

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4756499号  
(P4756499)

(45) 発行日 平成23年8月24日(2011.8.24)

(24) 登録日 平成23年6月10日(2011.6.10)

(51) Int.Cl. F I  
**G 1 0 L 15/08 (2006.01)** G 1 0 L 15/08 3 0 0 B  
**G 0 6 F 17/28 (2006.01)** G 0 6 F 17/28 V

請求項の数 4 (全 21 頁)

<p>(21) 出願番号 特願2005-238236 (P2005-238236)                  (22) 出願日 平成17年8月19日(2005.8.19)                  (65) 公開番号 特開2007-52307 (P2007-52307A)                  (43) 公開日 平成19年3月1日(2007.3.1)                  審査請求日 平成20年3月27日(2008.3.27)</p> <p>特許法第30条第1項適用 平成17年3月8日 社団法人日本音響学会発行の「日本音響学会2005年春季研究発表会 講演論文集-1-」に発表</p> <p>(出願人による申告)平成17年度独立行政法人情報通信研究機構、研究テーマ「大規模コーパスベース音声対話翻訳技術の研究開発」に関する委託研究、産業活力再生特別措置法第30条の適用を受ける特許出願</p>	<p>(73) 特許権者 393031586                  株式会社国際電気通信基礎技術研究所                  京都府相楽郡精華町光台二丁目2番地2                  (74) 代理人 100099933                  弁理士 清水 敏                  (72) 発明者 清水 徹                  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 ウェイキット・ロー                  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 中村 哲                  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</p> <p style="text-align: right;">最終頁に続く</p>
---	--

(54) 【発明の名称】 音声認識結果の検査装置及びコンピュータプログラム

(57) 【特許請求の範囲】

【請求項1】

音声認識処理により所定の入力音声から生成される音声認識結果の単語列を受けて、前記音声認識処理に後続する所定のフレーズベースの統計的自然言語処理の対象として、前記音声認識結果の単語列を構成する単語を受理すべきか否かを検査するための、音声認識結果の検査装置であって、

前記音声認識結果の検査装置は、前記自然言語処理において使用される統計的モデルの学習用のためのコーパスから所定の抽出方法で抽出されたフレーズの集合とともに用いられ、

前記音声認識結果の単語列を構成する単語にはそれぞれ前記音声認識処理により予め信頼度が付与され、

前記音声認識結果の検査装置は、

前記音声認識結果の単語列を構成する各単語に対し、当該単語を含む単語列であって前記音声認識結果の部分単語列をなす単語列のうち前記フレーズの集合内に一致するフレーズを持つ単語列の集合の関数として、前記自然言語処理に対する適合度を付与するための適合度付与手段と、

前記音声認識結果の単語列を構成する単語ごとに、前記適合度付与手段により当該単語に付与された適合度に応じて定められたしきい値と当該単語に付与された前記信頼度との比較により、当該単語を受理すべきか否かを決定するための決定手段とを含み、

前記適合度付与手段は、

前記フレーズの集合と、前記音声認識結果の部分単語列をなす単語列とを照合することにより、前記フレーズの集合内に一致するフレーズを持つ単語列を検出するための照合手段と、

前記音声認識結果の単語列を構成する各単語に対し、前記照合手段により検出された単語列のうち当該単語を含む単語列からなる集合をもとに、前記集合内に含まれる単語列のフレーズ長の最大値、平均値、若しくは和、又は、前記集合に含まれる要素の数に基づき、前記適合度を付与するための手段とを含む、音声認識結果の検査装置。

【請求項 2】

前記自然言語処理は、前記入力音声の言語と所定のターゲット言語とのフレーズベースの統計翻訳処理を含み、

前記フレーズの集合は、前記入力音声の言語と前記所定のターゲット言語とのバイリンガルコーパスから、前記統計翻訳処理のための翻訳モデルを学習する過程で抽出されたフレーズの集合を含む、請求項 1 に記載の音声認識結果の検査装置。

【請求項 3】

前記決定手段は、

前記適合度と前記しきい値とを対応付けて保持するための手段と、

前記音声認識結果の単語列を構成する単語ごとに、当該単語に付与された適合度をもとに、前記保持するための手段により保持された適合度としきい値とにしたがって、当該単語に対する前記しきい値を設定するための手段と、

前記音声認識結果の単語列を構成する単語ごとに、前記設定するための手段により設定された前記しきい値と当該単語に付与された前記信頼度との比較によって、当該単語を受理すべきか否かを決定するための比較手段とを含む、請求項 1 又は請求項 2 に記載の音声認識結果の検査装置。

【請求項 4】

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 3 のいずれかに記載の音声認識結果の検査装置として動作させる、コンピュータプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、音声認識結果の単語列を構成する単語を後続する処理の対象として受理すべきか否かを検査する装置に関し、特に、音声翻訳処理のための音声認識処理による結果の信頼度をもとに、後続する翻訳処理への適合性を考慮して検査を行なう装置に関する。

【背景技術】

【0002】

音声翻訳は、ある言語の音声を受けて、その発話内容を別の言語に翻訳して出力する処理である。音声翻訳を自動化・高性能化する技術は、音声言語処理技術の目標の一つといえる。音声翻訳は、一般的に次の二つの技術によって実現される。すなわち、発話音声から発話された単語又は文章を同定するための音声認識と、同定された単語又は文章を、別の言語の単語又は文章に変換するための機械翻訳とである。

【0003】

音声翻訳性能を向上させるためには、音声認識と機械翻訳との連携を緊密にすることが必要となる。そのための種々の技術が提案されている。例えば、非特許文献 1 には、音声認識と機械翻訳とに、統合された一つの統計モデルを用いる技術が開示されている。非特許文献 2 には、音声認識結果として複数の候補を生成し、それぞれの候補について機械翻訳を行なって、複数の翻訳結果の中から最適な候補を選択する技術が開示されている。

【0004】

高性能な音声翻訳を実現するためには、音声認識性能と機械翻訳性能とをそれぞれ向上させることも必要となる。機械翻訳性能を向上させるための技術として、統計翻訳 (Statistical Machine Translation: SMT) において、翻訳処理の単位を長くする技術が提案されている。例えば、非特許文献 3 には、単語又は接続した複数の単語によって構成

10

20

30

40

50

される「フレーズ」を処理の単位とした、フレーズベースでの統計翻訳技術が開示されている。この技術では、バイリンガルコーパスから統計翻訳用の統計モデル（以下、翻訳モデル）を学習する過程で、バイリンガルコーパスから自動的に抽出されるフレーズを、翻訳の単位として利用する。翻訳の処理の単位をフレーズにすることにより、単語を単位として翻訳を行なう場合より、単語同士が自然に接続する自然な翻訳結果を得ることができる。

【0005】

近年の研究の進展とコンピュータの性能の向上により、かなりの精度での音声認識が実現されるようになった。しかし、雑音、話者の変化、非文法的な発話等、音声認識の障害となる要因が多数あり、十分な音声認識の性能を達成することが難しい。そのため、音声認識結果を翻訳処理する前に、音声認識処理での音声認識誤りを的確に検出し訂正することが重要である。そのための技術として、音声認識結果として得られる単語列（又はそれを構成する単語）が信頼のおけるものであるか否かを評価する技術が開発されている。例えば、非特許文献4では、音声認識の結果として出力される単語グラフと、単語ごとに付与された単語事後確率とをもとに、一般化された単語事後確率（Generalized Word Posterior Probability: GWPP）を算出し、音声認識結果の各単語に関する評価に、GWPPを使用している。

【0006】

【非特許文献1】Y. ガオ、「カップリング対統一：音声 - 音声翻訳のためのモデリング技法」、Eurospeech 2003 予稿集、365 - 368 頁（2004年）（Y. Gao, "Coupling vs. unifying: Modeling technique for speech-speech translation", Proc. of Eurospeech 2003, pp.365-368 (2004)）

【非特許文献2】R. ツァン他、「N - ベスト音声認識仮説を用いて改良された音声言語翻訳」、ICSLP 2004 予稿集、1629 - 1632 頁（2004年）（R. Zhang et al., "Improved Spoken Language Translation Using N-Best Speech Recognition Hypothesis", Proc. of ICSLP 2004, pp.1629-1632 (2004)）

【非特許文献3】E. スミタ他、「用例翻訳（EBMT）、統計翻訳（SMT）、ハイブリッドなど：ATR 音声言語翻訳システム」、IWSLT 2004 予稿集、13 - 20 頁、（2004年）（E. Sumita et al., "EBMT, SMT, Hybrid and More: ATR Spoken Language Translation System" Proc. of IWSLT 2004, pp.13-20 (2004)）

