

姓名漢字表記を説明する対話システムの試作と評価

大山芳史* 浅野久子** 高木伸一郎**

*NTT コミュニケーション科学研究所

**NTT 情報通信研究所

〒239 神奈川県横須賀市光の丘1-1

ooyama@cslab.kecl.ntt.co.jp {hisako, takagi}@nttnly.isl.ntt.co.jp

あらまし ショーちゃんのショーは、「ヒツジ (羊) にハネ (羽)」で「翔」とわかるように、日本人の名前の綴りを自動的に説明し合成音声で出力する説明文生成の対話システム (EXPLANET) を構築した。本システムを用いると聞き手は多くの同音語の候補から目的の漢字が容易に判断でき、解が不明確な場合でもインタラクティブに別解を生成させることができ、文字あたり平均98.8%の伝達精度を達成した。また、別解を生成するモードでは、そうでないモード場合に比べて誤りの率は約1/3であった。

キーワード： 文生成、自然言語処理、対話文生成、音声出力システム、漢字説明

An Interactive Explainer of Kanji Characters used in Japanese Person's Names and Its Evaluation

Yoshifumi Ooyama* Hisako Asano** Shin-ichiro Takagi**

* NTT Communication Science Laboratories

**NTT Information and Communication Systems Laboratories

1-1, Hikari-no-oka, Yokosuka-shi, Kanagawa 239

Abstract The newly developed system can explain how to write a person's name automatically using the original readings and can give additional explanation that allows recipients to choose among many kanji character candidates. For example Mr. *Sho's* "spelling" is "sheep and wing", which indicates only one candidate from many homophones that have the same reading of the name *Sho*. Recipients can select other explanations if the original explanation is not clear for them. The accuracy in transmitting the kanji to human listener is about 98.8% per kanji character. The error rate is one third if the system provides another explanations.

Keywords: Text Generation, Natural Language Processing, Spoken Language Generation, Text-to-Speech System, Kanji Explanation

1. はじめに

日本語のメッセージを音声で伝えるとき、我々は同音語に注意して話を進める場合がある。特に姓名の綴りを正確に伝える場合は、正確な読みだけでなく、その文字を説明するフレーズを追加して対話するが多い。例えば「キョーコのキョーは、木にクチ」で「杏」と分かる。また、音声で文字を確実に伝達したい時や、特に聞き慣れない固有名詞の場合など、話し手と聞き手の間で、その確認のために上記説明と同じように書き言葉では現われない説明を追加する。通信用語として電報で使われていた「サクラのサ」などはその一種である。

複数の表現を用いて文字を音声入力する研究としては、英語で通常の読みに加えて、スペルを1文字ずつ言うことで、人名などを入力する試みがある[1]。日本語の場合、表意文字であり、その説明方法が多様で組み合わせも多い。筆者らは、その表現の多様性を利用して、同音語の漢字からターゲットとなる漢字を選択するインテリジェントなカナ漢字変換プログラム(TOPAZ)を作成し、漢字電報のオペレータ支援用として実用化している[2]。

一方、言い替えや書き換えによる話し言葉(音声)の生成の研究としては、新聞の見出しのように紙面の関係で凝縮された書き言葉を、モーラ数は長めで、より分かりやすい話し言葉の表現に書き換えて音声で読み上げる研究[3]、意味が比較的近い同音語に対して、例えば、「今秋→コンアキ」のように「今週」と区別させる専門家向けの新聞の校正支援の実用システム[4]がある。また、マルチモーダル生成の研究で、映像メディアであるアニメとコーディネートすることで、書き言葉の情報より簡易な表現で文やフレーズを生成し、音声出力する報告[5]がある。

我々は、電話による対話のように、音声メディアだけを用いる環境において、日本語の漢字のメッセージのうち、日本人の姓名の漢字表記を対象に、まず、システムが自動的に表記を説明するシステム(PLANET)を構築した[6][7]。今回さらに、インタラクティブに、別の表現の説明を聞くことのできるシステム(EXPLANET)を作成した。音声合成のソフトウェアはFLUET[8]を用いている。

