

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4960461号
(P4960461)

(45) 発行日 平成24年6月27日(2012.6.27)

(24) 登録日 平成24年3月30日(2012.3.30)

(51) Int.Cl.

F I

G 0 6 F 17/21 (2006.01)

G 0 6 F 17/21 5 5 0 J

請求項の数 14 (全 25 頁)

(21) 出願番号	特願2009-540463 (P2009-540463)	(73) 特許権者	500046438
(86) (22) 出願日	平成19年12月5日 (2007.12.5)		マイクロソフト コーポレーション
(65) 公表番号	特表2010-511966 (P2010-511966A)		アメリカ合衆国 ワシントン州 9805
(43) 公表日	平成22年4月15日 (2010.4.15)		2-6399 レッドモンド ワン マイ
(86) 国際出願番号	PCT/US2007/086560		クロソフト ウェイ
(87) 国際公開番号	W02008/070750	(74) 代理人	100077481
(87) 国際公開日	平成20年6月12日 (2008.6.12)		弁理士 谷 義一
審査請求日	平成22年12月3日 (2010.12.3)	(74) 代理人	100088915
(31) 優先権主張番号	11/633,788		弁理士 阿部 和夫
(32) 優先日	平成18年12月5日 (2006.12.5)	(72) 発明者	チャンフェン ガオ
(33) 優先権主張国	米国 (US)		アメリカ合衆国 98052 ワシントン
早期審査対象出願			州 レッドモンド ワン マイクロソフト
			ウェイ マイクロソフト コーポレーシ
			ョン インターナショナル パテンツ内
			最終頁に続く

(54) 【発明の名称】 ウェブベースのコロケーション誤りの校正

(57) 【特許請求の範囲】

【請求項 1】

コンピュータシステムにより実行する方法であって、

テキストサンプルに品詞タグ付けるステップであって、コーパスと比較される1つまたは複数のコロケーションは、動詞 - 名詞コロケーション、前置詞 - 名詞コロケーション、形容詞 - 名詞コロケーション、および動詞 - 副詞コロケーションの少なくとも1つを含むものが前記テキストサンプルから選択されるステップと、

前記テキストサンプルからの前記1つまたは複数のコロケーションをコーパスと比較するステップであって、前記1つまたは複数のコロケーションのそれぞれを含む1つまたは複数のクエリー語を使用して、1つまたは複数のウェブ検索を実行するステップを含み、検索が実行される前記1つまたは複数のコロケーションのそれぞれに対して、前記クエリー語の1つが、前記コロケーションに一致するよう事前に選択された閾値を満たす検索結果を与えるまで、または事前に選択された閾値を満たすことなく、前記コロケーションを含む前記クエリー語の全てが使用されるまで、前記コロケーションを含む前記1つまたは複数のクエリー語のそれぞれに対して検索が実行されるステップと、

前記コロケーションが、前記コーパス内で好ましくないかどうかを特定するステップであって、前記コロケーションが前記コーパス内で見出されない場合、または、前記コロケーションが、前記コーパス内に前記コロケーションのファジーマッチが有意に存在することを示す、事前に選択された閾値一致スコアより上のスコアを取れない場合、前記コロケーションは好ましくないとして特定するステップと、

10

20

前記コロケーションが好ましくないかどうかの表示を、出力装置を介して提供するステップと、

1つの前記好ましくないコロケーション内の1つの単語を置換するワイルドカードを用いて、1つまたは複数のクエリー語を作成するステップと、

前記クエリー語に対して単語コロケーションリファレンスを検索するステップと、

前記ワイルドカードを置換する候補の単語の割合が比較的高い検索結果を特定するステップと、

前記候補の単語を有する前記検索結果を、前記出力装置を介して潜在的に適切な単語コロケーションとして与えるステップと

を含むことを特徴とする方法。

10

【請求項2】

前記コーパスが、ウェブ上のコンテンツを含むことを特徴とする請求項1に記載の方法。

【請求項3】

前記コーパスが、ネットワーク上でアクセス可能なコンテンツを含むことを特徴とする請求項1に記載の方法。

【請求項4】

検索が実行されるコロケーションのそれぞれに対して、前記コロケーションを含むクエリー語が、前記コーパス内に前記クエリー語のファジーマッチが有意に存在することを示す、事前に選択された閾値一致スコアより上のスコアを取れない場合、前記コロケーションは好ましくないことを特徴とする請求項1に記載の方法。

20

【請求項5】

少なくとも1つの前記クエリー語が、前記コロケーションを含む文を含むことを特徴とする請求項1に記載の方法。

【請求項6】

少なくとも1つの前記クエリー語が、前記コロケーションを含む要約された文を含み、前記要約された文が、前記コロケーションを含む文から付属部分を取り除くことによって形成されることを特徴とする請求項1に記載の方法。

【請求項7】

少なくとも1つのクエリー語が、前記コロケーションを含むチャンクの対を含むことを特徴とする請求項1に記載の方法。

30

【請求項8】

少なくとも1つのクエリー語が、前記コロケーションを含む単語の対を含むことを特徴とする請求項1に記載の方法。

【請求項9】

前記テキストサンプルは、対象の言語であり、前記方法は、前記コーパスをスクリーニングして前記コロケーションを前記対象の言語で表示されるコンテンツのみと比較するステップをさらに含むことを特徴とする請求項1に記載の方法。

【請求項10】

前記クエリー語が、1つまたは複数の文テンプレート、チャンク・テンプレート、および単語テンプレートを含み、候補の単語を含む検索結果の割合に、文テンプレートに対しては最も高く、チャンク・テンプレートに対してはそれより低く、および単語テンプレートに対しては最も低い、クエリー・テンプレートの重み付けを乗ずることにより、前記ワイルドカードを置換する候補の単語の割合が比較的高い検索の結果を評価することを特徴とする請求項1に記載の方法。

40

【請求項11】

前記クエリー語が、1つまたは複数の動詞 - 名詞コロケーション、前置詞 - 名詞コロケーション、形容詞 - 名詞コロケーション、および動詞 - 副詞コロケーションを含み、前記ワイルドカードが、動詞 - 名詞コロケーション中の動詞、動詞 - 名詞コロケーション中の名詞、前置詞 - 名詞コロケーション中の前置詞、形容詞 - 名詞コロケーション中の形容詞

50

、または動詞 - 副詞コロケーション中の副詞として選択されることを特徴とする請求項 1 に記載の方法。

【請求項 1 2】

ユーザが、潜在的に適切な単語コロケーションの 1 つを選択して、これに対応する好ましくない単語コロケーションを置換することを可能にするステップをさらに含むことを特徴とする請求項 1 に記載の方法。

【請求項 1 3】

コンピュータシステムによって実行可能な命令を記憶したコンピュータ読み取り可能記憶媒体であって、前記命令が、請求項 1 乃至 1 2 のいずれかに記載の方法を前記コンピュータシステムに実行させることを特徴とするコンピュータ読み取り可能記憶媒体。

10

【請求項 1 4】

コンピュータシステムであって、前記コンピュータシステムに記憶されたコンピュータ実行可能命令が前記コンピュータシステムの処理装置によって実行されると、請求項 1 乃至 1 2 のいずれかに記載の方法を実行するように構成されたことを特徴とするコンピュータシステム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、ウェブベースのコロケーション誤りを校正する方法およびシステムに関する。

20

【背景技術】

【0002】

2 つの異なる言語間の翻訳は、個々の単語の意味が使用の文脈によって制約されるという難しい事態をしばしば伴う。人が自分の母国語ではない第二言語を学習するとき、正しく翻訳したように思われるのに、ネイティブスピーカーの間の典型的な使用において、特定のグループまたはコロケーションでは実際は使用されない単語が選択されることが一般的には多い。そのようなコロケーションは、理論的な意味では構文的にまたは文法的に正しいが、通常の使用からは外れており、ネイティブスピーカーにとっては本能的に奇妙にまたはぎこちなく思われるであろう。ある言語における単語の組合せは、しばしば単一の語彙項目を形成する設定パターンに一致し、同様の意味を持ち、同様の文法的関係で並べられる単語のグループでさえ、その言語のネイティブの使用範囲内のどの語彙項目にも入らないことがあり得る。これはその語彙項目が、代替の使用からその言語を制約するものであるとネイティブスピーカーが明確に事前に考えるかどうかに関わらない。

30

【0003】

例えば、第二言語として英語を学習する者は、「strong」および「powerful」を自分の母国語では 1 単語の 2 つの代替の翻訳とみなし、「to make」および「to do」を異なる単語の 2 つの代替の翻訳であるとみなすかもしれない。英語の使用にそれほど経験が無い学習者は、「I had a cup of strong tea」よりむしろ「I had a cup of powerful tea」と書くことがあり、また、「I made a plan」の代わりに「I did a plan」と書くことがある。コロケーションの同様の誤りは、自分がネイティブスピーカーでない任意の他の言語を学習中のどの言語のネイティブスピーカーも示す特徴である。そのような誤った単語のコロケーションは、典型的には長期間のその言語での文脈経験および対話実践の後でのみ避けることができ、典型的には辞書または他のタイプの素早く簡単にアクセスできる言語リファレンスを使用して解決することはできない。

40

【0004】

上記の説明は、一般的な背景情報のために提供するものであり、請求される主題の範囲を決定する助けとして使用されることは意図していない。

【発明の概要】

【発明が解決しようとする課題】

50

【 0 0 0 5 】

本発明は、ウェブベースのコロケーション誤りを校正する方法およびシステムに関する。

【課題を解決するための手段】

【 0 0 0 6 】

方法、コンピュータシステムおよびコンピュータシステム用の実行可能命令を符号化するソフトウェアを含む様々な実施形態において、コロケーション誤りを、ウェブ等のローカルなコーパスおよびネットワークベースのコーパスを使用して自動的に校正することが可能である。例えば、例示の一方法によると、テキストサンプルからの1つまたは複数のコロケーションをウェブのコンテンツ等のコーパスと比較する。そのコロケーションは、それらがコーパス内で好ましくないかどうか特定される。また、出力装置を介して、コロケーションがコーパス内で好ましくないかどうかについての表示が提供される。追加のステップとしては、次いで、潜在的に適切な単語コロケーションを検索すること、および、ユーザ出力を介してそれらを提供すること等がある。

10

【 0 0 0 7 】

上述した説明は、以下の「発明を実施するための形態」でさらに述べる概念を選択して簡略化した形式で紹介するために提供するものである。上述した説明は、請求の主題の重要な特徴または主要な特徴を確認することを意図していない。また、請求の主題の範囲を決定する補助として使用されることを意図してもいない。本請求の主題は、上述した任意のまたは全ての不都合を解決する実施に限定されない。

