

(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2005-228060

(P2005-228060A)

(43) 公開日 平成17年8月25日(2005.8.25)

(51) Int. Cl.⁷

G06F 17/27

F I

G06F 17/27

L

テーマコード (参考)

5B091

審査請求 未請求 請求項の数 9 O L (全 11 頁)

(21) 出願番号 特願2004-36308 (P2004-36308)

(22) 出願日 平成16年2月13日 (2004.2.13)

(出願人による申告) 平成15年度通信・放送機構、研究テーマ「大規模コーパススペース音声対話翻訳技術の研究開発」に関する委託研究、産業活力再生特別措置法第30条の適用を受ける特許出願

(71) 出願人 393031586

株式会社国際電気通信基礎技術研究所
京都府相楽郡精華町光台二丁目2番地2

(74) 代理人 100099933

弁理士 清水 敏

(72) 発明者 大野 誠寛

愛知県名古屋市昭和区檀溪通1-15-1
喜八マンション203

(72) 発明者 松原 茂樹

京都府相楽郡精華町光台二丁目2番地2
株式会社国際電気通信基礎技術研究所内

(72) 発明者 丸山 岳彦

京都府相楽郡精華町光台二丁目2番地2
株式会社国際電気通信基礎技術研究所内

最終頁に続く

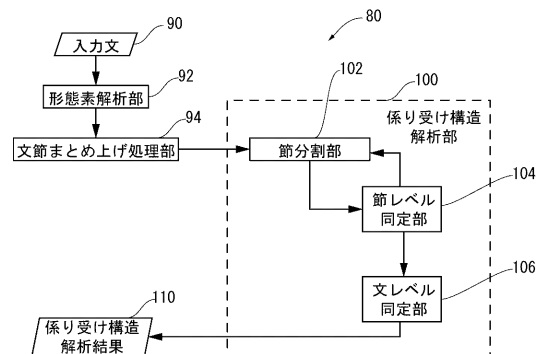
(54) 【発明の名称】 係り受け構造解析装置、係り受け構造解析方法、及びコンピュータで実行可能なプログラム

(57) 【要約】

【課題】 文の係り受け構造解析を効率的に行なう係り受け構造解析装置を提供する。

【解決手段】 入力文を構成する文節間の係り受け関係の構造を解析する係り受け構造解析部100は、入力文を、節ごとに分割するための節分割部102と、節分割部102によって分割された節の各々について、節を構成する文節間の係り受け関係の構造を同定するための節レベル同定部104と、入力文を構成する文節の内、その文節が属する節外の文節に係るものを選択し、選択した文節と節外の文節との間の係り受け関係の構造を同定するための文レベル同定部106とを含む。

【選択図】 図2



【特許請求の範囲】

【請求項 1】

入力文を構成する文節間の係り受け関係の構造を解析する係り受け構造解析装置であって、

前記入力文を、節ごとに分割するための節分割手段と、

前記節分割手段によって分割された節の各々について、当該節を構成する文節間の係り受け関係の構造を同定するための第 1 の同定手段と、

前記入力文を構成する文節の内、当該文節が属する節外の節を構成する文節に係るものを選択し、選択した前記文節と前記節外の節を構成する文節との間の係り受け関係の構造を同定するための第 2 の同定手段とを含む、係り受け構造解析装置。

10

【請求項 2】

前記入力文を構成する文節を、当該文節が属する節内の他の文節に係るものからなる第 1 の文節群と、当該文節が属する節外の文節に係るものからなる第 2 の文節群とに分類するための分類手段をさらに含み、

前記第 1 の同定手段は、前記節分割手段によって分割された節の各々について、当該節を構成する文節の内前記第 1 の文節群に含まれる第 1 の文節に係り文節とし、当該節を構成する文節の内前記第 1 の文節を除く文節のいずれかを受け文節とする、係り受け関係の構造を同定するための節レベル係り受け同定手段を含み、

前記第 2 の同定手段は、前記入力文を構成する文節の内、前記分類手段により第 2 の文節群に分類された文節を選択するための手段と、前記選択するための手段により選択された文節に係り文節とし、前記文節が属する節外の文節を受け文節とする、係り受け関係の構造を同定するための文レベル係り受け同定手段を含む、請求項 1 に記載の係り受け構造解析装置。

