

携帯情報端末用メモパッドと固有名詞の音声入力方式

鹿島 博晶 蒲谷 武正 河原 達也

京都大学大学院 情報学研究科 知能情報学専攻

〒606-8501 京都市 左京区 吉田本町

e-mail: kashima@kuis.kyoto-u.ac.jp

あらまし 携帯情報端末ではテキスト入力に音声を用いることは有用であり、これと併用し描画にペン入力を用いることでマルチモーダルな入力を実現するメモパッドを構築した。また携帯端末においては固有名詞の入力が重要であるが、一方音声入力を用いる上でエントリ数を不必要に増大させることは避けなければならない。これに対し本報告では、まずかなレベルのカバレッジの高い単語辞書を用い、そこでカバーできない単語については単漢字レベルのカバレッジが高くかつ人の想起しやすい単語を多く含んだ単語辞書を用いてカバーするという手法を提案する。この手法に基づき提案する辞書を Wnn4 の単語辞書および Web 上でフリーウェアとして公開されている単語辞書を用いて構成した。

キーワード 携帯情報端末、固有名詞、メモパッド、音声入力、単語辞書

A Method of Speech Input of Proper Nouns for PDA

Hiroaki Kashima Takemasa Kamatani Tatsuya Kawahara

Graduate School of Informatics

Kyoto University, Kyoto 606-8501, Japan

e-mail: kashima@kuis.kyoto-u.ac.jp

Abstract In PDA, it is useful to use speech as text input channels, and we developed a PDA Memo pad with multi-modal input channels. In PDA, proper nouns are frequently input, while the number of lexical entries must not be unnecessarily increased. We propose a method that at first the lexicon is used which contains many entries for the proper nouns and secondly, for the uncontained proper nouns, the lexicon is used which contains many single kanji characters of proper nouns and the words easily remembered. We constructed the lexicons on this method using the lexicons of Wnn4 and free software on the Web.

key words PDA, proper nouns, Memo pad, speech input, word lexicon

1 はじめに

近年シャープのザウルスシリーズや IBM の WorkPad など、携帯情報端末 (PDA: Personal Digital Assistant) が製品化されている。主にキーボードではなくペンで入力し、電子メールやスケジュール管理等を行うことができる。

その最大の特長として携帯性が挙げられ、これを考慮するとキーボード入力は不適當である。また歩行中などペンによる文字入力は困難であることから文字入力には音声を用いた入力が有用であると考えられる。以上のことにより携帯端末への音声認識器の実装が想定される。

携帯端末にはアドレス帳、スケジュール機能等の固有名詞 (人名、地名) の入力が頻繁に行われるアプリケーションが搭載されている。一方音声認識ではシステムに登録しておいた単語しか扱えないという問題があるが、これに対して未登録単語の処理に対して漢字レベルで解決する研究はなされていない。したがって携帯端末において音声入力を用いるには未登録の単語を漢字レベルで入力できる方式が必要である。

固有名詞の特性を考えると文脈などから情報を得ることは難しく、孤立単語音声による入力を想定する。この下では単語辞書の整備が特に重要であるが、特に未登録単語も扱えるよう辞書を構成する方式を提案する。

具体的には、カテゴリ毎に単語を登録して固有名詞辞書を構成し、これでカバーできない単語については、単漢字表記説明用単語辞書を用いて単漢字レベルでカバーするという方式を用いる。

2 携帯情報端末用音声入力メモパッド

2.1 マルチモーダル入力の意義

我々の日常のコミュニケーションでは、音声や指示ジェスチャといった、情報の伝達モードを単独で用いるのではなく、目的や場所に依じて臨機応変に使い分け、また組み合わせたりしている。人間とコンピュータのインターフェースのなかでも、特にこのような点に注目したものをマルチモーダルインターフェース [1][2] と呼ぶ。本メモパッドでは、音声による文字の入力とペン入力による手を用いた図像の描

