

(51) Int.Cl. ⁷	識別記号	F I	テ-マコード ⁸	(参考)
G10L 13/00		G06F 3/16	340	N 5B009
G06F 3/16	340	17/21	550	L 5D045
17/21	550	G10L 3/00		E

審査請求 有 請求項の数 8 O L (全12頁)

(21)出願番号	特願2001 - 324097 (P 2001 - 324097)	(71)出願人	301022471 独立行政法人通信総合研究所 東京都小金井市貫井北町 4 - 2 - 1
(22)出願日	平成13年10月22日 (2001.10.22)	(72)発明者	村田 真樹 東京都小金井市貫井北町 4 - 2 - 1 独立 行政法人通信総合研究所内
		(72)発明者	井佐原 均 東京都小金井市貫井北町 4 - 2 - 1 独立 行政法人通信総合研究所内
		(74)代理人	100090893 弁理士 渡邊 敏
		Fターム(参考)	5B009 ME21 RD00 5D045 AB01

(54) 【発明の名称】話し言葉による音声出力装置及びソフトウェア

(57) 【要約】

【課題】 主要言語間のみならず主要言語・非主要言語間における機械翻訳に用いることができる第3言語テキストの生成技術を創出すること。同時に、従来よりも高精度にテキストを生成することのできる生成技術を提供する。

【解決手段】 複数の対訳関係を有する言語テキストを入力し、両言語の対訳コーパスを用いることで、従来の単言語入力よりも高精度な第3言語テキストを生成する技術を実現する。入力後、解析過程、変換過程、生成過程の各過程を経て、目標言語文書を出力する。目標言語文書は、固有情報を自動獲得可能なため、大規模なコーパス等を必要としないことに特徴を有する。

書き言葉データ	話し言葉データ
本論文	え
では	今日は
単語の羅列	意味
を意味でソートする	ソート
と	に
いろいろなとき	について
に	述べ
便利である	ます
と	え
いうことについて記述する。	普通のソート
	って
	いう
	の
	は
	だいたい
	え
	50
	音順
	とか
	EUC
	コード
	順

【特許請求の範囲】

【請求項 1】書き言葉によって構成されたテキストを、話し言葉で読み上げる音声出力装置であって、該音声出力装置が、
 テキストをコンピュータ処理用の書き言葉データとして装置内に入力するテキスト入力手段と、
 該書き言葉データを、話し言葉データに変換する言葉変換手段と、
 該話し言葉データから音声データを作成する音声データ作成手段と、
 10 該音声データに基づき、音声を出力する音声出力手段とを有する構成において、
 該言葉変換手段が、
 少なくともその一部において同一意味内容を含む、書き言葉学習用データ及び話し言葉学習用データを用い、各言葉学習用データを所定の文字単位に分解した上、各言葉学習用データ間の不一致部分における各言葉間の書き換え可能確率が所定の値よりも高いものをパラフレーズとし、それをあらかじめ記載したパラフレーズテーブルを備えると共に、
 20 該パラフレーズテーブルを参照して書き言葉データを話し言葉データに変換する変換エンジンを備えたことを特徴とする音声出力装置。

【請求項 2】前記パラフレーズテーブルにおいて、1 個又は複数個の文字単位から構成される文字単位群に着目し、該着目文字単位群を前記不一致部分として捉えるとき、
 30 該着目文字単位群の前後に接続した前文字単位又は後文字単位の少なくともいずれかの、当該言葉学習用データ中における出現頻度に係る値又は、該値から所定の計算式によって算出された値に基づき、
 該着目文字単位群をパラフレーズテーブルに記載するかどうか、或いはパラフレーズテーブルへの記載の態様を決定する請求項 1 に記載の音声出力装置。

【請求項 3】前記パラフレーズテーブルにおいて、1 個又は複数個の文字単位から構成される文字単位群に着目し、該着目文字単位群を前記不一致部分として捉えるとき、
 40 該着目文字単位群が、当該言葉学習用データ中における出現頻度に係る値、或いは該値から所定の計算式によって算出された値に基づき、
 該着目文字単位群をパラフレーズテーブルに記載するかどうか、或いはパラフレーズテーブルへの記載の態様を決定する請求項 1 又は 2 に記載の音声出力装置。

【請求項 4】前記言葉変換手段が、1 個又は複数個の文字単位から構成される文字単位群が前記パラフレーズテーブルに記載されているか、或いはその記載の態様を判別し、
 その結果、該文字単位群が、書き言葉から話し言葉に変換可能な 1 個以上の変換後候補群を有すると判断される

場合に、
 該文字単位群の前後に近接する少なくとも 1 個の文字単位から成る前文字単位群及び後文字単位群が直前直後に出現する場合の該文字単位群と変換後候補群の当該言葉学習用データ中における出現頻度及び / 又は出現確率に係る値を算出し、その値が最も大きい 1 個の変換後候補を特定した上で、
 10 該文字単位群と、変換後候補との、出現頻度及び / 又は出現確率に係る値が後者の方が大きいときに該文字単位群を書き言葉から話し言葉に変換する請求項 1 ないし 3 に記載の音声出力装置。

【請求項 5】書き言葉によって構成されたテキストを、話し言葉で読み上げる音声出力ソフトウェアであって、該音声出力ソフトウェアが、
 テキストをコンピュータ処理用の書き言葉データとして装置内に入力するテキスト入力部と、
 20 該書き言葉データを、話し言葉データに変換する言葉変換部と、
 該話し言葉データから音声データを作成する音声データ作成部と、
 該音声データに基づき、音声を出力する音声出力部とを有する構成において、
 該言葉変換部が、
 少なくともその一部において同一意味内容を含む、書き言葉学習用データ及び話し言葉学習用データを用い、各言葉学習用データを所定の文字単位に分解した上、各言葉学習用データ間の不一致部分における各言葉間の書き換え可能確率が所定の値よりも高いものをパラフレーズとし、それをあらかじめ記載したパラフレーズテーブルを有すると共に、
 30 該パラフレーズテーブルを参照して書き言葉データを話し言葉データに変換する変換ルーチンを有することを特徴とする音声出力ソフトウェア。