20

【図面の簡単な説明】

【 0 0 0 8 】

【図1】例示の実施形態による方法を可能にするコンピュータシステムのためのユーザインターフェースを示す図である。

【図2】例示の実施形態による方法のフローチャートを示す図である。

【図3】例示の実施形態による方法のフローチャートを示す図である。

【図4】例示の実施形態による方法を可能にするコンピュータシステムのためのユーザインターフェースを示す図である。

【図5】幾つかの実施形態が実施され得るコンピュータ環境のブロック図である。

【図6】幾つかの実施形態が実施され得る別のコンピュータ環境のブロック図である。

30

【発明を実施するための形態】

【 0 0 0 9 】

図1は、例示の実施形態による、コロケーション誤りの校正の方法を可能にするコンピュータシステムのためのコロケーション誤りの校正インターフェース10を示す。コロケーション誤りの校正インターフェース10は、例えば、コンピュータシステム、およびコンピュータシステムにより実行可能に構成される実行可能命令を含む実施形態、ならびに方法の実施形態を可能にするコンテキストを例示する。以下の説明は、種々の例示の実施形態のさらなる詳細を提供する。ある例示の配置および付された名前、ならびにコンピュータ実施方法により比較されるテキストのサンプルが、この図面および後に続く図面に含まれるが、それらは、本明細書および請求の範囲に与えられる、様々な広い意味を例示し表示することを意図するものである。

40

【 0 0 1 0 】

図1の例示の実施形態において、コロケーション誤りの校正インターフェース10は、ソフトウェアアプリケーションに関連するグラフィックディスプレイパネル、すなわちコンピュータのモニタ等の出力装置上で、グラフィカルユーザインターフェースにおいて一度に開くことが出来る多くのグラフィックディスプレイパネルの1つを構成する。第二言語としての英語(ESL)を学習する学習者が入力した、校正用テキストサンプルを対象としている例示の一実施形態を参照すると、コロケーション誤りの校正インターフェース10は、タイトルバー12内で「ESL校正」と名前が付されている。ESLは、コロケーション誤り校正の例示の一実施形態についての典型的なアプリケーションを提供するが、

50

それは、誤りのあるコロケーションが、第二言語としての英語を学習する中級の学習者による誤りの多くを占めることが分かっているためである。

【 0 0 1 1 】

E S L 校正アプリケーションの例示の実施形態は、以下の説明において繰り返し参照されるが、単なる例であり、他の実施形態の広い範囲に適用できる広義の原理を示すことが理解されるべきである。例えば、以下に、図 1 に示すコロケーション誤りの校正インターフェース 1 0 を伴わずに動作する他の実施形態が記載され、その一方、さらに他の実施形態は、第二言語として任意の他の言語を学習する学習者を対象とし、また、他の実施形態は、例えば、自分の母国語を習得するために学習する子供を対象をより絞ることができる。

10

【 0 0 1 2 】

コロケーション誤りの校正インターフェース 1 0 は、その上部に、2 つの異なるタスクモードに対応する 2 つのタブを含む。2 つのタブとは、「E S L 文のロード」と名前が付されたタブ 1 1 (この場合も、特定の E S L 対象実施形態を対象とする)、および「クエリーの実行」と名前が付されたタブ 1 3 である。図 1 は、「E S L 文のロード」のタブ 1 1 が選択されたユーザインターフェース 1 0 を、ディスプレイ上の関連するボタン、見出し、テキストボックス、および他のユーザインターフェース要素 (または「ウィジェット」と共に示す。図 4 は、「クエリーの実行」のタブ 1 3 が選択されたコロケーション誤りの校正インターフェース 1 0 と、ディスプレイ上のそのタスクに関連するユーザインターフェース要素とを示すが、その説明はさらに以下で行う。

20

【 0 0 1 3 】

「E S L 文のロード」のタブ 1 1 の下のユーザインターフェースパネルは、様々な対話型の入力および出力ユーザインターフェース要素を含み、順次これらについて以下で説明する。これらは、通常、「文の分析」というタイトル見出しを有する上部分と、「コロケーションの分析」というタイトル見出しを有する下部分に分割される。

【 0 0 1 4 】

「文の分析」部分の上部に沿って、インターフェース要素の行があり、「構文解析済みファイルからのロード」ボタン 1 5、「生ファイルからのロード」ボタン 1 7、図では「V - N コロケーション」(以下で説明する)と現在書かれているコンボ・ボックス 1 9 の前に「コロケーションタイプ」と書かれた見出し、および、「コロケーションの抽出」ボタン 2 1 がある。それらのウィジェットの下には、左右および上下のスクロールバー 2 2、2 4、ならびに「元の文」と書かれたタイトルバー 2 0 を有する、テキストボックス 2 3 がある。その下には、「構文解析済みの文」という見出しのテキストボックス 2 5 がある。通常、これらのインターフェース要素は、誤り校正のために、テキストサンプルから文章をロードして、それらをネットワーク上で入手可能なコンテンツのコロケーションとの比較のために準備する、ユーザ機能を可能にする。

30

【 0 0 1 5 】

テキストボックス 2 3 は、E S L 学習者により書かれた文書等のテキストサンプルからの多数のサンプル文を含む。テキストボックス 2 3 上の上下スクロールバー 2 4 のトラフ (trough) 内にある圧縮されたサイズのスクロールサム 2 6 が示すように、現在表示される文章は、テキストボックスに現在ロードされている文章の全体のリストのごく一部であり、ユーザはこのリストに、スクロールサム 2 6 を移動させることによってアクセスできる。テキストボックス 2 3 に表示される文章は、第二言語として英語を学習する中級の学習者が犯しやすい典型的な間違いを表す。それらは、「I did a plan」、「I have recognized this person for years」、および「I had a cup of powerful tea」などである。

40

【 0 0 1 6 】

これらの文章は、「構文解析済みファイルからのロード」ボタン 1 5 または「生ファイルからのロード」ボタン 1 7 のどちらかを選択することにより、ユーザが処理を開始した後、別のアプリケーションで開かれた文書から手動でロードされるか、または、別の

50

文書または多数の文書から、E S Lタイプの誤りを有するものとして自動的に選択され、インポートされる。別の実施形態において、ワードプロセッシングアプリケーション等の異なるアプリケーションは、E S L校正トリガ有するが、これはデフォルトでアクティブであり得、E S Lタイプの誤りがワードプロセッシングプログラム内に入力されるテキスト内で検出された時に、自動的にコロケーション誤りの校正インターフェース10の利用を推奨するパネルを立ち上げることができる。

【0017】

ボタン15および17により示唆されるように、テキストサンプルは、アプリケーションにより取り上げられるときには、すでに構文解析はなされているかもしれないし、なされていないかもしれない。テキストサンプルがまだ構文解析されていない場合は、例えば、追加のステップが行われる前のローディングの後に構文解析され得る。構文解析された文が、テキストボックス25に示される。特に、これは、テキストボックス23に示される第二の文の構文解析済みのバージョンであり、周りを局所的にハイライトされ、ユーザにより、例えば、マウスで左クリックすることより、もしくはライトペンでタップすることにより、または何らかの他のユーザ入力機構により、選択されていることを示す。

【0018】

構文解析は、自然言語処理および自然言語理解における開発の盛んな分野である。構文解析は、例えば、品詞タグ付け、チャンキング、およびセマンティックラベリング等のタスクを含むことが出来る。図1に示される実施形態において、テキストボックス25内の文は、品詞タグ付け、およびチャンキングがなされている。他の実施形態は、何らかの非西欧の言語に対してより絞った対象とするものを含み、単語分割等の構文解析タスクの他の組合せを含むことが出来る。

【0019】

テキストボックス25内の構文解析済みの文において、元の文の各単語には、そのすぐ後に斜線（すなわち、右上がりスラッシュ）が続き、その単語に割り当てられる2または3文字の品詞タグからその単語を分離する。品詞タグ対して幾つかの例示のオプションを以下に挙げる。文の終わりのピリオドも、自身の斜線を伴い別のピリオドからそれを分離し、品詞タグと同等のタグの働きをし、句読点の機能を示す。1つまたは2つの（若しくは、他の例ではそれ以上の）グループにおける文中の単語も、各括弧で囲まれた組の内部の先頭に2文字のチャンクタイプのラベルを有する、チャンクの境界に対応する括弧で囲まれる。この文において表示された品詞タグは、人称代名詞を表す「P R P」、三人称単数形以外の現在形の動詞を表す「V B P」、過去分詞形の動詞を表す「V B N」、限定詞を表す「D T」、単数形普通名詞または代替可能普通名詞を表す「N N」、前置詞または従属禁止命令（subordinating injunction）を表す「I N」、および複数形の普通名詞を表す「N N S」を含む。例示の実施形態による品詞タグの例示のリストは、以下のように提供されるが、一方で、異なる定義および多かれ少なかれ特異性を伴う品詞タグの他のスキームが、他の実施形態において使用され得る。幾つかのタグを、現在の例示の実施形態について、英語において例示する。

【0020】

表1：品詞タグの例示のリスト

1. C C - 等位接続詞（例えば、「and」、「but」、「nor」、「or」、「yet」、「plus」、「minus」、「times」、「over」（分割にも同様に使用される））
2. C D - 基数
3. D T - 限定詞（冠詞および不定限定詞、例えば、「a」、「an」、「every」、「no」（冠詞として）、「the」、および、例えば、「another」、「any」、「some」、「each」、「either」、「neither」、「that」、「these」、「this」、「those」、ある使用での「all」、「they」を含む）
4. E X - 存在を表す語「there」

5 . F W	-	外国語	
6 . I N	-	前置詞または従属禁止命令	
7 . J J	-	形容詞	
8 . J J R	-	形容詞比較級	
9 . J J S	-	形容詞最上級	
10 . L S	-	リスト項目マーカ	
11 . M D	-	法動詞 (例えば、「can」、「could」、「may」、「might」、「must」、「shall」、「should」、「will」、「would」)	
12 . N N	-	単数形普通名詞、または代替可能普通名詞	10
13 . N N S	-	複数形普通名詞	
14 . N N P	-	単数形固有名詞	
15 . N N P S	-	複数形固有名詞	
16 . P D T	-	前限定詞	
17 . P O S	-	所有格語尾	
18 . P R P	-	人称代名詞	
19 . P R P \$	-	所有代名詞 (例えば、「's」、「s'」、「'」)	
20 . R B	-	副詞	
21 . R B R	-	副詞比較級	
22 . R B S	-	副詞最上級	20
23 . R P	-	不変化詞	
24 . S Y M	-	記号	
25 . T O	-	「To」	
26 . U H	-	感嘆詞	
27 . V B	-	動詞、原形	
28 . V B D	-	動詞、過去形	
29 . V B G	-	動詞、動名詞または現在分詞	
30 . V B N	-	動詞、過去分詞	
31 . V B P	-	動詞、三人称単数形以外の現在形	
32 . V B Z	-	動詞、三人称単数形現在形	30
33 . W D T	-	W h 限定詞	
34 . W P	-	W h 代名詞	
35 . W P \$	-	W h 代名詞所有格 (「whose」)	
36 . W R B	-	W h 副詞	

