

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4113814号
(P4113814)

(45) 発行日 平成20年7月9日(2008.7.9)

(24) 登録日 平成20年4月18日(2008.4.18)

(51) Int.Cl. F I
G06F 17/28 (2006.01) G O 6 F 17/28 U

請求項の数 12 (全 18 頁)

(21) 出願番号	特願2003-304965 (P2003-304965)	(73) 特許権者	393031586
(22) 出願日	平成15年8月28日 (2003. 8. 28)		株式会社国際電気通信基礎技術研究所
(62) 分割の表示	特願2000-331193 (P2000-331193) の分割	(74) 代理人	100086391
原出願日	平成12年10月30日 (2000.10.30)		弁理士 香山 秀幸
(65) 公開番号	特開2004-13913 (P2004-13913A)	(72) 発明者	菅谷 史昭
(43) 公開日	平成16年1月15日 (2004. 1. 15)		京都府相楽郡精華町光台二丁目2番地2
審査請求日	平成16年5月25日 (2004. 5. 25)	(72) 発明者	株式会社国際電気通信基礎技術研究所内
審査番号	不服2005-10902 (P2005-10902/J1)		安田 圭志
審査請求日	平成17年6月10日 (2005. 6. 10)	(72) 発明者	京都府相楽郡精華町光台二丁目2番地2
			株式会社国際電気通信基礎技術研究所内
		(72) 発明者	竹澤 寿幸
			京都府相楽郡精華町光台二丁目2番地2
			株式会社国際電気通信基礎技術研究所内

最終頁に続く

(54) 【発明の名称】 翻訳エージェントの翻訳能力評価装置

(57) 【特許請求の範囲】

【請求項1】

予め用意された複数の原言語テスト文を利用して翻訳エージェントの翻訳能力を評価する翻訳エージェントの翻訳能力評価装置であって、

各原言語テスト文に対応する正解翻訳文を記憶した第1記憶手段、

原言語テスト文とは異なる複数の原言語参照文を記憶した第2記憶手段、

各原言語参照文に対応する正解翻訳文を記憶した第3記憶手段、

各原言語テスト文毎に、原言語テスト文に対応する正解翻訳文と、第2記憶手段内の原言語参照文から検索された、原言語テスト文に類似する原言語参照文、に対応する正解翻訳文とを正解翻訳文候補とし、各正解翻訳文候補と翻訳エージェントによって得られた当該原言語テスト文に対する翻訳結果とを比較することにより、各原言語テスト文毎の翻訳正解率を算出する第1手段、ならびに

第1手段で算出された各原言語テスト文毎の翻訳正解率に基づいて翻訳エージェントの翻訳能力を評価する第2手段、

を備えていることを特徴とする翻訳エージェントの翻訳能力評価装置。

【請求項2】

第1手段は、

第2記憶手段内の原言語参照文から原言語テスト文に類似する原言語参照文を検索するa手段、

a手段で検索された1または複数の原言語参照文に対応する正解翻訳文と、当該原言語

テスト文に対応する正解翻訳文とを正解翻訳文候補とし、各正解翻訳文候補と翻訳エージェントによって得られた当該原言語テスト文に対する翻訳結果とを比較することにより、各正解翻訳文候補毎の翻訳正解率を算出する b 手段、および

b 手段で算出された各正解翻訳文候補毎の翻訳正解率のうち、最も翻訳正解率が高い翻訳正解率を、当該原言語テスト文に対する翻訳正解率として求める c 手段、

を備えていることを特徴とする請求項 1 に記載の翻訳エージェントの翻訳能力評価装置。

【請求項 3】

a 手段では、原言語テスト文と各原言語参照文との類似度を算出し、所定の類似度判定用閾値より類似度が高い原言語参照文を原言語テスト文に類似する原言語参照文として検索することを特徴とする請求項 2 に記載の翻訳エージェントの翻訳能力評価装置。

10

【請求項 4】

類似度は DP マッチングに基づいて算出されることを特徴とする請求項 3 に記載の翻訳エージェントの翻訳能力評価装置。

【請求項 5】

翻訳正解率は DP マッチングに基づいて算出されることを特徴とする請求項 1、2、3 および 4 のいずれかに記載の翻訳エージェントの翻訳能力評価装置。

【請求項 6】

第 2 手段は、第 1 手段で算出された各原言語テスト文毎の翻訳正解率の平均値を算出し、この算出結果に基づいて翻訳エージェントの翻訳能力を評価することを特徴とする請求項 1、2、3、4 および 5 のいずれかに記載の翻訳エージェントの翻訳能力評価装置。

20

【請求項 7】

第 2 手段は、第 1 手段で算出された各原言語テスト文毎の翻訳正解率と、所定のランク判別用域値とを比較することにより、翻訳エージェントによって得られた当該原言語テスト文に対する翻訳結果をランク分けし、ランク分け結果に基づいて翻訳エージェントの翻訳能力を評価することを特徴とする請求項 1、2、3、4 および 5 のいずれかに記載の翻訳エージェントの翻訳能力評価装置。

【請求項 8】

予め用意された複数の原言語テスト文を利用して翻訳エージェントの翻訳能力を評価する翻訳エージェントの翻訳能力評価装置であって、

30

各原言語テスト文に対応する正解翻訳文を記憶した第 1 記憶手段、

原言語テスト文とは異なる複数の原言語参照文を記憶した第 2 記憶手段、

各原言語参照文に対応する正解翻訳文を記憶した第 3 記憶手段、

各原言語テスト文に対応する翻訳能力別翻訳文を記憶した第 4 記憶手段、

各原言語テスト文毎に、原言語テスト文に対応する正解翻訳文と、第 2 記憶手段内の原言語参照文から検索された、原言語テスト文に類似する原言語参照文、に対応する正解翻訳文とを正解翻訳文候補とし、各正解翻訳文候補と翻訳エージェントによって得られた原言語テスト文に対する翻訳結果とを比較することにより、各原言語テスト文毎の第 1 翻訳正解率を算出する第 1 手段、

各原言語テスト文毎に、原言語テスト文に対応する正解翻訳文と、第 2 記憶手段内の原言語参照文から検索された、原言語テスト文に類似する原言語参照文、に対応する正解翻訳文とを正解翻訳文候補とし、各正解翻訳文候補と原言語テスト文に対応する翻訳能力別翻訳文とを比較することにより、各原言語テスト文毎の第 2 翻訳正解率を各翻訳能力別翻訳文セット毎に算出する第 2 手段、ならびに

40

第 1 手段で算出された各原言語テスト文毎の第 1 翻訳正解率と、第 2 手段で各翻訳能力別翻訳文セット毎に算出された各原言語テスト文毎の第 2 翻訳正解率とを各原言語テスト文毎に比較し、その比較結果に基づいて翻訳エージェントの翻訳能力を評価する第 3 手段

、

を備えていることを特徴とする翻訳エージェントの翻訳能力評価装置。

【請求項 9】

50

第 1 手段は、

第 2 記憶手段内の原言語参照文から原言語テスト文に類似する原言語参照文を検索する

a 手段、

a 手段で検索された 1 または複数の原言語参照文に対応する正解翻訳文と、当該原言語テスト文に対応する正解翻訳文とを正解翻訳文候補とし、各正解翻訳文候補と翻訳エージェントによって得られた当該原言語テスト文に対する翻訳結果とを比較することにより、各正解翻訳文候補毎の第 1 翻訳正解率を算出する b 手段、および

b 手段で算出された各正解翻訳文候補毎の第 1 翻訳正解率のうち、最も翻訳正解率が高い翻訳正解率を、当該原言語テスト文に対する第 1 翻訳正解率として求める c 手段を備えており、

