(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第4802327号
(P4802327)
(45)発行日　平成23年10月26日(2011.10.26)　(24)登録日　平成23年8月19日(2011.8.19)

(51)Int.Cl.　　F I
G06F 17/30 (2006.01)　　G06F 17/30 170A
G06F 17/28 (2006.01)　　G06F 17/30 350Z
　　G06F 17/28 X

請求項の数 6　(全 21 頁)

(21)出願番号　特願2006-66829(P2006-66829)
(22)出願日　平成18年3月11日(2006.3.11)
(65)公開番号　特開2007-241908(P2007-241908A)
(43)公開日　平成19年9月20日(2007.9.20)
審査請求日　平成19年11月13日(2007.11.13)

(73)特許権者　504139662
国立大学法人名古屋大学
愛知県名古屋市千種区不老町１番
(74)代理人　100085361
弁理士　池田　治幸
(72)発明者　松原　茂樹
愛知県名古屋市千種区不老町１番　国立大学法人名古屋大学内
(72)発明者　加藤　芳秀
愛知県名古屋市千種区不老町１番　国立大学法人名古屋大学内

審査官　岩田　淳

最終頁に続く

(54)【発明の名称】依存構造に基づく例文検索方法、プログラム、および例文検索プログラムを記録した記録媒体、ならびに例文検索装置

(57)【特許請求の範囲】
【請求項１】
予め蓄積された言語資料中の複数の例文と該例文を構成する単語間の依存関係についての情報である依存構造情報とから、検索条件に該当する例文を検索する例文検索方法であって、
コンピュータが、
検索しようとする文に含まれる複数の単語と該単語の順序とについての入力を受け付け、該入力された内容に基づく検索条件を設定する検索条件設定工程と、
前記複数の例文のそれぞれに対し、前記検索条件設定工程において設定された検索条件に含まれる複数の単語の全てが含まれるか否かを判断し、前記検索条件に含まれる複数の単語の全てが含まれる場合には、前記検索条件に含まれる複数の単語のそれぞれについて、単語間の依存関係を記述するための依存構造パターンの初期状態である初期依存構造パターンを生成する依存構造パターン初期化工程、および、前記例文において該初期依存構造パターンを含む２つの依存構造パターンが存在し、該２つの依存構造パターンにおける主辞が依存関係を有する場合に、該２つの依存構造パターンを結合する操作を実行することにより１つの依存構造パターンを生成する依存構造パターン結合工程を含む依存構造パターン生成工程と、
前記依存構造パターン生成工程によって、前記複数の例文のうち、前記検索条件に含まれる複数の単語のすべての単語間の依存構造を一の依存構造パターンで生成できた例文を前記検索条件に該当するものとして選択する文選択工程と

から構成されることを特徴とする例文検索方法。
【請求項２】
予め蓄積された言語資料中の複数の例文、該例文を構成する単語間の依存関係についての情報である依存構造情報、および、該例文を構成する単語の品詞とから、検索条件に該当する例文を検索する例文検索方法であって、
前記検索条件設定工程は、前記検索しようとする文に含まれる複数の単語のうち少なくとも１つの品詞を、前記単語に加えて、あるいは該単語に替えて前記検索条件として設定するものであり、
前記依存構造パターン生成工程は、前記複数の例文のそれぞれに対し、該例文に前記検索条件に含まれる複数の単語の全てが含まれ、かつ該複数の単語のうち少なくとも１つの単語について該単語に加えて設定された品詞が一致する場合、あるいは、該例文に前記検索条件に含まれる複数の単語の全てが含まれ、かつ該複数の単語のうち少なくとも１つの単語に替えて設定された品詞が一致する場合に、前記依存構造パターンを生成すること
を特徴とする請求項１の例文検索方法。
【請求項３】
請求項１または２に記載の方法をコンピュータに実行させる例文検索プログラム。
【請求項４】
請求項１または２に記載の方法をコンピュータに実行させる例文検索プログラムを記録したコンピュータ読み取り可能な記録媒体。
【請求項５】
予め蓄積された言語資料中の複数の例文と該例文を構成する単語間の依存関係についての情報である依存構造情報とから、検索条件に該当する例文を検索する例文検索装置であって、
検索しようとする文に含まれる複数の単語と該単語の順序とに基づく検索条件を設定する検索条件設定手段と、
前記複数の例文のそれぞれに対し、前記検索条件設定工程において設定された検索条件に含まれる複数の単語の全てが含まれるか否かを判断し、前記検索条件に含まれる複数の単語の全てが含まれる場合には、前記検索条件に含まれる複数の単語のそれぞれについて、単語間の依存関係を記述するための依存構造パターンの初期状態である初期依存構造パターンを生成する依存構造パターン初期化手段、および、前記例文において該初期依存構造パターンを含む２つの依存構造パターンが存在し、該２つの依存構造パターンにおける主辞が依存関係を有する場合に、該２つの依存構造パターンを結合する操作を実行することにより１つの依存構造パターンを生成する依存構造パターン結合手段を含む依存構造パターン生成手段と、
前記依存構造パターン生成手段によって、前記複数の例文のうち、前記検索条件に含まれる複数の単語のすべての単語間の依存構造を一の依存構造パターンで生成できた例文を前記検索条件に該当するものとして選択する文選択手段と
を有することを特徴とする例文検索装置。
【請求項６】
予め蓄積された言語資料中の複数の例文、該例文を構成する単語間の依存関係についての情報である依存構造情報、および、該例文を構成する単語の品詞とから、検索条件に該当する例文を検索する例文検索装置であって、
前記検索条件設定手段は、前記検索しようとする文に含まれる複数の単語のうち少なくとも１つの品詞を、前記単語に加えて、あるいは該単語に替えて前記検索条件として設定するものであり、
前記依存構造パターン生成手段は、該例文に前記検索条件に含まれる複数の単語の全てが含まれ、かつ該複数の単語のうち少なくとも１つの単語について該単語に加えて設定された品詞が一致する場合、あるいは、該例文に前記検索条件に含まれる複数の単語の全てが含まれ、かつ該複数の単語のうち少なくとも１つの単語に替えて設定された品詞が一致する場合に、前記依存構造パターンを生成すること

を特徴とする請求項５の例文検索装置。

【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、蓄積された言語資料中の複数の例文から、検索条件に該当する例文を検索する例文検索方法に関するものである。
【背景技術】
【０００２】
正しい英文を作成するために、用例を参照し模倣することが効果的であり、大量の例文の中から、参照に値する英文を容易に見つけることは、英文作成者にとって重要である。そのため、英文作成者は、適切な英文用例を見つけるために、「手がかりとなるフレーズ」を入力し、それを含む英文が出力される英文用例検索の環境を必要としている。
【０００３】
このような要求のもと、近年、大量の文が集積された大規模コーパス（ｃｏｒｐｕｓ；言語資料）の重要性はますます高まっており、言語現象の調査、外国語学習、自然言語処理システムの開発など様々な場面で言語資源として活用されている。そして、コーパスを効果的に活用するために、コーパスから任意の検索条件に該当する文を検索する様々なコーパス検索システムが提案されている。
【０００４】
しかしながら、多くのシステムは、キーワードベースの用例検索を実現するにすぎず、このキーワードベースの検索は、単純で直感的であるという利点があるものの、単純なマッチング、すなわち、コーパス中の各文に、検索条件として設定された単語が単に含まれるか否かによってなされる検索であり、構文構造などの言語的構造を活用した検索はできない。このため、複数の単語から成る構文を含む例文を検索したい場合であっても、その複数を単語を単に含むのみであって、目的とする構文として含まれるわけではない文も検索結果に含まれてしまうこととなる。
【０００５】
【特許文献１】特開平９－２６５４７６号公報
【非特許文献１】Ｃｏｒｌｅｙ，Ｓ．，Ｃｏｒｌｅｙ，Ｍ．，Ｋｅｌｌｅｒ，Ｆ，．Ｃｒｏｃｋｅｒ，Ｍ．，ａｎｄＴｒｅｗｉｎ，Ｓ．，”Ｆｉｎｄｉｎｇ Ｓｙｎｔａｃｔｉｃ Ｓｔｒｕｃｔｕｒｅ ｉｎ Ｕｎｐａｒｓｅｄ Ｃｏｒｐｏｒａ：Ｔｈｅ Ｇｓｅａｒｃｈ Ｃｏｒｐｕｓ Ｑｕｅｒｙ Ｓｙｓｔｅｍ”，Ｃｏｍｐｕｔｅｒｓ ａｎｄ ｔｈｅ Ｈｕｍａｎｉｔｉｅｓ，Ｓｐｒｉｎｇｅｒ Ｎｅｔｈｅｒｌａｎｄ，（オランダ），２００１年，３５巻，２号，ｐ．８１－９４
【非特許文献２】Ｒｅｓｎｉｋ，Ｐ．ａｎｄＥｌｋｉｓｓ，Ａ．”Ｔｈｅ Ｌｉｎｇｕｉｓｔ’ｓ Ｓｅａｒｃｈ Ｅｎｇｉｎｅ：Ａｎ Ｏｖｅｒｖｉｅｗ” ｔｈｅ ＡＣＬ Ｉｎｔｅｒａｃｔｉｖｅ Ｐｏｓｔｅｒ ａｎｄ Ｄｅｍｏｎｓｔｒａｔｉｏｎ Ｓｅｓｓｉｏｎｓ 予稿集，（米国），Ａｓｓｏｃｉａｔｉｏｎ ｆｏｒ Ｃｏｍｐｕｔａｔｉｏｎａｌ Ｌｉｎｇｕｉｓｔｉｃｓ，２００５年，ｐ．３３－３６
【発明の開示】
【発明が解決しようとする課題】
【０００６】
一方、これまでに、構文構造情報を利用したコーパス検索システムとしてＧｓｅａｒｃｈ（非特許文献１）やＬｉｎｇｕｉｓｔ’ｓ Ｓｅａｒｃｈ Ｅｎｇｉｎｅ（ＬＳＥ）（非特許文献２）といったシステムが提案されている。これらのシステムでは、後述する句構造を構成する品詞や当該句構造のタイプについての情報である句構造パターンを用いて、すなわち検索条件において設定されたまたは、解析によって得られた、複数の単語からなる単語列のまとまりを示す構造である句構造と、コーパスに付されたコーパスを構成する文における句構造の一致性を考慮してコーパスを検索する。Ｇｓｅａｒｃｈでは、ユー