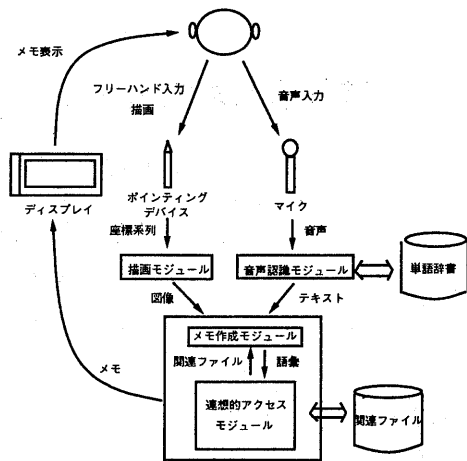


図 1: メモパッドシステムの構成図

画を同時に行えるようにすることで、マルチモーダルな入力システムを実現している。

2.2 メモパッドシステムの構成

システムの音声認識パーザには、本研究室で開発された Julian [4] を用いる。Julian を使ってマイクからの音声を認識した結果をスコア順で返し、スコアの一番高いものを入力として受けとり、ペンを使ってシート上の指示した場所に入力していく。音声入力によるコマンドの解釈等は行わないものとし、メニューの選択等はすべてマウス (実際の PDA ではペン入力) のみを用いて行う。

2.3 認識誤りの訂正

音声による文字の入力は文字の入力をマイクで行っているため、描画やメニュー選択をもう一方の手を使って文字入力とは独立に行えるというメリットや、タイピングのように、技術の習得を特に必要としないという利点がある。その一方で、認識精度がキーボードなどと比較すると悪く、また認識のためのディスプレイが避けられないという問題がある。

本システムでは認識精度の問題、すなわち認識誤りの問題に対しては、候補を複数選択できるようにすることで対処している。実際には、音声入力が行われた後、マウスの右クリックを 1 回おこなうと Julian

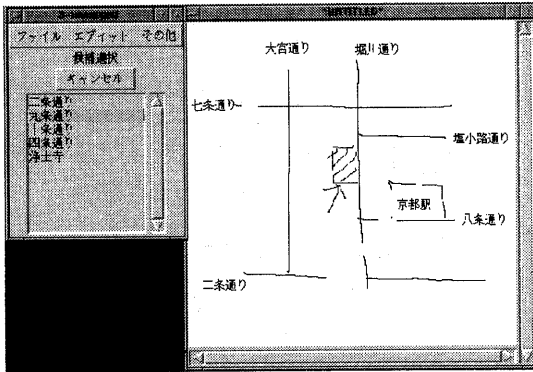


図 2: メモパッドの画面写真

の第1パスの結果が表示され、その時点でもし間違っていたら右ボタンをダブルクリックすることで複数候補が別画面上に表示されるようになっている。これを tcl/tk により実装した画面図を図 2 に示す。

2.4 メモパッドシステムの評価

一般的な PDA (シャープ製 ザウルス) の文字認識を用いた入力と比較した結果を示す。被験者 6 名に対し、まずシステムの説明を行ない、システムの入力を何回か行なってもらった後に、メモの入力を行なってもらった。評価の基準としては、入力に要した時間と認識誤りの回数 (誤り訂正回数) の 2 つを用い、6 名の被験者の平均をとった。タスクの入力時間を表 1 に、入力の訂正回数を表 2 に示す。

なお、両システムの認識時間はどちらも 2 秒以内におさまっており、今回はマシンの性能等に起因する差についてはほとんど無視できるものとして比較を行なった。

この結果、音声入力の方が手書き文字入力に比べておよそ半分の時間で入力をおこなうできた。これは、発話時間が、文字の入力時間と比較して短くすむことや、単語単位の制約を用いて認識を行なっているために、音声入力の方が認識誤りが少なくなったためであると考えられる。

文字入力では認識誤りの例として、「1」を「シ」に誤ったり、「目」を「自」に誤って認識するなど、少し字にくせのある人の入力を繰り返し誤る場合があり、これが入力時間や訂正回数が増えた主な原因

表 1: 入力時間

ユーザー	音声入力	文字入力
A	2分 22秒	5分 56秒
B	4分 37秒	6分 19秒
C	2分 28秒	12分 40秒
D	3分 27秒	7分 22秒
E	3分 8秒	4分 46秒
F	2分 53秒	7分 59秒
平均入力時間	3分 9秒	7分 50秒

表 2: 入力の訂正回数

ユーザー	音声入力	文字入力
A	2回	9回
B	2回	7回
C	0回	16回
D	2回	9回
E	5回	13回
F	0回	7回
平均訂正回数	2回	10回

となった。

3 固有名詞入力システム

既存のシステムで用いている従来の構成方式の単語辞書では、大部分の固有名詞がその性質上あらかじめ登録されていない。一方携帯端末にはアドレス帳など固有名詞が頻繁に入力されるアプリケーションが搭載されており、そのような従来の構成方式の単語辞書を用いるのではユーザに負担を与える。本固有名詞入力システムでは人名、地名の入力に対して高くカバーできるため、既存のシステムに統合することによりその語彙が大きく拡張されることが期待される。