10

第 2 手段は、

a 手段で検索された 1 または複数の原言語参照文に対応する正解翻訳文と、当該原言語テスト文に対応する正解翻訳文とを正解翻訳文候補とし、各正解翻訳文候補と当該原言語テスト文に対応する翻訳能力別翻訳文とを比較することにより、各正解翻訳文候補毎の第 2 翻訳正解率を翻訳能力別翻訳文セット毎に算出する d 手段、および

d 手段で算出された各正解翻訳文候補毎の第 2 翻訳正解率のうち、最も翻訳正解率が高い翻訳正解率を、当該原言語テスト文に対する第 2 翻訳正解率として求める e 手段を備えていることを特徴とする請求項 8 に記載の翻訳エージェントの翻訳能力評価装置。

【請求項 10】

第 3 手段は、

20

翻訳能力別翻訳文セット毎に、各原言語テスト文に対する第 1 翻訳正解率と第 2 翻訳正解率とを比較することにより、翻訳エージェントによって得られた翻訳結果が翻訳能力別翻訳文より優れている原言語テスト文の数（以下、優位文数という）を翻訳能力別翻訳文セット毎に算出する手段、および

翻訳能力別翻訳文セット毎に算出された優位文数に基づいて、翻訳エージェントの翻訳能力を算出する手段、

を備えていることを特徴とする請求項 8 および 9 のいずれかに記載の翻訳エージェントの翻訳能力評価装置。

【請求項 11】

予め用意された複数の原言語テスト文を利用して翻訳エージェントの翻訳能力を評価する翻訳エージェントの翻訳能力評価装置であって、

30

各原言語テスト文に対応する正解翻訳文を記憶した第 1 記憶手段、

原言語テスト文とは異なる複数の原言語参照文を記憶した第 2 記憶手段、

各原言語参照文に対応する正解翻訳文を記憶した第 3 記憶手段、

各原言語テスト文に対応する翻訳能力別翻訳文を記憶した第 4 記憶手段、

各原言語テスト文毎に、原言語テスト文に対応する正解翻訳文と、第 2 記憶手段内の原言語参照文から検索された、原言語テスト文に類似する原言語参照文、に対応する正解翻訳文とを正解翻訳文候補とし、各正解翻訳文候補と翻訳エージェントによって得られた当該原言語テスト文に対する翻訳結果とを比較することにより、各原言語テスト文毎の第 1 翻訳正解率を算出し、得られた各原言語テスト文毎の第 1 翻訳正解率から原言語テスト文

40

セットに対する平均翻訳正解率を算出する第 1 手段、

各原言語テスト文毎に、原言語テスト文に対応する正解翻訳文と、第 2 記憶手段内の原言語参照文から検索された、原言語テスト文に類似する原言語参照文、に対応する正解翻訳文とを正解翻訳文候補とし、各正解翻訳文候補と原言語テスト文に対応する翻訳能力別翻訳文とを比較することにより、各原言語テスト文毎の第 2 翻訳正解率を各翻訳能力別翻訳文セット毎に算出し、各翻訳能力別翻訳文セット毎に得られた各原言語テスト文毎の第 2 翻訳正解率から、各翻訳能力別翻訳文セット毎の平均翻訳正解率を算出する第 2 手段、

ならびに

第 1 手段で算出された原言語テスト文セットに対する平均翻訳正解率と、第 2 手段で算出された各翻訳能力別翻訳文セット毎の平均翻訳正解率とに基づいて、翻訳エージェント

50

の翻訳能力を評価する第3手段、

を備えていることを特徴とする翻訳エージェントの翻訳能力評価装置。

【請求項12】

第1手段は、

第2記憶手段内の原言語参照文から原言語テスト文に類似する原言語参照文を検索する
a手段、

a手段で検索された1または複数の原言語参照文に対応する正解翻訳文と、当該原言語
テスト文に対応する正解翻訳文とを正解翻訳文候補とし、各正解翻訳文候補と翻訳エー
ジェントによって得られた当該原言語テスト文に対する翻訳結果とを比較することにより、
各正解翻訳文候補毎の第1翻訳正解率を算出するb手段、

b手段で算出された各正解翻訳文候補毎の第1翻訳正解率のうち、最も翻訳正解率が高
い翻訳正解率を、当該原言語テスト文に対する第1翻訳正解率として求めるc手段、およ
び

各原言語テスト文に対する第1翻訳正解率の平均を算出することによって、原言語テス
ト文セットに対する平均翻訳正解率を算出するd手段、

を備えており、

第2手段は、

a手段で検索された1または複数の原言語参照文に対応する正解翻訳文と、当該原言語
テスト文に対応する正解翻訳文とを正解翻訳文候補とし、各正解翻訳文候補と原言語テス
ト文に対応する翻訳能力別翻訳文とを比較することにより、各正解翻訳文候補毎の第2翻
訳正解率を翻訳能力別翻訳文セット毎に算出するe手段、

e手段で算出された各正解翻訳文候補毎の第2翻訳正解率のうち、最も翻訳正解率が高
い翻訳正解率を、当該原言語テスト文に対する第2翻訳正解率として求めるf手段、およ
び

翻訳能力別翻訳文セット毎に原言語テスト文に対する第2翻訳正解率の平均を算出する
ことによって、各翻訳能力別翻訳文セット毎の平均翻訳正解率を算出するg手段、

を備えていることを特徴とする請求項11に記載の翻訳エージェントの翻訳能力評価装
置。

【発明の詳細な説明】

【技術分野】

【0001】

この発明は、翻訳エージェントの翻訳能力評価装置に関する。

【背景技術】

【0002】

翻訳エージェントの翻訳能力評価方法として、予め決められた評価基準に従い、A、B
、C、Dなどのランクを評価者が主観で付与し、例えばA、BまたはCランクの割合を翻
訳能力の尺度（翻訳率）とするといった方法が知られている。

【0003】

A、B、C、Dの各ランクは、例えば、完全訳（訳文だけでまったく問題なし）、部分
訳（訳文は少し情報が欠けている）、可能訳（訳文はかなり情報が欠けている）、不可能
訳（翻訳からは情報が想像もできない）に対応する。

【0004】

翻訳エージェントの翻訳結果と正解翻訳文とをDPマッチングを用いて比較すること
により、翻訳結果の正解率を算出する手法も提案されている（Takezawa他、"A New Evaluat
ion Method for Speech Translation Systems and A Case Study on ATR-MATRIX from Ja
panese to English", MT Summit VII, Sept. 1999 参照）。

【0005】

DPマッチングを用いた翻訳結果の正解率の計算方法について説明する。DPマッピン
グ（例えば、長尾真著「言語工学」、発行所；株式会社昭晃堂、1985 参照）では、2つ
の文の全体にわたってできるだけもっともらしいマッチングが行なわれる。DPマッピン

10

20

30

40

50

グの例を説明する。文は単語系列で表されるとする。

【0006】

日本語の "今日は天気です" という文1は、"今日"、"は"、"天気"、"です" と分割される。日本語の "今日雨ですね" という文2は "今日"、"雨"、"です"、"ね" と分割される。DPマッチングの結果、表1の単語の対応関係の結果が得られる。

【0007】

【表1】

文1	文2	対比
今日	今日	変化なし
は		削除
天気	雨	置換
です	です	変化なし
	ね	挿入

10

【0008】

文1の "は" は文2では削除され、文1の "天気" は文2では "雨" に置換され、文2には "ね" が挿入されている。たとえば、文2を評価対象文とし、文1を正解文とすると、文2の正解率(類似度)は、次式(1)のように定義される。式(1)において、Totalは文1の単語数、Subは置換された単語数、Delは削除された単語数、Insは挿入された単語数である。

20

【0009】

$$\text{Acc} = (\text{Total} - (\text{Sub} + \text{Del} + \text{Ins})) / \text{Total} \quad \dots (1)$$