聞き取り実験により表記の再生をしたところ、文字あたり平均98.8%の正解となった。また、別解を聞くことができる場合はそうでない場合と比べて、誤りは約1/3であった。

本システムは、姓名以外の漢字の説明にも応用でき、正確な読み上げでは情報が不足する場合のメディア変換機能として適用可能である。

2. 漢字説明のパターン

電話により音声で漢字を伝える人と人の対話を収集し分析すると、主に(1)文字や意味の集合を指定

した説明、(2)漢字の構成要素の列挙による説明、およびそれらの組み合わせで表現されている。

(1) 文字や意味の集合の指定による説明

・熟語の利用

「文章のブン」や「規則のキ」のように、「熟語の読み」+「の」+「漢字の読み」、で表現される。これらは、熟語を用いることで、使われている文字セットを限定する。

・用言による説明

漢字の部分が1文字であるため、「規則のキ」の「キ」のように、漢字の部分を指定する必要がない。例えば、「走→走る、大→大きい」など。

・概念の利用

漢字の読みだけでは、同音語がある場合、その語の属性や上位・下位概念、同意語・類語を用いて、「果物の柿」、「船の帆」のように説明する。外国語からの「ストーンの石」もこれにあたる。さらに、順序づけたり並列にして説明する場合がある。

「一、十、百、千の千」、「暑い寒い寒い」、「右、左の右」などがそうで、音声にする場合、アクセント句が独立になるため、聞き取りやすい表現となる。

・陽に指定された文字セットの利用

「旧字体のサワ」「カタカナのサキ」などが該当するが、「カタカナのズ」は「ヅ」との区別がつかないので、さらに説明が必要である。

(2) 文字の構造の利用

漢字を構成する要素を組み合わせで説明する方法で、構成要素の読みを「に」で接続する。「杏(上下に分けて)→キにクチ」「浩(左右に分けて)→サンズイに告げる」また、同じ構成要素からなる漢字は、「圭→ツチ2つ」のように個数で説明する。

これらは、目的の漢字を使った適当な熟語がないときに有効である。従って、小学校の低学年で習得するような漢字については、理解できないわけではないが、とまどう表現になる場合があり、熟語からの説明が一般的である(以下に示す?と○の例)

? ゴンベンにオノレ ○ニッキのキ ← 記
? クサカンムリにハイ ○クサバナのクサ ← 草

(1)が意味を伝える方法で(2)は筆順や筆記方法を伝える説明ととらえることもできる。特に聞き手側が、文字を思い出せなくてしかも手書きで記述する場合には上記?の例は有効の場合がある。

筆順や筆記方法を伝えるという意味では、仮名の場合に、「ズ」「ヅ」に対して「スに点々」「ツに点々」という説明や「ャュョ」を1文字ずつ伝える伝えるときの「カタカナの小さいヤ」の表現もこれにあたる。

(3) 組み合わせ、置き換え、および否定

(1)(2)の組み合わせも多く(例1)、置き換え(例2)も時々使われる。

- 例1: 伶→ニンベンに命令のレイ
例2: 瑜→楡(ニレ)の木を王に

また、否定の表現として、多義を区別できる情報を用いる説明も存在する(例3, 例4)。

例4で、「島」の文字に「山」は存在するが「嶋」と対比すると、「島」は前述の「草」のように通常分割せずにとらえるからと思われる。

- 例3: 沢→難しくないサワ (澤は沢より難しい)
例4: 島→山のないシマ

3 文の生成

3.1 既存の知識からの文生成

漢字電報受付オペレータ支援用のTOPAZ[2]では、「さんずいにつげる」のような説明表現を意味解析するために、個々の漢字毎に漢字の形状を記述した構成情報、また、オペレータ用に個々の漢字の説明情報として、ディスプレイ上にコンパクトに表示した情報などを漢字構成辞書に収録している(表1)。

