

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第5822432号
(P5822432)

(45) 発行日 平成27年11月24日 (2015.11.24)

(24) 登録日 平成27年10月16日 (2015.10.16)

(51) Int.Cl. F I
G O 6 F 17/28 (2006.01) G O 6 F 17/28 6 1 8

請求項の数 17 (全 33 頁)

(21) 出願番号	特願2009-504335 (P2009-504335)	(73) 特許権者	508299441
(86) (22) 出願日	平成19年4月9日 (2007.4.9)		ベシス テクノロジー コーポレイショ ン
(65) 公表番号	特表2009-533728 (P2009-533728A)		B A S I S T E C H N O L O G Y C O R P O R A T I O N
(43) 公表日	平成21年9月17日 (2009.9.17)		アメリカ合衆国 マサチューセッツ州 O 2 1 4 0 ケンブリッジ ケンブリッジ
(86) 国際出願番号	PCT/US2007/008666		パーク ドライブ 150
(87) 国際公開番号	W02007/117652		150 Cambridge Park Drive, Cambridge, M A 02140
(87) 国際公開日	平成19年10月18日 (2007.10.18)		
審査請求日	平成22年4月1日 (2010.4.1)	(74) 代理人	100130111
(31) 優先権主張番号	60/790,076		弁理士 新保 斉
(32) 優先日	平成18年4月7日 (2006.4.7)		
(33) 優先権主張国	米国 (US)		
(31) 優先権主張番号	11/784,449		
(32) 優先日	平成19年4月6日 (2007.4.6)		
(33) 優先権主張国	米国 (US)		

最終頁に続く

(54) 【発明の名称】 機械翻訳の方法およびシステム

(57) 【特許請求の範囲】

【請求項 1】

ソース言語から目的言語への機械翻訳するための方法であって、

(a) コンピュータの翻訳モジュールが、該ソース言語の少なくとも1つの文を受理するステップ；

(b) コンピュータの言語間の文法推定部が、言語間の解析モデルを生成し、該解析モデルは複数の候補となる解析結果を検索およびランク付けするためのパラメータとしてソース言語の単語、目的言語の単語、統語上のラベル及び役割ラベルを含み、検索・ランク付け結果から得られた複数の候補となる解析結果には該ソース言語文の単語と、該ソース言語文の単語に対応する目的言語の単語と、該単語の統語上のカテゴリを示す統語上のラベルと、該単語間の関係を示す役割ラベルを含む、ステップ；

(c) コンピュータの言語間のパーサ部が、該少なくとも1つの文について言語間の解析モデルに基づいて該候補となる解析結果から高いランクの解析結果を選択するステップ；および、

(d) コンピュータの再順序付けモジュールが、コンピュータのツリー・トランスフォーマ処理手段がコンピュータの目的言語パーサ部が生成した候補解析ツリーの逆規則を適用して、該目的言語の語順規則に従った該統語上のおよび役割ラベルを用いた該解析の再配置を行うステップ；を有し、

該言語間の解析モデルの生成において、

(e) パーサ・トレーニング・モジュールが、複数の目的言語文の解析ツリーであって

10

20

、該統語上のラベルで分類されるノードを有する目的言語ツリーバンクを受理するステップ；

(f) パーサ・トレーニング・モジュールが、目的言語文のための候補解析ツリーをランク付けするパラメータを含む、該目的言語ツリーバンクを用いた目的言語解析モデルを生成するステップ；

(g) 目的言語パーサ部が、複数の目的言語文およびそれらのそれぞれのソース言語の同等物を含む並列コーパスから目的言語文を受理するステップ；

(h) 目的言語パーサ部が、該複数の目的言語文の各々の文の候補解析ツリーのランク付けされたリストを該複数の目的言語文への該目的言語解析モデルを適用して生成するステップ；

(i) ツリー・トランスフォーマ処理手段が、該ソースおよび目的言語の言語学的な特性と関連する規則集合を適用することによる該候補解析ツリーを変換するステップ；

(j) 役割ラベラーでの該候補解析ツリーのノードへの言語学的な役割ラベルの割当てであって、該役割ラベルがそのそれぞれの候補解析ツリー内のノードの該言語学的な役割に対応するように候補解析ツリーのノードに対して定義を行うステップ；

(k) 生成物エクストラクタが、役割ラベラーから役割が分類されたツリーのリストを受容し抽出した生成物である文法制約を抽出するステップ；

(l) 言語間の文法推定部による該抽出された文法制約および該並列コーパスのソース言語文を用いた該言語間の解析モデルの推定であって、並列コーパスから各々の目的言語文のソース言語翻訳を受容し、これらの2つの入力ソースから目的言語単語、統語上のラベルおよび役割ラベルを含む候補解析をランク付けするのに十分なパラメータを含む解析モデルを生成するステップ

を含み、

コンピュータの翻訳モジュールが、ステップ (d) において目的言語の語順規則に従って再配置された結果を翻訳結果として出力する機械翻訳方法。

【請求項 2】

前記再配置 (d) が、前記語順規則と関連した規則集合の使用を含む請求項 1 の方法。

【請求項 3】

前記言語間の解析モデルの生成に係る各ステップ (e) ないし (l) において、単一の上部ノードから複数の語末ノードまで下流の分岐関係に配置されている複数のノードまでを解析の対象とする請求項 1 の方法。

【請求項 4】

前記複数のノードが、前記統語上のラベルおよび前記役割ラベルを含む複数の分類されたノード；該分類されたノードの下流に連結される複数の語末の前のノードであって、前記目的言語の単語をそれぞれ含む該語末の前のノード；および、該語末の前のノードに連結される前記複数の語末ノードであって、前記目的言語文の前記単語をそれぞれ含む前記語末ノードを含む請求項 3 の方法。

【請求項 5】

前記推定 (l) がインサイド・アウトサイドアルゴリズム又は C K Y アルゴリズムを使用して行われる請求項 1 の方法。

【請求項 6】

前記請求項 1 における受理 (g) が目的言語コーパスの受理を含み、前記請求項 1 における抽出 (k) が前記目的言語コーパスの前記解析ツリーからの要素間の依存関係に関する統計抽出を含み、前記依存関係が語彙的依存関係を含む請求項 1 の方法。

【請求項 7】

前記言語学的な特性が前記ソースおよび目的言語の形態的な構造を含む請求項 1 ないし 6 のいずれかに記載の方法。

【請求項 8】

前記言語学的な特性が前記ソースおよび目的言語の句構造を含む請求項 1 ないし 6 のいずれかに記載の方法。

10

20

30

40

50

【請求項 9】

目的言語にソース言語を機械翻訳するため、言語間の解析モデルを生成するための方法であって、

(a) パーサ・トレーニング・モジュールでの複数の目的言語文の解析ツリーを有する目的言語ツリーバンクの受理であって、該目的言語文の解析ツリーが、該目的言語での統語上のカテゴリによって分類されるノードを有するような受理；

(b) パーサ・トレーニング・モジュールが、該目的言語ツリーバンクを用いた、目的言語文のための候補解析ツリーをランク付けするパラメータを含む目的言語解析モデルの生成；

(c) 目的言語パーサ部が、並列コーパスから複数の目的言語文の受理であって、該並列コーパスが前記複数の目的言語文およびそれらのそれぞれのソース言語の同等物を含むような受理；

(d) 目的言語パーサ部が、前記複数の目的言語文の各々の文のための候補解析ツリーのランクが付けられたリストを生成するため、該複数の目的言語文への該目的言語解析モデルの適用；

(e) ツリー・トランスフォームでの該ソースおよび目的言語の言語学的な特性と関連している規則集合を適用することによる該候補解析ツリーの変換；

(f) 役割ラベラーでの該候補の解析ツリーのノードへの言語学的な役割ラベルの割当てであって、該役割ラベルがそのそれぞれの該候補の解析ツリー内のノードの該言語学的な役割に対応するように候補解析ツリーのノードに対して定義；

(g) 生成物エクストラクタが、役割ラベラーから役割が分類されたツリーのリストを受容し抽出した生成物である文法制約の抽出；

(h) 言語間の文法推定部による該抽出された文法制約および該並列コーパスのソース言語文を用いた該言語間の解析モデルの推定であって、並列コーパスから各々の目的言語文のソース言語翻訳を受容し、これらの2つの入力ソースから目的言語単語、統語上のラベルおよび役割ラベルを含む候補解析をランク付けするのに十分なパラメータを含む解析モデルを生成；

を含む方法。

【請求項 10】

前記推定 (h) がパーサおよび文法を含む言語間の文法推定モジュールによって生じられる請求項 9 の方法。

【請求項 11】

前記請求項 9 における受理 (c) が目的言語コーパスの受理を含み、前記請求項 9 における抽出 (g) が前記目的言語コーパスの前記解析ツリーからの要素間の依存関係に関する統計抽出を含み、前記依存関係が語彙的依存関係を含む請求項 9 の方法。

【請求項 12】

前記言語学的な特性が前記ソースおよび目的言語の形態的な構造を含む請求項 9 ないし 11 のいずれかに記載の方法。

【請求項 13】

前記言語学的な特性が前記ソースおよび目的言語の句構造を含む請求項 9 ないし 12 のいずれかに記載の方法。

【請求項 14】

ソース言語から目的言語への機械翻訳のためのシステムであって、

該ソース言語の少なくとも1つの文を受理するために構成される翻訳モジュール；

言語間の解析モデルを生成し、該解析モデルは複数の候補となる解析結果を検索およびランク付けするためのパラメータとしてソース言語の単語、目的言語の単語、統語上のラベル及び役割ラベルを含み、検索・ランク付け結果から得られた複数の候補となる解析結果には該ソース言語文の単語と、該ソース言語文の単語に対応する目的言語の単語と、該単語の統語上のカテゴリを示す統語上のラベルと、該単語間の関係を示す役割ラベルを含む言語間の文法推定部；

10

20

30

40

50

該少なくとも１つの文について言語間の解析モデルに基づいて該候補となる解析結果から高いランクの解析結果を選択する言語間のパーサ部；

および、コンピュータのツリー・トランスフォーマ処理手段がコンピュータの目的言語パーサ部が生成した候補解析ツリーの逆規則を適用して、該目的言語の語順規則に従った該統語上のおよび役割ラベルを用いた該解析の再配置を行う再順序付けモジュールを含み、

該翻訳モジュールが、該再順序付けモジュールにおいて目的言語の語順規則に従って再配置された結果を翻訳結果として出力する機械翻訳システムであって、

さらに、

複数の目的言語文の解析ツリーを有する目的言語ツリーバンクを受理するために構成されるパーサ・トレーニング・モジュールであって、該解析ツリーはそれらの統語上のカテゴリによって分類されるノードを有するパーサ・トレーニング・モジュール；

該目的言語ツリーバンクを使用して、目的言語の文のための候補解析ツリーランク付けするパラメータを含む目的言語解析モデルを生成するために構成される該パーサ・トレーニング・モジュール；

並列コーパスから複数の目的言語の文を受理するために構成される目的言語パーサ部であって、該並列コーパスは、該複数の目的言語の文、および、それらのそれぞれのソース言語同等物を含む該パーサ・トレーニング・モジュール；

該複数の目的言語文の各々の文の候補解析ツリーのランク付けされたリストを該複数の目的言語文への該目的言語解析モデルを適用して生成する目的言語パーサ部；

該ソースおよび目的言語の言語学的な特性と関連する規則集合を適用することによって該候補解析ツリーを変換するツリー・トランスフォーマ処理手段；

該候補解析ツリーのノードへの言語学的な役割ラベルの割当てであって、該役割ラベルがそのそれぞれの候補解析ツリー内のノードの該言語学的な役割に対応するように候補解析ツリーのノードに対して定義を行う役割ラベラー；

役割ラベラーから役割が分類されたツリーのリストを受容し抽出した生成物である文法制約を抽出する生成物エクストラクタ；

該抽出された文法制約および該並列コーパスのソース言語文を用いた該言語間の解析モデルを生成する言語間の文法推定部；

並列コーパスから各々の目的言語文のソース言語翻訳を受容し、これらの２つの入力ソースから目的言語単語、統語上のラベルおよび役割ラベルを含む候補解析をランク付けするのに十分なパラメータを含む解析モデル

を含むシステム。

【請求項 １５】

前記言語学的な特性が前記ソースおよび目的言語の形態的な構造を含む請求項 １４ に記載のシステム。

【請求項 １６】

前記言語学的な特性が前記ソースおよび目的言語の句構造を含む請求項 １４ に記載のシステム。

【請求項 １７】

前記請求項 １ ないし ８ のいずれかに記載の機械翻訳方法を実行処理するコンピュータで構成される自動翻訳機。

【発明の詳細な説明】

【技術分野】

【０００１】

本発明は自動機械翻訳に関し、ソースおよび目的言語間のテキストを機械翻訳する方法およびシステムに関する。

【背景技術】

【０００２】

転送ベースの翻訳システムは、ベネットおよびスローカム（１９８５）（「LRC機械翻

10

20

30

40

50

訳システム」、コンピュータ言語学、11巻、2-3号、4月-9月」) Bennett and Slocum, 1985, (“The LRC Machine Translation System,” Computational Linguistics, Volume 11, Numbers 2-3, April-September”)により例証されるように25年以上の間存在してきた。