ザは句構造パターンと文法をシステムに入力する。システムは入力された文法を用いてコーパス中の文を構文解析し、与えられた句構造パターンを持つ文を検索結果として提示する。ＬＳＥでは、ユーザはまず、探したい文の例を入力する。システムは、入力された例文を統計的構文解析により解析し、その解析結果をユーザに提示する。ユーザはこの構文解析結果を編集し、構造的なクエリを作成する。最終的にシステムは、このクエリにマッチする句構造を持つ文を検索結果として返す。これらのシステムでは、構文的情報を利用したコーパス検索を実現できる。しかし、検索にあたり、所定のパターンの入力を必要としたり、あるいは例文を入力する必要があるため、キーワードベースの検索システムのような、簡単で、直感的な検索を実現しているとは言いがたい。また、これらのシステムでは、クエリおよびコーパス中の文における複数の単語間の修飾および被修飾関係である係り受け関係が考慮されていない。

【０００７】

本発明は、以上の事情を背景として為されたものであり、その目的とするところは、特別な表現による検索条件を必要とすることなく、蓄積された言語資料中の複数の例文から、検索条件に含まれる単語を構文的に含む例文を検索する例文検索方法、コンピュータが実行可能な例文検索プログラム、および、その例文検索プログラムが記憶された記録媒体ならびに例文検索装置を提供するところにある。

【課題を解決するための手段】

【０００８】

かかる目的を達成するために、請求項１に係る方法発明の要旨とするところは、予め蓄積された言語資料中の複数の例文と該例文を構成する単語間の依存関係についての情報である依存構造情報とから、検索条件に該当する例文を検索する例文検索方法であって、コンピュータが、（ａ）検索しようとする文に含まれる複数の単語と該単語の順序とについての入力を受け付け、該入力された内容に基づく検索条件を設定する検索条件設定工程と、（ｂ）前記複数の例文のそれぞれに対し、前記検索条件設定工程において設定された検索条件に含まれる複数の単語の全てが含まれるか否かを判断し、前記検索条件に含まれる複数の単語の全てが含まれる場合には、前記検索条件に含まれる複数の単語のそれぞれについて、単語間の依存関係を記述するための依存構造パターンの初期状態である初期依存構造パターンを生成する依存構造パターン初期化工程、および、前記例文において該初期依存構造パターンを含む２つの依存構造パターンが存在し、該２つの依存構造パターンにおける主辞が依存関係を有する場合に、該２つの依存構造パターンを結合する操作を実行することにより１つの依存構造パターンを生成する依存構造パターン結合工程を含む依存構造パターン生成工程と、（ｃ）前記依存構造パターン生成工程によって、前記複数の例文のうち、前記検索条件に含まれる複数の単語のすべての単語間の依存構造を一の依存構造パターンで生成できた例文を前記検索条件に該当するものとして選択する文選択工程から構成されることを特徴とする。

【発明の効果】

【０００９】

このようにすれば、検索条件は検索しようとする文に含まれる複数の単語と該単語の順序とに基づいて設定され、前記複数の例文のそれぞれに対し、設定された検索条件に含まれる複数の単語の全てが含まれるか否かを判断し、前記検索条件に含まれる複数の単語の全てが含まれる場合には、前記検索条件に含まれる複数の単語のそれぞれについて、単語間の依存関係を記述するための依存構造パターンの初期状態である初期依存構造パターンが生成され、また、前記例文において該初期依存構造パターンを含む２つの依存構造パターンが存在し、該２つの依存構造パターンにおける主辞が依存関係を有する場合には、該２つの依存構造パターンを結合する操作を実行することにより１つの依存構造パターンが生成される。そのため、前記複数の例文のうち、前記検索条件に含まれる複数の単語のすべての単語間の依存構造を一の依存構造パターンで生成できた例文が前記検索条件に該当

するものとして選択されるので、好適に構文的に検索を行うことができる。

【００１０】

また、請求項２に係る発明によれば、好適には、予め蓄積された言語資料中の複数の例文、該例文を構成する単語間の依存関係についての情報である依存構造情報と、該例文を構成する単語の品詞から、検索条件に該当する例文を検索する例文検索方法であって、（ｄ）前記検索条件設定工程は、前記検索しようとする文に含まれる複数の単語のうち少なくとも１つの品詞を、前記単語に加えて、あるいは該単語に替えて前記検索条件として設定するものであり、（ｅ）前記依存構造パターン生成工程は、該例文に前記検索条件に含まれる複数の単語の全てが含まれ、かつ該複数の単語のうち少なくとも１つの単語について該単語に加えて設定された品詞が一致する場合、あるいは、該例文に前記検索条件に含まれる複数の単語の全てが含まれ、かつ該複数の単語のうち少なくとも１つの単語に替えて設定された品詞が一致する場合に、前記依存構造パターンを生成すること、を特徴とする。このようにすれば、例文の検索において、該例文に前記検索条件に含まれる複数の単語の全てが含まれ、かつ該複数の単語のうち少なくとも１つの単語について該単語に加えて設定された品詞が一致する場合、あるいは、該例文に前記検索条件に含まれる複数の単語の全てが含まれ、かつ該複数の単語のうち少なくとも１つの単語に替えて設定された品詞が一致する場合に前記依存構造パターンの生成が行なわれるので、一層正確な例文の検索が可能となり、あるいは、検索条件をあいまいにした例文の検索が可能となる。

【００１１】

また、請求項３に係る発明の要旨とするところは、上記請求項１または２に係る方法発明をコンピュータに実行させる例文検索プログラムであることを特徴とする。

【００１２】

また、請求項４に係る発明の要旨とするところは、上記請求項１または２に係る方法発明をコンピュータに実行させる例文検索プログラムが記憶された記録媒体であることを特徴とする。

【００１３】

また、請求項５に係る発明の要旨とするところは、予め蓄積された言語資料中の複数の例文と該例文を構成する単語間の依存関係についての情報である依存構造情報とから、検索条件に該当する例文を検索する例文検索装置であって、（ａ）検索しようとする文に含まれる複数の単語と該単語の順序とに基づく検索条件を設定する検索条件設定手段と、（ｂ）前記複数の例文のそれぞれに対し、前記検索条件設定工程において設定された検索条件に含まれる複数の単語の全てが含まれるか否かを判断し、前記検索条件に含まれる複数の単語の全てが含まれる場合には、前記検索条件に含まれる複数の単語のそれぞれについて、単語間の依存関係を記述するための依存構造パターンの初期状態である初期依存構造パターンを生成する依存構造パターン初期化手段、および、前記例文において、該初期依存構造パターンを含む依存構造パターンが２つ存在し、該２つの依存構造パターンにおける主辞が依存関係を有する場合に、該２つの依存構造パターンを結合する操作を実行することにより１つの依存構造パターンを生成する依存構造パターン結合手段を含む依存構造パターン生成手段と、（ｃ）前記依存構造パターン生成手段によって、前記検索条件に含まれる複数の単語のすべての単語間の依存構造を一の依存構造パターンで生成できた例文を前記検索条件に該当するものとして選択する文選択手段とを有することを特徴とする。このようにすれば、検索条件は検索しようとする文に含まれる複数の単語と該単語の順序とに基づいて設定され、前記複数の例文のそれぞれに対し、設定された検索条件に含まれる複数の単語の全てが含まれるか否かを判断し、前記検索条件に含まれる複数の単語の全てが含まれる場合には、前記検索条件に含まれる複数の単語のそれぞれについて、単語間の依存関係を記述するための依存構造パターンの初期状態である初期依存構造パターンが生成され、また、前記例文において、該初期依存構造パターンを含む依存構造パターンが２つ存在し、該２つの依存構造パターンにおける主辞が依存関係を有する場合には、該２つの依存構造パターンを結合する操作を実行することにより１つの依存構造パターンが生

成される。そのため、前記複数の例文のうち、前記検索条件に含まれる複数の単語のすべての単語間の依存構造を一の依存構造パターンで生成できた例文が前記検索条件に該当するものとして選択されるので、好適に構文的に検索を行うことができる。