【請求項 6】前記パラフレーズテーブルにおいて、1 個又は複数個の文字単位から構成される文字単位群に着目し、該着目文字単位群を前記不一致部分として捉えるとき、
 40 該着目文字単位群の前後に接続した前文字単位又は後文字単位の少なくともいずれかの、当該言葉学習用データ中における出現頻度に係る値又は、該値から所定の計算式によって算出された値に基づき、
 該着目文字単位群をパラフレーズテーブルに記載するかどうか、或いはパラフレーズテーブルへの記載の態様を決定する請求項 5 に記載の音声出力ソフトウェア。

【請求項 7】前記パラフレーズテーブルにおいて、1 個又は複数個の文字単位から構成される文字単位群に着目し、該着目文字単位群を前記不一致部分として捉えるとき、
 50 該着目文字単位群が、当該言葉学習用データ中における出現頻度に係る値又は、該値から所定の計算式によって

算出された値に基づき、
該着目文字単位群をパラフレーズテーブルに記載するか
否か、或いはパラフレーズテーブルへの記載の態様を決定
する請求項5又は6に記載の音声出力ソフトウェア。

【請求項8】前記言葉変換手段が、
1個又は複数個の文字単位から構成される文字単位群が
前記パラフレーズテーブルに記載されているか、或いは
その記載の態様を判別し、

その結果、該文字単位群が、書き言葉から話し言葉に変
換可能な1個以上の変換後候補群を有すると判断される
場合に、

該文字単位群の前後に近接する少なくとも1個の文字単
位から成る前文字単位群及び後文字単位群が直前直後に
出現する場合の該文字単位群と変換後候補群の当該言葉
学習用データ中における出現頻度及びノ又は出現確率に
係る値を算出し、その値が最も大きい1個の変換後候補
を特定した上で、

該文字単位群と、変換後候補との、出現頻度及びノ又は
出現確率に係る値が後者の方が大きいときに該文字単位
群を書き言葉から話し言葉に変換する請求項5ないし7
に記載の音声出力ソフトウェア。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、書き言葉で構成さ
れたテキストを話し言葉で音声出力する音声出力装置及
びそのソフトウェアに関する。より詳しくは、特に変換
手法に特徴を有する出力装置及びソフトウェアの提供を
図る技術に係るものである。

【0002】

【従来の技術】近時のコンピュータによる音声合成技術
の進歩にともなって、音声によって聴覚的に情報を伝達
出力する技術は、日常生活においても様々な場面で使わ
れている。例えば、カーナビゲーションシステムにおい
ては、運転中の運転者に対して音声で進路、渋滞情報等
の各種情報を伝えるために、運転に支障を及ぼさない、
音声で伝達するようになっている。また、文字を読みと
ることができない、盲人・弱視者向けのテキスト音読シ
ステムもある。該システムでは、テキストを入力すること
によって、文字を読み込みの辞書に従って読み上げるよ
うになっており、より聞き取りやすい読み方で読み上げ
る手法等が研究されている。

【0003】ところで、現在の音声合成技術は、主に与
えられたテキストをそのまま音読するため、聞き取りや
すい話し言葉で出力するためには、あらかじめ話し言葉
で構成されるテキストを用意しているのが一般的であ
る。このような方法では、上述のテキスト音読システム
で書籍や新聞等を入力しても、そのまま読み上げるだけ
のため、書き言葉特有の表現がそのまま音読され、不
自然な印象を与えるばかりでなく、非常に聞き取りに
くい問題がある。

【0004】これらの問題点を解消するため、従来の技
術としては、例えば特開2001-166787号公報
は、書き言葉のテキストを話し言葉の表現に置き換えて
話し言葉らしい韻律で読み上げ、かつ、話し言葉韻律に
あわせて規則韻律を調整することで、話し言葉として自
然にテキストを読み上げることができる音声合成装置を
開示している。該手法によると、確かにスムーズな韻律
によって聞き取りやすい音声合成を行うことはできるの
で、上記問題点に関する一定の解決方法にはなりうる。

10 【0005】しかし、上記発明では主に読み上げる全文
体の韻律等に重きをおいており、話し言葉の表現に置き
換える手法は、あらかじめ備えられたテキスト変換テー
ブルによって機械的に置き換えるのみであって、しかも
そのテキスト変換テーブルの作成については特徴的な手
法は提案されていない。

【0006】

【発明が解決しようとする課題】本発明は上記従来技術
において課題として残されていた書き言葉から話し言葉
への効果的な変換手法を開発し、該手法を備えた音声出
力装置やソフトウェアを提供することで、より自然な話
し言葉の実現に寄与することを目的としている。

【0007】

【課題を解決するための手段】本発明は、上記の課題を
解決するために、書き言葉によって構成されたテキスト
を話し言葉で読み上げる、次のような音声出力装置を提
供する。すなわち、音楽出力装置が、テキストをコンピ
ュータ処理用の書き言葉データとして装置内に入力する
テキスト入力手段と、該書き言葉データを話し言葉デー
タに変換する言葉変換手段と、該話し言葉データから音
声データを作成する音声データ作成手段と、該音声デー
タに基づき、音声を出力する音声出力手段とを有する。

30 【0008】そして、該言葉変換手段が、少なくともそ
の一部において同一意味内容を含む、書き言葉学習用デ
ータ及び話し言葉学習用データを用い、各言葉学習用デ
ータを所定の文字単位に分解した上、各言葉学習用デー
タ間の不一致部分における各言葉間の書き換え可能確率
が所定の値よりも高いものをパラフレーズとする。該パ
ラフレーズをあらかじめ記載したパラフレーズテーブル
を備えると共に、該パラフレーズテーブルを参照して書
き言葉データを話し言葉データに変換する変換エンジンを
備える。

40 【0009】本発明の音声出力装置のパラフレーズテー
ブルにおいて、1個又は複数個の文字単位から構成され
る文字単位群に着目し、該着目文字単位群を前記不一致
部分として捉えるとき、該着目文字単位群の前後に接続
した前文字単位又は後文字単位の少なくともいずれか
の、当該言葉学習用データ中における出現頻度に係る値
又は、該値からの計算式によって算出された値に基づ
き、該着目文字単位群をパラフレーズテーブルに記載す
るか否か、或いはパラフレーズテーブルへの記載の態様
50

を決定する構成でもよい。

【0010】また、前記パラフレーズテーブルにおいて、ある着目文字単位群を前記不一致部分として捉えるとき、該着目文字単位群が、当該言葉学習用データ中における出現頻度に係る値又は、該値から所定の計算式によって算出された値に基づき、該着目文字単位群をパラフレーズテーブルに記載するか否か、或いはパラフレーズテーブルへの記載の態様を決定する構成でもよい。

