

(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2005-11358
(P2005-11358A)

(43) 公開日 平成17年1月13日(2005.1.13)

(51) Int.Cl.⁷
G06F 17/28

F I
G O 6 F 17/28

T

テーマコード (参考)
5 B O 9 1

審査請求 未請求 請求項の数 30 O L (全 27 頁)

(21) 出願番号	特願2004-183026 (P2004-183026)	(71) 出願人	500046438 マイクロソフト コーポレーション アメリカ合衆国 ワシントン州 9805 2-6399 レッドモンド ワン マイ クロソフト ウェイ
(22) 出願日	平成16年6月21日 (2004. 6. 21)	(74) 代理人	100077481 弁理士 谷 義一
(31) 優先権主張番号	10/600, 297	(74) 代理人	100088915 弁理士 阿部 和夫
(32) 優先日	平成15年6月20日 (2003. 6. 20)	(72) 発明者	スティーブン ディー. リチャードソン アメリカ合衆国 98053 ワシントン 州 レッドモンド ノースイースト 80 ストリート 24550
(33) 優先権主張国	米国 (US)		
(31) 優先権主張番号	10/626, 925		
(32) 優先日	平成15年7月25日 (2003. 7. 25)		
(33) 優先権主張国	米国 (US)		

最終頁に続く

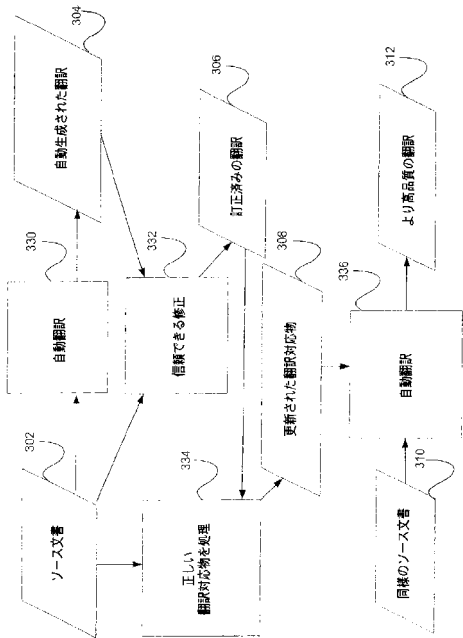
(54) 【発明の名称】 適応型機械翻訳

(57) 【要約】

【課題】 自動機械翻訳システムに情報を提供して、翻訳の精度を向上させるためのコンピュータに実装される方法を開示する。

【解決手段】 この方法は、ソーステキストの集まりを受け取ることを含む。このソーステキストの集まりに対応する試みの翻訳を、自動機械翻訳システムから受け取る。試みの翻訳中の少なくとも1つの誤りを訂正するよう構成されている訂正入力も受け取る。最終的に、自動機械翻訳システムによって生成されるその後の翻訳においてその誤りが繰り返される可能性を低減するよう自動機械翻訳システムに情報を提供する。

【選択図】 図3



【特許請求の範囲】**【請求項 1】**

自動機械翻訳システムに情報を提供して翻訳の正確さを向上させるためのコンピュータに実施される方法であって、

ソーステキストの集まりを受け取ることと、

前記ソーステキストの集まりに対応する試みの翻訳を前記自動機械翻訳システムから受け取ることと、

前記試みの翻訳の中の少なくとも 1 つの誤りの訂正を行うよう構成されている訂正入力を受け取ることと、

前記自動機械翻訳システムに情報を提供して前記自動機械翻訳システムによって生成されるその後の翻訳において前記誤りが繰り返されることになる可能性を低減することと
を備えることを特徴とする方法。 10

【請求項 2】

前記情報を提供することは、前記訂正入力を提供することを備えることを特徴とする請求項 1 に記載の方法。

【請求項 3】

前記情報を提供することは、前記自動機械翻訳システムに同化させるべき情報を提供することを備えることを特徴とする請求項 1 に記載の方法。

【請求項 4】

前記情報を提供することは、前記自動機械翻訳システムに関連する知識ソースに同化させるべき更新情報を提供することを備えることを特徴とする請求項 1 に記載の方法。 20

【請求項 5】

前記情報を提供することは、前記自動機械翻訳システムに関連する翻訳対応物に同化させるべき更新情報を提供することを備えることを特徴とする請求項 1 に記載の方法。

【請求項 6】

前記情報を提供することは、前記自動機械翻訳システムに関連する言語構造の集まりに同化させるべき更新情報を提供することを備えることを特徴とする請求項 1 に記載の方法。
。

【請求項 7】

前記情報を提供することは、前記自動機械翻訳システムに関連する対応する論理形式のデータベースに同化させるべき更新情報を提供することを備えることを特徴とする請求項 6 に記載の方法。 30

【請求項 8】

前記情報を提供することは、前記自動機械翻訳システムに関連する統計パラメータの集まりに同化させるべき更新情報を提供することを備えることを特徴とする請求項 1 に記載の方法。

【請求項 9】

前記情報を提供することは、前記自動機械翻訳システムに関連する構文解析情報の集まりに同化させるべき更新情報を提供することであって、前記構文解析情報はパーサがセグメントの集まりの解析を提供できるようにする情報であることを備えることを特徴とする
請求項 1 に記載の方法。 40

【請求項 10】

前記情報を提供することは、前記自動機械翻訳システムに関連する対応する語または句のグループの集まりに同化させるべき更新情報を提供することを備えることを特徴とする請求項 1 に記載の方法。

【請求項 11】

異なる自動機械翻訳システムに関連する知識ソースに同化させるべき更新情報を、ネットワークを介して伝送することであって、前記更新情報は、前記自動機械翻訳システムによって生成されるその後の翻訳において前記誤りが繰り返されることになる可能性を低減するよう構成されていることをさらに備えることを特徴とする請求項 1 に記載の方法。 50

【請求項 1 2】

前記情報を提供することは、1つまたは複数のセンテンスペアのバイリンガルコーパスを提供することを備えることを特徴とする請求項 1 に記載の方法。

【請求項 1 3】

前記訂正入力を受け取ることは、少なくとも1つの訂正指示を人間のトランスレータから受け取ることを備えることを特徴とする請求項 1 に記載の方法。

【請求項 1 4】

前記試みの翻訳を受け取ることは、前記機械翻訳システムが適切な翻訳を提供していないソーステキストの一部に対応する試みの翻訳を受け取ることを備えることを特徴とする請求項 1 に記載の方法。

10

【請求項 1 5】

自動機械翻訳システムの性能を向上させるためのコンピュータに実施される方法であって、

前記自動機械翻訳システムを利用して、ソーステキストの集まりの翻訳を生成することと、

前記ソーステキストの集まりと、前記翻訳の少なくとも一部とを信頼できる修正ソースに転送することと、

前記翻訳の少なくとも一部の中の誤りの指摘を前記信頼できる修正ソースから受け取ることと、

前記自動機械翻訳システムによって生成されるその後の翻訳について前記誤りが発生する可能性が少なくなるように前記自動機械翻訳システムをトレーニングすることと

20

を備えることを特徴とする方法。

【請求項 1 6】

前記翻訳に関する品質測定を表す信頼メトリックを生成することと、

少なくともある程度は前記信頼メトリックに基づいて、前記信頼できる修正ソースに転送される前記翻訳の一部を選択することと

をさらに備えることを特徴とする請求項 1 5 に記載の方法。

【請求項 1 7】

前記転送することは、前記自動機械翻訳システムが実施されているクライアントコンピューティングデバイスから、前記信頼できる修正ソースに関連するサーバコンピューティングデバイスに転送することを備えることを特徴とする請求項 1 5 に記載の方法。

30

【請求項 1 8】

前記転送することは、前記自動機械翻訳システムが実施されているサーバから、前記信頼できる修正ソースに関連するサーバコンピューティングデバイスに転送することを備えることを特徴とする請求項 1 5 に記載の方法。

【請求項 1 9】

前記自動機械翻訳システムをトレーニングすることは、前記自動機械翻訳システムに関連する知識ソースを更新することを備えることを特徴とする請求項 1 5 に記載の方法。

【請求項 2 0】

前記自動機械翻訳システムをトレーニングすることは、前記自動機械翻訳システムに関連する少なくとも1つの翻訳対応物を更新することを備えることを特徴とする請求項 1 5 に記載の方法。

40

【請求項 2 1】

前記自動機械翻訳システムをトレーニングすることは、前記自動機械翻訳システムに関連する言語構造の集まりを更新することを備えることを特徴とする請求項 1 5 に記載の方法。

【請求項 2 2】

前記自動機械翻訳システムをトレーニングすることは、前記自動機械翻訳システムに関連する対応する論理形式のデータベースを更新することを備えることを特徴とする請求項 2 1 に記載の方法。

50

【請求項 23】

前記自動機械翻訳システムをトレーニングすることは、前記自動機械翻訳システムに関連する統計パラメータの集まりを更新することを備えることを特徴とする請求項 15 に記載の方法。

【請求項 24】

前記自動機械翻訳システムをトレーニングすることは、前記自動機械翻訳システムに関連する構文解析情報の集まりを更新することであって、前記構文解析情報はパーサがセグメントの集まりの解析を提供できるようにする情報であることを備えることを特徴とする請求項 15 に記載の方法。

【請求項 25】

前記自動機械翻訳システムをトレーニングすることは、前記自動機械または句翻訳システムに関連する対応する語関連物の集まりを更新することを備えることを特徴とする請求項 15 に記載の方法。

【請求項 26】

前記自動機械翻訳システムをトレーニングすることは、前記誤りに基づくバイリンガルコーパスを前記自動機械翻訳システムに提供することと、前記自動機械翻訳システムが前記バイリンガルコーパスに基づいてそれ自体をトレーニングすることができるようにすることとを備えることを特徴とする請求項 15 に記載の方法。

【請求項 27】

自動機械翻訳システムの性能を向上させるための方法であって、
前記自動機械翻訳システムを利用してソーステキストの集まりの翻訳を生成することであって、信頼メトリックが前記翻訳の部分に関連していることと、
前記信頼メトリックを評価し、前記翻訳の低信頼部分を選択することと、
前記低信頼部分を、コンピュータネットワークを介して信頼できる修正ソースに伝送することと、
前記信頼できる修正ソースを利用して前記低信頼部分の訂正済みバージョンを生成することと、
前記低信頼部分の前記訂正済みバージョンに基づいて翻訳知識の更新済みデータベースを生成することと、
コンピュータネットワークを介して翻訳知識の前記更新済みデータベースを前記自動機械翻訳システムに伝送することと、
翻訳知識の前記更新済みデータベースを前記自動機械翻訳システムに組み込んで、前記自動機械翻訳システムが前記低信頼部分と同様のテキストをその後より高い正確さで翻訳できるようにすることと
を備えることを特徴とする方法。

【請求項 28】

前記信頼できる修正ソースを利用して訂正済みバージョンを生成することは、人間のトランスレータを利用することを備えることを特徴とする請求項 27 に記載の方法。

【請求項 29】

前記コンピュータネットワークを介して伝送することは、インターネットを介して伝送することを備えることを特徴とする請求項 27 に記載の方法。

【請求項 30】

第 1 の自己カスタマイズ自動機械トランスレータの性能を向上させるための方法であって、
前記第 1 の自己カスタマイズ自動トランスレータを第 1 のコンピューティングデバイス上に実施することと、
第 2 の自己カスタマイズ自動トランスレータを第 2 のコンピューティングデバイス上に実施することと、
信頼できる翻訳ソースを提供することと、
前記第 1 および第 2 のコンピューティングデバイス間の通信を可能にすることと、