【0021】

テキストボックス25内の構文解析済みの文に表示されるような構文解析されたテキストサンプルも、この図示においてはチャンキングされている。ここでは、括弧の対が、定義された各チャンクを囲み、チャンクタイプが各括弧の組の中身の先頭にラベルで示されている。チャンキングは、表面的で局所的な情報に基づき実行されることが可能な、比較的扱いやすい構文解析タスクである。チャンキングによって、文章を重複しない区分に分割して、各チャンクが1つの重要な主要語をこれに関連する単語と共に含むようにする。チャンキングは、従って、文をフレーズに分割することと考えられるが、具体的には、本例示の実施形態においては、フレーズの重複を防ぐために定義可能な最小のフレーズ単位に文を分割すると考えられる。チャンキングは、テキストの統語構造、並びにフレーズ間の関係または依存を識別可能になるようにする。例えば、1つの名詞句は、動詞句の主語であり得、また、第2の名詞句は動詞句の目的語であり得る。

【0022】

図1における例示のチャンクは、この良い例を提供する。第1のチャンクは、名詞句を表すNPがラベル付けされ、唯一の重要な主要語、文の主語である「I」に基づいている。第2のチャンクは、動詞句を表すVPがラベル付けされ、「have」および「rec

40

50

o g n i z e d」を含み、この2つの個々の動詞は一緒になって、動詞「t o r e c o g n i z e」の過去分詞形式の統合概念を形成する。第3のチャンクは、別のNPすなわち名詞句であり、主要語「p e r s o n」を、関連しサポートする単語と共に含み、文の目的語として単一概念「t h i s p e r s o n」を形成する。前置詞句を表すPPがラベル付けされた第4のチャンク、および別の名詞句を表すNPがラベル付けされた第5のチャンクは、それぞれ単一の単語を含む。ピリオドは、ピリオドとしてタグが付けられるが、チャンクからは除外される。

【0023】

コロケーション誤りの校正インターフェース10に関連する、またはコロケーション誤りの校正インターフェース10によって利用される構文解析システムは、自然言語処理の当業者にはよく知られるような方法で、正確にまた確実に品詞タグ付けおよびチャンキングを自動的に実行するように作りこむことが可能である。

10

【0024】

「ESL文のロード」タブ11の下ユーザインターフェースパネルの下部分は、タイトル見出し「コロケーションの分析」が付けられ、1組のテキストボックス27、29、31および33を含み、それぞれ、タイトルバーおよび少なくとも1つのスクロールバーと共に示される。テキストボックス27のタイトルバーには、「ESLコロケーション」と書かれ、テキストボックスは、テキストボックス25の文からの2つの連語配列されるチャンクである、動詞句「h a v e r e c o g n i z e d」、および文の目的語である名詞句「t h i s p e r s o n」を含むように示される。品詞タグならびにチャンキングラベルおよび括弧と共に示されるように、連語配列されるフレーズは「[V P h a v e / V B P r e c o g n i z e d / V B N] [N P t h i s / D T p e r s o n / N N]」と書かれる。コロケーションタイプ・コンボ・ボックス19が、動詞句およびそれに続く名詞句が互いに隣り合うように配置されるコロケーションを表示する選択肢である「V - Nコロケーション」に設定されているため、このコロケーションが、テキストボックス27内に与えられる。この設定に対応するコロケーションは、テキストボックス25の文から抽出されている。

20

【0025】

テキストボックス29は、タイトルバーに示されるように、コロケーション候補を含む。コロケーション候補は、テキストボックス27内のコロケーションと同じ名詞句「t h i s p e r s o n」を有するが、名詞句に先行する様々な異なる過去分詞形の動詞を有する。これらの候補の動詞は、過去分詞の形式で動詞句内の「h a v e」に続けて、動詞の主要語として、「r e c o g n i z e d」を置換する候補である。これには、「k n o w n」、「a d m i t t e d」、「t h o u g h t」等が含まれる。元のコロケーションをコーパスのコンテンツと比較することにより、元のコロケーションがコーパスのコンテンツには少ないかまたは存在せず、従って好ましくなく、誤りであつたらしいということが明らかになったため、これらの候補は元のコロケーションを含まない。

30

【0026】

コロケーション誤りの校正インターフェース10、または、これに関連するかもしくはこれと共に使用されるアプリケーションは、従って、元のコロケーションがコーパスのコンテンツには好ましくなく、誤りであつたらしいということを表示することができる。この表示を、例えば、図1の例示の実施形態のように、ユーザがテキストボックス33で見られるようにすることができる。テキストボックス33は、そのタイトルバーが示唆するように、コロケーションステータスを示す。これは、このコロケーションがコーパスのコンテンツには好ましくないか、または誤りであつたらしいということを表示し、好ましくないと判明したコロケーション内の置換される候補である動詞の主要語の下に波線32をつける。ユーザは、例えば、所望の置換コロケーションの上でダブルクリックすることなどにより、テキストボックス29内のコロケーション候補の1つを選択し、好ましくないと表示されたコロケーションと置換することができる。この表示は、また、例えば、別個のワードプロセッシングプログラム、ネットワークナビゲーションプログラム、または他

40

50

のアプリケーション内で、不適切なコロケーションにマークを付けること等により提供することができる。このようなマーク付けは、好ましくない、または誤ったコロケーションを、明るい色の波線などでハイライトしたり、または下線をつけたりすることにより、ユーザの注意を引くことができる。

【 0 0 2 7 】

テキストボックス 29 内のコロケーション候補は、テキストボックス 27 内のコロケーションを大きなコーパス内の利用可能なコンテンツと比較した結果として提供される。これは、ローカルなコーパス、および/または、分散リソース上でホストされるコンテンツ、およびこの例示の実施形態のワールドワイドウェブのようなネットワーク上でアクセス可能なコンテンツを含む、ネットワークベースのコーパスを含むことができる。他の実施形態においては、コーパスは、他の利用可能なネットワークのコンテンツ、例えば、イントラネット、広域ネットワーク、ローカルエリアネットワーク、または何らかの他のタイプのネットワーク等のコンテンツを含むことができる。ウェブをコーパスとして使用するウェブ検索由来のコンテンツは、例えばウォールストリートジャーナルコーパス等の、従来のコーパスの検索と一緒に使用することができる。従来のコーパス内におけるコロケーションの存在は、ウェブコンテンツにおけるコロケーションの存在と比べて特に重きを置かれるが、これはコーパスが、コーパスに比べてより広い様々なソース、およびよりカジュアルな書き物の内包を含み得るウェブコンテンツと比較すると、比較的に一貫して適切な使用を保証するためである。しかし、ウェブコンテンツは、任意の利用可能な従来の言語使用のコーパスに比べて、その膨大な量において十分な利点を提供する。テキストサンプル内の任意の所与のコロケーションに対応する適切なコロケーションは、従来のコーパスにはあまり存在しない。一方、少なくとも英語では、適切な使用で存在し得る任意の単語のコロケーションであればほとんどウェブ上で、または、以下でさらに説明するように、コロケーションに基づく異なるクエリー語の柔軟な検索により明らかになり得る、少なくとも構造的に等価のコロケーションを見出すことが可能であることが分かっている。また、少なくとも、例えば、第二言語としての英語を学習する学習者からの典型的なテキストサンプルに見られる、ほとんどの誤ったコロケーションに対する、適切な置換コロケーションは、ウェブで見出すことが可能であることが分かっている。ウェブ検索において特定の文またはコロケーションが見つからないということは、その文またはコロケーションが誤りであるということのより高い確信をもたらす。類似の代替コロケーションが対象のコロケーションより非常に高い頻度で現れるかどうかを確認して検索結果を評価しても良く、結果として誤りであるにもかかわらずコロケーションがウェブ上に存在する可能性のあることが、そのようにして明らかになるだろう。

【 0 0 2 8 】

本明細書書の作成時においては、ウェブで利用可能なほとんどのテキストのコンテンツは英語である。現在の人の言語の多くは、ウェブ上ではほとんど利用可能でなく、幾つかの共通で話される言語でさえ、ウェブコンテンツ内に現れることは比較的少ない。例えば、ヒンディー語は、現在世界で 3 番目に広く話されている言語であるが、そのウェブコンテンツは、話す人がその千分の一に満たないアイスランド語のものより少ないことが、少なくとも 1 つの調査で分かった。コロケーション誤りの校正の有効性は、校正されるべきテキストサンプル内のコロケーションに対する比較の基盤として利用可能な、コーパスのサンプルのサイズに部分的に依存する。従って、英語のテキストサンプルの校正には必要とされないであろう異なる言語において、ウェブベースのコロケーション誤りの校正には、特別な方法が使用されるであろう。

【 0 0 2 9 】

これらには、例えば、特に、テキストサンプルの言語に一致する対象言語のコンテンツに対する、ターゲティングもしくはスクリーニング、または、ウェブもしくはネットワーク上でアクセスされる他のコンテンツに加えて、比較の基盤として使用されるコーパスに対して、特別のネットワーク、コーパス、もしくはデジタル化されたライブラリを使用することが含まれる。これらの方法は、例えば、ユニコードに符号化されていないコンテン

ツにインデックスを付け、かつ検索することが可能な検索ユーティリティを使用すること、または、特別のキャラクタ符号化スキームもしくは変換フォーマットを使用して非標準符号化フォーマットのコンテンツを読み、インデックを付け、検索できるようにすることを含む。ローマ字（アルファベット）以外の筆記体系を使用する言語の大量のコンテンツは、幾つかの言語のウェブ上の大多数のコンテンツでさえも、非標準符号化フォーマットで存在するため、これらの非標準符号化フォーマットにインデックスをつけて検索することが可能な検索は、他のものよりも非常に多くのコンテンツを生じさせる傾向がある。これらのような方法は、種々の実施形態に従って、可能な最大範囲の言語に適用されるように、ウェブベースのコロケーション校正システムの性能に寄与する。

【0030】

ウェブコンテンツが検索されることを強制し、またはそれにバイアスをかける追加の方策は、URLのカテゴリまたは分類についての選好を適用することを含み得る。幾つかのURLは、具体的には、そのコンテンツが大量のライブラリまたは信頼できる適切な言語の使用のコーパスを含むことが知られている場合、選好する検索のため、および/または、このコンテンツ内で見出されるコロケーションにより重きを置くため、特別に列記され得る。これは、また、検索に選好的にバイアスをかけること、またはトップレベルドメイン（TLD）に基づき検索結果に重きを置くことを含む。例えば、追加の検索順序、または結果重み付けが、「com」または「org」ではなく、「edu」または「gov」のTLDを有するURLに対して適用され得る。英語以外の言語に対するコロケーション校正に関する検索では、対象の言語のウェブコンテンツについてのスクリーニング検索は、対象の言語が公用語であるかまたは実質的に使用されている国に割り当てられる国コードTLDに基づく検索または重みに、バイアスをかけることまたはそれらを制限することを含み得る。例えば、フランス語のテキストサンプルを校正することを対象とするコロケーション校正アプリケーションは、「fr」、「be」、「ch」もしくは「ca」（それぞれフランス、ベルギー、スイスおよびカナダの国コードTLD）といったTLD、または、他の公式もしくは実質的にフランス語圏の国に属する、他の国コードTLDを伴うウェブサイトで見出されるウェブコンテンツの検索に対して、バイアスをかけまたは制限することが可能である。

