

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第3924260号
(P3924260)

(45) 発行日 平成19年6月6日(2007.6.6)

(24) 登録日 平成19年3月2日(2007.3.2)

(51) Int. Cl. F I
G O 6 F 17/27 (2006.01) G O 6 F 17/27 E

請求項の数 6 (全 44 頁)

<p>(21) 出願番号 特願2003-108676 (P2003-108676)</p> <p>(22) 出願日 平成15年4月14日(2003.4.14)</p> <p>(65) 公開番号 特開2004-318300 (P2004-318300A)</p> <p>(43) 公開日 平成16年11月11日(2004.11.11)</p> <p>審査請求日 平成16年6月10日(2004.6.10)</p> <p>特許法第30条第1項適用 2003年3月18日 言語処理学会発行の「言語処理学会第9回年次大会発表論文集」に発表</p> <p>(出願人による申告) 国等の委託研究の成果に係る特許出願(平成15年度通信・放送機構、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条の適用を受けるもの)</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</p> <p>(74) 代理人 100099933 弁理士 清水 敏</p> <p>(72) 発明者 丸山 岳彦 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>(72) 発明者 柏岡 秀紀 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>(72) 発明者 熊野 正 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】 節境界検出装置、機械翻訳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】

文章に対して形態素解析を行なう事により得られた形態素列から、元の文章の節境界を検出するための節境界検出装置であって、

前記形態素列の中において、所定の形態素の並びのパタンを検出するための検出手段と、

前記パタンが検出された事に応答して、前記形態素列の中で、検出された前記パタン中の形態素の並びと所定の関係にある位置を節の境界に指定する予め定める処理を行なって前記形態素列を出力するための境界指定手段とを含み、

前記検出手段は、

前記形態素列を順次読込んでF I F O方式で記憶して出力するための一時記憶手段と、

前記一時記憶手段に記憶された形態素の配列の中に、前記所定の形態素の並びのパタンがある事を検出するための手段とを含み、

前記境界指定手段は、前記所定の形態素の並びのパタンがある事が検出された事に応答して、前記一時記憶手段の前記所定の形態素の並びのパタンまでを出力する様に前記一時記憶手段を制御するための手段と、

前記一時記憶手段から出力される前記所定の形態素の並びのパタンの末尾の形態素の直後に、節境界を示すマーカを挿入するための手段とを含む、節境界検出装置。

【請求項2】

文章に対して形態素解析を行なう事により得られた形態素列から、元の文章の節境界を

検出するための節境界検出装置であって、

前記形態素列の中において、所定の形態素の並びのパタンを検出するための検出手段と

前記パタンが検出された事に応答して、前記形態素列の中で、検出された前記パタン中の形態素の並びと所定の関係にある位置を節の境界に指定する予め定める処理を行なって前記形態素列を出力するための境界指定手段とを含み、

前記検出手段は、

前記形態素列を順次読込んでF I F O方式で記憶して出力するための一時記憶手段と、

前記一時記憶手段に記憶された形態素の配列の中に、複数個の前記所定の形態素の並びのパタンのうちの任意の一つがある事を検出するための手段とを含み、

10

前記境界指定手段は、前記任意の一つのパタンが検出された事に応答して、前記一時記憶手段中の、前記検出されたパタンまでをF I F O方式で出力する様に前記一時記憶手段を制御するための手段と、

前記任意の一つのパタンが検出された事に応答して、前記一時記憶手段から出力される前記パタンの末尾の形態素の直後に、検出されたパタンに対応した節境界ラベルを挿入するための手段とを含む、節境界検出装置。

【請求項 3】

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 又は請求項 2 に記載の節境界検出装置として動作させる、コンピュータプログラム。

【請求項 4】

20

入力される日本語の文章に対して形態素解析処理を行ない、得られる形態素列を出力するための形態素解析手段と、

請求項 1 に記載の節境界検出装置であって、前記形態素解析手段の出力を入力として受ける様に接続された節境界検出装置と、

前記節境界検出装置から出力される形態素列を、当該形態素列中の節境界によって節に分離するための節分離手段と、

前記節分離手段により分離された形態素列を入力とし、前記節分離手段から節を受けとった事に応答して、受けた節を翻訳するための機械翻訳手段とを含む、機械翻訳装置。

【請求項 5】

前記節分離手段は、

30

前記節境界検出装置からの出力をF I F O方式で一時記憶するための記憶手段と、

前記節境界検出装置から境界マークが出力された事に応答して、前記記憶手段に記憶された形態素列を前記機械翻訳手段に与え、機械翻訳を開始させるための手段とを含む、請求項 4 に記載の機械翻訳装置。

【請求項 6】

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 4 または請求項 5 に記載の機械翻訳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】

【0001】

【発明の属する技術分野】

40

この発明は、自然言語処理を適切に行なえる様にするための前処理を行なう装置に関し、特に、翻訳等の処理が適切に行なえる様に節単位に入力テキストを分離する前処理を行なうための節境界検出装置、そうした節境界検出装置を採用した機械翻訳装置、およびそれらのためのコンピュータプログラムに関する。

【0002】

【従来の技術】

近年、独話（複数の発話のまとめりであって、講演、ニュース等、発話者が1人のもの）を対象とした自然音声コーパスの構築が進んでいる。講演、ニュースまたは学会発表等、1人の話者が話しつつける独話は、対話（複数の発話のまとめりであって、2人の発話者が発話を交換するもの）よりも1文の長さが長くなったり、文の構造が複雑化したりする

50

という特徴を持つ事が知られている。

【0003】

図1に、典型的な独話の例であるテレビジョンのニュース(日本語)、及び典型的な対話の例である旅行会話(対訳形式の日本語部分)での1文あたりの形態素数及び文節数を示す。図1から分かる様に、1文あたりの形態素数及び文節数のいずれにおいても、独話の方が対話よりもはるかに多い。

【0004】

さらに、自発的な発話になるほど、明示的な文末表現の現れにくくなる傾向があり、文の境界を認識する事が困難になる。

【0005】

独話又は対話等の音声認識を行なったり、翻訳を行なったりする自然言語処理技術では、従来、「文」を基本的な処理単位とする場合が大半である。

【0006】

しかし、1文が長く、文末が確定しにくいという性格を持つ独話を自然言語処理の対象とする場合、文が長くなる事によって構文解析のあいまい性が爆発するという問題がある。また、文末がはっきりしないために、自然言語処理の対象となるものがはっきりせず、どこまで入力を待てば処理を始めることができるのかが分からないという問題が発生する。

【0007】

こうした問題は、独話を機械翻訳する場合にも現れる。独話を機械翻訳する場合、発話に追従して翻訳を出力する同時通訳としての運用が望ましい。しかし上記した様に独話では1文が長くなるため、解析が失敗したり、その結果として翻訳が失敗したりするという問題がある。仮に翻訳が成功したとしても、同時通訳としての追従性に欠けるという問題がある。また、文末が確定しにくい場合、どの時点でどの部分を対象に翻訳を開始するかを定めることは難しい。

【0008】

【非特許文献1】

益岡隆志・田窪行則著、「基礎日本語文法 改訂版」、くろしお出版、1992

【発明が解決しようとする課題】

従って、特に独話について、発話の中で各種の処理を漸進的に進めておくために、文とは異なる単位を発話の中で随時検出できる様にすることが望ましい。可能であれば、その処理単位は文よりも短いほうが望ましい。

【0009】

文よりも短い処理単位として、述語を中心としたまとまりである「節」を用いる事が望ましいと考えられる。節は、統語的・意味的にまとまった単位であり、翻訳または文の要約等の処理を節単位で行なうと有効であると考えられる。そこで、節境界を自動的に検出する手法が必要となる。

【0010】

節境界を検出する手法としてまず考えられるのは、構文解析器を用いて文を解析した結果から、節境界に相当する位置を特定する方法である。しかし、構文解析器は一般に入力として「文」を要求するものである。そのため、文末が入力されて構文解析が済むまでは、節境界の検出を始める事は難しい。この制約は、同時通訳の様に入力を漸進的に処理していく必要がある場合、望ましくない。漸進的な処理を行なうためには、発話の入力中であっても、局所的な情報のみから節境界の位置を検出できる事が望ましい。また、節境界により分離される節がどのようなものであるかを知る事ができれば、自然言語処理技術にとって有用なだけでなく、言語学的な分析にも応用できるため、より好ましい。

【0011】

従って、本発明の目的は、日本語の発話の中から、局所的な情報のみから節境界を随時検出する事ができる節境界検出装置を提供する事である。

【0013】

この発明のさらに他の目的は、日本語の発話の局所的な情報のみから節境界を検出し、当

10

20

30

40

50

該節境界により分離される節がどのような種類の節かを判定する事ができる節境界検出装置を提供する事である。

【0014】

この発明のさらに他の目的は、日本語の発話の中から節を随時検出し、節ごとに自動的に翻訳を行なう事ができる機械翻訳装置を提供する事である。

【0015】

【課題を解決するための手段】

本発明の第1の局面に係る節境界検出装置は、文章に対して形態素解析を行なう事により得られた形態素列から、元の文章の節境界を検出するための節境界検出装置であって、形態素列の中において、所定の形態素の並びのパターンを検出するための検出手段と、パターンが検出された事に応答して、形態素列の中で、検出されたパターン中の形態素の並びと所定の関係にある位置を節の境界に指定する予め定める処理を行なって形態素列を出力するための境界指定手段とを含む。

10

【0016】

好ましくは、境界指定手段は、パターンが検出された事に応答して、ある位置に節の境界を示す境界マーカを挿入して形態素列を出力するための手段を含む。

【0017】

さらに好ましくは、検出手段は、形態素列の中において、複数個のパターンのうちの任意の一つを検出するための手段を含む。

【0018】

境界指定手段は、任意の一つを検出するための手段により上記パターンのうちの任意の一つが検出された事に応答して、検出されたパターン中の形態素の並びと所定の関係にある位置に、検出されたパターンに対応して予め定められた節境界ラベルを挿入するためのラベル挿入手段を含んでもよい。

20

【0019】

節境界ラベル又は節境界マーカが挿入される位置は、検出されたパターン中の末尾の形態素の直後でもよい。

【0020】

好ましくは、検出手段は、形態素列を順次読込んでFIFO(First-In First-Out)方式で記憶して出力するための一時記憶手段と、一時記憶手段に記憶された形態素の配列の中に、所定の形態素の並びのパターンがある事を検出するための手段とを含み、境界指定手段は、所定の形態素の並びのパターンがある事が検出された事に応答して、一時記憶手段の所定の形態素の並びのパターンまでを出力する様に一時記憶手段を制御するための手段と、一時記憶手段から出力される所定の形態素の並びのパターンの末尾に、節境界を示すマーカを挿入するための手段とを含んでもよい。

30

【0021】

さらに好ましくは、検出手段は、形態素列を順次読込んでFIFO方式で記憶して出力するための一時記憶手段と、一時記憶手段に記憶された形態素の配列の中に、複数個の所定の形態素の並びのパターンのうちの任意の一つがある事を検出するための手段とを含み、境界指定手段は、任意の一つのパターンが検出された事に応答して、一時記憶手段中の、検出されたパターンまでをFIFO方式で出力する様に一時記憶手段を制御するための手段と、任意の一つのパターンが検出された事に応答して、一時記憶手段から出力されるパターンの末尾に、検出されたパターンに対応した節境界ラベルを挿入するための手段とを含む。

40

【0022】

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節境界検出装置として動作させるものである。

【0023】

本発明の第3の局面に係る機械翻訳装置は、入力される日本語の文章に対して形態素解析処理を行ない、得られる形態素列を出力するための形態素解析手段と、上記したいずれか

50

の節境界検出装置であって、形態素解析手段の出力を入力として受ける様に接続された節境界検出装置と、節境界検出装置から出力される形態素列を、当該形態素列中の節境界によって節に分離するための節分離手段と、節分離手段により分離された形態素列を入力とし、節分離手段から節を受けとった事に応答して、受けた節を翻訳するための機械翻訳手段とを含む。

【0024】

好ましくは、節境界検出装置は出力する形態素列の節境界に節境界マーカを挿入する機能を持ち、節分離手段は、節境界検出装置からの出力をFIFO方式で一時記憶するための記憶手段と、節境界検出装置から節境界マーカが出力された事に応答して、記憶手段に記憶された形態素列を機械翻訳手段に与え、機械翻訳を開始させるための手段とを含む。

10

【0025】

本発明の第4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機械翻訳装置として動作させるものである。

【0026】

【発明の実施の形態】

[第1の実施の形態]

節境界検出の原理

構文解析を行わずに節境界を検出するために、本実施の形態では、入力テキストを形態素解析し、形態素の局所的な接続関係のみを手掛かりとして節境界を検出する。そのために、形態素の局所的な接続関係をパターンに分類し、特定のパターンが検出された場合に節境界を特定するルールを作成し、このルールに従って自動的に節境界を特定する。この節境界検出ルールは、節境界の位置を発見するための形態素列パターンと、節境界の種類を問わず節境界ラベルの組とからなっている。構文解析は必要としない。

20

【0027】

構成

図2に、本実施の形態の節境界検出装置を採用した翻訳装置の機能ブロック図を示す。この実施の形態のシステムは、テキスト処理が可能な既存の言語処理系（具体的にはPerl処理系）を用い、節境界検出ルールをPerlの正規表現置換を用いたスクリプトの形式で実装している。

【0028】

図2を参照して、この翻訳装置30は、ユーザからの起動コマンド32に応答して、日本語の入力テキスト34を英語に機械翻訳し、結果を翻訳出力36として出力するためのものである。翻訳装置30は、入力テキスト34を形態素解析し、形態素列を出力するための形態素解析部54を含む。形態素解析部54としては、既存の形態素解析用プログラムを用いる事ができる。図3に、形態素解析用プログラムとしてよく知られているものの出力形式と出力例とを示す。図3の詳細については後述する。

30

【0029】

再び図2を参照して、翻訳装置30はさらに、上記した節境界検出ルールをPerlの正規表現命令列からなるスクリプトの形で実装したプログラム52と、形態素解析部54の出力する形態素列に対してプログラム52を適用する事により、節境界ごとに節ラベルが挿入された処理後のテキストを出力するための言語処理系56と、言語処理系56の出力をFIFO方式で一時的に蓄積するバッファ58と、言語処理系56から節ラベルが出力されるごとに、バッファ58に蓄積されたテキストを読み出して出力する事により、テキストを節に分離するためのテキスト分離部60と、テキスト分離部60から与えられるテキストを英語に翻訳して翻訳出力36として出力するための機械翻訳部62と、ユーザからの起動コマンド32に応答して入力テキスト34およびプログラム52を読み込み、形態素解析部54および言語処理系56等を起動するためのオペレーティングシステム(OS)50とを含む。ここで節の境界を示す情報を「節ラベル」と呼んでいるのは、節の種類を示す情報を含んでいるためである。この節ラベルは、そこに節境界が存在している事を表すものでもあり、節境界を示すマーカとしての役割も果たしている。