10

20

30

40

50

前記第2のコンピューティングデバイスにおいてソーステキストを受け取ることと、

前記第2のコンピューティングデバイスに、前記信頼できる翻訳ソースによって生成された前記試みの翻訳の訂正済みバージョンを供給することと、

前記第2の自己カスタマイズ自動トランスレータを利用して前記ソーステキストおよび前記試みの翻訳の訂正済みバージョンを処理して、前記ソーステキストと同様のテキストをその後より高い正確さで翻訳するように前記第1の自己カスタマイズ自動トランスレータを適応化するためのトレーニング情報を生成することと、

前記トレーニング情報を前記第2のコンピューティングデバイスから前記第1のコンピューティングデバイスに転送することと、

前記トレーニング情報を前記第1の自己カスタマイズ自動トランスレータの中に同化させて、前記第1の自己カスタマイズ自動トランスレータが前記ソーステキストと同様のテキストをその後より高い正確さで翻訳できるようにすることとを備えることを特徴とする方法。 10

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、機械翻訳に関する。より詳細には、本発明は、信頼できるソースから訂正済み翻訳を取得する通常のワークフロー内でユーザの自動機械翻訳システムの性能を体系的に向上させるための手段に関する。

【背景技術】

20

【0002】

インターネットなどの技術によって創造された国際的コミュニティの成長の結果、機械翻訳、より詳細には、自然言語テキストを翻訳するためのコンピュータシステムの利用が、近年、より広範に普及してきている。いくつかの事例では、機械翻訳は自動的に行うことができる。しかし、質の高い翻訳を作成するプロセスに、人とのやりとりを組み込む場合がある。一般的に言って、人的資源に依拠する翻訳はより正確ではあるが、完全に自動化されたシステムよりも時間およびコストの効果が低い。翻訳システムによっては、翻訳の正確さが極めて重要なときにのみ人とのやりとりに依拠することがある。一般に、人とのやりとりに関連する時間およびコストを、特別に正確な翻訳が所望されるたびに、投資しなければならない。 30

【0003】

完全に自動化された機械翻訳によって生成される翻訳の質は、一般に、このようなシステムに対する需要の増大に伴って向上していない。特定の領域（または対象）についてより高品質の自動翻訳を得るためには、機械翻訳システムをかなりカスタマイズする必要があることが一般に認識されている。カスタマイズには、一般に、所望の領域のテキストを翻訳するための専門的なボキャブラリおよび規則の追加が含まれる。このようなカスタマイズは、一般に、トレーニングを受けたコンピュータ言語学者によって行われる。彼らは、半自動化ツールを使用して、ボキャブラリ項目をオンライン辞書に追加し、また、一般に専門化された規則の書き言葉（specialized rule writing languages）で言語的に方向付けられた規則（linguistically oriented rules）を書く。この種のカスタマイズは、比較的成本がかかる。 40

【0004】

【特許文献1】米国特許出願第10/309,950号

【特許文献2】米国特許出願第09/899,755号

【特許文献3】米国特許第5,966,686号明細書

【発明の開示】

【発明が解決しようとする課題】

【0005】

概して、消費者にとって様々なソースから利用可能な翻訳サービスは、コスト効率の良 50

い高品質のカスタマイズされた翻訳を提供することができない。例えば、パッケージおよびウェブベースの翻訳システムが、一般の人々に現在入手可能である。しかし、これらの翻訳システムは、特定の領域または対象に合わせてカスタマイズすることが難しいか、または不可能である。商業グレードの翻訳システムも入手可能である。これらのシステムは、特定の領域に合わせてカスタマイズすることができるが、しかし、そのカスタマイズプロセスは単調であり、一般に非常に高価である。直接的な人ベースの翻訳サービス（すなわち、ウェブベースおよびメールオーダーベースの人による翻訳サービス）も利用可能である。しかし、人による翻訳には、一般に、翻訳すべき文書ごとに料金を支払う必要があり、終わることのない費用が必要となる。

【課題を解決するための手段】

10

【0006】

本発明の実施形態は、翻訳の正確さを向上させるために自動機械翻訳システムに情報を提供するためのコンピュータに実装される方法に関する。この方法は、ソーステキストの集まりを受け取るステップを含む。ソーステキストの集まりに対応する試みの翻訳を、自動機械翻訳システムから受け取る。試みの翻訳の中の少なくとも1つの誤りを訂正するよう構成されている訂正入力も受け取る。最終的に、自動機械翻訳システムに情報を提供して、自動機械翻訳システムによって生成されるその後の翻訳でその誤りが繰り返される可能性を低減する。

【発明を実施するための最良の形態】

【0007】

20

I. 例示の動作環境

本発明の様々な態様は、信頼できるソースから訂正済みの翻訳を取得する通常のワークフロー内での適応型機械翻訳のカプセル化に関する。しかし、本発明をより詳細に説明するのに先立って、本発明を実施することができる例示的な環境の実施形態について説明する。

【0008】

図1は、本発明を実施することができる適切なコンピューティングシステム環境100の一例を示す。コンピューティングシステム環境100は、適切なコンピューティング環境の一例にすぎず、本発明の使用または機能の範囲についていかなる限定を示唆するものではない。また、コンピューティングシステム環境100を、例示的な動作環境100に示したコンポーネントのいずれか1つまたは組合せに関していかなる依存性、または必要性を有するものとして解釈すべきでもない。

30

【0009】

本発明は、その他の多くの汎用または専用コンピューティングシステム環境または構成で動作可能である。本発明と共に使用するのに適する可能性のある周知のコンピューティングシステム、環境、および/または構成の例には、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、プログラム可能な民生用電子機器、ネットワークPC、ミニコンピュータ、メインフレームコンピュータ、電話システム、上記のシステムまたはデバイスのいずれかを含む分散コンピューティング環境などが含まれるが、これらに限られない。

40

【0010】

本発明を、コンピュータによって実行されるプログラムモジュールなどのコンピュータ実行可能命令という一般的な状況で説明することができる。一般に、プログラムモジュールは、特定のタスクを実行したり、または特定の抽象データ型を実施したりするルーチン、プログラム、オブジェクト、コンポーネント、データ構造などを含む。本発明はまた、タスクが通信ネットワークを介してリンクされているリモート処理デバイスによって実行される分散コンピューティング環境において実施するように設計されている。分散コンピューティング環境では、プログラムモジュールは、メモリ記憶デバイスを含むローカルおよびリモートのコンピュータ記憶媒体の両方に配置される。プログラムおよびモジュール

50

によって実行されるタスクについて、以下に、図の助けを借りて説明する。当業者は、これらの説明および図を、任意の形態のコンピュータ可読媒体上に書き込むことができるプロセッサ実行可能命令として実施することができる。

【0011】

図1を参照すると、本発明を実施するための例示的なシステムは、コンピュータ110の形態の汎用コンピューティングデバイスを含む。コンピュータ110のコンポーネントには、処理装置120、システムメモリ130、および、システムメモリを含む様々なシステムコンポーネントを処理装置120に結合するシステムバス121が含まれ得るが、これらに限定されない。システムバス121は、様々なバスアーキテクチャのいずれかを使用するメモリバスまたはメモリコントローラ、周辺バス、およびローカルバスを含む、いくつかのバス構造のいずれであってもよい。限定ではなく、例として、このようなアーキテクチャには、ISA (Industry Standard Architecture) バス、MCA (Micro Channel Architecture) バス、EISA (Enhanced ISA) バス、VESA (Video Electronics Standards Association) ローカルバス、およびメザニンバスとしても知られるPCI (Peripheral Component Interconnect) が含まれるが、これらに限定されない。

10

【0012】

コンピュータ110は、一般に、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る任意の利用可能な媒体とすることができ、揮発性媒体および不揮発性媒体、取外し可および取外し不可の媒体を含む。限定ではなく、例として、コンピュータ可読媒体には、コンピュータ記憶媒体および通信媒体が含まれる。コンピュータ記憶媒体には、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、またはその他のデータなどの情報を格納するためにいずれかの方法または技術で実施された揮発性および不揮発性の取外し可および取外し不可の媒体が含まれる。コンピュータ記憶媒体は、RAM、ROM、EEPROM、フラッシュメモリもしくはその他のメモリ技術、CD-ROM、デジタル多用途ディスク (DVD) もしくはその他の光記憶装置、磁気カセット、磁気テープ、磁気ディスク記憶装置もしくはその他の磁気記憶装置、または所望の情報を格納するために使用でき、コンピュータ110によってアクセスすることができる任意のその他の媒体が含まれるが、これらに限定されない。

20

30

【0013】

通信媒体は、一般に、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、またはその他のデータを、搬送波またはその他のトランスポートメカニズムなどの変調データ信号に具体化し、そして任意の情報配信媒体を含む。「変調データ信号」という用語は、信号中に情報を符号化するような方法でその特性の1つまたは複数を設定または変更した信号を意味する。限定ではなく、例として、通信媒体には、有線ネットワークまたは直接配線接続などの有線媒体、ならびに音響、RF、赤外線、およびその他の無線媒体などの無線媒体が含まれる。上記のいずれの組合せも、コンピュータ可読媒体の範囲内に含まれるべきである。

【0014】

システムメモリ130は、読取り専用メモリ (ROM) 131およびランダムアクセスメモリ (RAM) 132などの揮発性および/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。起動時などにコンピュータ110内の構成要素間の情報の転送を支援する基本ルーチンを収容する基本入出力システム (BIOS) 133は、一般に、ROM 131に格納されている。RAM 132は、一般に、処理装置120によって直ぐにアクセス可能であり、そして/または現在操作されているデータおよび/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1は、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、その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を示している。

40

【0015】

50

コンピュータ 110 は、その他の取外し可 / 不可の揮発性 / 不揮発性コンピュータ記憶媒体を含む場合もある。例にすぎないが、図 1 には、取外し不可の不揮発性磁気媒体との読取りまたは書込みを行うハードディスクドライブ 141、取外し可の不揮発性磁気ディスク 152 との読取りまたは書込みを行う磁気ディスクドライブ 151、および CD-R OM またはその他の光媒体などの取外し可の不揮発性光ディスク 156 との読取りまたは書込みを行う光ディスクドライブ 155 を示している。例示的な動作環境で使用できるその他の取外し可 / 不可の揮発性 / 不揮発性コンピュータ記憶媒体には、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、半導体 RAM、半導体 ROM などが含まれるが、これらに限定されない。ハードディスクドライブ 141 は、一般に、インターフェース 140 などの取外し不可の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、一般に、インターフェース 150 などの取外し可の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。

10

【0016】

上記に説明し、図 1 に示したドライブおよびそれらの関連す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およびその他のデータの格納をコンピュータ 110 に提供している。図 1 では、例えば、ハードディスクドライブ 141 を、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、その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 を格納しているものとして示してある。これらのコンポーネントは、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、その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 と同じ場合も異なる場合もあることに注意されたい。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、その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 には、少なくともそれらが異なるコピーであることを示すために、ここでは異なる番号を付与してある。

