(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第3727973号
(P3727973)
(45)発行日　平成17年12月21日(2005.12.21)　(24)登録日　平成17年10月7日(2005.10.7)

(51)Int.Cl.[7]　F I
G06F 17/22　G06F 17/22　514U

請求項の数 4　(全 19 頁)

(21)出願番号　特願平7-76581
(22)出願日　平成7年3月31日(1995.3.31)
(65)公開番号　特開平8-272786
(43)公開日　平成8年10月18日(1996.10.18)
審査請求日　平成14年4月1日(2002.4.1)

(73)特許権者　000001007
キヤノン株式会社
東京都大田区下丸子３丁目３０番２号
(74)代理人　100076428
弁理士　大塚　康徳
(74)代理人　100093908
弁理士　松本　研一
(72)発明者　福田　浩紀
東京都大田区下丸子３丁目３０番２号　キヤノン株式会社内

審査官　水野　恵雄

最終頁に続く

(54)【発明の名称】文書処理方法とその装置

(57)【特許請求の範囲】

【請求項１】

語の読みと表記と品詞とを含み、読みと表記が同一で品詞の異なる単語を別の単語として記憶した単語辞書から、各文節の語の読みに対応する表記語を検索する検索工程と、

語の組み合わせに関する用例を含む用例辞書を参照して、前記検索工程により得られた各文節に対応する表記語の中から、用例に該当する表記語を優先し、用例に該当する表記語がなければ、前記単語辞書で優先度の高い表記語を優先して、第１の表記語を決定する第１表記決定工程と、

ある文節の第１の表記語を同音の第２の表記語に変更する操作に応じて、当該文節の第１の表記語と他の文節の第１の表記語との組み合わせに対応する前記用例辞書中の用例に対して、前記第１表記決定工程における当該用例の適用を抑制するように設定する学習工程と、

前記用例辞書を参照して、前記第１表記決定工程で決定された各文節の第１の表記語に対して、前記用例辞書に記憶されている用例に該当する同音同表記の別の表記語があれば、当該別の表記語を新たな第１の表記語とするように修正する第１表記修正工程と、

前記第１表記修正工程における処理後の、ある文節の第１の表記語を同音の第２の表記語に変更する操作に応じて、当該文節の第２の表記語と他の文節の第１の表記語との組み合わせに対応する用例が前記用例辞書に存在し、かつ当該用例の適用を抑制するように設定されていれば、当該設定を解除する復活工程とを備えることを特徴とする文書処理方法。

【請求項２】
前記単語辞書における表記語の優先度の向上は、当該単語辞書での所定の検索順が先となる所にその表記語を再配置することによって行なわれることを特徴とする請求項１に記載の文書処理方法。
【請求項３】
語の読みと表記と品詞とを含み、読みと表記が同一で品詞の異なる単語を別の単語として記憶した単語辞書と、
語の組み合わせに関する用例を含む用例辞書と、
前記単語辞書から、各文節の語の読みに対応する表記語を検索する検索手段と、
語の組み合わせに関する用例を含む用例辞書を参照して、前記検索手段により得られた各文節に対応する表記語の中から、用例に該当する表記語を優先し、用例に該当する表記語がなければ、前記単語辞書で優先度の高い表記語を優先して、第１の表記語を決定する第１表記決定手段と、
前記用例辞書を参照して、前記第１表記決定手段により決定された各文節の第１の表記語に対して、前記用例辞書に記憶されている用例に該当する同音同表記の別の表記語があれば、当該別の表記語を新たな第１の表記語とするように修正する第１表記修正手段と、
ある文節の第１の表記語を同音の第２の表記語に変更する操作に応じて、当該文節の第１の表記語と他の文節の第１の表記語との組み合わせに対応する前記用例辞書中の用例に対して、前記第１表記決定手段による当該用例の適用を抑制するように設定する学習手段と、
前記第１表記修正手段による処理後の、ある文節の第１の表記語を同音の第２の表記語に変更する操作に応じて、当該文節の第２の表記語と他の文節の第１の表記語との組み合わせに対応する用例が前記用例辞書に存在し、かつ当該用例の適用を抑制するように設定されていれば、当該設定を解除する復活手段とを備えることを特徴とする文書処理装置。
【請求項４】
前記学習手段は、前記単語辞書での所定の検索順が先となる所に表記語を再配置することによって、当該単語辞書におけるその表記語の優先度を向上させることを特徴とする請求項３に記載の文書処理装置。
【発明の詳細な説明】
【０００１】
【産業上の利用分野】
本発明は、文字の変換を行う文書処理方法とその装置、特に、文字の変換における同音語の判定方法に関する。
【０００２】
【従来の技術】
従来、文字列の変換を行う文字処理方式として、仮名漢字変換方式が広く普及している。仮名漢字変換の結果、多数現れる同音語候補に対して、個々の単語の品詞の持つ文法的特性などから同音語の多義解消が行われてきたが、それだけでは対処できない場合が多かった。
【０００３】
そこで、意味的に共起関係がある単語と単語の組（すなわち、用例）を格納した用例辞書にある組み合わせを第一候補決定に利用することで、同音語の多義解消が行われてきた。更に、用例が存在する文節間において文節候補が変更された場合、文節間の状況を考慮し、適用されている用例の適用を禁止する処理（用例抑制）、または、一度適用が禁止されている用例を再び適用可能とする処理（用例復活）といった用例抑制学習が行われてきた。この用例抑制学習における用例復活処理では、ある文節（同音語）において同音語候補の変更が行われたとき、変更された同音語候補が用例が存在する１組の同音語候補と一致した場合、その用例は復活されるのが一般的である。例えば、用例『本は／厚い』が抑制されているとする。そして、文字列『本は熱い』が文書処理の出力として表示されており、ユーザが語『熱い』を『厚い』に変更したとする。この場合、文字列は『本は厚い』と

なり、２つの同音語『本は』、『厚い』は、用例『本は／厚い』と一致するため、用例『本／厚い』は復活が行われる。

【０００４】

ここで、また、『本は厚い』を『本は熱い』と変更したとする。この時、用例『本／厚い』を抑制する。また、単語『熱い』が単語学習、即ち、単語『熱い』の出現優先度を上げる。

次に、新たに『ほんはあつい』という文字列が入力されたとする。この状態では、用例『本／厚い』は抑制されており、かつ、『熱い』が単語学習されて優先度が高い状態であるため、変換結果として『本は熱い』が変換候補として選択される。

【０００５】

さてここで、『本は熱い』の『本は』に注目すると、名詞の『本』ではなく、接頭語等（名詞ではない）の『本』が抽出される可能性がある。これは、例えば、たまたま接頭語の『本』が単語辞書で最も高い優先度を持っていた場合、『ほんは』に対し接頭語の『本』が選択される可能性がある。

【０００６】

【発明が解決しようとする課題】

この状態で、選択された変換候補『本は熱い』を『本は厚い』と再び変更した場合、変更前の『本は熱い』における『本は』は名詞の『本』ではなく、用例の『本は（名詞）／厚い』の『本（名詞）』と異なるため、用例の復活処理は行われない。つまり、用例が存在する１組の文節間において、第一候補で決定された語の単語情報が用例情報の単語情報と比較して表示が同じでも単語ＩＤ（識別子）が異なる場合、つまり、表記は同じでも別の単語である場合は、その用例は復活することができないという問題を抱えていた。