40

50

【0030】

OS50、言語処理系56を備え、プログラム52を実行する事からも分かる様に、翻訳装置30は実質的にはコンピュータから構成されている。入力テキスト34および翻訳出力36はそれぞれ、標準入力および標準出力を示すが、本実施の形態では入力テキスト34は所定のファイルから与えられ、翻訳出力36も所定のファイルとして出力されるものとする。

【0031】

図3を参照して、形態素解析部54の出力形式80について説明する。形態素解析部54が出力する形態素は、出力形式80に示される様に、形態素の出現形と、その品詞と、その活用形と、出現したときの活用型とからなる。「活用形」とは、動詞、助動詞、形容詞等の活用のしかたの分類を示すものである。例は「五段活用」「下二段活用」の如きものである。活用型とは、出現した形態素が、各活用形の中でどの様な活用をされているかを示すものである。例は「未然形」「連用形」「連体形」等の如きものである。

10

【0032】

図3に、入力例82として「私は学校へ行きました」という一文を示す。これを形態素解析部54により形態素解析した結果を解析結果84に示す。解析結果84から明らかな様に、形態素解析部54は入力テキスト34を形態素解析し、出力形式80に従った形式の形態素列を出力する機能を持つ。

【0033】

図4に、節境界検出ルールの実装したPerlによるコマンドの例を示す。図4を参照して、Perlによる置換コマンドの一般形100は、置換を示すコマンド「s」と、置換の際に検索すべき検索文字列と、検索された文字列を置換すべき文字列と、置換の際のオプション機能を指定するオプション文字列とを「/(スラッシュ)」により区切った形式となっている。検索文字列及び置換文字列には、それぞれ「正規形」と呼ばれる形式を使用することができる。Perlに限らず、このような正規形が使用できる言語処理系は数多く存在する。正規形について必要な場合には以下で説明を加えてあるが、一般的な説明については各言語処理系の解説書等を参照されたい。

20

【0034】

図4の基本形102により、本実施の形態でPerlにより実装された節境界検出ルールの一般形を示す。基本形102は、一般形100において、検索文字列を形態素列パターン110とし、置換文字列を「\$1 ¥/節ラベル¥/」という置換文字列表現112とし、オプションとして「g」を指定したものである。

30

【0035】

形態素列パターン110は丸カッコで囲まれている。これは置換文字列表現112の中の「\$1」に対応する。置換文字列表現112中の「\$1」は、この部分を検索文字列中で丸カッコで囲まれた文字列の中で1番目のものと置換する事を示す。検索文字列の中で丸カッコで囲まれた文字列は形態素列パターン110しかないので、\$1は形態素列パターン110により置換される。

【0036】

置換文字列表現112の中の「¥」はエスケープ文字であり、この直後の一文字をコマンドの一部ではなく単なる文字として扱う事を示す。この例では、置換文字列が「/」を含んでいるのに対し、この「/」がコマンド中でも使用されているので、置換文字列中のスラッシュを単なる文字列として扱うために「¥」を使用している。節ラベルについては後述する。

40

【0037】

オプション「g」は、グローバルサーチを示す。すなわち、検索文字列で入力文字列を検索した結果、最初のマッチが見出されたところで検索を終了するのではなく、マッチがいくつ見出されても入力文字列の全体について検索および置換を行なう事を示す。

【0038】

すなわち、基本形102によれば、形態素列パターン110にマッチする形態素列があれば

50

、その形態素列の最後に「 / 節ラベル / 」で示される文字列が挿入される事になる。

【 0 0 3 9 】

図 4 には、具体的な節境界検出ルール第 1 の例 1 0 4 も示されている。この例 1 0 4 は、入力される形態素列中に「けれども」という出現形で「助詞 接続助詞」という品詞を持つものがあれば、その部分を全て「けれども / 並列節ケレドモ / 」という文字列で置換せよ、というものである。

【 0 0 4 0 】

同じく図 4 には、具体的な節境界検出ルール第 2 の例 1 0 6 も示されている。この例 1 0 6 は、入力される形態素列中に「連用タ接続」または「連用形」という活用型の形態素があり、その直後に「たら」という出現形で「助動詞」という品詞を持ち、「特殊・タ」という活用形で「仮定形」という活用型を持つパターンがあれば、それらを全て、その末尾に「 / 条件節タラ / 」という文字列を付加したもので置換せよ、というものである。検索文字列中の「 | 」は、「または」という意味を表す。

【 0 0 4 1 】

本実施の形態では、この様な節境界検出ルールとして 3 6 1 個のルールを用いている。全てのルールは、1 個から 3 個の接続する形態素から構成されるパターンを持つ。入力には読点が含まれていない事を想定し、パターンに読点は含めていない。

【 0 0 4 2 】

図 5 に、本実施の形態で検出される節の種類の一部を示す。本実施の形態では、非特許文献 1 に記載されている従属節の形態（補足節、副詞節、連体節、および並列節）を増補及び改編して作成したもので、合計 1 4 4 種類の節を用いる。これらの中には、統語的に大きな切れ目になると考えられる主題「は」、談話標識、および感動詞を検出するパターンも含まれている。本明細書では、これらも含めて「節境界」と考える事とする。

【 0 0 4 3 】

本実施の形態で用いられている節ラベルは、実際には図 5 に示したものをさらに細かく分類したものである。例えば、「タメ節」の下位には「タメ二節」「タメ二八節」という節境界が設定してある。これら下位の節境界を合計すると 1 4 4 種類となるという事である。

【 0 0 4 4 】

図 6 に、プログラム 5 2 の実際の形式を示す。図 6 を参照して、プログラム 5 2 は、Perl の書式に従って Perl の処理系へのパスを示す行（1 行目）を含む。2 行目は、入力されるテキストが存在する限り次の中かっこ「 { 」および「 } 」に囲まれた部分の処理を繰返し実行する事を示す命令である。この中かっこの中が、上記した節境界検出ルールの本体である。入力があると、ここに記載された全てのグローバル置換命令を実行し、置換後のテキストを末尾の「 p r i n t 」命令によって標準出力に出力し、次の入力に対する処理に移る。

【 0 0 4 5 】

図 7 に、図 2 に示す言語処理系 5 6 およびプログラム 5 2 によって実現される節境界検出処理の実態をフローチャート形式で示す。言語処理系 5 6 自体は図 7 に示すものと異なり汎用的な機能を備えたものであるが、ここではプログラム 5 2 を言語処理系 5 6 で実行する場合についてのみ、その動きを示す。後述する様に、プログラム 5 2 および言語処理系 5 6 により実現される処理を専用のプログラムで実装する場合には、その制御構造は例えばこの図 7 に示されたものの様になる。

【 0 0 4 6 】

図 7 を参照してこの処理は、関連するファイル（入力ファイルおよび出力ファイル等）をオープンするステップ 1 4 0 と、入力テキストファイルの 1 行目（改行コードまでの文字列）を読み込むステップ 1 4 2 と、ステップ 1 4 2 の結果、入力ファイルの末尾（ E O F : E n d O f F i l e ）に達したか否かを判定するステップ 1 4 4 とを含む。判定結果が Y E S であれば制御はステップ 1 6 2 に進み、さもなければ制御はステップ 1 4 6 に進む。なお、複数の入力ファイルに対して連続してこの処理を実行する事もできるが、こ

10

20

30

40

50

では説明を簡明にするために一つのファイルに対して処理を行なうものとする。

【 0 0 4 7 】

ステップ 1 4 6 では、初期処理を行なう。初期処理では、入力されたテキストに対し、節境界の検出処理を行なう上で妨げとなる様な要素を入力テキストから除去する処理を行なう。

【 0 0 4 8 】

続いてステップ 1 4 8 では、1 番目の置換コマンドのグローバル検索を行なう。ステップ 1 5 0 では、プログラム 5 2 内の全ての置換コマンドを実行したか否かを判定する。全て実行が終わってれば制御はステップ 1 5 8 に進む。さもなければ制御はステップ 1 5 2 に進む。

10

【 0 0 4 9 】

ステップ 1 5 2 では、検索の結果、置換コマンドの検索文字列の正規表現にマッチした部分があったか否かを判定する。マッチがあれば制御はステップ 1 5 4 に進む。さもなければ制御はステップ 1 5 0 に戻る。

【 0 0 5 0 】

ステップ 1 5 4 では、マッチがあった部分を全て置換文字列で置換する処理が行なわれる。全て置換が終わったらステップ 1 5 6 で処理を次の置換コマンドに進めて制御をステップ 1 5 0 に戻す。

【 0 0 5 1 】

ステップ 1 5 0 で全ての置換コマンドの実行が完了したと判定された場合、制御はステップ 1 5 8 に進む。ステップ 1 5 8 では、置換処理が完了した 1 行分のテキストを標準出力に書き出す処理が実行される。続いて入力テキストファイルの次の 1 行を読み込む。制御はこの後ステップ 1 4 4 に戻る。

20

【 0 0 5 2 】

一方、ステップ 1 4 4 で入力ファイルの E O F に到達したと判定された場合、ステップ 1 6 2 で関連のファイルを全てクローズして処理を終了する。

【 0 0 5 3 】

動作

この機械翻訳装置は以下の様に動作する。図 2 を参照して、ユーザが起動コマンド 3 2 を入力したものとする。起動コマンド 3 2 は、入力テキスト 3 4 とプログラム 5 2 とを特定する情報を含む。

30

【 0 0 5 4 】

O S 5 0 はこのコマンドに回答して形態素解析部 5 4 を起動し、入力テキスト 3 4 を開いて形態素解析部 5 4 で形態素解析を行なわせる。一方 O S 5 0 は、起動コマンド 3 2 により特定されるプログラム 5 2 を記憶装置から読出す。前述の通り、プログラム 5 2 の 1 行目にはこのプログラム 5 2 を実行するための言語処理系へのパスが記載されている。O S 5 0 はこのパスに従って言語処理系 5 6 を起動する。

【 0 0 5 5 】

形態素解析部 5 4 から出力される形態素列は言語処理系 5 6 に与えられる。言語処理系 5 6 は、この形態素列に対してプログラム 5 2 に含まれる節境界検出ルールを適用し、テキスト中の節境界に節ラベルを挿入する処理を行ない、結果をバッファ 5 8 に出力する。

40

【 0 0 5 6 】

テキスト分離部 6 0 は、言語処理系 5 6 から節ラベルが出力されるごとに、バッファ 5 8 に格納されたテキストを読出し、機械翻訳部 6 2 に与える。

【 0 0 5 7 】

機械翻訳部 6 2 は、与えられる節について機械翻訳を行ない、結果を翻訳出力 3 6 として出力する。

【 0 0 5 8 】

処理例

図 8 を参照して、テキスト 1 9 0 に対して節境界検出処理を行なった。その結果を処理後

50

のテキスト192として示す。処理後のテキスト192は、節境界に対応する形態素列パターンが検出された場所に挿入された節ラベルを含んでいる。たとえば「自主避難が呼びかけられている ×町の 地区では」という部分は「自主避難が呼びかけられている」という節と「 ×町の 地区では」以下の節とに分離されている。そして、「自主避難が呼びかけられている」という節には「連体節」という節ラベルが付されている。この節ラベルはスラッシュによって本文と区切られて挿入されている。

【0059】

性能評価のための実験

本実施の形態に係るプログラム52および言語処理系56により実装した節境界検出装置の性能を評価するために、性質の異なる複数のコーパスに対してルールを適用し、その結果を分析した。用意したコーパスの概略の規模を図9に示す。

10

【0060】

図9に示される様に、コーパスは全部で5つ用意した。そのうち3つは独話コーパスであり、2つは対話コーパスである。

【0061】

第1の独話コーパスは放送でのいわゆる解説番組を書き起こしたものである。第2の独話コーパスはテレビジョン放送でのニュースの原稿コーパスである。第3の独話コーパスは経済系の複数の新聞記事データベースである。一方、第1の対話コーパスは、出願人において準備したバイリンガルの旅行会話を題材とする模擬会話コーパスである。第2の対話コーパスは、海外旅行で用いられる典型的な表現を収集したコーパスである。

20

【0062】

図9を参照して、1文の長さは第2の独話コーパスが突出して長く、第1および第3の独話コーパスがこれに次ぐ事が分かる。これに比して対話コーパス中の文はいずれも極端に短い事が分かる。

【0063】

これらコーパスに上記した節境界検出処理を行なった。検出された節の数、1文に含まれる平均節数、各節に含まれる平均形態素数と平均文節数とを図10に示す。図10から、節境界検出処理によって検出された一つの節の長さ(形態素数および文節数)は、独話、対話を問わずコーパス間でほとんど差がない事が分かる。

【0064】

評価

さらに節境界検出装置の性能を評価するため、各コーパスから500文を選択し、人手で節境界の検出と判定とを行ない、正解データを作成した。上記した節境界検出装置による節境界検出処理の結果と正解データとを照合し、適合率と再現率とを求めた。その結果を図11に表形式で示す。

30

【0065】

図11を参照して、全てのコーパスにおいて、適合率と再現率ともに非常に高く、非常によい精度で節境界が検出されている事が分かる。この様により精度で節境界を検出し、節ごとに翻訳処理を行なう事で、機械翻訳の精度も高くなり、結果として良好な翻訳を得る事が可能になる。しかも上記した処理では、形態素列が所定の節境界パターンにマッチすれば節境界が検出できる。文末の入力が行なわれなくても漸進的に節の検出を行なう事ができる。そのため、同時翻訳等に適している。

40

【0066】

節境界検出ルールの実例

以下に、実験で実際に使用した節境界検出ルール(P e r lの置換コマンド形式)を示す。ここでは、ルールに相当する置換コマンドのみを示し、スクリプトの制御に属する部分は省略してある。また、実際のスクリプトにおいては1行で記載されるべきところを複数行に分けて記載した部分がある。