そのようなシステムは、3つの段階の翻訳を実行する。まず、ソース言語テキストがその統語上の構造を決定するために解析される。第2に、解析ツリーが目的言語の自然の順序に適合するように一セットの統語上の規則に基づいて再配置される。第3に、個々のソース言語の単語が目的言語の単語に変換される。

結果として生じる語列は、ソース言語の文の目的言語の翻訳である。このようなシステムの制限は、すべての情報が規則および辞書の複雑な集合において人手でコード化され、各々の言語対のためにコンピュータを利用した言語学者によって多くの時間および労力を必要とすることである。さらに、規則は予測不可能な方法で相互作用する可能性があり、ある時はあらゆる翻訳を妨げ、またあるときには誤訳を生じることがある。すべての文が翻訳され、正しい翻訳だけが生じられるように、規則間の相互作用を制御することは困難である。これらの困難は、一般に「範囲」および「過生成」の問題と呼ばれる。

【0003】

ソース-チャンネル・モデルと呼ばれる、翻訳への統計的アプローチは、1995年のブラウン他(米国特許5,477,451号)に記載されるBrown et al. 1995 (U.S. Patent 5,477,451)。(ソース-チャンネル・モデルの標準の記述において、ソースおよび目的という用語は、この文献の大部分におけるそれらの使用法とは逆転する;ソース-チャンネル文脈の「目的言語」は翻訳装置に入力される言語を指し、「ソース言語」は出力として生じられる言語である。標準専門用語集に合わせて、これらの用語はソース-チャンネル・モデルを議論する際にこの逆転された意味で使用されるが、この使用法は、ここでソース-チャンネル・モデルに制限される。)

この方法では、情報は訳文の例から自動的に得られ、手で作られた規則および辞書の必要性を取り除く。さらに、各々の考えられる翻訳は、確率値を割り当てられる。したがって、正しい翻訳だけが生じられるように、規則を配置する必要性はない。多数の翻訳が生じられてもよく、例えば最も高い可能性スコアの翻訳が選ばれる。

【0004】

それでもやはり、そのようなシステムは制限を有する。一つの制限は、単語順序がソースおよび目的言語の間でどのように異なるかを記載する歪みパラメータを含むチャンネル・モデルが、基礎的な文法の規則性を捕らえないことである。これらの規則性は、言語間の文法の構造を変換する場合(例えばSVO(主語-動詞-目的語の順序)言語からVSO(動詞-主語-目的語の順序)言語への変換の場合)、非常に有益である。この制限の結果は、訳文の単語の順序を決定するため、広範なかつ計算機的に不経済な調査が必要であることである。さらに、選択された順序が本当に文法的に正しいという保証はほとんどない。

【0005】

これらのアプローチの第2の制限は、典型的にn-gram言語モデルであるソース・モデルを含むことである。そのようなモデルは、すぐ近くの単語間の局所的符合の用途にのみ説明され、全体のフレーズが文法的に正しいかを決定する方法を提供しない。実際に文法的に正しいフレーズのため、表現される関係がもっともらしいかを決定する方法はない。

これらのアプローチが結合された場合であっても、ソースおよびチャンネル・モデルの制限は、非文法的な翻訳または意味が通っていない文法的な翻訳の実質的な可能性を生ずる。そのようなシステムは多くの場合、構文的に解析され、それらの入力の分析が可能であることに依存するので、そのようなふるまいは、下流の自動システムが(例えば事実のデータベースを抽出するため)翻訳されたテキストのさらなる処理を試みる場合、特に問題を含む可能性がある。

【0006】

ヤマダおよびナイト, 2003 (Yamada and Knight) (米国特許出願20030023423号) (US Patent Application 200330023423) は、ソース-チャンネル

10

20

30

40

50

構築を保持する一方、代替チャンネル・モデルを導入することによって、上述した制限を克服する試みる技術を記載する。過去の研究のように、この構築は、単語がソース言語から目的言語に翻訳される方法を記載する確率テーブルを含む。

過去の方法とは対照的に、この構築は、ソースが単語の単純な列というよりも統語上の解析ツリーであるようになる。確率テーブルはそれから、言語間の単語の順序の違いを説明するために、ツリーノードの考えられる置換であるモデルに導入される。追加的な確率テーブルは、ソース言語の対照物を有しない目的言語の単語の挿入であるモデルに導入される。このモデルに基づく復号器は、ヤマダおよびナイト, 2002 (Yamada and Knight, 2002) に記載される(「統語法ベースのMTのための復号器」コンピュータ言語学連合第40回年次総会の予稿)。(" A decoder for syntax-based MT, " Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics ") この復号器は、すぐ近くの単語間の局所的符合の用途のみを説明する n - gram 言語モデルを保持する。

【0007】

また別の技術は、「翻訳が文法的に説明できる」と考え、かつ、全体の翻訳工程を記載する一段式の分析モデルを定める代わりに、上記の方法のソースおよびチャンネル構築から完全に離れる。このアプローチの例は、ウー, 1997 (Wu) (「並列コーパスの確率的反転変換文法および二カ国語分析」コンピュータ言語学, 23(3):377-403) (" Stochastic Inversion Transduction Grammars and Bilingual Parsing of Parallel Corpora, " Computational Linguistics, 23(3):377-403 ")、アルシャウィ, 2001 (Alshawhi) (米国特許第6,233,544号)、および、メラメド, 2004 (Melamed) (「分析による統計的機械翻訳」(コンピュータ言語学連合第42回年次総会の予稿) (" Statistical Machine Translation by Parsing, " Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics ")) を含む。

【0008】

これらのアプローチの一般によくある特性は、ソースおよび目的言語の文を同時に表現する専門的な文法形式論の使用である:ウー(1997)Wuの場合、「反転変換文法」;アルシャウィ(2001)Alshawhiの場合、「先頭の変換部の収集」;メラメド(2004)Melamedの場合、「マルチテキスト文法」である。

【発明の開示】

【発明が解決しようとする課題】

【0009】

以上説明したように、先行技術の欠点に対処する機械翻訳システムおよび方法の必要性が存在する。本発明は、従来の技術が有する基礎的な文法の規則性を捕らえない点や、全体のフレーズが文法的に正しいかを決定することが出来ない点を解決することを目的としている。

【課題を解決するための手段】

【0010】

本発明の一態様において、ソース言語から目的言語までの機械翻訳の方法が提供される。方法は、ソース言語での文の受容と、各々が目的言語の単語で分類された要素、統語上のラベル、および要素の関係を示す役割ラベルを持つ候補解析の言語間の分析モデルでの検索と、統計学的なランク付けとを含む。方法は、文のための統計的に高ランクの解析の選択、および、ソース言語文の言語間の解析を生成するため、目的言語の語順規則による統語上および役割ラベルを用いた解析の再配置をさらに含む。

【0011】

10

20

30

40

50

本発明の別の態様において、ソース言語を目的言語に機械翻訳するための言語間の解析モデルを生成する方法が提供される。方法は、目的言語の文の解析ツリーを有する目的言語ツリーバンクの受容を含み、解析ツリーはそれらの統語上のカテゴリで分類されたノードを有している。目的言語解析モデルは目的言語ツリーバンクを用いて生成され、目的言語文の候補解析ツリーをランク付けするパラメータを含む。目的言語文の第2の集合は並列コーパスから受容され、並列コーパスは目的言語文およびそれらのそれぞれのソース言語の同等物を含んでいる。目的言語解析モデルは、その各々の文の候補解析ツリーのランク付けされたリストを生成するため、目的言語文の第2の集合に適用される。これらの候補解析ツリーは、ソースおよび目的言語の言語学的特性と関連している規則集合を適用することによって変換される。言語学的な役割ラベルは候補解析ツリーのノードに割り当てられ、そこで、役割ラベルは、そのそれぞれの解析ツリーの範囲内でノードの言語学的な役割に対応する。文法制約は、各々の候補解析ツリーの部分から抽出される。言語間の解析モデルはそれから、言語間の解析モデルが候補解析をランク付けするために十分なパラメータを含むように、抽出された文法制約および並列コーパスのソース言語文を用いて推定され、そこにおいて、パラメータは、ソース言語の単語、目的言語の単語、統語上のラベルおよび役割ラベルを含む候補解析の要素と関連がある。

10

【0012】

本発明のさらに別の態様において、ソース言語から目的言語までの機械翻訳のシステムが提供される。システムは、ソース言語の文を受容するために構成される翻訳モジュール、および、複数の候補解析を検索して統計的にランク付けするために構成される言語間の解析モデルを含む。候補の解析は各々ソース言語文の単語に対応する目的言語の単語で分類される要素を有し、統語上のラベルは要素の統語上のカテゴリを示し、役割ラベルは要素間の関係を示す。パーサは、文のための統計的に高ランクの解析を選ぶために構成される。再順序付けモジュールは、ソース言語文の言語間の解析を生成するため、目的言語の語順規則に従って、統語上のおよび役割ラベルを用いて少なくとも1つのランク付けされた解析を再配置するために構成される。

20

【0013】

さらに別の態様において、目的言語にソース言語を機械翻訳するため、言語間の解析モデルを生成するシステムが提供される。このシステムは、目的言語文の解析ツリーを有する目的言語ツリーバンクを受容するために構成されるパーサ・トレーニング・モジュールを含み、解析ツリーはそれらの統語上のカテゴリで分類されたノードを有する。パーサ・トレーニング・モジュールは、目的言語文のための候補解析ツリーをランク付けするパラメータを含む目的言語ツリーバンクを用いて、目的言語解析モデルを生成するために構成される。目的言語パーサは目的言語解析モデル、および、並列コーパスからの目的言語文の第2の集合を受容し、並列コーパスは、第2の複数の目的言語文およびそれらのそれぞれのソース言語の同等物を含む。目的言語パーサは、その各々の文のための候補解析ツリーのランク付けされたリストを生成するため、目的言語解析モデルを目的言語文の第2の集合に適用するように構成される。ツリー・トランスフォーマは、ソースおよび目的言語の言語学的な特性と関連した規則集合を適用することによって候補解析ツリーを変換するために構成される。役割ラベラーは、言語学的な役割ラベルを候補解析ツリーのノードに割り当てるように構成され、役割ラベルはそのそれぞれの解析ツリー内のノードの言語学的な役割に対応する。抽出モジュールは、各々の候補解析ツリーの部分から文法制約を抽出するために構成される。言語間の文法推定部は、言語間の解析モデルが候補解析をランク付けするのに十分なパラメータを含むように、並列コーパスの抽出された文法制約およびソース言語文を用いて言語間の解析モデルを予想するために構成され、そこにおいて、パラメータは、ソース言語の単語、目的言語の単語、統語上のラベルおよび役割ラベルを含む候補解析の要素と関連する。

30

40

【0014】

さらに別の態様において、ソース言語を目的言語に機械翻訳する製品が提供される。製品は、そこにおいて具体化されるコンピュータ可読のプログラムコードを有するコンピュ

50

ータが使用可能な媒体を含む。コンピュータが使用可能な媒体は、翻訳モジュールで、ソース言語の少なくとも1つの文を受容するためのコンピュータ可読のプログラムコードを有する。プログラムコードはまた、言語間の解析モデルで複数の候補解析の検索および統計的なランク付けも提供し、各々の解析はソース言語文の単語に対応する目的言語の単語で分類される要素を有し、統語上のラベルは要素の統語上のカテゴリを示し、役割ラベルは要素間の関係を示している。プログラムコードはまた、少なくとも1つの文のための統計的に高ランクの解析を選択するため;および、ソース言語文の言語間の解析を生成するため、目的言語の語順規則によって統語上のおよび役割ラベルを用いて、解析を再配置するために提供される。

【0015】

10

本発明のさらに別の態様において、ソース言語を目的言語に翻訳する自動翻訳機械が提供される。翻訳機械は、ソース言語の文を受容する手段および候補解析を検索し統計的なランク付けする手段を含む。候補の解析はソース言語文の単語に対応する目的言語の単語で分類される要素を各々有し、統語上のラベルは要素の統語上のカテゴリを示し、役割ラベルは要素間の関係を示している。文のための統計的に高ランクの解析を選ぶ手段、および、ソース言語文の言語間の解析を生成するため、目的言語の語順規則に従って統語上のおよび役割ラベルを用いて解析を再配置するための手段もまた提供される。

【発明を実施するための最良の形態】

【0016】

20

以下の詳細な説明において、この一部を形成し、実例として本発明が実施される特殊な実施態様を示される添付の図面が参照される。これらの実施態様は、当業者が本発明を実施できるのために十分詳細に記載され、他の実施態様が利用されてもよいと理解される。構造上の、手順のおよびシステムの変化が本発明の趣旨および範囲から逸脱することなく、作成されてもよいこともまた理解される。以下の詳細な説明はしたがって限定的なものではなく、本発明の範囲は添付の請求の範囲およびそれらの同等物によって定められる。説明の明瞭さのため、添付の図面において示される類似した特徴は、類似した参照番号で指示され、図面の代替的な実施態様で示す類似した特徴は、類似した参照番号によって示される。