【0011】1個又は複数個の文字単位から構成される文字単位群が前記パラフレーズテーブルに記載されているか、或いはその記載の態様を判別する。そして、該文字単位群が、書き言葉から話し言葉に変換可能な1個以上の変換後候補群を有すると判断される場合には次の構成としてもよい。

【0012】すなわち、該文字単位群の前後に近接する少なくとも1個の文字単位から成る前文字単位群及び後文字単位群が直前直後に出現する場合の該文字単位群と変換後候補群の当該言葉学習用データ中における出現頻度及び/又は出現確率に係る値を算出し、その値が最も大きい1個の変換後候補を特定した上で、該文字単位群と、変換後候補との、出現頻度及び/又は出現確率に係る値が後者の方が大きいときに該文字単位群を書き言葉から話し言葉に変換する。

【0013】本発明では、さらに上記と同様の構成を有する音声出力ソフトウェアを提供することもできる。該ソフトウェアは、上述の装置で実現したものと同一な手法を用いて、ソフトウェアとして実現するものである。

【0014】

【発明の実施の形態】以下、本発明の実施形態を図面に示しながら説述する。なお、本発明の実施においては、以下に限定されることなく、任意に変形、応用等を行うことが可能である。図1に本発明における音声出力装置(10)の概略図を示す。まず本装置(10)に書き言葉テキスト(11)を入力する。該書き言葉テキスト(11)は、例えば書籍や新聞等に記載された文章、文字放送やインターネット等で頒布送信される通信データ、CDROM等の媒体を介して提供される媒体データといった形態が一般的である。

【0015】上記のうち、通信データや媒体データは、コンピュータによる処理が可能な書き言葉データ(13)にそのまま用いることが可能であり、テキスト入力手段(12)としてはテキストフォーマット、例えば文字コードや改行コード、HTMLなどの特殊タグの除去等を行う機能を備えていればよい。そして、テキスト入力手段(12)によって、頒布送信されたデータを受信して必要に応じて変換する、或いは媒体から媒体データを読み出して必要に応じて変換することにより、本装置(10)で処理が可能な形態とする。

【0016】一方、上記のうち、書籍や新聞等に記載された文章は、実体的に印刷等をされたものであるため、

書き言葉データ(13)にそのまま用いることはできない。そこで、画像としてコンピュータに取り込むスキャナ、スキャンされた画像から文字認識を行うOCR装置或いはOCRソフトウェアを備え、OCR技術によって本装置(10)で処理可能な形態にする。これらOCRに係る技術は多様な方法が既知となっており、それらを随意利用することができる。

【0017】このように本発明では、テキスト入力手段(10)にスキャナやOCR装置を配設し、例えば文字を読めない物が所望の書籍をスキャナに読みとらせることで、自然な話し言葉による音声出力を得ることもできる。従来の点字翻訳等に比して、格段に低コストかつあらゆる書籍・新聞にも対応できるため、バリアフリー化にも寄与する。また、従来の同様の装置と比べても、自然な話し言葉による出力は、聞き取りやすく、疲労感を感じさせない読み聞かせが可能であり、その効果は極めて高い。

【0018】テキスト入力手段(10)によって書き言葉データ(13)を作成した後、該書き言葉データ(13)は本発明の中核をなす言葉変換手段(14)に送られる。言葉変換手段(14)では、該手段に備えられたパラフレーズテーブル(15)を参照しながら、書き言葉から話し言葉への変換が行われる。該変換についての詳細は後述する。

【0019】言葉変換手段(14)によって書き言葉データ(13)は話し言葉データ(16)に変換される。そして、音声として出力するためにその話し言葉データ(16)の読み方についての情報を付与し、電氣的に発声が行えるよう、例えば音声波形のデータが作成される。該作成を司るのが、音声データ作成手段(17)である。

【0020】音声データ(18)の作成においては、すでに多くの既知技術があるが、本発明においてもそれらの技術を用いることができる。特に、本発明では自然な話し言葉の出力を目的としており、イントネーションなど、韻律の制御を行うと好適である。すなわち、話し言葉データ(16)について、音声データ作成手段(17)が単語、形態素等の文字単位毎に読みだけでなく、それらの韻律、文字単位が組み合わさったときの韻律等の情報も含めた音声情報テーブルを備えてもよい。

【0021】音声データ作成手段(17)によって作成された音声データ(18)は、音声出力手段(19)から出力される。音声出力手段(19)は例えばデジタル・アナログ変換を行い、スピーカを介して音声出力を行う装置であり、これによって、本装置に入力された書き言葉のテキストは、音声として出力されることになる。

【0022】本発明における音声出力装置は、上記のような構成を備えるが、次に本発明の核心ともいべき、言葉変換手段につき、詳述する。本実施例においては、

10

20

30

40

50

言葉変換手段(14)で変換する際に書き言葉と話し言葉の対応関係を規定するパラフレーズテーブル(15)の作成にあたって、日本語の講演発表の音声を書き起こし、形態素情報を付与した日本語話し言葉コーパスを話し言葉データ(16)として、講演発表の元となる論文(予稿)の電子化データを書き言葉データ(13)として、それぞれ用いている。

【0023】そして、話し言葉データ(16)と書き言葉データ(13)の一致部分や不一致部分を調べ、書き言葉データ(13)から話し言葉データ(16)への変換規則を自動的に獲得している。さらに、獲得された変換規則を基に、話し言葉データ(16)から書き言葉データ(13)への自動変換を行うことを実現している。

【0024】はじめに、書き言葉データ(13)と話し言葉データ(16)について、不一致部分のとりかたについて説述する。まず、不一致部分をとる書き言葉と話し言葉のデータの形態素解析を行なって図2のように各形態素が各行にわかれた形にデータを変形する。なお、言葉データの分解に、形態素による区切りをおこなったが、本発明の実施においては必ずしも形態素による分解ではなく、辞書情報に基づく単語区切りなど、任意の分解方法によりいかなる文字単位にも分解することができる。

【0025】次に、これらのデータを照合し書き言葉と話し言葉のデータの不一致部分と一致部分の検出を行なう。この照合は、本実施例では簡単のため、UNIX(登録商標)コマンドのdiffを用いて行なっている。diffコマンドは、2つのファイル間の違いを探すコマンドであり、本発明では、書き言葉学習用データファイルと、話し言葉学習用データファイルを入力することで用いることができる。本コマンドは、空白の数の違いを無視したり、大文字小文字の違いを無視するほか、多様な比較項目の設定が可能であるため、本発明の実施において極めて好適である。もっとも、本発明の不一致部分と一致部分の検出には、該コマンドの使用に限定されることなく、専用の検出エンジンを備える構成でもよい。