【０００７】

本発明は、上記従来例に鑑みてなされたもので、読みと表記が同一で品詞の異なる単語を別の単語として記憶した単語辞書を用いたかな漢字変換において、一旦抑制した用例に対応する表記語に変更する操作がなされた場合に、同音同表記で用例に無い別の単語が存在してもそれに影響されずに、その抑制した用例を確実に復活させることができる文書処理方法及びその装置を提供する。

【０００９】

【課題を解決するための手段】

上記目的を達成するため、本発明の文書処理方法は、語の読みと表記と品詞とを含み、読みと表記が同一で品詞の異なる単語を別の単語として記憶した単語辞書から、各文節の語の読みに対応する表記語を検索する検索工程と、語の組み合わせに関する用例を含む用例辞書を参照して、前記検索工程により得られた各文節に対応する表記語の中から、用例に該当する表記語を優先し、用例に該当する表記語がなければ、前記単語辞書で優先度の高い表記語を優先して、第１の表記語を決定する第１表記決定工程と、ある文節の第１の表記語を同音の第２の表記語に変更する操作に応じて、当該文節の第１の表記語と他の文節の第１の表記語との組み合わせに対応する前記用例辞書中の用例に対して、前記第１表記決定工程における当該用例の適用を抑制するように設定する学習工程と、前記用例辞書を参照して、前記第１表記決定工程で決定された各文節の第１の表記語に対して、前記用例辞書に記憶されている用例に該当する同音同表記の別の表記語があれば、当該別の表記語を新たな第１の表記語とするように修正する第１表記修正工程と、前記第１表記修正工程における処理後の、ある文節の第１の表記語を同音の第２の表記語に変更する操作に応じて、当該文節の第２の表記語と他の文節の第１の表記語との組み合わせに対応する用例が前記用例辞書に存在し、かつ当該用例の適用を抑制するように設定されていれば、当該設定を解除する復活工程とを備えることを特徴とする。ここで、前記単語辞書における表記語の優先度の向上は、当該単語辞書での所定の検索順が先となる所にその表記語を再配置することによって行なわれる。

【００１１】

又、本発明の文書処理装置は、語の読みと表記と品詞とを含み、読みと表記が同一で品

詞の異なる単語を別の単語として記憶した単語辞書と、語の組み合わせに関する用例を含む用例辞書と、前記単語辞書から、各文節の語の読みに対応する表記語を検索する検索手段と、語の組み合わせに関する用例を含む用例辞書を参照して、前記検索手段により得られた各文節に対応する表記語の中から、用例に該当する表記語を優先し、用例に該当する表記語がなければ、前記単語辞書で優先度の高い表記語を優先して、第１の表記語を決定する第１表記決定手段と、前記用例辞書を参照して、前記第１表記決定手段により決定された各文節の第１の表記語に対して、前記用例辞書に記憶されている用例に該当する同音同表記の別の表記語があれば、当該別の表記語を新たな第１の表記語とするように修正する第１表記修正手段と、ある文節の第１の表記語を同音の第２の表記語に変更する操作に応じて、当該文節の第１の表記語と他の文節の第１の表記語との組み合わせに対応する前記用例辞書中の用例に対して、前記第１表記決定手段による当該用例の適用を抑制するように設定する学習手段と、前記第１表記修正手段による処理後の、ある文節の第１の表記語を同音の第２の表記語に変更する操作に応じて、当該文節の第２の表記語と他の文節の第１の表記語との組み合わせに対応する用例が前記用例辞書に存在し、かつ当該用例の適用を抑制するように設定されていれば、当該設定を解除する復活手段とを備えることを特徴とする。ここで、前記学習手段は、前記単語辞書での所定の検索順が先となる所に表記語を再配置することによって、当該単語辞書におけるその表記語の優先度を向上させる。

【００１６】

【実施例】

まず、本発明に係る一実施例の文書処理方法とその装置の処理上の一つのポイントを説明する。

かな漢字変換処理での文書処理方法のポイントは、まず、用例が存在する１組の語（文節間）において、少なくとも一方の語が同音語であって、その同音語に変更が行われた場合、同音語変更が行われなかった組を構成する語について単語学習を行う。

【００１７】

そして、次に新たな文字列を入力した場合、この文字列に対する第一漢字変換候補決定では、単語学習された同音語を優先的に第一候補として、それに対応する用例が存在する同音語を構成する単語と比較することで、表記は同じであるが単語ＩＤ（識別子）の異なる同音語が第一候補に選ばれることを避けることができる。

【００１８】

例えば、『本は厚い』（ここで、「本」は名詞であるとする）の『厚い』を『熱い』に変更した場合、用例『本は／厚い』が抑制された後、用例が存在する１組の語『本は／厚い』において変更が行われなかった語、つまり『本は』について単語学習が行われる（即ち、単語辞書での名詞の『本は』の優先度が上げられる）。

【００１９】

次に、新たな文字列『ほんはあつい』が入力されると、第１変換候補として『本は熱い』に変換される。この時の『本は』は、先の処理において名詞の「本」が単語学習されているので、名詞の『本は』が抽出される。よって、表記は同じであるが単語ＩＤの異なる接頭語等の『本は』が第一候補となることはなく、この後、『熱い』を『厚い』と変更すれば、用例『本は／厚い』に対して用例復活が行われる。

【００２０】

この処理を実行するために、本実施例では以下のように文書処理を実行する。まず、単語を含む文字列の読みを入力し、単語の読みと表記および各種情報を格納した単語辞書を検索することにより文節を抽出する。そして、抽出された文節の対について、共起関係を有する単語と単語の組である用例を格納した用例辞書を検索して用例候補を抽出する。次に、抽出された文節候補の中から、第一候補を決定する。そして、作成された各語について語ＩＤ（識別子）と、その語が同音語を有する場合に同音語の単語表記や、単語ＩＤを含む同音語情報を作成する。

【００２１】

次に、抽出された用例情報を元に、用例ＩＤ（識別子）、用例の対象となる語候補の単語

情報を持つ用例情報を作成する。そして、ユーザによって別の候補語が選択された時、その選択状態によって、選択用例の適用を禁止する処理（用例抑制）、または、適用を禁止された用例を適用可能とする処理（用例復活）を行う。さらに、前述の用例抑制学習が行われた後、用例を構成する１組の語において、候補の変更が行われなかった語について単語学習を行う。

【００２２】

以上の処理手順により、上述の問題を解決することができる。

以下、添付図面にしたがって、本発明に係る一実施例である文書処理装置の構成を説明する。

図１は、本発明における一実施例の文書処理装置のブロック図である。

１はＣＰＵであり、文書処理装置全体の主制御および仮名漢字変換等の文書処理を実行する。

【００２３】

２は記憶領域（ＲＡＭ）であり、同音語情報、用例情報等の各種情報の保存領域、また、各処理を行う作業領域である。

３はディスプレイであり、入力読み文字列や仮名漢字変換処理のなされた同音語候補等を表示する。

４はキーボードであり、アルファベットキー、ひらがなキー、カタカナキー等の文字記号入力キー（４－１）、および、変換を指示する変換キー（４－２）、表示された候補を選択する選択キー（４－３）等の各種のファンクションキーを備えている。

【００２４】

５はＲＯＭであり、各種処理プログラムを格納している。例えば、５－１にはＣＰＵ１が実行する仮名漢字変換の文書処理プログラムが格納されている。また、５－２には仮名漢字変換で用いられる単語辞書や、用例適用処理で使用する用例辞書、学習データを保存する学習辞書等のデータが格納されている。