20

【請求項 3】

前記分類手段は、前記節分割手段による分割の結果に基づき、前記入力文を構成する文節の内、節の最終の文節を第 2 の文節群とし、前記節の最終の文節以外の文節を第 1 の文節群とする分類を行なうための手段を含む、請求項 2 に記載の係り受け構造解析装置。

【請求項 4】

前記第 2 の同定手段は、前記第 1 の同定手段により同定された前記節の各々についての係り受け関係の構造を前提条件として、前記選択した前記文節と前記節外の節を構成する文節との間の係り受け関係の構造を同定するための条件付同定手段を含む、請求項 1 に記載の係り受け構造解析装置。

30

【請求項 5】

前記条件付同定手段は、

前記選択した前記文節との間で、前記前提条件となる係り受け関係との非交差性を満たす文節を、前記節外のいずれかの文節の中から選抜するための受け文節候補選抜手段と、

前記選択した文節と前記受け文節候補選抜手段により選抜された文節のいずれかとの間の係り受け関係からなる、前記係り受け関係の構造を同定するための手段とを含む、請求項 4 に記載の係り受け構造解析装置。

【請求項 6】

40

前記第 1 の同定手段は、

前記分割された節の各々について、当該節を構成する文節が、当該節を構成する他の文節に係る確率を、当該節を構成する文節の各々について計算するための第 1 の確率計算手段と、

前記第 1 の確率計算手段により計算された確率をもとに、前記節を構成する文節間の係り受け関係の構造を同定するための手段とを含む、請求項 1 に記載の係り受け構造解析装置。

【請求項 7】

前記第 2 の同定手段は、

前記入力文を構成する文節の内、当該文節が属する節外の文節に係るものを選択し、選

50

択した前記文節が前記節外の文節に係る確率を、前記選択された文節の各々について計算するための第2の確率計算手段と、

前記第2の確率計算手段により計算された確率をもとに、前記選択された文節と前記節外の文節との間の係り受け関係の構造を同定するための手段とを含む、請求項1に記載の係り受け構造解析装置。

【請求項8】

コンピュータにより実行されると、当該コンピュータを請求項1から請求項7のいずれかに記載の係り受け構造解析装置として動作させる、コンピュータで実行可能なプログラム。

【請求項9】

入力文を構成する文節間の係り受け関係の構造を解析する係り受け構造解析方法であって、

前記入力文を節ごとに分割するステップと、

前記分割するステップによって分割された節の各々について、当該節を構成する文節間の係り受け関係の構造を同定するステップと、

前記入力文を構成する文節の内、当該文節が属する節外の文節に係るものを選択し、選択した前記文節と前記節外の文節との間の係り受け関係の構造を同定するステップとを含む、係り受け構造解析方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、文を構成する文節の間の係り受け関係を同定する係り受け構造解析装置及び方法、並びにコンピュータで実行可能なプログラムに関する。特に、独話文のように一文が長く構造が複雑な文の係り受け構造解析を効率的に行なう係り受け構造解析装置に関する。

【背景技術】

【0002】

自然言語について機械翻訳等の処理を行なうためには、処理対象となる自然言語の文法的構造を解析し構文を明らかにした上で、その意味に関する解析を行なうことが必要となる。日本語のような言語の場合、文節間の構文上の関係は文節の順序によって決定されるのではなく、文節を構成する助詞及び助動詞、並びに語の意味などによって決定される。そのため、日本語をはじめとするこのような言語の構文解析を行なうための手法として、係り受け構造を解析する手法が広く用いられている。

【0003】

係り受け構造とは、ある文節が他の文節に係るという関係（以下、この関係を「係り受け関係」と呼ぶ。）によって構成された、文節間の関係構造である。一般的な係り受け解析では、2つの文節が係り受け関係を有する可能性を規定する係り受け規則と、係り受けに関する次の一般的な制約とに基づいて、係り受け構造を同定する手法が用いられている。すなわち、一般的な制約とは、