【0026】さきほどのデータでこれを行なうと、図3

$$P(S2) \simeq (d + 1) * \frac{\text{文字列}S2\text{の出現数}}{\text{文字総数}}$$

【0030】このときの不一致部分が確からしい確率、すなわちP(不一致、S1, S2)とすると、P(不一致、S1, S2)はS1, S2がともに図のような形であらわれにくい確率であると仮定すると、以下のような。(S1とS2が独立であることを仮定している。)

【式3】

$$P(\text{差分}, S1, S2) \simeq (1 - P(S1))(1 - P(S2))$$

【0031】このように求められた式2に明らかなよう

に示したような結果が得られる。表のセミコロンで始まる行は一致部分、不一致部分を示すためのもので、「;

」から「;

」までの部分は、書き言葉データ(13)でのみ出現したもの、「;

」までの部分は、話し言葉データ(16)でのみ出現したもの、「;

」までの部分は、書き言葉データ(13)と

話し言葉データ(16)でともに出現したものを意味する。ここで取り出したいものは、書き言葉と話し言葉の

10 違いであるので、「;

」までの部分となり、先ほどのデータの場合、不一致部分は図4のようになる。

【0027】図4の1行目の「本論文では単語の羅列を」「え今日は」は、書き言葉では「本論文では単語の羅列を」とあったが、話し言葉では「え今日は」となったということの意味する。しかし、このように対応付けを行っただけでは、書き換え規則となるような不一致は抽出できない。そこで、上記結果からさらに、確からしい書き言葉と話し言葉の不一致部分を抽出する。そのための条件として、本発明では次の2つの着想を得て、確からしい不一致部分の抽出を行う。

20 【0028】その1つは、珍しい(出現頻度の低い)文字列に囲まれた不一致部分ほど、パラフレーズとしては確からしいという着想である。ここでは文字列は形態素と置き換えて考える。この着想における「確からしい」につき、実際にコンピュータで処理を行うため、本発明では次のような確率値の算出方法を創出した。

【0029】不一致部分が図5のように、一致部分である文字列S1, S2にはさまれていて、S1とS2の間がd文字だけ離れているとする。このとき、S1およびS2からみて、d文字以内に図の方向にS2およびS1が現れる確率を、P(S1), P(S2)とすると、P(S1), P(S2)は近似的に以下のように表される。

【式1】

$$P(S1) \simeq (d + 1) * \frac{\text{文字列}S1\text{の出現数}}{\text{文字総数}}$$

【式2】

に、上記P(S1), P(S2)等が小さい、すなわち出現頻度が低いほど、不一致部分が確からしい確率は高まり、P(不一致、S1, S2)は大きな値となる。

【0032】本発明では、上記のような不一致部分の候補の取り出しについて、さらに以下のような改良を加えている。図6のように一致部分と不一致部分が出現しているときに、「不一致部分1」「一致部分1」「不一致部分2」、「不一致部分1」「一致部分1」「不一致部分2」「一致部分2」「不一致部分3」といっ

たものも不一致部分の候補とする。このため、上記の着想における「不一致部分」については、一致部分や不一致部分を含む文字列（文字単位やその集合）と考え、本発明ではこれを一致部分・不一致部分と表記している。

【0033】この改良は、単に「不一致部分1」だけでは、「一致部分0」「一致部分1」から求まるP（不一致）の値が小さくて取り出せないようなときも、“「不一致部分1」「一致部分1」「不一致部分2」を不一致部分と考えることで、「一致部分0」「一致部分2」から求まるP（不一致）の値が大きくなって取り出しう

【0034】本実施例では、この連結によって生成する不一致部分は、元の不一致部分を3個以下しか含まないものに限った。これは書き換えの規則を決定する際に、過剰に長い文字列とすると、一般的に適応しうるパラフレーズが抽出できなくなるためであり、本実施例では3個以下とすることで、概ね好適な抽出を行うことができた。同時に、これによって組み合わせの数を抑制し、演算量を減らし高速化を図ることができる。

【0035】さらに1つの着想は、複数箇所に出現した不一致部分ほど、パラフレーズとしては確からしいというものである。これは、複数箇所での確率を組み合わせることによって本装置に導入することができる。複数箇所のうち一か所でも正しければ、その不一致部分は正しいものとして抽出できると考える。

【0036】つまり、不一致部分が正しい事象は、任意のS1, S2に対してS1, S2に囲まれる不一致部分がすべて確からしくない場合の余事象なので、不一致部分が確からしい確率をP（不一致）とすると、それは以下の式で表される。（各不一致部分が独立であることを仮定している。）

【式4】

$$P(\text{差分}) \approx 1 - \prod_{S1, S2} (1 - P(\text{差分}, S1, S2))$$

不一致部分の取り出しは、diffの結果を上記の式3の値でソートし、その確率値の大きいものから取り出すことによって行なわれる。

【0037】次に、上記の手法による書き言葉データ（13）と話し言葉データ（16）の不一致部分の抽出について、実際の例を挙げて説述する。書き言葉、話し言葉のデータとしては、82編の学術講演の部分を利用する。書き言葉データは、論文であって、手作業による打ち込みがされたもの、82編、82編、352, 660文字を含む。話し言葉データとしては、上の論文に対応するもので口頭の講演を手作業によってデータ化したものである。話し言葉データは、330, 679文字を含む。

【0038】書き言葉データとして用いる論文データには、表題や著者名、所属なども含まれているが、これらは消したりせずにそのまま残して利用した。話し言葉デ

ータの方には、図7にあげるようなタグが埋め込まれていた。このタグの対処として、次の処理を行う。すなわち、基本的に各タグのリストの第二要素をタグのかわりに本文に埋め込む。例えば、“（Fあの）”の場合、「あの」を本文の該当箇所にいれる。ただし、セミコロンで区切られているものについては一番最後のものを、カンマで区切られているものについては、一番最初のものをを用いる。ファイラーや言い直しなどは省いた方がよいとも考えられるが、本実施例ではそういった表現も不一致部分として抽出することを目的として残すことにした。

