

【公報種別】特許法第17条の2の規定による補正の掲載
 【部門区分】第6部門第3区分
 【発行日】平成19年1月11日(2007.1.11)

【公開番号】特開2004-171575(P2004-171575A)
 【公開日】平成16年6月17日(2004.6.17)
 【年通号数】公開・登録公報2004-023
 【出願番号】特願2003-389979(P2003-389979)
 【国際特許分類】

G 0 6 F 17/28 (2006.01)

【F I】

G 0 6 F 17/28 U

【手続補正書】

【提出日】平成18年11月20日(2006.11.20)

【手続補正1】

【補正対象書類名】特許請求の範囲

【補正対象項目名】全文

【補正方法】変更

【補正の内容】

【特許請求の範囲】

【請求項1】

コンピュータで実行されるプログラムによって実施されるソース言語における句とターゲット言語における句との間の翻訳関係を特定する方法であって、前記コンピュータは、CPUと1つ又は複数の記憶手段とを備え、前記CPUが記憶手段からプログラムを読み出して実行するとき、前記方法は、

記憶手段に格納されている複数語ユニットの整列されたペアに対するアクセスを前記CPUが受け付けるステップであって、前記複数語ユニットは、1つが前記ソース言語におけるソースユニットであり、もう1つが前記ターゲット言語におけるターゲットユニットであり、前記ソース言語の句が前記ソースユニットにおいて特定されている、アクセスを受け付けるステップと、

前記ソース言語の句の仮定された翻訳である、前記ターゲットユニットにおける少なくとも1つの候補句を前記CPUが記憶手段に生成するステップと、

各候補句に関して、前記ソース言語の句内部の語と前記候補句内部の語の間との関連に基づく内部成分と、前記ソース言語の句外部の語と前記候補句外部の語の間との関連に基づく外部成分とを含むスコアを、前記CPUが計算し記憶手段に格納するステップと、

前記CPUが、前記スコアに基づいて前記ソース言語の句と前記候補句との間の翻訳関係を特定し、当該特定した結果を記憶手段に出力するステップと

を備えることを特徴とする方法。

【請求項2】

ペアは、前記ソースユニットにおける語と、前記ターゲットユニットにおける語とを含み、前記ソースユニットにおける語と前記ターゲットユニットにおける語とのペアの間に関連の度合いを示す個々の語の関連スコアを有する

ことを特徴とする請求項1に記載の方法。

【請求項3】

前記スコアを計算するステップは、

前記ソース言語の句内部の語と前記候補句内部の語との間の関連の度合いを示す内部スコアを計算するステップと、

前記ソース言語の句外部の前記ソースユニットにおける語と前記候補句外部の前記ターゲットユニットにおける語との間の関連の度合いを示す外部スコアを計算するステップと

を含む

ことを特徴とする請求項 2 に記載の方法。

【請求項 4】

前記スコアを計算するステップは、

前記内部スコアと前記外部スコアとを結合して合併スコアを得るステップを含む

ことを特徴とする請求項 3 に記載の方法。

【請求項 5】

前記翻訳関係を特定するステップは、

前記合併スコアに基づいて前記ソース言語の句と前記候補句との間の翻訳関係を特定するステップを含む

ことを特徴とする請求項 4 に記載の方法。

【請求項 6】

前記少なくとも 1 つの候補句を生成するステップは、

前記個々の語の関連スコアに基づいて前記候補句を生成するステップを含む

ことを特徴とする請求項 1 に記載の方法。

【請求項 7】

前記ソースユニットにおける各語は、その各語に最も強く関連しているヌル語を含む、前記ターゲットユニットにおける語を示す語の関連を有し、前記ターゲットユニットにおける各語は、その各語に最も強く関連している、ヌル語を含む、前記ソースユニットにおける語を示す語の関連を有し、前記アクセスを受け取るステップは、

最も強く関連している語ペアの出現のカウントを生成するステップを含む

ことを特徴とする請求項 5 に記載の方法。

【請求項 8】

前記候補句における大文字使用パターンを検出するステップと、

前記大文字使用パターンに基づいて前記合併スコアを調整するステップと

をさらに備えることを特徴とする請求項 5 に記載の方法。

【請求項 9】

前記大文字使用パターンを検出するステップは、

前記候補句における第 1 の語が大文字で始まる第 1 のパターンと、

前記候補句における前記第 1 の語は大文字で始まらないが、前記候補句における 1 つまたは複数の後続の語が大文字で始まる第 2 のパターンと、

前記候補句におけるいずれの語も大文字で始まらない第 3 のパターンと

の少なくとも 1 つを検出するステップを含む

ことを特徴とする請求項 8 に記載の方法。

【請求項 10】

前記大文字使用パターンに基づいて前記合併スコアを調整するステップは、

前記第 1 のパターンが検出された場合に第 1 の大文字使用スコアを適用するステップと

、

前記第 2 のパターンが検出された場合に第 2 の大文字使用スコアを適用するステップと

、

前記第 3 のパターンが検出された場合に第 3 の大文字使用スコアを適用するステップと

を含むことを特徴とする請求項 9 に記載の方法。

【請求項 11】

前記内部スコアを計算するステップは、

前記ソース言語の句内部の各語に関して、前記語の関連スコアに基づき、その各語に最も緊密に関連している、ヌル語を含む、前記候補句内部の語を特定するステップと、

前記候補句内部の各語に関して、前記語の関連スコアに基づき、その各語に最も緊密に関連している、ヌル語を含む、前記ソース言語の句内部の語を特定するステップと

によって内部語ペアを特定するステップを含む

ことを特徴とする請求項 3 に記載の方法。

【請求項 1 2】

前記内部スコアを計算するステップは、

各内部語ペアに関して、前記内部語ペアにおける 1 つの語の出現が前記内部語ペアにおける別の語を最も強く関連している語として有する確率を示す内部語ペアの確率を生成するステップと、