（1）係り受け関係は互いに交差しないこと（係り受けの非交差性）、

（2）文末の文節を除き、他の文節に係る文節（以下、「係り文節」と呼ぶ。）は、当該文節に係る対象となる文節（以下、「受け文節」と呼ぶ。）より前に置かれること（修節の後方性）、及び、

（3）文末の文節を除き、係り文節に対する受け文節は必ず存在し、かつ、1つしか存在しないこと（係り先の唯一性）である。

【0004】

話し言葉の内、複数の話者が交代で話す形式の「対話」文について、統計情報を用いて係り受け構造を解析する方法が、後掲の非特許文献1及び非特許文献2に記載されている。非特許文献1及び非特許文献2に記載の方法は、話し言葉において上記した一般的な制約に従わない係り受け関係が頻出することに注目した方法である。この方法では、係り受

10

20

30

40

50

け関係に関する上記した一般的な制約の内、係り受けの非交差性のみを満たすべき制約とし、残る修飾の後方性及び係り先の唯一性について統計情報を反映した考慮を行なうことにより、係り受け構造の解析を行なう。この方法により、係り受け関係に関する一般的な制約が必ずしも適用できるとは限らない話し言葉に対して頑健に対処することができる。

【0005】

【非特許文献1】マツバラ S.、ムラセ T.、カワグチ N.、イナガキ Y.：「日本語の自然な話し言葉の統計的係り受け解析」、第19回COLING予稿集、Vol. 1、640-645頁、2002. (Matsubara, S., Murase, T., Kawaguchi, N. and Inagaki, Y.: Stochastic Dependency Parsing of Spontaneous Japanese Spoken Language, Proc. of 19th COLING, Vol. 1, pp. 640-645, 2002)

【非特許文献2】大野 誠寛, 松原 茂樹, 河口 信夫, 稲垣 康善：「日本語音声対話文の統計的係り受け解析とその評価」、情報処理学会第65回全国大会講演論文集、Vol. 2、1-2頁、2003.

【非特許文献3】柏岡秀紀、丸山岳彦、田中英輝：「節境界と係り受け解析」、言語処理学会第9回年次大会論文集、117-120頁、2003.

【非特許文献4】丸山岳彦、柏岡秀紀、熊野正、田中英輝：「節境界自動検出ルール の作成と評価」、言語処理学会第9回年次大会論文集、517-520頁、2003.

【発明の開示】

【発明が解決しようとする課題】

【0006】

話し言葉は、上記した「対話」形式の他に、1人の話者のみが話す「独話」形式のものがある。独話文は、対話文に比べて一文が長く文の構造が複雑であるという特徴がある。一文が長くなると、一文に含まれる文節数も増え、可能な係り受け関係の組合せも増加する。また、文の構造が複雑になると、可能な係り受け関係の組合せの中から適切なものを選んで係り受け構造を同定することが困難になる。

【0007】

非特許文献1又は非特許文献2に記載の係り受け構造の解析方法は、対話文において非文法性に対して頑健に対処することが可能である。しかし、独話文のように一文の長さが長く文の構造が複雑な文を解析するには、長い解析時間を要することになるうえ、高い解析精度を達成することが困難である。

【0008】

それゆえに、本発明の目的は、文の係り受け構造解析を効率的に行なうことができる係り受け構造解析装置を提供することである。

【0009】

また、本発明の別の目的は、特に日本語の独話文など非文法性を有する係り受け構造が複雑な文の係り受け構造解析において、高い解析精度を維持しつつ、解析に要する時間を短縮することができる係り受け構造解析装置を提供することである。

【課題を解決するための手段】

【0010】

本発明の第1の局面に係る係り受け構造解析装置は、入力文を構成する文節間の係り受け関係の構造を解析する装置である。この係り受け構造解析装置は、入力文を、節ごとに分割するための節分割手段と、節分割手段によって分割された節の各々について、その節を構成する文節間の係り受け関係の構造を同定するための第1の同定手段と、入力文を構成する文節の内、その文節が属する節外の節を構成する文節に係るものを選択し、選択した文節と、その文節が属する節外の節を構成する文節との間の係り受け関係の構造を同定するための第2の同定手段とを含む。