【0067】

<ルールの開始>

50

【 0 0 6 8 】

【表 1】

並列節

s/(が_助詞-接続助詞_) /\$1 ¥/並列節カ¥/ /g;

s/((けれど|けれど|けども|けど)_助詞-接続助詞_) /\$1 ¥/並列節ケレド
モ¥/ /g;

s/¥/並列節ケレドモ¥/ (も_助詞-係助詞_) /\$1 ¥/並列節ケレドモ¥/ /g;

s/(し_助詞-接続助詞_) /\$1 ¥/並列節シ¥/ /g;

s/(¥/並列節(ガ|ケレドモ|シ)¥/) (([^_]*)_助詞-終助詞_) /\$3 ¥/並列節\$
2-助詞¥/ /g;

s/(で_助動詞_特殊・ダ_連用形) /\$1 ¥/並列節デ¥/ /g;

s/¥/並列節デ¥/ ((([^_]*)_助動詞_五段・ラ行アル_([^_]*))|(([^_]*)_動
詞-(自立|非自立)_五段・ラ行特殊_([^_]*))|(ごぞい_(動詞-自立|助動詞)_五段
・ラ行特殊_連用形)|(ない_形容詞-自立_形容詞・アウオ段_基本形)|(([^_]*)_
助動詞_特殊・ナイ_([^_]*))|(ない_動詞-自立_五段・ワ行促音便_連用形)|((は
|も)_助詞-係助詞_)) /\$1 /g;s/((基本形|命令形|仮定形|助詞-副助詞/並立助詞/終助詞_) ((とか_助
詞-並立助詞_)|(と_助詞-(格助詞-引用|接続助詞)_ か_助詞-副助詞/並立助
詞/終助詞_))) /\$1 ¥/並列節トカ¥/ /g;

s/((連用形|連用タ接続) たり_助詞-並立助詞_) /\$1 ¥/並列節タリ¥/ /g;

s/¥/並列節(トカ|タリ)¥/ ((で_助詞-格助詞-一般_)|(も_助詞-係助詞_)|
((とか_助詞-並立助詞_)|(と_助詞-(格助詞-引用|接続助詞)_ か_助詞-副助詞
/並立助詞/終助詞_))) /\$2 ¥/並列節\$1-助詞¥/ /g;# s/¥/(並列節([_^]*))¥/ (と_助詞-格助詞-(一般|引用)_) /\$3 ¥/\$1引用¥/
/g;

理由節

s/(から_助詞-接続助詞_) /\$1 ¥/理由節カラ¥/ /g;

【 0 0 6 9 】

【表 2】

10

20

30

40

s/((基本形 (の|んで)_助詞-接続助詞__)|(基本形 (の|ん)_名詞-非自立-一般__で_助動詞_特殊・ダ_連用形)|(基本形 (の|ん)_名詞-非自立-一般__で_助詞-格助詞-一般__)|(な_助動詞_特殊・ダ_体言接続 ((の|んで)_助詞-接続助詞__|(の|ん)_名詞-非自立-一般__で_助詞-格助詞-一般__))) /\$1 ¥/理由節ノデ¥/ /g;

s/(¥/理由節ノデ¥/) ¥/並列節デ¥/ /\$1 /g;

s/¥/理由節ノデ¥/ (([^_]*)_助詞-([^_]*)__|(ある|あり|あれ|あろ)_助動詞_五段・ラ行アル_([^_]*)) /\$1 /g;

10

s/¥/(理由節([^_]*))¥/ (と_助詞-格助詞-(一般|引用)__)/\$3 ¥/\$1引用¥/ /g;

s/¥/(理由節([^_]*)引用)¥/ (か_助詞-副助詞/並立助詞/終助詞__)/\$3 ¥/並列節トカ¥/ /g;

20

引用節

s/((基本形|助詞-副助詞/並立助詞/終助詞__|命令([^_]*)|助詞-終助詞__|感動詞__) (と|って)_助詞-格助詞-(引用|一般)__ /\$1 ¥/引用節¥/ /g;

s/((基本形|助詞-副助詞/並立助詞/終助詞__|命令([^_]*)|助詞-終助詞__|感動詞__) など_助詞-副助詞__ (と|って)_助詞-格助詞-(引用|一般)__ /\$1 ¥/引用節¥/ /g;

30

s/((基本形|助詞-副助詞/並立助詞/終助詞__|命令([^_]*)|助詞-終助詞__|感動詞__) (「|」)_記号-括弧閉__ (と|って)_助詞-格助詞-(引用|一般)__ /\$1 ¥/引用節¥/ /g;

s/((基本形|助詞-副助詞/並立助詞/終助詞__|命令([^_]*)|助詞-終助詞__|感動詞__) (「|」)_記号-括弧閉__ など_助詞-副助詞__ (と|って)_助詞-格助詞-(引用|一般)__ /\$1 ¥/引用節¥/ /g;

40

s/¥/引用節¥/ ((は|も)_助詞-係助詞__) /\$1 ¥/引用節¥/ /g;

【 0 0 7 0 】

【 表 3 】

s/¥/引用節¥/ (の_助詞-連体化__) /\$1 ¥/引用節トノ¥/ /g;

s/¥/引用節¥/ (([^_]*)_助詞-格助詞-一般__) /\$1 ¥/引用節-補足語¥/ /g;

s/¥/引用節¥/ (([^_]*)_助詞-([_^]*)) /\$1 ¥/引用節-助詞¥/ /g;

s/(か_助詞-副助詞/並立助詞/終助詞__) ¥/引用節-助詞¥/ (¥/並列節(トカ|トカ-助詞)¥/) /\$1 \$2 /g;

s/¥/引用節-(補足語|助詞)¥/ (([^_]*)_助詞-([_^]*)) /\$2 ¥/引用節-助詞¥/ /g;

10

s/¥/引用節-(補足語|助詞)¥/ (([^_]*)_助詞-格助詞-一般__) /\$2 ¥/引用節-助詞¥/ /g;

s/¥/引用節-(補足語|助詞)¥/ (([^_]*)_助詞-(係|副|終)助詞__) /\$2 ¥/引用節-助詞¥/ /g;

s/¥/引用節-助詞¥/ (¥/並列節(トカ|トカ-助詞)¥/) /\$1 /g;

s/¥/引用節¥/ ((いう_動詞-自立_五段・ワ行促音便_基本形)| (いっ_動詞-自立_五段・ワ行促音便_連用タ接続 た_助動詞_特殊・タ_基本形)) /\$1 /g;

20

s/((基本形|助詞-副助詞/並立助詞/終助詞__|命令([_^]*)|助詞-終助詞__) (と|って)_助詞-格助詞-(引用|一般__) (いう_動詞-自立_五段・ワ行促音便_基本形 (の|ん)_名詞-非自立-一般__) /\$1 ¥/引用節¥/ \$6 /g;

連体節トイウ

s/((基本形|助詞-副助詞/並立助詞/終助詞__|命令([_^]*)|助詞-終助詞__) ((という|といった|っていう)_助詞-格助詞-連語__|(など_助詞-副助詞__ といった_助詞-格助詞-連語__))) /\$1 ¥/連体節トイウ¥/ /g;

30

s/((基本形|助詞-副助詞/並立助詞/終助詞__|命令([_^]*)|助詞-終助詞__) と_助詞-格助詞-引用__ (いう_動詞-自立_五段・ワ行(促|ウ)音便_基本形|いっ_動詞-自立_五段・ワ行促音便_連用タ接続 た_助動詞_特殊・タ_基本形) /\$1 \$4 ¥/連体節トイウ¥/ /g;

40

s/((基本形|助詞-副助詞/並立助詞/終助詞__|命令([_^]*)|助詞-終助詞__) (「|」)_記号-括弧閉__ ((という|といった|っていう)_助詞-格助詞-連語__|(

【 0 0 7 1 】

【 表 4 】

など_助詞-副助詞__ といった_助詞-格助詞-連語__))) /\$1 ¥/連体節トイウ¥/ /g;

s/((基本形|助詞-副助詞/並立助詞/終助詞__|命令([^_]*)|助詞-終助詞__) (「|」)_記号-括弧閉__ と_助詞-格助詞-引用__) ¥/引用節¥/ (いう_動詞-自立_五段・ワ行促音便_基本形|いっ_動詞-自立_五段・ワ行促音便_連用夕接続 た_助動詞_特殊・タ_基本形) /\$1 \$5 ¥/連体節トイウ¥/ /g;

10

s/¥/連体節トイウ¥/ ((の|ん)_名詞-非自立-一般__|([^_]*)_助詞-([^_]*)_ |よう_名詞-非自立-助動詞語幹__) /\$1 /g;

間接疑問節

s/((基本形|(基本形 (の|ん)_名詞-非自立-一般__)|助動詞_文語・ベシ_体言接続)か_助詞-副助詞/並立助詞/終助詞__) /\$1 ¥/間接疑問節¥/ /g;

20

s/(¥/間接疑問節¥/) (どう_副詞-助詞類接続__ か_助詞-副助詞/並立助詞/終助詞__) /\$2 \$1 /g;

s/¥/間接疑問節¥/ ((と|って)_助詞-格助詞-(一般|引用)__ ¥/引用節¥/) /\$1 /g;

s/¥/間接疑問節¥/ (([^_]*)_助詞-格助詞-(一般|連語)__) /\$1 ¥/間接疑問節-補足語¥/ /g;

30

s/¥/(間接疑問節|間接疑問節-補足語)¥/ (((は|も)_助詞-係助詞__)|((など|とか)_助詞-副助詞__)|(と_助詞-格助詞-引用__ か_助詞-副助詞/並立助詞/終助詞__)) /\$2 ¥/間接疑問節-助詞¥/ /g;

s/¥/(間接疑問節|間接疑問節-(補足語|助詞))¥/ (([^_]*)_助詞-格助詞-(一般|引用|連語)__) /\$3 ¥/間接疑問節-助詞¥/ /g;

s/¥/(間接疑問節|間接疑問節-(補足語|助詞))¥/ (([^_]*)_助詞-(係|副|並立)助詞__) /\$3 ¥/間接疑問節-助詞¥/ /g;

40

s/¥/(間接疑問節|間接疑問節-(補足語|助詞))¥/ (([^_]*)_助詞-連体化__) /\$3 ¥/間接疑問節-助詞¥/ /g;

【 0 0 7 2 】

【表 5】

s/¥/間接疑問節-(補足語|助詞)¥/ ((いう_動詞-自立_五段・ワ行促音便_基本形| (いっ_動詞-自立_五段・ワ行促音便_連用タ接続 た_助動詞_特殊・タ_基本形)) ¥/連体節トイウ¥/) /\$2 /g;

s/¥/(間接疑問節|間接疑問節-(補足語|助詞))¥/ ((〔^_〕*)_記号-(〔^_〕*)_ |¥/(〔^¥/〕*)¥/) /\$3 /g;

s/¥/(間接疑問節|間接疑問節-(補足語|助詞))¥/ (¥/(〔^¥/〕*)¥/) /\$3 /g;

10

s/¥/間接疑問節¥/ ((〔^_〕*)_助詞-終助詞__) /\$1 /g;

テ節

s/((連用形|連用(タ|テ|デ)接続) (て|で)_助詞-接続助詞__) /\$1 ¥/テ節¥/ /g;

s/(基本形 (として|にあたって|につれて)_助詞-格助詞-連語__) /\$1 ¥/テ節¥/ /g;

20

s/(基本形 (「|」)_記号-括弧閉__ (として|にあたって|につれて)_助詞-格助詞-連語__) /\$1 ¥/テ節¥/ /g;

s/(基本形 (こと|もの)_名詞-非自立-一般__ (として|について)_助詞-格助詞-連語__) /\$1 ¥/テ節¥/ /g;

s/((にわたって|に関して|に際して|に従って|に対して|に当たって|をめぐって|をもって|を通して|を通じて|にかけて|によって)_助詞-格助詞-連語__) /\$1 ¥/テ節¥/ /g;

30

s/(((基本形 など_助詞-副助詞__) |依然_副詞-一般__) として_助詞-格助詞-連語__) /\$1 ¥/テ節¥/ /g;

s/(基本形 (「|」)_記号-括弧閉__ など_助詞-副助詞__ として_助詞-格助詞-連語__) /\$1 ¥/テ節¥/ /g;

s/に対し_助詞-格助詞-連語__ まして_副詞-一般__ /に_助詞-格助詞-一般__ 対し_動詞-自立_五段・サ行_連用形 まし_助動詞_特殊・マス_連用形 て_助詞-接続助詞__ ¥/テ節¥/ /g;

40

s/¥/テ節¥/ ((〔^_〕*)_形容詞-非自立_(〔^_〕*)_ (〔^_〕*)) /\$1 /g;

【 0 0 7 3 】

【 表 6 】

s/¥/テ節¥/ (から_助詞-格助詞-一般__) /\$1 ¥/テカラ節¥/ /g;
 s/¥/テカラ節¥/ (([^_]*)_助詞-([^_]*)) /\$1 ¥/テカラ節-助詞¥/ /g;
 s/¥/テ節¥/ (は_助詞-係助詞__) /\$1 ¥/テハ節¥/ /g;
 s/¥/テ節¥/ (も_助詞-係助詞__) /\$1 ¥/譲歩節テモ¥/ /g;
 s/¥/テ節¥/ (の_助詞-連体化__) /\$1 ¥/連体節テノ¥/ /g;
 s/(¥/(テ節|テハ節|譲歩節テモ|連体節テノ)¥/) ¥/([^_]*)-助詞¥/ /\$1 /g;

10

s/¥/テ節¥/ (か_助詞-副助詞__) /\$1 ¥/間接疑問節¥/ /g;

s/¥/テハ節¥/ (なく_(形容詞-自立_形容詞・アウオ段_連用テ接続|助動詞_特殊・ナイ_連用テ接続)) /\$1 /g;

譲歩節

20

s/(基本形 (のに_助詞-接続助詞__|(の_名詞-非自立-一般__ に_助詞-格助詞-一般__))) /\$1 ¥/譲歩節ノニ¥/ /g;

s/¥/譲歩節ノニ¥/ (は_助詞-係助詞__) /\$1 ¥/譲歩節ノニ-助詞¥/ /g;

s/¥/譲歩節ノニ¥/ (は_助詞-係助詞__) /\$1 /g;

条件節

30

s/(基本形 と_助詞-接続助詞__) /\$1 ¥/条件節ト¥/ /g;

s/¥/条件節ト¥/ (か_助詞-副助詞/並立助詞/終助詞__ ¥/並列節トカ¥/) /\$1 /g;

s/(仮定形 ば_助詞-接続助詞__) /\$1 ¥/条件節バ¥/ /g;

s/¥/条件節バ¥/ ((なら|なり)_動詞-非自立_五段・ラ行_([^_]*)) /\$1 /g;

s/((基本形|(の|ん)_名詞-非自立-一般__) なら_助動詞_特殊・タ_仮定形) /\$1 ¥/条件節ナラ¥/ /g;

40

s/¥/条件節ナラ¥/ (ば_助詞-接続助詞__) ¥/条件節バ¥/ /\$1 ¥/条件節ナラバ¥/ /g;

【 0 0 7 4 】

【 表 7 】

s/((連用タ接続|連用形) たら_助動詞_特殊・タ_仮定形) /\$1 ¥/条件節タラ
¥/ /g;

s/¥/条件節タラ¥/ (ば_助詞-接続助詞_) ¥/条件節バ¥/ /\$1 ¥/条件節タラ
バ¥/ /g;

s/¥/(条件節([_^]*))¥/ (と_助詞-格助詞-(一般|引用)_) /\$3 ¥/\$1-引用¥/
/g;