図２は、かな漢字変換に用いる辞書の構成を示した図である。

【００２５】

この辞書は、表記変換に用いる単語情報と、表記変換の際の第１候補を決定するための用例情報とで構成されている。

単語情報は、その見出し単語固有の単語ＩＤと、その単語の読みを格納する読み、単語の表記を格納する表記、単語の品詞を格納する品詞等の情報から構成されている。用例情報は、その見出し単語の持つ用例情報の個数、用例固有の用例ＩＤ、用例の相手の単語に付く付属語のうち用例が成立する条件を満たす代表的な付属語を格納した結合情報等から構成される。

【００２６】

用例を持たない単語は、用例情報の個数に"０"が格納されている。単語ＩＤ、用例ＩＤは、先頭の単語または用例を"１"とする連続番号である。

図２において、例えば、『厚い』という単語は"１００"という連続番号で区別され、読みが『あつい』、表記が『厚い』、品詞が『形容詞』であるという単語情報を持つ。そして、『板が厚い』、『紙が厚い』、『本が厚い』等の３０の用例を持つ。

【００２７】

図３は、入力した文字列を本実施例の文書処理装置で解析処理したときに生成する入力文字列に対応する文節情報と用例情報のデータ構成を示した図である。ここで、図３の（ａ）は、入力文字列が各文節に分解された文節情報の全体構成である。この文節情報は、後述する仮名漢字変換処理ステップＳ５－４（図５参照）で変換候補を決定するために生成される。

【００２８】

図３の（ｂ）は、各文節情報のデータ構造を示した図である。各文節情報は、その文節を他の文節と区別するための連続番号である文節ＩＤ、その文節が同音語を持つ場合での同音語情報の使用メモリーサイズを表すデータ長、格納されている同音語の候補の個数を格

納する候補個数、各同音語候補の情報を格納する候補単語情報から構成される。
【００２９】
この生成された文節情報で、候補単語情報（図３の（ｂ））の列の先頭の候補単語情報が第１の変換候補となる。
次に、図３の（ｃ）は候補単語情報の構成を示した図である。候補単語情報は、候補の単語表記、その単語のＩＤ（識別子）、その単語と用例のペアを構成する相手語のＩＤであるペア語ＩＤ、および、その単語に関する用例情報の格納先へのポインタを格納する用例情報ポインタ等から構成される。
【００３０】
ここで、ペア語ＩＤは、同音語に対しては、それら同音語の全てに関する用例に関して格納されるので複数個存在する。
図３の（ｄ）は、前述の文節情報からポインタ（用例情報ポインタ、図３の（ｃ））が張られて参照される用例情報の全体構成を示す。各用例情報は、図３の（ｃ）に示した各ペア語ＩＤに対応して配置される。
【００３１】
図３の（ｅ）は、図３の（ｄ）の各用例情報の詳細なデータ構造を示した図である。用例情報は、用例のＩＤ、用例の文頭側の語（文節）の単語情報（例えば、単語ＩＤ）である文頭側文節候補単語情報、そのペアの単語情報である文末側（対象）文節候補単語情報を備える。
次に、図４は学習データの構成を示す図である。
【００３２】
図４の（ａ）は学習データの全体構成を示した図である。学習データは"単語学習データ"と"用例学習データ"の２つから構成される。"単語学習へのポインタ"は、"単語学習データ列"の先頭を指すポインタである。また、"用例学習へのポインタ"は、"用例学習データ列"の先頭を指すポインタである。
図４の（ｂ）は、"単語学習データ列"のデータ構造を示した図である。ここで、"単語学習データ列"は"単語学習使用サイズ"と各単語学習データから構成される。
"単語ＩＤn学習データ"（ｎ＝１，２，．．．）は、各単語のＩＤ（識別子）に対応しており、学習されている単語には"１"、即ち、その単語は抑制されることを意味する。また、学習されていない単語に対しては"０"、即ち、その単語は抑制されないことを意味する。
【００３３】
図４の（ｃ）は、"用例学習データ列"のデータ構造を示す図である。"用例学習データ列"は"用例学習使用サイズ"と各"用例ＩＤn学習データ"（ｎ＝１，２，．．．）から構成される。
"用例学習使用サイズ"には、"用例学習データ列"のサイズを格納する。"用例学習データ列"の各用例学習データは、各用例ＩＤ（識別子）に対応しており、抑制されている使用禁止用例には"１"、抑制されていない使用可能用例には"０"が格納される。
【００３４】
上述の実施例の動作を図５～図８の各動作を示すフロー、および図９の図を元に説明する。
図５は本実施例の文字処理方法の手順を示すフローチャートである。
ステップＳ５－１において、キーボード４よりキーが押下されることによる割り込みが発生するのを待つ。キー入力割り込みが発生して、キーコードを入力するとステップＳ５－２においてキー判別し、キーの種類に応じてステップＳ５－３，ステップＳ５－４，ステップＳ５－５，ステップＳ５－６，ステップＳ５－７のいずれかのステップに分岐する。
【００３５】
ステップＳ５－３は、文字列読み入力キーが押下されたときの処理である。入力したコードを記憶領域２と表示バッファ６に格納する。
ステップＳ５－４は、変換キーが押下されたときの処理であり、ステップＳ５－３で入力

された、かな漢字変換の対象となる文字列を解析して漢字に変換して、その結果を記憶領域２と表示バッファ６に格納する。漢字に変換する際には、辞書（図１の５－２）に存在する用例情報を用いて、変換語の候補を選択して、その中から第一候補を決定する。

【００３６】

ステップＳ５－５は、次候補変換キー（図１の４－２）が押下されたときの処理であり、ステップＳ５－４によって選択された別の変換語候補を表示バッファ６に格納する。

ステップＳ５－６は、選択キー（図１の４－３）が押下されたときの処理であり、ユーザからのキー入力指示により画面表示されている変換語候補から１つを選択し、選択した変換語を対応する文書中に配置する。さらに、選択された変換語を学習する処理を行う。この学習とは、その語の単語辞書での選択優先度を上げる処理を行うことを意味する。

【００３７】

ステップＳ５－７は、その他のキー（例えば、カーソル移動キーなどの文書編集で用いるキー）が押下されたときの処理であり、同種の文字処理方式において一般に行われている処理であり、公知であるので特に記述しない。

ステップＳ５－８は、表示バッファ６に格納されたデータを表示するディスプレイ３に表示する処理である。この処理では、表示バッファ６中のデータ１文字を対応文字パターンに展開して表示する処理である。

【００３８】

次に、図６は、ステップＳ５－４の処理を詳細化したフローチャートである。ステップＳ６－１は、入力された仮名漢字変換の対象となる文字列を解析し、文節を抽出する処理である。この処理は、単語辞書（図１の５－２）を検索して解析を行い、文節として認定される候補のみを出力する処理であって、同種の文字処理方式において一般的に行われている処理であり、公知であるので特に詳細には記述はしない。

【００３９】

ステップＳ６－２は、ステップＳ６－１において出力された解析結果に対して、単語辞書中に格納されている用例のパターンが存在するかどうかをチェックし、用例のパターンが存在すれば、その用例の対象となる変換語の候補を優先候補として抽出する処理である。このステップでは、図３で示したデータ構造の文節情報および用例情報を作成する。そして、ここで生成された文節情報の内、候補単語情報（図３の（ｂ））が変換候補とする。これらの情報はまた、用例抑制学習を始めとする各処理で使用される内部的なデータとなる。