【0017】

30

図を参照すると、本発明の実施態様を示される。ベネットおよびスローカム(1985) Bennett and Slocumのような転送ベースのシステムとは異なり、これらの実施態様は、自動的に必要な詳細な情報をデータから学習し、例えば、各々の特定の言語のためにカスタマイズされる規則のような大量の手書きの規則の必要性を取り除く。これらの実施態様は各々のソース言語文のための広範囲にわたる考えられる翻訳を考慮し、確率を各々の候補に割り当て、それによって、上記で検討された従来のルールベース・システムの範囲および過生成の問題を克服する。これらの実施態様によって使用される比較的少ない規則は一般的な性質であり、興味のあるあらゆる特定の言語の文法入門を参照することによってすぐに決定することができる。

【0018】

40

本発明の実施態様は、ソース言語(例えばアラビア語)の単語列を目的言語(例えば英語)の単語列に翻訳するためのコンピュータによって動作する方法およびシステムを含む。これらの実施態様は、パラメータ値が少なくとも2つのデータ資源:並列コーパスおよび目的言語ツリーバンクから推定される統計的アルゴリズムに基づく。パラメータ推定値をさらに細かく区別するため、目的言語文(例えば目的言語コーパス)の広範な収集が任意に使われてもよい。これらの実施態様はまた、様々な規則集合に含まれる少量の一般的な言語学的な情報を使用してもよい。

【0019】

図32に示すように、解析ツリーを生じるためにソース言語文を解析することによって翻訳は達成される。典型的な統語上の解析ツリーのように、結果として生じる解析は、統語上のカテゴリで分類されるノードを含む。典型的な解析ツリーと異なり、葉を直接支配

50

しているノードの追加的な層が存在する;これらの語末の前のノードは、目的言語の単語で分類される。また、典型的な解析ツリーとは異なり、各々のノード- 語末の前のノードを除いて-は、その親ノード(例えばSBJをサブジェクトとするような)と関連してその言語学的な関係(すなわちその『役割』)を示す第2のラベルを含む。この種の解析ツリーは、言語間の解析ツリー(または、例えばその部分を指すために、単純に言語間の解析)とここで呼ばれる。そのような言語間の解析ツリーの例は、図2において示される。

【0020】

言語間の解析ツリーがソース言語文のために得られるならば、単純な集合の規則が目的言語の単語順序の規則に従ってノードを再配置するために使用される。この再配置は、言語間の解析ツリーにおける言語学的な関係(役割)ラベルによって容易にされる。ノードが再配置されるならば、規則の第2の集合が、再配置されたツリーによって提供される順序で目的言語の単語を抽出するために適用されてもよい。結果として生じる単語列は、ソース言語文の目的言語の翻訳である。

【0021】

代替的に、翻訳された出力が(例えば情報抽出システムによって)さらなる自動処理を意図する場合、再配置されたツリー- そのソース言語葉が取り除かれた-は、単語列の代わりに出力され、それによってソース言語文の目的言語の解析を提供する。

【0022】

例えば上述したブラウン他1995(Brown et al.,)のアプローチのような統計的モデルとは異なり、本発明の実施態様は、言語間の単語順序を変換する場合、文法的に正しい規則性を説明する(例えばSVO言語からVSO言語への変換)。さらにこれらの実施態様は、先頭の単語およびそれらの修飾子間の長距離の依存関係を説明する。これらの特性は、特にソースおよび目的言語が構造的に異種である場合、より文法的に正しくかつ意味的に首尾一貫した傾向の翻訳を導く。加えて、文法的に正しい規則性は検索空間を減少するのに適し、よりコンピュータ的に効率が良い翻訳処理を生じる。

【0023】

上述したヤマダおよびナイト, 2003(Yamada and Knight)のような統語法ベースのチャンネル・モデルとは異なり、本発明の実施態様は、別々のソースおよびチャンネル構成部分を有しない一段式の統計的モデルに基づく。それは、特定の置換または単語挿入のための確率テーブルを必要とせず;代わりに、単語の順序を変換し、少量の一般的な言語学的な情報を適用することによって挿入をより効率的に扱う。さらに、それはヤマダおよびナイト, 2002(Yamada and Knight)のように、専門の復号器を必要としない。復号化は、コリンズ, 97(Collins)の「統計的分析のために3つの生成語彙的モデル」(ACLの第35回年次総会の予稿) "Three Generative, Lexicalised Models for Statistical Parsing," Proceedings of the 35th Annual Meeting of the ACL"において開示されるものに類似した、統語分析のために一般に使用されるタイプ、例えば、LPCFG(語彙的確率文脈自由文法: Lexicalized Probabilistic Context-Free Grammar)のような従来の分析モデルを使用して実行されてもよい。加えて、モデルの語彙的特性は、別々のn-gram言語モデルの必要性を取り除き、単語間の長距離の依存関係を容易にする。

【0024】

ウー, 1997(Wu)、アルシャウィ, 2001(Alshawhi)およびメラメド, 2004(Melamed)の「分析して翻訳する」アプローチとは異なり、同期的な文法は必要とされない;文法は、まさに標準の単一言語の分析のように単純にLPCFGである。両方の言語の単語順序を同時に説明する必要性はない。むしろ、一般的な言語学的な規則の小さい集合と共に、言語間の解析において埋め込まれた統語上の関係情報は、目的言語翻訳が自然の標準的な順序において合成されることを可能にする。

【0025】

本発明の実施態様は、その統計的モデルを推定する際の以前の方法に追加的な利点を提

10

20

30

40

50

供する。過去の統語法ベースのモデルは、並列コーパスの目的言語側の自動的な解析により得られるツリーが正しいとみなす;実際は、存在しているパーサは相当数のエラーを作成する。さらに以前のモデルは、同一の言語学的な依存関係がソース言語文およびそれらの目的言語の翻訳において存在するとみなす。実際は、多くの言語学的な依存関係がソースおよび目的言語で分配される一方、相当数の差異がある。そのような仮定は、統計的翻訳モデルを推定する際の多数のアライメント・エラーを導く可能性がある。本発明の実施態様は、解析の不正確さおよび言語間のミスマッチな依存関係を説明し、より高品質な配列を導き、最終的により正確な翻訳を導く。

【0026】

データの貧弱さは、多くの場合統計的モデルを推定する際の、特に多数の個別の単語を含むモデルにおいて制限因子である。以前のソース-チャンネル構築は、大量の単一言語のテキストを使用して別々に推定された別々のn-gram言語モデルを導入することによって、部分的にこの制限を克服する。しかしながら、これらのモデルは、隣接した少数の単語の範囲に制限される。本発明の実施態様は、その分析モデルを細かく区別するために、目的言語の文の大量な収集を任意に利用することによってこの制限を克服する。別々のn-gramソース・モデルとは異なり、これらの実施態様の細かい推定は、解析モデルそれ自体の一部であり、したがって広く分離された単語間の依存関係を説明することができる。

【0027】

これらの多くの実施態様の特性は、それらのパラメータ値が並列コーパス、および、例えば、マーカス他, 1994 (Marcus et al.,) によるペンシルベニア大学のツリーバンク「英語の広範な注釈コーパスの構築:ペンシルベニア大学・ツリーバンク」、コンピュータ言語学、19巻、No.2」"Building a Large Annotated Corpus of English: The Penn Treebank," Computational Linguistics, Volume 19, Number 2"のような目的言語の統語上のツリーバンクのみを使用して推定されてもよいということである。ソース言語のためにツリーバンクは必要とされない。この特性は、ツリーバンクは英語には直ちに利用できるが、多くの他の言語には直ちに利用できないので、いくつかの外国語テキストを英語に翻訳するのに特に都合が良い。同様に、上記したように、広範な目的言語コーパスは、目的言語のパラメータをさらに細かく区別するために使用されてもよい。目的言語が英語である場合、実質的には無制限の量のテキストが利用できる。

【0028】

上に言及されるように、翻訳システムの出力がさらなる自動処理を意図するものである場合(例えば情報抽出システムによる)、これらの実施態様はソース言語文の目的言語解析を生じることができる。したがって、下流の処理部による非文法的な翻訳の解析で一般に直面する問題は、減少されるまたは取り除かれる。

【0029】

この開示において使われる場合、用語「コンピュータ」は、ワークステーション、サーバ、パーソナルコンピュータ、ノート・パソコン、パーソナル携帯情報機器(PDA)、ポケットPC、多機能電話、または、あらゆる他の適切なコンピュータを含むものとする。

【0030】

本発明を具体化しているシステムおよび方法は、以下を含むあらゆる適切な言語および技術でプログラムされることができるが、これらに制限されない。

C++; ビジュアル・ベーシック; ジャバ; VBスクリプト; Jスクリプト; BCMAスクリプト; DHTML; XML; CGI; ハイパーテキスト・マークアップ言語(HTML)、アクティブ・サーバーページ(ASP); およびジャバスクリプト。

以下を含むあらゆる適切なデータベース技術は使用されるが、これらに制限されない。

Microsoft AccessおよびIBM AS 400。

【0031】

ここで図を参照して、本発明の実施態様はより完全に記載される。

図 1 に示すように、例示的な翻訳システムは、3つの最高位の構成部分から構成されている：翻訳トレーニング・プログラム 110、言語間の解析モデル 108 および実行時翻訳プログラム 106。

【0032】

特定の実施態様において、翻訳トレーニング・プログラムは目的言語ツリーバンク 104 および並列コーパス 102 を受容し、言語間の解析モデル 108 を生じる。目的言語ツリーバンクは、統語上のラベルで注釈が付けられる目的言語文の解析を含み、解析の要素の統語上のカテゴリを示す。適切な目的言語ツリーバンクの例は、上述したペンシルバニア大学のツリーバンク（マークス他，1994 Marcus et al.）である。当業者に公知のいかなる数の統語上のラベルが、本発明の実施態様において使われてもよい。適切な統語上のラベルの例は、例えば、以下の表 1 および 2 において示されるペンシルバニア大学のツリーバンクによって使用されるものを含む。

【0033】

【表 1】

ペンシルバニア大学のツリーバンク・タグセット	
1. CC 等位接続詞	25. TO へ
2. CD カージナル数	26. UH 間投詞
3. DT 限定詞	27. VB 動詞、基本形
4. EX there 構文	28. VBD 動詞、過去時制
5. FW 外国語の単語	29. VBG 動詞、動名詞/現在分詞
6. IN 前置詞/下位	30. VBN 動詞、過去分詞 218z 結合
7. JJ 形容詞	31. VBP 動詞、非三人称単数現在
8. JJR 形容詞、比較級	32. VBZ 動詞、三人称単数現在
9. JJS 形容詞、最上級	33. WDT wh-限定詞
10. LS リストアイテム・マーカ	34. WP wh-代名詞
11. MD 法助動詞	35. WP\$ 所有 wh-代名詞
12. NN 名詞、単数または複数	36. WRB wh-副詞
13. NNS 名詞、複数	37. # ポンド記号
14. NNP 固有名詞、単数	38. \$ ドル記号
15. NNPS 固有名詞、複数	39. . 文の最後の句読点
16. PDT 前置限定詞	40. , カンマ
17. POS 所有格末尾	41. : コロン、セミコロン
18. PRP 人称代名詞	42. (左の括弧記号
19. PP\$ 所有代名詞	43.) 右の括弧記号
20. RB 副詞	44. " 直線の二重引用符
21. RBR 副詞、比較級	45. ` 左の一重引用符
22. RBS 副詞、最上級	46. " 左の二重引用符
23. RP 助詞	47. ' 右の一重引用符
24. SYM 記号	48. " 右の二重引用符

10

20

30

40

50

【 0 0 3 4 】

【表 2】

ペンシルバニア大学のツリーバンクの統語上のタグセット	
1. ADJP	形容詞句
2. ADVP	副詞句
3. NP	名詞句
4. PP	前置詞句
5. S	単純な平叙節
6. SBAR	従位接続詞によって導入される節または 0 (下記参照)
7. SBARQ	wh-語または wh-句によって導入される直接疑問
8. SINV	主・従反転を有する平叙文
9. SQ	wh-語または wh-句を排除する SBARQ の従構成要素
10. VP	動詞句
11. WHADVP	wh-副詞句
12. WHNP	wh-名詞句
13. WHPP	wh-前置詞句
14. X	未知のまたは不確実なカテゴリの構成要素

10

20

【 0 0 3 5 】

並列コーパス 1 0 2 は、ソース言語文および目的言語翻訳の文が整列された対を含む。加えて、広範な目的言語コーパス 1 1 6 は、言語間の解析モデルをさらに細かく区別するためにトレーニング・プログラムによって任意に受容されてもよい。広範な目的言語コーパスは使用される場合、一まとまりの注釈が付かない目的言語文を含む。