【００１４】

また、請求項６に係る発明によれば、好適には、予め蓄積された言語資料中の複数の例文、該例文を構成する単語間の依存関係についての情報である依存構造情報、および、該例文を構成する単語の品詞から、検索条件に該当する例文を検索する例文検索装置であって、（ｄ）前記検索条件設定手段は、前記検索しようとする文に含まれる複数の単語のうち少なくとも１つの品詞を、前記単語に加えて、あるいは該単語に替えて前記検索条件として設定するものであり、（ｅ）前記依存構造パターン生成手段は、該例文に前記検索条件に含まれる複数の単語の全てが含まれ、かつ該複数の単語のうち少なくとも１つの単語について該単語に加えて設定された品詞が一致する場合、あるいは、該例文に前記検索条件に含まれる複数の単語の全てが含まれ、かつ該複数の単語のうち少なくとも１つの単語に替えて設定された品詞が一致する場合に、前記依存構造パターンを生成することを特徴とする。このようにすれば、例文の検索において品詞の一致性についても考慮されることから、一層正確な例文の検索が可能となり、あるいは、検索条件をあいまいにした例文の検索が可能となる。

【発明を実施するための最良の形態】

【００１５】

以下、本発明の好適な実施の形態について図面を参照しつつ詳細に説明する。

【実施例】

【００１６】

図１は、本発明の一実施例である所謂コンピュータから主体的に構成される例文検索装置１０を示している。この例文検索装置１０は、よく知られたＣＰＵ、ＲＯＭ、ＲＡＭ、ＨＤＤ、入出力インターフェース等を有するコンピュータ本体１２、キーボードなどの入力操作装置１４およびＣＲＴ等の画像表示装置１６等を備えたコンピュータであり、ＣＰＵは入力操作装置１４の操作に応答して予め記憶されたプログラムを実行し、演算結果を画像表示装置１６の画面に表示させる。

【００１７】

入力操作装置１４は、キーボードあるいはマウス等により構成され、後述する検索条件設定手段３０において検索条件として設定入力される複数の単語の列をユーザが入力するのに用いられる。出力装置１６は、例えば画像表示装置であり、後述する検索結果格納手段３６に検索結果として格納された例文を適宜表示する。

【００１８】

図３は、その例文検索装置１０を、予め媒体２０に記憶された例文検索プログラムを該媒体２０から読み込み実行可能とすることにより得られる制御機能の要部を表す機能ブロック線図を模式的に示している。

【００１９】

検索条件設定手段３０は、入力操作装置１４によって入力された単語の列およびその品詞を、所定の形式に配列し、検索条件として後述する依存構造パターン生成手段３４に渡す。ここで、検索条件として入力される単語の一部は単語が特定されることなく単語の品詞のみが与えられてもよい。すなわち、検索条件設定手段３０によって設定される検索条件ｑは、

ｑ＝（ｑ１，ｑ２，…，ｑｍ）

のように表現され、ここで、ｑ１，ｑ２，…，ｑｍはｍ個の単語とその品詞との組をそれぞれ示している。また、単語を特定せず品詞のみを特定する場合には単語部分にはその旨を示すデータが記載される。

【００２０】

言語資料蓄積手段３２は、言語資料、いわゆるコーパスが記憶されたハードディスク等

の記憶装置に相当し、そこには複数の例文が蓄積されている。ここで、この複数の例文のそれぞれは、その例文を構成する単語間の依存関係が予め解析されており、依存構造情報としてその例文と共に蓄積されている。また、その例文を構成する単語の品詞についても予め解析されており、その例文と共に蓄積されている。上記複数の例文としては、たとえば、Marcus, M. P. and Santorini, B. and Marcinkiewicz, M. 著 Building a Large Annotated Corpus of English: the Penn Treebank,(Computational Linguistics誌, Vol. 19, No.2, pp.310-330, 1993）による英語新聞記事のコーパスなどが該当し、前記依存構造情報としては、このコーパスに対し、たとえば、Collins, M. 著 Head-Driven Statistical Models for Natural Language Parsing, Ph.D Dissertation, University of Pennsylvania, 1999に提案の方法に従ってコーパス中の各文を構成する単語間の依存関係を解析した結果が該当する。このようにして、言語資料蓄積手段３２に蓄積された言語資料Ｃには複数の例文ｓと、それに対応する依存構造情報Ｄが蓄積されている。すなわち、言語資料Ｃにｐ個の例文が蓄積されているとき、

Ｃ＝（ｓ１，Ｄ１，ｓ２，Ｄ２，…，ｓｐ，Ｄｐ）

のように表記される。ここで例文ｓは、ｎ個の単語により構成されるとき、

ｓ＝（ｗ１，ｗ２，…，ｗｎ）

のように表記され、ここで、ｗ１，…，ｗｎは単語とその品詞との組である。

【００２１】

上記言語資料Ｃに含まれるある例文ｓについて、その例文ｓと依存構造情報Ｄ、およびその例文を構成する単語の品詞の組ｗの例を図３に示す。図３のように、１つの例文に対し、その例文を構成する複数の単語間の依存関係が依存構造として矢印で示されている。また、便宜的に文頭から単語毎に１から始まる番号がその例文を構成する単語に付される。このとき、この依存構造の一つが、たとえば左からｉ番目の単語が左からｊ番目の単語に依存するものであるとき、この番号を用いてｉ→ｊの様に表記する。このようにすれば、図３の矢印で表された依存構造情報Ｄは、１→２、３→２、４→２、５→６、６→４、７→６、８→９、９→７のように表現される。

【００２２】

依存構造パターン生成手段３４は、言語資料蓄積手段３２において蓄積される複数の例文のそれぞれと、そのそれぞれの例文に対応する依存構造情報に基づいて、検索条件設定手段３０において設定された検索条件を構成する複数の単語について、その複数の単語間の依存関係を、後述する依存構造パターンによって最終的に一の依存構造パターンによって記述できるか否かを試みる。

【００２３】

依存構造パターンとは、検索条件設定手段３０において設定された検索条件ｑに含まれる複数の単語ｑ１，…，ｑｍが、言語資料蓄積手段３２中の言語資料Ｃに含まれる各例文ｓ１，…，ｓｐのそれぞれにおいて、どのような依存関係を有しつつ用いられているかを示すものであり、ｄ＝（ｈ，Ｌ，Ｒ）のように表記される。ここで、ｈは単語の位置、すなわち、各例文における先頭（左）から何番目の単語であるかを示す整数値であり、ＬおよびＲは依存構造パターンのリストである。Ｌに記載された依存構造パターンがある場合、その主辞が左からｈに依存することを意味しており、Ｒに記載された依存構造パターンがある場合、その主辞が右からｈに依存することを意味する。このとき、ＬおよびＲにそれぞれ依存構造パターンが重畳的に記載されることが可能であり、そのように記載されることにより、複数の単語の重畳的な依存関係を記載することができる。また、ｈに対してそれぞれ左または右から依存する単語が存在しないときには、ＬまたはＲにはその旨を表すεが記載される。

【００２４】

依存構造パターン生成手段３４は、以下に述べる初期化操作、結合操作、および補完操作の３つの操作を行う初期化操作手段、結合操作手段、および補完操作手段を備え、それら３つの手段により、依存構造パターンの生成を試みる。初期化操作手段は、言語資料中の各例文において検索条件に含まれる複数の単語および品詞の組の間の依存関係を記述す

る準備として、各単語および品詞の組に対応する初期依存構造パターンを各文ごとに生成する（初期化操作）。具体的には、言語資料Ｃ中のある例文ｓに対して、検索条件ｑに含まれる単語および品詞の組ｑｉ（１≦ｉ≦ｍ）のそれぞれがその例文ｓ中に一致するものがあるかを探し、その結果、例文ｓ中の単語とその品詞との組ｗｊ（１≦ｊ≦ｎ）と一致するならば、ｑｉに対する初期依存構造パターンとして（ｊ，ε，ε）を生成する。これを検索条件ｑに含まれるすべての単語とその品詞との組ｑｉについて行う。この結果、すべての単語とその品詞との組ｑｉについて初期依存構造パターンの生成を行うことができれば、その例文ｓは、少なくとも検索条件ｑに含まれる単語および品詞の組をその文中に含むものであると判断され、続いて結合操作および補完操作が試みられる。一方、すべての単語とその品詞との組ｑｉについて初期依存構造パターンの生成を行うことができなかった場合には、その例文ｓは検索条件ｑに含まれる単語とその品詞との組ｑｉの全てを含むものではないため、検索条件に合致する例文ではないと判断される。