【非特許文献4】Frank K. Soong 他、「単語の評価誤りを最小化するための最適な音響及び言語モデルの重み」、ICSLP 2004 予稿集、441 - 444 頁、（2004年）（Frank K. Soong et al., "Optimal Acoustic and Language Model Weights for Minimizing Word Verification Errors" Proc. of ICSLP 2004, pp.441-444 (2004)）

【発明の開示】

【発明が解決しようとする課題】

【0007】

非特許文献1に記載の技術では、音声認識と機械翻訳とに、統合された一つの統計モデルを用いている。しかしこの統合された統計モデルが、音声認識及び機械翻訳の両方に対して最適なモデルであるとは限らない。そのため、音声認識と機械翻訳とのいずれかの性能が犠牲になる恐れがある。実際、機械翻訳においては、非特許文献3に記載の技術のように、フレーズ等の長い単語列を処理単位として処理が可能であるのに対し、音声認識において、そのような長い処理単位での処理を行なうと、かえって認識性能を低下させるおそれがある。非特許文献2に記載の技術では、機械翻訳の処理を複数の候補について行なうことが必要になる。その分処理の量が増大する。

【0008】

そのため、音声認識と機械翻訳との連携を緊密なものにするためには、音声認識結果が機械翻訳に適したものであるか否かという観点から、音声認識結果を検証する必要がある

10

20

30

40

50

。しかしながら、非特許文献 1 ~ 3 には、このような評価を行なうための技術について、具体的な記載はない。

【 0 0 0 9 】

非特許文献 4 に記載の技術では、音声認識結果の各単語が認識結果として信頼のおけるものであるか否かについて評価が行なわれる。しかしながら、この技術では、「音声認識結果が機械翻訳に適しているか」という観点で評価が行なわれるものではない。そのためこの技術で高く評価された音声認識結果であっても、機械翻訳に全く適さない場合がある。

【 0 0 1 0 】

こうした問題は、音声認識と機械翻訳という組合せに限らず、音声認識とその結果を利用する自然言語処理全般との組合せの間にも生じうる。

【 0 0 1 1 】

それゆえに、本発明の目的は、音声認識結果の信頼度をもとに、後続する自然言語処理に適したものであるか否かという観点を導入して音声認識結果を検査できる、音声認識結果の検査装置を提供することである。

【 0 0 1 2 】

また、本発明の別の目的は、音声認識の性能及び機械翻訳の性能をそれぞれ高く保ちつつ、両者の処理の連携を緊密なものにするための音声認識結果の検査装置を提供することである。

【課題を解決するための手段】

【 0 0 1 3 】

本発明の第 1 の局面に係る音声認識結果の検査装置は、音声認識処理により所定の入力音声から生成される音声認識結果の単語列を受けて、音声認識処理に後続する所定のフレーズベースの自然言語処理の対象として、音声認識結果の単語列を構成する単語を受理すべきか否かを検査するための装置である。音声認識結果の検査装置は、自然言語処理のためのコーパスから所定の抽出方法で抽出されたフレーズの集合とともに用いられる。音声認識結果の単語列を構成する単語にはそれぞれ音声認識処理により予め信頼度が付与される。音声認識結果の検査装置は、音声認識結果の単語列を構成する各単語に対し、その単語を含む単語列であって音声認識結果の部分単語列をなす単語列のうちフレーズの集合内に一致するフレーズを持つ単語列の集合の関数として、自然言語処理に対する適合度を付与するための適合度付与手段と、音声認識結果の単語列を構成する単語ごとに、適合度付与手段により当該単語に付与された適合度に応じて定められたしきい値と当該単語に付与された信頼度との比較により、当該単語を受理すべきか否かを決定するための決定手段とを含む。

【 0 0 1 4 】

音声認識結果の単語列を構成する単語にはまず、適合度付与手段により適合度が付与される。適合度は、その単語を含む音声認識結果の部分単語列をなす単語列のうち、コーパスから抽出されたフレーズの集合内に一致するフレーズを持つものの集合の関数として与えられる。音声認識結果の単語列の中に、フレーズの集合内に一致するフレーズを持つ部分単語列があれば、その部分単語列にはコーパスとの適合性があると考えられる。すなわち、コーパスを利用して行なわれる後続の自然言語処理にその部分単語列が適していると考えられる。決定手段は、各単語を受理すべきか否かを、適合度に応じて定められたしきい値と信頼度との比較によって決定する。したがって、信頼度に基づく各単語の検査に、「コーパスを利用した自然言語処理に適しているか否か」という観点を導入することができ、音声認識の性能及び機械翻訳の性能を低下させることなく、音声認識と自然言語処理との連携を強化できる。

【 0 0 1 5 】

好ましくは、適合度付与手段は、フレーズの集合と、音声認識結果の部分単語列をなす単語列とを照合することにより、フレーズの集合内に一致するフレーズを持つ単語列を検出するための照合手段と、音声認識結果の単語列を構成する各単語に対し、照合手段によ

10

20

30

40

50

り検出された単語列のうちその単語を含む単語列からなる集合をもとに、所定の基準にしたがい適合度を付与するための手段とを含む。

【 0 0 1 6 】

照合手段が音声認識結果の部分単語列をなす単語列とフレーズの集合とを照合することによって、フレーズの集合内に一致するフレーズを持つ単語列が検出される。付与するための手段が、音声認識結果の単語列を構成する各単語に対し、検出された単語列のうちその単語を含む単語列の集合をもとに、適合度を付与する。こうすることにより、一致するフレーズを持つ単語列の検出と、各単語への適合度の付与とを効率的に行なうことができる。

【 0 0 1 7 】

好ましくは、適合度付与手段は、音声認識結果の単語列を構成する各単語に対し、その単語に対する単語列の集合に含まれる単語列の長さの関数として、自然言語処理に対する適合度を付与するための手段を含む。

【 0 0 1 8 】

音声認識結果の単語列を構成する単語ごとに、その単語を含みフレーズの集合内に一致するフレーズをもつ部分単語列の長さの関数として、適合度を付与する。したがってこの適合度により、一致したフレーズが存在する部分の長さにより処理性能が変化するような自然言語処理に、認識結果が適しているか否かを評価できる。

【 0 0 1 9 】

付与するための手段は、音声認識結果の単語列を構成する各単語に対し、その単語に対する単語列の集合に含まれる単語列のうち、その単語列を構成する単語数が最大のものの単語数を、適合度として付与してもよい。

【 0 0 2 0 】

一致したフレーズの長さが長いほど処理性能が向上するような自然言語処理との適合性を、この適合度によって表すことができる。

【 0 0 2 1 】

自然言語処理は、入力音声の言語と所定のターゲット言語とのフレーズベースの統計翻訳処理を含んでもよい。フレーズの集合は、入力音声の言語と所定のターゲット言語とのバイリンガルコーパスから、統計翻訳処理のための翻訳モデルを学習する過程で抽出されたフレーズの集合を含む。適合度付与手段は、音声認識結果の単語列を構成する各単語に対し、その単語に対する単語列の集合の関数として、自然言語処理に対する適合度を付与するための手段を含む。

【 0 0 2 2 】

検出されたフレーズをもとに付与される適合度は、バイリンガルコーパス及び当該コーパスから学習された翻訳モデルとの適合性を表すものとなる。したがって、音声認識結果の検査に、その音声認識結果の各単語がフレーズベースの統計翻訳に適しているかという観点を導入することができ、音声認識と統計翻訳との連携を強化できる。

【 0 0 2 3 】

決定手段は、適合度としきい値とを対応付けて保持するための手段と、音声認識結果の単語列を構成する単語ごとに、その単語に付与された適合度をもとに、保持するための手段により保持された適合度としきい値とにしたがって、その単語に対するしきい値を設定するための手段と、音声認識結果の単語列を構成する単語ごとに、設定するための手段により設定されたしきい値と単語に付与された信頼度との比較によって、その単語を受理すべきか否かを決定するための比較手段とを含んでもよい。

【 0 0 2 4 】

単語ごとに、その単語の適合度に応じて設定されたしきい値とその単語の信頼度とを比較してその単語を受理すべきか否かを決定することにより、音声認識結果を構成する各単語をコーパスとの適合性を考慮して検査できる。

【 0 0 2 5 】

保持するための手段が保持するしきい値は、適合度が高くなるにしたがい低くなるよう

10

20

30

40

50

選ばれてもよい。

【0026】

このようなしきい値が選ばれることにより、低い信頼度の単語でも、その単語を含む部分単語列が、後続の自然言語処理に適していれば受理されるようになる。部分単語列が後続のフレーズベースの自然言語処理に適していれば、その部分に対する自然言語処理に失敗する確率は低くなる。そのため結果として、音声認識処理とその結果を用いて行なわれるフレーズベースの自然言語処理とからなる一連の処理の性能が向上する。

【0027】

本発明の第2の局面に係るコンピュータプログラムは、コンピュータにより実行されると、当該コンピュータを上記したいずれかの音声認識結果の検査装置として動作させる。したがって、上記した音声認識結果の検査装置と同様の効果を得ることができる。