表1 TOPAZの漢字構成辞書[2]例

漢字	説明情報	構成情報 (要素1 Position 要素2)
翔	飛翔	(羊 R 羽)
浩	サンズイ-右-告	(サンズイ R 告)
杏	あんず/果実	(木 U 口)
郁	有-右-オオザト	(有 R オオザト)
芳	草冠-下-方	(クサカラムリ U 方)
備考	/:属性	R: 右側 U: 下側

これらの情報を使って説明文を音声出力する場合以下の課題がある。

(a)視覚的な支援情報は音韻・韻律情報が不足している。

(b)説明は漢字単位ではない場合がある。

例: 浩太郎は、サンズイに告げるに、普通の太郎これらの点を考慮して、部首名は単語登録を行って、

あらかじめ読みが既知の固有名詞などはその読みやアクセント型の指定が外部制御可能な、読み・韻律を付与するプログラム(AUDIOTEX [3])を用いるとともに、自分の名前についてどう説明するかを記述したデータ(姓名説明コーパス)を収集した。これを用いることにより、以下に示す効率的な説明文生成が可能となる。

(1) 複数文字の一括説明

- 例5: 普通のカトウです
例6: 川端康成のカワバタ
例7: 英雄(エイユウ)の英雄(ヒデオ)

カトウには「加糖、過当、加東」など同音語もあるが、姓名に限れば、「加藤、加東」となり、さらに出現頻度から見て、例5では「加藤」が選ばれる。これらは、名前に関する分野知識であり、この場合2文字がまとめて説明できるため、効率的といえる。著名な人名と同姓の場合、例6のような説明が使われる場合もある。

例7は、よく知られた一般語と同形で読みが異なる場合に有効である。「政治(マサハル)」も同様である。これは1文字毎に漢字の音と訓を用いるよりも、確実である。例えば、テイのサダでは、「定」と「貞」などの曖昧さが残る。

コーパスの収集の課題の一つは、網羅性であるが、漢字すべてについて適切な説明を収集することはコストがかかりすぎるため、文字単位で漢字構成辞書[2]等からの自動生成と併用した。

コーパスの収集の課題の一つは、網羅性であるが、漢字すべてについて適切な説明を収集することはコストがかかりすぎるため、文字単位で漢字構成辞書[2]等からの自動生成と併用した。

(2) 説明の接続

説明の表現としては、以下のケースが考えられる。

- (a) 表記のn文字までの説明と
表記のn+1文字以降の説明との接続
(b) 1個の漢字の構成要素の説明間の接続

ここで接続表現は上記下線の要素間の表現とする。

説明文 = <説明要素1><接続表現><説明要素2>

漢字が容易に構成要素に分けられる(b)の場合、例えば、偏や旁を組み合わせたフレーズを生成する。漢字約6500字の漢字構成辞書[2]をベースに構成要素に分けた説明フレーズを自動生成する。

人と人の対話では、この接続表現は、(a)(b)では「に」が最も多いが、「という字に」もあり、(b)では「(を)書いて」「の下に」「の右に」なども存在する。ただし、(b)のケースで、通常なら「に」で良いが、「の下に」や「の右に」を用いると曖昧さの削減が可能となる。これらは表1のPosition情報を用いて生成している。

また、「山偏に...」と表現することで、「山の右...」と等価であることから、「山偏の右に」のような重複した説明の生成を抑制している。これらより例8,9の2文字について区別を可能とした。

例8：嶋→ヤマヘンにトリ
例9：鳶→ヤマの下にトリ

また、説明要素1が用言で終了する場合、後方へ連体形で直接接続する。これにより、「明るい子」「青いに山」という表現ではなく、以下のような生成が可能となる。その他は「に」で接続する。

例10：明子→明るい子
例11：青山→青い山

(3) 音韻・韻律生成

(a) 読みとアクセント

垣根の'カキ'のように漢字一文字を表現する部分の生成は、以下のルールで行う。

- ・説明対象の漢字一文字が（連濁を除いて）その読みで自立語として存在すればその読みとアクセント型を優先する。例えば、垣根の垣(カキ')。
- ・自立語になければ読みの対応部分(連濁時は連濁)と未知語アクセント(1型)とする。例えば定規のギ。