【００２５】

結合操作手段は、２つの依存構造パターンが存在し、それらの主辞が依存関係を有する場合に、それら２つの依存構造パターンを結合することにより１つの依存構造パターンとする（結合操作）。この操作の様子を図示したのが図４である。図４において、三角形で表された記号５０は依存構造パターンを表している。具体的には、いま、検索条件の一部分であるｑｉ，…，ｑｊおよびｑｊ＋１，…，ｑｋに対する依存構造パターンとして、それぞれｄ＝（ｈ，Ｌ，Ｒ）およびｄ’＝（ｈ’，Ｌ’，Ｒ’）が存在し、かつ、依存構造パターンｄの右端の枝を可能な限りたどっていき、たどり先がなくなったところにある主辞の位置が、構造パターンｄ’の左端の枝を可能な限りたどっていき、たどり先がなくなったところにある主辞の位置よりも左にある場合において、（図４（ａ））、それらの主辞であるｈおよびｈ’の関係が、ｈがｈ’に依存する関係である場合、すなわちｈ→ｈ’で、かつＲ’＝εである場合、には、これらの２つの依存構造パターンを結合し、新たに、検索条件の一部分であるｑｉ，…，ｑｋに対する構造パターンとしてｄ’’＝（ｈ’，ｄＬ’，Ｒ’）を生成する（図４（ｂ））。また、ｈ’がｈに依存する関係、すなわちｈ’→ｈの場合には、検索条件の一部分であるｑｉ，…，ｑｋに対する構造パターンとしてｄ’’＝（ｈ，Ｌ，Ｒｄ’）を生成する（図４（ｃ））。ここで、結合操作が行われる条件としてＲ’＝εが必要とされるのは、同じ依存構造パターンを重複して生成しないようにするためである。この結合操作を繰り返すことにより、依存関係を有する複数の依存構造パターンは順次結合され、最終的に１つの依存構造パターンで記述されることとなる。

【００２６】

このように、結合操作により、検索条件ｑ中の各単語とその品詞との組のそれぞれについての初期依存構造パターンが、他の依存構造パターンと順次結合されることにより、最終的に１つの依存構造パターンとして記述される場合、それは、検索条件ｑ中に含まれるすべての単語とその品詞との組が、言語資料Ｃ中のその例文において、依存関係を持つ形で含まれることを意味している。したがって、すべての初期依存構造パターンを最終的に１つの依存構造パターンとして記述される場合には、その文は、検索条件に合致するものとされる。

【００２７】

続いて、補完操作手段について説明する。たとえば、単語とその品詞との組の２つを検索条件とする場合であって、その２つの単語には直接依存関係を持たない文を検索したい場合を考える。このような場合には、言語資料中のある例文について２つの初期依存構造パターンを作成できたとしても、それらが直接的には依存関係を有さないので、上述の結合操作を行うことができず、それらを１つの依存構造パターンで記述することはできない。しかしながら、それら２つの依存構造パターンの主辞が、例文には含まれるものの検索条件に含まれない単語を介して依存している場合には、その単語を主辞とする依存構造パターンの生成を行うことで、結合操作が可能となることがある。このような操作を行うのが補完操作手段である。また、この補完操作の様子を示したのが図５である。具体的には、検索条件の一部分であるｑｉ，…，ｑｊに対する依存構造パターンｄが存在し、その主

辞をｈとする。このとき、依存構造パターンｄの主辞ｈが依存する関係にあるｈ’が存在するとき、すなわちｈ→ｈ’であるとき、ｈ＜ｈ’、すなわち、例文ｓにおいてｈに対応する単語がｈ’に対応する単語よりも左にあるならば、ｑｉ，…，ｑｊに対する依存構造パターンとしてｄ＊＝（ｈ’＊，ｄ，ε）を生成する（図５（ａ））。また、ｈ＞ｈ’すなわち、例文ｓにおいてｈに対応する単語がｈ’に対応する単語よりも右にあるならば、ｑｉ，…，ｑｊに対する依存構造パターンとしてｄ＊＝（ｈ’＊，ε，ｄ）を生成する（図５（ｂ））。このとき、添字＊は、主辞ｈ’が補完操作により導入されたものであることを示す。このようにすれば、検索条件中の単語間に直接の依存関係がないような場合であっても補完操作により生成された依存構造パターンを用いて上記の結合操作を行うことによって、初期依存構造パターンから最終的に一つの依存構造パターンを生成することができることが期待できる。

【００２８】

ただし、補完操作を行った場合であっても、補完操作によって生成された依存構造パターンの主辞であるｈ’と、他の依存構造パターンの主辞との間に依存関係がない場合には、補完操作を１回行ってもその後に結合操作を行うことはできない。

【００２９】

一方で、この補完操作を無制限に繰り返すことができれば、複数個の単語を介して間接的に依存する２つの単語についても結合操作を行うことが可能となる。このように、複数回の補完操作を経て可能となった結合操作によって一の依存構造パターンが生成された場合、たとえ最終的に一の依存構造パターンが生成されたとしても、検索条件ｑに含まれるすべての単語とその品詞との組とが言語資料Ｃの中のその例文に含まれるとしても、検索条件ｑにおいて意図した依存関係とは異なる依存関係を持つ形で含まれる可能性がある。そして、この可能性は、補完操作を多く行う程高くなる。

【００３０】

そこで、この補完操作を行う回数を制限することが行われる。この制限は、たとえば、予め最大値のみを与えておき、依存構造パターン生成手段３４が自動的に結合操作を行いつつ、その制限の範囲内で適宜補完操作を行うようにすればよい。

【００３１】

図６は、例文検索装置１０の制御作動の要部を表すフローチャートである。以下本フローチャートに沿って例文検索装置１０の作動を説明する。

【００３２】

前記検索条件設定手段３０および検索条件設定工程に対応するステップ（以下「ステップ」を省略する。）ＳＡ１においては、検索条件となる複数の単語が入力操作装置１４等により入力され、これを上述の様式である検索条件ｑとされる。本実施例においては、たとえば、入力操作装置１４等により入力された検索条件ｑが「ｉｔ（代名詞），ｉｓ（ｂｅ動詞），ｆｏｒ（前置詞），ｔｏ（前置詞）」であった場合を考える。これは、これら４つの単語とその品詞との組を依存関係を有する状態で文中に含む例文を言語資料蓄積手段３２における言語資料Ｃから検索することを意味する。

【００３３】

ＳＡ２においては、図７または図９に示す依存構造パターン生成ルーチンが実行される。ここで、図７は依存構造パターン生成手段３４が上述の初期化操作工程および結合操作工程のみを行う場合の依存構造パターン生成ルーチンであり、図９は依存構造パターン生成手段３４が初期化操作工程、結合操作工程に加え、補完操作工程を行う場合の依存構造パターン生成ルーチンである。まず、図７の依存構造パターン生成ルーチンが採用される場合の例について説明する。図７におけるＳＢ１では言語資料蓄積手段３２に対応する記憶装置において、言語資料Ｃとして蓄積されている複数の例文の中から１の例文ｓが抽出される。また、併せて保存蓄積されている、その例文の依存構造情報およびその例文を構成する単語の品詞についての情報についても抽出される。たとえば、「Ｉｔ　ｉｓ　ｉｍｐｏｒｔａｎｔ　ｆｏｒ　ｕｓ　ｔｏ　ｈａｖｅ　ｓｕｃｈ　ｔｅｃｈｎｏｌｏｇｙ．」という文ｓがその依存構造情報Ｄと共に抽出される。この例文と依存構造情報、および品

詞についての情報は図３に示されるものである。

【００３４】

続くＳＢ２～ＳＢ６は依存構造パターン生成手段３４に対応する。ＳＢ２においては、上述の初期化操作が可能であるかが判定される。すなわち、検索条件を構成する単語とその品詞との組の全てが、抽出された例文に含まれるか否かが判断される。そして、このＳＢ２の判断が肯定されれば、前記初期化操作工程に対応するＳＢ３において初期化操作が行われ、ＳＢ２の判断が否定された場合には、その抽出された例文は依存構造パターンを生成し得ないものであるから検索条件を満たさないものと判断され、依存構造パターン生成ルーチンが終了させられる。本実施例においては、検索条件ｑに含まれる４つの単語およびその品詞の組はいずれも文ｓに含まれることから、ＳＢ２の判断は肯定され、ＳＢ３に移る。

【００３５】

ＳＢ３の初期化操作では、検索条件を構成する単語とその品詞との組の全てについて、初期依存構造パターンが生成される。本実施例においては、検索条件ｑに含まれる単語のうち、「ｉｔ（代名詞）」については、文ｓを構成する左から１番目の単語である「Ｉｔ（代名詞）」と一致することから、これに対応する初期依存構造パターンとして、

（１，ε，ε）　・・・（１）

が生成される。同様にして、検索条件ｑ中の「ｉｓ（ｂｅ動詞）」、「ｆｏｒ（前置詞）」、「ｔｏ（前置詞）」に対してそれぞれ初期依存構造パターン

（２，ε，ε）　・・・（２）

（４，ε，ε）　・・・（３）

（６，ε，ε）　・・・（４）

が生成される。なお、このとき、検索条件ｑに含まれる単語およびその品詞と、文ｓに含まれる単語およびその品詞との一致を照合する際には、単語の変化、例えば、三人称単数による動詞の変化や時制による動詞の変化、単数および複数による名詞の変化についても考慮し、これらの差異は同じ単語であるように認識する。