【発明を実施するための最良の形態】

【0028】

以下、図面を参照しつつ、本発明の一実施の形態について説明する。なお、以下の説明に用いる図面では、同一の物に同一の符号を付してある。それらの名称及び機能も同一である。したがって、それらについての説明は繰返さない。

【0029】

[概要]

本実施の形態では、音声認識と自然言語処理の一種であるフレーズベースの統計翻訳との組合せによる音声翻訳において、音声認識結果の各単語について、後続の翻訳処理の対象として受理すべきか否かを検査する。この検査においては、各単語について、その単語の信頼性を表す値（以下、この値を「信頼度」と呼ぶ。）としきい値との比較によって、当該各単語の処理対象としての合否を決める。本実施の形態では、信頼度としきい値との比較によって各単語を検査する問題に、音声認識結果の翻訳処理に対する適合性という観点での評価を導入する。すなわち、音声認識結果が翻訳に適しているかを評価し、この評価の結果に応じて単語ごとにしきい値を変化させる。以下、適合性の指標を「適合度」と呼ぶ。

【0030】

本実施の形態では、フレーズベースの統計翻訳で使用される翻訳モデルの学習過程で生成されるフレーズを、適合度を得るために利用する。ソース言語及びターゲット言語の多数の対訳文からなるバイリンガルコーパスから翻訳モデルを学習する際には、副産物として多数のフレーズが抽出される。それらは一つのフレーズの集合を形成する。この集合の要素であるフレーズは、バイリンガルコーパスにある文の一部であり、バイリンガルコーパスにより規定されるドメインに頻出するフレーズであるといえる。そのため、音声認識結果にこのようなフレーズと一致する部分があれば、その部分はドメインとの間に適合性があると考えられる。さらに、その部分が長ければ長いほど、その部分の適合性は高いと考えられる。そこで、本実施の形態では、音声認識結果に含まれる各単語の適合性に関する評価の尺度として、バイリンガルコーパスから抽出されるフレーズに一致する部分の長さを用いる。また、長さの尺度として単語数を用いる。すなわち、音声認識結果の各単語について、当該単語を含む音声認識結果の部分単語列であって、上記したフレーズの集合中に一致するフレーズを持つものからなる集合を考え、この集合に含まれる部分単語列の長さの最大値によって、当該単語の適合度を定める。なお、以下の説明では、上記フレーズの長さを、「フレーズ長」と呼ぶ。部分単語列もまた一つのフレーズであると考えられるから、部分単語列の長さも同じく「フレーズ長」と呼ぶ。音声認識結果内の、上記フレーズ集合内のフレーズのいずれかに一致する部分単語列を「一致部分」と呼ぶ。ある単語を含む全ての一致部分からなる集合のうち、フレーズ長が最大のもののフレーズ長を、「(その単語の)最大一致長」と呼び、「Lmax」で表す。

【0031】

[構成]

(音声翻訳システムの全体構成)

10

20

30

40

50

図 1 に、本実施の形態に係る音声翻訳システム 30 の構成をブロック図で示す。図 1 を参照して、音声翻訳システム 30 は、所定のバイリンガルコーパス 34 から翻訳モデルを学習するための翻訳モデル学習装置 54 と、翻訳モデル学習装置 54 により学習された翻訳モデルを保持するための翻訳モデル部 52 とを含む。翻訳モデル学習装置 54 は、翻訳モデルの学習の過程で、バイリンガルコーパス 34 からフレーズを抽出する機能を持つ。音声翻訳システム 30 はさらに、翻訳モデル学習装置 54 により抽出されるフレーズを蓄積するためのフレーズデータベース 56 を含む。

#### 【 0 0 3 2 】

音声翻訳システム 30 はさらに、音声翻訳と上述したしきい値の学習処理とからいずれかを選択する操作入力 42 に応じて、翻訳用音声 32 と学習のために予め用意された学習用音声 38A, ..., 38P とから処理の対象となる音声 46 を選択するためのセレクタ 44 と、セレクタ 44 により選択された音声 46 について音声認識処理を行ない、音声認識された単語列と当該単語列の各単語の信頼度とからなる音声認識結果 50 を出力するための音声認識装置 48 と、音声認識結果 50 の各単語の検査を、フレーズデータベース 56 内のフレーズに対するその単語の最大一致長  $L_{max}$  に応じて定められるしきい値とその単語の信頼度とをもとに行ない、音声認識結果 50 と各単語の最大一致長  $L_{max}$  及び合否の情報とからなる検査結果 76 を生成し出力するための検査装置 60 と、第 1 及び第 2 の出力を持ち、セレクタ 44 に連動し、検査結果 76 を操作入力 42 に応じて第 1 又は第 2 の出力に対して出力するためのスイッチ 78 とを含む。

#### 【 0 0 3 3 】

音声翻訳システム 30 はさらに、スイッチ 78 の第 1 の出力に接続されて、スイッチ 78 から検査結果 76 が与えられると、当該検査結果 76 に含まれる音声認識結果 50 に対し各単語の合否に応じた所定の前処理を施すための前処理装置 64 と、翻訳モデル部 52 を用いたフレーズベースの統計翻訳により、前処理装置 64 による前処理済の音声認識結果 66 から翻訳結果 36 への翻訳を行なうための翻訳装置 68 とを含む。本実施の形態では、前処理装置 64 は、検査結果 76 をもとに音声認識結果 50 の単語列から不合格の単語を除去して、前処理済の音声認識結果 66 を生成して出力する。

#### 【 0 0 3 4 】

音声翻訳システム 30 はさらに、スイッチ 78 の第 2 の出力に接続され、スイッチ 78 から学習用音声 38A, ..., 38P についての検査結果 76 が与えられると、当該検査結果 76 と学習用音声 38A, ..., 38P について予め用意された検査結果の正解（以下、「正解検査結果」と呼ぶ。）40A, ..., 40P とから、最大一致長  $L_{max}$  ごとにしきい値を学習するためのしきい値学習部 84 と、しきい値学習部 84 による学習により得られた最大一致長  $L_{max}$  ごとのしきい値からなるしきい値テーブルを保持するためのしきい値テーブル部 82 とを含む。スイッチ 78 はセレクタ 44 と連動している。すなわち、操作入力 42 に応じてセレクタ 44 が翻訳用音声 32 を選択しているときには、検査結果 76 を前処理装置 64 に出力し、セレクタ 44 が学習用音声 38A, ..., 38P を選択しているときには、検査結果 76 をしきい値学習部 84 に出力する。

#### 【 0 0 3 5 】

検査装置 60 は、フレーズデータベース 56 のフレーズに対する音声認識結果 50 の各単語の最大一致長  $L_{max}$  を求め、音声認識結果 50 及び各単語の最大一致長  $L_{max}$  からなる適合度の評価結果 72 を生成し出力するための適合度評価部 70 と、適合度の評価結果 72 に含まれる音声認識結果 50 の各単語の合否を、その単語の信頼度及び最大一致長  $L_{max}$  としきい値テーブル部 82 内のしきい値テーブルとを用いて決定し、適合度の評価結果 72 と各単語の合否とからなる検査結果 76 を生成し出力するための合否決定部 74 とを含む。

#### 【 0 0 3 6 】

（音声認識装置 48 による音声認識結果 50）

図 2 に、音声認識装置 48 が行なう音声認識処理の概要と、その結果音声認識装置 48 により出力される音声認識結果 50 の構成とを模式的に示す。図 2 を参照して、音声認識

10

20

30

40

50

装置 48 は、まず与えられた音声 46 の音声認識処理 100 を行ない、音声 46 を出力しうる発話内容を単語の組合せで表現した単語グラフ 102 を生成する。単語グラフ 102 は、単語に対応するパスからなる経路網で構成されたグラフであり、その各パスには音声認識装置 48 にて付与されたスコアが格納されている。この経路網を通して始点から終点まで進んだ場合の経路が、音声 46 を出力する単語列に対応する。またこの経路網をもとに各単語の事後確率と G W P P とが算出される。そこで音声認識装置 48 は、単語グラフ 102 の経路網から、各単語のスコアに基づいて、経路を選択する処理 104 を行なう。この処理 104 により選択された経路が本実施の形態に係る音声認識結果 50 となる。

【 0037 】

音声認識結果 50 は、単語グラフ 102 から選択された経路の単語  $W_1 \sim W_N$  からなる単語列 106 と、単語  $W_1 \sim W_N$  の各々について単語グラフ 102 から算出される前述の G W P P ( G W P P 1 ~ G W P P N ) からなる系列 108 とを含む。本実施の形態に係る音声翻訳システム 30 では、各単語について算出される G W P P を、音声認識結果 50 における当該単語の信頼度として使用する。

【 0038 】

( フレーズデータベース 56 )