【0031】

テキストボックス31は、コロケーション内で「this person」に先行する動詞句の主要語を形成するコーパスのコンテンツ内において見出される、全ての利用可能な動詞の長いリストの一部を示す。スクロールバーの何分の一かのサイズである、テキストボックス31の右側の上下スクロールバー内のスクロールサムにより示されるように、リストは、現在図1に示すテキストボックス31内に見える、利用可能な動詞の一部分より極めて長い。テキストボックス31にリストアップされる利用可能なコロケーションのほとんどが、テキストボックス29でのリストアップのように、コロケーション候補として選択されなかったが、それはこれらのコロケーションが、元のコロケーションと一致すること、およびそれを修正された形式で元の意味を与えるコロケーションと置換することの、考え得る正当性の基準の組に基づく、ファジーマッチのスコアリングスキームにおいてそれほど高いスコアを取れなかったからである。

【0032】

テキストボックス27内のコロケーションを比較し、コーパスのコンテンツ内でそのコロケーションが好ましくないかどうかを特定し、およびそのコロケーションに対して推奨する置換を提供する例示の方法が、残りの図面を参照してさらに詳細に説明される。

【0033】

図2は、図1に示すコロケーション誤りの校正インターフェース10の態様に類似する特定の態様である、例示の実施形態に従って、コロケーション校正方法50のフローチャートを示す。方法50の中心的なステップが、図2のフローチャート内に実線で輪郭を描いて示される。これらは、ウェブ検索のクエリー語として、コロケーションを含むコンテンツを提示することなどによって、テキストサンプルからの1つまたは複数のコロケーシ

10

20

30

40

50

ョンをコーパスと比較するステップ5 1、コロケーションがコーパスにおいて好ましくないかどうかを特定するステップ5 3、および、コロケーションが好ましくないかどうかの表示を、出力装置を介して提供するステップ5 5を含む。再帰的なフローチャートの矢印で示すように、ステップ5 1および5 3は繰り返し連続して実行され、テキストサンプルからのコロケーションとコーパスとの異なる種類の比較を実行することができ、その異なる種類の比較により示されるように、コロケーションがコーパスにおいて好ましくないかどうかを特定する。

【0034】

ステップ5 1、5 3、および5 5は、点線で輪郭を描いて表示されるように、追加のステップを伴い、追加のステップが先行するか、または追加のステップが後に続くことができる。これらは、例えば、ステップ6 1に示されるように、ユーザ入力もしくは選択または他の形式のテキストサンプルの指示を受け取ることなどにより、テキスト内の単語コロケーションの特定または指示を受け取ることを含む。それはまた、デフォルトの特徴を持ち、人によるテキスト入力を伴うワードプロセッシングアプリケーションまたは何らかの他のソフトウェアアプリケーションにおいて、誤ったコロケーションまたは非ネイティブの使用の他の指示に対して、入力が監視され、かつスクリーニングされ、また、ステップ6 3に示されるように、そのような指示が見つければ追加のステップが始動される。テキストサンプルは、また、ステップ6 5に示されるように、また図1に関連して例示的に説明されるように、構文解析されるが、例えば、本方法の他の実施形態は、既に構文解析済みのテキストサンプルに適用してもよい。

【0035】

ステップ6 7に示されるように、特定のタイプのコロケーションは、特に対象とされてコーパスと比較される。このステップは、コーパスと比較すること、および追加のコロケーション誤りの校正ステップを実行することのために、動詞 - 名詞コロケーション、前置詞 - 名詞コロケーション、形容詞 - 名詞コロケーション、または、動詞 - 副詞コロケーション等のコロケーションを選択することを含む。コロケーションのこれらの個々のカテゴリの任意の1つは、幾つかの実施形態において限定的に使用され得るが、任意のまたは全てのこれらのコロケーションタイプ、および/または追加のコロケーションタイプは、他の実施形態で使用され得る。ステップ6 7においてリストアップされる4つのコロケーションタイプを使用することが、典型的な非ネイティブの言語ユーザが犯す全てのコロケーションの誤りの相当な部分を占めるということが分かっている。これらの特定のコロケーションタイプを選択することは、例えば、ステップ6 5に示されるように、または図1に関連して説明されるように、事前の構文解析ステップによって、テキストサンプル内の品詞を最初に特定することを含む。

【0036】

コロケーションが、コーパスにおいて好ましくないかどうかを特定するステップ5 3は、異なるフォーマットのコロケーションに基づく様々な個々のクエリー語に対して、繰り返し実行され使用された全てのクエリー語に基づいて、コーパスと比較されるコロケーションがそのコーパスのコンテンツにおいて好ましくないかどうかの最終的な判定の形式とすることができる。

【0037】

例示の一実施形態は、図3のフローチャートにおける類似のステップ3 5 1および3 5 3にさらに示される。この実施形態において、図2のステップ5 1に示されるようにテキストサンプルからのコロケーションをコーパスと比較することは、ファジーマッチの方策を含めることができる。この比較は、図3においてステップ3 6 1、3 6 3、3 6 5、および3 6 7にそれぞれ示されるように、コロケーションを含む完全な文、要約された文(reduced sentence)、一対のチャンク、または一対の単語を含むクエリー語に対してコーパスを検索することを含むものである。要約された文とは、コロケーションを含む文から付属部分を取り除くことによって形成されるものであるが、ここで付属部分とは、例示の実施形態において、文中のコロケーションの保存には必要ではないと事前に定義され、文

10

20

30

40

50

から選択された単語である。例えば、誤ったコロケーションの文「I have recognized this person for years」に基づく要約された文のクエリー・テンプレートは、「have recognized this person」であり得る。チャンクの対および単語の対のクエリー語は、特定の単語またはコロケーションを含むチャンクが直接隣り合っているか、または、例えばそれらを引き離す1つまたは2つ程度の他の単語を伴う、同一の文内で近接する事前に選択された範囲の中にあるというものである。ファジーマッチ検索における、要約された文の検索、チャンクの対の検索、および単語の対の検索は、異なっているが等価な単語の順番であるという結果も検出し、または特定の機能語および他のタイプの単語に対してプレースホルダを使用する。その結果、例えば、「recognize this person」と「recognize that person」とは、ファジーマッチでは等価とみなされる。

10

【0038】

コロケーションを含むクエリー語が、コーパス内にクエリー語のファジーマッチが有意に存在することを示す、事前に選択された閾値一致スコアより上のスコアをとれない場合、コロケーションは、好ましくないと評価され得る。図2のステップ53に示されるように、コロケーションがコーパスにおいて好ましくないかどうかを特定することは、クエリー語に対して異なるタイプの検索を実行することを含むことができ、これは、図3のステップ371およびステップ373のそれぞれに示されるように、文または要約された文の一致を検索すること、または、図3のステップ375およびステップ377のそれぞれに示されるように、チャンクの対もしくは単語の対に対して見出す一致のための閾値のスコアを評価することを含む。文および要約された文が、非常に特有のものであるため、それらの内の1つに対する1つの一致または少数の一致で、コロケーションが好ましいもの、または誤りでないものとして示されるには十分であると考えることができる。チャンクの対または単語の対の一致は、これらがそれほど不自然ではなく、より一般的であるため、その言語の専門家ではないユーザが書いたウェブコンテンツが比較的まれに残っていることにより、誤りである場合でも、少なくとも幾つかの検索結果で見付つかることは大いに予測できる。従って、本実施形態によると、チャンクの対および単語の対の一致に対してより高い閾値を設定することができる。この閾値は、指示されたチャンクまたは単語の対の、相対的な普及（prevalence）を示すのに十分であり、コーパスにおいて相対的に少量の、好ましくない単語コロケーションと予想される指示ではなく、通常の使用に対応する単語コロケーションを示すものである。

20

30

【0039】

一致スコアは、特定の累積スコアを、例えばチャンクの対または単語の対の一致の存在それぞれに割り当てる基本カウンタを含み、十分な一致が見付き最小合計スコアを達成する場合、好ましくないものよりもより適切なコロケーションを単に示す。異なるクエリーを有する結果についてのスコアは、組み合わせられ、より特定されたクエリー語にさらに重み付けし、例えば、例示として、チャンクの対の一致のカウント当りに、単語の対の一致のカウント当たり10倍の重み付けをする。要約された文または全文もまた、コロケーションステータスを判定するのに、それ独自では十分ではなく、一致スコアの評価に含まれ得る。例えば、テキストサンプルからのコロケーションの対を含む要約された文で見付けられた各一致は、チャンクの対当りの20倍の一致重み付けのスコア、および単語の対当たり200倍の重み付けのスコアが割り当てられ得る。一致または一致スコアを評価するのにどんな方法が使用されても、その処理は、ステップ355において終了し、そのステップ355はステップ381に示されるように、コロケーションが好ましくないという表示、またはステップ383に示されるように、コロケーションが好ましいという表示を提供する。

40

【0040】

異なるクエリー・テンプレートの重み付け、および普通の使用のコロケーションを、好ましくないコロケーションから区別するために使用される閾値一致スコアは、コロケーション校正ソフトウェアの開発者またはユーザオブションとしてエンドユーザにより、より

50

大きいまたはより小さい検出感度に調整または調節することができる。ユーザは、閾値を低く下げて処理を早くしたいかもしれないが、問題のあるまたは不明確な好ましくない少数のコロケーションがくぐり抜ける可能性が大きい。またはユーザは、もっと完璧なものを求め、閾値を高く設定することができる。場合によっては注目すべきコロケーションを調べることに、より多くの時間を費やし、そのコロケーションが任意の可能性のある置換語に比較的類似した使用がなされていることが分かるということにもかかわらず、より良く確実にいかなる誤ったコロケーションもテキストサンプルから取り除くであろう。