【００４０】

ステップＳ６－３では、ステップＳ６－２で抽出された優先候補の中から、学習の結果なども考慮に入れて仮名漢字変換の第一候補を決定する。具体的には、生成された文節情報の内、候補単語情報（図３の（ｂ））の先頭のものを変換候補とする。

そして、第１の変換候補語を対応する文節に配置して、全体の文章とともに表示バッファ６に格納する。

【００４１】

以上、かな漢字変換処理を説明した。

次に、図７を参照して、図５のステップＳ５－６で示される選択学習処理の詳細を説明する。

ステップＳ７－１は、単語学習処理を行う。即ち、ステップＳ５－５でユーザからのキーボード指示によってステップＳ５ー４で選択した変換語を別の変換語に変更した時、変更された語に対応する単語学習データ（図４の（ｂ））に"１"を書き込む。これは、その変更された語を抑制することを意味する。ここで、"０"は、非抑制を意味する。

【００４２】

ステップＳ７－２は、用例抑制学習処理を行う。即ち、変更された語に関する用例を抑制するために、変更された語に対応する用例学習データ（図４の（ｃ））に"１"を設定する（用例の適用の禁止を意味する）。また、変更後の語に関する用例については、逆に、変更後の語に対応する用例学習データ（図４の（ｃ））には"０"を設定する（適用が禁止さ

れている用例を適用可能とすることを意味する）。
【００４３】
ステップＳ７－３では、その他の学習処理であり、例えば、文節の区切りを学習する区切り学習等の処理が含まれる。
以上、ステップＳ５ー６の選択学習処理について説明した。
次に、図９を参照して、具体例として、「ほんはあつい」をステップＳ５ー３（図５）で入力した場合での文書処理例を説明する。
【００４４】
まず、ステップＳ５－３で、文字列『ほんはあつい』が入力される（図９の（ａ）参照）と、その文字列がステップＳ５－８でディスプレイに表示され、次に、キーボードから変換キーを入力するとステップＳ５－４に進み、入力文字列に上述のかな漢字変換が適用される。そして、第一変換候補として『本は厚い』が決定される（図９の（ｂ））。このときの第一文節である『本は』は、名詞の「本」が抽出されている。
【００４５】
次に、キーボードから次候補変換キーが入力されると、ステップＳ５－５へ進み、変換候補をディスプレイに表示して、ユーザに選択させる。ユーザがキーボードを通じて例えば、『本は厚い』における『厚い』を別の変換候補『熱い』を選択して変更する（図９の（ｃ））。
次に、キーボードから選択キーが入力されると、ステップＳ５－６に進み、上述の選択学習処理が適用される。
【００４６】
まず、ステップＳ７－１で、単語学習が適用される。これは変更の行われた候補語に適用されるので、『熱い』に単語学習が行われ、単語学習データにおいて『熱い』の単語コードの該当する位置に"０"が書き込まれる。即ち、単語『熱い』
は非抑制の属性が付けられる。
【００４７】
次に、ステップＳ７－２では、用例抑制学習が行われる。これを図８を元に説明する。
図８は、用例抑制学習の処理手順を示すフローチャートである。
ステップＳ８－１は、処理の対象となるすべての変換語候補について処理を行うために、全ての語に関する処理が終了したか否かをチェックする。即ち、残っている処理する文節（語）数が０かどうかチェックし、０でなければ、ステップＳ８－２へ進む。
【００４８】
例えば、『本は熱い』では、『本は』、『熱い』の２つの文節語について処理を行う。
ステップＳ８－２では、処理の対象となる語候補に関する用例すべてについて処理を行うために、全ての用例について処理が終了したか否かをチェックする。そして、終了していなければステップＳ８ー３へ進む。これは、残っている処理する候補数が０かどうかチェックことによってなされる。
【００４９】
例えば、初めに、『本は』についての用例情報に関しては、図３の（ｅ）で示されるように用例が存在する１組の語間では、文末側の同音語にリンクして用例情報は存在している。この場合、語『本は』には用例情報は存在しないので、ステップＳ８－８に進む。
また、次の語『熱い』が処理対象の時は、ステップＳ８ー２の処理が行われると、語『熱い』への変更前の語には、用例『本は／厚い』が存在している。この場合は、ステップＳ８ー３へ進む。
【００５０】
ステップＳ８－８では、用例学習が行われたかどうかチェックし、用例学習が行われていればステップＳ８－９に進み、行われていなければステップＳ８－１に戻る。
ステップＳ８－９は、本発明に係る一実施例の一つのポイントとなる処理であり、用例学習が行われた場合、用例が存在する１組の語間において、語の変更が行われなかった語について単語学習を行う。例えば、用例『本が／厚い』においては、同音語『厚い』が『熱

い』に変更されている。この場合、語の変更が行われなかった用例のペアとなる語『本は』に関して単語学習を行う。これにより、名詞である語『本は』が単語学習される。即ち、名詞である語『本は』が、単語辞書中での選択優先度を上げる。選択優先度を上げる方法としては、例えば、単語辞書での記述順位を上げる。このことによって、検索は単語辞書の先頭から成されるのではじめに検索されるために実質上選択優先順位が上がる。

【００５１】

逆に、用例学習が行われない場合は、残り語数を"１"減らして、ステップＳ８－１に戻る。

ステップＳ８－３では、対象文節（語）について候補単語情報のチェックを行う。ここでは、用例情報の文末側（対象）文節候補単語情報における単語ＩＤ、つまり、用例『本は／厚い』の『厚い』における単語ＩＤと、その同音語『熱い』の単語情報における単語ＩＤとを比較する。この場合は、『厚い』と『熱い』の単語ＩＤは一致しない。

【００５２】

ステップＳ８－４では、文頭（相手）文節（語）について候補単語情報のチェックを行う。ここでは、用例情報の文頭側（相手）文節候補単語情報における単語ＩＤ、つまり、用例『本は／厚い』の『本は』における単語ＩＤと、その同音語『本は』の単語情報における単語ＩＤとを比較する。『本は』と『本は』共に名詞の「本」であるので、両者の単語ＩＤは一致する。

【００５３】

ステップＳ８－５では、ステップＳ８－３，ステップＳ８－４の処理結果を元に、用例抑制、用例復活の判定を行う。ここでは、対象文節が変更されたならば（ａ）、ステップＳ８ー６へ進み、用例の抑制処理を行う。また、対象文節とその相手文節のペアと一致する用例があれば（ｂ）、ステップＳ８ー７へ進み用例復活処理を行う。

【００５４】

この実施例では(a)の条件に該当するので、処理はステップＳ８－６に進む。ステップＳ８－６は、用例を抑制する処理であり、図４の（ｃ）における用例学習データの用例ＩＤに対応する位置に"１"を書き込む。これにより、用例『本は／厚い』は抑制されたこととなる。

ステップＳ８－７は、用例復活が行われる場合の処理である。ここでは、図４の（ｃ）における用例学習データの用例ＩＤに対応する位置に"０"を書き込む。

【００５５】

次に、ステップＳ５－３により、再び文字列『ほんはあつい』が入力される（図９（ｅ））。

ステップＳ５－４により、文字列『ほんはあつい』にかな漢字変換が適用される。ここで、前回までの処理により、同音語『熱い』が単語学習されている。また、用例『本は厚い』は抑制されている。更に、同音語『本は』（名詞）は単語学習されている。その結果、図９（ｆ）で示されるように『本は熱い』が第一候補として決定される。この場合の『本は』は名詞の「本」である。