20

【0017】

ユーザは、キーボード 162、マイクロフォン 163、およびマウス、トラックボール、タッチパッドなどのポインティングデバイス 161 などの入力デバイスを介して、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。その他の入力デバイス（図示せず）には、ジョイスティック、ゲームパッド、衛星アンテナ、スキャナなどが含まれる。これらおよびその他の入力デバイスは、システムバスに結合されているユーザ入力インターフェース 160 を介して処理装置 120 に接続されることが多いが、パラレルポート、ゲームポート、またはユニバーサルシリアルバス（USB）など、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もできる。モニタ 191 またはその他のタイプの表示デバイスも、ビデオインターフェース 190 など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、コンピュータはスピーカ 197 やプリンタ 196 などの他の周辺出力デバイスを含むこともでき、それらは出力周辺インターフェース 195 を介して接続することができる。

30

【0018】

コンピュータ 110 は、リモートコンピュータ 180 などの 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操作される。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、またはその他の一般的なネットワークノードであってよく、一般に、コンピュータ 110 に関連して上記に説明した構成要素の多くまたは全てを含む。図 1 に示す論理接続には、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 が含まれるが、他のネットワークを含むこともある。このようなネットワーク化環境は、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においてはごく普通である。

40

【0019】

LAN ネットワーク化環境において使用する場合、コンピュータ 110 を、ネットワー

50

クインターフェースまたはアダプタ 170 を介して LAN 171 に接続する。WAN ネットワーク化環境において使用する場合、コンピュータ 110 は、一般に、インターネットなどの WAN 173 を介して通信を確立するためのモデム 172 またはその他の手段を含む。モデム 172 は、内蔵または外付けであってよく、ユーザ入力インターフェース 160 またはその他の適切なメカニズムを介してシステムバス 121 に接続することができる。ネットワーク化環境においては、コンピュータ 110 に関連して説明した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記憶装置に記憶することができる。限定ではなく、例として、図 1 は、リモートアプリケーションプログラム 185 をリモートコンピュータ 180 に常駐するものとして示している。図示したネットワーク接続は例示的であり、コンピュータ間の通信リンクを確立するためのその他の手段を使用することができる。 10

【0020】

本発明は、図 1 を参照して説明したようなコンピュータシステム上で実行することができることに注意されたい。しかし、本発明は、サーバ、メッセージ処理専用のコンピュータ、または本発明の異なる部分が分散コンピューティングシステムの異なる部分で実行される分散システム上で実行することも可能である。

【0021】

図 2 は、本発明を実施することができる別の例示の適切なコンピューティング環境であるモバイルデバイス 200 のブロック図である。コンピューティングシステム環境 200 は適切なコンピューティング環境の別の例にすぎず、本発明の使用または機能の範囲についていかなる限定を示唆するものではない。また、コンピューティングシステム環境 200 を、図示したコンポーネントのいずれか 1 つまたはそれらの組合せに関して依存性、または必要性を有するものとして解釈すべきでもない。 20

【0022】

モバイルデバイス 200 は、マイクロプロセッサ 202、メモリ 204、入出力 (I/O) コンポーネント 206、およびリモートコンピュータまたはその他のモバイルデバイスと通信を行うための通信インターフェース 208 を含んでできる。一実施形態では、これらのコンポーネントは、適切なバス 210 を介して互いに通信を行うために結合されている。

【0023】

メモリ 204 は、モバイルデバイス 200 への全体的な電力が遮断されたときにメモリ 204 に格納された情報が失われないように、バッテリーのバックアップモジュール (図示せず) を備えたランダムアクセスメモリ (RAM) などの不揮発性電子メモリとして実装されている。メモリ 204 の一部は、好ましくは、プログラム実行のためのアドレス可能メモリとして割り振られ、メモリ 204 の別の部分は、好ましくは、ディスクドライブ上の記憶をシミュレートするためなどの記憶として使用されている。 30

【0024】

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 を含んでいる。動作中、オペレーティングシステム 212 は、メモリ 204 からプロセッサ 202 によって実行されることが好ましい。オペレーティングシステム 212 は、好ましい一実施形態では、マイクロソフトコーポレーションから市販されている WINDOWS (登録商標) CE ブランドである。オペレーティングシステム 212 は、モバイルデバイス用に設計されていることが好ましく、1 組の公開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メソッドを介してアプリケーションプログラム 214 によって利用することができるデータベース機能を実装している。オブジェクトストア 216 中のオブジェクトは、少なくとも部分的には、公開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メソッドへのコールに回答して、アプリケーションプログラム 214 およびオペレーティングシステム 212 によって維持管理されている。 40

【0025】

通信インターフェース 208 は、モバイルデバイス 200 が情報を送受信できるように 50

する数多くのデバイスおよび技術を代表している。これらのデバイスには、2、3例を挙げると、有線および無線モデム、衛星受信機、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して、データ交換することもできる。このような場合、通信インターフェース208は、赤外線トランシーバや、シリアルまたはパラレル通信接続とすることができ、それらの全てがストリーミング情報を伝送することができる。

【0026】

入出力コンポーネント206は、タッチ検知画面、ボタン、ローラ、マイクロフォンなどの様々な入力デバイス、ならびにオーディオジェネレータ、振動デバイス、ディスプレイなどの様々な出力デバイスを含む。上記のデバイスは例であり、それらが全てモバイルデバイス200上に存在する必要はない。さらに、本発明の範囲内で、その他の入出力デバイスをモバイルデバイス200に取り付けるか、または見つけることができる。

10

【0027】

II. 適応型機械翻訳サービスの概要

図3は、信頼できるソースから訂正済みの翻訳を取得する通常のワークフロー内での適応型機械翻訳を説明する概略流れ図である。

【0028】

統計的および用例に基づく技術を含む様々な機械学習技術を通して自動機械翻訳システムのカスタマイズを自動化するための研究がなされている。このような技術では、機械翻訳システムは、すでに翻訳済みのマテリアル(バイテキスト(b i t e x t)またはバイリンガルコーパス(b i l i n g u a l c o r p u s)と呼ばれることが多い)から翻訳対応物を学習することができる。このようなマテリアルには、1つの(ソース)言語によるセンテンスと、それに対応する別の言語による翻訳済み(ターゲット)センテンスが含まれている。さらに、このようなMT(m a c h i n e t r a n s l a t i o n)システムは、「比較」コーパス、またはソース言語とターゲット言語間の正確な翻訳ではないが、ソース言語とターゲット言語の両方で同様の概念およびイベントを記述しているテキストから、さらなる対応物を学習することができる。モノリンガルコーパスをさらに利用して、ターゲット言語で流暢な構文(c o n s t r u c t i o n)を学習することができる。本発明の一般的な一態様によれば、これらのカスタマイズ技術を、従来の文書管理環境内で適用し、利用することである。具体的には、システムユーザが文書を作成し、対応する翻訳を入手し、その翻訳を訂正するという通常の流れの中で、自動翻訳システムをトレーニングするためのデータを生成する。このトレーニングデータによって、ユーザの自動機械翻訳システムの体系的なカスタマイズが可能になる。

20

30

【0029】

図3を参照すると、本発明の実施形態は、ユーザが翻訳のためにユーザのコンピュータ上の(またはユーザに関連するサーバ上の)自動トランスレータにソース文書302を提出する、文書管理またはワークフロー環境内での適応型機械翻訳システムのカプセル化に関する。このアクションをブロック330で表してある。ソース文書302および自動生成された翻訳304が、検討および訂正のために、信頼できる修正ソース(すなわち、人間のトランスレータ)に伝送される。このアクションをブロック332で表してある。

40

【0030】

訂正済み翻訳306およびオリジナルのソース文書302が処理されて、更新され、正確であると考えられる翻訳対応物308の集まりが作成される。このアクションをブロック334で表してある。一実施形態によれば、対応物308は、ユーザによって維持管理されている自己カスタマイズ機械翻訳システムと並行して稼動している自己カスタマイズ機械翻訳システムによって生成される。一実施形態によれば、更新された翻訳対応物308は、更新済みデータベースの中に置かれ(あるいは、統計的な機械翻訳システムが使用されている場合には、統計パラメータの更新済みテーブルに反映される)、このデータベースは訂正済みの翻訳された文書とともにユーザに返送される。更新物は、ユーザの自動機械翻訳システムに同化される。ユーザが次に同様のテキストのマテリアル310の翻訳

50

を試みると、システムは、以前に訂正済みの文書と共に返送された更新物に基づいて、より高品質の翻訳 3 1 2 を自動的に生成する。このアクションをブロック 3 3 6 で表してある。このトレーニング、および本明細書に記載した全ての同様のトレーニングが、例えば、言語ペアの両方向（すなわち、スペイン語から英語、英語からスペイン語）でその後の翻訳に利益をもたらすことに注意されたい。

【 0 0 3 1 】

訂正済みの翻訳 3 0 6 およびソース文書 3 0 2 に基づいて多くの異なるタイプのトレーニングデータを生成できることに注意されたい。多くの異なるタイプのトレーニングデータを利用して、ユーザの自動翻訳システムを適応化することができる。翻訳対応物の更新は、本発明の範囲に含まれる一例にすぎない。任意の知識ソースの更新が範囲内に含まれる。任意の統計または用例ベースのトレーナのいかなる更新も範囲内に含まれる。以下に、具体例を詳細に説明する。

10

【 0 0 3 2 】

ユーザが様々な文書の自動翻訳を取得し、その結果を信頼できる事後校訂（すなわち、訂正および修正）に出すと、ユーザの自動翻訳システムは、同様の文書をより有効に翻訳できるようにそれ自体を徐々に適応化させる。コストのかかるカスタマイズの必要がなくなり、ユーザは、その後、より高品質の自動翻訳を享受する。ユーザの自動翻訳システムの適応化およびカスタマイズは、例えば、ユーザが品質の高い翻訳を取得する通常のルーチンに専念している間に「舞台裏で」行われる。

【 0 0 3 3 】

一実施形態によれば、自動生成された翻訳 3 0 4 は、翻訳全体および/またはその一部の品質を示す自動生成された信頼メトリック（*confidence metric*）を含む。信頼メトリックは、例えば、ユーザの見積もった出力の満足度に基づく。このような信頼メトリックの生成および利用については、本発明と同じ実体に譲渡され、参照によりその全体を本明細書に組み込む2002年12月4日に出願の「SYSTEM AND METHOD FOR MACHINE LEARNING A CONFIDENCE METRIC FOR MACHINE TRANSLATION」と題する米国特許出願（特許文献1）に記載されている。

20

【 0 0 3 4 】

図4は、記載した自己カスタマイズ機械翻訳システムに信頼メトリックがどのように組み込まれるかを示す流れ図である。ブロック402に従って、ユーザはソース文書の自動翻訳を得る。この文書は、その文書の全体および/またはその個々の部分の1つまたは複数に関する注記された信頼メトリック情報を含む。ブロック404に従って、ユーザは、事後校訂のために信頼の格付けが低い1つまたは複数の部分を選択する。これらの部分は、訂正のために信頼できる修正ソース（すなわち、人間のトランスレータ）に転送される。訂正された部分はオリジナルのソース文書と共に処理されて、更新済みの正確であると考えられる翻訳対応物が作成される。一実施形態によれば、この処理は、ユーザによって維持管理されている自己カスタマイズ機械翻訳システムと並行して稼動している自己カスタマイズ機械翻訳システムによって行われる。