【0039】上記の条件で書き言葉データ（13）と話し言葉データ（16）の不一致部分を、前記手法により抽出することを試みた。この結果抽出総数は図8のようになった。表の確率値は式4の値を意味する。また、不一致部分を式4の値でソートした結果の上位50個を図9に示す。図9の頻度は不一致部分の出現回数を意味する。

【0040】図9における「データ」「データー」の食い違いは、コーパスの定義によるもので、書き言葉では「データ」と書くが話し言葉で「データー」と伸ばして発音しやすいということを意味しているものではない。また、話し言葉で“<C>”が得られているが、これはコーパスにおいて単語の途中を意味するタグでこれが得られてもあまり意味はない。その他目立つものとしては、「え」「えー」などのフィラーが検出できていたり、「＝」は「は」と読むということがわかったり、話し言葉では「という」をいれてやわらかくいう場合があることがわかる。

【0041】抽出された不一致結果を分析したところ主に以下のものがあつた。

1. 表記の揺れ

表記の揺れの例を表7に示す。これはコーパスの定義にも関係するところであるが、脚注7にも書いたようなことを行なうと、極力話し言葉コーパスと書き言葉コーパスで揺れが生じないようにした方が望ましいことがわかる。

【0042】實際上、この表記の揺れについてはパラフレーズテーブルに記載されても支障は少ないと考えられるが、書き言葉コーパス側の典型的な表記の揺れ、例えば「データ」「データー」や「コンピュータ」「コンピューター」のような外来語における長音の表記などを、予め統一する変換を施すか、或いはパラフレーズテーブル（15）作成時に、両表記を同一と見なすこともできる。

【0043】同一と見なす方法としては、例えば、新聞の記事を書き言葉コーパスとして用いると、「データー」との表記が33個のとき、「データ」との表記は2042個という調査例がある。この場合、明らかに「データ」が書き言葉コーパスにおいては自然な表記で

あるので、残りの「データー」についても「データ」として見なすことができる。また、予め変換してもよい。これにより、学習用データとなる書き言葉コーパスの作成時に多少の揺れがあっても、典型的な表記の揺れについては補正することが可能であって、長音の有無による出現頻度の誤解を生じさせないようにすることもできる。

【0044】また、話し言葉学習用データとなる話し言葉コーパスについては、より自然な音声出力を得るために、作為的に自然な読み方ができるように修正を施してもよい。例えば、音声データ作成手段(17)の特性によっては、話し言葉データ(16)における「データー」を特に、「デェタァ」と表記する等によって、好適な話し言葉の出力が得られることもある。

【0045】2. 表記・読みを与えるもの
書き言葉コーパス上で、例えば記号や単位など、表記と読みの異なるものが抽出される。この例を図11にあげる。この図により「=」は「は」と読めばよいとか、「S」は「秒」を意味するときと記号「S」を意味するときがあるなどがわかる。本結果は、本発明による効果が強く表出しており、従来人手によって与えていたこれらの対応関係を、自動的に獲得し、さらに、前後の文字単位やその集合との関係から確率を算出して正確な書き換えを行うことができる。

【0046】3. 同義関係のもの
略同一な意味を有する同義関係にあるパラフレーズの抽出例を図12にあげる。論文に書いていたことをちょっと違って言ったり、書き言葉特有の表記を、話し言葉で自然な言い回しにするために、同義な意味を示す書き換え表現を獲得することができる。本態様についても、本発明の効果が強く現れる例である。なお、本実施例で対象としたデータが研究を述べた予稿と講演であったので、研究がらみの同義表現、例えば「論文」「研究」などが得られているように見受けられる。

【0047】ここで、本発明で特徴的なのは、「論文」と「研究」がそれが使われる場面によっては必ずしも書き言葉と話し言葉のパラフレーズには成りにくいことである。従って、従来の手法では、このような対応付けは、特に学術講演の書き換え用に作成しない限り、パラフレーズとして搭載されにくい。しかし、本発明では、これらを自動的に獲得するため、論文とその講演とのコーパスを用いて学習することで、容易に自然な話し言葉を出力できるようになる。

【0048】4. 口語調のもの
書き言葉を丁寧な表現に改めるなど、口語においてしばしば見られる形への変換がパラフレーズテーブル(15)に抽出された。図13はこの例である。「分かった」を「分かりました」にするなど、話し言葉で丁寧語にするものから、「。」と書いているところを「訳ですが」と文をつなげるものなど、非常に自然な書き換えが

期待される結果である。これらも従来の手法ではなかなか見付けにくい規則であり、本発明の有効性が示される。また、最後の行に「これ」が得られているが、これは「明瞭に発声したもの(これ)を」という形で使われていた。話し言葉において、「これ」などで指し示す表現が多い証である。

【0049】5. 省略をしているもの

この例を図14にあげる。話し言葉の方では「処理」を省いて言ってみたり、データの値を「11.25」を「11.3」に丸めて言ってみたりして、語数を少なくすることがある。これは、書き言葉においては繰り返し正確な表記を行っても不自然でないのに対し、話し言葉でそれを行うと、不自然で冗長な印象を与える。本発明では、これを効果的に抑止し、自然な省略、言い換えを行うことが可能となる。

【0050】6. 補完をしているもの

5. 省略しているものと逆の例であり、例えば書き言葉では短縮して表記しているものをより理解しやすいように補完する例である。図15に本例を示す。書き言葉では「損失の平均」となっていたが、「損失の値の平均」と「値」をいれてわかりやすいようにいいかえている。また、値も正確に「七十五デシベル」といっている場合もある。

【0051】7. コーパスの誤り検出に関わるもの

本例を図16にあげる。もともと、書き言葉データ(13)、話し言葉データ(16)自体に誤りがあった場合その部分が不一致として得られる場合がある。1行目のデータは、「速報」を「速記」と誤ったものと思われる。この誤りは論文を手作業でデータ化したときに生じたものと思われる。また、話し言葉データの方にも誤りが見受けられる。最後の行のデータは、「死活」と「生活」どちらでも正しそうな感じもする。実際の講演者自身どちらの気持ちで言っていたのかわからない場合もあるのではないかとと思われる。

【0052】本発明による手法を用いることで、上記のように興味深いパラフレーズテーブル(15)の作成を行うことができた。これらは従来の技術によっては、抽出が困難なものも多く、また用途に分けて膨大なテーブルを手作業で作ることはコスト的にも不可能に近い。その点、本手法は、コーパスを用意することで自動的にパラフレーズテーブル(15)を形成するため、低コストで、より自然な音声出力を可能にする画期的技術である。