【００５６】

次に、ステップＳ５－５により、『熱い』を『厚い』に同音語変更する（図９（ｇ））。その後、ステップＳ５－６により選択学習処理が行われる。

ステップＳ５－６では、前述と同様にして、初めにステップＳ７－１により『厚い』が単語学習される。次に、用例『本は／厚い』について、文頭側（本は）、文末側（厚い）の候補単語情報の比較が行われる。その結果、２つの同音語の単語情報における単語ＩＤは、用例情報の単語情報における単語ＩＤと一致するため、用例『本は厚い』は復活が行われる。またこの後同音語『本は』は単語学習される（図９（ｈ））。

【００５７】

尚、以上の説明において用いた辞書は単語情報と用例情報とを同一の見出し語の下に記述した単一の辞書構造であるが、単語情報と用例情報とを分離した各辞書テーブルとしても同様に処理することができる。

また、本実施例で用いた単語ＩＤおよび用例ＩＤは、先頭の単語または用例を１とする連続番号で表したが、各々の辞書の先頭からの相対アドレスを用いることもできる。
また、本実施例における単語学習データ、用例学習データの持ち方は、各学習データに学習が行われる場合は１を、学習が行われない場合は０を書き込んでいるが、学習データに、学習が行われる場合は単語ＩＤまたは用例ＩＤを書き込み、学習が行われない場合は単語ＩＤまたは用例ＩＤを削除するといった、単語ＩＤ、用例ＩＤの有無で示しても良い。

【００５８】

尚、本発明は、複数の機器から構成されるシステムに適用しても、１つの機器から成る装置に適用しても良い。また、本発明はシステム或は装置にプログラムを供給することによって達成される場合にも適用できることはいうまでもない。以上のように本発明によれば、用例が存在する１組の語（文節間）において，その同音語への変更が行われ用例抑制学習が適用された場合、語変更が行われなかった語について単語学習を行うことにより、次の入力における第一候補決定において、単語学習された同音語が優先的に第一候補となり、用例が存在する同音語を構成する単語と比較して、表記は同じであるが単語ＩＤの異なる同音語が第一候補に選ぶことはなくなるため、確実に用例の抑制または復活を実現することができる。

（実施例２）

第２の実施例では、かな漢字変換での用例抑制/復活での上述の問題を、実施例１とは異なる方法で解決する文書処理装置を提供する。

【００５９】

実施例２の文書処理装置の全体処理構成は、実施例１で図５を用いて説明した構成と同様である。
実施例２では、説明を簡潔に行うために、実施例１での構成と同様な構成部分については説明を割愛する。そして、異なる構成についてのみ以下に説明する。図１０は、ステップＳ５－４のかな漢字変換処理を詳細化したフローチャートであり、これは実施例１とは異なる処理である。

【００６０】

以下、図１０を参照して、かな漢字変換処理の手順を説明する。
ステップＳ１０－１は、入力された仮名漢字変換の対象となる文字列を解析し、文節としての切り出しを行い、抽出する処理である。ここでの処理は、単語辞書（図１の５－２）を検索して解析を行い、文節として認定される候補のみを出力する処理であって、同種の文字処理方式において一般的に行われている処理であり、説明を割愛する。

【００６１】

ステップＳ１０－２は、ステップＳ１０－１において出力された解析結果に対して、単語辞書中に格納されている用例のパターンが存在するかどうかをチェックし、用例のパターンが存在すれば、その用例の対象となる変換語の候補を優先候補として抽出する。
ステップＳ１０－３は、ステップＳ１０－２で抽出された優先候補の中から、学習の結果なども考慮に入れて、仮名漢字変換の第一候補を決定する。

【００６２】

ステップＳ１０－４は、図３で示される文節情報および用例情報を作成する処理である。これらの情報は用例抑制学習を始めとする各処理で使用される内部的なデータである。
ステップＳ１０－５は、図３の（ｃ）で示される文節情報のうちの候補単語情報の補正を行う処理である。第一候補の語の単語情報に基づいて、用例情報の探索を行い、その単語情報と比較してその表記が同じで単語ＩＤが異なる場合、同音語情報の持つ単語情報の単語ＩＤを用例情報の持つ単語情報の単語ＩＤと置き換えを行う。ここで、探索対象の用例は、抑制がかかっているものも含まれる。この処理の詳細については、図１１、図１２を参照して後述する。

【００６３】

ここで、図１４を参照して、入力文字列が『ほんはあつい』の時の具体的処理の流れを説明する。

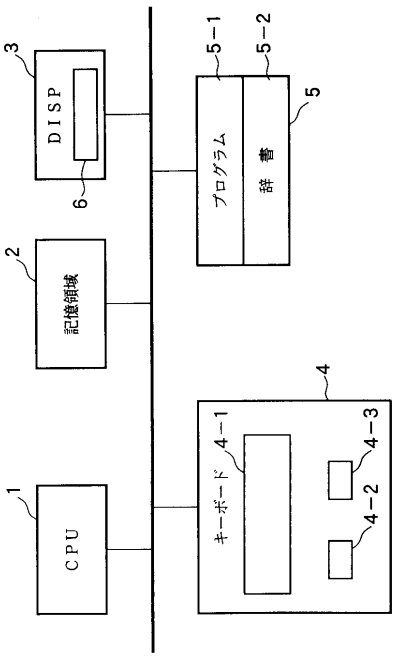
図１４（ａ）で示されるように、まず、入力文字列『ほんはあつい』が入力される（図５のステップＳ５－３）。
次に、ユーザ指示である変換キーを入力すると、ステップＳ５ー４へ進む。ステップＳ５ー４ではかな漢字変換処理を行う。この処理の詳細を図１０に示す。
【００６４】
図１０のステップＳ１０－１からステップＳ１０－３までの処理により、文節の抽出、用例の抽出、第一候補の決定が行われ、『本は厚い』が第一候補に決定される（図１４（ｂ）参照）。この時の語『本は』、『厚い』には、図１４（ｂ）で示されるような同音語情報が作成される。第一候補は、同音語情報の先頭に格納されるので、『本は』は、表記「本」、単語コード「４００」、その他の情報として品詞「名詞」を持つ。一方、『厚い』は、表記「厚」、単語コード「１００」、品詞「形容詞」を持つ。
【００６５】
次に、次候補変換キーが入力されると、図５のステップＳ５－５へ進み、次候補処理により『厚い』を『熱い』に変更したとする。
次に、選択キーが入力されると、図５のステップＳ５ー６へ進み、選択学習処理が行なわれる。ここで、用例『本は／厚い』は、後述する用例抑制学習（図１３参照）が行われて抑制される。
【００６６】
そして、その語他の文字入力が行われ、接頭語の『本』が単語学習されたりした場合、接頭語の『本』が単語辞書では、選択優先度が最も高くなっているとする。
この後、新たに文字列『ほんはあつい』が入力されるとする（図１４（ｄ））。
そして、変換キーの入力によりステップＳ５ー４のかな漢字変換処理へ進み、ここでは、文節に分解した後に用例検索を行うが、用例『本は／厚い』はすでに抑制されているので、対応する用例はない。そして、単語辞書の検索で、接頭語の『本』が最も選択優先度が高くなっているので、文節『ほんは』の『ほん』には接頭語の『本』が第１変換候補として選択される。また、同様に、文節『あつい』に対しては、単語辞書検索でたまたま選択優先度の高かった『熱い』が第１変換候補として選択されたとする。この時の単語辞書における同音語情報の様子を図１４（ｅ）に示す。図１４（ｅ）に示すように、先頭に接頭語の『本は』が格納されている。また、『あつい』に対する同音語も『熱』が先頭に置かれている。
【００６７】
ステップＳ１０ー５では、この状態で同音語情報の補正を行う。この同音語情報の補正処理の詳細を図１１のフローチャートで示す。
図１１を参照して、ステップＳ１１－１では、処理の対象となるすべての文節について以下の処理を繰り返し行うための制御を行うために、未処理の残り文節数のチェックを行い、残り文節があればステップＳ１１ー２へ進む。なければここでの補正処理を終了する。
【００６８】
例えば、『本は／熱い』という２文節で構成される文字列については２回ループ処理が行われるように制御する。
ステップＳ１１ー２では、対象文節に関する用例があるか否かをチェックする。
そして、対象となる文節に用例が存在する場合はステップＳ１１－３へ進み、なければここでの補正処理を終了する。
【００６９】
ステップＳ１１－３では、用例が存在する１組の文節対の文頭側の文節（語）について、その同音語情報の単語情報と用例情報の文頭側単語情報との間で、表記と単語ＩＤの比較を行う。その結果、表記が同じで単語ＩＤが異なる場合は、同音語情報における単語情報の単語ＩＤを用例情報における単語情報の単語ＩＤに置き換える。この処理は、図１２のフローチャートを参照して後で詳しく述べる。
【００７０】
ステップＳ１１－４は、用例が存在する１組の文節対の文末側の文節（語）について、ス