図 3 に、フレーズデータベース 56 ( 図 1 参照 ) のデータ構成を模式的に示す。図 3 を参照して、フレーズデータベース 56 には、ソース言語 ( 図 3 に示す例では日本語 ) の 1 又は複数の単語からなるフレーズ 120 A , ... , 120 Q と、それらのフレーズに対応するターゲット言語のフレーズ 122 A , ... , 122 Q とを含む。フレーズ 120 A , ... , 120 Q 及び 122 A , ... , 122 Q はいずれも、ソース言語又はターゲット言語の文法とは無関係にセグメンテーションされた単語列である。そのため、これらのフレーズは、文節等の文法的な単位との間に関連性を持たない。またそのフレーズ長は、フレーズによって異なる。これらのフレーズはいずれも、バイリンガルコーパス 34 内の文を所定の統計的手法でセグメンテーションすることにより得られるものである。そのためこれらのフレーズはいずれも、バイリンガルコーパス 34 が規定するドメインに頻出するフレーズといえる。

【 0039 】

( 適合度の評価結果 72 及び検査結果 76 )

検査装置 60 による適合度の評価と信頼度を用いた検査とは、単語単位で行なわれる。図 4 ( A ) 及び図 4 ( B ) にそれぞれ、適合度の評価結果 72 及び検査結果 76 のデータ構成を示す。図 4 ( A ) を参照して、適合度の評価結果 72 は、図 2 に示す音声認識結果 50 と、音声認識結果 50 における単語  $W_1 \sim W_N$  の最大一致長  $L_{max}$  からなる系列 140 とを含む。なお、本実施の形態においては、当該単語を含む一致部分からなる集合の要素数が 0 である場合、その単語の最大一致長  $L_{max}$  を 0 とする。図 4 ( B ) を参照して、検査結果 76 は、適合度の評価結果 72 と、単語  $W_1 \sim W_N$  の合否を表す検査結果列 144 とを含む。なお図 4 ( B ) では、検査結果列 144 において、単語  $W_1 \sim W_N$  の合否を「 O K ( 1 ) 」と「 N G ( 0 ) 」とによって表している。

【 0040 】

( しきい値テーブル部 82 の構成 )

図 5 に、しきい値テーブル部 82 に保持されるしきい値テーブルの構成を示す。図 5 を参照して、しきい値テーブル 160 は、それぞれ  $L_{max} = 0$ 、 $L_{max} = 1$ 、 $L_{max} = 2$ 、 $L_{max} = 3$ 、及び  $L_{max} > 3$  に対応する 5 種類のしきい値  $T_0$ 、 $T_1$ 、 $T_2$ 、 $T_3$ 、及び  $T_4$  を含む。しきい値  $T_0$ 、 $T_1$ 、 $T_2$ 、 $T_3$ 、及び  $T_4$  はそれぞれ、図 1 に示すしきい値学習部 84 による学習によって調整される。

【 0041 】

( 適合度評価部 70 の構成 )

図 6 に、適合度評価部 70 の機能的構成をブロック図で示す。図 6 を参照して、適合度評価部 70 は、入力された音声認識結果 50 を記憶するための認識結果記憶部 180 と、認識結果記憶部 180 に格納された音声認識結果 50 中の部分的な単語列 ( 以下、「部分

10

20

30

40

50



文」と呼ぶ。)とフレーズデータベース56内のフレーズ120A, ..., 120Q(図3参照)とを照合して、フレーズデータベース56内に一致するフレーズを持つ部分文を検出するための照合部182と、音声認識結果50の各単語について、照合部182による検出結果をもとに最大一致長Lmaxを算出するためのLmax算出部184と、Lmax算出部184により算出された各単語の最大一致長Lmaxを、音声認識結果50の当該単語についての適合度として付与し、認識結果記憶部180内に適合度の評価結果72を形成するための適合度付与部185と、適合度の評価結果72が形成されると認識結果記憶部180からこれを読み出して出力するための出力部186とを含む。

#### 【0042】

照合部182は、音声認識結果50の単語列106から全ての部分文とその部分文の単語列106内での位置を表す位置標識とを生成するための部分文生成部190と、生成された部分文及びその位置標識を記憶するための部分文記憶部192と、部分文記憶部192内の部分文に一致するフレーズをフレーズデータベース56内で探索し、一致するフレーズがあればその部分文の位置標識を一致部分の位置標識として出力するための探索部194と、探索部194から出力される位置標識を記憶するための一致部分記憶部196とを含む。一致部分記憶部196には、探索部194から出力された全ての位置標識が格納される。すなわち、一致部分記憶部196には、全ての一致部分の位置標識が格納される。

10

#### 【0043】

Lmax算出部184は、一致部分記憶部196内の位置標識をもとに、認識結果記憶部180内の単語列106の各単語について、当該単語を含む一致部分の集合を求める。Lmax算出部184はさらに、当該集合に含まれる一致部分の各々のフレーズ長を求め、そのフレーズ長から、当該単語についての最大一致長Lmaxを求めて、適合度付与部185に対し出力する。ここでは、ある単語について一致部分の集合が求められれば、当該単語に対する最大一致長が求められるという関数関係が存在する。

20

#### 【0044】

(合否決定部74の構成)

図7に、合否決定部74(図1参照)の機能的構成をブロック図で示す。図7を参照して、合否決定部74は、適合度評価部70から出力された適合度の評価結果72を記憶するための記憶部260と、記憶部260内の適合度の評価結果72から処理対象の単語を順次選択し、当該単語のGWPP及び最大一致長Lmaxを順次出力するための単語選択部270と、単語選択部270から出力された最大一致長Lmaxに応じたしきい値をしきい値テーブル160(図5参照)から選択して出力するためのしきい値設定部272と、GWPPとしきい値設定部272の出力するしきい値とを比較し、その結果を順次出力するための比較部274とを含む。合否決定部74はさらに、比較部274により出力された単語 $W_1 \sim W_N$ に関する比較の結果を、単語 $W_1 \sim W_N$ についての検査結果として記憶部260内の適合度の評価結果72に付与することにより、検査結果76を生成するための検査結果付与部264と、検査結果付与部264が検査結果76を生成すると、当該検査結果76を記憶部260から読み出して出力するための出力部268とを含む。

30

#### 【0045】

(しきい値学習部84の構成)

図8に、図1に示すしきい値学習部84の機能的構成をブロック図で示す。図8を参照して、しきい値学習部84は、スイッチ78(図1参照)からしきい値学習部84に与えられる検査結果76に、対応する正解検査結果40A, ..., 40Pを付与して正解付の学習用検査結果282を生成するための正解付与部280と、生成された正解付の学習用検査結果282を記憶するための学習用検査結果記憶部284と、学習用検査結果記憶部284に格納された正解付の学習用検査結果282から、各単語の検査結果の正誤を正解検査結果に基づいて判定してその結果288を出力するための正誤判定部286と、正誤判定の結果288を記憶するための正誤記憶部290とを含む。

40

#### 【0046】

50

しきい値学習部 84 はさらに、正誤判定部 286 による正誤の判定が完了すると、正誤記憶部 290 に格納された正誤判定の結果 288 をもとに、合否決定部 74 (図 1 参照) による検査の性能を評価し、その結果に応じて、正誤記憶部 290 内の情報を用いてしきい値テーブル 160 (図 5 参照) 内のしきい値 T0~T4 の値を調整するためのしきい値調整部 294 と、しきい値が調整されると、調整後のしきい値をもとに学習用検査結果記憶部 284 内の学習用検査結果 282 について検査を再実施するための再検査部 296 とを含む。

【0047】

学習用検査結果記憶部 284 に格納される学習用検査結果 282 は、学習用音声 38A, ..., 38P の各々に関する検査結果 76 と、学習用音声 38A, ..., 38P に対応して  
10  
予め用意された正解検査結果 40A, ..., 40P のうち、当該検査結果に対応するものとの組からなる。正解検査結果は、各単語の各検査結果に対応する正解を含む。正誤判定部 286 は、学習用検査結果記憶部 284 内の情報に変化が生じると、それに応じて正誤判定を開始する機能を持つ。この判定により生成される正誤判定の結果 288 は、正解検査結果の単語列に対応する G W P P の系列 108 (図 2 参照) 及び最大一致長 Lmax の系列 140 (図 4 参照) と、検査結果列 144 により表される各単語の検査結果の正誤を表す正誤標識からなる正誤標識列とを含む。正誤標識はそれぞれ、対応の単語についての検査結果が正解と一致したか否か、一致しなかった場合正解は何であったかを表す。すなわち、検査結果と正解とが一致したことを表す値と、合格 (OK) にすべき単語を不合格 (NG) にしていることを表す値と、不合格 (NG) にすべき単語を合格 (OK) にしている  
20  
ことを表す値とである。しきい値調整部 294 は、最大一致長 Lmax 別に、合否決定部 74 (図 1 参照) による検査の性能を、次に示す信頼度の判定誤り率 (Confidence Error Rate: CER) によって評価する機能を持つ。