【0011】

好ましくは、入力文を構成する文節を、その文節が属する節内の他の文節に係るものか

10

20

30

40

50

らなる第1の文節群と、その文節が属する節外の文節に係るものからなる第2の文節群とに分類するための分類手段をさらに含む。第1の同定手段は、節分割手段によって分割された節の各々について、節を構成する文節のうち第1の文節群に含まれる第1の文節に係り文節とし、節を構成する文節のうち第1の文節を除く文節のいずれかを受け文節とする、係り受け関係の構造を同定するための節レベル係り受け同定手段を含む。第2の同定手段は、入力文を構成する文節の内、分類手段により第2の文節群に分類された文節を選択するための手段と、選択するための手段により選択された文節に係り文節とし、文節が属する節外の文節を受け文節とする、係り受け関係の構造を同定するための文レベル係り受け同定手段を含む。

【0012】

10

より好ましくは、分類手段は、節分割手段による分割の結果に基づき、入力文を構成する文節の内、節の最終の文節を第2の文節群とし、節の最終の文節以外の文節を第1の文節群とする分類を行なうための手段を含む。

【0013】

第2の同定手段は、第1の同定手段により同定された節の各々についての係り受け関係の構造を前提条件として、選択した文節と節外の節を構成する文節との間の係り受け関係の構造を同定するための条件付同定手段を含んでもよい。

【0014】

より好ましくは、条件付同定手段は、選択した文節との間で、前提条件となる係り受け関係との非交差性を満たす文節を、節外のいずれかの文節の中から選抜するための受け文節候補選抜手段と、選択した文節と受け文節候補選抜手段により選抜された文節のいずれかとの間の係り受け関係からなる、係り受け関係の構造を同定するための手段とを含む。

20

【0015】

第1の同定手段は、分割された節の各々について、節を構成する文節が、節を構成する他の文節に係る確率を、節を構成する文節の各々について計算するための第1の確率計算手段と、第1の確率計算手段により計算された確率をもとに、節を構成する文節間の係り受け関係の構造を同定するための手段とを含んでもよい。

【0016】

第2の同定手段は、入力文を構成する文節の内、文節が属する節外の文節に係るものを選択し、選択した文節が節外の文節に係る確率を、選択された文節の各々について計算するための第2の確率計算手段と、確率計算手段により計算された確率をもとに、選択された文節と節外の文節との間の係り受け関係の構造を同定するための手段とを含んでもよい。

30

【0017】

本発明の第2の局面に係るプログラムは、コンピュータにより実行されると、そのコンピュータを本発明の第1の局面に係る係り受け構造解析装置として動作させる。

【0018】

本発明の第3の局面に係る係り受け構造解析方法は、入力文を構成する文節間の係り受け関係の構造を解析する方法である。この方法は、入力文を節ごとに分割するステップと、分割するステップによって分割された節の各々について、節を構成する文節間の係り受け関係の構造を同定するステップと、入力文を構成する文節の内、文節が属する節外の文節に係るものを選択し、選択した文節と節外の文節との間の係り受け関係の構造を同定するステップとを含む。

40

【発明を実施するための最良の形態】

【0019】

以下、図面を参照しつつ、本発明の一実施の形態を説明する。

【0020】

文は、節が1つ以上連接する構成をもつ。節は、述部を中心とする文節のまとまりであり、単文に相当する言語的単位である。図1に、文、節、及び文節、並びに文節間の係り受け構造の一例を概略的に示す。図1を参照して、例えば、係り受け構造の解析対象とな

50

る文 30 が、「この警察法という法律ですが警察を管理するという曖昧な言葉でしか公安委員会の役割を示していません。」というものであったとする。解析対象となる文 30 は、第 1 の節 40、第 2 の節 42、及び第 3 の節 44 によって構成される。文が複文又は重文である場合、その文には複数の節が含まれることになる。独話文のように 1 文が長く文の構造が複雑な文の場合、一文に含まれる節の数は多くなる。