また、強制的にアクセント型を指定したい場合にも、「文章のブン{1}」のように、「ブン」に対して、1型アクセントを採用するインターフェースを設けた[3]。

(b) ポーズ

同じ属性のものを列挙する場合、通常読み上げではアクセント結合が生じると不自然になることから、並列としてのアクセント、ポーズを付与するための、読点を付与している(例12)。

例12：千→「一、十、百、千の千」

3.2 聞き取り予備実験

姓名リストから、無作為に選んだ約200名分の試験データに対して説明文を生成し、被験者10名(A-J)による、聞き取り再現テストを行った。ここでは同じ姓名に対して一通りの説明しか生成しない[6][7]が、繰り返して再生することを許した。

表2はその結果であるが、一般に姓より同音語の多い名での誤りが多く、また個人差も大きく、最大で3倍以上の差があったが、共通した誤りをみると、以下の点が誤りの原因と考えられた。

- ・類似した音の語
ニチ(日)をミチ(未知)に、リュウキ(隆起)をユーキ(勇氣)に、というように、類似した音韻の語と誤る。

- ・アクセントの違う同音語

大正(0型)と大将(1型)はアクセントが違うが、被験者は確信がもてない。

- ・専門的な部首名

フデヅクリや、タマヘン(特に玉でなく王と記述する場合)は、一般的でないので意味不明となる。

表2 漢字再現のエラー(姓名各200件中)

	A	B	C	D	E	F	G	H	I	J
姓	7	8	6	4	4	3	9	10	5	3
名	14	14	14	9	7	17	17	20	8	5

この予備実験の結果から、一般的な説明の語彙を用いている姓名説明のコーパスを充実するとともに、漢字毎の説明で組み合わせられるものや出現度数の少ない説明は、コーパスから除くこととした。

4. システムの構成と処理の流れ

被験者によって、分野の知識、語彙の知識など個人差があることから、インタラクティブに別解を生成し、それを選べる対話システム(EXPLANET)を構築した。図1にシステムの構成を示す。

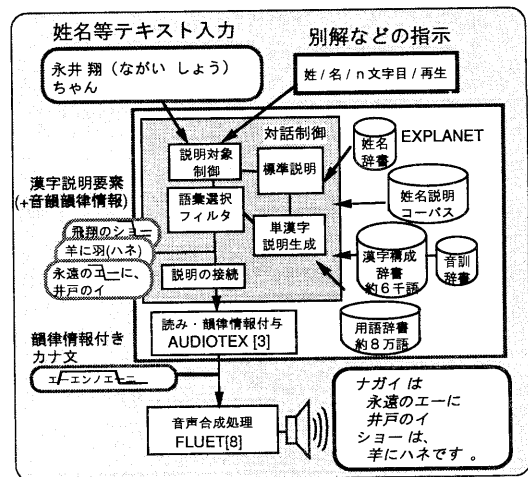


図1 EXPLANETのシステム構成

本システムは、WS上で、漢字説明の生成部は主にPerlで、読み韻律付与部[3]はCで、音声合成処理部は、NTTヒューマンインタフェース研究所音声情報研究部のテキスト合成ソフトウェア(FLUET[8])を用いて実現されている。

処理手順とモジュールの関係を以下に示す。

"()" は、モジュールを示す。

・ 説明対象の選択 (説明対象制御&標準説明)

漢字などの表記の列が入力されると、姓名の説明コーパスを検索し、最長マッチで検索を行う。姓や名の単位や3文字姓名の2文字でヒットすればその説明を用いて、そうで無い場合は、漢字毎に単漢字説明により説明文を生成する。また、別解の指示があった場合、説明対象の漢字の切り出しを行う。

・ 文/フレーズの生成 (単漢字説明生成)

1つの漢字を複数の構成要素に分け、それらの漢字を用いた説明を生成する場合、単純な音読みや訓読みではわかりにくい場合を考えて、その漢字を修飾する埋め込み文を生成する。また、漢字の対象を指示する音韻情報を生成し、フレーズを組み立てる。

・ 説明の順位 (語彙選択フィルタ)