【００３６】

ＳＢ４においては、依存構造パターンについて、結合操作が可能かどうかが判断される。すなわち、存在する依存構造パターンのそれぞれの主辞と、文ｓの依存構造情報とが比較され、依存関係にある２つの主辞の組があるかが判断される。ＳＢ４において依存関係にある主辞の組があると判断されれば、その依存関係にある２つの主辞についての依存構造パターンを結合すべくＳＢ５に移る。なお、依存関係にある主辞の組が複数ある場合には、依存関係の末端にある単語、すなわち、他の単語から依存されることのない単語に相当する主辞が依存する依存関係から結合操作を実行する。一方、依存関係にある主辞がないと判断された場合には、なし得るすべての結合操作を完了したとして、ＳＢ６に移る。

【００３７】

本実施例においては、ＳＢ３において生成された４つの初期依存構造パターンの主辞１、２、４、６について依存構造情報Ｄを参照すると、１→２、４→２、６→４の３つの依存関係があることがわかる。ここで、この３つの依存関係のうち、１→２および６→４の依存関係において依存元に相当する１および６には他の依存関係の依存先とはなっていない一方で、４→２については６→４の依存関係の依存先となっているため４→２の依存関係についての結合操作よりも１→２および６→４の依存関係の結合操作のほうが先に行われることとなる。この場合において、１→２の依存関係と６→４の依存関係に対応する結合操作はいずれが先に行われても良いが、例えば、文ｓのより左に依存元を有する依存関係に対応する結合操作を先に行うよう定めるようにすればよい。以上より、ＳＢ４においては、１→２の依存関係に対応する初期依存構造パターン、すなわち１および２を主辞とする初期依存構造パターンを結合することができるとしてＳＢ５に移る。

【００３８】

結合操作工程に対応するＳＢ５においては、ＳＢ４において結合操作を行う様に判断された２つの依存構造パターンが結合される。本実施例においては、上記（１）で記述され

た１を主辞とする初期依存構造パターンと上記（２）で記述された２を主辞とする初期依存パターンが結合され、依存構造パターン
（２，（１，ε，ε），ε）　・・・（５）
が生成される。

【００３９】
続いてＳＢ４に戻り、再度結合可能な依存構造パターンがないかが判断される。本実施例においては、ＳＢ５において（１）、（２）の初期依存パターンが結合された結果、（３）、（４）の初期依存パターンおよび（５）の依存構造パターンが存在している。すなわち、その主辞は４、６、２であるから、文ｓにおける依存構造情報Ｄに基づいて６→４、４→２の依存関係があるが、上述のように４→２の依存関係よりも６→４の依存関係に対応する依存構造パターンの結合操作が先に行われる。したがって、６→４の依存関係に対応する初期依存構造パターン、すなわち６および４を主辞とする初期依存構造パターンを結合することができるとしてＳＢ５に移る。

【００４０】
ＳＢ５においては、先の場合と同様にして結合操作が行われる。すなわち、（３）で記述された４を主辞とする初期依存構造パターンと（４）で記述された６を主辞とする初期依存パターンが結合され、依存構造パターン
（４，ε，（６，ε，ε））　・・・（６）
が生成される。

【００４１】
そして、再度ＳＢ４に戻り、結合可能な依存構造パターンがないかが判断される。この段階においては、（５）および（６）で記述された２つの依存構造パターンが存在しており、その主辞は２および４である。ここで、文ｓの依存構造情報Ｄによれば４→２の依存関係があるので、この２つの依存構造パターンを結合することができるとしてＳＢ５に移る。

【００４２】
ＳＢ５においては、先の場合と同様にして結合操作が行われる。すなわち、（５）で記述された２を主辞とする依存構造パターンと（６）で記述された４を主辞とする依存パターンが結合され、依存構造パターン
（２，（１，ε，ε），（４，ε，（６，ε，ε））　・・・（７）
が生成される。

【００４３】
ＳＢ４に戻り、再度結合可能な依存構造パターンがないかが判断される。この段階においては、（７）で記述された１つの依存構造パターンしか残っておらず、これ以上の結合操作を行うことはできない。したがって、ＳＢ４の判断が否定され、ＳＢ６に移る。

【００４４】
ＳＢ６においては、ＳＢ４～ＳＢ５における結合操作の結果、検索条件ｑに対する依存構造パターンが生成できたかが判断される。具体的には、複数の初期依存構造パターンから結合操作により１つの依存構造パターンが生成されたか否かによて判断される。この判断が肯定された場合には、文ｓは検索条件ｑを満たすものとして、ＳＢ７に移る。一方、この判断が否定された場合には、文ｓは検索条件ｑを満たさないものとして、依存構造パターン生成ルーチンを終了する。本実施例においては、上記（１）～（４）として生成された初期依存構造パターンが、結合操作の結果、（７）で記述された１つの依存構造パターンとして生成されているため、この判断が工程され、ＳＢ７に移る。

【００４５】
文選択手段３６および文選択工程に対応するＳＢ７においては、ＳＢ６における判断が肯定された文ｓについて、検索条件ｑを満たすものとして、その文ｓが検索結果格納手段３８に格納される。

【００４６】
図６に戻って、ＳＡ３においては、言語資料Ｃ中の全ての文についてＳＡ２の依存構造

パターン生成ルーチンが実行されたかが判断され、この判断が否定される場合にはＳＡ２に戻る。一方、この判断が肯定された場合にはＳＡ４に移る。これにより、言語資料Ｃ中のすべての文について依存構造パターンの生成を試みる。

【００４７】

ＳＡ４においては、検索条件ｑに合致する文として検索結果格納手段３８に格納された文が出力装置１６を通して出力される。

【００４８】

例文検索装置１０がこのように作動することにより、検索条件ｑに含まれる単語と品詞の組が単に含まれるのみならず、それらの単語がその内部において依存関係を有する文を検索結果とすることができる。例えば、上記ＳＢ１において、別の文ｓ’「Ｉｔ ｉｓ ｃｌｅａｒ ｗｈｅｔｈｅｒ ｓｕｐｐｏｒｔ ｆｏｒ ｔｈｅ ｐｒｏｐｏｓａｌ ｗｉｌｌ ｂｅ ｂｒｏａｄ ｅｎｏｕｇｈ ｔｏ ａ ｓｅｒｉｏｕｓ ｃｈａｌｌｅｎｇｅ．」が抽出された場合を考える。この文を構成する単語間の依存関係と単語の品詞を表したのが図８である。この場合、検索条件ｑの４つの単語および品詞の組はいずれも文ｓ’に含まれることからＳＢ２の判断が肯定される。したがって、ＳＢ３において初期化操作が行われ、

（１，ε，ε）　・・・（８）
（２，ε，ε）　・・・（９）
（６，ε，ε）　・・・（１０）
（１３，ε，ε）　・・・（１１）

の４つの初期依存構造パターンが生成される。続くＳＢ４において、これらの主辞と文ｓ’の依存構造情報Ｄを比較すると１→２の依存構造があることがわかるので、ＳＢ５において（８）および（９）の初期依存構造パターンが結合操作され、

（２，（１，ε，ε），ε）　・・・（１２）

が生成される。

【００４９】

続いて再度ＳＡ４に戻って、他に結合可能な依存構造パターンがないかが判断されるが、このとき（１０）、（１１）、（１２）の主辞である６、１３、２には依存関係がなく、これ以上結合操作を行うことはできないと判断される。従って、ＳＢ６に移り、検索条件に対応する依存構造パターンが生成できたかが判断されるが、このとき、依存構造パターンは（１０）、（１１）、（１２）の３つが存在しており、１つの依存構造パターンで記述できていない。従って、ＳＢ６の判断が否定され、文ｓ’は検索結果に含まれることなく依存構造パターン生成ルーチンが終了させられる。すなわち、検索条件ｑに含まれるすべての単語を単に含むのみであって、それらの単語が文中において依存関係を有さない状態で存在する場合には、その文は検索結果に含まれることがない。

【００５０】

続いて、別の実施例について説明する。以下の説明において、実施例相互に共通する部分には同一の符号を付して説明を省略する。

【００５１】

図９は、図６における依存構造パターン生成ルーチンの別の作動を表すフローチャートであり、図７のフローチャートに代えて用いられるものである。本図のフローチャートにおいては、依存構造パターン生成ルーチンが初期化操作、結合操作に加え、補完操作を行う点において図７のフローチャートと相違する。なお、上述のように、補完操作を行う場合には、操作者が補完操作を行う回数の上限を予め設定しておく必要がある。本実施例では、この上限として例えば２回が設定される。

【００５２】

検索条件設定手段３０および検索条件設定工程に対応するＳＡ１においては、検索条件となる複数の単語が入力操作装置１４等により入力され、これを上述の様式である検索条件ｑとされる。本実施例においては、たとえば、入力操作装置１４等により入力された検索条件ｑ’が「ｃｏｍｂｉｎｅｓ（動詞），ａｎｄ（接続詞）」であった場合を考える。

これは、これら２つの単語と品詞の組を依存関係を有する状態で文中に含む例文を言語資料蓄積手段３２における言語資料Ｃから検索することを意味する。

【００５３】

続くＳＡ２においては、依存構造パターン生成ルーチンとして図９のフローチャートが実行される。図９におけるＳＣ１は、前述のＳＢ１同様、言語資料蓄積手段３２に対応し、言語資料Ｃとして蓄積されている複数の例文の中から１の例文ｓ’’が抽出される。また、併せて保存蓄積されている、その例文の依存構造情報およびその例文を構成する単語の品詞についての情報についても抽出される。たとえば、「Ｏｐｅｒａ　ｃｏｍｂｉｎｅｓ　ｍｕｓｉｃ　ａｎｄ　ｄｒａｍａ．」という文ｓ’’がその依存構造情報Ｄと共に抽出される。この例文と依存構造情報、および品詞についての情報は図１０に示されるものである。

