形態素のうち応要素に該当するものを抽出する応要素抽出手段と、

前記呼要素抽出手段で抽出した応要素と前記応要素抽出手段で抽出した応要素との語間距離を前記語番号付与手段で付与された語番号に基づいて算出する語間距離演算手段と、

10

20

30

40

50

前記語間距離演算手段による演算結果に対応する平均語間距離を有する一の応要素を前記呼応ペアデータベースから特定する応要素特定手段とを具備する自然文曖昧性解消装置として機能させることを特徴とする自然文曖昧性解消プログラム。

【請求項 5】

前記コンピュータをさらに、

前記入力受付手段で受け付けた日本語テキストデータを出力する出力手段と、前記呼要素抽出手段で抽出した呼要素と前記応要素抽出手段で抽出した応要素との関連性を示す情報を出力する関連性情報出力手段と、

前記応要素特定手段で特定した応要素を明示する応要素明示情報を出力する応要素明示手段と具備する自然文曖昧性解消装置として機能させる請求項 4 記載の自然文曖昧性解消プログラム。

10

【請求項 6】

前記出力手段、関連性情報出力手段、応要素明示手段を、それぞれディスプレイ等の表示装置に日本語テキストデータ、関連性情報、応要素明示情報を表示出力するように機能させる請求項 5 記載の自然文曖昧性解消プログラム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、コンピュータに入力済の日本語自然文からなるテキストデータについて、その自然文が表す意味の曖昧性を解消する装置及びそのプログラムに関するものである。

20

【0002】

【従来の技術】

中世以前の日本語には、係助詞と文末の活用形とが形態的な呼応関係を持つ「係り結び」の用法が存在したが、「係り結び」が消滅した現代の日本語の文章の場合、述語が文末に置かれるため、文の終末まで進まないとその文章の内容が確定しない。そのため、長文で複雑な内容の文章では、その内容が肯定的なのか否定的なのか、或いは疑問を表しているのかが文末まで読まないことには明らかにならない。すなわち、同一の自然文中に一つの「係り」となる語に対応する「結び」となる語が複数出現した場合、その文の意味を理解する上でどの「結び」語が真に「係り」語に対応しているかを的確に把握することが難しいといえる。ここで、現代日本語の文構造の研究において、現代語ではある種の副詞などが古語の係助詞と似た役割を果たしており、後続要素を予告しているとの示唆がなされている（例えば、非特許文献 1 参照）。例えば「たぶん……だろう」や「おそらく……だろう」などといった組み合わせは、呼応関係を形成する先行要素（呼要素）及び後続要素（応要素）のペアとして内省や直感である程度予測がつくと考えられると指摘されている（例えば、非特許文献 2 参照）。

30

【0003】

【非特許文献 1】

大野 晋，「係り結びの研究」，第 1 版，岩波書店，1993 年 1 月 12 日，p 350 - 351

【非特許文献 2】

益岡 隆志，「モダリティの文法」，第 1 版，くろしお出版，1991 年 5 月 25 日，p 29 - 46

40

【0004】

【発明が解決しようとする課題】

ところが、このような呼応ペアについては未だ体系立てた研究がなされておらず、上述した文献や教科書等においても少数の呼応ペアが例示されるに留まっているのが現状である。すなわち、現代日本語における「係り結び」の研究では、内省や直感では予測し得ないような呼応ペアが不足しているために、ある語とそれと共に現れる（共起する）語とが本当に呼応関係にあるのか否かを明らかにするには基礎的データが余りにも少ないといわざるを得ない。したがって、ユーザが自分でコンピュータに入力した日本語の文章が正しい

50

呼応関係にある「係り結び」の表現を行っているのか検証を行ったり、或いはユーザが他者から受け取った日本語自然文がの意味を理解するには、そのユーザ自身の直感や内省に頼るしかないのが現状である。このような問題は、入力された日本語自然文が長文であればあるほど顕著となる。

【0005】

そこで本発明は、以上のような問題に鑑みて、日本語文における「係り結び」を形成する呼応ペアを蓄積したデータベースを利用することで、入力済の日本語自然文からなるテキストデータにおいて、真に正しいと推測される呼応関係を見出し、当該自然文の曖昧性を解消しようとするものである。