コーパスより抽出したもので度数の少ないものは除き、残ったなかから熟語や属性を用いる説明を優先し、形状に分解した説明の優先度は下げている。また、例えば「大空さん」の2文字目を説明するときに、「大空の空」という説明にならないように、説明対象に当たる熟語を使用しないための語彙をフィルタさせる機能を持たせた。

・ 接続 (説明の接続)

漢字毎の説明、姓と名の説明を接続し、1文として説明文を作成する (漢字内の接続は、単漢字説明で行う)。

・ 対話制御処理 (対話制御)

システムが最初に生成する説明文に対して、姓単位の別解、名単位の別解、および文字位置に着目して、N文字目の説明の別解を個々に聞くことができるように制御するほか、対話の状態を1つ前に戻す機能がある。

5. 実験と考察

実験方法：

試験データ：姓名リストから、新たに約200名 (男女同数) の試験データを選び、これを男女名50ずつの2組 (N1, N2) に分け、別解を聞くことのできる対話モード(a)とそうでない非対話モード(b)を設定し、被験者4名がN1, N2とモード(a)(b)を組み合わせて (同じ名前を2つのモードで試験はしない)、該当する漢字をWPに投入し再現する実験を行った。ユーザインタフェース：被験者からの入力手段としては、別解の機能を図2のように割り当てた。図2の画面を見ながら、スピーカより説明文を聞いて、

姓名の表記をWPに入力する。(a)では、曖昧さがあると判断した場合には、メニューをマウスでクリックすることにより別解を聞くことができる。(b)では、9の再生のみ可能とした。対話のシーケンス例を図3に示す。

0	処理の開始
1	1 番目文字の別解
2	2 番目文字の別解
...	...
6	6 番目文字の別解
7	姓の別解
8	名の別解
9	繰り返し再生
*	1つ前の状態に戻る

図2 ユーザインタフェース (画面)

<数字：入力> <音声合成された説明文>

深井和裕さん
 ファイは、深い井戸です。
 カズヒロは平和のワに、
 裕福のユーです。

ケース1：

3(文字目)? (和) 昭和のワです。
 3(文字目)? (和) ノギヘンにクチです。
 4(文字目)? (裕) コロモヘンに谷です。

ケース2：

8(名の別解)? 昭和のワに、コロモヘンに谷です。

図3 説明シーケンス例

実験結果：

被験者がWPへ投入した漢字データとオリジナルの漢字データを比較し、誤った文字数をまとめたものが表3である。

表3 伝達誤りの文字数

モード	非対話 別解なし	対話 別解あり
姓	11/792(1.38%)	3/792(0.38%)
名	38/764(4.97%)	16/764(2.09%)

誤りは、どちらのモードも名の方が姓より多かった。モード間では、対話モードのほうが、非対話モードより誤りが、約1/4~1/2と少なく、対話モードでは文字当たり平均98.8%の文字伝達正解率を達成した。

また、対話モードで、被験者が別解など何のキーを押したか (処理希望したか) の度数を表4に示す。

表4 キーを押した回数 (394件データ中)

キ	1	2	3	4	5	6	7	8	9
	先頭からN文字目の別解生成							姓	名
回	10	25	49	34	5	0	65	213	16

名の部分の別解生成依頼も姓に比べて3倍以上多

く、日本人の姓は2文字が多いことから、N文字目の別解生成も、名の部分が多いといえる。また、誤ったデータに対して、対話実験終了後のヒアリングや再試験を行うことで、その原因として図4に示すものが想定された。

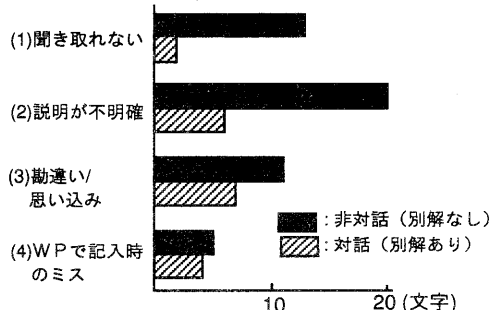


図4 漢字伝達誤りの主な理由 (1,556文字中)