上記の例では、 $\text{Acc} = (4 - (1 + 1 + 1)) / 4 = 1/4 = 0.25$ となる。挿入、置換、削除の各誤りが無い場合には、正解率100%となり、誤りに応じて正解率は低下する。

【0010】

日英翻訳結果の翻訳正解率の計算も、同様に行なわれる。つまり、原言語である日本語に対応した英語の正解の翻訳結果と、評価対象の英語の翻訳結果とをDPマッチングすることにより、正解率を求める。翻訳正解率が100%であれば、原言語である日本語に対応した英語の正解の翻訳結果と、評価対象の英語の翻訳結果は完全に一致しているため、評価対象の英語の翻訳結果は正しいと判断できる。

30

【発明の開示】

【発明が解決しようとする課題】

【0011】

上記従来におけるDPマッチングを用いた翻訳正解率の計算では、評価対象の翻訳結果と正解翻訳とのマッチングに基づく距離を単に計算しているだけなので、翻訳結果が正解とは異なる表現を採用している場合は、たとえ正解であっても翻訳正解率が著しく劣化してしまう。したがって、表現の多様性を許容できない欠点がある。更に、翻訳正解率と従来の主観評価に基づく翻訳率などの評価尺度との関連性が明らかでないという欠点がある。

40

【0012】

この発明は、評価対象の翻訳結果の表現の多様性を許容しながら、高い精度で翻訳エージェントの翻訳能力を評価することができる翻訳エージェントの翻訳能力評価装置を提供することを目的とする。

【0013】

また、この発明は、従来、主観的に行なっていたランク評価の全部または一部を、自動

50

的に行なうことが可能となる翻訳エージェントの翻訳能力評価装置を提供することを目的とする。

【課題を解決するための手段】

【0014】

この発明は、予め用意された複数の原言語テスト文を利用して翻訳エージェントの翻訳能力を評価する翻訳エージェントの翻訳能力評価装置であって、各原言語テスト文に対応する正解翻訳文を記憶した第1記憶手段、原言語テスト文とは異なる複数の原言語参照文を記憶した第2記憶手段、各原言語参照文に対応する正解翻訳文を記憶した第3記憶手段、各原言語テスト文毎に、原言語テスト文に対応する正解翻訳文と、第2記憶手段内の原言語参照文から検索された、原言語テスト文に類似する原言語参照文、に対応する正解翻訳文とを正解翻訳文候補とし、各正解翻訳文候補と翻訳エージェントによって得られた当該原言語テスト文に対する翻訳結果とを比較することにより、各原言語テスト文毎の翻訳正解率を算出する第1手段、ならびに第1手段で算出された各原言語テスト文毎の翻訳正解率に基づいて翻訳エージェントの翻訳能力を評価する第2手段を備えていることを特徴とする。

10

【0015】

第1手段は、たとえば、第2記憶手段内の原言語参照文から原言語テスト文に類似する原言語参照文を検索するa手段、a手段で検索された1または複数の原言語参照文に対応する正解翻訳文と、当該原言語テスト文に対応する正解翻訳文とを正解翻訳文候補とし、各正解翻訳文候補と翻訳エージェントによって得られた当該原言語テスト文に対する翻訳結果とを比較することにより、各正解翻訳文候補毎の翻訳正解率を算出するb手段、およびb手段で算出された各正解翻訳文候補毎の翻訳正解率のうち、最も翻訳正解率が高い翻訳正解率を、当該原言語テスト文に対する翻訳正解率として求めるc手段を備えている。

20

【0016】

a手段では、たとえば、原言語テスト文と各原言語参照文との類似度を算出し、所定の類似度判定用閾値より類似度が高い原言語参照文を原言語テスト文に類似する原言語参照文として検索する。

【0017】

類似度は、たとえば、DPマッチングに基づいて算出される。翻訳正解率は、たとえば、DPマッチングに基づいて算出される。

30

【0018】

第2手段では、たとえば、第1手段で算出された各原言語テスト文毎の翻訳正解率の平均値を算出し、この算出結果に基づいて翻訳エージェントの翻訳能力を評価する。

【0019】

第2手段では、第1手段で算出された各原言語テスト文毎の翻訳正解率と、所定のランク判別用域値とを比較することにより、翻訳エージェントによって得られた当該原言語テスト文に対する翻訳結果をランク分けし、ランク分け結果に基づいて翻訳エージェントの翻訳能力を評価する。

【0020】

この発明は、予め用意された複数の原言語テスト文を利用して翻訳エージェントの翻訳能力を評価する翻訳エージェントの翻訳能力評価装置であって、各原言語テスト文に対応する正解翻訳文を記憶した第1記憶手段、原言語テスト文とは異なる複数の原言語参照文を記憶した第2記憶手段、各原言語参照文に対応する正解翻訳文を記憶した第3記憶手段、各原言語テスト文毎に、原言語テスト文に対応する正解翻訳文と、第2記憶手段内の原言語参照文から検索された、原言語テスト文に類似する原言語参照文、に対応する正解翻訳文とを正解翻訳文候補とし、各正解翻訳文候補と翻訳エージェントによって得られた原言語テスト文に対する翻訳結果とを比較することにより、各原言語テスト文毎の第1翻訳正解率を算出する第1手段、各原言語テスト文毎に、原言語テスト文に対応する正解翻訳文と、第2記憶手段内の原言語参照文から検索された、原言語テスト文に類似する原言語参照文、に対応

40

50

する正解翻訳文とを正解翻訳文候補とし、各正解翻訳文候補と原言語テスト文に対応する翻訳能力別翻訳文とを比較することにより、各原言語テスト文毎の第2翻訳正解率を各翻訳能力別翻訳文セット毎に算出する第2手段、ならびに第1手段で算出された各原言語テスト文毎の第1翻訳正解率と、第2手段で各翻訳能力別翻訳文セット毎に算出された各原言語テスト文毎の第2翻訳正解率とを各原言語テスト文毎に比較し、その比較結果に基づいて翻訳エージェントの翻訳能力を評価する第3手段を備えていることを特徴とする。

【0021】

第1手段は、たとえば、第2記憶手段内の原言語参照文から原言語テスト文に類似する原言語参照文を検索するa手段、a手段で検索された1または複数の原言語参照文に対応する正解翻訳文と、当該原言語テスト文に対応する正解翻訳文とを正解翻訳文候補とし、各正解翻訳文候補と翻訳エージェントによって得られた当該原言語テスト文に対する翻訳結果とを比較することにより、各正解翻訳文候補毎の第1翻訳正解率を算出するb手段、およびb手段で算出された各正解翻訳文候補毎の第1翻訳正解率のうち、最も翻訳正解率が高い翻訳正解率を、当該原言語テスト文に対する第1翻訳正解率として求めるc手段を備えている。

10

【0022】

第2手段は、たとえば、a手段で検索された1または複数の原言語参照文に対応する正解翻訳文と、当該原言語テスト文に対応する正解翻訳文とを正解翻訳文候補とし、各正解翻訳文候補と当該原言語テスト文に対応する翻訳能力別翻訳文とを比較することにより、各正解翻訳文候補毎の第2翻訳正解率を翻訳能力別翻訳文セット毎に算出するd手段、およびd手段で算出された各正解翻訳文候補毎の第2翻訳正解率のうち、最も翻訳正解率が高い翻訳正解率を、当該原言語テスト文に対する第2翻訳正解率として求めるe手段を備えている。

20

【0023】

第3手段は、たとえば、翻訳能力別翻訳文セット毎に、各原言語テスト文に対する第1翻訳正解率と第2翻訳正解率とを比較することにより、翻訳エージェントによって得られた翻訳結果が翻訳能力別翻訳文より優れている原言語テスト文の数(以下、優位文数という)を翻訳能力別翻訳文セット毎に算出する手段、および翻訳能力別翻訳文セット毎に算出された優位文数に基づいて、翻訳エージェントの翻訳能力を算出する手段を備えている。