【0053】また、音声認識技術と融合させ、新聞記事の内容を話し言葉で話す話者の声を自動認識し、当該新聞記事との一致部分・不一致部分を抽出してパラフレーズテーブル(15)を形成すれば、話者の話し方の特徴を捉え、かつ新聞記事特有の書き言葉表現も自然な話し言葉表現に書き換え可能なパラフレーズテーブル(15)が極めて容易に作成できる。

【0054】本発明は、このように形成されたパラフレーズテーブル(15)を有する言葉変換手段(14)を備えているが、さらに変換エンジンによって、実際の変換処理を行う。次にこの方法について詳述する。上記で、多くの書き言葉と話し言葉の不一致部分のデータを収集することができたが、この不一致部分のデータは書き言葉から話し言葉への変形規則とみることでもできる。

【0055】ここで用いる変形規則は、前節で獲得した不一致部分のうち、式4の値でソートを行なったデータにおいて頻度1の事例が現れる直前のものまでとした。これは頻度1の事例はサンプルとして小さいため、信頼性が低いためである。これによって得られる規則の数は240個であった。書き言葉から話し言葉への変換エンジンが備えるアルゴリズムは以下のものとした。本アルゴリズムを図17に示す。

【0056】1. 入力として与えられる書き言葉のデータを文字単位に分解(例えば、形態素解析。以下、形態素を例に挙げる)して、文字単位列(形態素列)に分解する。

2. 文頭の形態素から順に、形態素ごとに以下の処理を行なう。

(a) 現在の形態素で始まる形態素列 S (形態素を一つも持たない場合、つまり空文字列も含む)と、240個の不一致データの書き言葉の文字列 A_i が一致した場合、その不一致部分のデータ R_i が規則として用いられ、その不一致データの話し言葉の部分(文字列 B_i)が、書き換え後表現の候補となる。また、 S の前接 k -gramの形態素列を S_{1i} 、 S の後節 k -gramの形態素列を S_{2i} とする。

(b) 各書き換え後表現の候補 B_i に対して、話し言葉コーパスでの S_{1i} 、 B_i 、 S_{2i} の文字列の頻度を求め、この頻度が最も大きかったときの i を m とする。

(c) 話し言葉コーパスでの S_{1m} 、 S_{2m} の文字列の頻度を求め、この値よりも、 S_{1m} 、 B_m 、 S_{2m} の文字列の頻度の方が大きいとき、 A_m を B_m に書き換え、処理を次の形態素に移す。

ただし、 k は定数である。

【0057】本アルゴリズムは、話し言葉コーパスでの頻度が大きくなるように書き換える、つまり、話し言葉コーパスで出てきやすい表現に書き換えることをしている。ある文章を入力テキストとして、実際に上記のアルゴリズムで書き言葉から話し言葉に変形を試みた。 $k=1$ のとき(前後の形態素の環境が1gramの場合)の結果を図18に、 $k=2$ のとき(同、2gramの場合)の結果を図19に示す。 $k=1$ では頻度を求める環境が短く精度が悪いがそれでも、「え」をいれたり「本稿」を「本研究」と言い換えたりという話し言葉らしい結果が得られている。

【0058】 $k=2$ では精度はよくほとんど誤りがなかった。「という」や「ま」や「あー」をいれていて、い

かにも話し言葉にふさわしい表現になっている。本実施例では行っていないが、上記アルゴリズムについては、さらに次のような改良も可能である。

【0059】すなわち、各文字列の頻度の部分を、その文字列を x とすると、与えられた入力データを環境にもつときの x が話し言葉コーパスに出現する事象の確率としてもよい。また、上記アルゴリズムは環境としては前後 k 形態素(文字単位)のものを固定で用いるものとなっているが、該形態素数(文字単位数)を可変したり、構文的な素性など広範な情報を用いて確率を求める構成でもよい。

【0060】本発明は、上記のような言葉変換手段(14)を有し、その結果書き換えられた話し言葉データ(16)から音声出力を行うものである。書き言葉データ(13)を出力するだけでは上記で得られたパラフレーズテーブル(15)を効果的に用いることはできず、音声として出力することで初めて効果的に上記手法を利用することができる。

【0061】例えば、新聞の音読を行う装置として利用すると、新聞特有の言い回しであっても自然な話し言葉に変換出来る上、話者の特徴をそれに反映することもできる。すなわち、話し方には人それぞれに特徴があり、用いる用語や言い換えの好みによって、その人なりの個性がでる。本発明では、話し言葉と書き言葉の一致部分・不一致部分を自動的に獲得することができるため、そのような個性をあえて分析しなくとも、必然的に話者の個性が反映されるのである。このような書き換えが可能な装置は従来にはないものであり、本発明が最も効果を奏する点でもある。

【0062】本発明の実施形態としては、ソフトウェアによる提供も可能である。上記装置における各手段はいずれも一般に流通するパーソナルコンピュータ等によってハードウェア的には実現が可能のものであり、本発明をソフトウェアで提供し、それをインストールすることで、上記装置と同様の効果を有することができる。

【0063】

【発明の効果】本発明は、以上の構成を備えるので、次の効果を奏する。すなわち、本発明に係る音声出力装置によれば、書き言葉学習用データと話し言葉学習用データから、各言葉間の書き換え可能確率に基づいて作成されたパラフレーズテーブルを備えるので、自然な話し言葉による音声出力が可能となる。特にパラフレーズテーブルを自動的に獲得することもできるため、用途に合わせた、又は話者の特徴を反映させたパラフレーズテーブルが容易に作成でき、従来の変換では成し得なかった興味のある書き換えも可能になった。

【0064】このパラフレーズテーブルの作成に当たっては、着目する文字単位群をパラフレーズテーブルに記載するか否かを書き言葉学習データ中の出現頻度やそれから算出した値によって決定することもできるので、コ

ンピュータ処理に適した方法で作成することができ、好適である。

【0065】さらに本発明はソフトウェアの形態としての提供も可能であり、市販のパーソナルコンピュータ等に装備することで容易に本発明の効果を享受しうる。これにより、低コストで高性能な話し言葉による音声出力が可能となり、同時に書き言葉の入力方法も選択肢が広がり有効である。