【0048】

【数 1】

$$CER = \frac{\text{誤って不合格にした単語数} + \text{誤って合格にした単語数}}{\text{認識された単語の総数}} \times 100\%$$

しきい値調整部 294 はさらに、CER がそれぞれ最小になるように、しきい値 T0~T4 の各々を調整する機能を持つ。  
30

【0049】

[動作]

本実施の形態の音声翻訳システム 30 は、以下のように動作する。

【0050】

(翻訳モデルの学習及びフレーズデータの抽出)

図 1 を参照して、音声翻訳システム 30 は、翻訳又はしきい値の学習を行なう前に、予め翻訳モデルの学習を行なう。すなわち、翻訳モデル学習装置 54 は、パイリンガルコーパス 34 に格納されている対訳文から翻訳モデルの学習を行ない、得られた翻訳モデルを翻訳モデル部 52 に格納する。翻訳モデル学習装置 54 はこの学習過程において、パイリンガルコーパス 34 内に含まれる文から、ソース言語のフレーズ 120A, ..., 120Q  
40  
及びターゲット言語のフレーズ 122A, ..., 122Q (図 3 参照) を生成する。生成されたフレーズの各々は、フレーズデータベース 56 に格納される。

【0051】

(しきい値の学習)

以下、しきい値学習部 84 がしきい値を学習する動作について説明する。図 1 を参照して、しきい値の学習が選択され、その選択に対応する操作入力 42 が音声翻訳システム 30 に与えられると、セレクタ 44 が、処理対象の音声 46 として学習用音声 38A, ..., 38P を選択する。選択された音声 46 は、音声認識装置 48 により音声認識結果 50 に変換される。さらに検査装置 60 により音声認識結果 50 に対する検査が実行され、検査結果 76 (図 4 (B) 参照) が検査装置 60 より出力される。しきい値を学習する場合、  
50

スイッチ 78 は、検査結果 76 をしきい値学習部 84 に出力する。なお、音声認識装置 48 及び検査装置 60 の動作の詳細については、後述する。

【0052】

図 8 を参照して、しきい値学習部 84 の正解付与部 280 に検査結果 76 が与えられると、正解付与部 280 は、与えられた検査結果 76 に、正解検査結果 40A, ..., 40P から、この検査結果 76 に対応するものを付与する。その結果、正解付の学習用検査結果 282 が生成される。生成された正解付の学習用検査結果 282 は、学習用検査結果記憶部 284 に格納される。

【0053】

学習用検査結果記憶部 284 に正解付の学習用検査結果 282 が格納されると、正誤判定部 286 は、検査結果列 144 と対応の正解検査結果との比較により、各単語についての検査結果の正誤を判定し、正誤標識列を生成する。そして正誤判定部 286 は、正誤標識列と、学習用検査結果 282 内の G W P P の系列 108 及び最大一致長 L max の系列 140 とから、正誤判定の結果 288 を形成し、正誤記憶部 290 に格納する。

【0054】

正誤判定部 286 が、学習用検査結果記憶部 284 内にある正解付の学習用検査結果 282 に対する以上の処理を終了すると、しきい値調整部 294 に対して終了信号を与える。しきい値調整部 294 は、正誤記憶部 290 内にある正誤判定の結果 288 をもとに、適合度ごとに C E R を算出し、適合度ごとの C E R がそれぞれ最小になるように、しきい値テーブル 160 内のしきい値 T0 ~ T4 を調整する。

【0055】

しきい値テーブル 160 内のしきい値が変更されると、再検査部 296 は、変更後のしきい値をもとに、学習用検査結果記憶部 284 に格納された学習用の検査結果に関する再検査を実施して、学習用検査結果記憶部 284 内に格納された検査結果を変更する。正誤判定部 286 は、学習用検査結果記憶部 284 内の情報に変化が生じると、再度正誤判定を行なう。以上のような動作を繰返し、C E R が最小化していれば、しきい値の調整を終了する。このような一連の動作により、しきい値テーブル 160 内のしきい値 T0 ~ T4 は、対応する適合度の単語について検査を行なった場合の検査誤りの最も少ないしきい値となる。

【0056】

(音声認識)

再び図 1 を参照して、翻訳用音声 32 に対する音声翻訳に対応する操作入力 42 が音声翻訳システム 30 に与えられると、音声翻訳システム 30 は、音声翻訳を行なうための一連の動作を開始する。この場合、セクタ 44 は、処理対象の音声 46 として翻訳用音声 32 を選択する。

【0057】

セクタ 44 から音声認識装置 48 に音声 46 が与えられると、音声認識装置 48 は図 2 を用いて前述した一連の処理を実行する。すなわち、図 2 を参照して、音声認識装置 48 はまず、与えられた音声 46 の音声認識処理 100 を行なって、単語グラフ 102 を生成する。音声認識装置 48 はさらに、単語グラフ 102 の経路網を構成する各単語について、事後確率を算出し、さらに G W P P を算出する。続いて音声認識装置 48 は、経路選択 104 を行なって単語列 106 を生成すると共に、当該単語列 106 に対応する G W P P の系列 108 を生成する。音声認識装置 48 は、生成された単語列 106 及び対応する G W P P の系列 108 の組を音声認識結果 50 として出力する。

【0058】

(適合度の評価)

以下に、検査装置 60 の適合度評価部 70 (図 1 参照) が、音声認識結果 50 から適合度の評価結果 72 (図 4 (A) 参照) を生成する動作について説明する。図 6 を参照して、適合度評価部 70 に音声認識結果 50 が入力されると、当該音声認識結果 50 は、認識結果記憶部 180 に格納される。部分文生成部 190 は、認識結果記憶部 180 内の音声

10

20

30

40

50

認識結果 50 の単語列 106 (図 2 参照) から、部分的な単語列とその位置標識とからなる部分文を生成し、部分文記憶部 192 に格納する。部分文記憶部 192 に全ての部分文が格納されると、探索部 194 は、部分文の各々について次の処理により、各部分文に一致するフレーズの探索を行なう。すなわち、探索部 194 はまず、処理対象の部分文を選択し、当該部分文と同じフレーズをフレーズデータベース 56 で探索する。同じフレーズが存在すれば、この部分的な単語列は一致部分の一つとなる。この場合、探索部 194 は、処理対象の部分文の位置標識を一致部分記憶部 196 に格納する。同じフレーズがなければ、別の部分文を選択する。探索部 194 による探索が完了すると、一致部分記憶部 196 には、一致部分の各々の単語列 106 内での位置が、格納された位置標識により特定できるようになる。

10

## 【0059】

全ての一致部分の位置標識が一致部分記憶部 196 に格納されると、Lmax 算出部 184 は、認識結果記憶部 180 内の音声認識結果 50 における単語列 106 中の単語  $W_1 \sim W_N$  について次の処理を行ない、各単語の適合度を決定する。すなわちまず、処理対象の単語  $W_n$  を選び、単語  $W_n$  の位置を特定する。単語  $W_n$  の位置が含まれる全ての位置標識を一致部分記憶部 196 の中で選ぶ。この結果、実質的に単語  $W_n$  に関する一致部分の集合が作成される。選ばれた位置標識をもとに、単語  $W_n$  を含む全ての一致部分についてフレーズ長を求める。すなわち、単語  $W_n$  に関する一致部分の集合の要素の全てについてフレーズ長を求める。求められたフレーズ長のうち最大のものを探すことにより最大一致長  $L_{max}$  を求めて、適合度付与部 185 に出力する。

20

## 【0060】

以上の処理により決定された各単語の適合度は、図 6 に示す適合度付与部 185 に与えられる。適合度付与部 185 は、認識結果記憶部 180 内の単語  $W_1 \sim W_N$  にその単語の最大一致長  $L_{max}$  を適合度として付与する。これにより認識結果記憶部 180 内に、図 4 (A) に示す適合度の評価結果 72 が形成される。適合度付与部 185 は、単語  $W_1 \sim W_N$  についてこの処理が完了すると、終了信号を出力部 186 に与える。出力部 186 は、これに应答して、認識結果記憶部 180 内の情報を読み出して出力する。その結果、適合度の評価結果 72 が出力されることになる。

## 【0061】

(合否の決定)

30

以下に、合否決定部 74 (図 1 参照) が、適合度の評価結果 72 と信頼度とをもとに各単語を検査する動作について説明する。図 7 を参照して、合否決定部 74 に適合度の評価結果 72 が入力されると、当該適合度の評価結果 72 は、記憶部 260 に記憶される。単語選択部 270 は、適合度の評価結果 72 における単語  $W_1 \sim W_N$  の中から処理対象の単語  $W_n$  を順次選択し、当該単語  $W_n$  の適合度と  $GWPP$  とをそれぞれ、しきい値設定部 272 と比較部 274 とに与える。