【 0 0 4 1 】

図3におけるクエリー語の順番は、より特有のものから、より一般的なもの、およびより制約されないものへと進むため、各連続するステップからの検索結果を収集することに、前のステップより時間がかかりやすい。検索結果を戻す全てのステップは、ほんの一瞬を必要とすると思われるが、システムの性能はさらにより高効率化され、一致の十分な表示が見付かると、すぐに次第に時間がかかってゆく残りのステップを経由することなく、ステップ351およびステップ353の処理を終了することにより、全体の時間を減らす。所定の文または要約された文に対して多くの一致が見付けられた場合、例えば、チャンクの対または単語の対の検索を実行するために必要な比較的長い瞬間を費やすことを、回避することが出来る。異なるクエリー語を使用する同一のコロケーションの検索は、従って、クエリー語の1つが、コロケーションに一致する事前に選択された閾値を満たす検索結果を与えるまで、または、コロケーションを含む全てのクエリー語が、事前に選択された閾値を満たすことなく使用されるまで、連続して続行される。

【 0 0 4 2 】

図2に示されるさらなるステップに戻り、例示の一実施形態によると、特定のコロケーションが好ましくないかどうかの表示が与えられた後、ステップ71、73、75および77に示されるように、コロケーション誤りの校正アプリケーションが作動して、さらに、候補の置換語を検索および提供して、誤ったコロケーションを修正する。これは、ステップ71に示されるように、1つまたは複数の好ましくないコロケーション内の1つまたは複数の単語を置換するワイルドカードを用いて、クエリー語を作成することを含み得る。ワイルドカードは、誤っている可能性があることが示される疑わしい単語の代わりに、クエリー語の中に挿入されるオープンなプレースホルダとして役立つことができ、ここで、プレースホルダは、検索されているコーパス内の、クエリー語の残りの部分の疑わしい単語の位置に挿入される任意の単語に対応する。ワイルドカードは、単語の1つを、誤りが疑われる単語の代わりとなる、アスタリスク等の記号と置き換えることによって検索に使用することができる。

【 0 0 4 3 】

単語コロケーションリファレンスは、ステップ73に示されるように、ワイルドカード要素を含むもの等の、可能性のあるコロケーション候補を提供するクエリー語に対して検索されることができる。単語コロケーションリファレンスは、さらに、ウェブまたは別のネットワークベースのもしくはローカルなコーパスであることが可能であり、また、例えばウェブの検索結果をスクリーニングして単語コロケーション辞書に含まれないコロケーションを削除することなどによって、特別の単語コロケーション辞書リファレンスを含むことができる。

【 0 0 4 4 】

次に、コロケーションリファレンスの検索は、ワイルドカードプレースホルダを表す記号の代わりに、任意の単語またはフレーズを使用するテキストの区分を有する結果を戻すことができる。例えば、図1のコロケーション候補のテキストボックス29を再度参照すると、そこにリストアップされる幾つかのコロケーション候補は、全てフレーズ「this person」に先行するワイルドカード記号を用いたワイルドカード検索から戻されたものである。検索語は、例えば「* this person」であり、ここでアスタリスク「*」はワイルドカードを形成し、検索結果は、「known this person」、「admitted this person」、「thought th

10

20

30

40

50

i s p e r s o n」などを含む。検索ツールによりワイルドカードの単語を示すのに使用される記号の選択は、任意であり、アンパサンド、クエスチョン・マーク、またはその他が適当であろう。

【0045】

ワイルドカードは、同一のコロケーションに基づく一連のクエリー語のそれぞれに使用され、より大きく制約されたものからより小さく制約されたものへ、コロケーションに基づく一連のクエリー・テンプレートを順次通って進み、元々なされていたのと同様に対象のコロケーションの正当性または好ましくない程度を評価する。例えば、誤ったコロケーションの文「I have recognized this person for years」に対して、一連のワイルドカードクエリー語が含むことができるものは、完全な文であって、ワイルドカードでコロケーション内の動詞を置き換えたもの、すなわち「I have * this person for years」、要約された文の形式に基づく1つまたは複数のクエリー語、すなわち「I have * this person」、「have * this person for years」、「I have * this person」、コロケーションを含有するチャンクの対に基づく1つまたは複数のクエリー語、すなわち「* this person for years」、「* this person」、および、直接隣り合っているもの、または、1つ程度の中間の単語によって引き離されるような近接したもの、例えば「* ~ person」、として検索されるいずれかの、単語の対としてのコロケーションに対する、1つまたは複数のクエリー語である。ここで波線「~」は、例示の一実施形態において、任意に選択されているが、検索ツールによって、近接するプレースホルダとして解釈される。このプレースホルダは、ゼロもしくは1つまたは場合によっては複数の単語をその場所において表すことが可能で、ワイルドカードおよび特別の単語を引き離すことを示している。従って、ワイルドカードをクエリー語内で使用することは、テキストサンプルからのコロケーションを、ウェブ上、または他のネットワークベースのもしくはローカルなコーパスで利用可能なコンテンツ内の、潜在的に類似のコロケーションと、様々なレベルで比較するための効果的な方法の1つの例示を提供する。

【0046】

限定される組の選好的な候補に制限される特別なワイルドカードを使用することもでき、または、結果がスクリーニングされるコロケーションリファレンスが、好ましいコロケーション候補の表示を含むことができる。例えば、特定のコロケーション候補は、好ましくはテキストサンプルの筆者の母国語からの逆翻訳の処理に基づく。ユーザが、自分の母国語を入力するという選択ができるか、または、コロケーション校正ソフトウェアが、筆者の母国語を示すような文章サンプル内の手掛かりを評価するツールを使用することができ、例えば、ソフトウェアはユーザに母国語を確認するよう指示する。他の手掛かりには、例えば、コンピュータ上にインストールされた他のソフトウェアで使用されている言語オプション、またはコンピュータの物理的な場所を示すものが考えられる。例えば、ソフトウェアが、オペレーティングシステムを中国語バージョンで作動させているコンピュータ上で、作動していることを検出した場合、または、ソフトウェアが、物理的には北京にあるものとリファレンスが示すサーバを介して、コンピュータがインターネットに接続していることを検出し、一方でテキストサンプルが英語である場合、ソフトウェアは、選好的な処理または優先処理を、ネイティブの中国語話者に典型的であるコロケーション候補に割り当てる。

【0047】

筆者の母国語が分かっている場合、コロケーション校正ソフトウェアは、誤ったコロケーションが、筆者の母国語を文章サンプルの母国語でない言語へ不適切にマッピングすることに基づくかどうかを考慮することができる。これには、例えば、母国語の文法規則を母国語でない言語に適用するといったマッピングエラー、偽同族語または間違っただグループの言語(f a l s e f r i e n d)の誤用、母国語が複数形を区別しない場合に複数形を省略すること等が含まれ得る。

10

20

30

40

50

【0048】

例えば、「I have recognized this person for years」を含むテキストサンプルの筆者が、ネイティブの中国語話者であると表示される。コロケーション校正ソフトウェアは、この表示に応答して、これを中国語に翻訳して入力し英語に戻された候補の翻訳をチェックすることができる。「To recognize」は(漢語ピンイン(hanyu pinyin)で与えると)標準中国語の単語

【0049】

【表1】

「rénshì」

【0050】

に翻訳され得るが、英語では「to recognize」または「to know」として翻訳される単語である。従って、逆翻訳は「know」が「recognize」の考え得る置換の候補であることを示し、「I have known this person for years」を与える。別の例として、ネイティブのフランス語話者であると示される筆者による英語のテキストサンプルが、「I commanded the faith and onions」という文を含み、「commanded」および「faith」の両方に好ましくない単語コロケーションのタグが付けられている。これは、フランス語で

【0051】

【表2】

「J'ai commandé le foi et des oignons」

【0052】

と翻訳され、

【0053】

【表3】

「commandé」

【0054】

が適切に「ordered」として逆翻訳され、「le foi」が「le foie」と綴りが直され「the liver」として逆翻訳し、よってソフトウェアは、検索結果により生成される任意の他のものの中のこれらのコロケーション候補を優先させ、追加の重み付けをし、置換のコロケーション候補を示唆し、文「I ordered the liver and onions」を与えることができる。

【0055】

いずれの場合も、母国語からの逆翻訳を使用して、まず候補置換語の検索を制限し、またはそれをコロケーションリファレンスとして使用して、検索によって生成される可能性のある置換候補の間で評価し、および優先付けする。

【0056】

ステップ73に示されるように、クエリー語に対して単語コロケーションリファレンスを検索した後、本例示の実施形態においては、ステップ75を実行し、ワイルドカードを置換する候補の単語の割合が比較的高い検索の結果を特定する。これにより、例示の一実施形態において、コロケーションを含む1文全体、要約された文、チャンクの対、または単語の対に対する一致を与える検索結果の間で、異なる重み付けが異なる一致のタイプに対して再度使用される。合計スコアは、各タイプのクエリー語からのスコア成分の重み付けされた和であり得、ここで各スコア成分は、例えばそのクエリー語の検索結果の数と、そのクエリー語に割り当てられる重みの積である。このような方法を使用する比較的高い合計スコアは、本例示の実施形態において、ワイルドカードを置換する候補の単語の比較的高い割合の効果的な基準として解釈され、誤ったコロケーションを修正するための考え得る可能性として役立つ。

【0057】

ワイルドカードは、特にクエリー語内の品詞またはチャンクタイプのある部分を対象と

10

20

30

40

50

する。例えば、動詞 - 名詞コロケーションに基づく 1 組のクエリー語内で、各コロケーションの動詞は、ある組のクエリー語のワイルドカードで置換され、また各コロケーションの名詞は、別の組のクエリー語のワイルドカードで置換され得る。前置詞 - 名詞コロケーションでは、特に前置詞は、ワイルドカードでの置換のために選択されるコロケーションの一部であり得るが、それは、意図される文の意味に名詞がより大きく関連し、また選択される特定の前置詞が、より一般的にその言語の中間学習者による誤ったコロケーション使用をされやすいからである。前置詞 - 名詞コロケーションに基づき、前置詞をクエリー語内のワイルドカードによる置換の対象とすることは、従って、所望の置換コロケーションを生成するために必要な時間を少なくする方策をサポートすることができる。同様に、ワイルドカードは、形容詞 - 名詞コロケーション内の形容詞および動詞 - 副詞コロケーション内の副詞として選好的に選択されるが、それは、この場合もやはり、意図される意味に、名詞および動詞がより固定されやすく、一方、形容詞および副詞が、それらが修飾する名詞および動詞よりも誤ったコロケーションに陥りやすいからである。品詞またはチャントタイプの他の特定の部分は、他の実施形態において、クエリー語内のワイルドカードによる選好的なまたは限定的な置換に対して選択される。

【 0 0 5 8 】

検索結果の中で、疑わしい単語よりも有意に大きな割合を有する可能性のある置換語は、コロケーションの明らかに好ましくない部分に対して置換を示唆するものとして示され得る。これは、適切な単語コロケーションの辞書、または上記で説明したように、筆者の母国語に基づき示唆される逆翻訳との相互参照等のコロケーションリファレンスに含めるための、コロケーションと可能性のある置換語の相互参照の後でなされる。