3.1 入力方式の概要

本システムにおける固有名詞の入力過程は、3のようになっている。入力する固有名詞のカテゴリが指定しているという前提の下で入力単語の読み「かし

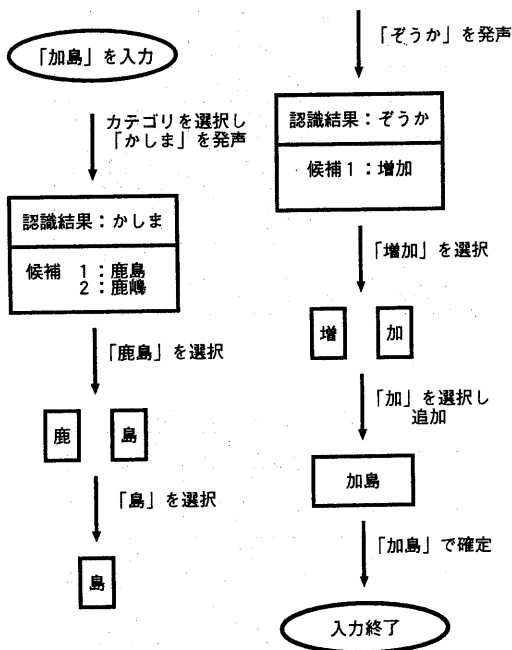


図 3: 固有名詞入力 of 具体例

ま」を発声すると、その読みがかな文字として認識され、そのかな文字に対する候補単語である同音異表記の単語が出力される。その中に、意図した単語「加島」と同じものは無いが、その一部の単漢字を含む「鹿島」を選択し、さらに必要な文字「島」のみを選択する。次に、「加」を含む普通名詞「増加」の読みを発声し、候補単語の中から「増加」を選択する。この単語から必要な文字「加」を選択して「島」に追加し、「加島」を確定する。

ここで、固有名詞の読みを発声する際、意図したもの異なるかな文字として認識された場合は、何度か発話を繰り返す。一定回数繰り返しても正しく認識されない場合は、音声認識エンジンの単語辞書にそのかな文字が登録されていないものと判断し、他の単語で入力を行うものとする。

3.2 システムの構成

まず入力された音声进行かな文字に変換する処理を行う。かな表記の認識結果を出力することによりユーザに対し誤認識なのか意図した漢字表記がないのか

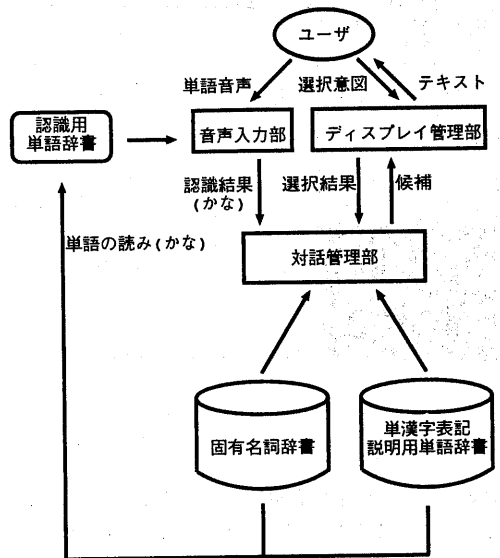


図 4: システム全体の構成図

を明示する。孤立単語音声の入力を想定しており、認識エンジンとして本研究室で開発された Julian [4] を用いる。辞書の語彙数が 1 万 - 1 万 5 千程度になるため、音響モデルには前後の音素コンテキストを考慮した状態数 2000 で 1 状態あたり 16 連続混合分布をもつ triphone モデルを用いる。認識用単語辞書としては、固有名詞辞書あるいは単漢字表記説明用単語辞書に存在する単語の読みとなっているかな文字から構成されたものとなる。認識用法としては、名詞を一単語認識するように記述されたものとなっている。

4 単語辞書の構成

アドレス帳などのアプリケーションを想定しているため、固有名詞として人名、地名を考える。したがって単語辞書としては名字用、名前用、地名用を構成した。

4.1 固有名詞辞書

人名(名字、名前)用としてまず、かな漢字変換辞書の Wnn4 付随の辞書の中から人名に用いられると示されていた単語と地名と人名双方に用いられる

表 3: Web 上から収集した人名単語

	Web 名字	Web 名前	Web 人名
単語数	17076	30497	44077
かなレベル	11768	6501	19236
Web 人名全単語			
単語数	60153		
かなレベル	22744		