【0021】

第 1 の節 40、第 2 の節 42、及び第 3 の節 44 はそれぞれ、複数の文節 50 A 及び 50 B、文節 52 A 及び 52 B、並びに文節 54 A、54 B、54 C、54 D、及び 54 E によって構成される。第 1 の節 40 を構成する文節について係り受け関係をみると、文節 50 A が係る文節すなわち文節 50 A の受け文節は、第 1 の節 40 内にある文節 50 B と
10 なる。これに対して、文節 50 B の受け文節は、第 1 の節 40 と異なる第 3 の節 44 内の文節 54 E となる。

【0022】

文 30 を構成する文節間の係り受け関係は、係り文節と受け文節とが共に同一の節内に存在する係り受け関係 60 A、60 B、…、60 F と、係り文節と受け文節とが互いに他と異なる節内に存在する係り受け関係 70 A 及び 70 B とに分類することができる。本明細書では、文節間の係り受け関係の内、前者の係り受け関係 60 A、60 B、…、60 F を「節レベルの係り受け関係」と呼ぶ。また、後者の係り受け関係 70 A、70 B を「文
20 レベルの係り受け関係」と呼ぶ。

【0023】

なお、非特許文献 3 によると、実際の独話文では、約 90 % の節において文節間の係り受け関係に以下のような性質があることが報告されている。すなわち、「文は一つ以上の節の接続であり、各節を構成する文節は、節の最終の文節を除き、節内の文節に係る」ということである。

【0024】

本発明の一実施の形態として以下に説明する係り受け構造解析装置は、複数の文節によって構成される入力文を 1 以上の節に分割し、各節において、節レベルの係り受け構造を同定する。さらに同定した節レベルの係り受け関係に基づいて、文レベルの係り受け解析を行なう。

【0025】

図 2 に、本実施の形態に係る係り受け構造解析装置 80 の機能的構成をブロック図形式で示す。図 2 を参照して、係り受け構造解析装置 80 は、解析の対象となる入力文 90 を形態素解析する形態素解析部 92 と、形態素解析部 92 の出力する形態素列を文節ごとにまとめる文節まとめ上げ処理部 94 と、文節まとめ上げ処理部 94 の出力する、形態素及び文節に関する情報を受け、入力文 90 を構成する文節間の係り受け構造の解析結果 110 を出力する係り受け構造解析部 100 とを含む。

【0026】

係り受け構造解析部 100 は、文節まとめ上げ処理部 94 の出力を受け、入力文 90 中の節の境界を同定し、入力文 90 を複数の節に分割する節分割部 102 と、節分割部 102 による分割で生成された節の各々について、節レベルの係り受け関係を同定する節レベル同定部 104 と、節レベル同定部 104 による解析結果に基づき、文レベルの係り受け関係を同定する文レベル同定部 106 とを含む。

【0027】

節分割部 102 は、入力文 90 の形態素解析部 92 による解析結果に基づき節境界解析を行なうことにより節の分割を行なうものである。非特許文献 4 には、形態素列パターンに関する節境界検出ルールをもとに、入力文 90 の局所的な形態素解析結果から節境界を検出する手法が記載されている。節分割部 102 は、例えば、非特許文献 4 に記載の方法を用いて節境界解析を行ない、節と節との境界を検出し、入力文 90 を節に分割するものであってもよい。

【0028】

10

20

30

40

50

図 3 に、係り受け構造解析部 100 による係り受け構造の解析過程を模式的に示す。係り受け構造解析部 100 の動作の詳細について図 2 及び図 3 を参照しつつ説明する。

【0029】

図 3 を参照して、文節まとめ上げ処理部 94 は、複数の文節 142 A, ..., 142 P が接続した文節列 140 を図 2 に示す節分割部 102 に与える。

【0030】

節分割部 102 は、与えられた文節列 140 について、図 2 に示す形態素解析部 92 による形態素解析結果をもとに節境界解析を行ない、与えられた文節列を複数の節 150 A, ..., 150 M に分割する。分割によって生成された節 150 A は、複数の文節 152 A, ..., 152 N が接続した文節列 154 A を含む。同様に、節 150 B, ..., 150 M はそれぞれ、文節列 154 B, ..., 154 M を含む。 10