【 0 0 5 9 】

検索の結果は、1 つまたは幾つかの可能性のあるコロケーションの候補の単語であって、最高の一致スコアを有することが判明するような単語を有するが、図 2 のステップ 77 に示されるように、潜在的に適切な単語コロケーションとしてユーザ出力を介して与えることができる。これは、図 1 のテキストボックス 29 に表示されるコロケーション候補に類似し、例えば、ユーザがコロケーション候補の 1 つを選択し、誤りであると表示されるコロケーションを置換することが可能になる。時には、単一のコロケーションが、誤ったコロケーションに対する適切な置換としての可能性を十分に良好に示すであろう。これは、図 4 に示す場合である。

【 0 0 6 0 】

図 4 は、例示の実施形態による、コロケーション校正方法を可能にするコンピュータシステムのユーザインターフェース 410 を示す。ユーザインターフェース 410 は、その上部に図 1 のコロケーション誤りの校正インターフェース 10 と同じタブを含む。具体的には、「E S L 文のロード」と名前が付されたタブ 11 (この場合も、例示的に特定の E S L の実施形態を対象とする)、および「クエリーの実行」と名前が付されたタブ 13 である。図 1 は、「E S L 文のロード」のタブ 11 が選択されたユーザインターフェース 10 を示すが、図 4 は、タブ 13 が選択されたユーザインターフェース 410 を、ディスプレイ上の関連するボタン、見出し、テキストボックス、および他のユーザインターフェース要素と共に示す。これらは、「E S L 構文解析済みの文」と名前が付された上部分と、「クエリーのレベル」と名前が付された下部分に分割される。ユーザインターフェース 410 は、コロケーション誤りの校正の例示の実施形態をサポートし、特定のタイプのクエリーに関わり、それらのクエリーで実行された検索の結果をユーザインターフェースを介して提供することを対象とする。

【 0 0 6 1 】

上部の「E S L 構文解析済みの文」部分は、図 1 のテキストボックス 25 と類似のテキストボックス 25 を含み、テキストボックス 25 は、ここでは同じ構文解析済みの文を含む。この部分は、また、構文解析済みの文テキストボックス 25 の下にボタンの行 41 を含む。これらは、構文解析済みの文に基づき検索を実行する、様々な異なる例示のクエリー・テンプレートのためのボタンを含む。特に、本例示の実施形態においては、上記の説

明によると、「クエリーの作成」、例えば完全な文または要約された文に対して、異なる文に基づくクエリーを参照する「S - クエリー I の提示」および「S - クエリー I I の提示」、異なるタイプのチャンクの対のクエリーについての「C - クエリー I の提示」および「C - クエリー I I の提示」、並びに、コロケーションの個々の単語の対に基づくクエリーについての「W - クエリーの提示」、という名前が付されたボタンがある。

【 0 0 6 2 】

ユーザインターフェース 4 1 0 の下部分は、現在の置換コロケーション候補の単語に対して呼び出され得るいくつかのクエリーの結果のテキストボックスの例示として、文レベルのクエリー用のテキストボックス 4 3、チャンクの第 1 のレベルのクエリー用のテキストボックス 4 5、およびチャンクの第 2 のレベルのクエリー用のテキストボックス 4 7を含む。各テキストボックス 4 3、4 5 および 4 7 は、また、そのコロケーション候補のクエリー検索結果のヒット数、および、右側に、各クエリーのタイプに対応するコーパスからの幾つかのクエリー検索結果を表示する別個のテキストボックスとともに表示されている。その図により示唆されるように、本例示の例においては、校正されているコロケーションにおいて各クエリー・テンプレートのレベルで、「known」が「recognized」の適切な置換語として強く表示される。これは、クエリー・テンプレートの段階的なスキームが与えた冗長性が、この場合は、1 つの特定の置換候補に対する説得力のあるケースを提示するには十分すぎたことを例示する。次に、ユーザは、自分の所望の各クエリー・テンプレートのヒット数を考慮し、置換コロケーション候補の 1 つを選択して、これが対応する好ましくない単語コロケーションを置換することができる。

【 0 0 6 3 】

図 1 および図 4 は、専用のアプリケーションの実施形態に関連するグラフィックディスプレイパネルを示すが、別の例示の実施形態は、ワードプロセッシングプログラム、ウェブブラウザ、Eメールアプリケーション、プレゼンテーションプログラム、または、テキストもしくはは自然言語の別の形式のユーザ入力を伴う他のアプリケーション等の、別のアプリケーションと接続して機能しえる。そのような他のプログラムまたはアプリケーションは、別個の誤りコロケーション校正プログラムの実施形態と協働するか、または、例えば誤りコロケーション校正の実施形態から成る、それ自身のモジュールを含むことができる。コロケーション誤りの校正ツールは、異なる実施形態においては、例えば他のアプリケーションのツールメニューの下でアクセス可能であり、または別の例として、ユーザが単語の上で右クリックするとポップアップメニュー内の選択肢としてアクセス可能であり得る。

【 0 0 6 4 】

さらに別の動作モードにおいて、他のプログラムは、誤りコロケーション校正モジュールまたはプログラムを、デフォルトの設定で自動的に作動させることができ、また、誤りのある、または好ましくないコロケーションを、モニタ上の単語に付けられるマーキングを用いて自動的に表示することができる。マーキングは、例えばコロケーションの下で青いギザギザの線の形式をとり、そのコロケーションを誤りのあるまたは好ましくないものとして示すことができる。さらに別の動作モードにおいて、別個のモジュールまたはプログラムが作動し、例えば、ユーザ入力が、ユーザがその入力言語のネイティブユーザではないことを示すものを含むかどうかを検出し、非ネイティブの使用の表示は、誤りコロケーションモジュールまたはプログラムの起動のトリガとなる。

【 0 0 6 5 】

図 5 は、種々の実施形態を実装可能な適切なコンピュータシステム環境 1 0 0 の一例を例示する。例えば、種々の実施形態は、ソフトウェアアプリケーション、モジュール、または、コンピュータシステム環境 1 0 0 により実行可能な、およびコンピュータシステム環境 1 0 0 を構成して異なる実施形態に含まれる種々のタスクまたは方法を実行させる、他の形式の命令として実装される。コロケーション誤りの校正の実施形態を具現化するソフトウェアアプリケーションまたはモジュールは、任意の様々なプログラミング言語もしくははスクリプト言語、またはプログラミング環境もしくははスクリプト環境で開発され得る

。例えば、C #、F #、C ++、C、P a s c a l、V i s u a l B a s i c、J a v a (登録商標)、J a v a (登録商標) S c r i p t、D e l p h i、E i f f e l、N e m e r l e、P e r l、P H P、P y t h o n、R u b y、V i s u a l F o x P r o、L u a、または任意の他のプログラミング言語で書くことができる。また新しいプログラミング言語および他の形式の作成可能命令の開発が続けられ、そこではさらなる実施形態が容易に開発され得るということも想定される。

【 0 0 6 6 】

例示の一実施形態によると、コンピュータシステム環境 1 0 0 は、テキスト内の単語コロケーションの表示を受け取るのに応答して、コロケーション誤りの校正タスクを実行するように構成され得る。次に、コンピュータシステム環境 1 0 0 は、表示される単語コロケーションに関連する 1 つまたは複数のクエリー・テンプレートのそれぞれに対して、ウェブ検索を実行することができる。使用される種々のクエリー・テンプレートは、文、要約された文、チャンクの対、および / または個々の単語の対を含むが、これらの内の任意のものが単語コロケーションを含み得る。コンピュータシステム環境 1 0 0 は、次に、各クエリー・テンプレートに対するウェブ検索の結果が、単語コロケーションが通常の使用に対応していることを示すかどうか、または、それが好ましくないかもしくは考え得る誤りを示しているかどうかを評価することができる。通常の使用がなされているかどうかは、文を含むクエリー・テンプレートが正確に一致すること、または事前に選択された閾値より大きい一致スコアによって示すことができる。次に、システムは、ユーザが認識可能な出力装置を介したコンピュータシステム環境 1 0 0 の出力の一部として、コロケーション誤りの校正方法の実施形態の結果として、単語コロケーションが通常の使用に対応するか、または好ましくないため誤った使用であるとして示されるかどうかを示すことができる。

【 0 0 6 7 】

図 5 に示すコンピュータシステム環境 1 0 0 は、種々の実施形態を実行し、種々の実施形態からの出力を提供する、適切なコンピュータ環境の一例に過ぎず、請求される主題の使用または機能の範囲に任意の制限を示唆することは意図されていない。コンピュータ環境は、例としての動作環境 1 0 0 に示されるコンポーネントの任意の 1 つまたは組合せに関して依存も要求も有するものとして解釈されるべきではない。

【 0 0 6 8 】

実施形態は、多数の他の汎用もしくは専用のコンピュータシステム環境または構成で動作する。種々の実施形態の使用に適切な、既知のコンピュータシステム、環境、および / または構成の例は、パーソナルコンピュータ、サーバコンピュータ、ハンドヘルドまたはラップトップデバイス、マルチプロセッサシステム、マイクロプロセッサベースのシステム、セットトップボックス、プログラム可能家庭用電化製品、ネットワーク P C、ミニコンピュータ、メインフレームコンピュータ、電話技術システム、任意の上記のシステムまたはデバイスを含む分散コンピュータ環境等を含むが、これに限定されない。

【 0 0 6 9 】

実施形態は、プログラムモジュール等の、コンピュータによって実行されるコンピュータ実行可能命令の一般的なコンテキストで説明され得る。一般に、プログラムモジュールは、特定のタスクを実行し、または特定の抽象データ型を実装するルーチン、プログラム、オブジェクト、コンポーネント、データ構造等を含む。幾つかの実施形態は、通信ネットワークを通じてリンクされるリモート処理装置によりタスクが実行される、分散コンピュータ環境において実践されるように設計される。分散コンピュータ環境において、プログラムモジュールは、メモリストレージデバイスを含む、ローカルおよびリモート両方のコンピュータ記憶媒体内に置かれる。本明細書に記載されるように、そのような実行可能命令は、媒体に格納され、コンピュータシステムの 1 つまたは複数のコンポーネントによって読み込まれ実行されることが可能にであり、それによって新しい能力を有するコンピュータシステム構成する。

【 0 0 7 0 】

図5を参照すると、幾つかの実施形態を実装する例としてのシステムは、コンピュータ110の形式の汎用のコンピューティングデバイスを含む。コンピュータ110のコンポーネントは、プロセッシングユニット120、システムメモリ130、および、システムメモリを含む種々のシステムコンポーネントをプロセッシングユニット120に連結するシステムバス121を含むことができるが、これに限定されない。システムバス121は、メモリバスまたはメモリコントローラ、周辺機器用バス、および任意の様々なバスアーキテクチャを使用するローカルバスを含む、任意の数タイプのバス構造である。限定ではなく例として、上記アーキテクチャは、業界標準アーキテクチャ（ISA）バス、マイクロ・チャンネル・アーキテクチャ（MCA）バス、拡張ISA（EISA）バス、VESA（Video Electronics Standards Association）ローカルバス、および、メザンバスとしても既知のPCI（Peripheral Component Interconnect）バスを含む。