30

【0024】

この発明は、予め用意された複数の原言語テスト文を利用して翻訳エージェントの翻訳能力を評価する翻訳エージェントの翻訳能力評価装置であって、各原言語テスト文に対応する正解翻訳文を記憶した第1記憶手段、原言語テスト文とは異なる複数の原言語参照文を記憶した第2記憶手段、各原言語参照文に対応する正解翻訳文を記憶した第3記憶手段、各原言語テスト文に対応する翻訳能力別翻訳文を記憶した第4記憶手段、各原言語テスト文毎に、原言語テスト文に対応する正解翻訳文と、第2記憶手段内の原言語参照文から検索された、原言語テスト文に類似する原言語参照文、に対応する正解翻訳文とを正解翻訳文候補とし、各正解翻訳文候補と翻訳エージェントによって得られた当該原言語テスト文に対する翻訳結果とを比較することにより、各原言語テスト文毎の第1翻訳正解率を算出し、得られた各原言語テスト文毎の第1翻訳正解率から原言語テスト文セットに対する平均翻訳正解率を算出する第1手段、各原言語テスト文毎に、原言語テスト文に対応する正解翻訳文と、第2記憶手段内の原言語参照文から検索された、原言語テスト文に類似する原言語参照文、に対応する正解翻訳文とを正解翻訳文候補とし、各正解翻訳文候補と原言語テスト文に対応する翻訳能力別翻訳文とを比較することにより、各原言語テスト文毎の第2翻訳正解率を各翻訳能力別翻訳文セット毎に算出し、各翻訳能力別翻訳文セット毎に得られた各原言語テスト文毎の第2翻訳正解率から、各翻訳能力別翻訳文セット毎の平均翻訳正解率を算出する第2手段、ならびに第1手段で算出された原言語テスト文セットに対する平均翻訳正解率と、第2手段で算出された各翻訳能力別翻訳文セット毎の平均翻訳正解率とに基づいて、翻訳エージェントの翻訳能力を評価する第3手段を備えているこ

40

50

とを特徴とする。

【0025】

第1手段は、たとえば、第2記憶手段内の原言語参照文から原言語テスト文に類似する原言語参照文を検索するa手段、a手段で検索された1または複数の原言語参照文に対応する正解翻訳文と、当該原言語テスト文に対応する正解翻訳文とを正解翻訳文候補とし、各正解翻訳文候補と翻訳エージェントによって得られた当該原言語テスト文に対する翻訳結果とを比較することにより、各正解翻訳文候補毎の第1翻訳正解率を算出するb手段、b手段で算出された各正解翻訳文候補毎の第1翻訳正解率のうち、最も翻訳正解率が高い翻訳正解率を、当該原言語テスト文に対する第1翻訳正解率として求めるc手段、および各原言語テスト文に対する第1翻訳正解率の平均を算出することによって、原言語テスト文セットに対する平均翻訳正解率を算出するd手段を備えている。

10

【0026】

第2手段は、たとえば、a手段で検索された1または複数の原言語参照文に対応する正解翻訳文と、当該原言語テスト文に対応する正解翻訳文とを正解翻訳文候補とし、各正解翻訳文候補と原言語テスト文に対応する翻訳能力別翻訳文とを比較することにより、各正解翻訳文候補毎の第2翻訳正解率を翻訳能力別翻訳文セット毎に算出するe手段、e手段で算出された各正解翻訳文候補毎の第2翻訳正解率のうち、最も翻訳正解率が高い翻訳正解率を、当該原言語テスト文に対する第2翻訳正解率として求めるf手段、および翻訳能力別翻訳文セット毎に原言語テスト文に対する第2翻訳正解率の平均を算出することによって、各翻訳能力別翻訳文セット毎の平均翻訳正解率を算出するg手段を備えている。

20

【発明の効果】

【0027】

この発明によれば、評価対象の翻訳結果の表現の多様性を許容しながら、高い精度で翻訳エージェントの翻訳能力を評価することができるようになる。また、この発明によれば、従来、主観的に行なっていたランク評価の全部または一部を、自動的に行なうことが可能となる。

【発明を実施するための最良の形態】

【0028】

以下、図面を参照して、この発明の実施例について説明する。

【実施例1】

30

【0029】

図1、図2および図3を参照して、この発明の第1の実施の形態について説明する。

【0030】

図1は、翻訳エージェントの翻訳能力評価を行なうシステムの基本的な構成を示している。ここでは、日本語を英語に翻訳する日英機械翻訳システムの翻訳能力を評価する場合を例にとって説明する。

【0031】

日英機械翻訳システムの翻訳能力を評価するために、テスト文の集合が日本語テストセット1として用意されているとともに、日本語の各テスト文に対応した英語の正解が英語テストセット2として用意されているものとする。ここでは、各日本語テスト文を $j(i)$ ($i=1,2,3, \dots, n$) で表し、各日本語テスト文に対応した英語の正解文を $e(i)$ ($i=1,2,3, \dots, n$) で表す。

40

【0032】

一方、日本語テスト文とは別に、複数の日本語文（以下、日本語参照文という）とその英語翻訳結果である英文とが用意されており、各日本語参照文と対応する英文とは対応付けられているものとする。日本語参照文の集合を日本語データベース3と呼び、各日本語参照文に対応する英文の集合を英語データベース4と呼ぶことにする。

【0033】

日本語テキストセット1内の日本語テスト文 $j(i)$ は、日英機械翻訳システム6に入力される。日英機械翻訳システム6は、入力された日本語テスト文 $j(i)$ を翻訳し、その翻

50

訳結果 7 を出力する。

【 0 0 3 4 】

また、日本語テキストセット 1 内の日本語テスト文 $j(i)$ は、類似日本語文検索器 5 にも入力される。類似日本語文検索器 5 は、入力された日本語テスト文 $j(i)$ に類似する日本語参照文を、日本語データベース 3 から検索する。

【 0 0 3 5 】

この検索は、次のようにして行なわれる。まず、日本語テスト文 $j(i)$ と日本語データベース 3 内の各日本語参照文とを比較し、上記式 (1) を用いて、各日本語参照文毎に、日本語参照文の日本語テスト文 $j(i)$ に対する類似度 (正解率) を算出する。予め定めた類似度判別用閾値よりも正解率が高い日本語参照文を、日本語テスト文 $j(i)$ に類似した日本語参照文 (以下、類似日本語参照文という) として特定する。

10

【 0 0 3 6 】

そして、特定した類似日本語参照文に対応する英文を、英語データベース 4 から検索し、検索した英文を、日本語テスト文 $j(i)$ に対する英文正解候補として追加する。したがって、日本語テスト文 $j(i)$ に対する英文正解候補は、日本語テスト文 $j(i)$ に対応する正解文 $e(i)$ と、類似日本語文検索器 5 によって検索された類似日本語参照文に対応する英文との集合となる。

【 0 0 3 7 】

翻訳正解率計算器 8 は、日本語テスト文 $j(i)$ の翻訳結果 7 と各英文正解候補とを比較し、上記式 (1) を用いて、各英文正解候補毎に日本語テスト文 $j(i)$ の英文正解候補に対する翻訳正解率を算出する。そして、翻訳正解率の最大値を当該日本語テスト文 $j(i)$ の翻訳正解率とする。

20

【 0 0 3 8 】

このようにして、日本語テキストセット 1 内の各日本語テスト文 $j(i)$ に対する翻訳正解率を算出する。そして、全ての日本語テスト文 $j(i)$ に対して求められた翻訳正解率の平均をテストセット翻訳正解率として求める。得られたテストセット翻訳正解率を、機械翻訳システムの翻訳能力評価の尺度として用いる。

