

【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6月21日(2007.6.21)

【公開番号】特開2005-100335(P2005-100335A)
【公開日】平成17年4月14日(2005.4.14)
【年通号数】公開・登録公報2005-015
【出願番号】特願2004-151965(P2004-151965)
【国際特許分類】

G 0 6 F 17/28 (2006.01)

【F I】

G 0 6 F 17/28 Z

【手続補正書】

【提出日】平成19年5月7日(2007.5.7)

【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】

第1の言語の文と第2の言語の文とからなる対訳を複数個含む対訳コーパスとともに用いられる機械翻訳装置であって、

前記第1の言語の入力文を受け、前記入力文に対する前記第2の言語の翻訳文の初期候補を準備するための初期候補準備手段と、

前記初期候補に対して所定のオペレータを適用して得られる修正候補文を探索し、得られた修正候補文の中で統計的に予め定められる所定の条件を充足するものを前記入力文に対する翻訳文として出力するための候補修正手段とを含む、機械翻訳装置。

【請求項2】

前記初期候補準備手段は、前記第1の言語の入力文を受け、前記入力文との類似度に関連する所定の条件を充足する前記第1の言語の文を含む対訳文を、前記初期候補として前記対訳コーパス中から検索するための初期候補選択手段を含む、請求項1に記載の機械翻訳装置。

【請求項3】

前記初期候補選択手段が前記初期候補を選択できないことを検出するための検出手段と、

前記検出手段により前記初期候補選択手段による前記初期候補の選択ができないことが検出されたことに応答して、前記入力文を文よりも小さな部分単位に分割し、分割された部分単位ごとに前記対訳コーパスから検索された訳を組合せることにより、前記初期候補を準備するための手段とをさらに含む、請求項2に記載の機械翻訳装置。

【請求項4】

前記初期候補準備手段は、前記第1の言語の入力文を受け、前記入力文に対して所定の機械翻訳方式により機械翻訳を行なって前記初期候補を準備するための機械翻訳手段を含む、請求項1に記載の機械翻訳装置。

【請求項5】

前記初期候補準備手段は、前記第1の言語の入力文を受け、予め定める用例コーパスを用いて用例翻訳を行なうことにより前記入力文に対応する訳文を生成するための用例翻訳手段を含む、請求項1に記載の機械翻訳装置。

【請求項6】

前記用例コーパスは前記対訳コーパスである、請求項5に記載の機械翻訳装置。

【請求項 7】

前記初期候補準備手段は、前記初期候補を予め定める複数個準備する、請求項 1 ~ 請求項 6 のいずれかに記載の機械翻訳装置。

【請求項 8】

前記初期候補選択手段は、

前記対訳コーパスに含まれる前記第 1 の言語の各文と前記入力文との間で、前記対訳コーパスに含まれる前記第 1 の言語の文を文書とみなして定義される文書頻度を用いた所定の類似尺度を算出するための類似尺度算出手段と、

前記対訳コーパスに含まれる前記第 1 の言語の各文と前記入力文との間の編集距離を算出するための編集距離算出手段と、

前記類似尺度算出手段により算出された類似尺度、及び前記編集距離算出手段により算出された編集距離に基づいて定義されるスコアを算出し、当該スコアが所定の条件を充足する対訳文を前記初期候補として選択するためのスコア算出手段とを含む、請求項 2 に記載の機械翻訳装置。

【請求項 9】

前記類似尺度算出手段は、前記対訳コーパスに含まれる前記第 1 の言語の各文と前記入力文との間で、以下の式に従って tf/idf 基準 $P_{tf/idf}$ を算出するための手段を含み、

【数 1】

$$P_{tf/idf}(J_k, J_0) = \sum_{i: J_{0,i} \in J_k} \frac{\log(N/df(J_{0,i}))/\log N}{|J_0|}$$

ただし J_0 は入力文、 $J_{0,i}$ は入力文 J_0 の i 番目の単語、 $df(J_{0,i})$ は単語 $J_{0,i}$ に対する文書頻度、 J_k は k 番目の前記第 1 の言語の文 ($1 \leq k \leq N$)、 N は対訳コーパス内の全ての対訳文の数を、それぞれ示す、請求項 8 に記載の機械翻訳装置。

【請求項 10】

前記編集距離算出手段は、入力文 J_0 と前記第 1 の言語の文 J_k の間の DP (Dynamic Programming) マッチングを行なって編集距離 $dis(J_k, J_0)$ を算出するための手段を含み、編集距離 $dis(J_k, J_0)$ は以下の式

【数 2】

$$dis(J_k, J_0) = I(J_k, J_0) + D(J_k, J_0) + S(J_k, J_0)$$

ただし k は $1 \leq k \leq N$ なる整数、 $I(J_k, J_0)$ 、 $D(J_k, J_0)$ 、及び $S(J_k, J_0)$ はそれぞれ文 J_0 を文 J_k に変形させる際に必要な語の挿入、削除、及び置換の数、により定められる、請求項 9 に記載の機械翻訳装置。

【請求項 11】

前記スコア算出手段は、前記第 1 の言語の文 J_k に対し、前記類似尺度算出手段により算出された tf/idf 基準 $P_{tf/idf}$ 、及び前記編集距離算出手段により算出された編集距離 $dis(J_k, J_0)$ に基づいて次の式

【数 3】

$$score = \begin{cases} (1.0 - \alpha) \left(1.0 - \frac{dis(J_k, J_0)}{|J_0|}\right) + \alpha P_{tf/idf}(J_k, J_0) & (dis(J_k, J_0) > 0 \text{ のとき}) \\ 1.0 & (\text{その他のとき}) \end{cases}$$