【0071】

コンピュータ110は、典型的には、様々なコンピュータ可読媒体を含む。コンピュータ可読媒体は、コンピュータ110によってアクセスでき、揮発性および不揮発性媒体、着脱可能および着脱不可能な媒体の両方を含む、任意の利用可能な媒体であることが可能である。限定ではなく例として、コンピュータ可読媒体は、コンピュータ記憶媒体および通信媒体を含むことができる。コンピュータ記憶媒体は、コンピュータ可読命令、データ構造、プログラムモジュール、または他のデータ等の情報を格納する任意の方法または技術で実施される、揮発性および不揮発性、着脱可能および着脱不可能の両方の媒体を含む。コンピュータ記憶媒体には、RAM、ROM、EEPROM、フラッシュメモリ、もしくは他のメモリ技術、CD-ROM、デジタル多用途ディスク（DVD）、もしくは他の光ディスクストレージ、磁気カセット、磁気テープ、磁気ディスクストレージ、もしくは他の磁気ストレージデバイス、または、所望の情報を格納するのに使用でき、コンピュータ110によってアクセスできる任意の他の媒体が含まれるが、これに限定されない。通信媒体は、典型的には、搬送波または他の転送機構などの変調されたデータ信号内のコンピュータ可読命令、データ構造、プログラムモジュール、または他のデータを具現化し、任意の情報配信媒体を含む。用語「変調されたデータ信号」は、信号内で情報を符号化する方法でその特性の1つまたは複数を設定または変更された信号を意味する。限定ではなく例として、通信媒体は、有線ネットワークまたは直接配線接続などの有線媒体と、音響、RF、赤外線、および他の無線媒体などの無線媒体を含む。上記の任意の組合せも、コンピュータ可読媒体の範囲に含まれるべきである。

【0072】

システムメモリ130は、読み出し専用メモリ（ROM）131およびランダムアクセスメモリ（RAM）132等の揮発性および/または不揮発性メモリの形態のコンピュータ記憶媒体を含む。起動中などに、コンピュータ110内の要素の間での情報の転送を助ける基本ルーチンを含む基本入出力システム133（BIOS）が、典型的にはROM131に格納される。RAM132は、典型的には、プロセッシングユニット120に即座にアクセス可能、かつ/またはプロセッシングユニット120によって現在作動中のデータおよび/またはプログラムモジュールを含む。限定ではなく例として、図5に、オペレーティングシステム134、アプリケーションプログラム135、他のプログラムモジュール136、およびプログラムデータ137を例示する。

【0073】

コンピュータ110は、また、他の着脱可能/着脱不可能の、揮発性/不揮発性のコンピュータ記憶媒体も含むことができる。単に例として、図5は、着脱不可能な不揮発性磁気記憶媒体から読み取りまたはこれに書き込むハードディスクドライブ141、着脱可能な不揮発性磁気ディスク152から読み取りまたはこれに書き込む磁気ディスクドライブ151、および、CD-ROMまたは他の光媒体などの着脱可能な不揮発性光ディスク156から読み取りまたはこれに書き込む光ディスクドライブ155を示す。例示的な動作環境で使用できる他の着脱可能/着脱不可能の、揮発性/不揮発性のコンピュータ記憶媒

10

20

30

40

50

体は、磁気テープカセット、フラッシュメモリカード、デジタル多用途ディスク、デジタルビデオテープ、ソリッドステートRAM、ソリッドステートROM等を含むが、これに限定されない。ハードディスクドライブ141は、通常は、インターフェース140等の着脱不可能なメモリインターフェースを介してシステムバス121に接続され、磁気ディスクドライブ151および光ディスクドライブ155は、典型的には、インターフェース150などの着脱可能なメモリインターフェースによってシステムバス121に接続される。

【0074】

上で説明し、図5に示されたドライブおよびそれに関連するコンピュータ記憶媒体は、コンピュータ110のコンピュータ可読命令、データ構造、プログラムモジュール、および他のデータのストレージを提供する。図5において、例えば、ハードディスクドライブ141が、オペレーティングシステム144、アプリケーションプログラム145、他のプログラムモジュール146、およびプログラムデータ147を格納するものとして図示されている。これらのコンポーネントを、オペレーティングシステム134、アプリケーションプログラム135、他のプログラムモジュール136、およびプログラムデータ137と同一のまたは異なるもののいずれかとすることができることに留意されたい。オペレーティングシステム144、アプリケーションプログラム145、他のプログラムモジュール146、およびプログラムデータ147は、ここで、最低限でも異なるコピーであることを示すために異なる符号を与えられている。

【0075】

ユーザは、キーボード162、マイク163、および、一般にマウス、トラックボール、またはタッチパッドのようなポインティングデバイス161などの入力装置を介して、コンピュータ110にコマンドおよび情報を入力することができる。他の入力装置（図示せず）は、ジョイスティック、ゲームパッド、衛星放送用パラボラアンテナ、スキャナ等、を含むことができる。これらおよび他の入力装置は、しばしば、システムバスに連結されるユーザ入力インターフェース160を介してプロセッシングユニット120に接続されるが、パラレルポート、ゲームポート、またはUSB(universal serial bus)等の他のインターフェースおよびバス構造によって接続することができる。モニタ191または他のタイプのディスプレイ装置は、また、ビデオインターフェース190等のインターフェースを介してシステムバス121に接続される。モニタに加えて、コンピュータはまた、スピーカー197およびプリンタ196等の他の周辺出力装置を含み得るが、これらは出力周辺インターフェース195を介して接続され得る。

【0076】

コンピュータ110は、リモートコンピュータ180等の1つまたは複数のリモートコンピュータへの論理接続を使用して、ネットワーク化された環境で動作する。リモートコンピュータ180は、パーソナルコンピュータ、ハンドヘルドデバイス、サーバ、ルータ、ネットワークPC、ピアデバイス、または他の一般的なネットワークノードであり得、典型的には、コンピュータ110に関して上記した要素の多くまたはすべてを含む。図5に示される論理接続は、ローカルエリアネットワーク(LAN)171および広域ネットワーク(WAN)173を含むが、他のネットワークを含むこともできる。そのようなネットワーク環境は、オフィス、企業規模のコンピュータネットワーク、イントラネット、およびインターネットでは一般的なものである。

【0077】

LANネットワーク環境で使用される場合、コンピュータ110は、ネットワークインターフェースまたはネットワークアダプタ170を介してLAN171に接続される。WANネットワーク環境で使用される場合、コンピュータ110は、典型的には、インターネットなどのWAN173を介する通信を確立する、モデム172または他の手段を含む。モデム172は、内蔵または外付けであり得るが、ユーザ入力インターフェース160または他の適正な機構を介してシステムバス121に接続され得る。ネットワーク化された環境では、コンピュータ110に関連して示されるプログラムモジュールまたはその一

部を、リモートメモリストレージデバイスに格納することができる。限定ではなく例として、図5は、リモートコンピュータ180に存在するものとしてリモートアプリケーションプログラム185を示す。図示のネットワーク接続は例示的であり、コンピュータの間の通信リンクを確立する他の手段を使用できることは理解されるであろう。

【0078】

図6は、別の例示の実施形態による、一般的なモバイルコンピューティング環境のブロック図を示す。この環境は、モバイルコンピューティングデバイスと、そのモバイルコンピューティングデバイスにより読み取り可能であり、モバイルコンピューティングデバイスにより実行可能な実行可能命令を含む媒体とを含む。例示の実施形態によると、図6は、モバイルデバイス201を含むモバイルコンピューティングシステム200のブロック

10

【0079】

メモリ204は、バッテリーバックアップモジュール（図示せず）を有するランダムアクセスメモリ（RAM）等の不揮発性電子メモリとして実装され、モバイルデバイス200への総電源がシャットダウンされた場合に、メモリ204に格納された情報が消えないようにする。メモリ204の一部は、例示的には、プログラム実行用のアドレス可能メモリとして割り付けられ、メモリ204の他の部分は、例示的には、格納のために使用されデ

20

【0080】

メモリ204は、オペレーティングシステム212、アプリケーションプログラム214、ならびにオブジェクトストア216を含む。動作中、オペレーティングシステム212は、例示的には、プロセッサ202によってメモリ204から実行される。オペレーティングシステム212は、例示の一実施形態において、本件特許出願人から市販されているWINDOWS（登録商標）CEブランドのオペレーティングシステムである。オペレーティングシステム212は、例示的には、モバイルデバイス用に設計され、一組の露出されたアプリケーションプログラミングインターフェースおよび方法を介して、アプリケー

30

【0081】

通信インターフェース208は、モバイルデバイス200が情報を送受信することを可能にする多数の装置および技術を表す。その装置には、2、3例を挙げると、有線および無線のモデム、衛星放送受信機、および、放送チューナが含まれる。モバイルデバイス200は、また、直接コンピュータに接続され、これとデータ交換することが可能である。そのような場合、通信インターフェース208は、赤外線トランシーバー、またはシリアルもしくはパラレルの通信接続であり得、それらは全てストリーミング情報を転送する

40

【0082】

入出力コンポーネント206には、タッチセンサ式スクリーン、ボタン、ローラ、およびマイク等の様々な入力装置が含まれ、同様に、オーディオジェネレータ、バイブレーションデバイス、およびディスプレイを含む様々な出力装置が含まれる。上記に挙げた装置は、例示のためのものであり、モバイルデバイス200上に全てが存在することが必要なわけではない。加えて、他の入出力装置を、モバイルデバイス200に取り付けるか、またはモバイルデバイス200に見付けることができる。

【0083】

モバイルコンピューティングシステム200はまた、ネットワーク220を含む。モバ

50

イルコンピューティングデバイス 201 は、例示的には、例えばインターネット、広域ネットワーク、またはローカルエリアネットワークのような、ネットワーク 220 との無線通信を、通信インターフェース 208 と無線インターフェース 222 との間の適切なプロトコルの電磁信号 299 の送受信により行う。無線インターフェース 222 は、例えば無線ハブもしくは携帯電話アンテナであり得、または任意の他の信号インターフェースであり得る。無線インターフェース 222 は、例示的にコンピューティングリソース 224 および 226 で表される、多様な追加のコンピューティングリソースへの、ネットワーク 220 を介したアクセスを提供する。当然、任意の場所の任意の数のコンピューティングデバイスが、ネットワーク 220 との通信接続を行うことができる。幾つかの例示の実施形態において、コンピューティングデバイス 201 は、コンピューティングデバイス 201 が母国語でない言語校正クエリーおよび他のタスクを実行可能にする実行可能命令のような、メモリコンポーネント 204 の媒体上に格納される実行可能命令の使用が可能にされる。