10

連用節

s/((動詞-([_^]*)|助動詞)_([_^]*)_ (連用形|連用ニ接続|命令 e)) (([_^]*
)_((名詞-([_^]*)_)|(接続詞_)|(連体詞_)|(副詞-([_^]*)_)|(接頭詞-([_^]*
_)|((動詞|形容詞)-自立_([_^])_([_^]*))|(記号-(括弧開|アルファベット)
)|感動詞|未知語_)) /\$1 ¥/連用節¥/ \$6 /g;

20

s/(((で_(助動詞_特殊・ダ_連用形|助詞-(格助詞-一般|接続助詞)_)は_助
詞-係助詞_)|(は_助詞-係助詞_)|で_助動詞_特殊・ダ_連用形) なく_(助動詞_
特殊・ナイ_連用テ接続)) (([_^]*)_((名詞-([_^]*)_)|(接続詞_)|(連体詞_)
|(副詞-([_^]*)_)|(接頭詞-([_^]*)_)|((動詞|形容詞)-自立_([_^]*)_([_^]*
)|(記号-(括弧開|アルファベット))|(未知語_))) /\$1 ¥/連用節¥/ \$8 /g;

30

s/(ず_助動詞_特殊・又_連用ニ接続 に_助詞-格助詞-一般_) /\$1 ¥/連用節
¥/ /g;

s/(に_助詞-格助詞-一般_) ¥/連用節¥/ (は_助詞-係助詞_) /\$1 \$2 ¥/連
用節¥/ /g;

s/¥/連用節¥/ (([_^]*)_名詞-接尾-([_^]*)_) /\$1 /g;

s/¥/連用節¥/ (¥/(並列節デ|理由節ノデ|テ節)¥/) /\$1 /g;

40

s/(基本形 に_助詞-格助詞-一般_) /\$1 ¥/連用節その他¥/ /g;

形容詞連用節

【 0 0 7 5 】

【 表 8 】

s/((が|は)_助詞-(格助詞-([^_]*)|係助詞) __ なく_形容詞-自立_形容詞・
アウオ段_連用テ接続) (([^_]*)_((名詞-([^_]*))|(接続詞__)|(連体詞__)|(
副詞-([^_]*))|(接頭詞-([^_]*))|(動詞-自立_([^_]*)_([^_]*)|(記号-(括
弧開|アルファベット)__)|(未知語__))) /\$1 ¥/形容詞連用節¥/ \$5 /g;

連用節 その他

s/(基本形 に_助詞-格助詞-一般__) ¥/連用節その他¥/ (も_助詞-係助詞__
関わら_動詞-自立_五段・ラ行_未然形 ず_助動詞_特殊・又_連用ニ接続) ¥/連用
節¥/ /\$1 \$2 ¥/連用節その他¥/ /g;

10

s/基本形 にもかかわらず_接続詞__ /基本形 に_助詞-格助詞-一般__ も_助
詞-係助詞__ かかわら_動詞-自立_五段・ラ行_未然形 ず_助動詞_特殊・又_連用
ニ接続 ¥/連用節その他¥/ /g;

s/¥/間接疑問節-補足語¥/ (関わら_動詞-自立_五段・ラ行_未然形 ず_助動
詞_特殊・又_連用ニ接続) ¥/連用節¥/ /\$1 ¥/連用節その他¥/ /g;

20

s/(基本形 (とともに|と共に)_助詞-格助詞-連語__) /\$1 ¥/連用節その他¥/
/g;

s/(を_助詞-格助詞-一般__ はじめ_名詞-副詞可能__) /\$1 ¥/連用節その他¥
/ /g;

30

s/(の_名詞-非自立-一般__ を_助詞-格助詞-一般__ 始め_動詞-自立_一段_
連用形) ¥/連用節¥/ (([^_]*)_((名詞-([^_]*))|(接続詞__)|(連体詞__)|(
副詞-([^_]*))|(接頭詞-([^_]*))|((動詞|形容詞)-自立_([^_]*)_([^_]*))|(
記号-(括弧開|アルファベット)__)|(未知語__))) /\$1 ¥/連用節その他¥/ \$2 /g;

s/(に対し_助詞-格助詞-連語__) (([^_]*)_((名詞-([^_]*))|(接続詞__)|
(連体詞__)|副詞-([^_]*))|(接頭詞-([^_]*))|((動詞|形容詞)-自立_([^_]
)_([^_]))|(記号-(括弧開|アルファベット)__)|(未知語__))) /\$1 ¥/連用節¥/
\$2 /g;

40

【 0 0 7 6 】

【 表 9 】

s/(と_助詞-接続助詞_) ¥/条件節ト¥/ (同時に_副詞-一般_) /\$1 \$2 ¥/連用節その他¥/ /g;

s/¥/連用節¥/ ((次第|しだい)_名詞-(副詞可能|非自立-副詞可能)_) /\$1 ¥/連用節その他¥/ /g;

s/(基本形 にせよ_動詞-自立_一段_命令 y o) /\$1 ¥/譲歩節命令形¥/ /g;

s/(基本形 に_助詞-格助詞-一般_) ¥/連用節その他¥/ (しろ_動詞-自立_サ変・スル_命令 r o) /\$1 \$2 ¥/譲歩節命令形¥/ /g;

10

s/¥/譲歩節命令形¥/ (と_助詞-格助詞-(一般|引用)_) /\$1 /g;

s/(¥/連用節その他¥/) (([^_])*_助詞-([^_])*_) /\$1 \$2 /g;

20

s/(お_接頭詞-名詞接続_ ([^_])*_動詞-自立_([^_])*_連用形) ¥/連用節¥/ (([^_])*_動詞-自立_サ変・スル_([^_])*_) /\$1 \$4 /g;

ナド節

s/(基本形 など_助詞-副助詞_) /\$1 ¥/ナド節¥/ /g;

s/¥/ナド節¥/ (の_助詞-連体化_) /\$1 ¥/連体節ナドノ¥/ /g;

s/¥/ナド節¥/ (([^_])*_助詞-格助詞-一般_) /\$1 ¥/ナド節-補足語¥/ /g;

30

s/¥/(ナド節|ナド節-(引用|補足語))¥/ (([^_])*_助詞-([^_])*_) /\$3 ¥/ナド節-助詞¥/ /g;

s/¥/ナド節-(助詞|補足語)¥/ (¥/([^_])*¥/) /\$2 /g;

s/¥/ナド節¥/ (といった_助詞-格助詞-連語_) /\$1 /g;

ナガラ節

40

s/(連用形 ながら_助詞-接続助詞_) /\$1 ¥/ナガラ節¥/ /g;

s/¥/ナガラ節¥/ (も_助詞-係助詞_) /\$1 ¥/ナガラモ節¥/ /g;

s/¥/ナガラ節¥/ (と_助詞-格助詞-(引用|一般)_) /\$1 ¥/ナガラ節-引用¥/

【 0 0 7 7 】

【 表 1 0 】

/g;

s/¥/(ナガラ節|ナガラ節-引用)¥/ (([^_]*)_助詞-([^_]*)_) /\$2 ¥/ナガラ節-助詞¥/ /g;

ツツ節

s/(連用形 つつ_助詞-接続助詞_) /\$1 ¥/ツツ節¥/ /g;

s/¥/ツツ節¥/ (([^_]*)_助詞-([^_]*)_) /\$1 ¥/ツツ節-助詞¥/ /g;

10

ダケ節

s/(基本形 だけ_助詞-副助詞_) /\$1 ¥/ダケ節¥/ /g;

s/¥/ダケ節¥/ (に_助詞-格助詞-一般_) /\$1 ¥/ダケニ節¥/ /g;

s/¥/ダケ節¥/ (が_助詞-格助詞-一般_) /\$1 ¥/ダケ節-補足語¥/ /g;

s/¥/ダケ節¥/ (と_助詞-格助詞-(一般|引用)_) /\$1 ¥/ダケ節-引用¥/ /g;

20

s/¥/(ダケ節|ダケ節-引用)¥/ (([^_]*)_助詞-([^_]*)_) /\$2 ¥/ダケ節-助詞¥/ /g;

s/¥/ダケニ節¥/ (([^_]*)_助詞-([^_]*)_) /\$2 ¥/ダケニ節-助詞¥/ /g;

s/(¥/連体節-形式名詞¥/ ([^_]*)_名詞-非自立-一般_ だけ_助詞-副助詞_ に_助詞-格助詞-一般_) /\$1 ¥/ダケニ節¥/ /g;

30

マデ節

s/(基本形 まで_助詞-副助詞_) /\$1 ¥/マデ節¥/ /g;

s/¥/マデ節¥/ (に_助詞-格助詞-一般_) /\$1 ¥/マデニ節¥/ /g;

s/¥/マデニ節¥/ (は_助詞-係助詞_) /\$1 ¥/マデニハ節¥/ /g;

s/¥/マデ節¥/ (の_助詞-連体化_) /\$1 ¥/連体節マデノ¥/ /g;

40

s/¥/マデ節¥/ (は_助詞-係助詞_) /\$1 ¥/マデハ節¥/ /g;

s/¥/マデ節¥/ (も_助詞-係助詞_) /\$1 ¥/マデモ節¥/ /g;

s/¥/マデ節¥/ (と_助詞-格助詞-(一般|引用)_) /\$1 ¥/マデ節-引用¥/ /g;

【 0 0 7 8 】

【 表 1 1 】

s/¥/マデモ節¥/ (なく_形容詞-自立_形容詞・アウオ段_連用テ接続) /\$1 ¥/
連用節その他¥/ /g;

ヨウ節, ヨウニ節

s/(基本形 よう_名詞-(非自立-助動詞語幹|接尾-一般)_) /\$1 ¥/ヨウ節¥/
/g;

s/¥/ヨウ節¥/ (な_助動詞_特殊・タ_体言接続) /\$1 ¥/連体節ヨウナ¥/ /g;

10

s/¥/ヨウ節¥/ (に_助詞-(格助詞-一般|副詞化)_) /\$1 ¥/ヨウニ節¥/ /g;

s/(か_助詞-副助詞/並立助詞/終助詞__ の_助詞-連体化__) ¥/間接疑問
節-助詞¥/ (よう_名詞-非自立-助動詞語幹__ に_助詞-副詞化__) /\$1 \$2 ¥/ヨウ
ニ節¥/ /g;

s/¥/ヨウニ節¥/ (と_助詞-格助詞-(一般|引用)_) /\$1 ¥/ヨウニ節-引用¥/
/g;

20

s/¥/ヨウ節¥/ (([^_]*)_助詞-([^_]*)) /\$1 ¥/ヨウ節-助詞¥/ /g;

s/¥/(ヨウニ節|ヨウニ節-引用)¥/ (([^_]*)_助詞-([^_]*)) /\$2 ¥/ヨウニ
節-助詞¥/ /g;

s/¥/ヨウ節¥/ (([^_]*)_助動詞-([^_]*)([^_]*)) /\$1 /g;

s/¥/ヨウナ節¥/ (の_名詞-非自立-一般) /\$1 /g;

ヨリ節

30

s/(基本形 より_助詞-格助詞-一般) /\$1 ¥/ヨリ節¥/ /g;

s/¥/連体節トイウ¥/ (より_助詞-格助詞-一般) /\$1 ¥/ヨリ節¥/ /g;

s/¥/ヨリ節¥/ (([^_]*)_助詞-([^_]*)) /\$1 ¥/ヨリ節-助詞¥/ /g;

s/(¥/ヨリ節-助詞¥/) (([^_]*)_助詞-([^_]*)) /\$2 \$1 /g;

連体節

40

s/((動詞-([^_]*)|助動詞-([^_]*)_基本形) (([^_]*)-((名詞-([^_]*)) |
(接続詞) | (連体詞) | (副詞-([^_]*)) | (接頭詞-([^_]*)) | (形容詞-自立-
([^_]*)([^_]* | (未知語) | (記号-(括弧開|アルファベット)_____))) /\$1 ¥/連体

【 0 0 7 9 】

【 表 1 2 】

節¥/ \$5 /g;

s/(文語・(〔^_〕*)_体言接続) ((〔^_〕*)_名詞-(〔^_〕*)_) /\$1 ¥/連体節¥/ \$

3 /g;

s/((と_助詞-格助詞-引用__|(と_助詞-格助詞-引用__ ¥/(〔^¥/〕*)¥/)) ((い
う_動詞-自立_五段・ワ行促音便_基本形)| (いっ_動詞-自立_五段・ワ行促音便_
連用タ接続 た_助動詞_特殊・タ_基本形))) ¥/(連体節|連体節-形式名詞)¥/ (¥/
連体節トイウ¥/) /\$1 \$9 /g;

10

s/((に|対する|にあたる|に関する)_助詞-格助詞-連語__) /\$1 ¥/連体節¥/ /
g;

s/(か_助詞-副助詞/並立助詞/終助詞__ ((に|対する|にあたる|に関する)_
助詞-格助詞-連語__) ¥/連体節¥/) ¥/間接疑問節-補足語¥/ /\$1 /g;

20

s/(を_助詞-格助詞-一般__ はじめ_名詞-副詞可能__) ¥/連用節その他¥/ (と_助詞-格助詞-一般__ (する_動詞-自立_サ変・スル_基本形|(し_動詞-自立_サ
変・スル_連用形 た_助動詞_特殊・タ_基本形))) /\$1 \$2 /g;

s/((動詞-(〔^_〕*)|助動詞)_ (〔^_〕*)_基本形 (「|」)_記号-括弧閉__) ((〔^_
〕*) ((名詞-(〔^_〕*)_)|(接続詞__)|(連体詞__)|(副詞-(〔^_〕*)_)|(接頭詞-(〔^_
〕*))|(形容詞-自立_(〔^_〕*)_ (〔^_〕*)_|(未知語__)|(記号-(括弧開|アルファベ
ット)__)))) /\$1 ¥/連体節¥/ \$6 /g;

30

s/(((〔^_〕*)_動詞-自立_(〔^_〕*)_連用形 (〔^_〕*)_形容詞-非自立_(〔^_〕*)_
基本形) ((〔^_〕*)_ ((名詞-(〔^_〕*)_)|(接続詞__)|(連体詞__)|(副詞-(〔^_〕*)_)
|(接頭詞-(〔^_〕*)_)|(形容詞-自立_(〔^_〕*)_ (〔^_〕*)_|(未知語__)|(記号-(括弧
開|アルファベット)__)))) /\$1 ¥/連体節¥/ \$6 /g;

40

s/¥/連体節¥/ ((の|ん|こと|事|もの|もん|ところ|所|はず|つもり|わけ|訳

【 0 0 8 0 】

【 表 1 3 】

)_名詞-非自立-(一般|副詞可能)_) /¥/連体節-形式名詞¥/ \$1 /g;

形容詞連体節

s/(の_助詞-格助詞-一般__ ([^_]*)_形容詞-自立_([^_]*)_基本形) (([^_]*
)_((名詞-([^_]*))|(接続詞__)|(連体詞__)|(副詞-([^_]*))|(接頭詞-([^_]
)_)|(形容詞-自立_([^_])_([^_]*)|(未知語__)|(記号-(括弧開|アルファベッ
ト)_____)) /\$1 ¥/形容詞連体節¥/ \$4 /g;