【 0 0 3 6 】

30

実行時翻訳プログラム 1 0 6 は言語間の解析モデル 1 0 8 および一まとまりのソース言語文 1 0 0 を受容し、目的言語翻訳 1 1 2 を生成する。任意に、実行時翻訳プログラムは目的言語解析ツリー 1 1 4 を生じる。目的言語解析ツリー 1 1 4 は、例えば図 2 において示されるように目的言語翻訳の統語上の解析を含む。加えてまたは代替的に、ソース言語の単語は、以下にさらに詳細に検討されるようにさらなる処理のために使われる可能性のある目的言語の解析を残すため、解析ツリー 1 1 4 から取り除かれてもよい。

【 0 0 3 7 】

(翻訳トレーニング・プログラム)

ここで図 3 に戻って、翻訳トレーニング・プログラム 1 1 0 が詳細に記載される。

特定の実施態様に示すように、3つの構成部分は、目的言語文を構文的に解析するメカニズムを提供する:パーサ・トレーニング・プログラム 1 2 0、目的言語分析モデル 1 2 2 および目的言語パーサ 1 2 4。構成部分のこの組合せは、LPCFG (語彙的確率文脈自由文法)パーサを実行する。コリンズ, 1 9 9 7 (Collins)により例えば開示される LPCFGパーサが、このために使われてもよい。必要ではないが、特定の実施態様において、パーサは、単一の最高スコアリング解析のみよりも、考えられる解析のランク付けn-bestリストを生じる。例えば、シャーニアクおよびジョンソン, 2 0 0 5 (Charniak and Johnson)「粗いものから細かいものへのn-best分析およびMaxEntの区別的な再ランク付け」:コンピュータ言語学第43回年次総会の予稿"Coarse-to-fine n-best parsing and MaxEnt discriminative reranking," Proceedings of th

40

50

e 43rd Annual Meeting of the Association for Computational Linguistics”に開示されるように、解析ツリーのn-bestリストを生じる適切な技術は、周知である。

【0038】

並列コーパス102からの目的言語文は、目的言語パーサ124に到達する。目的言語分析モデル122を適用して、目的言語パーサは、各々の目的言語文のための候補解析のリストを生じる。このリストは、相当高いスコア（例えば最も高いスコアリング解析の50%以内）を有するすべての他の解析と同様に最も高いスコアリング解析を含む。

【0039】

候補解析のリストは、ソース言語により近い形状に各々のツリーを変換するため、単純な、一般的な、言語学的な規則を適用するツリー・トランスフォーマ126によって受容される。（言語学的な規則が単純かつ一般的である一方、それらは言語により特定であり、各々の言語対のために異なる規則が必要である。ここで提供される例において、ソース言語はアラビア語であり、目的言語は英語である。）これらの変換は単語順序の差異を説明せず、むしろ1)形態学、および2)フレーズ構造における差異を説明する。規則は、厳密に決定的ではない:いくつかの代替的な目的言語構造物のうちの1つを用いて、ソース言語要素が表現可能である場合、ツリー・トランスフォーマ126は、各々の可能性のための代替的な出力を生じる。したがって、入力として受容される各々のツリーのために、ツリー・トランスフォーマ126は、出力として代替的な変換されたツリーのいくつかを典型的に生じる。以下の表3は、プロトタイプ（例示的な）のアラビア語から英語への翻訳システムにおいて使用される一セットの10の規則を記載する。

【0040】

10

20

【表 3】

規則 #	規則
1001	NP が冠詞(a, the)または所有代名詞(his, her, ...)で始まる場合、その冠詞/代名詞を先頭の名詞に付ける*。
1002	句が接続詞(and)を含む場合、接続詞を以下のノードの先頭の単語に付ける。
1003	VP が重ねられた VP に続いて不定詞(to)、助動詞、または、法助動詞で構成される場合、動詞群をフラットにし、および不定詞/助動詞/法助動詞を主動詞に付ける。
1004	VP が連辞動詞(is, are, ...)によって率いられ、形容詞、名詞、または、文の述部(ADJP, NP, S, ...)に続く場合、繋合詞を続く述部の先頭の単語に付ける。
1005	NP が重ねられた中心の NP を含む場合、単数レベルに NP をフラットにする。
1006	S が VP によって率いられる場合、VP 層を取り除き、その内容を S に隠す。
1007	PP が NP に続く前置詞を含む場合、2 つの選択肢を作成する。選択肢 1 において、前置詞句を不変のままにする。選択肢 2 において、前置詞を NP の先頭の単語に付ける。
1008	すべての所有格の NP のため、アポストロフィが先行するすべてのノードを含む追加的な NP を挿入する。2 つの選択肢を作成する。選択肢 1 において、さらなる変換を実行しない。選択肢 2 において、アポストロフィを挿入された NP の先頭の単語に付ける。
1009	S の主語が代名詞である場合、2 つの選択肢を作成する。選択肢 1 において、別々の代名詞を保持する。選択肢 2 において、代名詞を先頭の動詞に付ける。
1010	S の目的語が人称代名詞(him, her, ...)である場合、代名詞を先頭の動詞に付ける。

10

20

30

*どこに 2 つの単語が付けられても、それらの POS タグもまた付けられる。

【 0 0 4 1 】

図 5 から 1 4 は、例によって各々の規則の操作を図示する。ツリー・トランスフォーマの全体の操作は、図 4 および 1 5 において図示される。例えば図 4 において示される解析ツリーが提供され、ツリー・トランスフォーマ 1 2 6 は、いくつかの代替的な出力ツリーを生じ、その 1 つが図 1 5 において示される。

【 0 0 4 2 】

図 3 をまた参照すると、役割ラベラー 1 2 8 は、ツリー・トランスフォーマ 1 2 6 から変換されたツリーのリストを受容する。役割ラベラーは、言語学的な役割（例えば SBJ、OBJ）を各々の受容されたツリーの各々の構成要素に割り当てる。プロトタイプ・システムにおいて使用される言語学的な役割の例は、表 4 において提供される。

40

【 0 0 4 3 】

【表 4】

役割レベル	記述
HEAD	句、または、節の意味論的な先頭の単語。
SBJ	文の主語。
OBJ	文の目的語。
SYNHEAD	句、または、節の統語上の先頭。統語上の先頭が意味論的な先頭とは異なる場合にのみ例えば前置詞句で使われる。
PPMOD	前置詞句修飾子。
SMOD	文の修飾子、例えば関係詞節。
PREMOD	上記のラベルの 1 つで適用されない先頭の前に発生するあらゆるノード。
POSTMOD	上記のラベルの 1 つで適用されない先頭の後ろに発生するあらゆるノード。
MIDMOD	第 1 の助動詞および先頭ノード間の文または節において発生するノード、例えば副詞。

10

20

【0044】

図 15 および 27 は、役割ラベラーの操作の例を提供する。入力として図 15 において示される変換されたツリーが提供され、役割ラベラー 128 は、出力として図 27 において示されるツリーを生成する。

【0045】

役割ラベラー 128 の操作は、目的言語パーサ 124 によって決定される（予測される）ように先頭の構成要素に注意することから始める。（コリンズ, 97 のような LPCFG パーサは、先頭中心であり、暗に各々の構成要素の先頭を識別する。ここで、用語「先頭」は、標準言語学的な感覚において使用され、直観的に句または節の最も重要な単語を指す。当業者は、この先頭の情報がそれに対して言及する解析ツリーを増やすことによって、または、それを助動詞データ構造に格納することによって単純に保持されることができるとを考慮して本開示を認識する。）規則の集合はそれから、各々の構成要素の言語学的な役割を決定する役割ラベラー 128 によって適用される。以下の表 5 は、プロトタイプのアラビア語から英語へのシステムにおいて使われる 11 の規則の例示的な集合を記載する。

30

【0046】

【表 5】

規則 #	カテゴリへの適用	規則
2001	All	パーサによって先頭として予測されるすべてのノードに役割 HEAD を割り当てる。
2002	S	先頭に先行する（もしあれば）最も右の非時間的*な NP を見つけ、それを役割 SBJ と分類する。
2003	S	先頭に先行する（もしあれば）最も左の非時制的な NP を見つけ、それを役割 OBJ と分類する。
2004	PP	PP が NP または S に続く前置詞を含む場合、その前置詞を SYNHEAD として、および、NP/S を HEAD として分類する（規則 2001 に優先する）。
2005	SBAR	SBAR が S に続く補文標識を含む場合、その補文標識を SYNHEAD として、および、S を HEAD として分類する（規則 2001 に優先する）。
2006	NP	NP が所有格である場合、'(または り)を HEAD として、および、先行する NP を SYNHEAD として分類する（規則 2001 に優先する）。
2007	All	役割 PPMOD をすべての PP に割り当てる。
2008	All	役割 SMOD をすべての SBAR に割り当てる。
2009	S	最初の助動詞および先頭動詞（未変換の解析の）の間で発生し、上記の規則の 1 つに適用されないあらゆるノードに、役割 MIDMOD を割り当てる。
2010	All	先頭ノードに先行し、上記の規則の 1 つに適用されないあらゆるノードに、役割 PREMOD を割り当てる。
2011	All	先頭ノードに続き、上記の規則の 1 つに適用されないあらゆるノードに、役割 POSTMOD を割り当てる。

*時間的 NP は、時を表す単語（例えば月曜日、朝、昨日、週など）が先頭にあるものである。

【 0 0 4 7 】

図 1 6 から 2 6 は、例によって各々の規則の操作を図示する。

タイプ a の生成物エクストラクタ 1 3 0 は、役割ラベラー 1 2 8 から役割が分類されたツリーのリストを受容する。そのようなリストは、潜在的に大きい。例えば、単一の入力文のために、目的言語パーサ 1 2 4 は、5 0 の信頼できそうな解析ツリーのリストを生じる可能性がある。ツリー・トランスフォーマ 1 2 6 がそれから各々の考えられる解析のために平均 1 0 の変換されたツリーを生じる場合、タイプ a の生成物エクストラクタ 1 3 0 は、5 0 0 の役割が分類されたツリーのリストを受容する。

【 0 0 4 8 】

タイプ a の生成物エクストラクタ 1 3 0 の目的は、語末にない生成物を並列コーパス 1 0 2 の特定の文と関連している各々の役割が分類されたツリーから抽出し、受容されたツリーのこの全体のリストから生成物の結合を取ることによって、生成物の単一の集合を構

10

20

30

40

50

成することである。生成物のこの集合は、受容されたツリーの全体のリストを生成することができる簡潔な文法を定める。(大部分の生成物がn-bestリストの多数のツリーで分配されるので、生成物の数は、ツリーの数よりずっと遅く増加する。)図28は、生成物を表すために使用する例示的なデータ構造を示す。(これらの生成物は、以下に検討される他のタイプの生成物から区別するために『タイプa生成物』と呼ばれる。)

【0049】

これらのタイプa生成物は、まさに前述のLPCFGのように語彙的である。さらに、ユニット生成物、すなわち右側(RHS)が単一の要素だけを含む生成物は、それらの親の生成物に折り畳まれる。特に、各々のRHS要素は、親ノードおよび次の分岐または語末の前のノード間の中間ノードを記載するパスを含む。(パスの一番上のノードの下ノードは、1つの考えられる役割:HEADのみを有する。したがって、パスの要素の役割情報を明白に表す必要性はない。)したがって、上部生成物を除いて、タイプaユニット生成物はない(上部生成物は、TOPのLHSカテゴリおよびツリーの一番上のノードである単一のRHS要素を有する)。図29は、例のツリーおよびそれから抽出されるタイプa生成物の集合を示す。生成物142の要素1のユニット生成パスに留意のこと。タイプa生成物は、標準の文脈自由の文法の生成物があらゆる典型的な解析ツリーから抽出されるのと同じ方法で解析ツリーから抽出されることができると、当業者は認識する。

【0050】

図3に戻ると、言語間の文法推定部132は、タイプaの生成物エクストラクタ130からタイプa生成物の集合を受容し(168、図31)、各々の集合は単一の目的言語の文、すなわち、並列コーパス102から派生する。加えて、言語間の文法推定部132は示されるように、並列コーパス102から各々の目的言語文のソース言語翻訳(170、図31)を受容する。これらの2つの入力ソースから、言語間の文法推定部132は言語間の解析モデル108を生成する。モデル108はしたがって、候補解析をランク付けするのに十分なパラメータを含み、候補解析のパラメータ関連する要素は、ソース言語の単語、目的言語の単語、統語上のラベルおよび役割ラベルを含む。

【0051】

上で言及されるように、言語間の文法推定部132は、言語間の解析モデル108をさらに細かく区別するため、任意に広範な目的言語コーパス116を受容してもよい。推定部132の例は、以下に詳述する。

【0052】

例えば、言語間の文法を生じるための言語間の文法推定部132の操作は、強制的な文法誘導式として処理される。特に、推定部132は、それらの目的言語対照物から引き出される制約の集合を受けてソース言語トレーニング文を生成することができるLPCFG文法(ここでタイプb文法と呼ぶ)を誘導するように構成される。この文法誘導は、ベーカー, 79(Baker)('音声認識のための訓練可能文法'、米国音響学会の春季会議の予稿、547-550ページ)("Trainable grammars for speech recognition," Proceedings of the Spring Conference of the Acoustical Society of America, pages 547 - 550")によって記載される周知のインサイド・アウトサイドアルゴリズムを使用して生成されてもよい。