【図面の簡単な説明】

【図1】本発明による音声出力装置の概念図である。

【図2】書き言葉データと話し言葉データの形態素への分割を示す表である。

【図3】書き言葉データと話し言葉データのdiffコマンドによる結果を示す表である。

【図4】不一致部分の抽出結果を示す表である。

【図5】不一致部分の出現モード図である。

【図6】不一致部分の拡張を説明する説明図である。

【図7】話し言葉データに使用されているタグを示す表である。

【図8】不一致部分の抽出数を示す表である。

【図9】書き言葉データと話し言葉データの照合結果の例を示す表である。

【図10】表記の揺れの例を示す表である。

【図11】表記・読みを与えるものの例を示す表である。

る。

【図12】同義関係のもの例を示す表である。

【図13】口語調のもの例を示す表である。

【図14】省略をしているもの例を示す表である。

【図15】補完をしているもの例を示す表である。

【図16】誤り検出の例を示す表である。

【図17】変換エンジンのアルゴリズムの説明図である。

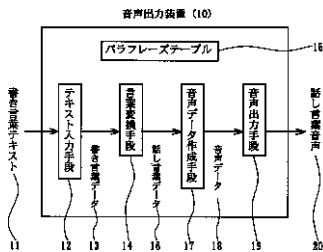
【図18】書き言葉データから話し言葉データへの変形例(1gramの場合)を示す表である。

【図19】書き言葉データから話し言葉データへの変形例(2gramの場合)を示す表である。

【符号の説明】

- 10 音声出力装置
- 11 書き言葉テキスト
- 12 テキスト入力手段
- 13 書き言葉データ
- 14 言葉変換手段
- 15 パラフレーズテーブル
- 16 話し言葉データ
- 17 音声データ作成手段
- 18 音声データ
- 19 話し言葉音声

【図1】



【図2】

書き言葉データ	話し言葉データ
本論文では単語の羅列を意味でソートするといろいろなときに便利であるということについて記述する。	え今日はソートについて述べますえ普通のソートっていうのはだいたい50音順とかEUCコード順

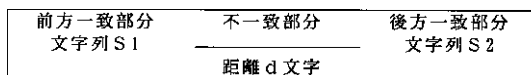
【図3】

; ▲▲▲▲▲▲▲ ; 本論文 ; ▼ ; ▲▲▲▲▲▲▲ ; 意味 ; ▼ ; ▲▲▲▲▲▲▲ ; ソート ; ▼ ; ▲▲▲▲▲▲▲ ; について ; ▼ ; ▲▲▲▲▲▲▲ ; 述べ ; ▼ ; ▲▲▲▲▲▲▲ ; ます ; ▼ ; ●●●●●●● ; え ; ▼ ; ▲▲▲▲▲▲▲ ; 普通の ; ▼ ; ▲▲▲▲▲▲▲ ; ソート ; ▼ ; ▲▲▲▲▲▲▲ ; っていう ; ▼ ; ▲▲▲▲▲▲▲ ; は ; ▼ ; ●●●●●●● ; だいたい ; ▼ ; ●●●●●●● ; え ; ▼ ; ▲▲▲▲▲▲▲ ; 50 ; ▼ ; ●●●●●●● ; 音順 ; ▼ ; ▲▲▲▲▲▲▲ ; とか ; ▼ ; ●●●●●●● ; EUC ; ▼ ; ●●●●●●● ; コード ; ▼ ; ●●●●●●● ; 順	; ▼▼▼▼▼▼▼ ; ソート ; ▼ ; ▲▲▲▲▲▲▲ ; 意味 ; ▼ ; ▲▲▲▲▲▲▲ ; ソート ; ▼ ; ▲▲▲▲▲▲▲ ; について ; ▼ ; ▲▲▲▲▲▲▲ ; 述べ ; ▼ ; ●●●●●●● ; え ; ▼ ; ▲▲▲▲▲▲▲ ; 普通の ; ▼ ; ▲▲▲▲▲▲▲ ; ソート ; ▼ ; ▲▲▲▲▲▲▲ ; っていう ; ▼ ; ▲▲▲▲▲▲▲ ; は ; ▼ ; ●●●●●●● ; だいたい ; ▼ ; ●●●●●●● ; え ; ▼ ; ▲▲▲▲▲▲▲ ; 50 ; ▼ ; ●●●●●●● ; 音順 ; ▼ ; ▲▲▲▲▲▲▲ ; とか ; ▼ ; ●●●●●●● ; EUC ; ▼ ; ●●●●●●● ; コード ; ▼ ; ●●●●●●● ; 順
---	---

【図4】

書き言葉データ	話し言葉データ
本論文では単語の羅列をするといろいろなときに便利であるということに	え今日は

【図5】



【図6】

【図10】

一致部分0 不一致部分1 一致部分1 不一致部分2 一致部分2 不一致部分3 一致部分3

書き言葉	話し言葉
データ	データー
クラスタ	クラスター
パラメータ	パラメーター
モダリティ	モダリティー
データベース	データーベース
係る	掛かる
超える	越える
すべて	全て
ため	為
行う	行なう
言い換え	言い換え

【図7】

タグ	説明	使用例
(F)	フィラー・感情表出系感動詞	(F あの)
(D)	言い直し	(D こ) これ
(W)	いい間違い	(W ミダリ; ヒダリ)
(?)	聞き取りなどに自身がない	(? タオングー)
		(? あのー、あんのー)
(M)	音や言葉の引用	(M わ)は(M は)と表記
(O)	外国語や古語, 方言など	(O ザッツファイン)
(R)	個人名など	(R 小林)さんが
(A)	基本型で漢字仮名以外の文字を使用する場合	(A イーユー; EU)
(K)	何らかの原因で漢字表記できなくなった場合	(K たち(Fんー)ばな; 權)
(S)	未登録の口語表現が出現した場合	(S こりゃ)
(笑), (泣), (咳), (あくび)	非言語音との共起	(笑 なにそれ)
(L)	ささやき声や独り言などの小さな声	(L アレコレナンダッケ)

【図8】

【図11】

	抽出数
確率 99.99%以上	1,011
確率 99.9%以上	3,245
確率 99%以上	7,846
確率 90%以上	14,777
総数	72,835
頻度 2以上	421