30

【 0 0 3 5 】

ブロック406に従って、更新された翻訳対応物が、訂正された翻訳済みの部分（または訂正された翻訳済みの文書全体）と共にユーザに返送される。ブロック408に従って、更新物がユーザの自動機械翻訳システム中に同化される。ユーザが次に同様のテキストのマテリアルの翻訳を試みると、ユーザの自動機械翻訳システムは、より高品質の翻訳を生成する。

40

【 0 0 3 6 】

III. 特定の応用例

図5Aおよび5Bは、適応型機械翻訳システムの上述の実施形態の特定の応用例のブロック図である。これらの特定の応用例は例にすぎず、本発明の使用または機能の範囲についていかなる限定を示唆するものではない。また、これらの特定の応用例を、図示したコンポーネントのいずれか1つまたはそれらの組合せに関するいかなる依存性、または必要

50

性を有するものとして解釈すべきでもない。

【0037】

図5Aは、コンピューティング環境500のブロック図である。ユーザ502は、コンピューティングデバイス504を操作して、コンピュータネットワーク505（すなわち、インターネット）を介して、信頼できる修正ソース506とやりとりすることができる。ソース506は、例えば、コンピューティングデバイス上で実施された翻訳サービスであり、ネットワーク505を介してコンピューティングデバイス504およびそのユーザ502に提供される。

【0038】

コンピューティングデバイス504、ならびに修正ソース506が実装されているコンピューティングデバイスは、様々な既知のコンピューティングデバイスのいずれかであってよく、これらには図1および2に関連して説明したコンピューティングデバイスのいずれかが含まれるが、これらに限定されない。ネットワーク505を介したコンピューティングデバイス504と修正ソース506との間の通信は、様々な既知のネットワーク通信方法のいずれかを利用して行うことができ、これらには図1および2に関連して説明した方法のいずれかが含まれるが、これらに限定されない。一実施形態によれば、コンピューティングデバイス504は、サーバに実装される修正ソース506との無線ネットワークでの通信のために構成されたクライアント無線モバイルデバイスである。別の実施形態によれば、コンピューティングデバイス504は、サーバに実装実施される修正ソース506とのインターネットでの通信のために構成されたクライアントパーソナルコンピュータである。これらは、本発明の範囲に含まれる多くの特定の実施形態のうちの2つにすぎない。

【0039】

コンピューティングデバイス504は、自動翻訳システム508を含む。ユーザ502は、例えば、対応する自動翻訳の生成のために、テキストサンプルをシステム508に提出する。ユーザ502が、翻訳システム508によって生成された翻訳の1つまたは複数部分に満足していない（すなわち、ユーザが指示された低信頼メトリックスに満足していない）と想定し、自動翻訳が、ソース文書のコピーと共に修正ソース506に提出される。この自動翻訳はソース506で訂正される。一実施形態によれば、人間のトランスレータ510が自動翻訳を訂正する。別の実施形態によれば、信頼できる自動化システムが訂正を行う。訂正済みの翻訳は、ユーザ502への引き渡しのためにコンピューティングデバイス504に返送される。

【0040】

自動翻訳システム508を適応化させるために利用できるトレーニングデータの集まりを生成するために、トレーニングジェネレータ512を利用して、自動翻訳、訂正済みの翻訳、および/またはソース文書进行处理する。トレーニングジェネレータ512は、修正ソース506、またはコンピューティングデバイス504、または別個のアクセス可能な独立したロケーション（すなわち、独立したアクセス可能なサーバ）に記憶されたコンポーネントである。トレーニングジェネレータ512が修正ソース506に格納されている場合、生成されたトレーニング情報は、例えば、関連する訂正済み翻訳と共に自動翻訳システム508に転送される。トレーニングジェネレータ512がコンピューティングデバイス504に格納されている場合、情報は直接システム508中に実装される。トレーニングジェネレータ512を修正ソース506と共に格納することによって、コンピューティングデバイス504に課される記憶および処理要件が低減される。また、この構成によれば、トレーニングジェネレータ512を集中ロケーションから維持管理し、操作することが可能になる。

【0041】

一実施形態によれば、自動翻訳システム508の適応化を容易にするために、トレーニングジェネレータ512が信頼できる修正ソース506とコンピューティングデバイス508の両方にある。トレーニングジェネレータ512のペアは、例えば、同じか、または

ほぼ同様である。トレーニングジェネレータ 512 のペアは、例えば、自己カスタマイズ機械翻訳システム（このようなシステムについては図 6 に関連して説明する）に関連付けられている。修正ソース 506 で事後校訂が完了した後、生成された訂正済み翻訳は、オリジナルのソーステキストと共に、例えば、修正ソース 506 上で実施された自己カスタマイズ機械翻訳システムの「トレーニング」フェーズによって処理される。このトレーニングフェーズの間に正しい翻訳対応物が学習される。この対応物は、更新済みデータベースの中に置かれ（あるいは、統計的なシステムが使用されている場合には、統計パラメータの更新済みテーブルに反映される）、これはコンピューティングデバイス 504 上で実施された機械翻訳システムのバージョンに送られる。次いで、更新物は、ユーザのコンピュータ上の自己カスタマイズシステムのバージョン（または、以下に説明するように、サーバ上に維持管理されているバージョン）に自動的に同化される。ユーザが次に同様のテキストの MATERIAL の翻訳を試みると、その翻訳システムは、以前に訂正済みの文書と共に返送された更新物に基づいて、より高品質の翻訳を自動生成する。

10

【0042】

一実施形態によれば、信頼できる修正ソース 506 は、ネットワーク 505 上で動作しているサーバに関連付けられる。トレーニングジェネレータ 512 は、同じサーバ上で維持管理され、操作される。修正ソース 506 に関連してユーザ 502 に提供された翻訳およびトレーニング情報は、例えば、必須ではないが、支払いベース（すなわち、時間毎または申し込みベースによる支払い）で提供される。

【0043】

20

図 5B は、コンピューティング環境 520 のブロック図である。図 5A の構成要素と同じ、または同様である図 5B の構成要素には、同じまたは同様の参照番号を付けてある。図 5B では、1 人または複数のユーザ 502 が、サーバ 524 に接続可能な 1 台または複数のコンピューティングデバイスとやりとりしている。自動翻訳システム 508 は、例えばユーザ 502 に関連付けられており、サーバ 524 上に格納され、維持管理されている。サーバ 524 は、ネットワーク 505 に接続可能である。ユーザ 502 は、コンピューティングデバイス 522 を操作して、ネットワーク 505 に同様に接続可能な信頼できる修正ソース 506 とのやりとりを可能にしている。修正ソース 506 は、例えば、ネットワーク 505 を介して、コンピューティングデバイス 504 経由でユーザ 502 に提供される翻訳サービスである。

30

【0044】

システム 520 は、システム 500 と同じ方法で動作するが、しかし、自動翻訳システム 508 は、可能性として、複数のコンピューティングデバイスによってアクセスされて、1 人または複数の個々のユーザ 502 のための自動翻訳を行うことができる。したがって、翻訳システム 508 は、複数のユーザによって提出された文書に関連したトレーニング情報を用いて適応および更新することができる。翻訳システム 508 の翻訳の正確さは、複数のユーザ 502 に対応するように進化していくことになる。これは、複数のユーザに単一の領域または分野の対象の範囲内で文書を生成し、翻訳させる共通のコネクションを持っている場合（すなわち、彼らが同じ業界、同じ会社などで働いている場合）には特に望ましい。

40

【0045】

IV. 自動カスタマイズを利用する機械翻訳システムを用いた特定の応用例

ここまでは自動翻訳システム 508 について一般的に説明してきた。システム 508 の正確な詳細は、本発明にとって重要ではない。さらに、説明したトレーニングデータを翻訳システム 508 がどのように同化させるかについての正確なスキームについては述べていない。本発明は、トレーニングデータの任意の 1 つの特定のタイプに限定されるものではなく、また、データを同化させるための任意の 1 つの方法に限定されるものでもない。しかし、特定の自動翻訳システム、およびトレーニングデータを同化させるための対応するスキームについて、図 6 に関連して説明する。

【0046】

50

いくつかの自動翻訳システムは、システムをカスタマイズするための自動化技術を利用して、以前に未知であったボキャブラリに対する翻訳を適応させる（すなわち、専門領域についての翻訳を適応させる）ことが知られている。本発明の実施形態は、このような翻訳システムの関連においてうまく適用可能である。このようなシステムについては、参照によりその全体が本明細書に組み込まれる、本発明と同じ実体に譲渡された2001年7月5日出願の「SCALEABLE MACHINE TRANSLATION SYSTEM」と題された米国特許出願（特許文献2）に記載されている。組み込まれた参照文献に記載のシステムの部分について、図6に関連して説明する。

【0047】

図6に関連する自動翻訳システムを説明するのに先立って、論理形式（logical form）の簡単な説明が役立つ可能性がある。論理形式、ならびにそれらを生成するためのシステムおよび方法についての完全かつ詳細な説明を、Heidornらに対する1999年10月12日発行の「METHOD AND SYSTEM FOR COMPUTING SEMANTIC LOGICAL FORMS FROM SYNTAX TREES」と題された米国特許（特許文献3）に見ることができる。しかし、簡単に言えば、論理形式は、入力テキストに関して形態素および統語解析（morphological and syntactic analysis）を実行して、文法関係（grammatical relations）で増強された従来の句構造解析（phrase structure analyses）を生成することによって生成される。統語解析は、論理形式を導出するためにさらなる処理を受ける。その論理形式は、テキスト入力中の内容語間のラベル付けされた依存を記述するデータ構造である。論理形式は、特定の統語上の交替（例えば、能動／受動）を標準化（normalize）し、センテンス間の照応関係（intrasentential anaphora）および長距離依存（long distance dependencies）を解消する。論理形式はグラフとして表すことができ、これにより、論理形式の構成要素を直感的に理解するのに役立つ。しかし、当業者には理解されるように、コンピュータ可読媒体上に格納した場合、論理形式はグラフを表すものとして直ぐに理解されず、むしろ（依存）木として理解される場合がある。

【0048】

論理関係（logical relation）は、以下のような、方向関係タイプ（directional relation type）で結び付けられた2つの単語からなる。

Logical Subject, Logical Object,
Indirect Object;
Logical Nominative, Logical Complement, Logical Agent;
Co Agent, Beneficiary;
Modifier, Attribute, Sentence Modifier;
Prepositional Relationship;
Synonym, Equivalence, Apposition;
Hypernym, Classifier, Subclass;
Means, Purpose;
Operator, Modal, Aspect, Degree Modifier, Intensifier;
Focus, Topic;
Duration, Time;
Location, Property, Material, Manner, Measure, Color, Size;
Characteristic, Part;
Coordinate;
User, Possessor;

S o u r c e , G o a l , C a u s e , R e s u l t ;
および D o m a i n .