【0031】

節分割部 102 は、第 1 の節 150 A を節レベル同定部 104 に与える。節レベル同定部 104 は、与えられた節 150 A 中の文節列 154 A に関する節レベルの係り受け構造の同定を行なう。

【0032】

節レベルでの係り受け構造の同定は、例えば以下のように行われる。文節列 154 A が、 n_1 個の文節 b_i^{-1} (ただし、 $1 \leq i \leq n_1$) を含むものとする。まず、節レベル同定部 104 は、異なる 2 つの文節 b_k^{-1} 及び文節 b_l^{-1} (ただし、 $1 \leq k < l \leq n_1$) について、文節 b_k^{-1} が文節 b_l^{-1} に係る確率を求めるために、文節 b_k^{-1} 及び文節 b_l^{-1} の属性を準備する。 20
ここで準備される属性は、文節における自立語の原形及びその自立語の品詞を含む。また節レベル同定部 104 は、文節 b_k^{-1} が付属語を伴うものであるか否かを判定して、文節 b_k^{-1} の係りの種類を識別する。この際、文節 b_k^{-1} が付属語を伴っているならば、係りの種類は、付属語の語形、品詞、及び活用形により判別される。付属語を伴っていないならば、文節を構成する最後の形態素の品詞及び活用形により判別される。節レベル同定部 104 は、上記した属性、係りの種類、及び文節 b_k^{-1} と文節 b_l^{-1} との文節間距離に基づき、共起頻度関数を用いて、文節 b_k^{-1} が文節 b_l^{-1} に係る確率を計算し、文節間の係り受け関係の各々が独立であるという仮定に基づき、動的計画法を用いて最尤の係り受け構造 160 A (図 3 参照) を同定する。

【0033】

節レベル同定部 104 は、節レベルの係り受け構造を同定するにあたり、他の文節列に属する文節に係る文節、すなわち文レベルの係り受け関係における係り文節について、係り受け関係の同定を行なわない。上記した非特許文献 3 によると、実際の独話文では、約 90% の節において、各節を構成する文節は、節の最終の文節を除き節内の文節に係る。節レベル同定部 104 は、与えられた文節列 154 A の最終文節 152 N を係り文節とする係り受け関係についての同定を行なわずに、与えられた文節列 154 A についての係り受け構造の同定を終了する。また、節レベル同定部 104 による係り受け構造の同定において、受け文節がないもの又は不明なものを、文レベルの係り受け関係における係り文節の候補としてもよい。 30

【0034】

図 2 を参照して、節レベル同定部 104 が与えられた文節列 154 A (図 3 参照) についての同定を終了したことに応答して、節分割部 102 は、文節列 154 A からなる節 150 A の次の節 150 B を構成する文節列 154 B を、節レベル同定部 104 に与える。節分割部 102 及び節レベル同定部 104 はこれら一連の動作を繰返すことにより、入力文 90 を構成する節 150 A, ..., 150 M の各々について、図 3 に示す節レベルの係り受け構造 160 A, ..., 160 M の同定を行なう。同定された節レベルの係り受け構造 160 A, ..., 160 M の情報は、節レベル同定部 104 から、図 2 に示す文レベル同定部 106 に与えられる。 40

【0035】

文レベル同定部 106 は、節レベルの係り受け構造 160 A, ..., 160 M の情報を受 50

け、文レベルの係り受け構造の同定を開始する。

【0036】

まず、文レベル同定部106は、係り受け関係の同定対象となる文節を選択する。節レベル同定部104により係り受け関係の同定が行われなかった文節は、文レベルの係り受け関係において他の文節に係る可能性がある文節であるため、このような文節は係り受け関係の同定対象として選択される。また、各節の最終文節も同定対象として選択される。