図4の(3),(4)は、モードによる差は比較的少ないが、それらに比べて(1),(2)では、対話モードは非対話モードと比べて誤りが、それぞれ1/6, 1/3と格段に減少しており、別解を対話により生成する効果があると言える。

説明が不明確なものとしては、表5に示す類似または同音のものへの間違いがあった。これらは、使われても不自然さはない語彙なので、語彙選択としての検討も必要である。特に句レベルで同音となる語彙の自動的な判定も必要と考えられる。

表5 聞き誤り例

音声出力したデータ	誤って聞こえた解釈	読み
地名の津 清流 見本市	地名の地 声優/西友 日本一	類似
優良可 卓球のタク 恭賀新年のキョー	有料化 宅急のタク 今日が新年の今日	同一

誤りには至らなかったが、説明が分かりにくいものとして以下のものがあつた。

例13: 右紀子→右に...に子供の子

例14: 沢 →簡単なサワの旧字体

例15: 植木 →植物のショックに、植物の木

例13は、「右」が、位置関係を示す「右」と誤解されやすい。例14は、「沢→簡単な沢」と「澤→沢の旧字体」を結合したもので、相反する属性が使われ、かつ修飾語のスコープが不明確となった。例15の最初の「植物」は、「熟語による説明」の熟語であり、次の「植物」は、木の上位概念の「植物」であるが、同じ語が別の働きで使われており、理解が容易ではない。

例16: 「チホ (千穂)」か「シホ (志穂)」か

逆に、説明情報が有効だった例として、名前が不明だったが、説明文で補足できたケースがあつた。例16では、どちらか不明だったが、「セン (千)」で「チ」とわかつた。

また、コーパス収集とその利用に関しては、均質性の課題がある。「〇〇子」を、ある場合は「子供の子」と言ったり、ある場合は省略し、均質性に欠ける場合があり、被験者には不自然と感じられた。

6. まとめ

本稿では、漢字の表記を説明する対話システム (EXPLANET) の構成とそれを使った聞き取りの実験を行った結果を報告した。

本システムは、対話により漢字の候補を精度よく伝達できることから、電話での各種情報案内、新聞やメールの読み上げなど幅広い応用が考えられる。

人と人の対話でも、聞き間違いや言い替えは頻繁に発生することことから、忠実な読み上げに加えて、聞き取りにくい表現を積極的に判定して言い替えを行うユーザフレンドリなシステムの研究を進める。

<参考文献>

- [1] Hermann Hild, Alex Waibel, "Integrating Spelling into Spoken Dialog Recognition", EUROSPEECH'95.
- [2] Yoshifumi Ooyama, Tsuneaki Kato and Kenzi Imamura, "Japanese Kanji Name Input System using Spoken-style Transcribed Japanese Analysis", IEEE CAIA'95, pp. 2-8.
- [3] Koji Matsuoka, Eiji Takeishi, Hisako Asano, Ryomi Ichii, and Yosifumi Ooyama. "Natural Language Processing in a Japanese Text-to-Speech system for written-style texts" IEEE IVTTA, pp. 33-36, 1996.
- [4] 池原悟、安田恒雄、島崎勝美、高木伸一郎、"日本文訂正支援システム (REVISE)". NTT 研究実用化報告, Vol. 36, pp. 1159-1167, 1987.
- [5] Shimei Pan, Kathleen McKeown, "Spoken Language Generation in a Multimedia System", ICSP'96, pp.374-377.
- [6] 大山芳史、浅野久子、松岡浩司、"音声出力による姓名漢字表記の説明方式"、第52回情処全大,1996.
- [7] Yoshifumi Ooyama, Hisako Asano, and Koji Matsuoka, "Spoken-style Explanation Generator for Japanese Kanji using a Text-to-speech system", ICSP'96, pp.1369-1372.
- [8] Kazuo Hakoda, Tomohisa Hirokawa, Hajime Tsukada, Yuki Yoshida, and Hideyuki Mizuno, "Japanese Text-to-Speech Software based on Wave Form Concatenation Method", AVIOS '95.