前記内部語ペアの確率を結合するステップと

をさらに含むことを特徴とする請求項 1 1 に記載の方法。

【請求項 1 3】

前記外部スコアを計算するステップは、

前記ソース言語の句の外部の、前記ソースユニットにおける各語に関して、前記語の関連スコアに基づき、その各語に最も緊密に関連している、ヌル語を含む、前記候補句の外部の前記ターゲットユニットにおける語を特定するステップと、

前記候補句の外部の前記ターゲットユニットにおける各語に関して、前記語の関連スコアに基づき、その各語に最も緊密に関連している、ヌル語を含む、前記ソース言語の句の外部の前記ソースユニットにおける語を特定するステップと

によって外部語ペアを特定するステップを含む

ことを特徴とする請求項 1 2 に記載の方法。

【請求項 1 4】

前記外部スコアを計算するステップは、

各外部語ペアに関して、前記外部語ペアにおける 1 つの語の出現が前記外部語ペアにおける別の語を最も緊密に関連している語として有する確率を示す外部語ペアの確率を生成するステップと、

前記外部語ペアの確率を結合するステップと

をさらに含むことを特徴とする請求項 1 3 に記載の方法。

【請求項 1 5】

前記合併スコアを計算するステップは、

前記内部語ペアの確率と前記外部語ペアの確率とを結合するステップを含む

ことを特徴とする請求項 1 4 に記載の方法。

【請求項 1 6】

前記候補句を生成するステップは、

前記ソース言語の句における語に最も強く関連している前記ターゲットユニットにおけるターゲット言語の語を特定するステップと、

前記ソース言語の句における語が最も強く関連している前記ターゲットユニットにおけるターゲット言語の語を特定するステップと

を含むことを特徴とする請求項 1 に記載の方法。

【請求項 1 7】

前記候補句を生成するステップは、

特定されたターゲット言語の語で始まり、特定されたターゲット言語の語で終わる語のシーケンスに限定された前記ターゲットユニットにおける語のシーケンスとして候補句を生成するステップをさらに含む

ことを特徴とする請求項 1 6 に記載の方法。

【請求項 1 8】

前記候補句を生成するステップは、

大文字で始まるターゲット言語の語で始まり、特定されたターゲット言語の語で終わる候補句をさらに生成するステップをさらに含む

ことを特徴とする請求項 1 7 に記載の方法。

【請求項 1 9】

前記句の翻訳関係を特定するステップは、

候補句に関する前記合併スコアに基づき、前記コーパス全体にわたってどれだけ頻繁に候補句が前記ソース言語の句の翻訳として生成されているかに基づいて、各候補句に関連

する前記合併スコアを変更して新しい合併スコアを得るステップを含むことを特徴とする請求項5に記載の方法。

【請求項20】

複数語ソースユニットにおける特定されたソース言語の句に対する複数語ターゲットユニットにおける句の翻訳を特定するためのシステムであって、前記システムは、CPUと1つ又は複数の記憶手段とを備え、前記CPUが記憶手段からプログラムを読み出して実行するときに、前記プログラムは、

前記ソース言語の句内部の語とターゲット言語の句内部の語との間の語の関連、および前記ソース言語の句外部の語と前記ターゲット言語の句外部の語との間の語の関連に基づいて、1つまたは複数の候補句、および各候補句に関するスコアを前記CPUが生成するように構成された、個々の単語の関連モデルを備えたことを特徴とするシステム。

【請求項21】

前記ソースユニットと前記ターゲットユニットとは、整列されたコーパスの一部であり、前記プログラムは、

前記CPUが、前記コーパス全体にわたって前記ソース言語の句に対して生成された他の候補句に基づいて前記スコアを変更して、変更されたスコアを得るように構成されたクロスセンテンスモデル

をさらに備えたことを特徴とする請求項20に記載のシステム。

【請求項22】

前記プログラムは、

前記CPUが、前記ソース言語の句の翻訳としての前記候補句に関連する信頼性レベルを示す所望の信頼性メトリックに、変更された前記スコアを変換するように構成された変換モデル

をさらに備えたことを特徴とする請求項21に記載のシステム。

【請求項23】

コンピュータで実行されるプログラムによって実施されるソース言語の複数語ソースユニットにおける特定された句の仮定された翻訳としてターゲット言語の複数語ターゲットユニットにおける候補句を生成する方法であって、前記コンピュータは、CPUと1つ又は複数の記憶手段とを備え、前記CPUが記憶手段からプログラムを読み出して実行するときに、前記方法は、

記憶手段に格納されている前記ソース言語の句における語に最も強く関連している前記ターゲットユニットにおける第1のターゲット言語の語を前記CPUが特定するステップと、

記憶手段に格納されている前記ソース言語の句における語に最も強く関連している前記ターゲットユニットにおける第2のターゲット言語の語を前記CPUが特定するステップと、

第1のターゲット言語の語または第2のターゲット言語の語で始まり、第1のターゲット言語の語または第2のターゲット言語の語で終わる句として、前記候補句を前記CPUが記憶手段に生成するステップと

を備えることを特徴とする方法。

【請求項24】

前記候補句を生成するステップは、

大文字で始まる語で始まり、第1のターゲット言語の語または第2のターゲット言語の語で終わる句として、追加の候補句を生成するステップ

をさらに備えることを特徴とする請求項23に記載の方法。