【 0 0 3 9 】

ところで、上記のようにして求められたテストセット翻訳正解率は、機械翻訳システムの翻訳能力評価の尺度の 1 つであるが、従来の主観評価 (A B C D ランク評価) との対応関係が不明である。これらの対応関係の算出方法を次に説明する。

30

【 0 0 4 0 】

図 2 は、各日本語テスト文 $j(i)$ に対して求められた翻訳正解率と、各日本語テスト文 $j(i)$ の翻訳結果に対して行なわれた主観評価結果 (A B C D ランク評価結果) との関係を示している。なお、図 2 は、日英機械翻訳システム 6 として、A T R 音声翻訳通信研究所が研究開発した機械翻訳機 (T D M T) を用いた場合の実験結果を示している。

【 0 0 4 1 】

図 2 の A、B、C、D は、評価ランクを示している。図 2 の縦軸は各日本語テスト文の翻訳正解率を表している。

【 0 0 4 2 】

A ランクが付与された日本語テスト文 (翻訳結果) の翻訳正解率は分布しているがその平均翻訳正解率は高く、D ランクが付与された日本語テスト文 (翻訳結果) の平均翻訳正解率は低い。そこで、A ランクと B ランク以下 (B、C、D) とを判別できるようなランク判別用閾値を設定すれば、ランク判別用閾値より大きな翻訳正解率に属する日本語テスト文 (翻訳結果) を A ランクと見做し、ランク判別用閾値以下の翻訳正解率に属する日本語テスト文 (翻訳結果) を B ランク以下のランクと見做すことができる。

40

【 0 0 4 3 】

A ランクと B ランク以下とを判別できるようなランク判別用閾値を決定する方法としては、主観評価によって A ランクが付与された日本語テスト文 (翻訳結果) の平均翻訳正解率と、主観評価によって B ランク以下のランク B、C、D が付与された日本語テスト文 (

50

翻訳結果)の平均翻訳正解率の中間値を選択する等、判別分析と呼ばれている手法を利用することができる。

【0044】

AランクとBランク以下とを判別できるようなランク判別用閾値を用いて、各日本語テスト文をAランクとBランク以下とにランク分け(A/B, C, D)し、主観評価結果を用いてそのランク分けが正しい割合(判別率)を算出した。同様に、Bランク以上のランクA, Bと、Cランク以下のランクC, Dとを判別できるようなランク判別用閾値を用いて、各日本語テスト文をBランク以上とCランク以下とにランク分け(A, B/C, D)し、そのランク分けが正しい割合(判別率)を算出した。

【0045】

また、Cランク以上のランクA, B, Cと、Dランクとを判別できるようなランク判別用閾値を用いて、各日本語テスト文をCランク以上とDランクとにランク分け(A, B, C/D)し、そのランク分けが正しい割合(判別率)を算出した。さらに、AランクとBランク、BランクとCランク、CランクとDランクとをそれぞれ判別できるようなランク判別用閾値を用いて、各日本語テスト文を各ランク毎にランク分け(A/B/C/D)し、そのランク分けが正しい割合(判別率)を算出した。

【0046】

このような判別率の算出を、類似日本語文検索器5で用いられる類似度判別用閾値を10%から100%までの範囲で10%ずつ変化させて行なった結果を、図3に示す。図3の横軸は類似度判別用閾値を示し、縦軸は判別率を示している。また、図3の右端には、日本語テスト文に対して類似日本語文を検索しない従来例の場合の実験結果を示している。

【0047】

図3から、類似度判別用閾値を100%とした場合であっても、日本語テスト文に対して類似日本語参照文が検索されることがあるため、従来例に比べて判別率が高く、表現の多様性を許容する効果があることが理解できる。また、類似度判別用閾値を適切に選択することにより、従来例に比べて、15%も、判別率が向上させることが可能である。

【0048】

また、図3から、ランク判別用閾値を用いてランク分けを行なった場合に、AランクとBランク以下とのランク分け、Bランク以上とCランク以下とのランク分けおよびCランク以上とDランクとのランク分けのうち、AランクとBランク以下とのランク分けを行なった場合の判別率が最も高いことが分かる。Aランクに属すると判別された日本語テスト文の数の日本語テスト文総数に対する割合を翻訳能力の尺度(翻訳率)として用いれば、主観評価の80%という高い精度で翻訳能力評価を自動的に行なえるようになる。

【0049】

なお、A, B, C, Dを別々のランクで高い精度で評価する場合には、Aのランク判定を上述の方法で自動判定し、その後、B, C, Dのランク判別を評価者が行うこともできる。このようにすれば、高い翻訳率精度を維持しながら、評価者が評価する日本語テスト文の数を削減できるので、評価者の負担が軽減され、効率的な評価を行なうことが可能となる。

【実施例2】

【0050】

以下、図4および図5を参照して、この発明の第2の実施の形態について説明する。

【0051】

図4は、翻訳エージェントの翻訳能力評価を行なうシステムの基本的な構成を示している。ここでは、日本語を英語に翻訳する日英機械翻訳システムの翻訳能力を評価する場合を例にとって説明する。

【0052】

第1の翻訳正解率計算器12は、第1の実施の形態の翻訳正解率計算器8と同様に、日本語テスト文j(i)に対する日英機械翻訳システムの翻訳結果(評価対象の翻訳結果)11

10

20

30

40

50

と日本語テスト文 $j(i)$ に対する英文正解候補とに基づいて、日本語テスト文 $j(i)$ 毎に翻訳正解率を計算する。日本語テスト文 $j(i)$ に対する英文正解項候補は、日本語テスト文 $j(i)$ に対応する正解文 $e(i)$ と、類似日本語文検索器5（図1参照）によって検索された類似日本語参照文に対応する英文との集合からなる。

【0053】

参照翻訳データベース13には、各日本語テスト文毎に能力別の複数の翻訳結果が保存されている。英語を例にとると、参照翻訳データベース13には TOEIC (Test of English for International Communication) 試験のスコアが異なる複数の TOEIC 受験者の各日本語テスト文毎の翻訳結果 $TOEIC(i,t)$ ($i=1,2,3, \dots, n, t=1,2,3, \dots, m$) が格納されている。

10

【0054】

第2の翻訳正解率計算器14は、日本語テスト文 $j(i)$ に対する参照翻訳データベース13内の翻訳結果 $TOEIC(i,t)$ と日本語テスト文 $j(i)$ に対する英文正解候補とに基づいて、第1の翻訳正解率計算器12と同様な手法で、日本語テスト文 $j(i)$ 毎に翻訳正解率を計算する。日本語テスト文 $j(i)$ に対する英文正解項候補は、日本語テスト文 $j(i)$ に対応する正解文 $e(i)$ と、類似日本語文検索器5（図1参照）によって検索された類似日本語参照文に対応する英文との集合からなる。

【0055】

第1の翻訳正解率計算器12によって算出される日本語テスト文 $j(i)$ 毎の翻訳正解率を $A1(i)$ とする。第2の翻訳正解率計算器14によって算出される受験者毎の日本語テスト文 $j(i)$ 毎の翻訳正解率を $A2(i,t)$ とする。

20

【0056】

ここでは、参照翻訳データベース13に格納されている複数の受験者の各日本語テスト文毎の翻訳結果のうち、一人の受験者の各日本語テスト文毎の翻訳結果を用いて1セット分の日本語テスト文 $j(i)$ に対する処理が行なわれた後、次の受験者の各日本語テスト文毎の翻訳結果を用いて1セット分の日本語テスト文 $j(i)$ に対する処理が行なわれるというように、各受験者の翻訳結果セット毎に1セット分の日本語テスト文 $j(i)$ に対する処理が行なわれる。

【0057】

一対比較器15は、第1の翻訳正解率計算器12によって算出される翻訳正解率 $A1(i)$ と、第2の翻訳正解率計算器14によって算出される翻訳正解率 $A2(i,t)$ とを比較し、それらの優劣を判定する。一対比較器15は、閾値を Th とすると、たとえば、次の条件式(i), (ii), (iii)を用いて優劣を判定する。