【 0 0 4 9 】

論理形式は、センテンスやその一部など、単一のテキスト入力を表す、関連した論理関係のデータ構造である。論理形式は、最低、1つの論理関係からなり、構造関係（すなわち、統語上および意味上の関係）、特に入力ストリング中の重要な語の間の項および/または修飾関係（ a r g u m e n t a n d / o r a d j u n c t r e l a t i o n ）を表現する。

【 0 0 5 0 】

統語解析から論理形式を構築する特定のコードは、例えば、機械翻訳システムが動作している様々なソース言語およびターゲット言語の間で共用される。この共用アーキテクチャは、異なる言語からの論理形式セグメントの位置合せのタスクを大きく簡略化する。と言うのは、2つの言語における一見識別可能な構造は、しばしば同様または同一の論理形式表現に倒れ込むためである。

【 0 0 5 1 】

こうした背景を心に留めて、図6は、本発明の一態様による機械翻訳システム600のアーキテクチャのブロック図である。システム600は、規則ベースおよび統計的技術を用例ベースのトランスファと組み合わせるデータ駆動型の機械翻訳システムである。このシステムは、データから直接、語彙および句の翻訳（ l e x i c a l a n d p h r a s a l t r a n s l a t i o n s ）の知識を学習することができる。システム600の

10

20

【 0 0 5 2 】

機械翻訳システム600は、バイリンガルの対応するテキストから翻訳する方法を自動的に学習するよう構成される。このシステムは、特定のテキストについて、そのセンテンスおよびそれに対応する人の翻訳を処理することによってカスタマイズすることができ、その結果、そのテキストと同様のマテリアルに対して、その後の翻訳がより高品質になる。機械翻訳システム600はまた、全体の翻訳および/またはその一部の品質を指す、組み込まれた信頼スコアをうまく提供するよう構成される。

【 0 0 5 3 】

システム600は、構文解析コンポーネント604および606、統計的単語関連付け学習コンポーネント608、論理形式（ L F ）位置合せコンポーネント610、語彙知識ベース（ L K B ）構築コンポーネント612、バイリンガル辞書614、辞書マージコンポーネント616、トランスファマッピングデータベース618、ならびに更新済みバイリンガル辞書620を含む。トレーニングおよび翻訳の実行時の間、システム600は、解析コンポーネント622、マッチングコンポーネント624、転送コンポーネント626および/または生成コンポーネント628を利用する。一実施形態によれば、構文解析コンポーネント604および解析コンポーネント622は同じコンポーネントであり、または少なくとも互いに同一である。

30

【 0 0 5 4 】

バイリンガルコーパスを使用して、システムをトレーニングする。バイリンガルコーパスは、位置合せされた翻訳済みセンテンス（例えば、スペイン語などのソースまたはターゲット言語による、人が作成した翻訳物と1対1で対応している、英語などの別のソースまたはターゲット言語のセンテンス）を含む。バイリンガルコーパスにおける翻訳「センテンス」は、実際の完全なセンテンスに限定されず、代わりにセンテンスセグメントの集まりとすることができることに注意されたい。トレーニングの間、センテンスは、位置合せされたバイリンガルコーパスからシステム600中に、ソースセンテンス630（翻訳すべきセンテンス）、またターゲットセンテンス632（ソースセンテンスの翻訳）として、提供される。構文解析コンポーネント604および606は、位置合せされたバイリンガルコーパスからのセンテンスを構文解析して、ソース論理形式634およびターゲッ

40

50

ト論理形式 6 3 6 を生成する。

【 0 0 5 5 】

構文解析の間、センテンス中の語は、標準化された語形式（見出語（lemma））に変換され、統計的単語関連付け学習コンポーネント 6 0 8 に提供することができる。学習コンポーネント 6 0 8 によって、単一語および複数語の両方の関連物について、それぞれの信頼できるセットが得られるまで繰り返し仮定を立て、採点される。統計的単語関連付け学習コンポーネント 6 0 8 は、学習した単一語翻訳ペア 6 3 8 ならびに複数語ペア 6 4 0 を出力する。

【 0 0 5 6 】

複数語ペア 6 4 0 は辞書マージコンポーネント 6 1 6 に提供される。そのコンポーネントを使用して追加エントリをバイリンガル辞書 6 1 4 に追加して、更新済みバイリンガル辞書 6 2 0 を形成する。新規のエントリは複数語ペア 6 4 0 を表す。

【 0 0 5 7 】

単一語翻訳ペア 6 3 8 は、ソース論理形式 6 3 4 およびターゲット論理形式 6 3 6 と共に論理形式位置合せコンポーネント 6 1 0 に提供される。簡単に言えば、コンポーネント 6 1 0 はまず、ソースおよびターゲットの論理形式 6 3 0 および 6 3 6 中のそれぞれのノード間に一時的な対応物を確立する。これは、バイリンガル語彙目録（例えばバイリンガル辞書）6 1 4 からの翻訳ペアを使用して行われる。このバイリンガル語彙目録を、統計的単語関連付け学習コンポーネント 6 0 8 からの単一および複数語翻訳ペア 6 3 8、6 4 0 で増強することができる。可能な対応物を確立した後、位置合せコンポーネント 6 1 0 は、語彙上および構造上の考慮の両方に従って論理形式ノードの位置合せを行い、語および/または論理形式のトランスファマッピング 6 4 2 を作成する。

【 0 0 5 8 】

基本的に、位置合せコンポーネント 6 1 0 は、バイリンガル辞書情報 6 1 4、および単一および複数語ペア 6 3 8、6 4 0 を使用して、論理形式間にリンクを引く。トランスファマッピングは、ソースおよびターゲット論理形式 6 3 4 および 6 3 6 中で見つけられる頻度に基づいて、任意選択としてフィルタされ、語彙知識ベース構築コンポーネント 6 1 2 に提供される。

【 0 0 5 9 】

フィルタリングは任意選択ではあるが、一例では、トランスファマッピングがトレーニングデータ中に少なくとも 2 回見られない場合には、トランスファマッピングデータベース 6 1 8 を構築するために使用されない。ただし、その他の任意の所望の頻度をフィルタとして使用することもできる。出現頻度以外にその他のフィルタリング技術を使用することもできる。例えば、トランスファマッピングが入力センテンスの完全な構文解析から形成されているかどうかに基づいて、また、トランスファマッピングを作成するために使用される論理形式が完全に位置合せされているかどうかに基づいて、トランスファマッピングをフィルタすることができる。

【 0 0 6 0 】

コンポーネント 6 1 2 は、トランスファマッピングデータベース 6 1 8 を構築し、このトランスファマッピングデータベースは、基本的にはある言語の語および/または論理形式を第 2 の言語の語および/または論理形式にリンクするトランスファマッピングを収容する。こうして作成されたトランスファマッピングデータベース 6 1 8 により、システム 6 0 0 が実行時翻訳のために構成される。翻訳実行時の間、翻訳されることになるソースセンテンス 6 5 0 が解析コンポーネント 6 2 2 に提供される。解析コンポーネント 6 2 2 はソースセンテンス 6 5 0 を受け取り、このソースセンテンス入力に基づいてソース論理形式 6 5 2 を作成する。

【 0 0 6 1 】

ソース論理形式 6 5 2 は、マッチングコンポーネント 6 2 4 に供給される。マッチングコンポーネント 6 2 4 は、リンクされた論理形式 6 5 4 を得るために、ソース論理形式 6 5 2 をトランスファマッピングデータベース 6 1 8 中の論理形式にマッチさせようと試み

る。複数のトランスファマッピングが、ソース論理形式 6 5 2 の（複数）部分にマッチすることがある。マッチングコンポーネント 6 2 4 は、マッチする見出語、品詞（*parts of speech*）、およびその他の特徴情報を有するマッチトランスファマッピングの最良のセットについて、データベース 6 1 8 の中をサーチする。このベストマッチのセットは、所定のメトリックに基づいて見つけれられる。例えば、より大きい（より具体的な）論理形式を有するトランスファマッピングが、より小さい（より一般的な）論理形式を有するトランスファマッピングよりも好ましい場合がある。等しいサイズの論理形式を有するマッピングの間で、マッチングコンポーネント 6 2 4 は、例えば、より頻度の高いマッピングを好むことがある。マッピングはまた、何ら互いにコンフリクトしないという条件で、ソース論理形式 6 5 2 の重複する部分にマッチすることもある。マッピングのあるセットが集合的に、代替のマッピングセットよりも多くの入力センテンスを対象とする場合、例えば、このセットの方が好ましい場合がある。

10

【0062】

マッチするトランスファマッピングのセットが見つかり、マッチングコンポーネント 6 2 4 は、ソース論理形式 6 5 2 中のノード上に、トランスファマッピングによって受け取られた対応するターゲットの語または論理形式セグメントのコピーヘリンクを作成して、リンクされた論理形式 6 5 4 を生成する。複数語マッピングについてのリンクは、対応するセグメントのルートノードをリンクし、次いでその複数語マッピングに加わっているその他のソースノードにアスタリスクをリンクすることによって表される。このようなマッピングの対応する個々のソースおよびターゲットノード間のサブリンクも、例えば、転送中に使用のために作成することができる。転送コンポーネント 6 2 6 は、マッチングコンポーネント 6 2 4 からリンクされた論理形式 6 5 4 を受け取り、ターゲット翻訳の基礎を形成することになるターゲット論理形式 6 5 6 を作成する。これは、ソース論理形式 6 5 2 のノード上のリンクによってポイントされたターゲット論理形式セグメントが組み合わされているリンクされた論理形式 6 5 4 のトップダウンの全検索を実行することによって行われる。複雑である可能性のある複数語マッピングのために論理形式セグメントを共に組み合わせる場合、マッチングコンポーネント 6 2 4 によって個々のノード間に設定されるサブリンクを使用して、修飾語句（*modifier*）などの正しい付加ポイント（*attachment point*）を決定する。必要に応じて、デフォルトの付加ポイントが使用される。

20

30

【0063】

適用可能なトランスファマッピングが見つからない場合、ソース論理形式 6 5 2 のノードおよびそれらの関係は、単にターゲット論理形式 6 5 6 にコピーされる。それでも、デフォルトの単一語翻訳をトランスファマッピングデータベース 6 1 8 中で見つけ、ターゲット論理形式 6 5 6 の中に挿入することができる。しかし、何も見つからない場合は、翻訳を、例えば更新済みバイリンガル辞書 6 2 0 から得ることができる。この辞書は、位置合せの期間に使用されたものである。

【0064】

生成コンポーネント 6 2 8 は、例えば、ターゲット論理形式 6 5 6 からターゲットストリング（または出力ターゲットセンテンス）6 5 8 にマップする規則ベースのアプリケーション非依存の生成コンポーネントである。生成コンポーネント 6 2 8 は、例えば、入力論理形式のソース言語に関する情報を何も持たず、転送コンポーネント 6 2 6 によって渡された情報だけで動作する。生成コンポーネント 6 2 8 はまた、例えば、この情報をモノリンガル（例えば、ターゲット言語の）辞書と共に使用して、ターゲットセンテンス 6 5 8 を生成する。このように、1つの汎用生成コンポーネント 6 2 8 があれば、各言語に十分である。

40

【0065】

このように、論理形式を異なる言語間でマッチさせることができるように、システム 6 0 0 が情報を様々な言語から共用される共通の論理形式に構文解析することがわかる。このシステムはまた、トランスファマッピングデータベースを構築するにあたって単純なフィ

50

ルタリング技術を利用して、雑音の多いデータ入力を取り扱うことができる。それゆえ、システム 600 は、多くのセンテンスペアを使用して、自動的にトレーニングすることができる。