テップＳ１１－３と同様の処理を行う。次にステップＳ１１ー１へ戻り残っている文節に関する処理を繰り返す。
以上の手順で同音語情報の補正処理を行うことができる。
図１２は、図１１におけるステップＳ７－３又は、ステップＳ７－４の処理の詳細を示したフローチャートである。
【００７１】
ステップＳ１２－１では、同音語候補の数だけ以下の処理を行う。図１４（ｅ）では、例えば、語『本は』の候補単語情報は『本（接頭語）』、『本（名詞）』、『本（助数詞）』、『奔（名詞）』の４つが存在するため、この４つについて以下の処理を繰り返すための制御を行う。
ステップＳ１２－２では、文節情報の候補単語情報と、用例情報の単語情報を比較する。
【００７２】
用例情報の単語情報には、用例『本は（名詞）／厚い（形容詞）』が格納されている。
ステップＳ１２－３では、ステップＳ１２－２における比較で、表記が一致すればステップＳ１２ー３へ進み、一致しなければステップＳ１２ー５へ進む。
文節情報の第一候補の表記は『本（接頭語）』であり、用例情報の単語情報・表記『本（名詞）』と一致する。
【００７３】
ステップＳ１２－４では、表記が一致した単語情報について、単語ＩＤの比較を行い、一致すればステップＳ１２－６へ進み、一致しなければステップＳ１２ー５へ進む。
文節情報の単語情報『本（接頭語）』の単語ＩＤは４０１である。用例情報の単語情報『本（名詞）』の単語ＩＤは４００である。この場合、両者は一致しないので、処理はステップＳ１２－５に進む。
【００７４】
ステップＳ１２－５では、次の語情報を処理対象とし、ステップＳ１２ー１へ戻る。
図１４（ｅ）の例では、同音語情報の２番目の候補として『本（名詞）』が格納されている。同音語情報『本（名詞）』は、ステップＳ１２－１～ステップＳ１２－４の処理の結果、用例情報の単語情報『本（名詞）』と、表記、単語コードが共に一致すると判断されるので、次に処理はステップＳ１２－６に進む。
【００７５】
ステップＳ１２－６では、表記、単語コードが共に一致する単語情報について、文節情報の第一候補の単語情報における単語ＩＤを、用例情報の単語情報における単語ＩＤと置き換える。これにより、図１４（ｆ）で示されるように、同音語情報の第一候補（先頭）には、単語ＩＤ４００の『本』が格納される。これで同音語情報の補正処理が終了する。
【００７６】
次に、図１４（ｇ）に示すように、語『熱い』を『厚い』に変更する。『厚い』が決定されると、図５のステップＳ５－６により選択学習処理が行われ、用例抑制学習が行われる。
図１３は用例抑制学習の詳細を示したフローチャートである。
ステップＳ１３－１では、すべての語候補について処理を行うための繰り返し処理制御を、未処理の語数をチェックすることによっておこなう。即ち、未処理の残り語数が０でなければ、ステップＳ１３ー２へ進む。また、０であれば、ステップＳ１３ー１へ戻る。
【００７７】
例えば、例文の『本／厚い』の場合は、同音語『本は』、『厚い』について２度ループ処理するように制御する。
ステップＳ１３－２は、対象となる語候補に関する存在する全ての用例情報について以下の処理を繰り返すように、未処理の残り用例候補数をチェックすることによって制御する。
【００７８】
図３（ｅ）に示したように、用例が存在する１組の文節間において、その用例情報は文末

側の語に持たせている。用例『本は／厚い』は、文末側の同音語「厚い」の用例情報ポインタによって示されている。
図１４の例では語『本は』に用例候補は存在しない。よって、処理はステップＳ１３－１に戻り、次の語『厚い』に対しての処理が行われる。この語『厚い』には用例情報が存在している。
【００７９】
ステップＳ１３－３では、用例情報の文頭（相手）文節の候補単語情報における単語ＩＤと同音語の単語ＩＤとの比較を行う。
図１４では、用例情報の文頭文節候補単語情報は『本は（名詞）』であり、同音語の候補単語情報は『本は（名詞）』である。よって、両者の単語ＩＤは一致する。
【００８０】
ステップＳ１３－４では、用例情報の対象文節（文末）の候補単語情報における単語ＩＤと同音語の単語ＩＤとの比較を行う。図１４では、用例情報の対象文節は『厚い』であり、同音語の候補単語情報は『厚い』である。よって、両者の単語ＩＤは一致する。
ステップＳ１３－５では、ステップＳ１３－３，ステップＳ１３－４及びその他の条件を考慮し、用例情報の抑制、復活の判定を行う。ここでは、対象文節が変更されたならば（ａ）、ステップＳ１３ー６へ進み、用例の抑制処理を行う。また、対象文節とその相手文節のペアと一致する用例があれば（ｂ）、ステップＳ１３ー７へ進み用例復活処理を行う。
【００８１】
ステップＳ１３－６は、対象となる用例情報を抑制する（その用例を使用禁止とする）処理であり、用例学習データの用例ＩＤに対応する位置に１を書き込む。
ステップＳ１３－７は、対象となる用例情報を復活する（使用禁止であった用例を使用可能とする）処理であり、用例学習データの用例ＩＤに対する位置に０を書き込む。
【００８２】
図１４の例では、文頭、文末共に単語ＩＤが一致するので、条件(b)により、用例『本は／厚い』は復活処理が行われる。
しかしながら、図１０のステップＳ１０－５による補正処理が行われなければ、図１４（ｅ）から（ｆ）への変更は行われず、同音語情報の第一候補は単語ＩＤ４０１の『本（接頭語）』のままである。ステップＳ１３－３による処理では両者の単語ＩＤは一致せず、用例の復活は行われない。
【００８３】
このように、第一候補で決定された同音語の単語情報が用例情報の単語情報と比較して表記が同じで単語ＩＤが異なる場合でも、同音語情報の補正処理ステップＳ１０－５を行うことにより、表記、単語ＩＤ共に一致した同音語情報が作成され、用例の復活が可能となる。
尚、以上の説明において用いた辞書は単語情報と用例情報を同一の見出し語の下に記述した単一の辞書構造であるが、単語情報と用例情報とを分離した各辞書テーブルとしても同様に処理することができる。
【００８４】
また、本実施例で用いた単語ＩＤおよび用例ＩＤは、先頭の単語または用例を１とする連続番号で表したが、各々の辞書の先頭からの相対アドレスを用いることもできる。
また、本実施例における単語学習データ、用例学習データの持ち方は、各学習データに学習が行われる場合は１を、学習が行われない場合は０を書き込んでいるが、学習データに、学習が行われる場合は単語ＩＤまたは用例ＩＤを書き込み、学習が行われない場合は単語ＩＤまたは用例ＩＤを削除するといった、単語ＩＤ、用例ＩＤの有無で示してもよい。
【００８５】
また、上述の文書処理例では、入力文書はひらがなであったが、これはカタカナでもよいことは言うまでもない。
また、上述の文書処理例では、入力文書はひらがなであったが、これはカタカナや漢字を