10

s/(の_助詞-格助詞-一般__ ([^_]*)_形容詞-自立_([^_]*)_基本形) ¥/形容
詞連体節¥/ ((の|ん|こと|事|もの|もん|ところ|所|はず|つもり|わけ|訳)_名
詞-非自立-(一般|副詞可能)_) /\$1 ¥/形容詞連体節-形式名詞¥/ \$4 /g;

s/((が_助詞-格助詞-一般__|の_助詞-連体化__) ない_形容詞-自立_形容詞
・アウオ段_基本形) (([^_]*)_((名詞-([^_]*))|(接続詞__)|(連体詞__)|(副
詞-([^_]*))|(接頭詞-([^_]*))|(形容詞-自立_([^_]*)_([^_]*)|(未知語__)
|(記号-(括弧開|アルファベット)_____)) /\$1 ¥/形容詞連体節¥/ \$3 /g;

20

s/((が_助詞-格助詞-一般__|の_助詞-連体化__) ない_形容詞-自立_形容詞
・アウオ段_基本形) ¥/形容詞連体節¥/ ((の|ん|こと|事|もの|もん|ところ|所|
はず|つもり|わけ|訳)_名詞-非自立-(一般|副詞可能)_) /\$1 ¥/形容詞連体節-
形式名詞¥/ \$3 /g;

30

形容動詞連体節

s/(の_助詞-格助詞-一般__ ([^_]*)_名詞-形容動詞語幹__ な_助動詞_特殊
・ダ_体言接続) (([^_]*)_((名詞-([^_]*))|(接続詞__)|(連体詞__)|(副詞-([
^_]*))|(接頭詞-([^_]*))|(形容詞-自立_([^_]*)_([^_]*)|(未知語__)|(記
号-(括弧開|アルファベット)_____)) /\$1 ¥/形容動詞連体節¥/ \$3 /g;

s/(の_助詞-格助詞-一般__ ([^_]*)_名詞-形容動詞語幹__ な_助動詞_特殊
・ダ_体言接続) ¥/形容動詞連体節¥/ ((の|ん|こと|事|もの|もん|ところ|所|は
ず|つもり|わけ|訳)_名詞-非自立-(一般|副詞可能)_) /\$1 ¥/形容詞連体節-形
式名詞¥/ \$3 /g;

40

【 0 0 8 1 】

【 表 1 4 】

s/¥/(連体節|形容詞連体節|形容動詞連体節)¥/ (([^_])*_名詞-(非自立|特殊)-助動詞語幹__)/\$2 /g;

s/¥/(連体節|形容詞連体節|形容動詞連体節)¥/ (よう_名詞-接尾-一般__)/\$2 /g;

ノダ

s/¥/(連体節|形容詞連体節|形容動詞連体節)-形式名詞¥/ ((の|ん)_名詞-非自立-一般__ (([^_])*_(動詞-([^_])*|助動詞)_([^_])*_([^_])*|か_助詞-副助詞/並立助詞/終助詞__|かも_助詞-副助詞__|([^_])*_助詞-終助詞__))/ \$2 /g ;

10

補足節

s/¥/(形容詞連体節|形容動詞連体節|連体節)-形式名詞¥/ ((の|ん|こと)_名詞-非自立-一般__ ([^_])*_助詞-(格助詞-([^_])*|係助詞)__) /\$2 ¥/補足節¥/ /g;

20

s/¥/(形容詞連体節|形容動詞連体節|連体節)-形式名詞¥/ ((の|ん|こと)_名詞-非自立-一般__ (([^_])*_助詞-格助詞-引用__|って_助詞-格助詞-連語__)) /\$2 ¥/補足節¥/ /g;

30

s/((([^_])*_形容詞-自立_([^_])*_基本形|([^_])*_名詞-形容動詞語幹__な_助動詞_特殊・ダ_体言接続)(の|ん)_名詞-非自立-一般__ ([^_])*_助詞-(格助詞-([^_])*|係助詞))/ \$1 ¥/補足節¥/ /g;

s/¥/(形容詞連体節|形容動詞連体節|連体節)-形式名詞¥/ ((の|ん|こと)_名詞-非自立-一般__ ([^_])*_助詞-並立助詞__)/\$2 ¥/補足節-並立¥/ /g;

40

【 0 0 8 2 】

【 表 1 5 】

s/¥/補足節¥/ ¥/補足節¥/ /¥/補足節¥/ /g;

s/¥/補足節¥/ (([^_]*)_助詞-(格助詞-([^_]*)|係助詞|並立助詞)__) /\$1 ¥
/補足節¥/ /g;

s/¥/補足節¥/ ((の|ん)_名詞-非自立-一般__ ([^_]*)_助詞-(格助詞-([^_]*
)|係助詞|並立助詞)__) /\$1 ¥/補足節¥/ /g;

s/¥/補足節¥/ (([^_]*)_助詞-連体化__) /\$1 ¥/連体補足節¥/ /g;

s/¥/連体補足節¥/ (¥/([^¥/]*¥/) /\$1 /g;

10

s/(基本形 (の|ん)_名詞-非自立-一般__ に_助詞-格助詞-一般__) ¥/補足節
¥/ (¥/譲歩節ノニ¥/) /\$1 \$3 /g;

s/((基本形| (な_助動詞_特殊・タ_体言接続)) (の|ん)_名詞-非自立-一般__
で_助詞-格助詞-一般__) ¥/補足節¥/ (¥/理由節ノデ¥/) /\$1 \$5 /g;

s/(([^_]*)_助詞-格助詞-連語__) ¥/補足節¥/ (¥/([^¥/]*¥/) /\$1 \$3 /g;

s/(([^_]*)_助詞-格助詞-連語__ ([^_]*)_助詞-([^_]*__) ¥/補足節¥/ (¥/
([^¥/]*¥/) /\$1 \$5 /g;

20

s/(基本形 (の|ん)_名詞-非自立-一般__ で_助動詞_特殊・タ_連用形 は_助
詞-係助詞__) /\$1 ¥/補足節¥/ /g;

s/¥/補足節¥/ (ない_助動詞_特殊・ナイ_基本形) /\$1 /g;

s/(基本形 (の|ん)_名詞-非自立-一般__ ([^_]*)_助詞-格助詞-連語__) ¥/
補足節¥/ /\$1 /g;

30

s/((の_名詞-非自立-一般__ を_助詞-格助詞-一般__) ¥/補足節¥/ (始め_動
詞-自立_一段_連用形 ¥/連用節その他¥/) /\$1 \$2 /g;

s/(((によって|にわたって|に関して|に際して|に従って|に対して|に当た
て|をめぐって|をもって|を通して|を通じて|について)_助詞-格助詞-連語__) ¥
/([^¥/]*¥/) (¥/テ節¥/) /\$1 \$4 /g;

s/(((によって|にわたって|に関して|に際して|に従って|に対して|に当た
て|をめぐって|をもって|を通して|を通じて|について)_助詞-格助詞-連語__ ([
^_]*)_助詞-([^_]*__) ¥/([^¥/]*¥/) (¥/(テ節|テハ節|譲歩節テモ|連体節テノ
))¥/) /\$1 \$6 /g;

40

s/(((によって|にわたって|に関して|に際して|に従って|に対して|に当た

【 0 0 8 3 】

【 表 1 6 】

て|をめぐって|をもって|を通して|を通じて)_助詞-格助詞-連語__ ¥/テ節¥/) ¥
/([^\\$]*)¥/ /\$1 /g;

連用節その他

s/¥/連体節-形式名詞¥/ (こと_名詞-非自立-一般__ なく_(形容詞-自立_形容詞・アウオ段_連用テ接続|助動詞_特殊・ナイ_連用テ接続)) /\$1 ¥/連用節その他¥/ /g;

10

s/((([^\\$]*)_助詞-([^\\$]*)__ 関係_名詞-サ変接続__ なく_(形容詞-自立_形容詞・アウオ段_連用テ接続|助動詞_特殊・ナイ_連用テ接続)) /\$1 ¥/連用節その他¥/ /g;

s/¥/(連体節|形容詞連体節|形容動詞連体節)¥/ ((一方|上|うえ|中|なか|以上|反)_名詞-非自立-副詞可能__|(あまり|半面)_名詞-一般__) /\$2 ¥/連用節その他¥/ /g;

20

s/((一方|上|うえ|中|なか|以上)_名詞-非自立-副詞可能__|(あまり|半面)_名詞-一般__) ¥/連用節その他¥/ ((で|に)_助詞-格助詞-一般__) /\$1 \$4 ¥/連用節その他¥/ /g;

s/((一方|上|うえ|中|なか|以上)_名詞-非自立-副詞可能__|(あまり|半面)_名詞-一般__) ¥/連用節その他¥/ ((([^\\$]*)_助動詞_特殊・(デス|ダ)_([^\\$]*)) /¥/連体節¥/ \$1 \$4 /g;

30

s/((一方|上|うえ|中|なか|以上)_名詞-非自立-副詞可能__|(あまり|半面)_名詞-一般__) ¥/連用節その他¥/ ((([^\\$]*)_助詞-([^\\$]*)__ /¥/連体節¥/ \$1 \$4 /g;

s/(((一方|上|うえ|中|なか|以上)_名詞-非自立-副詞可能__|(あまり|半面)_名詞-一般__) で_助詞-格助詞-一般__) ¥/連用節その他¥/ ((([^\\$]*)_助詞-([^\\$]*)__ /¥/連体節¥/ \$1 \$5 /g;

40

s/(こと_名詞-非自立-一般__ に_助詞-格助詞-一般__) ¥/補足節¥/ (より_動詞-自立_五段・ラ行_連用形) ¥/連用節¥/ /\$1 \$2 ¥/連用節その他¥/ /g;

【 0 0 8 4 】

【 表 1 7 】

s/(こと_名詞-非自立-一般__ により_助詞-格助詞-連語__) ¥/補足節¥/ /\$1 ¥/連用節その他¥/ /g;

s/(にわたり_助詞-格助詞-連語__) /\$1 ¥/連用節その他¥/ /g;

s/¥/(連体節|形容詞連体節|形容動詞連体節)¥/ (かわり_名詞-一般__) (([^_]*)(名詞-([^_]*))|(接続詞__)|(連体詞__)|(副詞-([^_]*))|(接頭詞-([^_]*))|((動詞|形容詞)-自立_([^_]*)([^_]*)|(記号-(括弧開|アルファベット))|(未知語__))) /\$2 ¥/連用節その他¥/ \$3 /g;

10

s/¥/(連体節|形容詞連体節|形容動詞連体節)¥/ (かわり_名詞-一般__ に_助詞-格助詞-一般__) (([^_]*)(名詞-([^_]*))|(接続詞__)|(連体詞__)|(副詞-([^_]*))|(接頭詞-([^_]*))|((動詞|形容詞)-自立_([^_]*)([^_]*)|(記号-(括弧開|アルファベット))|(未知語__))) /\$2 ¥/連用節その他¥/ \$3 /g;

20

s/¥/連体節-形式名詞¥/ (ん_名詞-非自立-一般__ じゃ_助詞-副助詞__) /\$1 ¥/並列節デハ¥/ /g;

タメ節

s/¥/(連体節|連体節トイウ|形容詞連体節|形容動詞連体節)¥/ ((ため|為)_名詞-非自立-副詞可能__) /\$2 ¥/タメ節¥/ /g;

30

s/¥/タメ節¥/ (の_助詞-連体化__) /\$1 ¥/連体節タメノ¥/ /g;

s/¥/タメ節¥/ (に_助詞-格助詞-一般__) /\$1 ¥/タメニ節¥/ /g;

s/¥/タメニ節¥/ (は_助詞-係助詞__) /\$1 ¥/タメニハ節¥/ /g;

s/¥/(タメ|タメニ)節¥/ (([^_]*)_助詞-([^_]*)) /\$2 ¥/\$1節-助詞¥/ /g;

s/¥/タメニ節¥/ (([^_]*)_助詞-([^_]*)) /\$1 ¥/タメニ節-助詞¥/ /g;

s/¥/((タメ|タメニ)節-助詞)¥/ (([^_]*)_助詞-([^_]*)) /\$3 ¥/\$1¥/ /g;

40

ママ節

s/¥/(連体節|形容詞連体節|形容動詞連体節)¥/ (まま_名詞-非自立-副詞可

【 0 0 8 5 】

【 表 1 8 】

能__) /\$2 ¥/ママ節¥/ /g;

s/¥/ママ節¥/ (と_助詞-格助詞-一般__) /\$1 ¥/ママ節-引用¥/ /g;

s/¥/ママ節¥/ (で_助詞-格助詞-一般__) /\$1 ¥/ママデ節¥/ /g;

s/¥/ママ節¥/ (([^_]*)_助詞-格助詞-一般__) /\$1 ¥/ママ節-補足語¥/ /g;

s/¥/(ママ|ママデ)節¥/ (([^_]*)_助詞-([^_]*__) /\$2 ¥/\$1節-助詞¥/ /g;

目的節

10

s/(基本形 に_助詞-格助詞-一般__) ¥/連用節その他¥/ ((は|も)_助詞-係助詞__) /\$1 \$2 ¥/目的節¥/ /g;

ホド節

s/(基本形 ほど_助詞-副助詞__) /\$1 ¥/ホド節¥/ /g;

s/¥/ホド節¥/ (に_助詞-格助詞-一般__) /\$1 ¥/ホドニ節¥/ /g;

20

s/¥/ホド節¥/ (の_助詞-連体化__) /\$1 ¥/連体節ホドノ¥/ /g;

s/¥/(ホド|ホドニ)節¥/ (([^_]*)_助詞-([^_]*__) /\$2 ¥/\$1節-助詞¥/ /g;

ホカ節

s/¥/(連体節|形容詞連体節|形容動詞連体節)¥/ ((ほか|他)_ (名詞-非自立-副詞可能|名詞-副詞可能)__) /\$2 ¥/ホカ節¥/ /g;

s/¥/ホカ節¥/ (に_助詞-格助詞-一般__) /\$1 ¥/ホカニ節¥/ /g;

30

s/¥/(ホカ|ホカニ)節¥/ (([^_]*)_助詞-([^_]*__) /\$2 ¥/\$1節-助詞¥/ /g;

時間節トキ

s/¥/(連体節|形容詞連体節|形容動詞連体節)¥/ ((時|とき)_名詞-非自立-副詞可能__) (([^_]*_((名詞-([^_]*__)|(接続詞__)|(連体詞__)|(副詞-([^_]*__)|(接頭詞-([^_]*__)|(形容詞-自立_([^_]*_([^_]*__)|(未知語__)|(記号-(括弧開|アルファベット)__)))))) /\$2 ¥/時間節トキ¥/ \$4 /g;