【0066】

図 3、4、5 A および 5 B に記載した適応型自動翻訳システムに戻ると、記載のシステム 600 は、例えば、ユーザの適応型自動翻訳システム（すなわち、翻訳システム 508）として実装することができる。一実施形態によれば、システム 600 によって生成された翻訳の少なくとも一部が、訂正のために、例えば、信頼できる修正ソース（すなわち、ソース 506）に送られる（すなわち、ユーザが修正のために信頼メトリックが低い部分を選択する）。施された訂正に基づいてトレーニング情報が生成される（トレーニングジェネレータ 512 によって生成されるトレーニング情報）。システム 600 は、このトレーニングデータを受け取り、処理する。一実施形態によれば、システム 600 は、施された訂正に対応するバイリンガルコーパスを処理する。翻訳システム 600 のユーザは、その後の同様のテキストに対してより高品質の翻訳を得る。

10

【0067】

一実施形態によれば、ユーザの自動翻訳システムの適応化を容易にするために、システム 600 は、信頼できる修正ソースとユーザのコンピューティングデバイス（または関連サーバ）の両方に常駐する。システム 600 のこのペアは、例えば、互いに並行して稼動する。修正ソースで事後校訂が完了した後、生成された訂正済み翻訳は、オリジナルのソーステキストと共に、例えば、修正ソース上で実施されたシステム 600 のバージョンの「トレーニング」フェーズによって処理される。このトレーニングフェーズの間、正しい翻訳の対応物が学習される。次いで、対応（物）が更新済みデータベース中に入れられ、このデータベースはユーザのコンピューティングデバイス（または関連サーバ）上に実装されたシステム 600 のバージョンに送られる。更新物は、訂正済み翻訳と共に、または独立して送ることができる。更新物は、システム 600 のユーザ側のバージョンの中に自動的に同化される。ユーザが次に同様のテキストマテリアルの翻訳を試みると、ユーザのシステム 600 は、前に訂正済みの文書と共に返送された更新物に基づいて、より高品質の翻訳を自動的に生成する。

20

【0068】

トレーニング情報に基づくシステム 600 の更新は、様々な方法のいずれかによって行うことができ、どの特定の手法も本発明にとって重要ではない。システム 600 に提供されるトレーニングデータは、適応化を行うのに適した様々な異なる形態が可能である。前述したように、一実施形態によれば、トレーニングデータはバイリンガルコーパス（すなわち、図 6 のセンテンスペア 630 および 632）である。別の実施形態によれば、トレーニングジェネレータ（すなわち、図 5 A および 5 B のジェネレータ 512）は、施された訂正に基づいて、パーサ 604 および / またはパーサ 606 に対する更新物を生成し、システム 600 に供給する（すなわち、更新によって、将来、XY は X として扱われるべきであるなどと指示される）。別の実施形態によれば、トレーニングジェネレータは、翻訳システム 600 によって維持管理された単一語のペアについて施された変更に基づいて、更新物を生成する。別の実施形態によれば、トレーニングジェネレータは、施された訂正に基づいて、トランスファマッピングデータベース 618 のための更新物を生成する。別の実施形態によれば、トレーニングジェネレータは、施された訂正に基づいて、直接または間接的に、トランスファマッピングデータベース 618 を再構築する。いずれの知識ソースの更新も本発明の範囲に含まれる。

30

40

【0069】

M i n d N e t は、翻訳システム 600 に関連する論理形式の言語構造データベース（すなわち、トランスファマッピングデータベース 618）などの構造を説明するために、この業界で使用されている総称である。M i n d N e t という用語は、ワシントン州、レッドモンドのマイクロソフト社によって造られた。本発明の一実施形態によれば、信頼できる修正ソースによってなされた訂正に基づいてシステム 600 を適応化するためのトレ

50

ニング情報の利用には、MindNetの操作（すなわち、更新）が伴う。この更新のプロセスは、ユーザのシステム上で（またはユーザに関連するサーバ上で）、または修正ソースに関連するシステム上で遠隔的に行うことができる。

【0070】

図7は、MindNetが更新される本発明の一実施形態を示す流れ図である。ブロック702に従って、ユーザのMindNetが、翻訳およびオリジナルテキストと共に信頼できる修正ソース（すなわち、サーバ上で実施される）に送られる（すなわち、クライアントマシンから）。翻訳に必要な訂正が施された後（ブロック704）、MindNetを再構築して、訂正を反映させる（ブロック706）。次いで、再構築されたMindNetは、訂正済みの翻訳マテリアルと共にユーザに送られる（すなわち、クライアントマシンに返送される）（ブロック708）。ブロック710に従って、再構築されたMindNetが、ユーザの自動翻訳システム内に組み込まれる。更新されたMindNetは、その後の翻訳に利用される。ユーザの翻訳システムの記載した遠隔的な更新は、MindNet以外のデータ構造に関連して行うことができることに注意されたい。

10

【0071】

図8は、ユーザのマシンから離れることなく（または、ユーザの関連サーバから離れることなく）MindNetが更新される別の実施形態を説明する流れ図である。ブロック802に従って、信頼できる修正ソースは、ユーザから翻訳マテリアルと対応するオリジナルテキストを受け取る（ブロック802）。必要に応じて訂正が施され（ブロック802）、対応するMindNetの追補が編集される（ブロック804）。ブロック806に従って、クライアントは、訂正済み翻訳と共に、自身のMindNet中にロードし、編集すべき追補を受け取る（ブロック808）。ブロック810によって表された実施形態に従って、ユーザのMindNetは、所定数の追補が収集されるまでは更新されない。ユーザの翻訳システムの記載したローカルの更新は、MindNet以外のデータ構造に関連して行うことができることに注意されたい。

20

【0072】

一実施形態によれば、複数の追補がサーバ上、すなわち、信頼できる訂正が施されるサーバ上につなが合わされるか、または収集される。所定数の追補が収集されると、ユーザは自身のMindNetを、再構築し、返送してもらうためにサーバに送る。ユーザのMindNetを更新するためのその他のスキームも、本発明の範囲内に含まれる。

30

【0073】

本発明の別の態様によれば、記載の適応型機械翻訳プロセスは、ユーザおよび信頼できる修正ソースが同じ1つであるシステム内に実装することができる。図3のプロセスフローは、このような実施形態と一致している。言い換えれば、図3のフローは、例えば信頼できる修正ソースであるユーザが、ソース文書の少なくとも一部を翻訳のために自分自身のコンピュータ上（またはユーザに関連するサーバ上）の自動トランスレータに提出する文書管理またはワークフロー環境内に適応型機械翻訳システムがカプセル化されている本発明の実施形態を対象としている。次に、このような実施形態について、図3を参照して説明する。

【0074】

ソース文書302の少なくとも一部の提出を、ブロック330で表してある。ユーザは、例えば、ソース文書302に関連する言語に関して信頼できるトランスレータである。ソース文書302情報、ならびに対応する自動生成された翻訳304が、検討および訂正のためにユーザ/訂正者に提示される。このアクションをブロック332で表してある。

40

【0075】

訂正済み翻訳306およびオリジナルのソース文書302が処理されて、更新され、正確であると考えられる翻訳対応物308の集まりを作成する。このアクションをブロック334で表してある。一実施形態によれば、更新された翻訳対応物308は更新済みデータベース中に配置される（あるいは、統計的機械翻訳システムが使用されている場合には、統計パラメータの更新テーブルに反映される）。更新物は、ユーザの自動機械翻訳シス

50

テム中に同化される。ユーザが次に同様のテキストデータ 310 の翻訳を試みると、システムは、前に訂正済みの文書に基づいて生成された更新物に基づいて、より高品質の翻訳 312 を自動的に生成する。このアクションをブロック 336 で表してある。このトレーニングが、言語ペアの両方向（すなわち、スペイン語から英語、英語からスペイン語）においてその後の翻訳に利益をもたらすことに注意されたい。

【0076】

訂正済み翻訳 306 およびソース文書 302 に基づいて多くの異なるタイプのトレーニングデータを生成できることを強調しておきたい。多くの異なるタイプのトレーニングデータを利用して、ユーザの自動翻訳システムを適応化することができる。翻訳対応物を更新することは、本発明の範囲に含まれる一例にすぎない。どのような知識ソースの更新もこの範囲内に含まれる。どのような統計的または用例ベースのトレーナの更新もこの範囲内に含まれる。具体的な例は、他の実施形態に関連して上述している。

10

【0077】

本発明の別の態様によれば、記載の適応型機械翻訳プロセスは、信頼できる翻訳ソースであるユーザによって操作される専門化された翻訳ソフトウェアに関連して利用することができる。人間のトランスレータ（すなわち、職業翻訳者、アマチュア翻訳者など）が、専門化された翻訳ソフトウェアを採用して必要とされる翻訳作業の量を低減していることが知られている。専門ソフトウェアを利用する人間のトランスレータは、そのソフトウェアなしでも正確に翻訳するのに必要な知識を備えているのが普通である。このソフトウェアを利用して、単に、所与の文書を翻訳するのに必要なキーストロークの数を減らす。

20

【0078】

専門化された翻訳ソフトウェアの実施例には、翻訳すべきセンテンス（またはセンテンス群）（すなわち、翻訳される文書から取り出されたセンテンスまたはセンテンス群）を、以前の翻訳済みのセンテンス（またはセンテンス群）のデータベースと比較するように構成されるものがある。マッチが見つかり、そのマッチした翻訳が自動的に取り出すことができる。このような場合、ユーザは、マニュアルによる翻訳の負担がある程度軽減されることになる。

【0079】

ターゲットセンテンスに利用可能な正確なマッチがない場合、専門化された翻訳ソフトウェアの実施例には、似ているが同一ではないセンテンスである「ファジーマッチ」を取り出すよう構成されるものがある。ユーザは、ファジーマッチを拒否して、最初からセンテンスを翻訳することができ、あるいは、ファジーマッチを正しい形に修正することができる。多くの場合、ファジーマッチを修正した方が最初から翻訳するよりも作業が少なくて済む（すなわち、キーストロークが少ない）ことになる。

30

【0080】

専門化された翻訳ソフトウェアの実施例には、自動翻訳システムと協同して、これらに限定されないが、利用できる正確な翻訳またはファジー翻訳が利用できないソーステキストセンテンスなど、翻訳すべき特定のセンテンスに対して自動機械翻訳を提供するよう構成されるものがある。ユーザは、機械翻訳を拒否し、センテンスを最初から翻訳することができ、あるいは、機械翻訳を正しい形に修正することができる。多くの場合、機械翻訳を修正した方が最初から翻訳するよりも作業が少なくて済む（すなわち、キーストロークが少ない）ことになる。

40

【0081】

本発明の一態様によれば、記載の専門化された翻訳ソフトウェアのユーザは、実質的には信頼できる翻訳ソースである。したがって、ユーザがファジーまたは機械翻訳を訂正すると、訂正に対応する情報を利用して、ソフトウェアに関連する機械翻訳システムをトレーニングまたは更新することができる。このように、その後の翻訳のために翻訳システムの効率および精度が改善される。機械翻訳システムのトレーニングまたは更新は、本明細書に記載の方法またはその他の任意の方法と同様に達成することができる。