ただし α は調整パラメータ、により定義されるスコアを算出するための手段と、

前記スコアを算出するための手段により算出されたスコアの大きなものから順番に予め定める個数の対訳文を前記初期候補として選択するための手段とを含む、請求項 10 に記載の機械翻訳装置。

【請求項 12】

前記初期候補選択手段により検索された対訳文の中に、前記スコアが 1 の前記第 1 の言語の文を持つ対訳文があるか否かを判定し、スコアが 1 の前記第 1 の言語の文を持つ対訳文があるときに、当該対訳文を前記訳文として選択するための手段をさらに含む、請求項 1

1 に記載の機械翻訳装置。

【請求項 1 3】

前記機械翻訳装置は、前記第 2 の言語の言語モデル、及び前記第 2 の言語から前記第 1 の言語への翻訳モデルに接続されて使用され、

前記候補修正手段は、

前記スコアが 1 の前記第 1 の言語の文を持つ対訳文がないと判定されたことに応答して、前記初期候補選択手段により選択された対訳文を第 1 の候補文として、当該第 1 の候補文に対して所定のオペレータを適用して得られる修正候補文を探索するための探索手段と

、
前記探索手段により探索された修正候補文の各々の尤度を前記言語モデル及び前記翻訳モデルに基づいて算出するための尤度算出手段と、

前記探索手段により探索される修正候補文の尤度に対し山登り法を適用して、最高の尤度を持つ修正候補文を選択するための手段とを含む、請求項 1 2 に記載の機械翻訳装置。

【請求項 1 4】

前記機械翻訳装置は、前記第 2 の言語の言語モデル、及び前記第 2 の言語から前記第 1 の言語への翻訳モデルに接続されて使用され、

前記候補修正手段は、

前記初期候補準備手段により準備された初期候補に対して所定のオペレータを適用して得られる修正候補文を探索するための探索手段と、

前記探索手段により探索された修正候補文の各々の尤度を前記言語モデル及び前記翻訳モデルに基づいて算出するための尤度算出手段と、

前記探索手段により探索された修正候補文の尤度に対し山登り法を適用して、最高の尤度を持つ修正候補文を選択するための手段とを含む、請求項 1 ~ 請求項 1 2 のいずれかに記載の機械翻訳装置。

【請求項 1 5】

コンピュータにより実行されると、当該コンピュータを請求項 1 ~ 請求項 1 4 のいずれかに記載の機械翻訳装置として動作させる、機械翻訳コンピュータプログラム。

【請求項 1 6】

請求項 1 5 に記載のコンピュータプログラムによりプログラムされたコンピュータ。

【手続補正 2】

【補正対象書類名】明細書

【補正対象項目名】0038

【補正方法】変更

【補正の内容】

【0038】

機械翻訳装置は、初期候補選択手段により検索された対訳文の中に、スコアが 1 の第 1 の言語の文を持つ対訳文があるか否かを判定し、スコアが 1 の第 1 の言語の文を持つ対訳文があるときに、当該スコアが 1 の対訳文を訳文として選択するための手段をさらに含んでもよい。

【手続補正 3】

【補正対象書類名】明細書

【補正対象項目名】0039

【補正方法】変更

【補正の内容】

【0039】

好ましくは、機械翻訳装置は、第 2 の言語の言語モデル、及び第 2 の言語から第 1 の言語への翻訳モデルに接続されて使用され、候補修正手段は、スコアが 1 の第 1 の言語の文を持つ対訳文がないと判定されたことに応答して、初期候補選択手段により選択された対訳文を第 1 の候補文として、当該第 1 の候補文に対して所定のオペレータを適用して得られる修正候補文を探索するための探索手段と、探索手段により探索された修正候補文の各

々の尤度を言語モデル及び翻訳モデルに基づいて算出するための尤度算出手段と、探索手段により探索される修正候補文の尤度に対し山登り法を適用して、最高の尤度を持つ修正候補文を選択するための手段とを含む。

【手続補正 4】

【補正対象書類名】明細書

【補正対象項目名】0048

【補正方法】変更

【補正の内容】

【0048】

【数 4】

$$P_{tf/idf}(J_k, J_0) = \sum_{i: J_{0,i} \in J_k} \frac{\log(N / df(J_{0,i})) / \log N}{|J_0|}$$

ただし J_0 は入力文、 $J_{0,i}$ は入力文 J_0 の i 番目の単語、 $df(J_{0,i})$ は入力文 J_0 の i 番目の単語 $J_{0,i}$ に対する文書頻度、 N は対訳コーパス 34 内の全ての対訳文の数を、それぞれ示す。文書頻度 $df(J_{0,i})$ とは、入力文 J_0 中の i 番目の単語 $J_{0,i}$ が出現する文書（本実施の形態では文）の数のことをいう。

【手続補正 5】

【補正対象書類名】明細書

【補正対象項目名】0049

【補正方法】変更

【補正の内容】

【0049】

初期候補選択部 32 はさらに、対訳コーパス 34 に含まれている各対訳文 (J_k, E_k) の第 1 の言語の文 J_k について、入力文 J_0 との間の DP (Dynamic Programming) マッチングを行なって編集距離 $dis(J_k, J_0)$ を算出するための編集距離算出部 52 と、 tf/idf 算出部 50 により算出された tf/idf 基準 $P_{tf/idf}$ に基づいて、後述する式に従って各文のスコアを算出するためのスコア算出部 54 とを含む。