40

s/¥/(連体節|形容詞連体節|形容動詞連体節)¥/ ((時|とき)_名詞-非自立-副詞可能__ に_助詞-格助詞-一般__ は_助詞-係助詞__) /\$2 ¥/時間節トキニハ¥/

【 0 0 8 6 】

【 表 1 9 】

/g;

s/¥/(連体節|形容詞連体節|形容動詞連体節)¥/ ((時|とき)_名詞-非自立-副詞可能__ に_助詞-格助詞-一般__) /\$2 ¥/時間節トキニ¥/ /g;

s/¥/(連体節|形容詞連体節|形容動詞連体節)¥/ ((時|とき)_名詞-非自立-副詞可能__ の_助詞-連体化__) /\$2 ¥/時間節トキノ¥/ /g;

s/(¥/(連体節|形容詞連体節|形容動詞連体節)¥/ (時|とき)_名詞-非自立-副詞可能__ ([^_]*_助詞-([^_]*)) ¥/時間節トキニ¥/ (([^_]*_助詞-([^_]*)_) /\$1 \$6 /g;

10

時間節アト

s/¥/(連体節|形容詞連体節|形容動詞連体節)¥/ ((後|あと)_名詞-非自立-副詞可能__) /\$2 ¥/時間節アト¥/ /g;

20

s/¥/時間節アト¥/ (に_助詞-格助詞-一般__) /\$1 ¥/時間節アトニ¥/ /g;

s/¥/時間節アト¥/ (で_助詞-格助詞-一般__) /\$1 ¥/時間節アトデ¥/ /g;

s/¥/時間節アト¥/ (の_助詞-連体化__) /\$1 ¥/時間節アトノ¥/ /g;

s/¥/時間節(アト|アトニ|アトデ)¥/ (([^_]*_助詞-([^_]*)) /\$2 ¥/時間節\$1-助詞¥/ /g;

時間節マエ

s/¥/(連体節|連体節トイウ|形容詞連体節|形容動詞連体節)¥/ ((前|まえ)_名詞-副詞可能__) /\$2 ¥/時間節マエ¥/ /g;

30

s/¥/時間節マエ¥/ (に_助詞-格助詞-一般__) /\$1 ¥/時間節マエニ¥/ /g;

s/¥/時間節マエ¥/ (の_助詞-連体化__) /\$1 ¥/時間節マエノ¥/ /g;

s/¥/時間節(マエ|マエニ)¥/ (([^_]*_助詞-([^_]*)) /\$2 ¥/時間節\$1-助詞¥/ /g;

40

時間節

s/¥/(連体節|連体節トイウ|形容詞連体節|形容動詞連体節)¥/ ((際|さい|う

【 0 0 8 7 】

【 表 2 0 】

ち|瞬間|あいだ|間|たび)_名詞-(非自立-副詞可能|副詞可能)_ /\$2 ¥/時間節
その他¥/ /g;

s/((際|さい|うち|瞬間|たび)_名詞-(非自立-副詞可能|副詞可能)_ ¥/時間節
その他¥/ ((に_助詞-格助詞-一般_)|(は_助詞-係助詞_)|(に_助詞-格助
詞-一般_ は_助詞-係助詞_)) /\$1 \$4 ¥/時間節その他¥/ /g;

s/¥/(連体節|連体節トイウ)¥/ ((折り|おり)_名詞-一般_) /\$2 ¥/時間節そ
の他¥/ /g;

10

s/((折り|おり)_名詞-一般_) ¥/時間節その他¥/ ((に_助詞-格助詞-一般_)
)|(は_助詞-係助詞_)|(に_助詞-格助詞-一般_ は_助詞-係助詞_) /\$1 \$3 ¥/
時間節その他¥/ /g;

s/¥/時間節その他¥/ (([^_]*)_助詞-([^_]*_) /\$1 ¥/時間節その他-助詞¥
/ /g;

20

s/¥/テ節¥/ (以来_名詞-副詞可能_) /\$1 ¥/時間節イライ¥/ /g;

s/¥/時間節イライ¥/ (([^_]*)_助詞-([^_]*_) /\$1 ¥/時間節イライ¥/ /g;

s/(基本形 につれて_助詞-格助詞-連語_) ¥/テ節¥/ /\$1 ¥/時間節その他¥/
/g;

s/(基本形 に_助詞-格助詞-一般_) ¥/連用節その他¥/ ((連れ|つれ)_動詞-
自立_一段_連用形 て_助詞-接続助詞_) ¥/テ節¥/ /\$1 \$2 ¥/時間節その他¥/ /g
;

30

時間節イマ

s/¥/(連体節|形容詞連体節|形容動詞連体節)¥/ ((今|いま)_ (名詞-副詞可能
|接頭詞-名詞接続)_ (([^_]*)_((名詞-([^_]*_)|(接続詞_)|(連体詞_)|(
副詞-([^_]*_)|(接頭詞-([^_]*_)|(動詞|形容詞)-自立_([^_]*_)_([^_]*_)|(
未知語_)|(記号-(括弧開|アルファベット)_)))) /\$2 ¥/時間節イマ¥/ \$5 /g;

40

s/¥/(連体節|形容詞連体節|形容動詞連体節)¥/ ((今|いま)_ (名詞-副詞可能
|接頭詞-名詞接続)_ (は|も)_助詞-係助詞_) (([^_]*)_((名詞-([^_]*_)|(

【 0 0 8 8 】

【 表 2 1 】

接続詞__)|(連体詞__)|(副詞-([^_]*)__)|(接頭詞-([^_]*)__)|((動詞|形容詞)-
 自立_([^_]*)_([^_]*)|(未知語__)|(記号-(括弧開|アルファベット)__))) /\$2
 ¥/時間節イマ-助詞¥/ \$6 /g;

s/¥/時間節(イマ|イマ-助詞)¥/ (([^_]*)_助詞-([^_]*)__)/\$2 /g;

条件節バアイ

s/¥/(連体節|形容詞連体節|形容動詞連体節|連体節([^_]*)¥/ ((場合|ばあ
 い)_名詞-(非自立-副詞可能|副詞可能)__ (([^_]*)((名詞-([^_]*)__)|(接続
 詞__)|(連体詞__)|(副詞-([^_]*)__)|(接頭詞-([^_]*)__)|((動詞|形容詞)-自立
 ([^]*)_([^_]*)|(未知語__)|(記号-(括弧開|アルファベット)__)))) /\$3 ¥/条
 件節バアイ¥/ \$6 /g;

10

s/¥/(連体節|形容詞連体節|形容動詞連体節|連体節([^_]*)¥/ ((場合|ばあ
 い)_名詞-(非自立-副詞可能|副詞可能)__ の_助詞-連体化__)/\$3 ¥/連体節バ
 アイノ¥/ /g;

20

s/¥/(連体節|形容詞連体節|形容動詞連体節|連体節([^_]*)¥/ ((場合|ばあ
 い)_名詞-(非自立-副詞可能|副詞可能)__ ((に|で)_助詞-格助詞-一般__|で_助
 動詞_特殊・ダ_連用形)(は|も)_助詞-係助詞__)/\$3 ¥/条件節バアイ-助詞¥/ /
 g;

s/¥/(連体節|形容詞連体節|形容動詞連体節|連体節([^_]*)¥/ ((場合|ばあ
 い)_名詞-(非自立-副詞可能|副詞可能)__ ((に|で)_助詞-格助詞-一般__|(は|も
)_助詞-係助詞__))/ \$3 ¥/条件節バアイ-助詞¥/ /g;

30

s/¥/条件節(バアイ|バアイ-助詞)¥/ (([^_]*)_助詞-([^_]*)__)/\$2 ¥/条件
 節バアイ-助詞¥/ /g;

条件節カギリ

s/¥/(連体節|形容詞連体節|形容動詞連体節|連体節([^_]*)¥/ ((限り|かぎ
 り)_名詞-非自立-副詞可能__ (([^_]*)((名詞-([^_]*)__)|(接続詞__)|(連体
 詞__)|(副詞-([^_]*)__)|(接頭詞-([^_]*)__)|((動詞|形容詞)-自立_([^_]*)_([
 ^_]*)|(未知語__)|(記号-(括弧開|アルファベット)__)))) /\$3 ¥/条件節カギリ¥

40

【 0 0 8 9 】

【 表 2 2 】

/ \$5 /g;

s/¥/(連体節|形容詞連体節|形容動詞連体節|連体節([[^]_]*))¥/ ((限り|かぎり)_名詞-非自立-副詞可能__ の_助詞-連体化__) /\$3 ¥/連体節カギリノ¥/ /g;

s/¥/(連体節|形容詞連体節|形容動詞連体節|連体節([[^]_]*))¥/ ((限り|かぎり)_名詞-非自立-副詞可能__ ((に|で)_助詞-格助詞-一般__ |で_助動詞_特殊・タ_連用形) (は|も)_助詞-係助詞__) /\$3 ¥/条件節カギリ-助詞¥/ /g;

s/¥/(連体節|形容詞連体節|形容動詞連体節|連体節([[^]_]*))¥/ ((限り|かぎり)_名詞-非自立-副詞可能__ (に|で)_助詞-格助詞-一般__ |(は|も)_助詞-係助詞__) /\$3 ¥/条件節カギリ-助詞¥/ /g;

10

s/¥/条件節(カギリ|カギリ-助詞)¥/ (([[^]_]*)_助詞-([[^]_]*)) /\$2 ¥/条件節カギリ-助詞¥/ /g;

条件節モノノ

20

s/(ものの_助詞-接続助詞__ |(もの_名詞-非自立-一般__ の_助詞-連体化__) /\$1 ¥/条件節モノノ¥/ /g;

条件節トコロ

s/(た_助動詞_(無変化|特殊・タ)_基本形) ¥/連体節-形式名詞¥/ (ところ_名詞-非自立-副詞可能__) (([[^]_]*)_((名詞-([[^]_]*)) | (接続詞__) | (連体詞__) | (副詞-([[^]_]*)) | (接頭詞-([[^]_]*)) | (形容詞-([[^]_]*)_([[^]_]*)_([[^]_]*)) | (記号-(括弧開|アルファベット__) | (未知語__))) /\$1 \$3 ¥/条件節トコロ¥/ \$4 /g;

30

s/(た_助動詞_(無変化|特殊・タ)_基本形) ¥/連体節-形式名詞¥/ (ところ_名詞-非自立-副詞可能__ で_助詞-格助詞-一般__) /\$1 \$3 ¥/条件節トコロデ¥/ /g;

s/¥/条件節(トコロ|トコロデ)¥/ (([[^]_]*)_助詞-([[^]_]*)) /\$2 ¥/条件節\$1-助詞¥/ /g;

40

結果節

【 0 0 9 0 】

【 表 2 3 】

s/¥/(連体節|形容詞連体節|形容動詞連体節)¥/ (結果_名詞-副詞可能__) ((
 [^_]*)((名詞-([^_]*))|(接続詞__)|(連体詞__)|(副詞-([^_]*))|(接頭詞-
 ([^_]*))|((動詞|形容詞)-([^_]*)([^_]*)([^_]*))|(記号-(括弧開|アルフ
 アベット))|(未知語__))) /\$2 ¥/条件節ケツカ¥/ \$3 /g;

体言止

s/(([^_]*_名詞-([^_]*))(((これ|これら|それ|それら|こちら|そちら)
 _名詞-代名詞-一般__)|((この|その|こういう|こういった|こうした|そういう|
 そういった|そうした|いわゆる)_連体詞__)|((こ|そ)れら_名詞-代名詞-一般__
 の_助詞-連体化__)|(さらに|また)_副詞-(一般|助詞類接続)__|(そして|あるい
 は|それから|つまり)_接続詞__)) /\$1 ¥/体言止¥/ \$4 /g;

10

s/((こと|もの)_名詞-非自立-一般__)(([^_]*_名詞-([^_]*))_副詞-(一
 般|助詞類接続)__|接続詞__|形容詞-自立_([^_]*)([^_]*))) /\$1 ¥/体言止¥/ \$
 3 /g;

20

s/¥/体言止¥/ (¥/体言止¥/|や_名詞-接尾-一般__|なく_形容詞-自立_形容詞
 ・アウオ段_連用テ接続) /\$1 /g;

s/(((場合|時|前|まえ|後|あと|際|さい|折り|上|うえ|限り|当初|中|なか|
 うち|まま|結果|一方|ため|最近|少し|今回)_名詞-([^_]*))|(([^_]*_名詞-(
 代名詞-一般|非自立-副詞可能|副詞可能__)|(や_名詞-接尾-一般__)) ¥/体言止
 ¥/ /\$1 /g;

30

従属文

s/¥/連体節¥/ (([^_]*_接続詞|副詞-一般__)) ¥/従属文¥/ \$1 /g;

s/(基本形)(([^_]*_感動詞__)) /\$1 ¥/従属文¥/ \$2 /g;

40

s/(基本形)((それ|これ)_名詞-代名詞-一般__|(その|この)_連体詞__)) /\$1
 ¥/従属文¥/ \$2 /g;

s/((動詞-([^_]*)|助動詞)_([^_]*_基本形)(([^_]*_動詞-([^_]*)([^_]*

【 0 0 9 1 】

【 表 2 4 】

)_([^_]) /\$1 ¥/従属文¥/ \$5 /g;

s/(ませ_助動詞_特殊・マス_未然形 ん_助動詞_不変化型_基本形) ¥/連体節 ¥/ /\$1 ¥/従属文¥/ /g;

s(((([^_]*)_助動詞_([^_]*)_連用形|は_助詞-係助詞_) ない_形容詞-自立_形容詞・アウオ段_基本形) (([^_]*)_ (名詞-代名詞__|接続詞__|副詞-([^_]*)_ __)) /\$1 ¥/従属文¥/ \$5 /g;

s/(です_助動詞_特殊・デス_基本形|だ_助動詞_特殊・ダ_基本形) (([^_]*)_ _ (接続詞__|連体詞__)) /\$1 ¥/従属文¥/ \$2 /g;

10

s/(です_助動詞_特殊・デス_基本形|だ_助動詞_特殊・ダ_基本形) ¥/連体節 ¥/ (([^_]*)_ 名詞-([^_]*)_ __) /\$1 ¥/従属文¥/ \$2 /g;

s¥/従属文¥/ (当初_名詞-副詞可能__) /¥/連体節¥/ \$1 /g;

s¥/従属文¥/ ((いわば|言わば|最も|もう)_副詞-一般__) /¥/連体節¥/ \$1 /g;

20

s/(だ_助動詞_特殊・ダ_基本形) ¥/従属文¥/ (そう_名詞-特殊-助動詞語幹_ _ だ_助動詞_特殊・ダ_基本形) /\$1 \$2 /g;

s¥/従属文¥/ (¥/([^\\$]*)¥/) /\$1 /g;

主題

s(((([^_]*)_ 名詞-([^_]*)_ __) は_助詞-係助詞__) /\$1 ¥/主題ハ¥/ /g;