30

(i) if $A1 > A2 + Th$ then $A1$ が優位と判定

(ii) if $|A1 - A2| < Th$ then $A1$ と $A2$ とが同等と判定

(iii) if $A2 > A1 + Th$ then $A2$ が優位と判定

これらの判定結果は、一対比較集計器16に送られる。一対比較集計器16は、一人の受験者の翻訳結果に基づいて1セット分の日本語テスト文 $j(i)$ に対する優位判定が行なわれる毎に、当該1セット分の日本語テスト文 $j(i)$ に対する優位判定結果に基づいて、日英機械翻訳システムによって得られた翻訳結果が、TOEIC 受験者の翻訳結果より優れている原言語テスト文の数（優位文数）を算出する。

40

【0058】

ここでは、優位文数は、1セット分の日本語テスト文 $j(i)$ に対する日英機械翻訳システムの翻訳結果のうち、 $A1$ が優位と判定された評価対象の翻訳結果の数と、同等と判定された評価対象の翻訳結果の数の1/2とを加算した値として定義される。以上により、受験者毎（TOEIC のスコア毎）の優位文数が求められる。

【0059】

TOEIC のスコア毎の優位文数は、翻訳能力均衡点計算器17に送られる。翻訳能力均衡点計算器17は、TOEIC のスコア毎の優位文数を回帰分析することによってTOEIC のスコアに対する優位文数を表す回帰直線を求め、その回帰直線上で優位文数が日本語テスト文

50

の総数の半分になる点に対応するTOEIC のスコアを求める。そして、求めたTOEIC のスコアを、日英機械翻訳システムのTOEIC のスコア（翻訳能力の評価値）とする。

【 0 0 6 0 】

図 5 は、一対比較集計器 1 6 によって求められたTOEIC のスコア毎の優位文数、および翻訳能力均衡点計算器 1 7 によって求められた回帰直線を示している。

【 0 0 6 1 】

なお、図 5 は、日英機械翻訳システムとして、A T R 音声翻訳通信研究所が研究開発した機械翻訳機(TDMT)を用いた場合の実験結果を示している。また、この例では、日本語テスト文の総数は 3 3 0 である。

【 0 0 6 2 】

図 5 の折れ線 a 1 は、第 2 の実施の形態によって得られたTOEIC のスコア毎の優位文数を示し、直線 a 2 は折れ線 a 1 のデータから求められた回帰直線を示している。

【 0 0 6 3 】

また、折れ線 b 1 は、第 1 および第 2 の翻訳正解率計算器 1 2、1 4 が日本語テスト文に対する類似日本語参照文を利用することなく翻訳正解率を算出した場合のスコア毎の優位文数を示し、直線 b 2 は折れ線 b 1 のデータから求められた回帰直線を示している。また、折れ線 c 1 は、日英機械翻訳システムの翻訳結果と参照翻訳データベース内の翻訳結果との比較を評価者が主観的に行なった場合の、TOEIC のスコア毎の優位文数を示し、直線 c 2 は折れ線 c 1 のデータから求められた回帰直線を示している。

【 0 0 6 4 】

優位文数は、受験者のTOEIC スコアが高くなるのにもなって少なくなる傾向がある。図 5 では、第 2 の実施の形態によって求められた日英機械翻訳システムのTOEIC スコアは 7 3 0 点となっている。日英機械翻訳システムの翻訳結果と参照翻訳データベース内の翻訳結果との比較を評価者が主観的に行なった場合の日英機械翻訳システムのTOEIC スコアは 7 0 8 点となり、類似日本語参照文を利用しない翻訳正解率に基づいて日英機械翻訳システムのTOEIC スコアを求めると 5 8 0 点となる。

【 0 0 6 5 】

第 2 の実施の形態によって求められた日英機械翻訳システムのTOEIC スコアは、類似日本語参照文を利用しない翻訳正解率に基づいて日英機械翻訳システムのTOEIC スコアに比べて、日英機械翻訳システムの翻訳結果と参照翻訳データベース内の翻訳結果との比較を評価者が主観的に行なった場合の日英機械翻訳システムのTOEIC スコアにはるかに近似したものとなり、第 2 の実施の形態によれば主観的評価に近い評価結果が自動的に得られることがわかる。

【 0 0 6 6 】

なお、参照翻訳データベース 1 3 内の翻訳結果を翻訳正解率計算器 1 4 に入力する際に、当該翻訳結果の精度が低下するように誤差を付加するようにしてもよい。

【実施例 3】

【 0 0 6 7 】

以下、図 6 および図 7 を参照して、この発明の第 3 の実施の形態について説明する。

【 0 0 6 8 】

図 6 は、翻訳エージェントの翻訳能力評価を行なうシステムの基本的な構成を示している。ここでは、日本語を英語に翻訳する日英機械翻訳システムの翻訳能力を評価する場合を例にとって説明する。

【 0 0 6 9 】

参照翻訳データベース 2 3 は、第 2 の実施形態の参照翻訳データベース 1 3 と同じであり、TOEIC試験のスコアが異なる複数のTOEIC 受験者の各日本語テスト文毎の翻訳結果 $TOEIC(i, t)$ ($i=1, 2, 3, \dots, n, t=1, 2, 3, \dots, m$) が格納されている。

【 0 0 7 0 】

第 1 のテストセット翻訳正解率計算器 2 2 は、第 1 の実施の形態の翻訳正解率計算器 8 と同様に、日本語テスト文 $j(i)$ に対する日英機械翻訳システムの翻訳結果（評価対象の翻

10

20

30

40

50

訳結果) 2 1 と日本語テスト文 $j(i)$ に対する英文正解候補とに基づいて、テスト日本語文全体に対する翻訳正解率(評価対象の翻訳結果に対するテストセット翻訳正解率)を計算する。日本語テスト文 $j(i)$ に対する英文正解候補は、日本語テスト文 $j(i)$ に対応する正解文 $e(i)$ と、類似日本語文検索器 5 (図 1 参照) によって検索された類似日本語参照文に対応する英文との集合からなる。

【0071】

第 2 のテストセット翻訳正解率計算器 2 4 は、日本語テスト文 $j(i)$ に対する参照翻訳データベース 1 3 内の翻訳結果 $TOEIC(i, t)$ と日本語テスト文 $j(i)$ に対する英文正解候補とに基づいて、第 1 のテストセット翻訳正解率計算器 2 2 と同様な手法で、受験者毎 (TOEIC のスコア毎) にテスト日本語文全体に対する翻訳正解率(参照翻訳結果に対するテストセ

10

【0072】

回帰直線計算器 2 5 は、第 2 のテストセット翻訳正解率計算器 2 4 によって求められた TOEIC のスコア毎の参照翻訳結果に対するテストセット翻訳正解率から、回帰分析によって TOEIC のスコアに対するテストセット翻訳正解率を表す回帰直線を計算する。

【0073】

翻訳能力計算器 2 6 は、第 1 のテストセット翻訳正解率計算器 2 2 によって求められた評価対象の翻訳結果に対するテストセット翻訳正解率と、回帰直線計算器 2 5 によって求められた回帰直線とに基づいて、日英機械翻訳システムの翻訳能力を計算する。

【0074】

図 7 は、第 2 のテストセット翻訳正解率計算器 2 4 によって求められた TOEIC のスコア毎の参照翻訳結果に対するテストセット翻訳正解率および回帰直線計算器 2 5 によって求められた回帰直線を示している。

20

【0075】

なお、図 7 は、日英機械翻訳システムとして、ATR 音声翻訳通信研究所が開発した機械翻訳機 (TDMT) を用いた場合の実験結果を示している。また、この例では、日本語テスト文の総数は 330 である。