【0006】

【課題を解決するための手段】

すなわち、本発明は、日本語の自然文中において共起する2つの語のうち係り結びを形成する呼要素と応要素とを対にした呼応ペアの集合を、当該呼応ペアを含む元の自然文中における呼要素と応要素との平均語間距離を付加した状態で格納してある呼応ペアデータベースを検索することによって、入力された日本語自然文テキストにおける呼応関係の曖昧性を解消するコンピュータからなる自然文曖昧性解消装置、並びに当該コンピュータを自然文曖昧性解消装置として機能させるためのプログラムである。

【0007】

図1に基本的な機能構成図を実線で示すように、この自然文曖昧性解消装置Aは、日本語テキストデータの入力を受け付ける入力受付手段1と、入力受付手段1で受け付けた日本語テキストデータについて形態素解析を実行する形態素解析手段3と、形態素解析手段3で形態素解析された各形態素に先頭から順に語番号を付与する語番号付与手段4と、形態素解析手段3で形態素解析された形態素から所定の呼要素を抽出する呼要素抽出手段5と、呼応ペアデータベースDB1から、前記呼要素抽出手段5で抽出した呼要素に対応する応要素を検索し、その検索結果に基づいて形態素解析手段3で形態素解析された形態素のうち応要素に該当するものを抽出する応要素抽出手段6と、呼要素抽出手段5で抽出した呼要素と前記応要素抽出手段6で抽出した応要素との語間距離を前記語番号付与手段4で付与された語番号に基づいて算出する語間距離演算手段8と、語間距離演算手段8による演算結果に対応する平均語間距離を有する応要素を前記呼応ペアデータベースDB1から特定する応要素特定手段9とを具備していることを特徴とするものである。

【0008】

ここで、呼要素とは、現代日本語文において「係り結び」を構成する2単語のうち「係り」に該当する語を意味する。呼要素となり得る単語には、例えば『基礎日本語文法 改訂版』（益岡隆志、田窪行則、くろしお出版、1992年）の分類によると、「提題助詞」、「取り立て助詞」、「陳述の副詞」が該当する。

具体的な対象語としては、「こそ」、「しか」、「さえ」、「は」、「も」、「ばかり」、「のみ」、「すら」、「なら」、「くらい（ぐらい）」、「だけ」、「なんて」、「けっして（決して）」、「おそらく（恐らく）」、「たぶん（多分）」、「ぜひ（是非）」、「まるで」、「もし」、「きっと」等の語を挙げることができるが、必ずしもこれらに限定されるわけではい。そして、「結び」に該当する「応要素」は、現代日本語自然文において「呼要素」が出現した場合にそれと同時に出現する単語である。本発明では、「呼要素」と「応要素」とが同一文中に同時に出現することを「共起する」と定義するとともに、この「共起」する「呼要素」と「応要素」との組み合わせを「共起ペア」と呼び、その「共起ペア」のうち「呼要素」と「応要素」とがこの順で出現することを「呼応する」と定義するとともに、この「呼応」する「呼要素」と「応要素」との組み合わせを特に「呼応ペア」と呼ぶものとする。そして、呼応ペアデータベースDB1には、多数の現代日本語の自然文データを原文データとして抽出した呼要素と応要素とが対をなす「呼応ペア」として格納されている。さらに、さらに、これら呼応ペアのデータは、呼要素ごとに分類されており、各呼応ペアには、呼応ペアデータベースを生成する基礎となった多数の原文において当該呼要素と応要素との語間距離の平均値が付与されている。この語間距離は