含んだものでもよいことは言うまでもない。
以上説明したように本実施例によれば、用例が存在する１組の文節間において、第一候補で決定された同音語情報が用例情報の単語情報と比較し、表記が同じでも単語ＩＤが異なる場合、同音語情報の単語情報における単語ＩＤを用例情報の単語情報における単語ＩＤに置き換えることにより、用例復活処理の際、対象となる同音語において、表記が同じでも単語ＩＤが異なる、つまり表記が同じである別の単語であるということがなくなり、用例復活処理を確実に行うことができる。
【００８６】
【発明の効果】
以上説明したように、本発明によれば、読みと表記が同一で品詞の異なる単語を別の単語として記憶した単語辞書を用いたかな漢字変換において、一旦抑制した用例に対応する表記語に変更する操作がなされた場合に、同音同表記で用例に無い別の単語が存在してもそれに影響されずに、その抑制した用例を確実に復活させることができるという効果がある。
【図面の簡単な説明】
【図１】本発明における一実施例の文書処理装置のブロック図である。
【図２】本発明の一実施例の辞書構成の例を示す図である。
【図３】本発明の一実施例の文書処理で生成する文節情報と用例情報のデータ構造を示す図である。
【図４】本発明の一実施例の文書処理で生成する学習データのデータ構造を示す図である。
【図５】本発明の一実施例の文書処理全体の処理手順を示すフローチャートである。
【図６】本発明の一実施例の文書処理での仮名漢字変換部の処理手順を示すフローチャートである。
【図７】本発明の一実施例の文書処理での選択学習処理部の処理手順を示すフローチャートである。
【図８】本発明の一実施例の文書処理での用例抑制学習の処理手順を示すフローチャートである。
【図９】本発明の一実施例の文書処理手順の理解を容易にするための具体例による処理の流れを示す図である。
【図１０】本発明の第２の実施例の文書処理での仮名漢字変換部の処理手順を示すフローチャートである。
【図１１】図１０のステップＳ１０－５での同音語情報補正の処理手順を示すフローチャートである。
【図１２】図１０のステップＳ１１－３、ステップＳ１１－４の詳細処理手順を示すフローチャートである。
【図１３】本発明の第２の実施例の文書処理手順での用例抑制学習の処理手順を示す図である。
【図１４】本発明の第２の実施例の文書処理手順の理解を容易にするための具体例による処理の流れを示す図である。

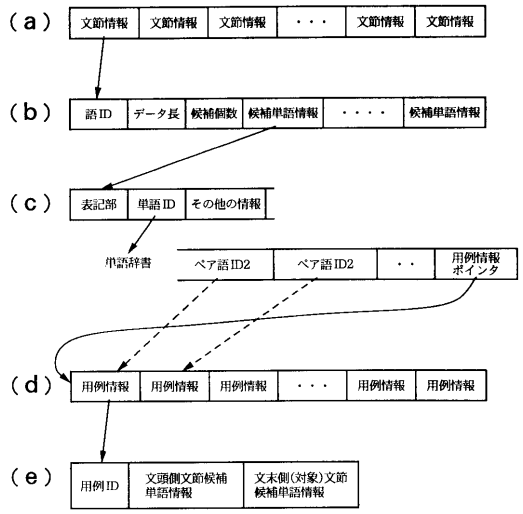
【図１】



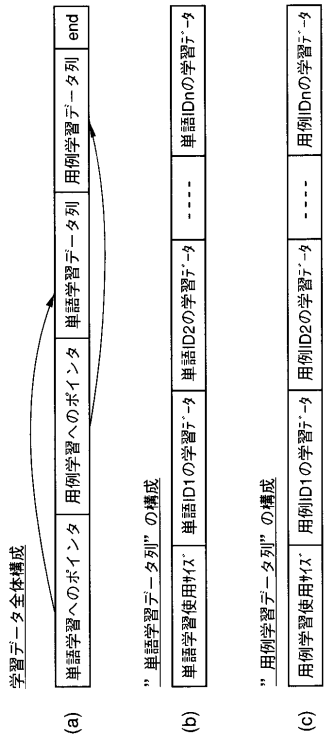
【図２】

単語情報					用例情報				
ＩＤ	読み部	表記部	品詞	その他の情報	個数	ＩＤ	用　例	結合情報	その他の情報
100	あつ　い	厚	形容		30	12	板	が	
						13	紙	が	
						14	胸	が	
						15	本	が	
≈	≈	≈	≈	≈	≈	≈	≈	≈	≈
300	か　う	買	動詞		2	123	酒	を	
						124	喧嘩	を	
301	か　う	飼	動詞		1	125	動物	を	
350	か　つ	勝	動詞		1	126	喧嘩	に	
4567	たつ	立	動詞		20				
4678	はな　す	放	動詞		1	543	動物	を	