【００５４】

続くＳＣ２～ＳＣ９は依存構造パターン生成手段３４に対応する。ＳＣ２においては、ＳＢ２同様、初期化操作が可能であるかが判定される。ＳＣ２における判断が肯定されれば、ＳＣ３において初期化操作が実行され、ＳＣ２における判断が否定されれば、その文は検索条件を満たさないものと判断され、依存構造パターン生成ルーチンが終了させられる。本実施例においては、検索条件ｑ’に含まれる２つの単語および品詞の組はいずれも文ｓ’’に含まれることからＳＣ２の判断は肯定され、ＳＣ３に移る。

【００５５】

初期化操作工程に対応するＳＣ３においては，ＳＢ３同様、初期化操作が行われる。すなわち、検索条件ｑ’に含まれる単語であるｃｏｍｂｉｎｅｓとａｎｄのそれぞれについて、初期依存構造パターン

（２，ε，ε）　・・・（１３）

（４，ε，ε）　・・・（１４）

が生成される。

【００５６】

ＳＣ４においては、ＳＢ４同様、複数の依存構造パターンの中に、結合操作が可能なものがあるかどうかが判断される。本判断が肯定される場合はＳＣ５に移り、結合操作工程に対応するＳＣ５においては、ＳＢ５同様、依存構造パターンの結合操作が行われる。また、本判断が否定されれば、ＳＣ６に移る。本実施例においては、文ｓ’’の依存構造情報Ｄを参照すると、２→４も４→２の依存関係もない。すなわち、初期依存パターンの主辞である２と４の間には直接の依存関係は存在しない。したがって、ＳＣ４の判断が否定され、ＳＣ６に移る。

【００５７】

ＳＣ６においては、検索条件ｑ’に対する依存構造パターンが生成できたかが、複数の初期依存構造パターンから１つの依存構造パターンが結合操作により生成されたかによって判断される。この判断が肯定されれば、文ｓ’’は検索条件ｑ’を満たすものとしてＳＣ１０に移る。一方、この判断が否定された場合には、続くＳＣ７に移る。本実施例においては、この時点で（１３）および（１４）で記述される２つの依存構造パターンが存在するので、ＳＣ６の判断は否定されＳＣ７に移る。

【００５８】

ＳＣ７においては、これまでに行った補完操作の回数がその回数の制限値よりも小さいかを判断する。本判断が肯定されれば、さらに補完操作を行うことが可能であるとしてＳＣ８に移る。一方本判断が否定される場合は、文ｓ’’に対してそれ以上の補完操作は行うことができないとして、依存構造パターン生成ルーチンは終了させられる。本実施例においては、この時点で未だ補完操作を行っておらず、また、補完操作の上限は２回と設定されていることから、ＳＡ７における判断は肯定されＳＣ８に移る。

【００５９】

ＳＣ８においては、依存構造パターンにおいて、補完操作が可能かどうかを判定する。これは、文ｓ’’の依存構造情報Ｄをもとに、依存構造パターンの主辞ではない文ｓ’’

の１つの単語を介して間接的に依存構造パターンの主辞が依存関係をとるような単語が文ｓ’’中に存在するか否かによって判断する。具体的には、本実施例においては、依存構造パターンの主辞は２および４であるが、文ｓ’’の依存構造情報Ｄにおいては、２→４も４→２も存在しない。しかしながら、依存構造情報Ｄには４→３という依存関係が存在するため、補完操作が可能であると判断される。以上のように、本実施例においては、ＳＣ８の判断が肯定され、ＳＣ９に移る。

【００６０】

補完操作工程に対応するＳＣ９においては、ＳＣ８において補完が可能であると判断される根拠となった依存関係に基づいて、上述の補完操作が行われる。本実施例においては、初期依存構造パターン（１４）に対して、４→３という依存関係に基づいて補完操作を行うと、依存構造パターン

（３＊，ε，（４，ε，ε））　・・・（１５）

が生成される。

【００６１】

続いて、再びＳＣ４に戻り、依存構造パターンについて結合操作が可能かが判断される。本実施例においては（１３）および（１５）の依存構造パターンが存在しており、その主辞は２および３である。ここで、文ｓ’’の依存構造情報Ｄには３→２が含まれることから、これらの依存構造パターンは結合操作が可能であるとしてＳＣ４の判断は肯定される。

【００６２】

ＳＣ５においては、ＳＣ４において結合可能であると判断された２つの依存構造パターンが結合操作により結合される。本実施例においては、（１３）で記述された２を主辞とする依存構造パターンと（１５）で記述された３を主辞とする依存構造パターンが結合され、依存構造パターン

（２，ε，（３＊，ε，（４，ε，ε））　・・・（１６）

が生成される。

【００６３】

その後、再度ＳＣ４に戻り、さらに結合可能な依存構造パターンがないかが判断される。この段階においては、（１６）で記述された１つの依存構造パターンしか残っていないため、これ以上の結合操作を行うことはできないとして、ＳＣ４の判断は否定され、ＳＣ６に移る。

【００６４】

ＳＣ６においては、検索条件ｑ’に対する依存構造パターンが生成できたかが、複数の初期依存構造パターンから１つの依存構造パターンが結合操作により生成されたかによって再度判断される。本実施例においては、この時点においては、（１３）および（１４）で記述される２つの初期依存構造パターンから、補完操作および結合操作によって（１６）で記述される１つの依存構造パターンが生成されたので、この判断が肯定され、文ｓ’’は検索条件ｑ’を満たすものとしてＳＣ１０に移る。

【００６５】

文選択手段３６に対応するＳＣ１０においては、ＳＣ６における判断が肯定された文ｓ’’について、検索条件ｑ’を満たすものとして、その文ｓ’’を検索結果格納手段３８に格納する。

【００６６】

続いて、ＳＣ７に移り、更に補完操作が可能かどうかについて、すでに行った補完回数と予め設定されたその上限値の大小により判断される。本実施例においては、すでに行った補完操作の回数が１回である一方、補完操作の上限回数が２回と設定されていることから、本判断は肯定され、ＳＣ８に移る。

【００６７】

ＳＣ８においては、先に行ったＳＣ８と同様、依存構造パターンにおいて、補完操作が可能かどうかを判定し、本判断が肯定されればＳＣ９に移り、再度の補完操作を実行する

。一方本判断が否定されれば、これ以上の補完操作は行うことができないとして、本ルーチンは終了させられる。本実施例においては、依存構造パターンの主辞は２であるが、文ｓ’’の依存構造情報Ｄにおいては、２が依存する単語が存在しないことから、補完操作の対象となる主辞ｈ’＊が存在せず、補完操作をなし得ない。したがって、ＳＣ８の判断が否定され、これ以上の補完操作は行うことができないとして、本ルーチンが終了させられる。

【００６８】

図６に戻って、ＳＡ３においては、言語資料Ｃ中の全ての文についてＳＡ２の依存構造パターン生成ルーチンが実行されたかが判断され、この判断が否定される場合にはＳＡ２に戻る。一方、この判断が肯定された場合にはＳＡ４に移る。これにより、言語資料Ｃ中のすべての文について依存構造パターンの生成を試みる。

【００６９】

ＳＡ４においては、検索条件ｑ’に合致する文として検索結果格納手段３８に格納された文が出力装置１６を通して出力される。

【００７０】

表１および表２は、本実施例の例文検索装置１０と、従来技術である、検索条件ｑに含まれる単語を含む文を取りだす単純な方法による例文検索装置とについて、その検索結果の精度、再現率およびそれらの調和平均の値を比較するものである。ここで、この比較における言語資料Ｃとしては、上述のthe Penn Treebank を使用し、その言語資料中の文の依存構造情報は、上述のCollins, M. 著 Head-Driven Statistical Models for Natural Language Parsing, Ph.D Dissertation, University of Pennsylvania, 1999に提案の方法に従って解析した結果を用いた。

【００７１】

【表１】

補完操作の上限	精度	再現率	調和平均
0 回	100.0% (17/17)	81.0% (17/21)	89.5% (34/38)
1 回	90.5% (19/21)	90.5% (19/21)	90.5% (38/42)
2 回	70.0% (21/30)	100.0% (21/21)	82.4% (42/51)
3 回	60.0% (21/35)	100.0% (21/21)	75.0% (42/56)
（従来の方法）	27.3% (21/77)	100.0% (21/21)	42.9% (42/98)

【００７２】

表１は、検索条件を「ｉｔ（代名詞），ｉｓ（ｂｅ動詞），ｆｏｒ（前置詞），ｔｏ（前置詞）」とした場合の検索結果について、その精度、再現率およびそれらの調和平均を算出し、比較したものである。ここで、精度、再現率およびそれらの調和平均はそれぞれ、