【0082】

50

図 9 は、専門化された翻訳ソフトウェアを含む、本発明の実施形態の一応用例のブロック図である。例示の応用例は、例にすぎず、本発明の使用または機能の範囲についていかなる限定を示唆するものではない。また、この特定の応用例を、図示したコンポーネントのいずれか 1 つまたはそれらの組合せに関していかなる依存性、または必要性を有するものとして解釈すべきでもない。

【0083】

図 9 を参照すると、ユーザ / 訂正者 902 は、専門化された翻訳システム 910 (すなわち、専門化された翻訳ソフトウェア)、自動翻訳システム 912、およびトレーニングジェネレータ 914 (すなわち、上述のトレーニングジェネレータ 512 と同一または同様のもの) が実施されたコンピューティングデバイス 904 とやりとりする。コンピューティングデバイス 904 は、様々な既知のコンピューティングデバイスのいずれかとすることができ、これには図 1 および 2 に関連して説明したいずれかのものが含まれるが、これらに限定されない。一実施形態によれば、コンピューティングデバイス 904 はパーソナルコンピュータである。

10

【0084】

ユーザ 902 は、ソース文書の翻訳に関連する作業の少なくとも一部を取り除くためにシステム 910 に依存しているトランスレータ (すなわち、職業翻訳者またはアマチュア翻訳者) である。専門化された翻訳システム 910 は、ソース文書の翻訳にあたってユーザ 902 を支援するよう構成された専門化された翻訳システムである。ユーザ 902 は、例えば、対応する翻訳を生成するにあたっての支援のために、ソース文書の少なくとも一部をシステム 910 に提出する。自動翻訳システム 912 は、提供されたテキストについて自動的に導出された機械翻訳を提供するよう構成されている。専門化された翻訳システム 910 は、解析中のソース文書テキストの自動翻訳を探索し、翻訳システム 912 から受け取るよう構成されている (すなわち、システム 910 は、正確な翻訳またはファジー翻訳のマッチを生成できない場合にシステム 912 に依存する)。

20

【0085】

専門化された翻訳システム 910 に関連する、以前に翻訳済みのセンテンスのどのデータベースも、システム 912 によって生成された自動翻訳に基づいて更新することができることに注意されたい (すなわち、自動翻訳は、潜在的に正確マッチまたはファジーマッチになる)。機械翻訳は、「オンデマンド」で (すなわち、ユーザの要求により) 提供することができることに注意されたい。あるいは、機械翻訳を前処理ステップの間に生成し、その他の以前に翻訳済みのセンテンスと共に格納する (すなわち、その他の潜在的な正確マッチまたはファジーマッチと共に格納する) ことができる。前処理ステップの間に、以前に翻訳済みのセンテンスのデータベースを、正確またはファジーマッチが存在しないセンテンスを用いて更新することができる。それゆえ、機械翻訳は「オンデマンド」で、または前もって提供することができる (次いで、その他の以前に翻訳済みのセンテンスと共に格納することができる)。

30

【0086】

ユーザ 902 が、翻訳システム 912 によって生成された翻訳の 1 つまたは複数の部分に満足していない (すなわち、ユーザが、指示された低信頼メトリックで満足していない) 場合、自動翻訳は、例えば、訂正のためにユーザ 902 に提示される (すなわち、ユーザ 902 が信頼できると考えられる修正ソースとなる)。訂正済み翻訳 922 は、例えば、訂正プロセスから結果として得られる。トレーニングジェネレータ 914 は、自動翻訳システム 912 を適応化させるために利用することができるトレーニングデータの集まりを生成するために、自動翻訳、訂正済み翻訳、および / またはソース文書进行处理のために利用される。トレーニングジェネレータ 914 は、コンピューティングデバイス 904 上に格納されるか、あるいは別個ではあるがアクセス可能な独立したコンピューティングロケーションに記憶された (すなわち、独立したアクセス可能なサーバ上に記憶された) コンポーネントである。トレーニングジェネレータ 914 が別個のコンピューティングロケーションに格納されている場合、生成されたトレーニング情報は、例えば、自動翻訳シ

40

50

ステム 912 に返送される。トレーニングジェネレータ 914 がコンピューティングデバイス 904 で格納されている場合、情報は直接、システム 912 中に実装される。トレーニングジェネレータ 914 をコンピューティングデバイス 904 で格納することにより、記憶および処理要件が低減される。自動翻訳システム 912 とトレーニングジェネレータ 914 との間のトレーニング関係は、例えば、自動翻訳システム 508 およびトレーニングジェネレータ 512 に関連して上記に説明した実施形態のいずれかと同様である。

【0087】

一実施形態によれば、2人以上のユーザ 902 がコンピューティングデバイス 904 および専門化された翻訳システム 910 とやりとりして、より高品質の翻訳を集合的に生成することができる。別の実施形態によれば、ユーザ 902 は、コンピューティングデバイス 904 に直接（図示のように）またはコンピュータネットワークを介してアクセスすることができる。別の実施形態によれば、ジェネレータ 914 によって生成されたトレーニング材料または更新材料は、更新システム 912 に利用されるだけでなく、少なくとも 1 つの追加自動機械翻訳システムを更新するためにコンピュータネットワークを介して転送することもできる。例えば、トレーニング材料または更新材料を、同化のために、単一の追加の自動機械翻訳システムに直接、転送することができる。しかし、その代わりとして、材料を集中サーバに転送して、その後、同化のために複数の機械翻訳システムに（すなわち、支払い申し込みベースで）配信することができる。あるいは、材料を集中サーバに転送して、その後、同化のために大規模組織（すなわち、企業）に関連する複数の機械翻訳システムに配信することができる。

10

20

【0088】

本発明について特定の実施形態を参照して説明してきたが、当業者は、本発明の趣旨および範囲から逸脱することなく、形態および詳細に変更を加えることができることを認識されよう。

【図面の簡単な説明】

【0089】

【図 1】本発明を実施することができる 1 つの例示的な環境を示すブロック図である。

【図 2】本発明を実施することができる別の例示的な環境を示すブロック図である。

【図 3】本発明に従って適応型機械翻訳サービスを例示する概略流れ図である。

【図 4】適応型機械翻訳サービスに照らして信頼メトリックスの利用を例示する流れ図である。

30

【図 5 A】本発明の実施形態の 1 つの具体的な応用例を示すブロック図である。

【図 5 B】本発明の実施形態の別の具体的な応用例を示すブロック図である。

【図 6 A】本発明を実施することができる機械翻訳アーキテクチャを示すブロック図である。

【図 6 B】本発明を実施することができる機械翻訳アーキテクチャを示すブロック図である。

【図 7】ユーザの翻訳システムが遠隔的に更新される一実施形態を示す流れ図である。

【図 8】ユーザの翻訳システムがローカルに更新される一実施形態を示す流れ図である。

【図 9】本発明の実施形態の別の具体的な応用例を示すブロック図である。

40

【符号の説明】

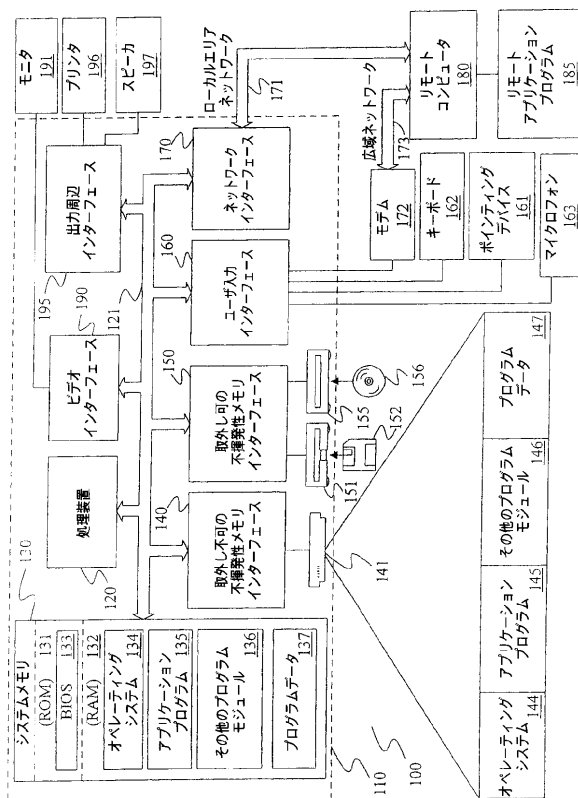
【0090】

- 100 コンピューティングシステム環境
- 200 モバイルデバイス
- 302 ソース文書
- 304 自動生成された翻訳
- 306 訂正済みの翻訳
- 308 更新された翻訳対応物
- 310 同様のソース文書
- 312 より高品質の翻訳

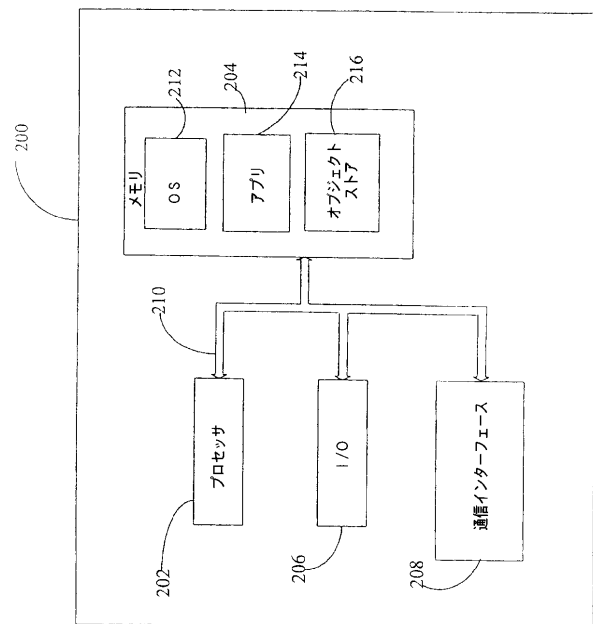
50

- 3 3 0 自動翻訳
- 3 3 2 信頼できる修正
- 3 3 4 正しい翻訳対応物进行处理
- 3 3 6 自動翻訳
- 5 0 0 コンピューティング環境
- 5 0 5 コンピュータネットワーク
- 5 2 0 コンピューティング環境
- 6 0 0 機械翻訳システム