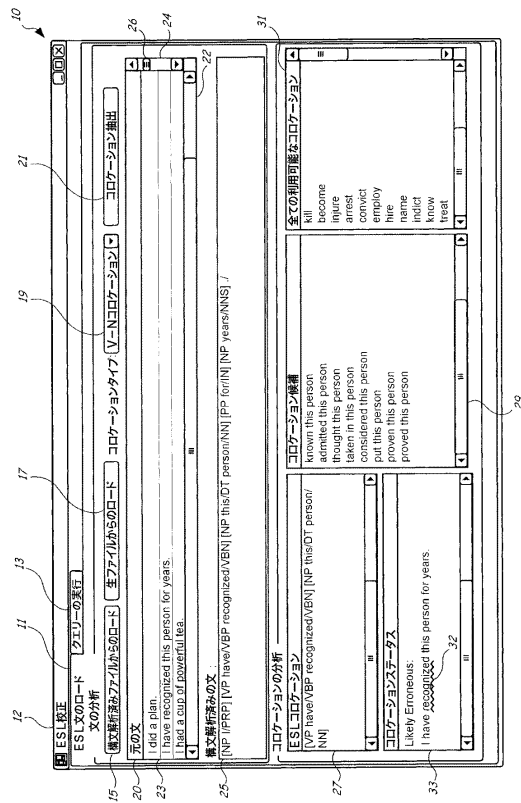
10

【0084】

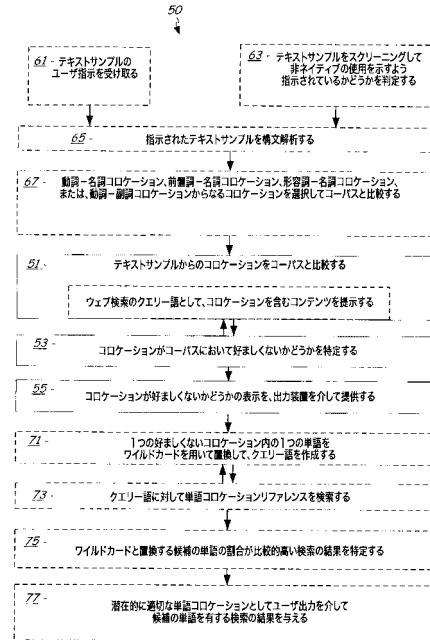
本主題が、構造の特徴および/または方法論的な動作に特有の言葉で記載されたが、添付の特許請求の範囲で定義される主題は、必ずしも上述の特有の特徴または動作に限定されないことは理解されよう。むしろ、上述の特有の特徴および動作は、本主題を実現するための例示の形として開示される。特別な例として、用語「コンピュータ」、「コンピューティングデバイス」または「コンピュータシステム」は、本明細書において便宜上、時に単独で使用されることがあるが、これらのそれぞれは、任意のコンピューティングデバイス、コンピュータシステム、コンピュータ環境、モバイルデバイス、または他の情報処理コンポーネントもしくはコンテキストに言及し、任意の個々の解釈に限定されないことがよく理解されるであろう。別の特別な例として、多くの実施形態が、特許出願時には広く知られた例示の要素で提示されるが、コンピュータ技術における多くの新しい手法が、ユーザインターフェース、ユーザ入力方法、コンピュータ環境、およびコンピューティング方法等の態様において、異なる実施形態の要素に影響を及ぼすであろうこと、および、特許請求の範囲により定義される要素が、これらおよび他の革新的な進歩に従って具現化される一方で、依然として本明細書の特許請求の範囲により定義される要素に一致しかつその要素に包含されたままであることが想定される。

20

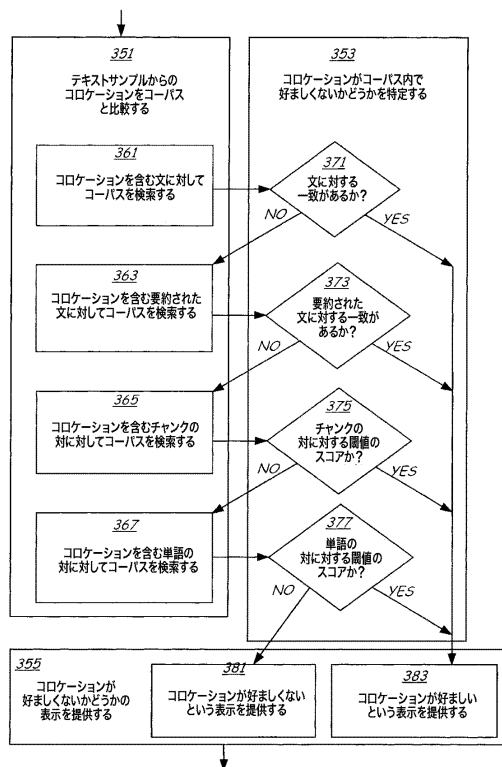
【図 1】



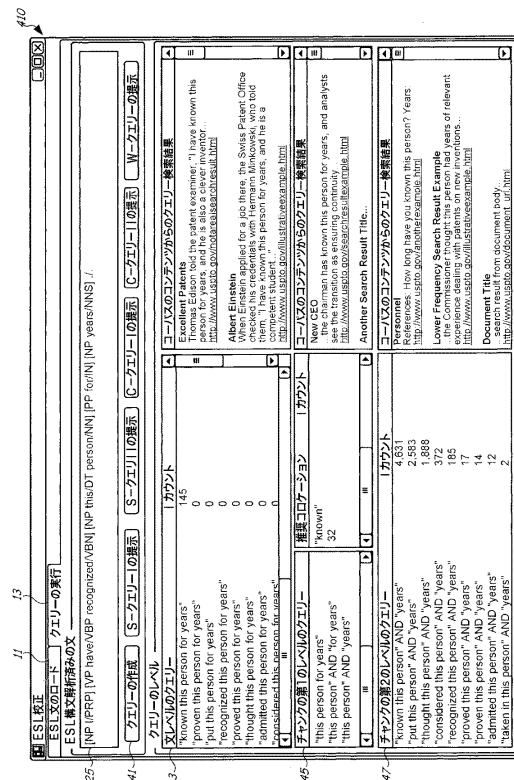
【図 2】



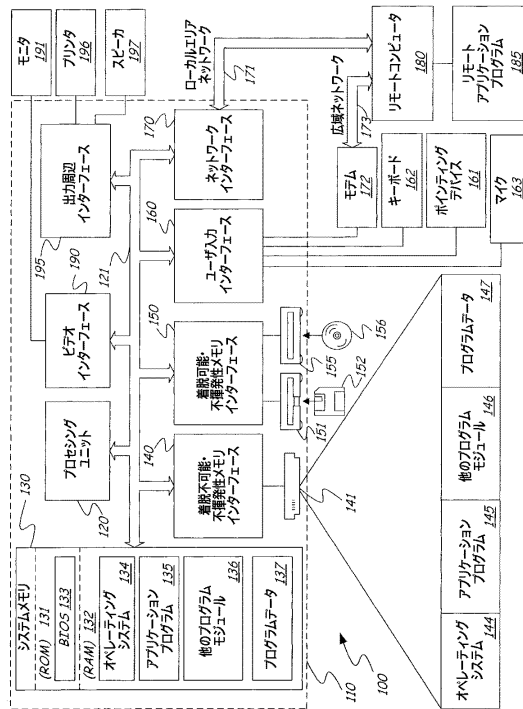
【図 3】



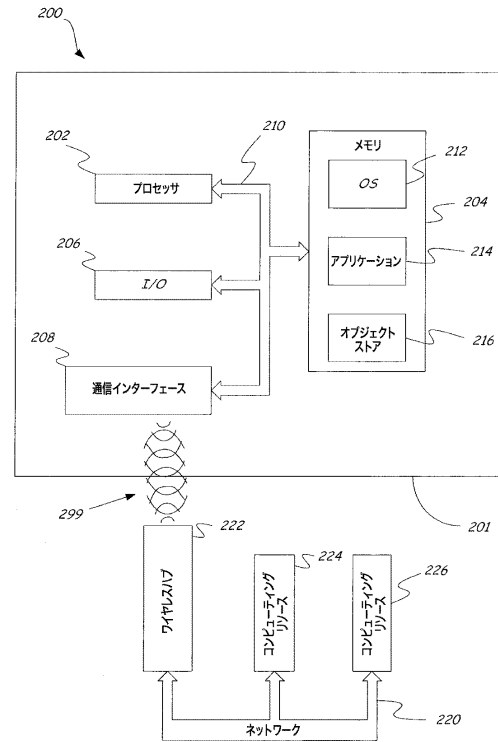
【図 4】



【図 5】



【図 6】



フロントページの続き

- (72)発明者 ウィリアム ビー・ドラン
アメリカ合衆国 98052 ワシントン州 レッドモンド ワン マイクロソフト ウェイ マ
イクロソフト コーポレーション インターナショナル パテンツ内
- (72)発明者 シャオ・ウェン ホン
アメリカ合衆国 98052 ワシントン州 レッドモンド ワン マイクロソフト ウェイ マ
イクロソフト コーポレーション インターナショナル パテンツ内
- (72)発明者 ミン チョウ
アメリカ合衆国 98052 ワシントン州 レッドモンド ワン マイクロソフト ウェイ マ
イクロソフト コーポレーション インターナショナル パテンツ内

審査官 長 由紀子

- (56)参考文献 特開2004-038969(JP,A)
特開2001-101186(JP,A)
特開平10-069483(JP,A)
T Robb, Google as a Quick'n Dirty Corpus Tool, TESL-EJ, 2003年 9月30日, Volume
7, Number 2, [平成23年8月19日検索], URL, <http://www.tesl-ej.org/wordpress/issues/volume7/ej26/ej26int/>
C Shei et.al, An ESL Writer's Collocational Aid, Computer Assisted Language Learning
, Division of Informatics, Unibersity of Edinburgh, 2000年 2月29日, Volume 13
, p.167-182, [平成23年8月19日検索], URL, http://scholar.google.co.jp/scholar?hl=ja&q=an+esl+writer%27s+collocational+aid&btnG=%E6%A4%9C%E7%B4%A2&lr=&as_ylo=&as_vis=0
大鹿 広憲 外3名, Googleを活用した英作文支援システムの構築, DEWS2005論
文集 [online], 日本, (社)電子情報通信学会データ工学研究専門委員会, 2005
年 5月 2日, DEW20054B-i8
大鹿 広憲 外3名, 検索エンジンを使った翻訳サポートシステムの構築, 情報処理学会研究報
告, 日本, 社団法人情報処理学会, 2004年 7月15日, 第2004巻第72号, p.585-591

- (58)調査した分野(Int.Cl., DB名)
G06F 17/21 -28