【0053】

特定の実施態様において、誘導されたタイプb文法が負う特殊な制約は、以下から成る:
(1) 受容されたタイプa生成物がそれぞれ役割 $r_1, r_2 \dots r_n$ で先頭 h に取り付けられる修飾子 $m_1, m_2 \dots m_n$ を含む場合、誘導された文法は、同じ役割を使用している取付け物の同じ集合を許可するが、タイプa生成物に記載されるのと同じ順序で修飾子が取付けられる必要はない。その代わりに、修飾子のあらゆる順序は許可される。
(2) 誘導された文法のあらゆる特定の派生のため、受容されたタイプa生成物のいずれにもおいて発生しない制限された数の取付けは許可される。すなわち、 m が h に取り付けられるタイプa生成物がない場合であっても、修飾子 m が先頭 h に取り付けられてもよい。そのよ

10

20

30

40

50

うな場合、取付け物に割り当てられる役割は、「挿入される」。そのような取付けは、ミスマッチがソースおよび目的言語依存関係の間で発生する場合に文法誘導を容易にする。概算手順を扱いやすく保つため、より多くのものがいくつかの実施態様において提供されてもよいけれども、ツリーにつき1つだけのそのような挿入された取付けが現在許可される。

(3) 受容されたタイプa生成物において発生しているすべての修飾子が、誘導された文法に存在する必要があるわけではない;いくつかは省略されてもよい。例えば、特定のタイプa生成物において、修飾子m1、m2およびm3は、先頭hに取り付けられてもよいが、強制的な文法は、m2およびm3だけが先頭hに取り付けられる構成要素を許可する。

(4) 各々の修飾子は、構成要素内で一度だけ取り付けることが許可される。例えば、役割SBJを有する修飾子(NP、代表的な)は、一度だけ先頭(S、強調される)に取り付けるのが許可される。

10

【0054】

前述の制約を実施するため、図31に示すようにタイプb文法162から切り離されるパーサ160を含む文法推定部132のための構成が使われてもよい。形式的に、特定の实施態様において、タイプ-b文法162は、1)一セットの可変部分、2)一セットの語末、3)一セットの見込みに基づく生成物、および4)可変のユニークなTOPから構成される4つの要素からなる。システムにおいて使用される他のタイプの文法からこれを区別するため、これをタイプb文法と呼び、その構成部分をタイプb可変部分、タイプb語末、およびタイプb生成物と呼ぶ。

20

【0055】

タイプb可変部分は、例えば従来の文法で予想されるように単純な記号ではなく、むしろ、追加的に状態情報を維持する複合的な目的語である。この状態情報は、様々な形の制約を実施するために使用される。例えば、各々のタイプb可変部分は、構成要素にまだ取り付けられない残りの修飾子を追跡する「補足」領域を含み、それによって、各々の修飾子が一度だけ付くことができる制約の施行を許可する。タイプb可変部分において使用されるデータ・フィールドの簡潔な説明は、以下の表6において提供される。

【0056】

【表 6】

領域	記述
カテゴリ	非語末記号（例えば S、VP、NP、など）
先頭の単語	（LPCFG のような）この構成要素の先頭の単語。
補足集合	パーサ状態情報:まだ表されないタイプ a 生成物の残りの要素。
左の端	パーサ状態情報:ここまでのところ構成に取り付けられる最も左の修飾子。形式は、（カテゴリ、役割）である。
右の端	パーサ状態情報:ここまでのところ構成に取り付けられる最も右の修飾子。形式は、（カテゴリ、役割）である。
A 生成物	パーサ状態情報:この構成要素が導き出すタイプ a 生成物。
失った構成要素	パーサ状態情報:ソースおよび目的言語間の依存関係の差異を説明するために使われる。形式は、（カテゴリ、先頭の単語）である。失った構成要素は、タイプ a 生成物予測が取り付けられるべきものであるが、省略されている。
挿入された構成要素	パーサ状態情報:ソースおよび目的言語間の依存関係の差異を説明するために使われる。形式は、（カテゴリ、先頭の単語）である。挿入された構成要素は、タイプ a 生成物によって予測されないけれども、取り付けられたものである。
全体のミスマッチ	パーサ状態情報:解析のどこにでも発生するミスマッチされた取付けの総数。最大は、プロトタイプ・システムにおけるものである。

【 0 0 5 7 】

タイプb語末は、1以上のソース言語の単語の連続した列である（プロトタイプ・システムにおいて、語末は長さが最高3つの単語である）。1以上の単語の列である語末を許可することによって、単一の目的言語の単語は、多数のソース言語の単語に整列されることができる。

【 0 0 5 8 】

図30に示すように、タイプb生成物は3つのタイプで構成される:上部生成物（tp-生成物）、バイナリ分岐生成物（br-生成物）、および葉生成物（lf-生成物）。文法推定（推定部132、図3）の間、パーサ160は、文法162から生成物（164で）を求め、代わりにそのリクエストを満たすタイプb生成物（166で）の（おそらく空の）リストを受容する。文法162の変わった特性は、パーサからのリクエストを満たすために必要とされると、その要素は「高速で」作成されることである。したがって、文法162は、すべての制約は、その応答を返す前に満たされ、派生の現行の状態を反映する可変部分を動的に作成することができると確認することができる。

【 0 0 5 9 】

これらの実施態様において、各々のリクエストのために、パーサ160は文法162にRHS（右側）を通過させ、代わりに特定されたRHSを有する生成物166のリストを受容する。この配置は、インサイド・アウトサイドアルゴリズムと同様に、例えば周知のCKYアルゴリズムのような様々な下からの分析戦略を支持する。3つのタイプのリクエストが支持される。:

・上部生成物のリクエスト（B__Variable v r）- 上部生成物（もし存在すれば）を返す。ここでTOP→v r。

・分岐生成物のリクエスト (`B__Variable vr1` , `B__Variable vr2`) - 分岐生成物のリストを返す (もしあれば)、ここで `vr1` は `vr2` に付く、または `vr2` は `vr1` に付く。

・葉生成物のリクエスト (`B__Terminal tr`) - ソース言語の単語 `tr` を生じる葉生成物のリストを返す。

付録Aは、各々のこれらのリクエストされたタイプの実施態様を記載する図示された擬似コードを提供する。

【 0 0 6 0 】

特定の実施態様において、言語間の文法推定部 1 3 2 は、3つの段階において作動する。段階 1 の間、インサイド・アウトサイドアルゴリズムが、LF_生成のみの概算の確率モデル (すなわち単語の翻訳確率) を推定するために使用される。これらの概算の確率は、次の推定段階の間、より多量の枝刈りを可能にする。必要に応じて、LF_生成のための確率は、ブラウン他, 1 9 9 5 (Brown et al.) によって記載される例えばモデル 1 のような定形状態モデルからの推定値で段階 1 の前に初期化されてもよい。そのような初期設定は、収束の速度を上げるのを助ける傾向がある。段階 2 の間、インサイド・アウトサイドアルゴリズムは、すべての3つのb_生成タイプ: LF_生成と同じく、TP_生成、BR_生成のための確率を決定するために使用される。(実施例によっては、段階 2 はモデル精度の潜在的劣化を犠牲にして、完全に省略されてもよい。) 段階 1 および段階 2 のため、インサイド・アウトサイド推定の繰返し (1 7 2、図 3 1) 典型的に使用される (例えば少なくとも 2 または 3 回)。段階 3 の間、従来の確率CKY解析アルゴリズムが、 - 任意に段階 2 から
のモデル・パラメータで - 各々の文のための単一の最高スコアリング解析を抽出するために使われてもよい。結果は、図 2 において例えば示される言語間の解析ツリーのツリーバンクである。このツリーバンクから、周波数は言語間の解析モデル 1 0 8 の確率を演算するために抽出される。(原則として、それは、周波数を抽出する前に 1 -best 解析のツリーバンクを作成するのに厳密には必要でない; 段階 2 の最終的なインサイド・アウトサイドパスから直接に周波数を抽出することは可能である。しかしながら、1 -best ツリーバンクの作成により、最終的なモデルのサイズが本質的に減少される傾向がある。) これらの3つの段階の特殊な例は、次に記載される。

【 0 0 6 1 】

段階 - 1 確率モデル

代表的な実施態様において、標準設定の分岐確率は 1 . 0 で固定され、したがって、この段階において効果を有しない。ミスマッチされた依存関係の分岐確率は、固定したペナルティ (例えば 0 . 0 5) に設定される。上部確率もまた、1 . 0 で固定される。LF_生成確率のみが、この段階の間更新される。これらの確率のためにユニグラム確率モデルが使用される。

【 0 0 6 2 】

段階 - 2 確率モデル

この段階において、分岐および上部確率もまた、更新される。分岐生成物は、3つの部分に分解される: 1) 先頭予測、2) 修飾子予測、および 3) 修飾子見出し語予測。以下の公式は、可変部分前のカテゴリ (`prev_category`) および前の役割 (`prev_role`) に関連する部分的に構成された構成要素を拡張する生成物のため、前のカテゴリおよび前の役割は、分岐方向次第で、先頭可変部分のどちらかの左端または右端から取られる。新規な構成要素を始める生成物のため、前のカテゴリおよび前の役割は両方ゼロである。

【 0 0 6 3 】

段階 - 3 確率モデル

段階 3 の間、言語間のツリーバンクが作成され、言語間の解析モデル 1 0 8 が生成される。このモデルは、2つの部分を有する。

【 0 0 6 4 】

第 1 の部分は、目的言語の用語およびそのカテゴリ (すなわち統語上のラベル) を提供され、1 以上のソース言語の単語を生成するモデルである。このモデルのための確率は、

上記のLF_生成のものと同一でもよい。

【 0 0 6 5 】

第 2 の部分は、以下のことを除いて、コリンズ, 9 7 (Collins) に記載されるモデル 1 に類似したLPCFGである：

- 統語上のカテゴリおよびラベルと共に、パスおよび役割は同時に予測される。
- 距離特徴は、取り除かれる。
- 一次のマルコフ特性- 前のカテゴリおよび前の役割は、加えられる。

【 0 0 6 6 】

完全性のため、モデルの詳細を以下に記載する。

先頭予測：

10

$P(hc, hp | pc, pt, pw)$

ここで、

- hcは、先頭カテゴリである。
- hpは、先頭パスである。
- pcは、両親カテゴリである。
- ptは、親品詞タグである。
- pwは、親単語である。

修飾子予測：

$P(mc, mt, mw, mp, mr | pc, pt, pw, hc, pm)$

20

ここで、

- mcは、修飾子カテゴリである。
- mtは、修飾子タグである。
- mwは、修飾子の単語である。
- mpは、修飾子パスである。
- mrは、修飾子の役割である。
- pc、pt、pwおよびhcは、上記の先頭の予測場合と同じである。
- pmは、前の修飾子（先頭の次に最も近い）の統語上のカテゴリである。

【 0 0 6 7 】

修飾子予測は、それから 2 つの部分に分けられる：1) カテゴリ、タグ、パスおよび役割の予測、および、2) 見出し語の予測。

30

実際には、このモデルのすべての確率は、容易にされるべきである。容易にする技術をととは、例えば、コリンズ, 9 7 (Collins) の文献において周知である。

【 0 0 6 8 】

（広範な目的言語コーパスから統計によるモデルの任意の改善）

上に言及されるように、広範な目的言語コーパスから導き出される統計は、直接に上記のモデルに加えることができる。修飾子予測が 2 つの部分に分けられ、見出し語を予測部分は前の修飾子に依存していないので、修飾子-見出し語予測の統計は、修飾子順序に反応しない。したがって、修飾子順序が異なるツリーから、見出し語統計を組み込むことの問題はない。特に、目的言語解析ツリーから、修飾子-見出し語統計を組み込むことの障

40

【 0 0 6 9 】

図 3 に示すように、広範な目的言語コーパス 1 1 6 は、提供される場合、目的言語パーサ 1 2 4 によって解析される。特定の実施態様において、1 -best な解析のみが各々の文のために使用される。次に、ツリー・トランスフォーマ 1 2 6 は、前述のようにソース言語により近い形式に各々のツリーを変換する。規則が各々の解析のために多数のツリーを導く場合、すべての生じたツリーが、使われてもよい。役割ラベラー 1 2 8 はまた、以前に記載されるように、役割ラベルを変換されたツリーに加える。最終的に、語彙のカウント抽出器 1 3 6 は、上に検討されたように、言語間の文法推定部 1 3 2 により生成される統計に類似した 2 つの語彙の統計を抽出する。これらの統計はそれから、上記の段階 3 の

50

間、生じられる推定部 1 3 2 によって、言語間のツリーバンクから抽出される（例えば 2 つの語彙の）統計に付加される。結合された統計は、それから言語間の解析モデル 1 0 8 の（2 つの語彙の）確率を計算するために使用される。