【0037】

選択された文節に係る文節は、選択された文節を含む節内ではなく、それ以外の節のいずれかに含まれる文節である。例えば図3に示す文節列150Aを構成する文節列154Aの最終文節152N, 162N, ...に係り文節の同定対象とした場合、この文節に対する受け文節は、文節列154A以外の文節列154B, ..., 154Mの中に存在する文節162Lである。文レベル同定部106は、選択された文節152Aとその文節が属する節以外の節に属する文節162Lを受け文節とする係り受け関係によって構成される、文レベルの係り受け構造を同定する。

10

【0038】

文レベルでの係り受け構造の同定は、例えば以下のように行われる。文レベル同定部106は、節レベル同定部104が節レベルの係り受け構造を同定するのに用いた方法と同様の方法で、選択された文節及び選択した文節の受け文節の候補の属性、選択された文節の係りの種類、並びに選択された文節と受け文節の候補との文節間距離に基づき、共起頻度関数を用いて、選択された文節が受け文節の候補に係る確率を計算し、動的計画法を用いて最尤の係り受け構造170（図3参照）を同定する。ただし、文レベル同定部106は、節レベルの係り受け構造160A, ..., 160Mは既に確定したものであるとみなして、最尤の係り受け構造170を同定する。例えば、図3に示す文節152Nと文節列154Bに含まれる文節162Lとの間の係り受け関係が、文節列154Bの節レベルの係り受け構造160Bを構成するいずれかの係り受け関係との間で交差する場合、この係り受け関係は、係り受け関係の探索対象から除外される。

20

【0039】

文レベル同定部106は、節レベルの係り受け構造160A, ..., 160Mと文レベルの最尤の係り受け構造とを合わせて、入力文90の係り受け構造解析結果110として出力する。

30

【0040】

以上のように、係り受け構造解析部100（図2参照）は、節レベルの係り受け構造の同定処理と文レベルの係り受け構造の同定処理とを分けて行なう。節レベルの係り受け構造の同定を行なう際には、節を構成する文節の中で、係り受け関係の同定を行なう。節を構成する文節数は少ないため、係り文節と受け文節との組合せの数も少ない。よって、文全体を対象として係り受け関係の同定を行なう場合より、係り受け関係の探索範囲は絞込まれ、解析に要する計算量は少なくなり、効率的に処理を行なうことができる。また、文レベルの係り受け関係を同定する際には、同定した節レベルの係り受け構造を前提とするため、受け文節となる文節の候補が絞込まれる。よって、文レベルの係り受け構造の同定においても、解析に要する計算量は少なくなり、効率的に処理を行なうことができる。

40

【0041】

なお、ブロック図形式で説明した各機能部は、いずれもコンピュータハードウェア及び当該コンピュータ上で実行されるプログラムにより実現できる。このコンピュータとしては、汎用のハードウェアを有するものを用いることができる。また、上で説明した装置の各機能ブロックは、この明細書の記載に基づき、当業者であればプログラムで実現することができる。そうしたプログラムもまた1つのデータであり、記憶媒体に記憶させて流通させることができる。