10

20

30

40

50

、原文を形態素解析した場合に、隣接する形態素同士の距離を1として求められている。

【0009】

したがって、このように自然文曖昧性解消装置Aとしてコンピュータを機能させることで、入力済の日本語自然文から「係り結び」すなわち呼応関係にある語を選び出す。そして、呼要素とそれに対応する一以上の応要素との各語間距離と、呼応ペアデータベースDB1に格納されている当該呼要素と当該応要素との平均距離とを比較して、対応する1つの応要素を選出することで、その自然文における呼応関係を明確にして、ユーザによる文章の検証や理解を補助することができる。ここで、「入力」とは、キーボードやマウスやタブレット等の入力装置を用いてコンピュータに文字データを入力することや、他のコンピュータ等の外部の機器等から入力を受けることなどを意味する。また、「平均語間距離に対応する語間距離を有する応要素」とは、平均語間距離と同一又は近似する語間距離を有する応要素や、平均語間距離と一定の演算式関係にある語間距離を有する応要素など、平均語間距離と所定の関係を有する語間距離を有する応要素を意味している。

10

【0010】

特に、この自然文曖昧性解消装置Aを利用するユーザに対する文章理解や文章検証の補助をより利便性の高い者とするためには、コンピュータを、入力受付手段1で受け付けた日本語テキストデータを出力する出力手段2と、呼要素抽出手段5で抽出した呼要素と前記応要素抽出手段6で抽出した応要素との関連性を示す情報を出力する関連性情報出力手段7と、応要素特定手段9で特定した応要素を明示する応要素明示情報を出力する応要素明示手段10としてさらに機能させることが好ましく、さらに視覚的に使い勝手もよいものとするには、出力手段2、関連性情報出力手段7、応要素明示手段10を、それぞれディスプレイ等の表示装置に日本語テキストデータ、関連性情報、応要素明示情報を表示出力するものとするといよい。

20

【0011】

【発明の実施の形態】

以下、本発明の一実施形態を、図面を参照して説明する。

【0012】

この実施形態に係る自然文曖昧性解消装置Aは、図1に機能構成を示したように、日本語自然文から「係り結び」を形成する呼要素と応要素とのペアである呼応ペアデータを収集した呼応ペアデータベースDB1を利用して、入力済の日本語自然文を表すテキストデータにおける一の呼要素と一以上の応要素とを見出し、その文章の意味を示す一の応要素を特定して、当該日本語自然文の検証やユーザによる理解を補助するためのものである。この自然文曖昧性解消装置Aは、呼応ペアデータベースDB1を内蔵し又は外部に接続して検索することができる状態にあるコンピュータにより構成されるものである。このコンピュータは、図2に概略的な機器構成図を示すように、バス線等で電気的に接続されたCPU101、メモリ102等の内部機器を有しており、これらにハードディスク等の記憶装置103、ディスプレイ等の表示装置104、キーボードやマウス等の入力装置105、各種通信インターフェース106等の外部機器を備えた通常のパーソナルコンピュータ等からなり、例えば外部に呼応ペアデータベースDB1を通信線を介して接続してある。なお、この呼応ペアデータベースDB1に格納されるデータは、ハードディスク等の記憶装置に格納させることもできる。

30

40

【0013】

そして、記憶装置103に記録した自然文曖昧性解消プログラムをCPU101が読み出してメモリ102に記憶させ、当該CPU101が前記プログラムに従った処理を行い、メモリ102、ハードディスク等の記憶装置103、モニタ等の表示装置104、キーボードやマウス等の入力装置105、各種通信インターフェース106等の機器を駆動させることによって、このコンピュータは、自然文曖昧性解消装置Aとして機能する。ここで、自然文曖昧性解消装置Aの機能とは、図1に示したように、入力受付手段1、出力手段2、形態素解析手段3、語番号付与手段4、呼要素抽出手段5、応要素抽出手段6、関連性情報出力手段7、語間距離演算手段8、応要素特定手段9、応要素明示手段10を指す

50

。特に、入力受付手段 1 は、キーボードやマウス等の入力装置 105 からユーザによって入力されたデータを受け付けてメモリ 102 に格納する機能を有するものであり、出力手段 2、意味情報出力手段 4 及び確定意味情報出力手段 6 は、ディスプレイ等の表示装置 104 に文字を表示させる機能を有するものである。