## 【0062】

単語  $W_n$  の適合度が与えられると、しきい値設定部 272 は、しきい値テーブル 160 (図 5 参照) から、その適合度に対応するしきい値を読み出して、比較部 274 に与える。例えば、図 4 (B) に示す  $n$  番目の単語  $W_n$  の最大一致長  $L_{max}$  は 1 である。図 5 に示すしきい値テーブル 160 において、 $L_{max} = 1$  の単語に関するしきい値は  $T1$  である。したがって、しきい値設定部 272 は、しきい値  $T1$  を比較部 274 に与える。比較部 274 は、単語  $W_n$  の  $GWPP$  としきい値とを比較する。単語  $W_n$  の  $GWPP$  がしきい値  $T1$  以上であれば、合格 (OK) を表す第 1 の値を出力する。さもなければ不合格 (NG) を表す第 2 の値を出力する。

40

## 【0063】

以上のようにして比較部 274 から出力される値が、順次図 7 に示す検査結果付与部 264 に与えられると、検査結果付与部 264 は、その値をもとに検査結果列 144 (図 4 (B) 参照) を形成して、記憶部 260 内の適合度の評価結果 72 に付与する。これにより記憶部 260 内に、図 4 (B) に示す検査結果 76 が形成される。検査結果付与部 26

50

4は、この処理が完了すると、終了信号266を出力部268に与える。出力部268は、これに応答して記憶部260内の情報を読み出して出力する。その結果、検査結果76が図1に示すスイッチ78に出力される。スイッチ78は、検査結果76の入力を受けると、その時点で選択されている処理に対応する出力に対して、当該検査結果76を出力する。この場合スイッチ78は、検査結果76を前処理装置64に対して出力する。

【0064】

(前処理及び翻訳)

図1に示す前処理装置64は、スイッチ78から検査結果76の入力を受けると、検査結果76をもとに、前処理を行なう。本実施の形態では、検査結果が不合格(NG)の単語を、音声認識結果の単語列 $W_1 \sim W_N$ から除去して、翻訳用の音声認識結果を生成する。翻訳装置68は、翻訳モデル部52に記憶されているフレーズベースの翻訳モデルを用いた統計翻訳により、翻訳結果36を生成する。

【0065】

[実験]

本実施の形態に係る音声翻訳システム30における単語の検査性能を評価するために、しきい値の学習実験と、当該学習実験により得られたフレーズテーブルを用いての音声認識結果の検査実験とを行なった。

【0066】

学習実験及び検査実験では、出願人により作成された旅行会話基本表現コーパス(Basic Travel Expression Corpus: BTEC)を使用して、学習実験用のデータセット及び検査実験用のデータセットを作成した。以下、これらをそれぞれ「学習セット」及び「テストセット」と呼ぶ。学習セットは、BTECの1016種類の発声からなり、当該データセット全体での単語数は延べ7215単語である。テストセットは、BTECの3060種類の発声からなり、当該データセット全体での単語数は延べ21005単語である。

【0067】

本実験では、音声認識装置48として、出願人において開発されたものを使用した。本実験では、フレーズデータベース56もまた、出願人により予め作成されたものを用いた。このフレーズデータベース56に格納されたフレーズは異なり数約80万フレーズであった。フレーズ長L別の異なり数の内訳を、次のテーブル1に示す。

【0068】

【表1】

テーブル1: フレーズ長別の異なり数

フレーズ長L	異なり数
L > 3	325,500
L = 3	232,090
L = 2	195,050
L = 1	40,950
計	793,590

なお、このフレーズデータベース56において、1フレーズあたりの平均単語数は3.48単語であった。

【0069】

学習実験ではまず、学習セットに含まれる各発声に対する音声認識を行ない、各音声認識結果について単語ごとにLmaxを算出し、さらに検査結果の正解を付与した。そして、正解に対する検査結果のCERが最小となるよう、Lmaxごとにしきい値の学習を行なった。この学習実験の結果得られたしきい値テーブルを次のテーブル2に示す。

10

20

30

40

50

【 0 0 7 0 】

【表 2】

テーブル2: 学習実験結果

Lmax	しきい値
Lmax > 3	0.34
Lmax = 3	0.52
Lmax = 2	0.62
Lmax = 1	0.65
Lmax = 0 (一致せず)	0.89

テーブル 2 を参照して、Lmax が大きな値になるにしたがい、しきい値が低くなった。特に Lmax = 0 の単語に対するしきい値は、Lmax = 1 のいずれの単語に対するしきい値よりもはるかに高い値になった。これにより、本実施の形態の検査装置による検査では、検査結果に次のような傾向が生じることが分かる。すなわち、一致するフレーズが長くなるにしたがい G W P P が低い単語まで合格とする傾向、及び一致するフレーズがない単語については非常に高い G W P P のもののみを許容する傾向である。これらの傾向は、Lmax が大きな単語ほど翻訳に適しているという当初の予測に合致する。

【 0 0 7 1 】

検査実験では、テストセットに含まれる各発話に対する音声認識を行ない、各音声認識結果について、予め単語ごとに検査結果の正解とを付与した。この検査実験では、比較のため、次の 2 種類の方法で検査を行なった。すなわち一方は、本実施の形態の検査方法であり、テーブル 2 に示すしきい値テーブルを使用した検査である。他方は、Lmax に関係なく全単語共通のしきい値を用いた検査である。以下、これらの検査方法をそれぞれ、単に「検査 1」、「検査 2」と呼ぶ。なお検査 2 では、全単語共通のしきい値を 0.62 に設定した。この値は、Lmax によらずにしきい値固定とし、CER が最小となるように調整されたしきい値である。2 種類の検査方法による検査結果についてそれぞれ、Lmax 別に CER を求めた。また、テストセット全体の CER も求めた。テストセットに占める単語の割合と、2 種類の検査方法による検査結果の各々についての CER とを、次のテーブル 3 において Lmax 別に示す。

【 0 0 7 2 】

【表 3】

テーブル3: テストセットの認識結果についての検査実験結果

Lmax	テスト セットに 占める 割合 (%)	検査結果			
		検査1 (しきい値可変)		検査2 (しきい値固定)	
		しきい値	CER (%)	しきい値	CER (%)
Lmax > 3	39.9	0.34	1.2	0.62	1.4
Lmax = 3	31.8	0.52	3.3		3.3
Lmax = 2	22.0	0.62	8.9		8.9
Lmax = 1	5.5	0.65	14.8		15.2
Lmax = 0 (一致せず)	0.8	0.89	6.6		7.2
テストセット全体	100.0	-	4.3	-	4.5

テーブル 3 を参照して、Lmax = 1 の場合、検査 1 及び検査 2 の両方において、Lmax が大きな値になるにしたがい CER が低くなった。また、検査 1 及び検査 2 の両方において、

$L_{max} = 0$  の単語についての C E R は、 $L_{max} = 1$  又は 2 の単語についての C E R より低くなった。

【 0 0 7 3 】

検査 1 及び検査 2 の検査結果についての C E R を比較すると、 $L_{max} > 3$ 、 $L_{max} = 1$ 、及び  $L_{max} = 0$  の単語について、検査 1 の検査結果に対する C E R は、検査 2 の結果に対する C E R より比べて低下した。テストセット全体についての C E R を比較すると、検査 1 及び検査 2 に対する C E R はそれぞれ、4.3 (%) 及び 4.5 (%) であった。したがって、G W P P を評価尺度とする検査において、各単語の合否の判定基準として最適なしきい値は、 $L_{max}$  に応じて変化することが明らかとなった。さらに、学習により最適化されたしきい値を合否の判定基準として使用することにより、 $L_{max}$  が大きな単語及び  $L_{max}$  が小さな単語の両方、並びに検査対象の単語全体において、検査性能に改善が見られることが明らかとなった。このようにして性能の改善された検査方法で音声認識結果を検査し前処理を行なうことにより、音声認識と統計翻訳との連携が強化されるため、音声翻訳の処理の全体的な性能の向上が期待できる。

【 0 0 7 4 】

[ コンピュータによる実現 ]

本実施の形態の音声翻訳システム 30 は、コンピュータハードウェアと、そのコンピュータハードウェアにより実行されるプログラムと、コンピュータハードウェアに格納されるデータとにより実現される。図 9 はこのコンピュータシステム 330 の外観を示し、図 10 はコンピュータシステム 330 の内部構成を示す。

【 0 0 7 5 】

図 9 を参照して、このコンピュータシステム 330 は、コンピュータ 340 と、モニタ 342 と、キーボード 346 と、マウス 348 と、マイクロフォン 370 と、スピーカ 372 とを含む。コンピュータ 340 は、C D - R O M (コンパクトディスク読出専用メモリ) ドライブ 350 及び F D (フレキシブルディスク) ドライブ 352 を有する。

