

音声による属性情報付加を用いたかな漢字変換候補の選択手法

梅澤 猛[†] 大澤 範高[†]千葉大学大学院融合科学研究科[‡]

1. はじめに

同音語が多いという日本語の特性上、かな漢字変換の際に大量の変換候補が表示され、候補選択が困難になることがある。筆者らは、一般的なキーボード入力によるかな漢字変換を前提として、多くの変換候補の中から適切なものを容易に選択する手法を提案している[1]。変換候補の選択にあたっては、入力したい候補についての属性情報を補助的に与えることにより、候補の並び順を段階的に変更して入力確定を容易にする。属性情報としては候補に含まれる漢字の異音や部首名などを利用し、音声によって付加を行うことでキー操作の連続性を妨げない入力を可能としている。

本稿では、属性情報として用いる情報の網羅性と妥当性について検討するため、自分自身の氏名をどのように説明するのかについて追加アンケートを行いより多くの事例の収集と傾向の分析を行った。

2. 音声を用いた属性情報付加の利点

提案手法による文字入力の1例として、ユーザが「優季」と入力したいとするときの操作手順は次の通りである(図1)。まず、①かな情報である「YUUKI」を従来通りキーボード入力する。②IME機能により変換候補群が提示される。提示された候補群の中から「優季」を容易に見つけ出せない場合、③属性情報として「優」の訓読みである「やさしい」を音声によって入力する。④変換候補群の中から、「優」を含むものを上位とする並び替えを行う。このとき、もし新たに提示された変換候補群からも「優季」を見つげ出すことが難しければ③、④を繰り返し、さらに「季」を含む語として「きせつ」と音声入力するなどして所望の変換候補をさらに上位とすることも可能である。そして、変換候補の

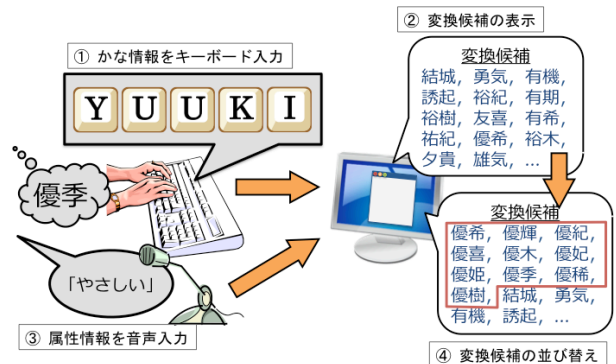


図1 提案手法による文字入力例

最上位(あるいは上位数件目)に並び替えられた候補「優季」を選択することで文字入力を完了する。

既存の環境で入力する別の方法に、当該漢字を含む別単語を入力して不要な部分を削除するものがある。「優季」の例でいえば、「やさしい」と入力して「優しい」と変換し、このうち「しい」を削除する。続いて同様に「季節」として「節」を削除することで入力が完了する。しかし、この場合には必要以上のタイピングが生じ、入力した文字の一部を削除する無駄も生じる。さらに、キーボードからの入力文字列が変換結果の本来の読みと一致しないため、学習辞書の機能があっても、「ゆうき」との入力で「優季」と変換できるようなにはならない。一方、提案手法では所望の変換結果の読みと入力文字列と一致するため、辞書の学習機能との親和性が高く、一度入力した文字列を属性情報なしで繰り返し変換することができる。

3. 人名漢字の入力方法調査

変換候補の並び替えに用いる属性情報について、効果が期待できるものを検討するため、人名入力時の漢字表記の表現方法をアンケート調査した。工学部の1年生161名に対して、自分の姓名を口頭で伝えるとしたらどのように表現するかを問い、150名分の有効回答を得た。

回答内容を精査したところ、1字に対して複数通りの回答があったものを含め、538例の属性情

A speech-based candidate selection using attribute information in Kana-Kanji conversion

[†] Takeshi UMEZAWA, Noritaka OSAWA, Graduate School of Advanced Integration Science, Chiba University