【0076】

図 7 において、横軸は TOEIC 受験者のスコアを示し、縦軸はテストセット翻訳正解率を示している。

30

【0077】

図 7 の折れ線 a 1 は、第 3 の実施の形態の第 2 のテストセット翻訳正解率計算器 2 4 によって求められた TOEIC のスコア毎の参照翻訳結果に対するテストセット翻訳正解率を示し、直線 a 2 は折れ線 a 1 のデータから求められた回帰直線を示している。値 A は、第 3 の実施の形態の第 1 のテストセット翻訳正解率計算器 2 2 によって算出された評価対象の翻訳結果に対するテストセット翻訳正解率を示している。

【0078】

また、折れ線 b 1 は、日本語テスト文に対する類似日本語参照文を利用することなく算出された TOEIC のスコア毎の参照翻訳結果に対するテストセット翻訳正解率を示し、直線 b 2 は折れ線 b 1 のデータから求められた回帰直線を示している。値 B は、日本語テスト文に対する類似日本語参照文を利用することなく算出された評価対象の翻訳結果に対する

40

【0079】

直線 a 2 で示される回帰直線と、第 1 のテストセット翻訳正解率計算器 2 2 によって求められた評価対象の翻訳結果に対するテストセット翻訳正解率 A との交点に対応する TOEIC のスコアが、第 3 の実施の形態によって求められた日英機械翻訳システムの TOEIC のスコアである。図 7 の場合は、日英機械翻訳システムの TOEIC のスコアは 680 点となっている。

【0080】

同様にして、類似日本語文を利用しない翻訳正解率に基づいて日英機械翻訳システムの

50

TOEIC スコアを求めると 650 点となる。なお、日英機械翻訳システムの翻訳結果と参照翻訳データベース内の翻訳結果との比較を評価者が行った主観評価実験によると日英機械翻訳システムの TOEIC スコアは 708 点となった。

【0081】

第3の実施の形態によって求められた日英機械翻訳システムの TOEIC スコアは、類似日本語参照文を利用しない翻訳正解率に基づいて日英機械翻訳システムの TOEIC スコアに比べて、主観的に評価を行なった場合の日英機械翻訳システムの TOEIC スコアにはるかに近似したものとなり、第3の実施の形態によれば主観的评价に近い評価結果が自動的に得られることがわかる。

【0082】

なお、参照翻訳データベース23内の翻訳結果をテストセット翻訳正解率計算器24に入力する際に、当該翻訳結果の精度が低下するように誤差を付加するようにしてもよい。

【0083】

上記各実施の形態によれば、次のような利点がある。

(1) 原言語テスト文に類似する類似原言語参照文として誤った用例が検索されたとしても、誤って検索された用例に対応する類似対訳文に対する翻訳正解率は低い値となるため、当該翻訳正解率は採用されなくなる。このため、原言語テスト文に類似する類似原言語参照文の検索誤りの影響が少なく、高い精度の翻訳評価が可能である。

(2) 原言語テスト文に類似すると検索された類似原言語参照文に対応する類似対訳文は、人手などで作られる精度の高い対訳であるので、単語の選択、並びが正確である。このため高い精度の翻訳評価が可能である。

(3) 従来において主観評価を行なうことによって決定していた A B C D 等のランク分けを、自動的に行なうことが可能となる。特に、高い翻訳正解率の場合には A ランクの主観評価との対応関係が強いので、A ランクのテスト文については高い精度で自動判別を行なうことが可能となる。

(4) 各テスト文に対する翻訳結果を、A B C D ランクにランク分けさせたい場合に、高い翻訳正解率の場合には A ランクの主観評価との対応関係が強い性質を利用することにより、A ランクのテスト文については自動的な判別を行い、A ランク以下のランクのテスト文については、従来の主観による高い精度の評価を行なうことにより、ランク分けに要する評価者の負担が軽減される。

(5) 異なる人間の翻訳能力に対応した参照翻訳データベースを利用する場合、各人間の翻訳能力に対応した翻訳正解率と評価対象の翻訳結果の翻訳正解率をテスト文毎に一对比較し、その比較結果の優劣が均衡する参照翻訳データベースの人間能力を翻訳対象の翻訳能力とすることにより翻訳自動評価が実現できる。また、参照翻訳データベースの翻訳正解率と優位文数の間の回帰直線を利用し、高い精度で均衡点を求める構成も実現できる。

(6) 異なる人間の翻訳能力に対応した参照翻訳データベースを利用する場合、各人間の翻訳能力それぞれに対応したテストセット平均翻訳正解率と、評価対象の翻訳結果に対するテストセット平均翻訳正解率とを比較し、両者が均衡する参照翻訳データベースの人間能力を翻訳対象の翻訳能力とすることにより翻訳自動評価が実現できる。この際、各人間の翻訳能力それぞれに対応したテストセット平均翻訳正解率から人間の翻訳能力に対するテストセット平均翻訳正解率を表す回帰直線を求め、この回帰直線と評価対象の翻訳結果に対するテストセット平均翻訳正解率の交点から、高い精度で均衡点を求めることもできる。

【図面の簡単な説明】

【0084】

【図1】第1の実施の形態による翻訳エージェントの翻訳能力評価を行なうシステムの基本的な構成を示すブロック図である。

【図2】各日本語テスト文 $j(i)$ に対して求められた翻訳正解率と、各日本語テスト文 $j(i)$ の翻訳結果に対して行なわれた主観評価結果 (A B C D ランク評価結果) との関係を示すグラフである。

10

20

30

40

50

【図3】類似度判別用閾値と判別率との関係を示すグラフである。

【図4】第2の実施の形態による翻訳エージェントの翻訳能力評価を行なうシステムの基本的な構成を示すブロック図である。

【図5】一対比較集計器16によって求められたTOEICのスコア毎の優位文数、および翻訳能力均衡点計算器17によって求められた回帰直線を示すグラフである。

【図6】第3の実施の形態による翻訳エージェントの翻訳能力評価を行なうシステムの基本的な構成を示すブロック図である。

【図7】第2のテストセット翻訳正解率計算器24によって求められたTOEICのスコア毎の参照翻訳結果に対するテストセット翻訳正解率および回帰直線計算器25によって求められた回帰直線を示すグラフである。