30

s(((([^_]*)_ 助詞-(格助詞-([^_]*)_ |副助詞) __) は_助詞-係助詞__) /\$1 ¥/ 主題ハ¥/ /g;

s(((([^_]*)_ (未知語|記号-(括弧閉|アルファベット)) __) は_助詞-係助詞_ __) /\$1 ¥/主題ハ¥/ /g;

s/(という_助詞-格助詞-連語__ の_名詞-非自立-一般__ は_助詞-係助詞_ __) /\$1 ¥/主題ハ¥/ /g;

40

s/(という_助詞-格助詞-連語__ の_名詞-非自立-一般__ は_助詞-係助詞_ __) ¥/補足節¥/ /\$1 /g;

【 0 0 9 2 】

【 表 2 5 】

s/(で_助動詞_特殊・ダ_連用形 は_助詞-係助詞__) ¥/(補足節|主題ハ)¥/ ((ない|なく)_助動詞_特殊・ナイ_(〔^_]*)) /\$1 \$3 /g;

s/(で_助詞-格助詞-一般__ は_助詞-係助詞__) ¥/(補足節|主題ハ)¥/ ((〔^_]*)_助動詞_特殊・ナイ_(〔^_]*)) /\$1 \$3 /g;

s/(ん_名詞-非自立-一般__ じゃ_助詞-副助詞__) ¥/並列節デハ¥/ ((〔^_]*)_助動詞_特殊・ナイ_(〔^_]*)) /\$1 \$2 /g;

s/¥/主題ハ¥/ (¥/(〔^¥/]*¥/)) /\$1 /g;

10

感動詞

s/^(〔〔^_]*)_感動詞__) /\$1 ¥/感動詞¥/ /g;

s/¥/ ((〔〔^_]*)_感動詞__) /¥/ \$1 ¥/感動詞¥/ /g;

s/¥/感動詞¥/ (が_助詞-接続助詞__ ¥/並列節ガ¥/) /\$1 /g;

s/(どうも_感動詞__) ¥/感動詞¥/ (ありがとう_感動詞__) /\$1 \$2 /g;

20

s/((〔〔^_]*)_感動詞__) ¥/感動詞¥/ (ごさい_助動詞_五段・ラ行特殊_連用形 (ます_助動詞_特殊・マス_基本形|まし_助動詞_特殊・マス_連用形 た_助動詞_特殊・タ_基本形))) ¥/(連体節|連体節-形式名詞)¥/ /\$1 \$3 ¥/感動詞¥/ /g;

s/((〔〔^_]*)_感動詞__) ¥/感動詞¥/ (ごさい_助動詞_五段・ラ行特殊_連用形 (ます_助動詞_特殊・マス_基本形|まし_助動詞_特殊・マス_連用形 た_助動詞_特殊・タ_基本形))) /\$1 \$3 ¥/感動詞¥/ /g;

30

s/(¥/(〔^¥/]*¥/) ¥/従属文¥/ /\$1 /g;

談話標識

s/^(〔〔^_]*)_接続詞__) /\$1 ¥/談話標識¥/ /g;

s/¥/ ((〔〔^_]*)_接続詞__) /¥/ \$1 ¥/談話標識¥/ /g;

40

s/^((また|なお|まず|さらに|更に|つまり|例えば|このように|その際|まずは|実のところ|つまるところ|特に|やはり|結局|先ほど|先程|さきほど|取り合えず|取りあえず|ただ)_ (副詞-(一般|助詞類接続)|名詞-副詞可能)__ | (もう_副

【 0 0 9 3 】

【 表 2 6 】

詞-一般__ 一つ_名詞-一般__)|(その_連体詞__ 一方_名詞-非自立-副詞可能__
で_(助動詞_特殊・ダ_連用形|助詞-格助詞-一般__))) /\$1 ¥/談話標識¥/ /g;

s/¥/ ((また|なお|まず|さらに|更に|つまり|例えば|このように|その際|ま
ずは|実のところ|つまるところ|特に|やはり|結局|先ほど|先程|さきほど|取り
合えず|取りあえず|ただ)_ (副詞-(一般|助詞類接続)|名詞-副詞可能)__ |(もう__
副詞-一般__ 一つ_名詞-一般__)|(その_連体詞__ 一方_名詞-非自立-副詞可能__
で_(助動詞_特殊・ダ_連用形|助詞-格助詞-一般__))) /¥/ \$1 ¥/談話標識¥/ /g
;

10

s/^((その_連体詞__ (際|上)_名詞-非自立-副詞可能__) |(この_連体詞__ よ
う_名詞-非自立-助動詞語幹__ に_助詞-副詞化__) |(ご存じ_名詞-一般__ の_助
詞-連体化__ よう_名詞-非自立-助動詞語幹__ に_助詞-副詞化__)) /\$1 ¥/談話
標識¥/ /g;

20

s/¥/ ((その_連体詞__ (際|上)_名詞-非自立-副詞可能__) |(この_連体詞__
よう_名詞-非自立-助動詞語幹__ に_助詞-副詞化__) |(ご存じ_名詞-一般__ の_
助詞-連体化__ よう_名詞-非自立-助動詞語幹__ に_助詞-副詞化__)) /¥/ \$1 ¥/
談話標識¥/ /g;

s/((そう_副詞-助詞類接続__ だ_助動詞_特殊・ダ_基本形 と_助詞-格助詞-
引用__) ¥/引用節¥/ (すれ_動詞-自立_サ変・スル_仮定形 ば_助詞-接続助詞__)
/\$1 \$2 ¥/談話標識¥/ /g;

30

s/^((その|この)_連体詞__ 結果_名詞-副詞可能__) (([^_]*)((名詞-([^_]
))|(接続詞__)|(連体詞__)|(副詞-([^_]))|(接頭詞-([^_]*))|(形容詞-
自立-([^_]*)([^_]*)|(未知語__)|(記号-(括弧開|アルファベット)__)))) /\$1
¥/談話標識¥/ \$3 /g;

40

s/^((と_助詞-格助詞-引用__ 言い_動詞-自立_五段・ワ行促音便_連用形 ま
ず_動詞-自立_五段・サ行_基本形 の_名詞-非自立-一般__ は_助詞-係助詞__) ¥
/補足節¥/ /\$1 ¥/談話標識¥/ /g;

【 0 0 9 4 】

【 表 2 7 】

s/^(と_助詞-格助詞-引用__ いう_動詞-自立_五段・ワ行促音便_基本形) ¥/
連体節-形式名詞¥/ (わけ_名詞-非自立-一般__ で_助動詞_特殊・ダ_連用形) ¥/
並列節デ¥/ /\$1 \$2 ¥/談話標識¥/ /g;

s/^(ご覧_名詞-一般__ の_助詞-連体化__ よう_名詞-非自立-助動詞語幹__
に_助詞-(副詞化|格助詞-一般)__) /\$1 ¥/談話標識¥/ /g;

s/¥/ (ご覧_名詞-一般__ の_助詞-連体化__ よう_名詞-非自立-助動詞語幹__
に_助詞-(副詞化|格助詞-一般)__) /¥/ \$1 ¥/談話標識¥/ /g;

10

s/¥/談話標識¥/ (([^_])*_助詞-(格助詞-([^_])*|係助詞|副助詞)__ ([^_]*
)_助詞-(格助詞-([^_])*|係助詞|副助詞|連体化)__) /\$1 ¥/談話標識¥/ /g;

s/¥/談話標識¥/ (([^_])*_助詞-(格助詞-([^_])*|係助詞|副助詞)__) /\$1 ¥
/談話標識¥/ /g;

s/¥/談話標識¥/ (([^_])*_(助詞-連体化__|助動詞_([^_])*_([^_]*))) /\$1
/g;

20

s/(¥/談話標識¥/) ¥/([^¥/]*)¥/ /\$1 /g;

引用マーカ

s/(「_記号-括弧開__) /¥/引用開始¥/ \$1 /g;

s/¥/ ¥/引用開始¥/ /¥/ /g;

s/^¥/引用開始¥/ //g;

30

s/(」_記号-括弧閉__) /\$1 ¥/引用終了¥/ /g;

s/¥/引用終了¥/ ¥/引用開始¥/ /¥/引用終了-開始¥/ /g;

s/¥/文末¥/ (」_記号-括弧閉__ ¥/引用終了¥/) /\$1 /g;

引用部の処理

40

s/(¥/([^¥/]*)¥/) ((と|って)_助詞-格助詞-(引用|一般)__) /\$3 ¥/引用節¥
/ /g;

s/(¥/([^¥/]*)¥/) (など_助詞-副助詞__ (と|って)_助詞-格助詞-(引用|一

【 0 0 9 5 】

【 表 2 8 】

一般)__) /\$3 ¥/引用節¥/ /g;

s/(¥/([[^]¥/])*¥/) ((「 | 』)_記号-括弧閉__ (と|って)_助詞-格助詞-(引用|
一般)__) /\$3 ¥/引用節¥/ /g;

s/(¥/([[^]¥/])*¥/) ((「 | 』)_記号-括弧閉__ など_助詞-副助詞__ (と|って)
_助詞-格助詞-(引用|一般)__) /\$3 ¥/引用節¥/ /g;

s/¥/引用節¥/ ¥/([[^]¥/])*¥/ /¥/引用節¥/ /g;

s/¥/([[^]¥/])*¥/ ¥/引用節¥/ /¥/引用節¥/ /g;

10

終助詞のノ

s/¥/(連体節-形式名詞|形容詞連体節)¥/ (の_名詞-非自立-一般__ (。_記
号-句点__ | ?_記号-一般__ | (¥ | ¥)_記号-括弧閉__)) /\$2 /g;

アスペクト関連

20

s/¥/テ節¥/ (([[^]_])*_動詞-非自立_([[^]_])*_([[^]_]*)) /\$1 /g;

s/¥/テ節¥/ (([[^]_])*_助動詞_(下二([[^]_])*|形容詞([[^]_])*|五段([[^]_])*|特
殊・タ|特殊・タイ|特殊・ナイ|特殊・ヌ|特殊・マス|不変化型|文語([[^]_]*))_([[^]_]*)) /\$1 /g;

30

s/¥/テ節¥/ (ない_形容詞-自立_形容詞・アウオ段_基本形) /\$1 /g;

s/¥/テ節¥/ (おり_動詞-自立_五段・ラ行_連用形) /\$1 /g;

間投句すべて削除.

s/(です_助動詞_特殊・デス_基本形 ね_助詞-終助詞__ ¥/間投句¥/) / /g;

デスネのパーコレート

40

s/(¥/[[^]_]*¥/) (です_助動詞_特殊・デス_基本形 ね_助詞-終助詞__) ¥/間
投句¥/ /\$2 \$1 /g;

s/(¥/[[^]_]*¥/) (ね_助詞-終助詞__) /\$2 \$1 /g;

【 0 0 9 6 】

【 表 2 9 】

文末

s/(¥/([[^]¥/]*)*¥/) ((¥|¥)_記号-括弧閉__ (。_記号-句点__|?_記号-一般__)) /\$3 /g;

s/(¥/([[^]¥/]*)*¥/) (。_記号-句点__|?_記号-一般__) /\$3 /g;

s/(。_記号-句点__|?_記号-一般__) /\$1 ¥/文末¥/ /g;

s/(¥/([[^]¥/]*)*¥/) ((¥|¥)_記号-括弧閉__) /\$3 \$1 /g;

s/((¥|¥)_記号-括弧閉__) (——_記号-一般__) /\$1 ¥/従属文その他¥/
\$3 /g;

s/¥/([[^]¥/]*)*¥/ (——_記号-一般__ 。_記号-句点__ ¥/文末¥/) /\$2 /g;

s/(。_記号-句点__|?_記号-一般__ (¥|¥)_記号-括弧閉__) ¥/文末¥/ (。
_記号-句点__) /\$1 \$3 /g;

s/((¥|¥)_記号-括弧閉__) ¥/([[^]¥/]*)*¥/ (。_記号-句点__) /\$1 \$4 /g;

s/(?_記号-一般__) ¥/文末¥/ (。_記号-句点__) /\$1 \$2 /g;

s/(。_記号-句点__) ¥/文末¥/ (?_記号-一般__) /\$1 \$2 /g;

s/((¥|¥)_記号-括弧閉__) ¥/([[^]¥/]*)*¥/ ((と|って)_助詞-格助詞-(引用|一般__)|(など_助詞-副助詞__ (と|って)_助詞-格助詞-(引用|一般__)) /\$1
\$4 ¥/引用節¥/ /g;

s/((¥|¥)_記号-括弧閉__) ¥/([[^]¥/]*)*¥/ ((という|といたった|っていう)
_助詞-格助詞-連語__|(など_助詞-副助詞__ といたった_助詞-格助詞-連語__)) /\$1
\$4 ¥/連体節トイウ¥/ /g;

s/((¥|¥)_記号-括弧閉__) ¥/([[^]¥/]*)*¥/ (と_助詞-格助詞-引用__ (い
う_動詞-自立_五段・ワ行(促|ウ)音便_基本形|いつ_動詞-自立_五段・ワ行促音
便_連用タ接続 た_助動詞_特殊・タ_基本形)) /\$1 \$4 ¥/連体節トイウ¥/ /g;

s/(¥/引用節¥/) ¥/引用節¥/ /\$1 /g;

s/(¥/連体節トイウ¥/) ¥/連体節トイウ¥/ /\$1 /g;

【 0 0 9 7 】

【 表 3 0 】

10

20

30

40

s/(¥/([[^]¥/]*)*¥/) ((¥ | ¥))_記号-括弧閉_) /\$3 /g;

s/((¥ | ¥))_記号-括弧閉_) (¥/([[^]¥/]*)*¥/) (¥/([[^]¥/]*)*¥/) /\$1 \$5 /g;

s/((¥ | ¥))_記号-括弧閉_) ((¥「 | ¥『)_記号-括弧開_) /\$1 ¥/連続引用
¥/ \$3 /g;

s/(¥/([[^]¥/]*)*¥/) (¥/([[^]¥/]*)*¥/) /\$3 /g;

< ルールの終了 >

【 0 0 9 8 】

なお、本実施の形態では、言語処理系 5 6 から節境界ラベルが出力されるごとに、テキスト分離部 6 0 がバッファ 5 8 から形態素列を讀出して機械翻訳部 6 2 に与え、それによって機械翻訳部 6 2 による機械翻訳がスタートする。しかし本発明はその様な実施の形態に限定されるわけではない。たとえば言語処理系 5 6 の出力を全て一旦バッファ 5 8 に記憶し、その後バッファ 5 8 の内容を節境界ラベルにより節ごとに分離して機械翻訳部 6 2 に与える様にしてもよい。