【 0 0 7 6 】

図 10 を参照して、コンピュータ 340 は、C D - R O M ドライブ 350 及び F D ドライブ 352 に加えて、ハードディスク 354 と、C P U (中央処理装置) 356 と、F D ドライブ 352、C D - R O M ドライブ 350、ハードディスク 354、及び C P U 356 に接続されたバス 366 と、バス 366 に接続され、ブートアッププログラム等を記憶する読出専用メモリ (R O M) 358 と、バス 366 に接続され、プログラム命令、システムプログラム、及び作業データ等を記憶するランダムアクセスメモリ (R A M) 360 とを含む。コンピュータ 340 はさらに、バス 366、マイクロフォン 370、及びスピーカ 372 とに接続されたサウンドボード 368 を含む。ここでは示さないが、コンピュータ 340 はさらにローカルエリアネットワーク (L A N) への接続を提供するネットワークアダプタボードを含んでもよい。

【 0 0 7 7 】

コンピュータシステム 330 に音声翻訳システム 30 としての動作を行なわせるためのプログラムは、C D - R O M ドライブ 350 又は F D ドライブ 352 に挿入される C D - R O M 362 又は F D 364 に記憶され、さらにハードディスク 354 に転送される。又は、プログラムは図示しないネットワークを通じてコンピュータ 340 に送信されハードディスク 354 に記憶されてもよい。プログラムは実行の際に R A M 360 にロードされる。プログラムは、C D - R O M 362 から、F D 364 から、又はネットワークを介して、直接に R A M 360 にロードしてもよい。なお、バイリンガルコーパス 34 (図 1 参照) は例えばハードディスク 354 に記憶され、翻訳モデル部 52 の学習時にその必要部分が適宜 R A M 360 に読込まれる。学習により得られる翻訳モデル及びその過程で抽出されるフレーズもまた、例えばハードディスク 354 に記憶され、必要に応じて必要部分が適宜 R A M 360 に読込まれる。

【 0 0 7 8 】

上記プログラムは、コンピュータ 340 を本実施の形態の音声翻訳システム 30 として

動作を行なわせる複数の命令を含む。この動作を行なわせるのに必要な基本的機能のいくつかはコンピュータ340上で動作するオペレーティングシステム(OS)若しくはサードパーティのプログラム、又はコンピュータ340にインストールされる各種ツールキットのモジュールにより提供される。したがって、このプログラムは本実施の形態のシステム及び方法を実現するのに必要な機能全てを必ずしも含まなくてよい。このプログラムは、命令のうち、所望の結果が得られるように制御されたやり方で適切な機能又は「ツール」を呼出すことにより、上記した音声翻訳システム30としての動作を実行する命令のみを含んでいればよい。コンピュータシステム330の動作は周知であるので、ここでは繰返さない。

**【0079】**

上記実施の形態では、各単語についての信頼度としてGWPPを用いた。しかし、検査装置による検査に用いることのできる信頼度は、GWPPには限定されない。音声認識結果の単語列中の各単語に、所定の評価尺度で信頼度が付与されていれば、当該付与されている信頼度を用いて各単語を検査することができる。ただし、その場合、当該付与されている信頼度を用いて検査を行なうために、予めしきい値の学習をLmax別に行なっておくことが必要である。又は何らかの学習によりLmax別に決定されたしきい値をしきい値テーブルとして設定しておくことが必要となる。

**【0080】**

また、フレーズデータベースには、ソース言語及びターゲット言語のフレーズの対が格納されていた。しかし、本発明はこのような実施の形態には限定されない。検査装置60は、ソース言語のフレーズを用いて適合性の評価を行なう。そのため、予めソース言語のフレーズのみを格納したデータベースを用意しておき、当該データベースを用いて適合性の評価を行なうようにしてもよい。

**【0081】**

本実施の形態では、フレーズとの照合の対象として、音声認識結果から全ての組合せで部分文を作成した。しかし、本発明はこのような実施の形態には限定されない。例えば、フレーズデータベースにおいて、各フレーズの出現頻度等の確率情報が得られるならば、当該確率情報に基づき、音声認識結果にフレーズ単位のセグメンテーションを行なってもよい。このようにすることにより、フレーズのオーバーラップを回避して、効率的にLmaxを求めることができる。また、本実施の形態では、適合度としてLmaxを用いたが、さら

**【0082】**

なお、上記した実施の形態では、音声認識結果の各単語について、一致部分の集合を求め、集合に含まれる一致部分のフレーズ長の最大値(最大一致長 $L_{max}$ )に基づいて適合度を定めている。しかし、音声認識結果の各単語と、後続するフレーズベースの自然言語処理との適合度を定める方法はこれには限定されない。あるフレーズの集合が与えられた場合、音声認識結果のある単語を含む部分単語列であって、当該フレーズの集合に含まれるものからなる集合を求めれば、当該集合の関数として適合度を定めることができる。

**【0083】**

上記実施の形態では、この関数は「集合を構成する一致部分のうち、フレーズ長が最大のもののフレーズ長」であった。それ以外にも、例えば、部分単語列の集合の要素数、平均のフレーズ長、各要素のフレーズ長の和、等に基づいて適合度の算出を行なうことができる。

**【0084】**

上記実施の形態では、前処理装置64は、検査結果をもとに不合格の単語を音声認識結果の単語列から排除することにより、翻訳用の単語列を生成した。しかし、本実施の形態の検査装置60による検査結果を用いた前処理は、このようなものには限定されない。例えば、検査結果が不合格の単語が一つでも存在すれば、音声認識結果そのものを棄却するようにしてもよい。この場合、音声認識結果の棄却に回答して所定のエラー信号を発行するようにすると便利である。また例えば、音声認識装置がN-ベストの音声認識結果を出

10

20

30

40

50



力するようにし、前処理装置がN - ベストの音声認識結果の各々に対する検査結果に基づき、翻訳に最適な音声認識結果を選択するようにしてもよい。逆に、翻訳に使用される翻訳モデル及びその学習に用いられるバイリンガルコーパスのドメインが、認識すべき音声の言語及びドメインに適合しているかを評価することも可能である。さらに、その評価をもとに、適切なバイリンガルコーパス及び翻訳モデルを選ぶことにより、認識すべき音声に対する音声翻訳の性能の向上が期待できる。

【0085】

今回開示された実施の形態は単に例示であって、本発明が上記した実施の形態のみに制限されるわけではない。本発明の範囲は、発明の詳細な説明の記載を参酌した上で、特許請求の範囲の各請求項によって示され、そこに記載された文言と均等の意味及び範囲内のすべての変更を含む。

【図面の簡単な説明】

【0086】

【図1】音声翻訳システム30の全体構成を示すブロック図である。

【図2】音声認識装置48による処理の概要と音声認識結果50のデータ構成とを示す模式図である。

【図3】フレーズデータベース56の構成を示す概略図である。

【図4】適合度の評価結果付の音声認識結果72及び検査結果76のデータ構成を示す図である。

【図5】しきい値テーブル160のデータ構成を示す図である。

【図6】適合度評価部70の機能的構成を示すブロック図である。

【図7】合否決定部74の機能的構成を示すブロック図である。

【図8】しきい値学習部84の機能的構成を示すブロック図である。

【図9】本発明の一実施の形態に係る音声翻訳システム30を実現するコンピュータシステムの外観図である。

【図10】図9に示すコンピュータのブロック図である。

【符号の説明】

【0087】

- 30 音声翻訳システム
- 34 バイリンガルコーパス
- 44 セレクタ
- 48 音声認識装置
- 52 翻訳モデル部
- 54 翻訳モデル学習装置
- 56 フレーズデータベース
- 60 検査装置
- 64 前処理装置
- 68 翻訳装置
- 70 適合度評価部
- 74 合否決定部
- 78 スイッチ
- 82 しきい値テーブル部
- 84 しきい値学習部
- 160 しきい値テーブル
- 180 認識結果記憶部
- 182 照合部
- 184 Lmax算出部
- 185 適合度付与部
- 186, 268 出力部
- 190 部分文生成部