【0042】

今回開示された実施の形態は単に例示であって、本発明が上記した実施の形態のみに制限されるわけではない。本発明の範囲は、発明の詳細な説明の記載を参酌した上で、特許

50

請求の範囲の各請求項によって示され、そこに記載された文言と均等の意味及び範囲内でのすべての変更を含む。

【図面の簡単な説明】

【0043】

【図1】文、節、及び文節、並びに文節間の係り受け構造の一例を概略的に示す図である。

【図2】本実施の形態に係る係り受け構造解析装置80の構成を示すブロック図である。

【図3】係り受け構造解析部100による係り受け構造の解析過程を模式的に示す図である。

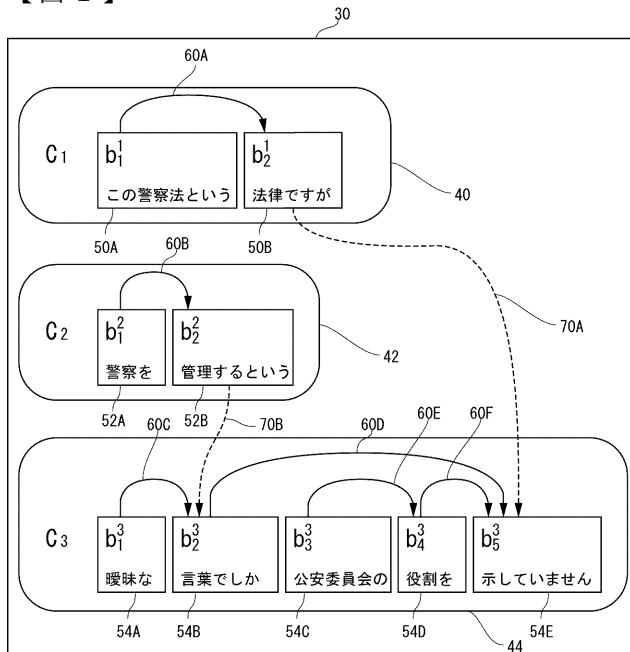
【符号の説明】

【0044】

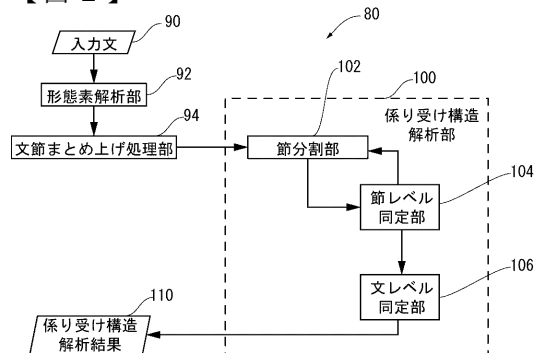
40、42、44、150A、…、150M 節、80 係り受け構造解析装置、90 入力文、92 形態素解析部、94 文節まとめ上げ処理部、100 係り受け構造解析部、102 節分割部、104 節レベル同定部、106 文レベル同定部、110 係り受け構造解析結果、160A、…、160M 節レベルの係り受け構造、170 文レベルの係り受け構造

10

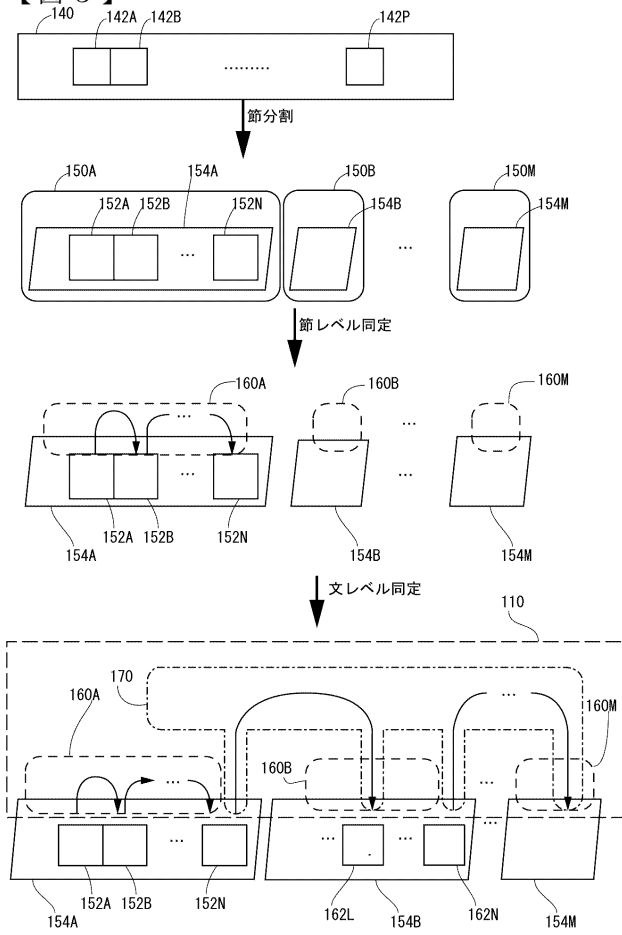
【図1】



【図2】



【図 3】



フロントページの続き

(72)発明者 柏岡 秀紀

京都府相楽郡精華町光台二丁目 2 番地 2 株式会社国際電気通信基礎技術研究所内

Fターム(参考) 5B091 AA15 CA02 CA05