【0014】

また、呼応ペアデータベース DB 1 は、例えば新聞記事等から収集した膨大な日本語自然文のデータを原文データとしてこれを形態素解析し、その形態解析後のデータに基づいて同一文中で共起した一对の語（共起ペア）を補完類似度の演算により求め、その共起ペアのなかから「呼要素」「応要素」の順で原文データ中に出現したものについて信頼度を演算することによって得られた「呼応ペア」を格納したものである。ここで「信頼度」とは、呼応ペアごとに着目して、ある呼要素が出現する原文数に対する当該呼応ペアが出現する原文数の割合を意味し、最終的に得られた呼応ペアはこの信頼度が所定値以上を示したものである本実施形態では原文データとして「毎日新聞記事データ」と「日経新聞記事データ」の各 10 年分を利用している。ここで、呼要素は、本実施形態では「提題助詞」、「取り立て助詞」、「陳述の副詞」に分類される語、具体的には、「こそ」、「しか」、「さえ」、「は」、「も」、「ばかり」、「のみ」、「すら」、「なら」、「くらい（ぐらい）」、「だけ」、「なんて」、「けっして（決して）」、「おそらく（恐らく）」、「たぶん（多分）」、「ぜひ（是非）」、「まるで」、「もし」、「きっと」を採用して予め漸進的文解釈支援プログラムに記述してあり、呼応ペアデータベース DB 1 ではこれら呼要素ごとに呼応ペアが分類されている。そして、この呼応ペアデータベース DB 1 においては、各原文データを形態素解析して得られる各形態素に先頭から語番号を付与しておき、各呼応ペアについて、当該呼応ペアが出現した原文データにおいて応要素の語番号と呼要素の語番号との差を語間距離とし、その呼応ペアが出現した全原文データにおける語間距離の平均値を「平均語間距離」として付与した状態でデータを格納してある。なお、ここでは隣接する形態素間の語間距離を 1 としている。図 3 に、呼要素「決して」に関する呼応ペアデータベース DB 1 の内容の一部を示す。同図中欄には、呼要素「決して」に対応する応要素が記述されており、各応要素の右欄には、呼要素「決して」と当該応要素との平均語間距離が記述されている。

【0015】

自然文曖昧性解消装置 A の動作は、上述した入力受付手段 1、出力手段 2、形態素解析手段 3、語番号付与手段 4、呼要素抽出手段 5、応要素抽出手段 6、関連性情報出力手段 7、語間距離演算手段 8、応要素特定手段 9、応要素明示手段 10 に対応して次のように行われる。

【0016】

この自然文曖昧性解消装置 A は、図 4 のフローチャートに示すように、まず、入力受付手段 1 の機能によってキーボード等の入力装置 105 から日本語自然文のテキストデータの入力を受け付けると (S1)、そのテキストデータを一時的にメモリ 102 に格納しつつ出力手段 2 の機能によって即時的にディスプレイ等の表示装置 104 に当該文字列を表示していく (S2)。次に、入力を受け付けたテキストデータをメモリ 102 から形態素解析手段 3 が読み出して形態素解析を実行し、当該テキストデータを品詞分解する (S3)。ここで、形態素解析プログラムには、例えば日本語形態素解析プログラムとして、「JUMAN」(<http://www.lab25.kuee.kyoto-u.ac.jp/nl-resource/juman.html>) や「茶筌」(<http://chasen.aist-nara.ac.jp/>) を用いることができる。そして、語番号付与手段 4 の機能によって、各形態素に対して文頭から順に語番号を付与する (S4)。また、呼要素抽出手段 5 の機能により、形態素解析手段 3 の形態素解析結果から得られた形態素から上述した呼要素に該当する形態素を抽出する (S5)。この抽出された形態素は、一時的にメモリ 102 に格納される。さらに、応要素抽出手段 6 の機能によって、先に抽出してある呼要素について呼応ペアデータベース DB 1 を検索して、当該呼要素に対応する応要素に該当する応要素となる一以上の形態素を、ステップ S3 で形態素解析され