10

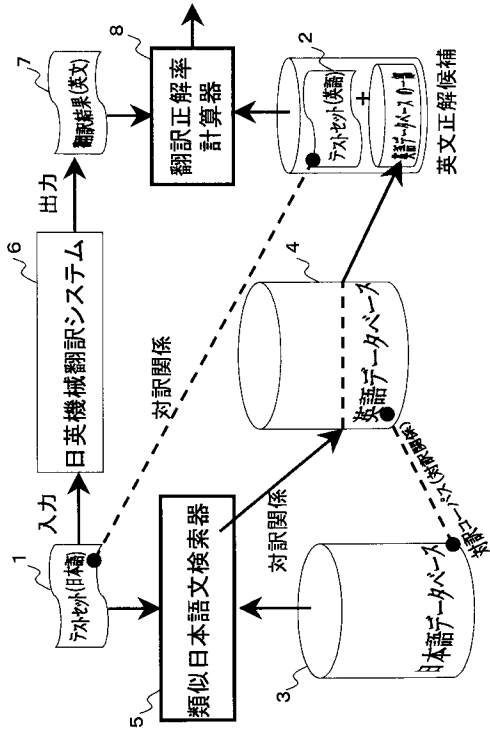
【符号の説明】

【0085】

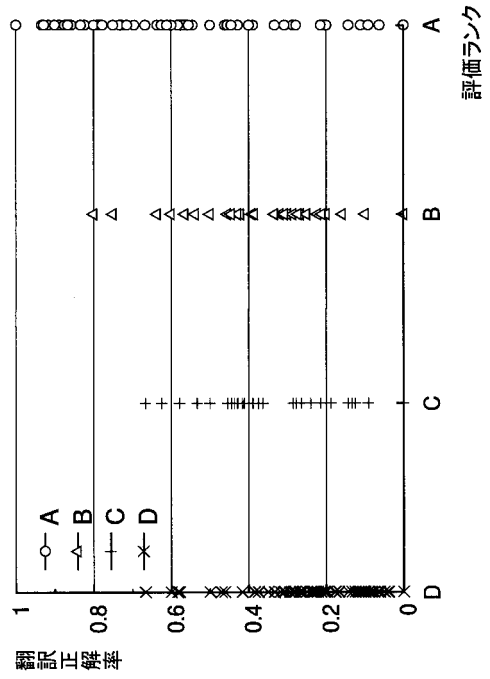
- 3 日本語データベース
- 4 英語データベース
- 5 類似日本語文検索器
- 8 翻訳正解率計算器
- 12 第1の翻訳正解率計算器
- 13 参照翻訳データベース
- 14 第2の翻訳正解率計算器
- 15 一対比較器
- 16 一対比較集計器
- 17 翻訳能力均衡点計算器
- 22 第1のテストセット翻訳正解率計算器
- 23 参照翻訳データベース
- 24 第2のテストセット翻訳正解率計算器
- 25 回帰直線計算器
- 26 翻訳能力計算器

20

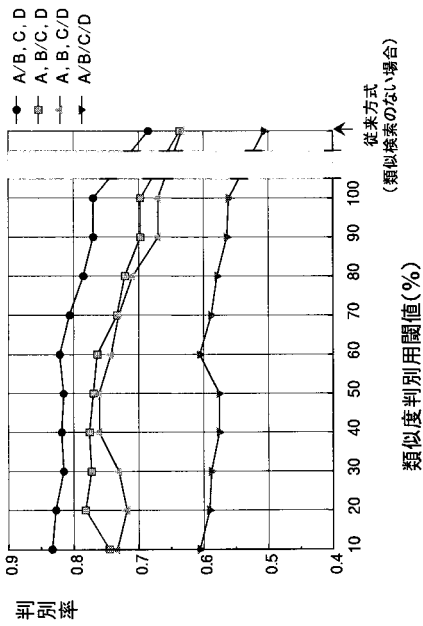
【図1】



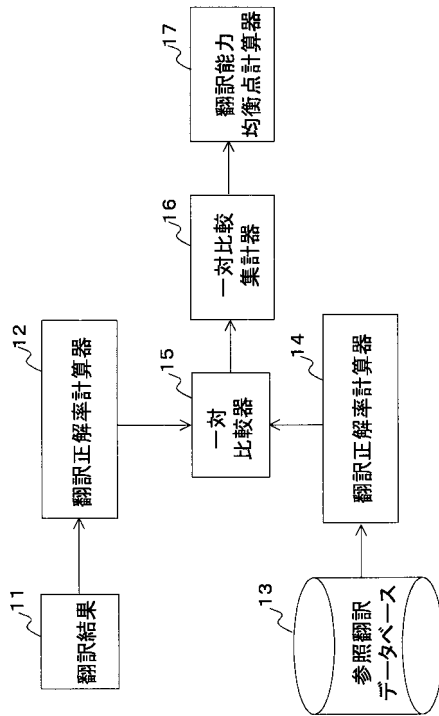
【図2】



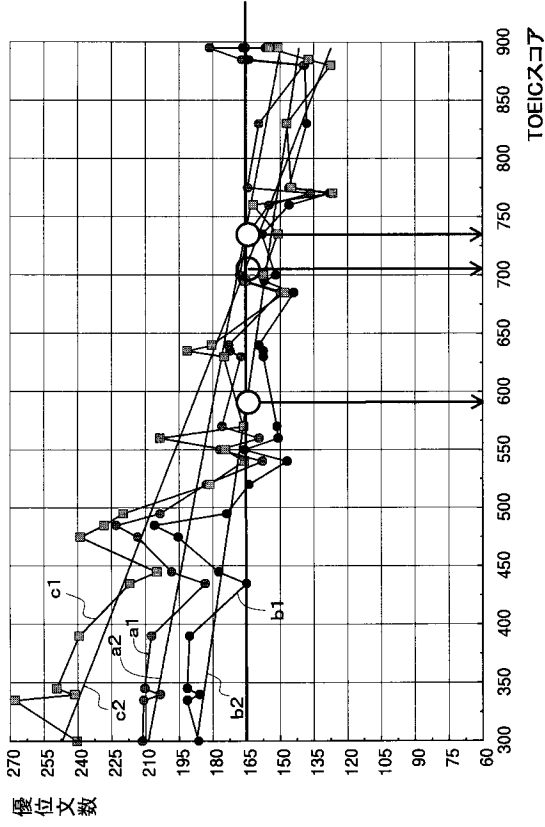
【図3】



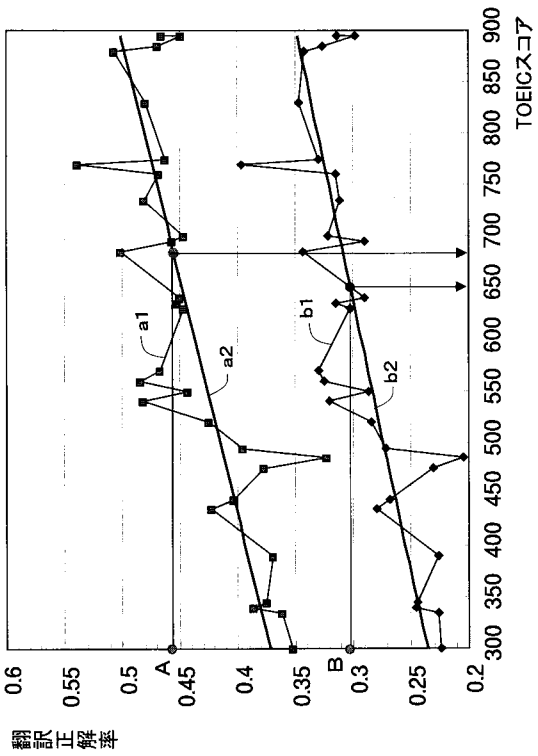
【図4】



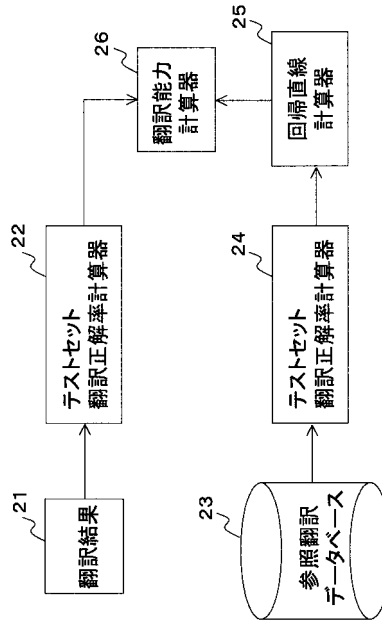
【図5】



【図7】



【図6】



フロントページの続き

(72)発明者 山本 誠一

京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内

合議体

審判長 立川 功

審判官 多賀 実

審判官 長 由紀子

(56)参考文献 「外国語作文における自動検索手法について」, 情報処理学会研究報告99-CH-41-1, Vol.99, p.1-p.6

「英作文における自動検索システムの構築と評価」, 情報処理学会論文誌, Vol.40, No.12, p.4388-p.4395(1999.12.15)

「英作文学習支援システムの作成」, 情報処理学会研究報告2000-NL-139-6, Vol.2000, No.86, p.41-p.48(2000.09.22)

(58)調査した分野(Int.Cl., DB名)

G06F17/21-17/30