精度　＝　検出された正解の数　／　検出結果の数

再現率　＝　検出された正解の数　／　真の正解の数

調和平均　＝（検出された正解の数×２）　／　（検出結果の数＋真の正解の数）

で表される数値をパーセント表示したものであり、検索装置の検索結果の精度を表す指標として通常用いられるものである。特に調和平均（「Ｆ値」ともいう。）を用いて評価される。なお、補完操作の上限を０回としたものは、依存構造パターン生成ルーチンとして図７のフローチャートを用いる場合に相当する。

【００７３】

表１によれば、補完操作の上限を０～３回のいずれとした場合であっても従来の手法による例文検索装置よりはよいＦ値が得られており、精度のよい検索が行われていることが確認される。

【００７４】

【表２】

補完操作の上限	精度	再現率	調和平均
0 回	0.0% (0/ 2)	0.0% (0/11)	0.0% (0/13)
1 回	55.0% (11/20)	100.0% (11/11)	71.0% (22/31)
2 回	42.3% (11/26)	100.0% (11/11)	59.5% (22/37)
3 回	32.4% (11/24)	100.0% (11/11)	62.9% (22/35)
（従来の方法）	24.5% (11/45)	100.0% (11/11)	39.3% (22/56)

【００７５】

表２は、検索条件を「ｃｏｍｂｉｎｅｓ（動詞），ａｎｄ（接続詞）」とした場合の検索結果について、表１と同様に比較したものである。この場合も従来の週報による例文検索装置よりは補完操作の上限に関わらずよいＦ値が得られ、精度のよい検索が行われていることが確認される。

【００７６】

このようにすれば、検索条件設定手段３０および検索条件設定工程ＳＡ１において設定される検索条件ｑは検索しようとする文に含まれる複数の単語と該単語の順序とに基づいて設定され、依存構造パターン生成手段３４および依存構造パターン生成工程ＳＢ２～ＳＢ６またはＳＣ２～ＳＣ９によって複数の単語の依存構造を表現する依存構造パターンｄの生成が試みられ、前記複数の例文のうち、該複数の例文のそれぞれに対し、前記検索条件に含まれるすべての単語間の依存構造を一の依存構造パターンで生成できたものが文選択手段３６および文選択工程ＳＢ７またはＳＣ１０によって前記検索条件ｑに該当するものとして選択されるので、好適に、予め蓄積された言語資料Ｃ中の複数の例文とその例文を構成する単語間の依存関係についての情報である依存構造情報Ｄとから、構文的に例文の検索を行うことができる。

【００７７】

また、言語資料蓄積手段３２は、蓄積された複数の例文のそれぞれについてその複数の例文を構成する各単語の品詞について予め解析し、その複数の例文のそれぞれと、その例文を構成する単語の品詞および単語間の依存構造Ｄについての情報をあわせて蓄積するものであり、検索条件設定手段３０および検索条件設定工程ＳＡ１は、検索しようとする文中の複数の単語の品詞を前記検索条件として設定するものであり、さらには、該複数の単語の一部については、単語が特定されることなく品詞のみが与えられること、依存構造パターン生成手段３４および依存構造パターン生成工程ＳＢ２～ＳＢ６またはＳＣ２～ＳＣ９は、前記例文中の単語の品詞の一致性についても考慮するものであるようにすれば、例文の検索において品詞の一致性についても考慮されることから、予め蓄積された言語資料Ｃ中の複数の例文とその例文を構成する単語間の依存関係についての情報である依存構造情報Ｄ、およびその単語の品詞から、一層正確な例文の検索が可能となる。

【００７８】

以上、本発明の実施例を図面に基づいて詳細に説明したが、本発明はその他の態様においても適用される。

【００７９】

例えば、本実施例においては、検索条件として単語およびその品詞が与えられたが、その一部については単語は必要とせず品詞のみを指定してもよい。検索条件の一部として品詞のみが与えられた場合には、上述の依存構造パターン生成手段３４における初期化操作手段において、その検索条件の一部である品詞が一致する単語が例文中に存在する場合に、初期依存構造パターンを生成し、その後の操作は同様に行うことによって、検索を行うことができる。このようにすれば、検索条件の一部を、単語を特定せず、品詞の一致のみとすることにより、検索条件の一部をあいまいにした検索をすることが可能となる。また、検索結果として、操作者が意図しない依存構造情報を有するにもかかわらず検索結果として抽出された例文が多い場合には、その検索における検索条件に加えて、単語を特定せ

ず品詞のみを特定する検索条件を付加して再度検索することにより、単語を特定せず品詞のみに基づく単語間の依存関係に基づいた検索が可能となり、特定の構文構造を有する例文のみに検索結果を絞り込むことができる。その結果として所望の検索結果を得ることが可能となる。

【００８０】

また、逆に、品詞は与えられず、単語のみから成る検索条件であってもよい。この場合、上述の依存構造パターン生成手段３４における初期化操作手段において、検索条件中に含まれる単語が例文中に含まれれば、その品詞の一致を要件とせず初期化操作を行い、その後の操作は同様に行うことによって検索を行うことができる。このとき、言語資料蓄積手段３２において蓄積される言語資料Ｃには、その言語資料に含まれる複数の文をそれぞれ構成する単語の品詞についての情報が含まれていたが、その単語の品詞についての情報を含まないものであってもよい。

【００８１】

上述の補完操作を行うにあたり、操作者が補完操作の回数の上限を予め定めておき、依存構造パターン生成手段３４がその範囲内において適宜補完操作を行うようにしたが、これに限られず、例えば、依存構造パターン生成手段３４がその上限を定めるようにしてもよく、あるいは、操作者が一旦定めた上限の条件下で例文検索装置１０を実行し、その結果に応じて適宜その上限の値を増減させることも可能である。

【００８２】

また、本実施例においては、言語資料蓄積手段３２は例文検索装置１０としてのコンピュータ１２の内部に配置されたが、これに限られず、例えば、ネットワークを介して別の場所に設けられた他のコンピュータ内に配置されてもよい。また、入力操作装置１４および出力装置１６についても、例文検索装置１０としてのコンピュータ１２に取り付けられたものが用いられたが、これに限られず、ネットワークを介して接続された他のコンピュータの入力操作装置や出力装置が用いられてもよい。

【００８３】

また、本実施例においては、例文検索装置１０は英語について例文の検索を行ったが、これに限られず、上述の依存構造情報のように単語間の依存関係を記述することができる言語であれば他の言語についても適用することができる。

【図面の簡単な説明】

【００８４】

【図１】本発明の実施例による例文検索装置の構成の概略を示す図である。

【図２】本発明の実施例による例文検索装置の制御機能の要部の概要を表す機能ブロック線図である。

【図３】言語資料蓄積手段に蓄積される例文と、その例文を構成する単語の依存関係および単語の品詞を表した図である。

【図４】結合操作の概要を表した図である。

【図５】補完操作の概要を表した図である。

【図６】本発明の例文検索装置の作動を表すフローチャートである。

【図７】図５における依存構造パターン生成ルーチンを表すフローチャートである。

【図８】言語資料蓄積手段に蓄積される別の例文と、その例文を構成する単語の依存関係および単語の品詞を表した図である。

【図９】図６のフローチャートに代えて用いられる、別の依存構造パターン生成ルーチンを表すフローチャートである。

【図１０】言語資料蓄積手段に蓄積される別の例文と、その例文を構成する単語の依存関係および単語の品詞を表した図である。

【符号の説明】

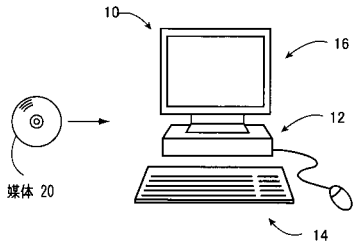
【００８５】

１０：例文検索装置（コンピュータ）

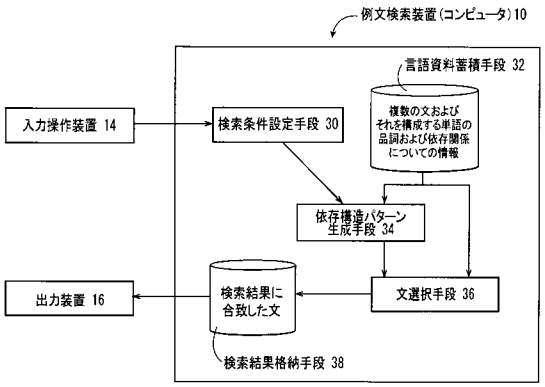
２０：媒体

３０：検索条件設定手段
３４：依存構造パターン生成手段
３６：文選択手段
ＳＡ１：検索条件設定工程
ＳＢ２～ＳＢ６，ＳＣ２～Ｓ９：依存構造パターン生成工程
ＳＢ７，ＳＣ１０：文選択工程