と示されていた単語を入手した。ただし、Wnn では人名を名字と名前に分類していない。また、地名用については、Wnn から地名に用いられると示されていた単語と Web 上でフリーウェアとして公開されていた地名辞書（全国地名テキスト MS-IME95 用）[5] 中の都道府県、市区郡町村名の単語を追加した。

人名用の固有名詞辞書としては、入力時の負担を軽減するためかなレベルのカバレッジが大きいことが必要である。下記のカバレッジ比較実験の結果により Wnn4 に付随の人名辞書のみでは不足であるため、人名については Web 上でフリーウェアとして公開されていた人名辞書 [6][7][8][9][10][11][12] を収集した。

またカテゴリ毎に単語を収集する必要があるが、一般的な人名辞書は姓、名の区別がなく人名として一括して登録されているため、この種の辞書に対しては姓、名の分割が必要である。これを解決するために名字として登録されている単語群のかな表記と名前として登録されている単語群のかな表記の二つの集合を用いて、名字あるいは名前のみを用いられる単語群のかな表記集合を構成した。

Web 上から収集した単語の概要については、表 ?? に示す。

4.2 単漢字表記説明用単語辞書

単漢字の読みの単独発声は音声認識において誤認識する可能性が高い。また全ての単漢字にその表記を説明する情報（「翔」はヒツジ「羊」にハネ「羽」など）を付加する作業は非常にコストがかかる [3]。したがって意図する単漢字を含む単語を入力し、そこから意図する単漢字を選択することを考える。

表記説明用単語辞書の構成においては、

- 誤認識を減らすため語彙数を少なくする
- 固有名詞に用いられる単漢字の表記を全て含

表 4: 単語を用いた単漢字表記説明の例

単漢字	単漢字を含む単語
眞	眞一(しんいち)、眞(まこと)、眞紀(まき)...
衛	衛星(えいせい)、護衛(ごえい)、防衛(ぼうえい)...

表 5: カバレッジ調査に用いたデータ

	名字	名前
単語数	1450	1450
かなレベル	842	549
単漢字数	410	349

むようにカバレッジを高くする

- 人が想起しやすい使用頻度の高い語を含んでいる

ということを考慮した。具体的にはまず単語として固有名詞辞書の語彙に含まなくても、その個々の表記は一般的な単語においてよりも人名あるいは地名においてよく用いられる表記であろうという考察から、固有名詞辞書を用いた。また、表記を説明するのに用いる単語として他の人名、地名が思い付きがたい表記も考えられるため、一般的に用いられる単語を語彙に持つかな漢字変換辞書 Wnn4 に付随の基本辞書から名詞のみを用いた。表記説明に用いる単語の例を表 4 に示す。ただし基本辞書の名詞に関しては固有名詞の単漢字を説明する上で必要と考えられる単語だけを用いた。すなわち基本辞書の単語のうち人名あるいは地名には用いられないような表記しか含まない単語は不必要であるため、固有名詞辞書の単語に含まれる単漢字を持つ名詞のみ用いた。さらに基本辞書に付加されている頻度情報が 0 とされている単語は想起しづらいと考えられるため頻度 1 以上のものだけを用いた。以上のような固有名詞辞書を基にした方式で単漢字表記説明用単語辞書を構成した。

4.3 人名に対するカバレッジの比較予備実験

堂下研究室の過去の名簿データと Web 上でフリーで公開されていた 1998 年のプロ野球の人名辞書 [13] のデータ（表 5）を用いた。

表 6: 名字に対するかなカバレッジの比較

	Wnn 人名	名字用	全人名
単語数	4756	19602	60936
かなレベル	3493	12974	23269
かなカバレッジ (%)	74.9	93.7	99.2

Wnn 人名: Wnn4 付随の人名辞書
 名字用: Wnn 人名から名前を除き Web 名字を追加
 全人名: Wnn 人名 + Web 人名全単語

表 7: 名前に対するかなカバレッジの比較

	Wnn 人名	名前用	全人名
単語数	4756	30778	60936
かなレベル	3493	6646	23269
かなカバレッジ (%)	78.3	99.1	99.4