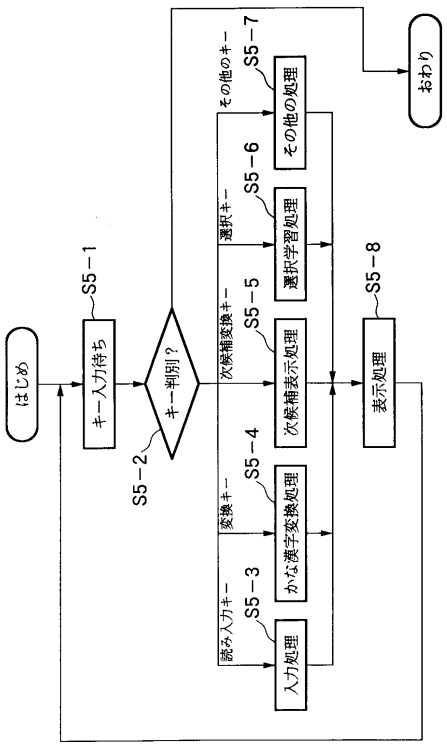
【図３】



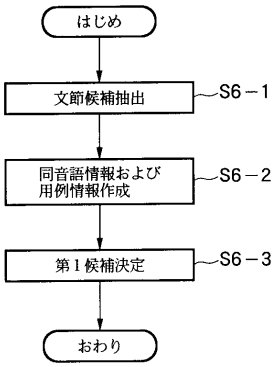
【図４】



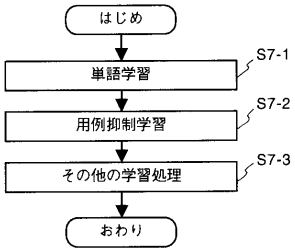
【図５】



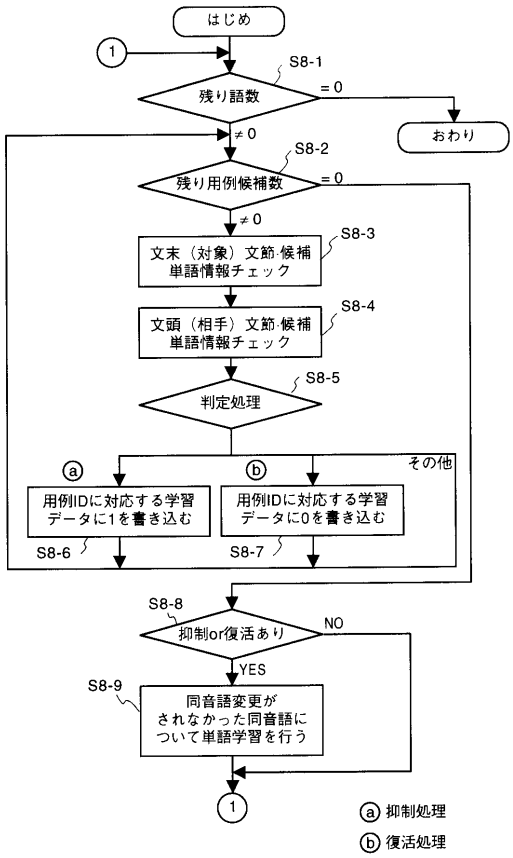
【図６】



【図７】



【図８】



【図９】

(a) 入力 ほ ん は あ つ い

(b) 変換 本 は 厚 い

「本は」（名詞）が第一候補に抽出される

(c) 次候補変換 「厚 い」 ⇨ 「熱 い」

(d) 選択学習処理 単語学習 「熱い」が単語学習される

用例抑制学習 「本は／厚い」が抑制される

「本は」（名詞）が単語学習される

(e) 再入力 ほ ん は あ つ い

(f) 変換 本 は 熱 い

「本は」（名詞）が第一候補に抽出される

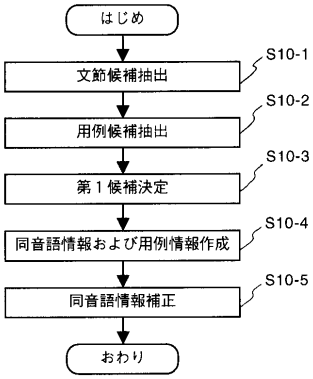
(g) 次候補変換 「熱 い」 ⇨ 「厚 い」

(h) 選択学習処理 単語学習 「厚い」が単語学習される

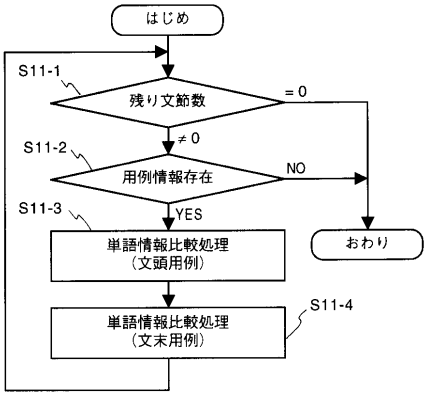
用例抑制学習 「本は／厚い」が復活される

「本は」（名詞）が単語学習される

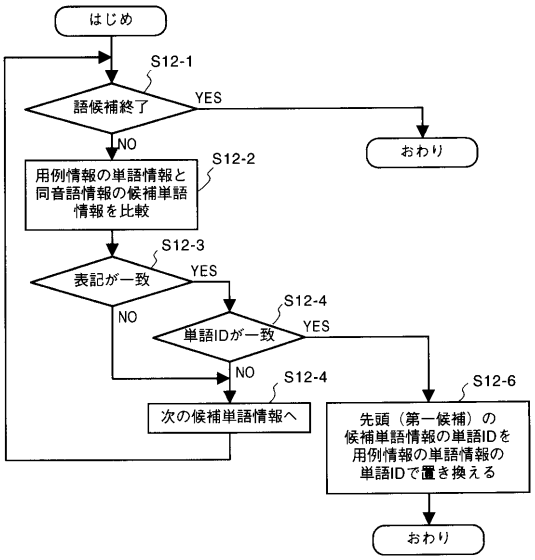
【図１０】



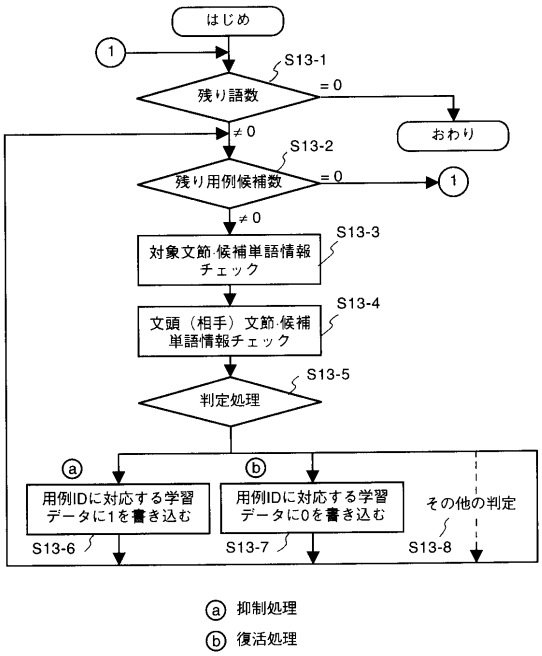
【図１１】



【図１２】



【図１３】



ⓐ 抑制処理
ⓑ 復活処理

【図１４】

(a) 入力　ほ　ん　は　あ　つ　い

(b) 変換　本　は　厚　い

「本は」の同音語情報

同音語情報

表記	単語ＩＤ	品詞
本	400	名詞
本	401	接頭語
本	402	助数詞
弁	701	名詞

「厚い」の同音語情報

表記	単語ＩＤ	品詞
厚	100	形容詞
熱	101	形容詞
暑	102	形容詞
篤	103	形容詞

(c) 次候補変換　「厚　い」　⇨　「熱　い」

(d) 再入力　ほ　ん　は　あ　つ　い

(e) 変換　本　は　熱　い

「本は」の同音語情報

同音語情報

表記	単語ＩＤ	品詞
本	401	接頭語
本	400	名詞
本	402	助数詞
弁	701	名詞

「熱い」の同音語情報

表記	単語ＩＤ	品詞
熱	101	形容詞
厚	100	形容詞
暑	102	形容詞
篤	103	形容詞

(f)

「本は」の同音語情報

同音語情報

表記	単語ＩＤ	品詞
本	400	名詞
本	401	接頭語
本	402	助数詞
弁	701	名詞

「熱い」の同音語情報

表記	単語ＩＤ	品詞
熱	101	形容詞
厚	100	形容詞
暑	102	形容詞
篤	103	形容詞

(g) 次候補変換　「熱　い」　⇨　「厚　い」

フロントページの続き

(56)参考文献 特開平０３－０４１５５１（ＪＰ，Ａ）
特開平０５－１０８６２２（ＪＰ，Ａ）
特開平０４－１１１０４３（ＪＰ，Ａ）
特開平０８－２７２７９３（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.[7]，ＤＢ名)
G06F 17/22 514