た形態素から抽出する（S6）。この抽出された応要素も一時的にメモリ102に格納される。ここで、関連性情報出力手段7の機能により、メモリ102から呼要素と応要素とを読み出して、関連性情報として呼要素を明示するマーク（一例として、当該呼要素を印で囲むマークや当該呼要素に付された下線）と、当該マークから応要素に向けて延びる矢印マークとを表示装置104に出力して、ステップS2で表示装置104に表示させた文字列に重ね合わせて表示させる（S7）。その一方、語間距離演算手段8の機能によって、メモリ102に格納されている各応要素の語番号から応要素の語番号を差し引く演算を実行し、各応要素についての語間距離を算出する（S8）。この演算結果である語間距離も対応する応要素に関連づけてメモリ102に格納される。さらに、応要素特定手段9の機能によって、各応要素に関する語間距離と、呼応ペアデータベースDB1に格納されている当該応要素の平均語間距離とを比較し、ここでは平均語間距離に最も近い値を示した語間距離に対応する応要素を、ステップS5で抽出した呼要素に対応する応要素として確定する（S9）。最後に、応要素明示手段10の機能によって、ステップS9で確定された応要素を明示するための応要素明示情報を表示装置104に表示出力する（S10）。ここでは応要素明示情報を表示する処理の一例として、当該応要素に下線を引くとともに、呼要素から当該応要素へ延びる矢印マークと他の応要素へ延びる矢印マークとを区別するために、該当する矢印マークを実線で表して他の矢印マークを破線で表す方法を採用しているが、矢印マークを色分けするなどの他の処理を行ってもよい。

10

【0017】

ここで、図5を参照して、入力後にディスプレイ等の表示装置104に表示された日本語自然文における曖昧性解消の具体例を挙げて説明する。なお、以下の例は、呼要素「決して」に関するものである。まず、図5(a)に示すように、「道のりは決して楽ではない。」という日本語自然文が入力装置105を利用して入力され表示装置104に表示される場合、上述した自然文曖昧性解消装置Aの動作のフローチャートに従って、この自然文について形態素解析が行われ各形態素に語番号が付与されたうえで、呼要素「決して」が抽出される。なお、同図に示した「|」（縦線）は、隣接する形態素間の区切りを示している。そして、図3に示したような呼応ペアデータベースDB1を参照して、呼要素「決して」に対応する応要素として「楽」「で」「は」「ない」が抽出されると、関連性情報として呼要素「決して」から各応要素「楽」「で」「は」「ない」に対して矢印が付され、これが表示装置104に表示される。次いで、呼要素「決して」と各応要素との語間距離が算出されて、その語間距離に最も近い平均語間距離を有する応要素「ない」が呼応ペアデータベースDB1から確定されたうえで、前記関連性情報のうち「決して」から「ない」に向かう矢印が応要素明示情報として太線で表示装置104に表示される。したがって、この自然文曖昧性解消装置Aを使用するユーザは、入力した又は入力された自然文中に正しい「係り結び」表現が用いられており、そのなかでも「決して」と「ない」とが正しい呼応関係にあることが分かる。なお、このような短い自然文に限らず、例えば図5(b)に示すような比較的長い自然文の場合であっても、上述した手順と同様にして、「決して」と「ない」とが正しい呼応関係にあることを明示することができる。

20

30

【0018】

以上のようにして、本実施形態では、入力済の日本語自然文のテキストデータ中から、当該文中において真に呼応するという妥当する呼要素と応要素とを明示して、複数の呼応関係による自然文の曖昧さを解消することができる。そのため、この自然文曖昧性解消装置Aを利用するユーザが、自分で入力したテキストデータに基づく文章が正しい呼応関係を有しているか否かを検証したり、他者から受け取ったテキストデータに基づく文章の真の意味を理解することに大いに役立つことになる。このことは、この自然文曖昧性解消装置Aや自然文曖昧性解消プログラムを、日本語入力装置や日本語入力プログラムにおいて自然文の検証装置又は検証プログラムの一部として応用することができることになる。

40

【0019】

なお、本発明は上述した実施形態に限られるものではなく、「決して」以外の呼要素についても同様にして、入力後の日本語自然文から正しい「係り結び」関係にある呼要素と応

50

要素とを抽出し、その日本語自然文の曖昧性を解消することができる。また、例えば、呼要素「きっと」に対応する応要素「違い」と応要素「ない」とを組み合わせた複数の形態素からなる応要素「違いない」を生成しておき、これとその平均語間距離とを呼応ペアデータベースに格納しておけば、上述した実施形態と同様に、より実用的に真の呼要素と応要素との組み合わせを入力済のテキストデータから抽出することができる。