Wnn 人名: Wnn4 付随の人名辞書
 名前用: Wnn から名字を除き Web の名前を追加
 全人名: Wnn 人名 + Web 人名全単語

表 6, 7 より Wnn の人名辞書をそのまま用いるとカバレッジが低く、また収集した全人名を用いるとカバレッジは高くなるが単語数が不必要に増大する。したがって Wnn の人名から名前 (名字) にしか用いられない読みを持つ単語を除き、Web から収集した名字 (名前) に用いられる単語を追加して人名用固有名詞辞書を構成した。予備実験結果よりこの辞書のかなカバレッジは十分高いことが示された。

これらの固有名詞辞書に対し同じデータを用いて漢字レベルのカバレッジを調べた結果は表 8 のようになり、約 10% の単語は直接入力できないことが示された。つまり固有名詞を単漢字レベルでカバーす

表 8: 名字、名前に対する漢字カバレッジ

	名字用	名前用
単語数	17969	30484
かなレベル	12586	6580
漢字カバレッジ (%)	90.3	94.1

表 9: 全単語辞書の概要

	Wnn 全単語	Web 全単語
単語数	35436	64088
かなレベル	23065	26569

表 10: 単漢字レベルのカバレッジ比較 (名字)

	全単語辞書	名字説明 1
単語数	91353	26235
かなレベル	43539	19008
単漢字数	3768	2680
単漢字カバレッジ (%)	99.9	99.9
	名字説明 2	Wnn より構成
単語数	18694	11956
かなレベル	14348	9538
単漢字数	2456	2083
単漢字カバレッジ (%)	99.5	98.0

全単語辞書: Wnn 全単語 + Web 全単語
 名字説明 1: 固有名詞辞書の名字用を用いた
 名字説明 2: 1 から頻度 5 以下を除いた固有名詞辞書を用いた
 Wnn より構成: Wnn 人名を固有名詞辞書として用いた

る単語辞書として単漢字表記説明用単語辞書が必要であることが示された。

次に単漢字のカバレッジの比較予備実験を行った。ここで固有名詞の頻度情報を用いているが、これは Wnn の人名辞書に付与されていた頻度と先述した Web 上から入手した辞書のうち頻度情報をもつ辞書からかなレベルの頻度辞書を構成したもの (単語数: 21690 語, かなレベル: 13400 語) を参照している。ただし、この頻度辞書では姓、名の分割はなされていない。また全単語辞書は Wnn4 に付随の単語辞書に含まれる全ての単語と Web 上から収集した人名、地名全単語を加えたものである (表 9)。単漢字カバレッジの比較予備実験の結果は表 10, 11 のようになった。

この結果より人名用の単漢字表記説明用単語辞書として、名字説明 2 と名前説明 2 を提案する。これらの提案辞書は全単語辞書に比べ語彙数が 25% から 40% に削減されているがカバレッジについては大き

表 11: 単漢字レベルのカバレッジ比較 (名前)

	全単語辞書	名前説明 1
単語数	91353	38705
かなレベル	43539	13041
単漢字数	3768	2622
単漢字カバレッジ (%)	99.9	99.9
	名前説明 2	Wnn より構成
単語数	25637	11956
かなレベル	10547	9538
単漢字数	2456	2083
単漢字カバレッジ (%)	99.8	97.7

全単語辞書：Wnn 全単語 + Web 全単語

名前説明 1：固有名詞辞書の名前用を用いた

名前説明 2：1 から頻度 5 以下を除いた固有名詞辞書を用いた

Wnn より構成：Wnn 人名を固有名詞辞書として用いた

表 12: 地名用単語辞書の概要

	固有名詞辞書	単漢字説明用辞書
単語数	6249	13459
かなレベル	6143	12208
単漢字数	--	2062

な低下は見られない。また Wnn 人名を固有名詞辞書として構成した単漢字表記説明用単語辞書は語彙数はやや少ないが単漢字カバレッジが低下する。つまり提案辞書の語彙数が適切に削減されていることが示された。

以上より人名入力の際に音声認識システムで用いる辞書として、提案した固有名詞辞書と単漢字表記説明用単語辞書は有用であることが示された。

なお地名用については、地名用の固有名詞辞書と基本辞書の名詞のうちその単漢字を含む頻度 1 以上の単語群を用いて構成し表 12 のようになった。ただし予備実験は行っていない。