【0070】

（実行時翻訳プログラム 1 0 6）

図 3 2 を参照すると、言語間の解析モデル 1 0 8 が推定されるならば、翻訳は LPCFG 以上の分析をほとんど必要としない。示されるように、ソース言語文 1 0 0 は、言語間のパーサ 1 8 0 に到達する。言語間のパーサはまた、前述のように言語間の解析モデル 1 0 8 を受容する。分析が進行する前に、パーサの図は、1 以上のソース言語単語の各々の連続した群の考えられる翻訳が入れられ、ここで、考えられる翻訳およびそれらの確率は、上記した LF_生成が提供される。（最高 3 つの単語が例示的な / プロトタイプ・システムにおいて使われてもよい。）したがって図は、1）目的言語の単語、2）それらの統語上のおよび役割ラベル、および 3）ソース言語単語を生じているそれらの単語 / ラベルの対の確率が入れられる。この点で、生じているツリーが各々のノードの役割情報を含むことを除いて、単一言語の場合に関するよう分析は進行する。

【0071】

ツリー再順序付け部 1 8 2 は、目的言語の自然の単語順序にツリーを再配置するためにその役割情報を使用する。しかしながら、ツリー再順序付け部 1 8 2 は最初に、ツリー・トランスフォーマ 1 2 6 によって作成される形態的な変化を元に戻すために逆規則を適用する（例えば、先頭名詞からの限定詞を分ける、動詞群を個別的な単語に切るなど）。形態的な変化が元に戻るならば、ツリー再配置は進行する。目的言語が英語の場合の、プロトタイプ・システムにおいて、再配置規則は次の通りである：

- SBJ ノードは、直接に先頭動詞に先行する。
- OBJ ノードは、直接に先頭動詞に続く。
- POSTMOD ノードは、もしあれば、OBJ の後の先頭ノードに続く。
- PPMOD ノードは、もしあれば、POSTMOD の後の先頭ノードに続く。
- SMOD ノードは、もしあれば、PPMOD の後の先頭ノードに続く。
- MIDMOD ノードは、動詞群の最初の動詞の後に配置される。
- PREMOD ノードは、もしあれば、SBJ に先行する先頭ノードの前に配置される。言語間の解析ツリーの先頭の前に発生するあらゆる PREMOD ノードの順序は、それらを先頭の前の位置へ移動する前に逆転される。
- SYNHEAD ノードは、もしあれば、PREMOD の前に先頭ノードに先行する。
- 限定詞および所有代名詞は、名詞句の先頭に配置される。
- 前置詞は、前置詞句の前部に移動される。

【0072】

ソース言語文が解析され、再順序付けされるならば、翻訳工程は、本質的に完結される。従来の翻訳が必要とされる場合、語末の前の抜粋工程 1 8 4 は語末の前のノードを抽出し、訳文を生ずる。代替的に、目的言語の解析ツリーが必要とされる場合、葉の除去工程 1 8 6 は、ソース言語語末を取り除き、目的言語解析を残す。

【0073】

ここで記載される実施態様の 1 つに関して記載されるあらゆる特徴は、本発明の趣旨および範囲から逸脱することなく、ここで記載される他の実施態様に使用されると理解される。

【0074】

前述の実施態様がコンピュータまたは他の高性能装置で動作するソフトウェア要素として記載されてきたけれども、名目上、あらゆるここで記載される様々なモジュールまたは工程は、本発明の趣旨および範囲から逸脱することなくソフトウェア、ハードウェア、または、そのあらゆる組合せで実行されてもよいとさらに理解される。

【0075】

先行する明細書において、本発明は、その特殊な例示的な実施態様に関して記載されて

10

20

30

40

50

きた。以下に続く請求項に記載のより広い本発明の精神と範囲から逸脱することなく、様々な変更態様と改変がそれに対して作成されてもよいことは明らかである。明細書および図面はしたがって、これに限定されるということではなく例示の意味であるとみなされることになる。