【 0 0 9 9 】

また、本実施の形態では、節境界を示す形態素列のパターンが検出されると、その末尾に節境界ラベルを挿入している。しかし本発明はその様な実施の形態には限定されず、そのパターンと所定の関係にある位置に節境界ラベルを挿入する様にしてもよい。例えば、形態素列のパターン中の末尾以外の部分に節境界ラベルを挿入すべき場合もあるかもしれない。パターンの末尾以外の場所、たとえばその一つ前に節境界ラベルを挿入する様にしてもよい。この場合、節に分離するときには節境界ラベルの次の形態素までを一つの節とすればよい。また、一箇所だけでなく 2 箇所以上に節境界ラベルを挿入する様にしてもよい。たとえば節境界に対応するパターンの先頭と末尾とに節境界の開始ラベルと終了ラベルとをそれぞれ挿入する様にしてもよい。

【 0 1 0 0 】

さらに、上記した実施の形態では、入力テキスト 3 4 の各行について最初にまとめて読み込み、節境界検出処理を行なっている。しかし本発明はその様な実施の形態に限定されるわけではない。例えば、形態素を順次一時記憶装置に F I F O 方式で記憶し、記憶された形態素列の中に所定のパターンを満足するものがあれば、そこで節境界を検出する様にしてもよい。この場合には、一時記憶装置に記憶された形態素列を当該パターンまで順次出力し、その末尾に当該パターンに対応する節境界ラベルを挿入する様にすればよい。

【 0 1 0 1 】

[第 2 の実施の形態]

上記した第 1 の実施の形態の翻訳装置 3 0 は、節境界を検出するために、予め所定のプログラム言語 (P e r l) によりプログラムされたプログラム 5 2 と、そのプログラム言語の処理系である言語処理系 5 6 とを用いている。しかし本発明はその様な実施の形態に限定される訳ではない。汎用の言語処理系を用いる代わりに、専用のプログラムを用いる事もできる。その場合、節境界ルールについては適宜追加、変更または削除が可能となる様に、ルールのみをデータベース化しておく事が考えられる。

【 0 1 0 2 】

図 1 2 に、この実施の形態に係る節境界検出装置を採用した、コーパスの統計処理装置の機能的ブロック図を示す。この装置は、処理対象のコーパスに対し、前述した節境界検出処理を行ない、その結果として得られた各節の節ラベルの種類を統計処理し、それによってコーパスの性格を調べる事を可能とするものである。

【 0 1 0 3 】

図 1 2 を参照して、このコーパスの統計処理装置 2 0 0 は、コーパス 2 0 2 を入力として、コーパス 2 0 2 に含まれる各文を節ラベル付の節に分離し、その結果を統計処理する機能を持つ。コーパスの統計処理装置 2 0 0 は、コーパス 2 0 2 を入力とし、その各文を形

10

20

30

40

50

態素解析して形態素列を出力するための形態素解析部 210 と、形態素解析部 210 の出力する形態素列に対して節境界検出処理を行ない、節境界にその直前までの節の種類を表す節ラベルを挿入してテキストとして出力する処理を行なうための節境界検出部 212 と、節境界検出部 212 から出力される節境界検出後のテキスト 214 内の節ラベルに対して統計的処理を行ない、統計出力 204 を出力するための統計処理部 216 とを含む。

【0104】

節境界検出部 212 は、節境界検出ルールをデータベース化したルールデータベース（ルールDB）232 と、形態素解析部 210 から出力される形態素列に対し、ルールDB 232 に格納されている節境界検出ルールを適用し、実施の形態 1 の置換命令と同様の処理を行なって、節境界に節ラベルを挿入したテキスト列として出力するための置換処理部 230 とを含む。

10

【0105】

置換処理部 230 としては、実施の形態 1 の Perl 処理系と同様、正規形を処理できる様な性能を持つものが好ましい。その場合、ルールDB 232 に格納されるルールの検索文字列に相当する部分を正規表現で表現する事ができるので、ルールDB 232 の容量を小さくし、かつ処理対象をもれなく適切に処理する事が可能となる。

【0106】

置換処理部 230 もコンピュータとソフトウェアとで実現できる。その場合の置換処理部 230 を実現するソフトウェアの構成は、図 7 に示したフローチャートと同様となる。

【0107】

形態素解析部 210 としては、実施の形態 1 で用いたプログラム 52 と同じものを用いる事ができる。また、統計処理部 216 で行なう統計処理は、目的に応じて適切なものを準備すればよい。たとえば、前述した節ごとの平均形態素数、平均文節数、節の種類分布等を、テキスト 214 に含まれる節ラベルに基づいて計算により求める事ができる。

20

【0108】

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および範囲内の全ての変更を含む。

【図面の簡単な説明】

30

【図 1】 独話と対話との相違を示すための図である。

【図 2】 本発明の第 1 の実施の形態に係る翻訳装置の機能的ブロック図である。

【図 3】 節境界検出ルールの一般形及び例を説明するための図である。

【図 4】 節境界検出ルールを実装した Perl のコマンド形式を説明するための図である。

【図 5】 第 1 の実施の形態の節境界検出ルールで検出可能な節の種類を説明するための図である。

【図 6】 第 1 の実施の形態の装置で節境界検出ルールを実装した Perl スクリプトの構成を示す図である。

【図 7】 第 1 の実施の形態の装置のプログラム 52 および言語処理系 56 により実現される、節境界検出処理の制御構造を説明するためのフローチャートである。

40

【図 8】 節境界検出処理の結果例を示す図である。

【図 9】 第 1 の実施の形態による節境界検出処理の性能評価に用いたコーパスの概略規模を表形式で示す図である。

【図 10】 第 1 の実施の形態による節境界検出処理の結果を、コーパス別に表形式で示す図である。

【図 11】 第 1 の実施の形態による節境界検出処理の性能評価の結果を表形式で示す図である。

【図 12】 本発明の第 2 の実施の形態に係るコーパスの統計処理装置の機能的ブロック図である。

50

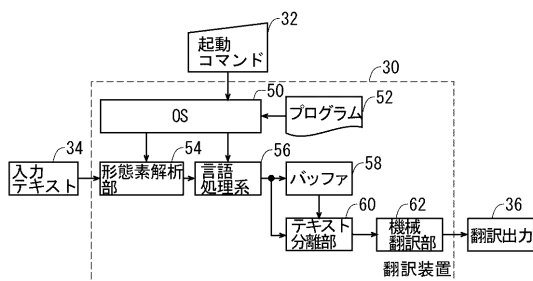
【符号の説明】

30 翻訳装置、50 オペレーティングシステム（OS）、52 プログラム、54、
 210 形態素解析部、56 言語処理系、60 テキスト分離部、62 機械翻訳部、
 200 コーパスの統計処理装置、212 節境界検出部、216 統計処理部、230
 置換処理部、232 ルールデータベース（ルールDB）

【図1】

	ニュース	旅行会話
形態素数/文	47.46	11.72
文節数/文	17.07	3.90

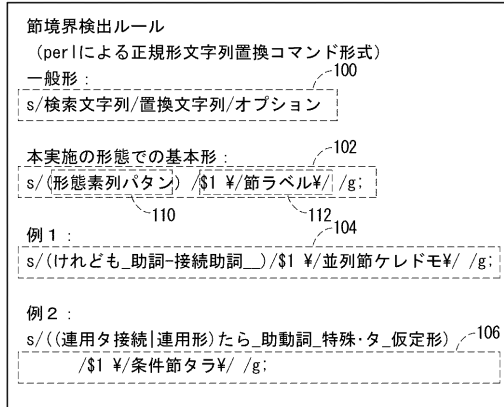
【図2】



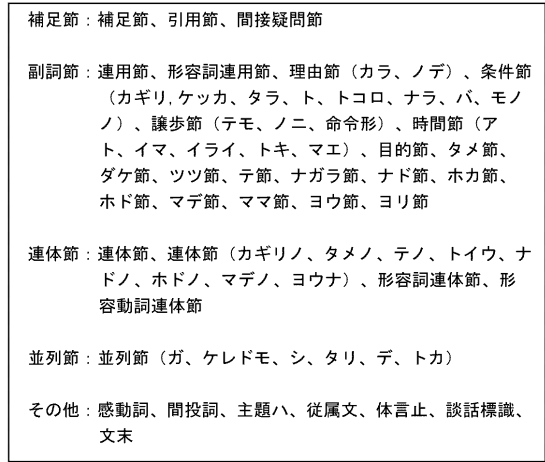
【図3】

形態素出力	80		
出現形_品詞_活用形_活用型			
入力例	82		
私は学校へ行きました			
形態素列出力例	84		
出現形	品詞	活用形	活用型
私	名詞-代名詞-一般		
は	助詞-係助詞		
学校	名詞-一般		
へ	助詞-格助詞-一般		
行き	動詞-自立	五段・カ行促音便	連用形
まし	助動詞	特殊・マス	連用形
た	助動詞	特殊・タ	基本形

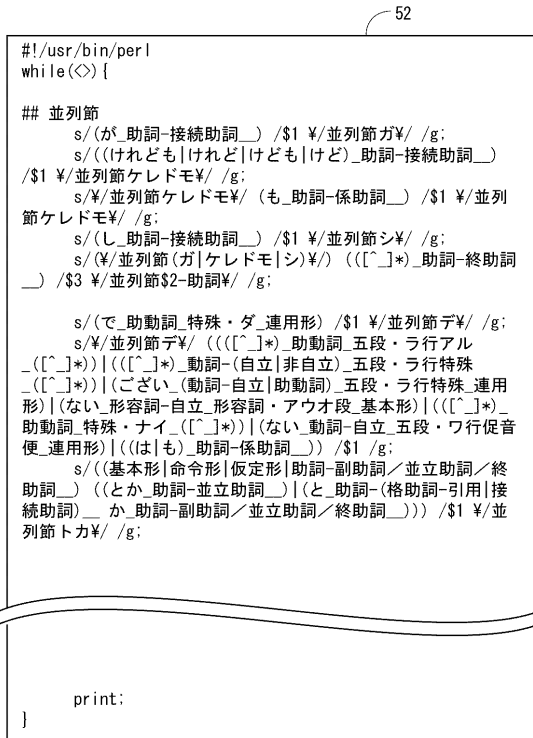
【 図 4 】



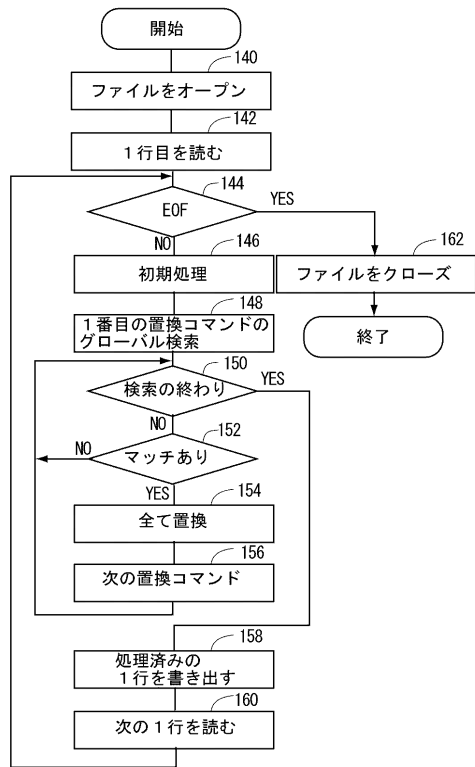
【 図 5 】



【 図 6 】



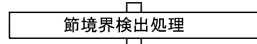
【 図 7 】



【 図 8 】

190

自主避難が呼びかけられている〇×町の▽▽地区ではほとんどの商店が休業する中営業しているコンビニエンスストアに食料などを買求める人が集中しています。〇×町役場の近くで営業を続けているコンビニエンスストアには車で十分あまりかかる避難所からも住民が訪れ食料品や生活用品などを買求めています。この店では◎□山の火山活動が活発になってからミネラルウォーターや下着類を多めに仕入れたということです。しかし道路の通行止めで仕入の車が来れなくなったため弁当や生鮮食品が品切れになったほか乾電池や生活用品なども在庫がほとんどなくなりました。



192

自主避難が呼びかけられている/連体節/〇×町の▽▽地区では/主題ハ/ほとんどの商店が休業する中/連用節その他/営業している/連体節/コンビニエンスストアに食料などを買求める人が集中しています。/文末/〇×町役場の近くで営業を続けている/連体節/コンビニエンスストアには/主題ハ/車で十分あまりかかる/連体節/避難所からも住民が訪れ/連用節/食料品や生活用品などを買求めています。/文末/この店では/主題ハ/◎□山の火山活動が活発になってから/テカラ節/ミネラルウォーターや下着類を多めに仕入れたという/連体節トイウ/ことです。/文末/しかし/談話標識/道路の通行止めで仕入の車が来れなくなったため/タメ節/弁当や生鮮食品が品切れになったほか/ホカ節/乾電池や生活用品なども在庫がほとんどなくなりました。/。文末/

【 図 9 】

	形態素数	文数	形態素数/文	文節数/文
第1の独話コーパス	577.9K	19.8K	29.14	11.24
第2の独話コーパス	75.6M	1.6M	47.46	17.07
第3の独話コーパス	499.5M	18.6M	26.83	8.92
第1の対話コーパス	255.2K	21.8K	11.72	3.90
第2の対話コーパス	1.4M	174.2k	7.87	2.71

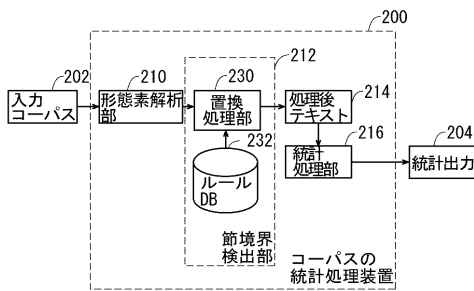
【 図 10 】

	節数	節数/文	形態素数/節	文節数/節
第1の独話コーパス	93.0K	4.69	6.21	2.40
第2の独話コーパス	10.2M	6.41	7.40	2.66
第3の独話コーパス	62.4M	3.35	8.00	2.66
第1の対話コーパス	45.2K	2.08	5.64	1.87
第2の対話コーパス	262.4K	1.51	5.23	1.80

【 図 11 】

	適合率	再現率
第1の独話コーパス	97.49%	97.07%
第2の独話コーパス	97.02%	96.71%
第3の独話コーパス	98.00%	97.24%
第1の対話コーパス	98.91%	99.69%
第2の対話コーパス	99.11%	98.01%

【 図 12 】



フロントページの続き

(72)発明者 田中 英輝

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

審査官 和田 財太

(56)参考文献 新森昭宏他, 手がかり句を用いた特許請求項の修辞構造解析, 情報処理学会研究報告2002 - NL - 149, 日本, 2002年 5月24日, Vol. 2002, No. 44, p. 65 - p. 72

古瀬蔵他, 構成素境界解析を用いた多言語話し言葉翻訳, 自然言語処理, 日本, 言語処理学会, 1999年 7月10日, Vol. 6, No. 5, p. 63 - p. 91

(58)調査した分野(Int.Cl., DB名)

G06F 17/27-17/28