【0020】

その他、各部の具体的構成や機能についても上記実施形態に限られるものではなく、本発明の趣旨を逸脱しない範囲で種々変形が可能である。

【0021】

【発明の効果】

以上に詳述したように、本発明に係る自然文曖昧性解消装置又はそのための自然文曖昧性解消プログラムによれば、入力済のテキストデータに基づく日本語自然文中において、複数の呼応関係が出現してその文章の意味が曖昧となった場合に、当該自然文中の呼要素と応要素との語間距離のうち、当該呼要素と当該応要素との平均的な語間距離に対応する呼応関係にある応要素を機械的に選出するようにしているので、最も確からしい呼応関係が明らかとなって曖昧性が解消され、ユーザが自然文の意味を理解したり自然文において正しい呼応関係を用いているかを検証したりするのに非常に役立つ。また、日本語入力プログラムや日本語検証プログラム等に適用することで、本発明は、それらの信頼性を向上することも可能である。

10

【図面の簡単な説明】

【図1】本発明及びその一実施形態に係る自然文曖昧性解消装置の機能構成を概略的に示す図。

【図2】同実施形態に係る自然文曖昧性解消装置を構成するコンピュータの概略的機器構成図。

【図3】同実施形態において利用される呼応ペアデータの一部を示す図。

【図4】同自然文曖昧性解消装置の動作の概観を示すフローチャート。

【図5】同自然文曖昧性解消装置を利用したテキストデータの曖昧性解消の一具体例を示す図。

20

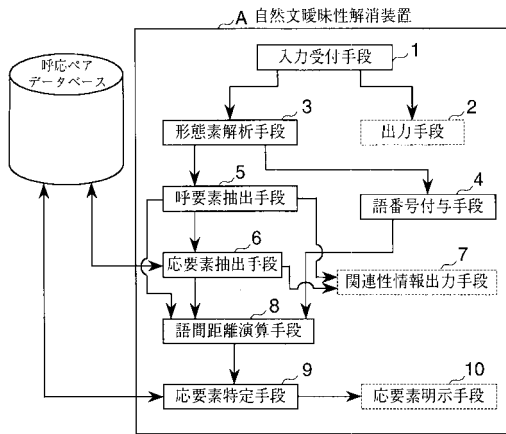
【符号の説明】

- A ... 自然文曖昧性解消装置
- DB 1 ... 呼応ペアデータベース
- 1 ... 入力受付手段
- 2 ... 出力手段
- 3 ... 形態素解析手段
- 4 ... 語番号付与手段
- 5 ... 呼要素抽出手段
- 6 ... 応要素抽出手段
- 7 ... 関連性情報出力手段
- 8 ... 語間距離演算手段
- 9 ... 応要素特定手段
- 10 ... 応要素明示手段