【図面の簡単な説明】

【0076】

【図1】本発明の実施態様の高水準の機能的なブロック線図である。

【図2】図1の実施態様によって使用される例示的な言語間の解析ツリーである。

【図3】図1の一部の実施態様の詳細な機能的なブロック線図である。

【図4】図1の実施態様によって使用される例示的な目的言語解析ツリーである。

10

【図5】本発明のツリー-トランスフォーマ・モジュールの実施態様によって使用される様々な規則の例である。

【図6】本発明のツリー-トランスフォーマ・モジュールの実施態様によって使用される様々な規則の例である。

【図7】本発明のツリー-トランスフォーマ・モジュールの実施態様によって使用される様々な規則の例である。

【図8】本発明のツリー-トランスフォーマ・モジュールの実施態様によって使用される様々な規則の例である。

【図9】本発明のツリー-トランスフォーマ・モジュールの実施態様によって使用される様々な規則の例である。

20

【図10】本発明のツリー-トランスフォーマ・モジュールの実施態様によって使用される様々な規則の例である。

【図11】本発明のツリー-トランスフォーマ・モジュールの実施態様によって使用される様々な規則の例である。

【図12】本発明のツリー-トランスフォーマ・モジュールの実施態様によって使用される様々な規則の例である。

【図13】本発明のツリー-トランスフォーマ・モジュールの実施態様によって使用される様々な規則の例である。

【図14】本発明のツリー-トランスフォーマ・モジュールの実施態様によって使用される様々な規則の例である。

30

【図15】図5から14までのツリー-トランスフォーマによって生成される変換されたツリーの例である。

【図16】本発明の役割ラベラーの実施態様によって使用される様々な規則の例である。

【図17】本発明の役割ラベラーの実施態様によって使用される様々な規則の例である。

【図18】本発明の役割ラベラーの実施態様によって使用される様々な規則の例である。

【図19】本発明の役割ラベラーの実施態様によって使用される様々な規則の例である。

【図20】本発明の役割ラベラーの実施態様によって使用される様々な規則の例である。

【図21】本発明の役割ラベラーの実施態様によって使用される様々な規則の例である。

【図22】本発明の役割ラベラーの実施態様によって使用される様々な規則の例である。

【図23】本発明の役割ラベラーの実施態様によって使用される様々な規則の例である。

40

【図24】本発明の役割ラベラーの実施態様によって使用される様々な規則の例である。

【図25】本発明の役割ラベラーの実施態様によって使用される様々な規則の例である。

【図26】本発明の役割ラベラーの実施態様によって使用される様々な規則の例である。

【図27】図16から26までの役割ラベラーによって生成される役割ラベル化されたツリーの例である。

【図28】本発明の実施態様によって使用されるタイプaの製造データ・フィールドの例である。

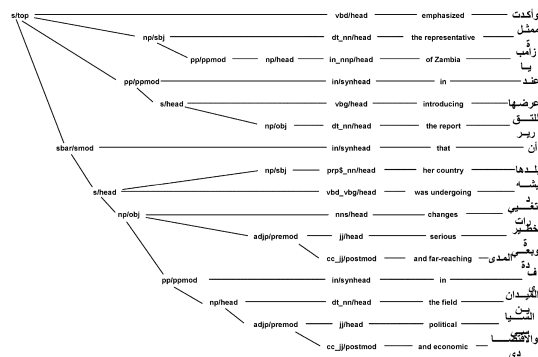
【図29】本発明のタイプaの生成物エクストラクタの例である。

【図30】本発明の実施態様によって使用されるタイプbの文法要素の例およびデータ・フィールドである。

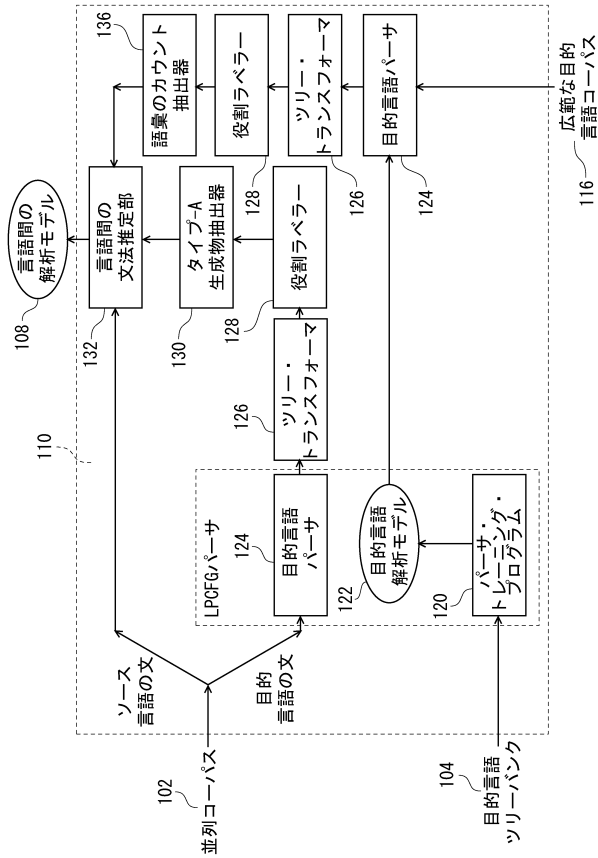
50

【図 3 2】図 3 2 は、図 1 の実施態様の実行時翻訳プログラム部分の機能的なブロック線図である。

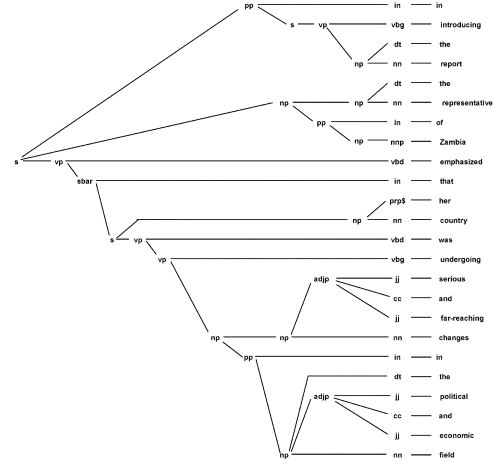
【 図 2 】



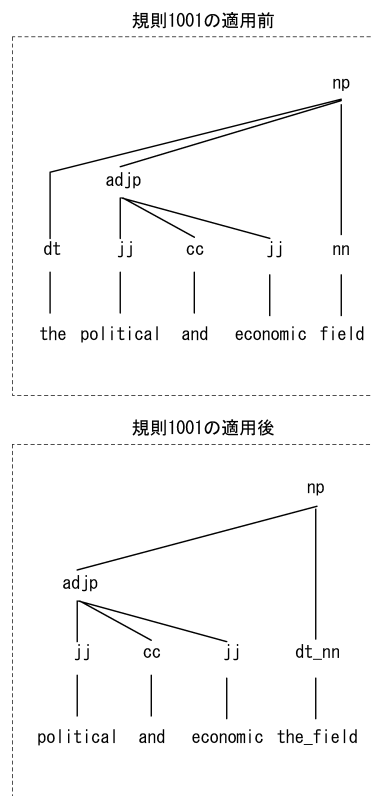
【図3】



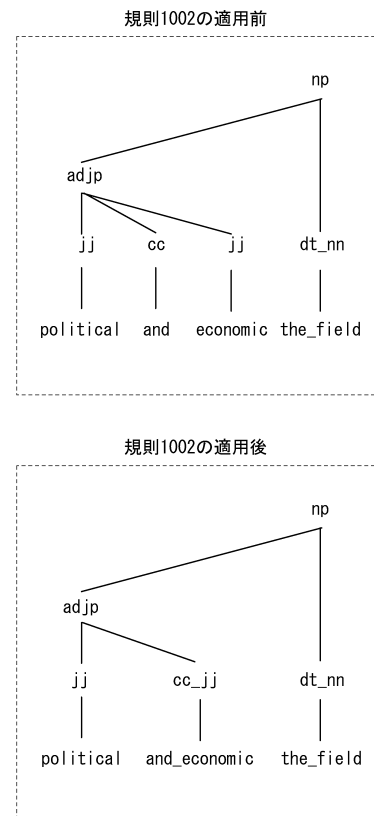
【図4】



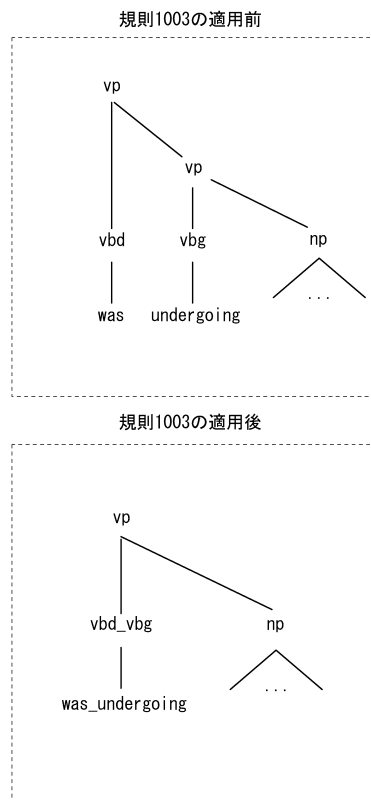
【図5】



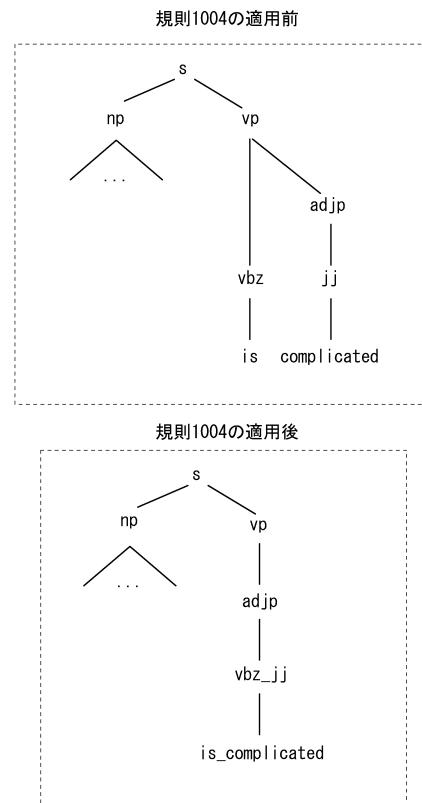
【図6】



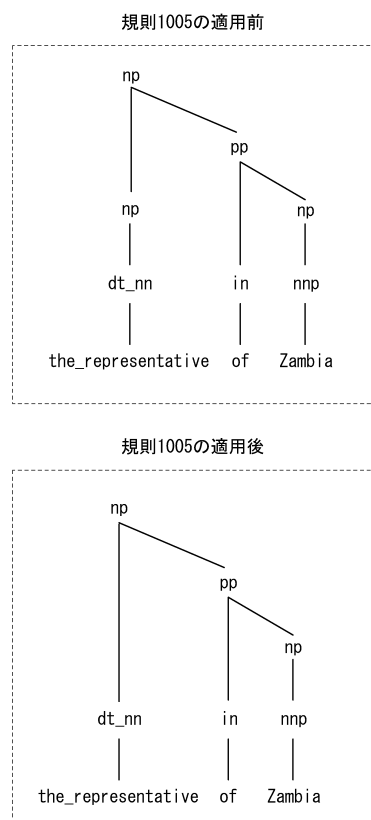
【図 7】



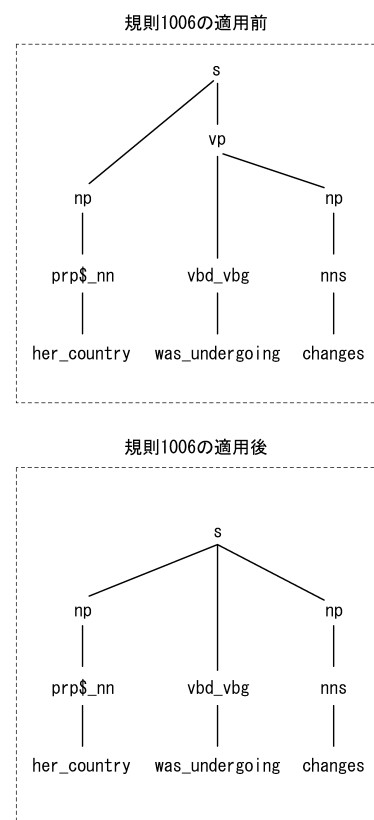
【図 8】



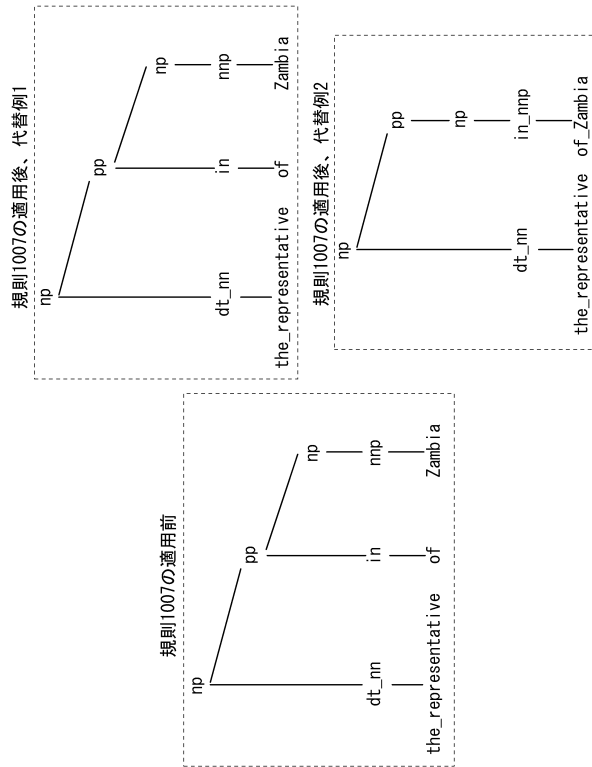
【図 9】



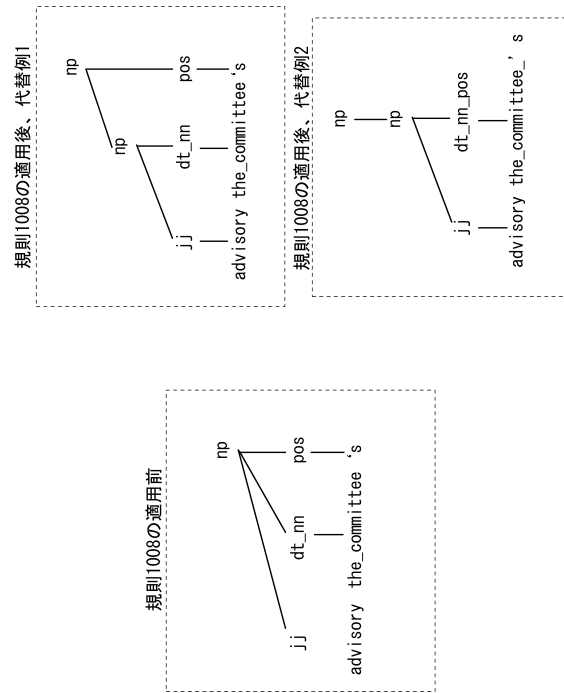
【図 10】



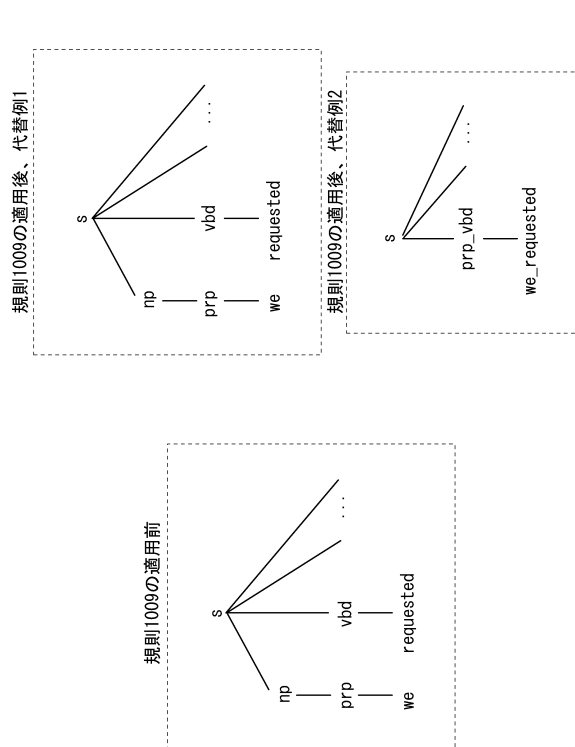
【図 1 1】



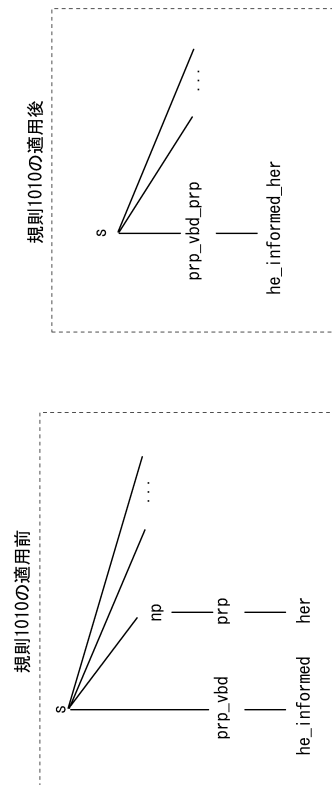
【図 1 2】



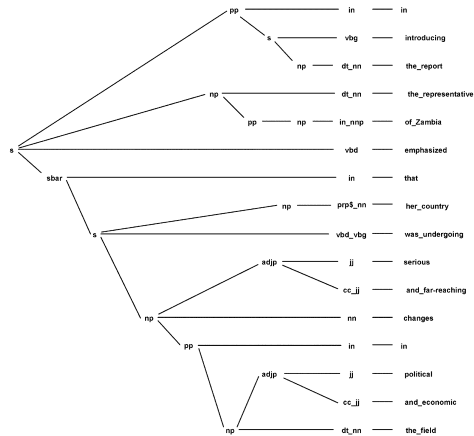
【図 1 3】



【図 1 4】

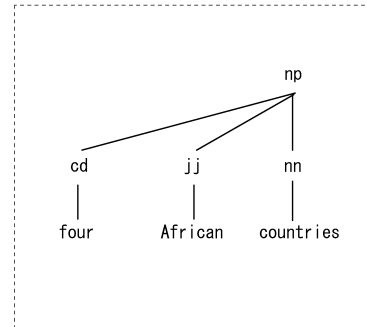


【図 15】

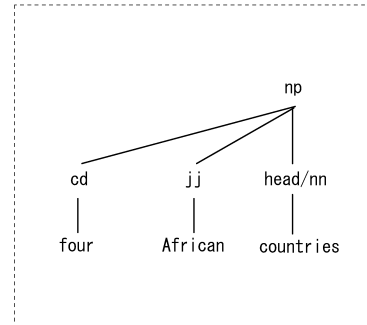


【図 16】

規則2001の適用前

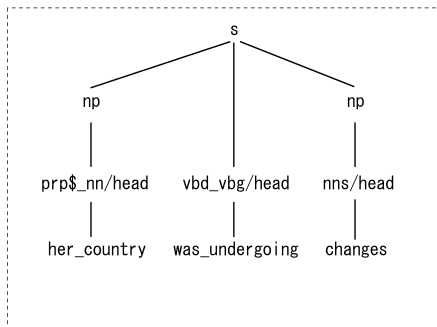


規則2001の適用後

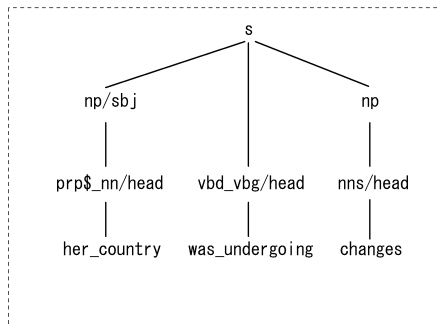


【図 17】

規則2002の適用前

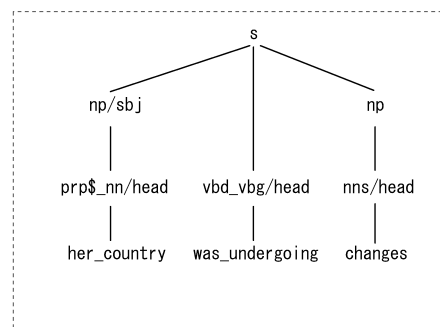


規則2002の適用後

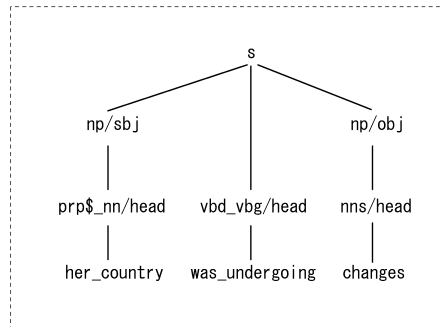


【図 18】

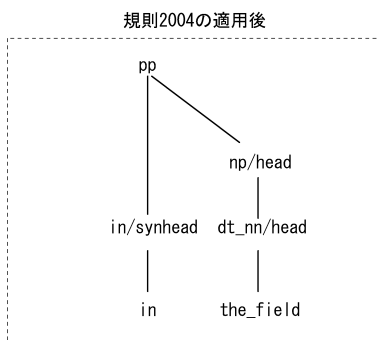
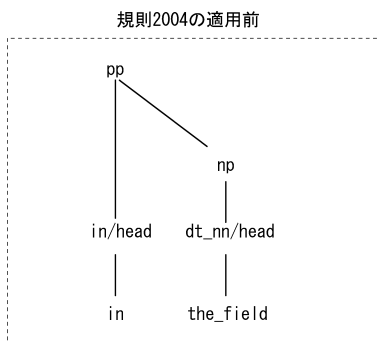
規則2003の適用前