【図１】



【図２】

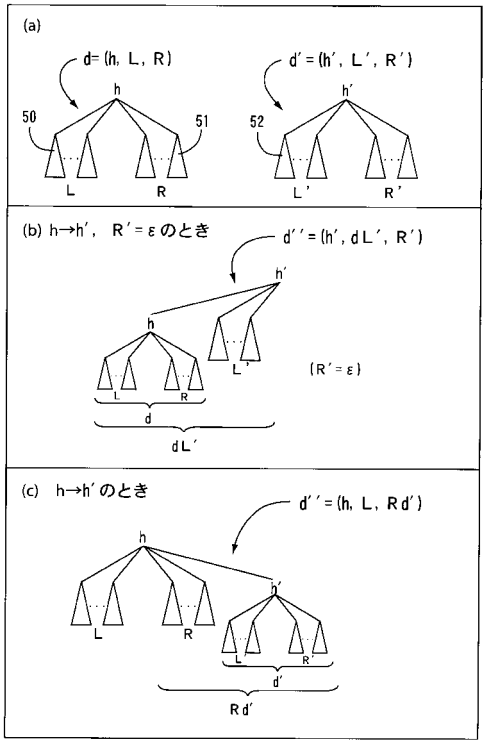


【図３】

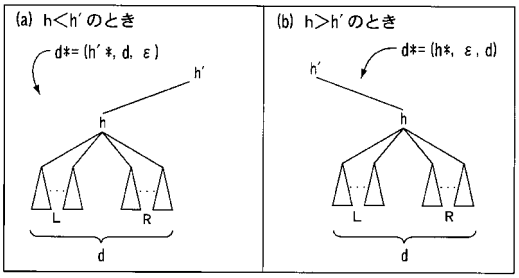
依存構造情報D: 1→2
3→2
4→2
5→6
6→4
7→6
8→9
9→7

文s=(w1, w2, w3, w4, w5, w6, w7, w8, w9)
w1= it, 代名詞
w2= is, be動詞
w3= important, 形容詞
w4= for, 前置詞
w5= us, 代名詞
w6= to, 前置詞
w7= have, 動詞
w8= such, 形容詞
w9= technology, 名詞

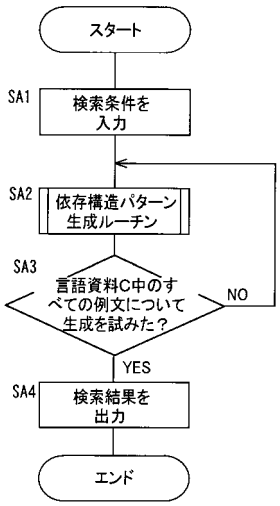
【図４】


(a)
d=(h, L, R)
d'=(h', L', R')
50
51
52
(b) h→h', R'=εのとき
d''=(h', dL', R')
(R'=ε)
dL'
(c) h→h'のとき
d''=(h, L, Rd')
Rd'

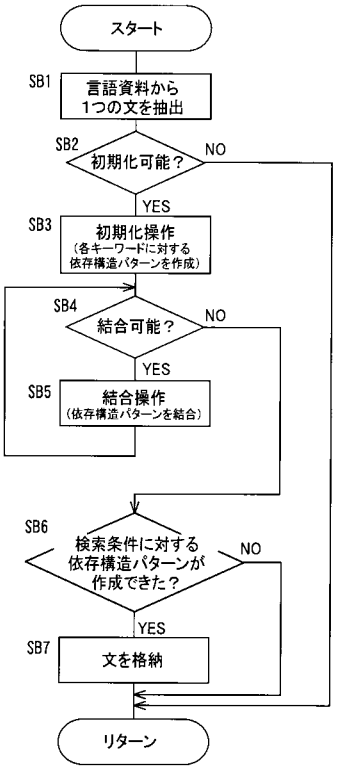
【図５】


(a) h<h'のとき
d*=(h'*, d, ε)
(b) h>h'のとき
d*=(h*, ε, d)

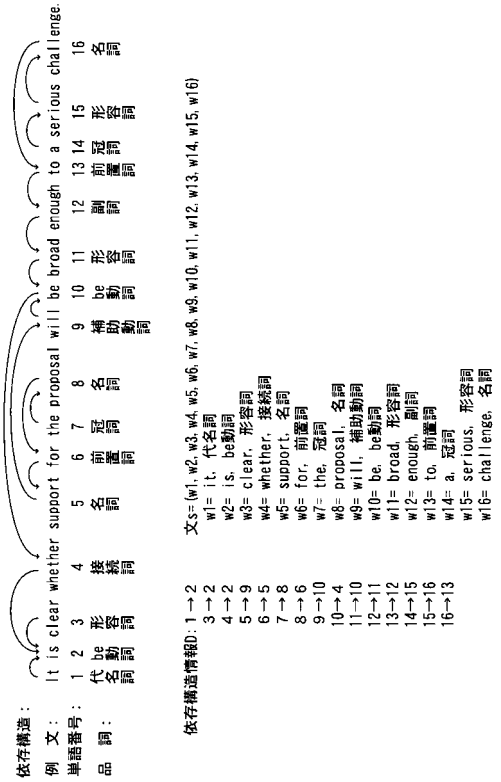
【図６】


スタート
SA1
検索条件を入力
SA2
依存構造パターン生成ルーチン
SA3
言語資料C中のすべての例文について生成を試みた？
NO
YES
SA4
検索結果を出力
エンド

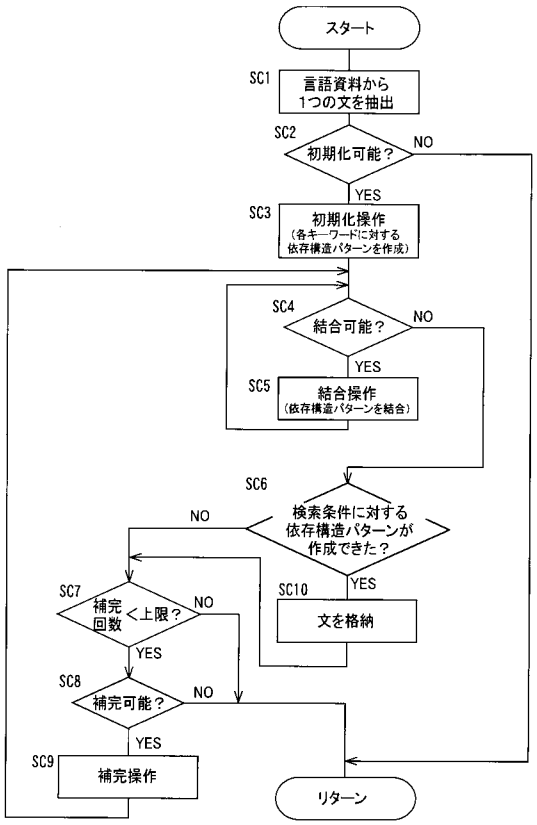
【図７】


スタート
SB1
言語資料から1つの文を抽出
SB2
初期化可能？
NO
YES
SB3
初期化操作
(各キーワードに対する依存構造パターンを作成)
SB4
結合可能？
NO
YES
SB5
結合操作
(依存構造パターンを結合)
SB6
検索条件に対する依存構造パターンが作成できた？
NO
YES
SB7
文を格納
リターン

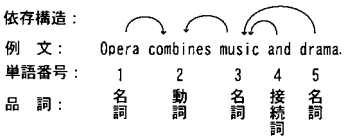
【図８】


依存構造：
例 文：It is clear whether support for the proposal will be broad enough to a serious challenge.
単語番号：1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
品 詞：代名詞 be動詞 形容詞 接続詞 名詞 前置詞 冠詞 名詞 補助動詞 be動詞 形容詞 副詞 前置詞 冠詞 形容詞 名詞
依存構造情報D：1→2
3→2
4→2
5→9
6→5
7→8
8→6
9→10
10→4
11→10
12→11
13→12
14→15
15→16
16→13
文s=(w1, w2, w3, w4, w5, w6, w7, w8, w9, w10, w11, w12, w13, w14, w15, w16)
w1= it, 代名詞
w2= is, be動詞
w3= clear, 形容詞
w4= whether, 接続詞
w5= support, 名詞
w6= for, 前置詞
w7= the, 冠詞
w8= proposal, 名詞
w9= will, 補助動詞
w10= be, be動詞
w11= broad, 形容詞
w12= enough, 副詞
w13= to, 前置詞
w14= a, 冠詞
w15= serious, 形容詞
w16= challenge, 名詞

【図９】



【図１０】



依存構造情報D：1→2
3→2
4→3
5→3

文s=(w1, w2, w3, w4, w5)
w1= opera, 名詞
w2= combine, 動詞
w3= music, 名詞
w4= and, 接続詞
w5= drama, 名詞

フロントページの続き

(56)参考文献 兵頭安昭，外1名，係り受け構造の照合に基づく用例検索システムＴＷＩＸ，電子情報通信学会論文誌，日本，社団法人電子情報通信学会，１９９４年 5月２５日，第Ｊ７７－Ｄ－２巻，第５号，第1028-1030ページ
加藤芳秀，外3名，主辞情報付き文脈自由文法に基づく漸進的な依存構造解析アルゴリズム，電子情報通信学会技術研究報告，日本，社団法人電子情報通信学会，２００４年 4月１８日，第102巻，第31号，第29-36ページ，ＣＯＭＰ２００２－１～８ コンピュテーション

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ １７／３０
Ｇ０６Ｆ １７／２８
ＪＳＴＰｌｕｓ／ＪＭＥＤＰｌｕｓ／ＪＳＴ７５８０（ＪＤｒｅａｍＩＩ）