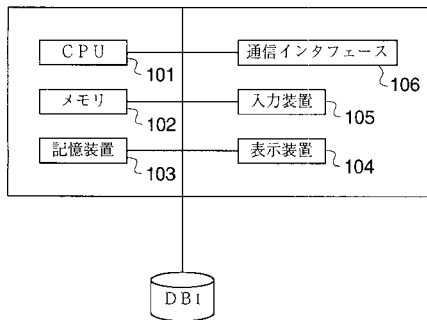
30

40

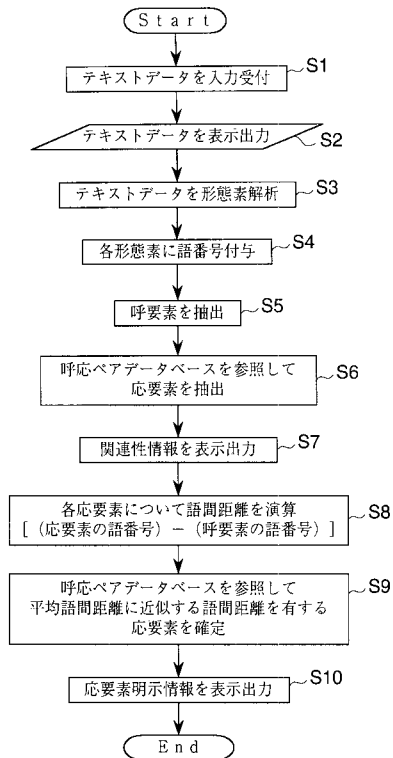
【 図 1 】



【 図 2 】



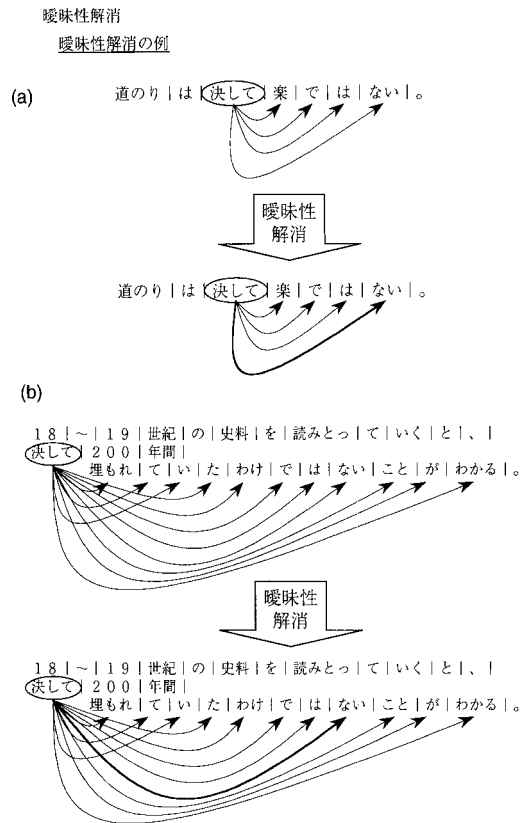
【 図 4 】



【 図 3 】

呼要素候補 (対象語)	応要素候補	距離平均
決して (副詞-一般)	ない (助動詞/特殊・ナイ)	6.965429
決して (副詞-一般)	だ (助動詞/特殊・ダ)	9.940159
決して (副詞-一般)	と (助動詞-格助詞-引用)	14.58871
決して (副詞-一般)	こと (名詞-非自立-一般)	11.709452
決して (副詞-一般)	ない (形容詞-自立/形容詞)	7.116818
決して (副詞-一般)	て (助詞-接続助詞)	16.992443
決して (副詞-一般)	もの (名詞-非自立-一般)	9.02517
決して (副詞-一般)	ある (助動詞/五段・ラ行アル)	16.159379
決して (副詞-一般)	の (名詞-非自立-一般)	16.732897
決して (副詞-一般)	わけ (名詞-非自立-一般)	8.07582
決して (副詞-一般)	いう (動詞-自立/五段)	16.143636
決して (副詞-一般)	ん (助動詞/不変化型)	8.781193
決して (副詞-一般)	ます (助動詞/特殊・マス)	13.294299
決して (副詞-一般)	よう (名詞-非自立-助動詞語幹)	15.973487
決して (副詞-一般)	思う (動詞-自立/五段・ワ行促音便)	12.661195
決して (副詞-一般)	う (助動詞/不変化型)	15.550137
決して (副詞-一般)	できる (動詞-自立/一段)	9.31457
決して (副詞-一般)	です (助動詞/特殊・デス)	17.007337
決して (副詞-一般)	ある (動詞-自立/五段・ラ行)	17.045036
決して (副詞-一般)	いる (動詞-非自立/一段)	17.537399

【 図 5 】



フロントページの続き

(72)発明者 榊山 享子

東京都小金井市貫井北町4 - 2 - 1 独立行政法人通信総合研究所内

(72)発明者 井佐原 均

東京都小金井市貫井北町4 - 2 - 1 独立行政法人通信総合研究所内

Fターム(参考) 5B091 AA15 CA02 CA05 EA25