10

20

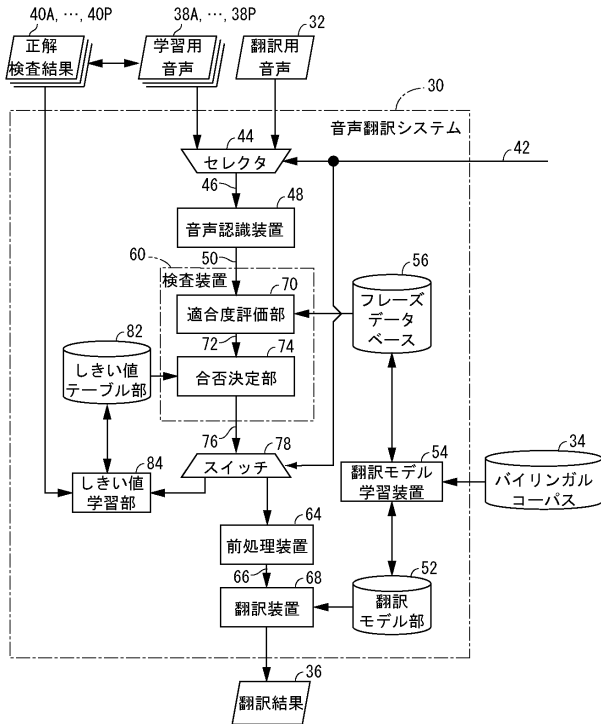
30

40

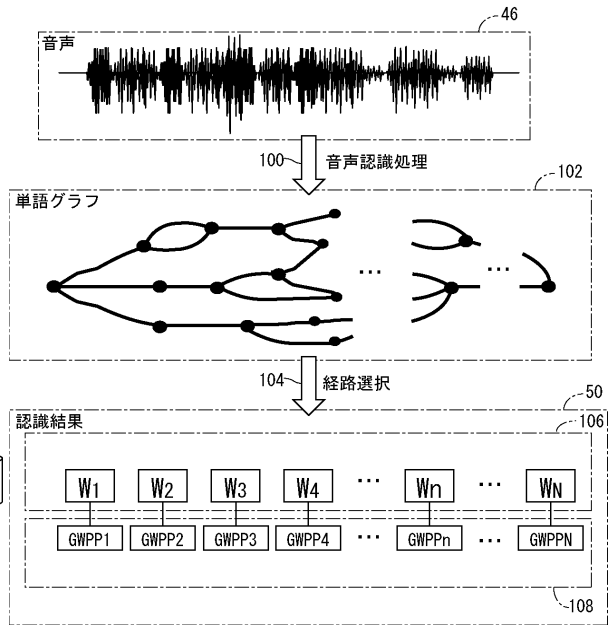
50

- 1 9 2 部分文記憶部
- 1 9 4 探索部
- 1 9 6 一致部分記憶部
- 2 6 0 記憶部
- 2 6 4 検査結果付与部
- 2 7 0 単語選択部
- 2 7 2 しきい値設定部
- 2 7 4 比較部
- 2 8 0 正解付与部
- 2 8 4 学習用検査結果記憶部
- 2 8 6 正誤判定部
- 2 9 0 正誤記憶部
- 2 9 4 しきい値調整部
- 2 9 6 再検査部

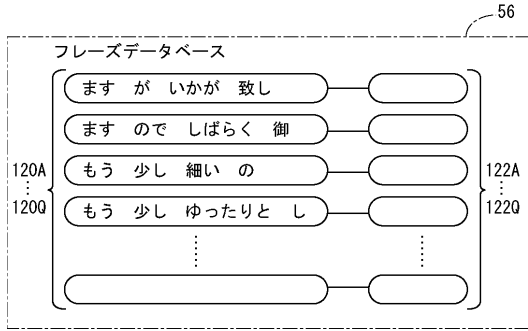
【図1】



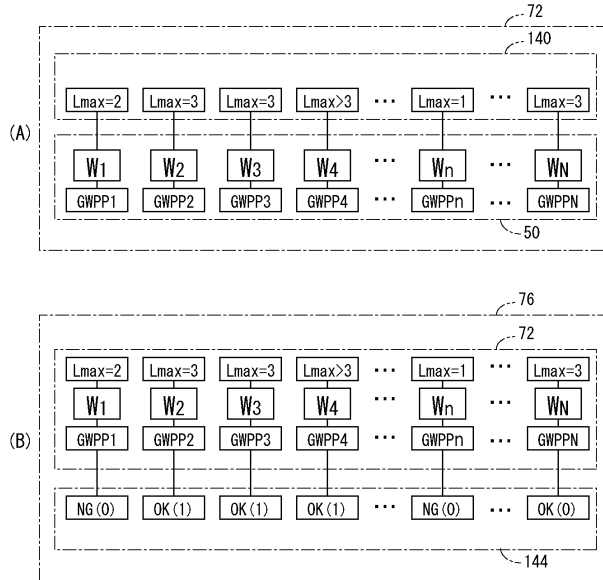
【図2】



【図3】



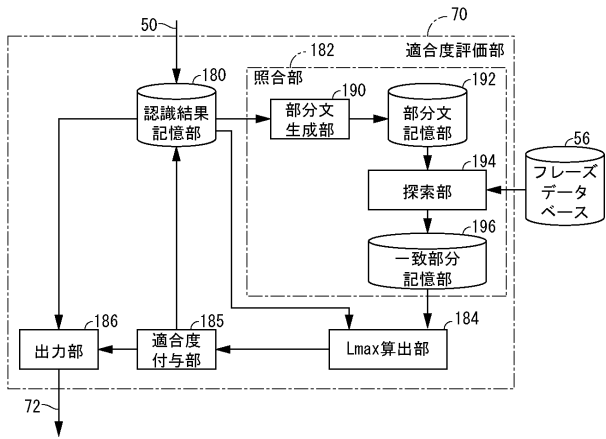
【図4】



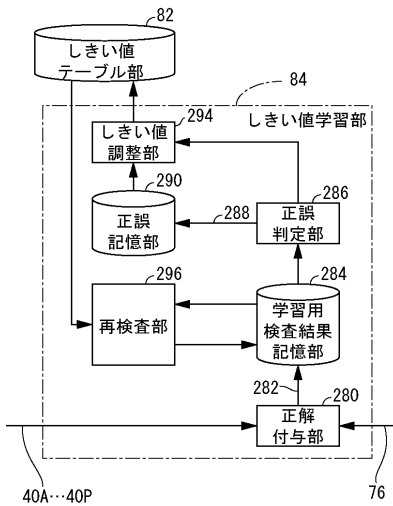
【図5】

Lmax	しきい値
Lmax > 3	T4
Lmax = 3	T3
Lmax = 2	T2
Lmax = 1	T1
Lmax = 0 (一致せず)	T0

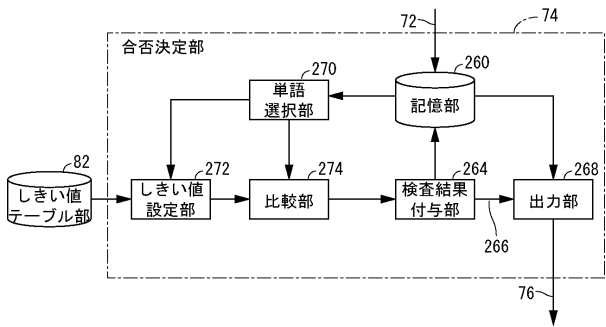
【図6】



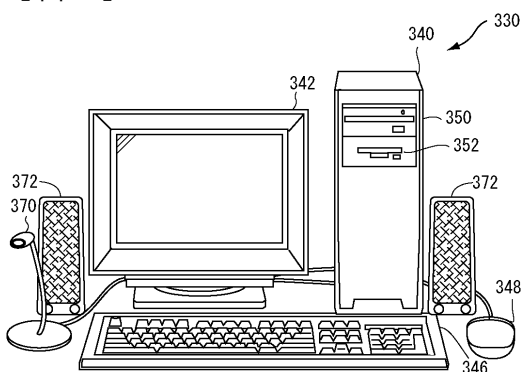
【図8】



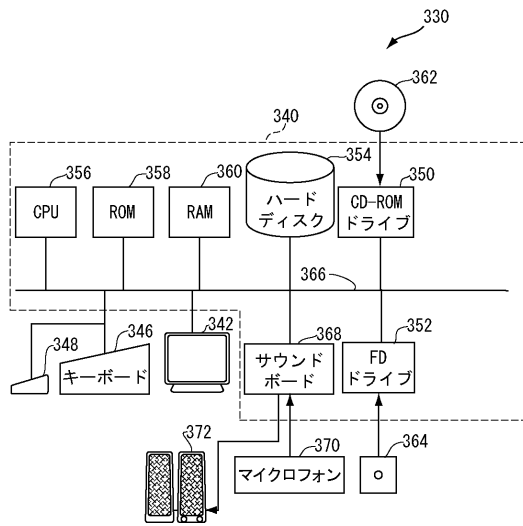
【図7】



【図9】



【図10】



---

フロントページの続き

審査官 毛利 太郎

- (56)参考文献 特開2000-148757(JP,A)  
特開2000-214881(JP,A)  
特開2000-200275(JP,A)  
特開2004-355226(JP,A)  
特表2007-514992(JP,A)

(58)調査した分野(Int.Cl., DB名)

G10L 15/00 - 15/28  
G10L 17/00  
G06F 17/00 - 17/60