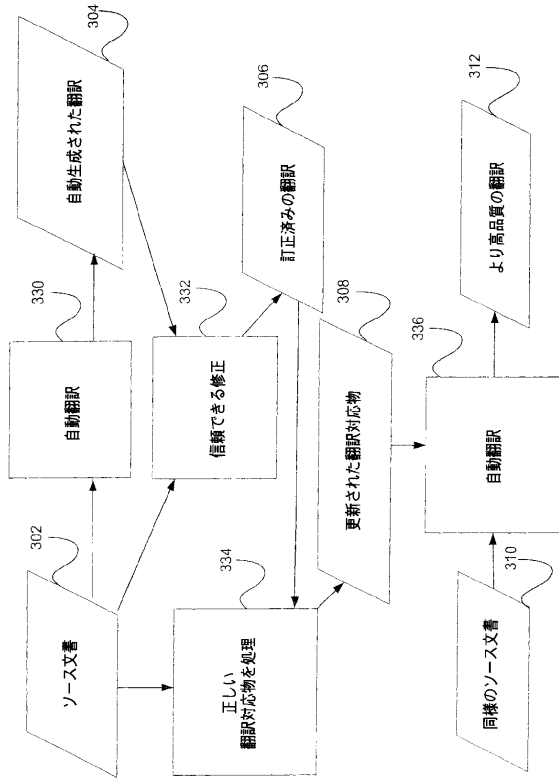
【図 1】



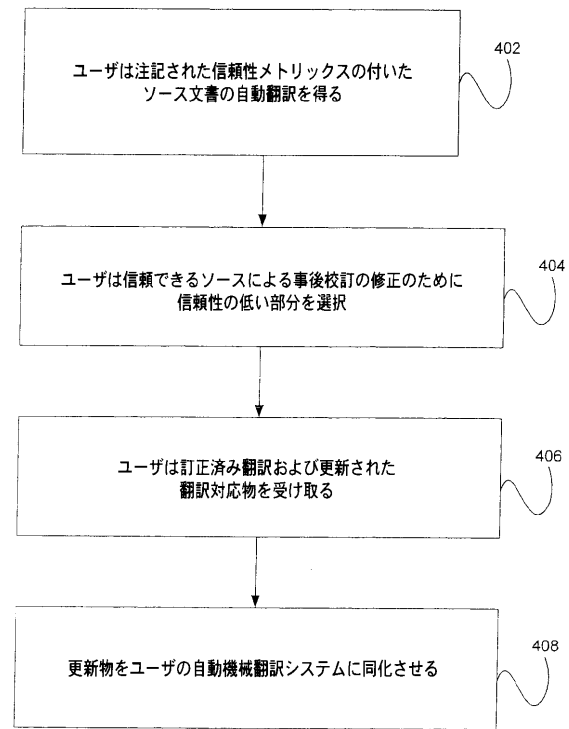
【図 2】



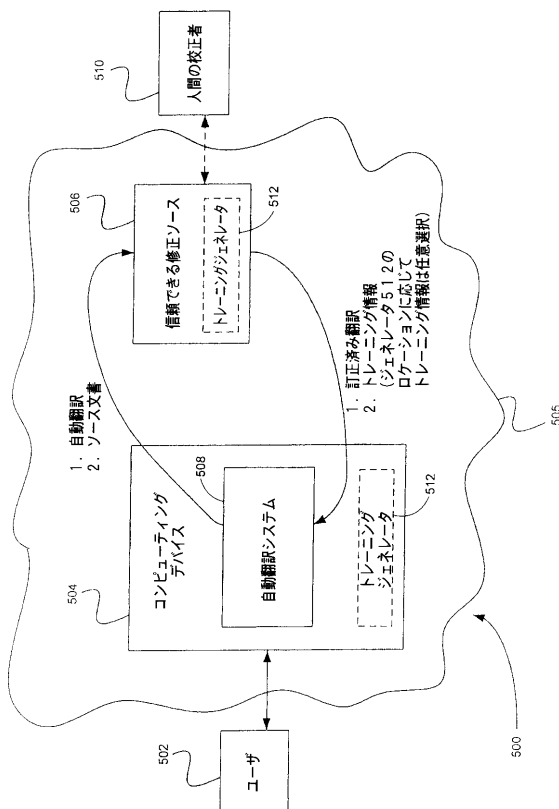
【図 3】



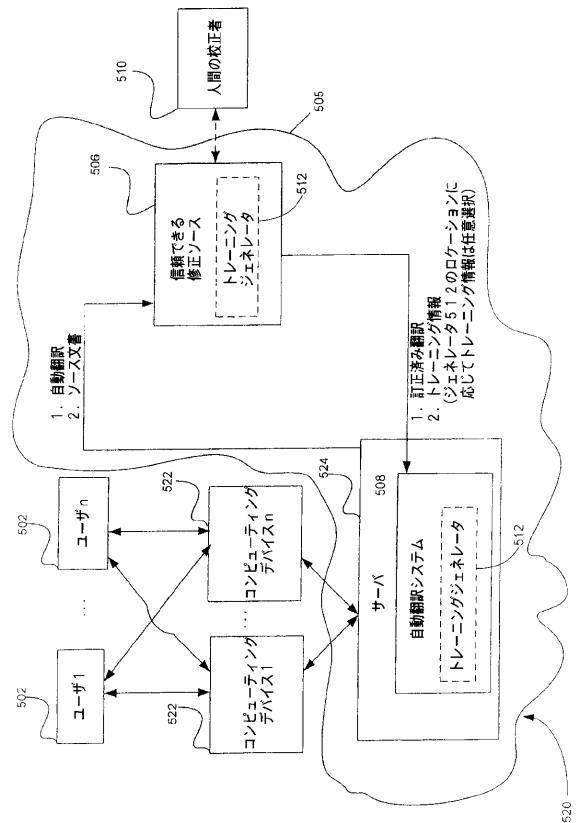
【図 4】



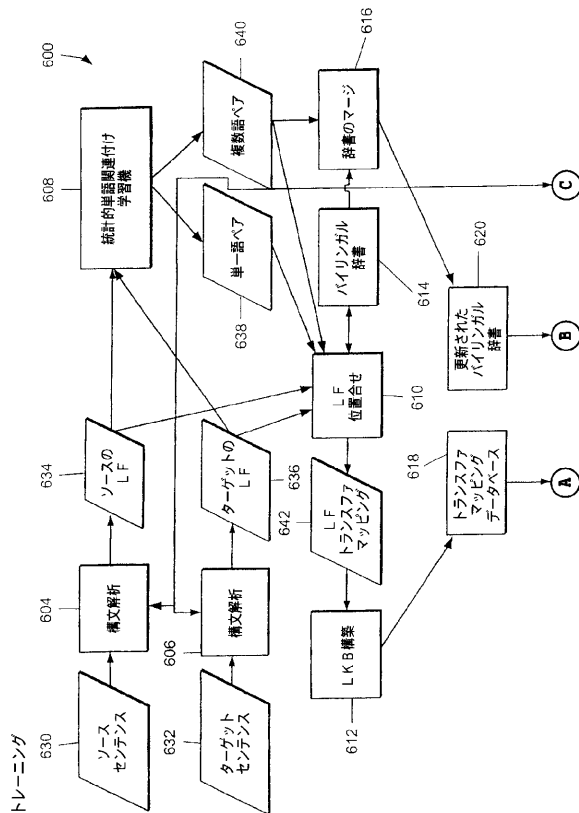
【図 5 A】



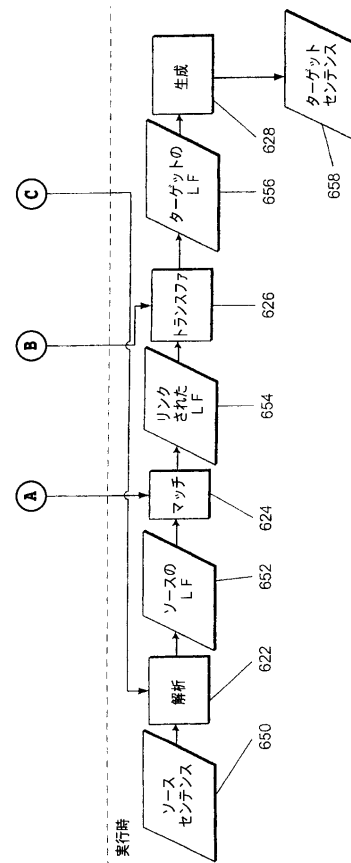
【図 5 B】



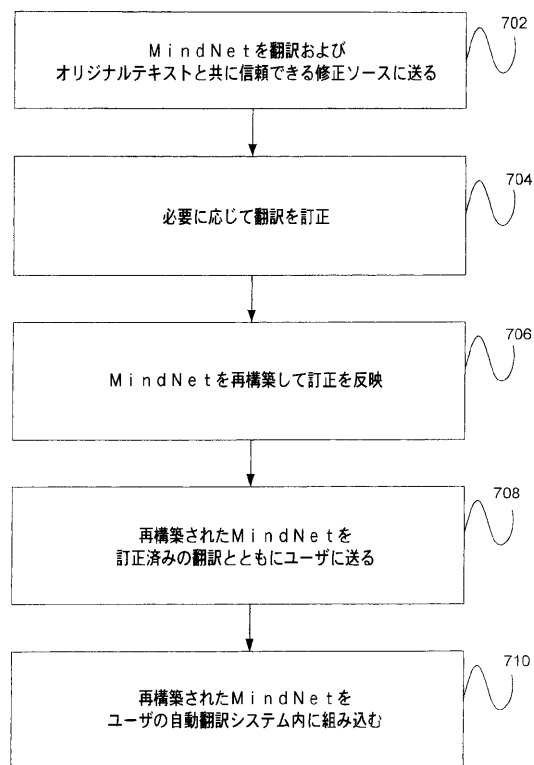
【図 6 A】



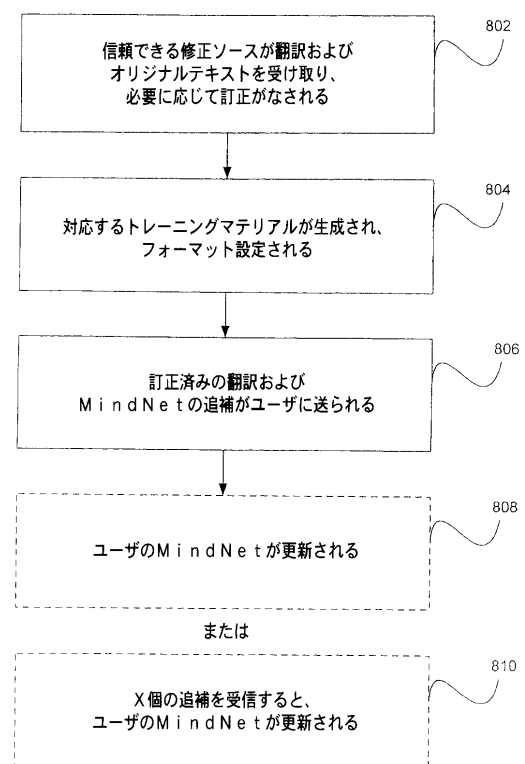
【図 6 B】



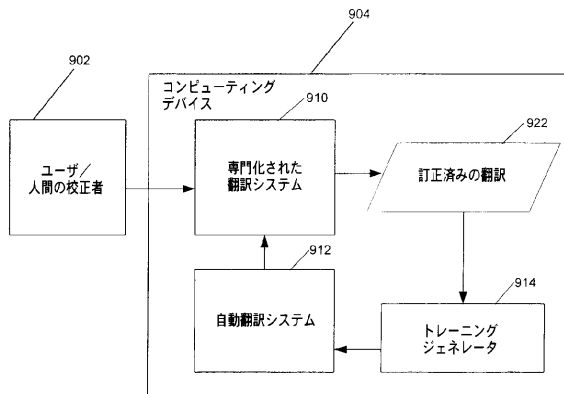
【図 7】



【図 8】



【 図 9 】



フロントページの続き

(72)発明者 リチャード エフ . ラシッド

アメリカ合衆国 9 8 0 5 3 ワシントン州 レッドモンド ノースイースト 8 4 ストリート
2 1 4 2 5

Fターム(参考) 5B091 AA03 CC03 CC16 DA04 DA09 EA01