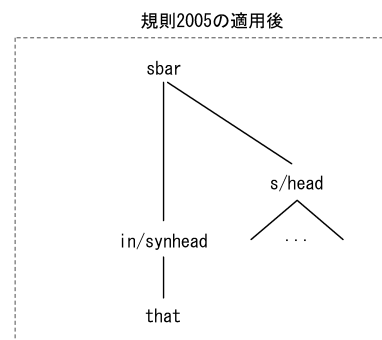
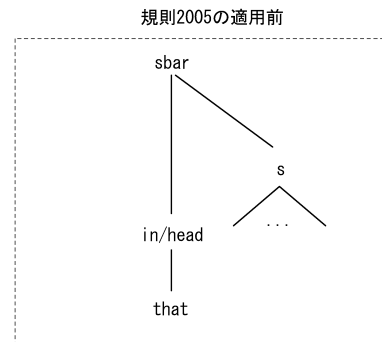
規則2003の適用後



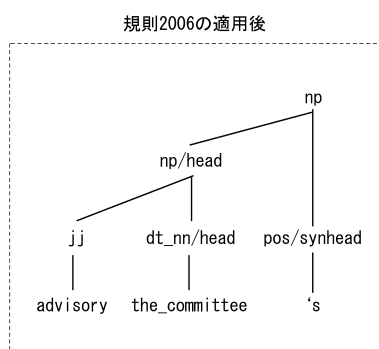
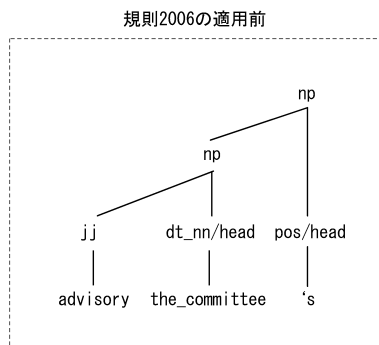
【図 19】



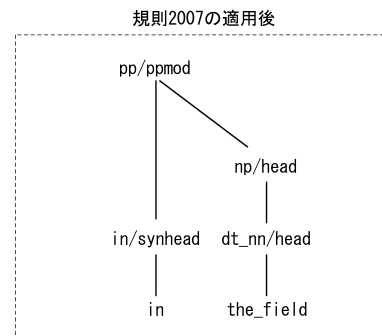
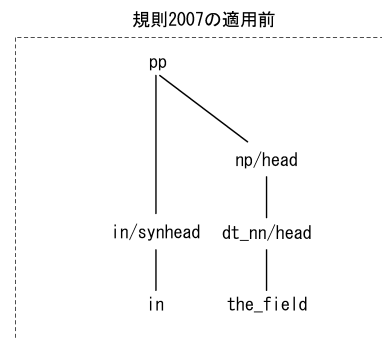
【図 20】



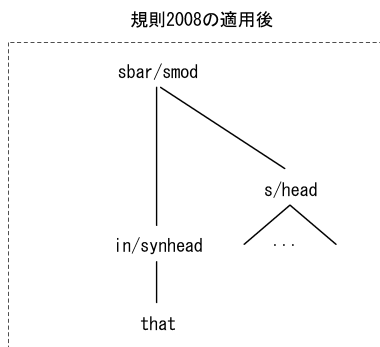
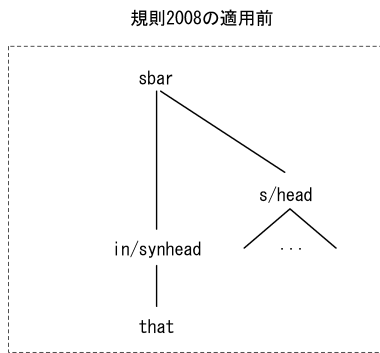
【図 21】



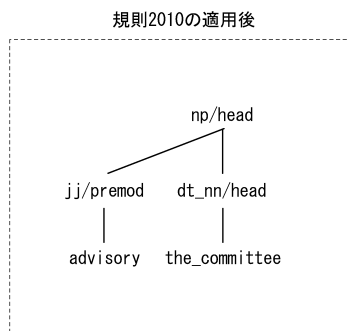
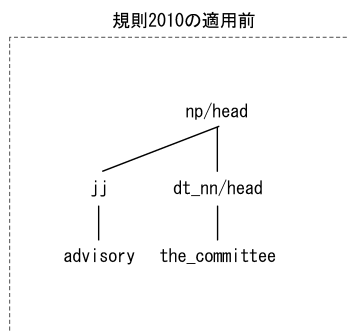
【図 22】



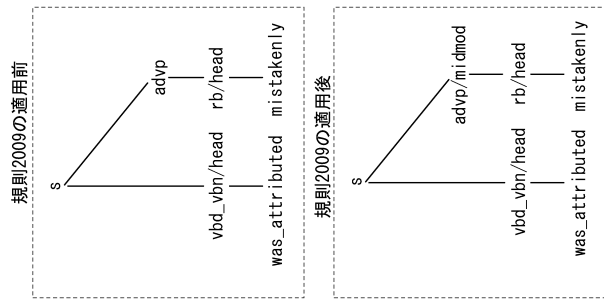
【図 23】



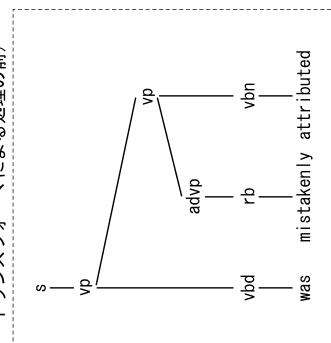
【図 25】



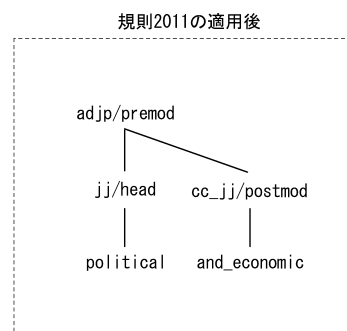
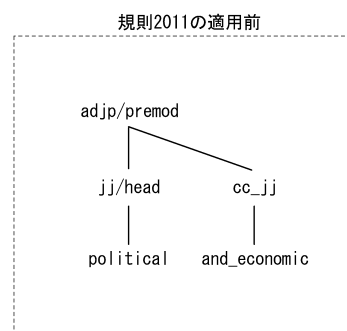
【図 24】



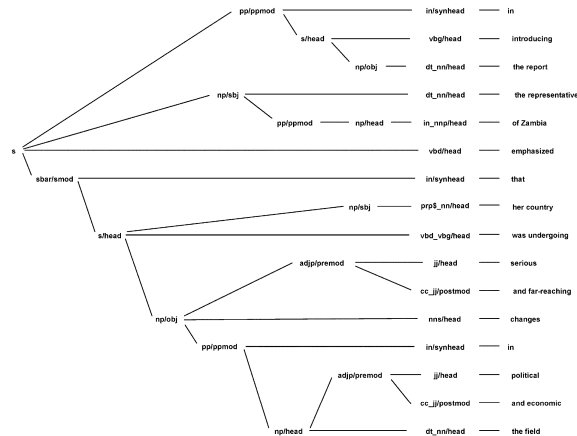
未変換の解析（ツリー・
トランスフォーマーによる処理の前）



【図 26】



【図 27】

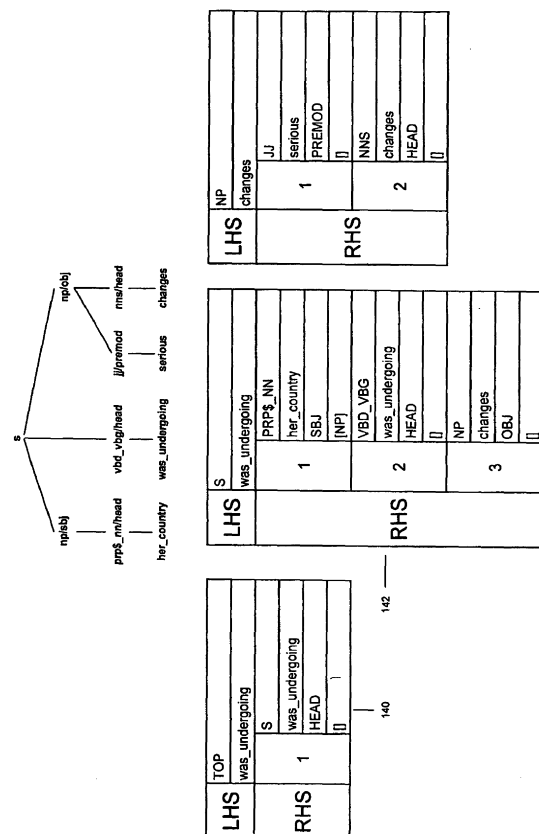


【図 28】

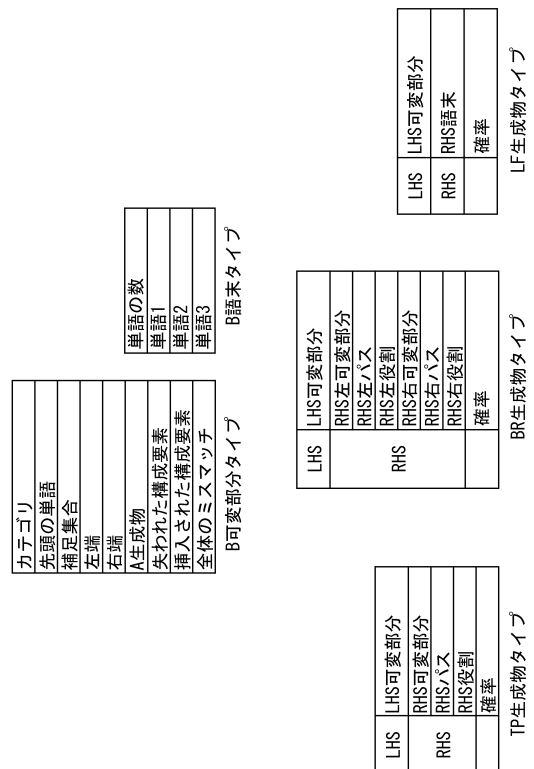
A生成物タイプ

LHS	LHSカテゴリ
	LHS見出し語
RHS	1
	RHS-1カテゴリ
	RHS-1見出し語
	RHS-1役割
	RHS-1パス
	2
	RHS-2カテゴリ
	RHS-2見出し語
	RHS-2役割
	RHS-2パス
	...
	n
	RHS-nカテゴリ
	RHS-n見出し語
	RHS-n役割
	RHS-nパス

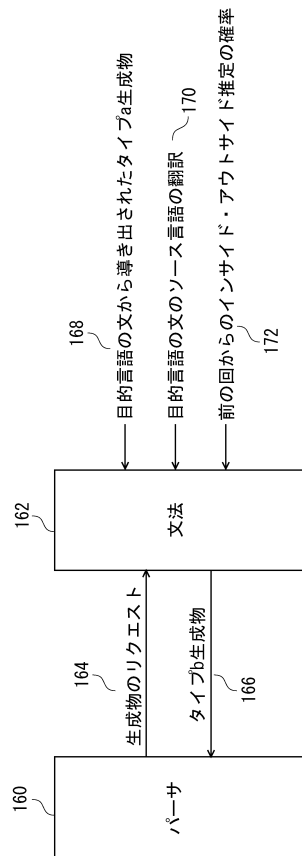
【図 29】



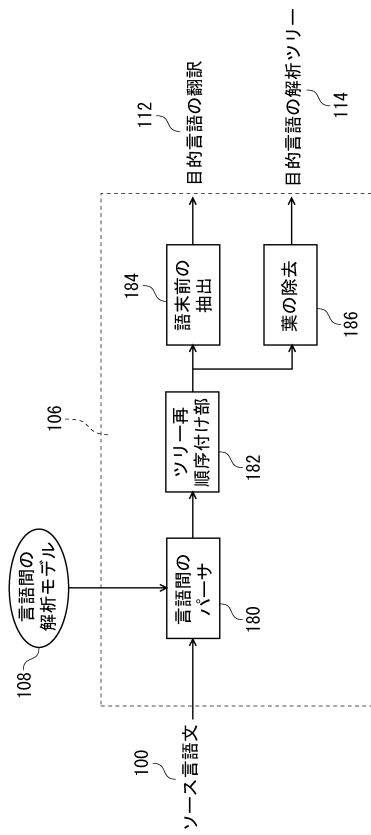
【図 30】



【図 3 1】



【図 3 2】



フロントページの続き

(72)発明者 ミラー スコット

アメリカ合衆国 マサチューセッツ州 02468 ワーバン ウッドワード ストリート 322

審査官 長 由紀子

(56)参考文献 特開2005-285129(JP,A)

特開平05-189481(JP,A)

特表2004-501429(JP,A)

FEI XIA and Michael Mccord, Improving a Statistical MT System with Automatically Learned Rewrite Patterns, IN PROCEEDINGS OF COLING 2004, 2004年 8月27日, URL, <http://www.mt-archive.info/Coling-2004-Xia.pdf>

MICHAEL COLLINS, et.al., Clause Restructuring for Statistical Machine Translation, PROCEEDINGS OF THE 43RD ANNUAL MEETING ON ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 2005年 6月30日, URL, <http://acl.lidc.upenn.edu/P/P05/P05-1066.pdf>

(58)調査した分野(Int.Cl., DB名)

G06F 17/20-28