書き言葉	話し言葉
-	は
2	二
零	ゼロ
-gram	グラム
s	秒
s	S
Hebb	ヘブ

【図12】

【図13】

書き言葉	話し言葉	書き言葉	話し言葉
および	と	した。	という
や	とか	、	いたしました
論文	研究	、	ですね
・異なり	異なりで	られる。	です
各	それぞれ	や	られます
i 番目のターム	ターム I	や	っていう
述語	動詞	いる。	とか
識別	認識	分かった。	こう
異なれば、	違えば	ない。	非常にこう
		。	います
		れた。	分かりました
		であり、	ません
		ことである。	訳ですが
		ある。	っていうの
		なくても、	れるんですが
		ないと	であって
			訳ですけども
			ありますけれども
			なくとも、
			ないと
			これ

【図14】

【図15】

前方一致部分	書き言葉	話し言葉	後方一致部分	前方一致部分	書き言葉	話し言葉	後方一致部分
スムージング	処理		を	に対する損失の		値の	平均として
各	C (V) _{k}		素片	会話に		関しましては全然	不便はない
スポーツニュース	における	の	会話部分を	音圧レベル	70 dB	七十五デシベル	で提示
平均時間が	11.25	11.3	分まで				

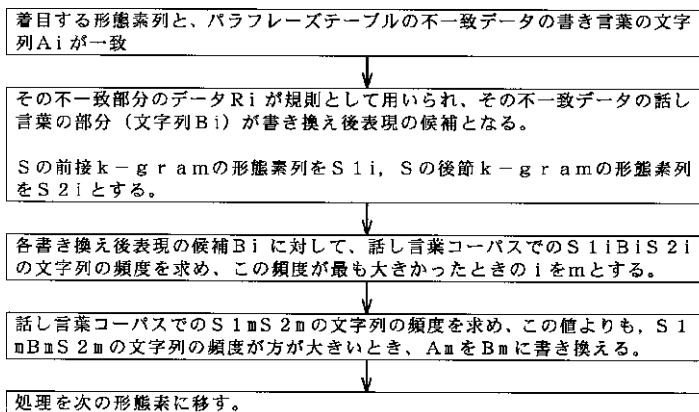
【 図 9 】

頻度	前方一致部分の例	書き言葉	話し言葉	後方一致部分の例
182	I PALの形容詞	、		形容動詞の
72	を母国語		の	キーワードで効率よく検索
43	ばたばた	、		ばたばた
49	の拡張である		え	識別的特徴抽出法
56	L R表への複数の		えー	接続制約の組み込みによる一般化LR法の
54	引き上げでNTT株は百万円になる	」		と予言したとか
39	に属するベクトル	の		和の距離に
28	本文間のハイパーリンク		を	自動生成
19	修飾要素	を		取り込む
21	ロストファイル群から導出した絞り込み語は	「		米映画
22	名詞の繰り返しの場合	、	えー	先頭の名詞のみを余剰後の挿入
20	Yは)		Xと事情が違う
21	経験的損失は全学習	データ	データー	に対する損失の
11	および用語の	ガ	ガ	格情報を付与したコーパスの作成
19	計算機科学の学会発表論文		(C)	25534表題対からなる
13	大量のコーパスを用いて	、	え	機械翻訳により
15	窓関数		に	より切り出された
12	原言語と目標言語	と		の間の同時進行性が
10	叙述を示す言葉を補って	1	一	つの文を
10	雑音スペクトル	の	を	減算
14	その結果を人手		で	修正していくプロジェクトについて
20	また頻度も少なくない		という	ことから
10	番組	を	の	最初から最後まで
10	合計強度がN	-{		ε
11	記事の		その	構造と特徴
16	ニュース番組自動字幕化の	ため	為	の音声認識システムを
10	割近い	クラスター	クラスター	が意味
6	ベクトルを異なる	k	K	個の集合にクラスタリングした
8	VQコードブックの	2	二	種類の話者モデル
7	適合率	=	は	翻訳結果が
8	その文中で		ん	話題となっている要素
9	十分良い精度で		は	推定できないという問題が
6	連結学習と		お	混合数を倍増する操作を繰り返す
11	ピッチの上昇		というもの	が有音声
7	基礎的		な	検討を
5	生起制限	の	が	厳しい形態素も
5	狭帯域	/		広帯域CELP
7	深さDの		おー	上位概念に抽象化
7	をコンテキストに持つ		ような	言語モデルに
14	手法		というの	を提案
4	かぎ括弧部分が	係る	掛かる	語が叙述を示す言葉ではないが
7	状態遷移をセットにして遺伝子操作を	行う	行なう	ことを提案し
4	ε	}		以下のリンクを検出
4	時間的に逆向きの遷移	は	が	起こらないという
5	科学論文における要旨	-		本文間のはハイパーリンク
4	最も結果	が	の	良かった混合数
4	良い意味合い	」		悪い意味合い
6	概念が	「	あー	支配する概念数
4	が持つて	いる	る	常識を
6	性能		が	改善

【 図 16 】

前方一致部分	書き言葉	話し言葉	後方一致部分
書き言葉データでの誤り			
ニュース	速記	速報	記事
日本語が	述べ	延べ	178091語
日本語複合	名刺	名詞構造	解析法
話し言葉データでの誤り			
かぎ括弧の種類を	表層的な	表装的な	情報のみで
どちらかのデータが誤り、もしくは不明			
マイクロホン 社会の	及び生体 死活	および声帯 生活	アンプ にかかわる問題

【図17】



【図18】

近年、パラフレーズに関する知識獲得の研究が重要視されつつある。本稿では、同義のテキストを照合し、その照合結果を用いてパラフレーズに関する知識を自動獲得することを試みた。この自動獲得の実験を辞書定義文、新聞記事タイトル・本文対、講演テキストにおいて行なったところ、同義のテキストの照合による方法がパラフレーズ獲得にある程度役に立つことがわかった。(中略)。そのシステムでは、基本的には、与えられた質問文の答えが書いてありそのような文を探し出し、その答えが書いてありそのような文と質問文の類似度が大きくなるよう双方を書き換えて照合し、答えが書いてありそのような文での、質問文の疑問詞に対応している箇所を答えとして出力するシステムである。

【図19】

これらは、同義な言い換えを示す、パラフレーズに関する知識がきちんとした形で整備されていないことによる。複数の国語辞典を用意してその定義を利用することが考えられる。ある程度よさそうな同義・類義表現を抽出することを試みる。対応づけの誤りである場合もあり、同義表現としてはふさわしくない対が多い。このパターンを頻度でソートした結果を表に示す。