5 評価実験

図 4 をもとに Perl/Tk により実装を行った。その画面写真を図 5 に示す。

このシステムを用い、話者 5 名による評価実験を

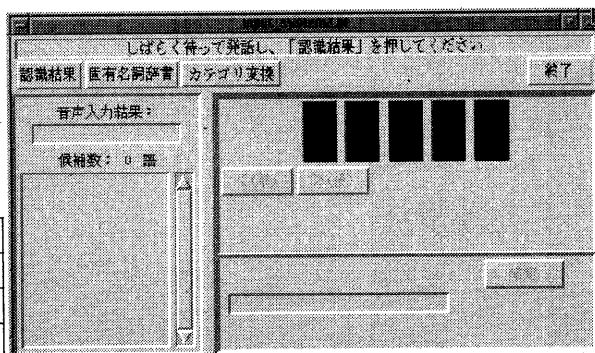


図 5: 固有名詞入力システムの画面写真

表 13: 平均発声回数の比較実験結果

	名字用	名前用	地名用
従来辞書	3.20 回	3.18 回	4.56 回
提案辞書	2.18 回	2.35 回	4.06 回

行った。各話者には比較予備実験で用いたデータの中から 20 単語の固有名詞 (名字 8 単語, 名前 8 単語, 地名 4 単語) を入力してもらい、予備実験で用いたような従来の構成方式の単語辞書と提案辞書とで一単語入力当りの発声回数にどの程度の差が生じるのかを調べ、表 13 のような結果を得た。ただしここでは 3 回発声して入力されない場合、被験者はその読みが登録されていないものと判断し他の単語を入力するものとした。

実験の結果、固有名詞を一単語入力する際に提案辞書を用いた方がより少ない発声回数で済むことが示された。

6 まとめ

本稿では、携帯情報端末においてテキスト入力時に音声を用いることを考え、マルチモーダルな入力を用いたメモパッドと固有名詞の入力方式を提案した。

実際に話者による評価実験の結果、メモパッドを用いると従来の文字認識と比較し文字入力の平均時間が半分に削減された。また固有名詞入力システムに実装した提案辞書を用いると従来の構成方式の単語辞書に比べ一単語入力当り発声回数が削減され

本稿で提案したシステムおよび単語辞書の有用性が示された。今後は現在手動で行っている固有名詞辞書と単漢字表記説明用単語辞書の自動切り替えの設計、実装を行う予定である。

参考文献

- [1] 池田徹志, 佐々木哲之, 荒木雅弘, 堂下修司. 音声・ジェスチャ・画像を統合したマルチモーダル情報の理解, 人工知能学会研究会資料, SIG-SLUD-9703-1, (1996).
- [2] 野村淳一, 佐々木哲之, 荒木雅弘, 堂下修司. マルチモーダル作図システムにおける文脈知識を利用した発話理解, 人工知能学会研究会資料, SIG-SLUD-9703-1, (1997).
- [3] 大山芳史, 浅野久子, 高木伸一郎. 姓名漢字表記を説明する対話システムの試作と評価. 情処研報, SP14-8, pp.53-59, (1996).
- [4] 李晃伸, 河原達也, 堂下修司. 文法カテゴリ対制約を用いたA*探索に基づく大語彙連続音声認識パーザ. 情処学会論文誌, Vol.40, No.4, pp.1374-1381, (1999).
- [5] 全国地名テキスト MS-IME95 用,
<http://www.vector.co.jp/vpack/browse/software/data/writing/sn044245.html>.
- [6] MS-IME95 用 人名テキスト,
<http://www.vector.co.jp/vpack/browse/software/data/writing/sn066952.html>.
- [7] ATOK10/11 用 人名テキスト,
<http://www.vector.co.jp/soft/data/writing/se066951.html>.
- [8] 人名辞書 (WXIII 用),
<http://www.vector.co.jp/soft/data/writing/se008149.html>.
- [9] WX 用 人名辞書 (名前),
<http://www.vector.co.jp/soft/data/writing/se011404.html>.
- [10] ATOK8,9 用 人名テキスト,
<http://www.vector.co.jp/soft/data/writing/se018539.html>.
- [11] ATOK 用人名辞書,
<http://www.vector.co.jp/soft/data/writing/se011392.html>.
- [12] gerodic (人名に関する固有名詞辞書),
<ftp.wg.omron.co.jp>.
- [13] 1998 年プロ野球人名辞書,
<http://www.vector.co.jp/vpack/browse/software/data/writing/sn067818.html>.