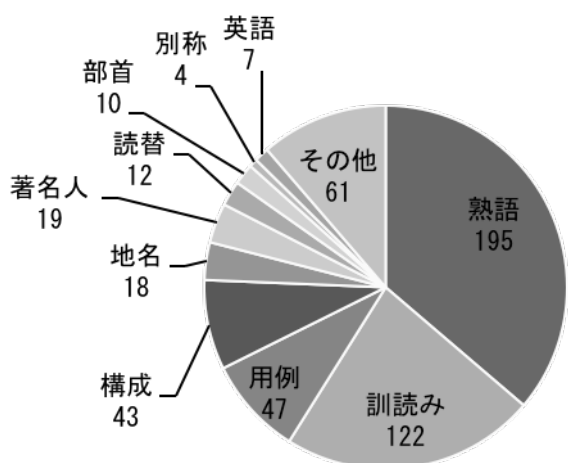


図 2 人名漢字の入力方法調査結果

報を得ることができた。属性情報の種類ごとの回答数を図 2 に示す。

- 熟語
介護の「介」、開拓の「拓」など同音語が多い字に対して典型的な熟語を示して特定を行うもののほか、克己心の「克」、崇拝の「崇」のように単独では説明が難しい字を伝えようとするものなどを含め、全体の中で最も使用例が多くみられた。
- 訓読み
みやびの「雅」、なりの「也」など訓読みすることで類似の漢字と区別する例。音読みの場合には複数の候補がある文字に対し、訓読みをすることで一意に候補が定まる例が多く見られた。
- 用例
本の「葉」などその文字を使った用例を示すもの。
- 構成
さんずいに光るの「洗」などの文字を構成する要素を説明するもののほか、紫の下糸を木に変えた「柴」のように基準になる漢字を差分とともに示すもの。
- 地名
大阪の「阪」、奈良の「奈」など地名の一部を示すもの。
- 著名人
オリエンタルラジオと同じ「敦彦」など良く知られた人物を引き合いに出すもの。
- 読替

優子を「優しい子」、賢人を「賢い人」とするなど、読み替えるもの。

- 部首
てへんの「拓」など類似の字の中から特定するための部首を示すもの。
- 別称
ハシゴ高の「高」、繰り返し記号の「々」など特殊な文字を表すもの。
- 英語
castle の「城」など平易な英単語で言い換えることで漢字を特定するもの。
- その他
その他にみられた例としては、漢数字の「一」という表現、浅い深いの「浅」という対になるものを挙げるもの、顔についている「口」などの説明や、たべもの「栗」など種別を示すものがみられた。これ以外に特徴的だった例として、比較的一般的な苗字について”普通の”「中原」という表現や、旧字体が存在する字に対して難しい方の「澤」、簡単な方の「沢」という表現があった。

4. まとめ

音声を用いた補助的情報の付加によって、人名入力時の仮名漢字変換の候補選択を行う際に用いる属性情報を検討するために、実際に自身の名前を説明する際に用いる情報を調査した。日常生活において、繰り返し説明する機会があることから、得られた回答からは当該の氏名を容易に特定できる表現の工夫がみられた。

音声を使った情報付加の観点からは、部首、別称、英語、その他として見られた例については、対象となる漢字に対して予め属性を登録しておくなどの対応が必要になり、属性情報として活用するには課題が残るが、該当する漢字については活用が容易な表現のものも多く含まれ実用上の障害は低いと考えられる。

参考文献

- [1] 河原直人, 梅澤猛, 大澤範高: 仮名漢字変換における音声を用いた情報付加による候補選択手法, 情報処理学会第 74 回全国大会, 6ZA-5 (2012).
- [2] 梅澤猛, 大澤範高: 人名入力における音声を用いた属性情報付加によるかな漢字変換の候補選択, 情報処理学会第 78 回全国大会, 5D-07 (2016).
- [2] 森信介, 土屋雅稔, 山地治, 長尾真: 確率的モデルによる仮名漢字変換, 情報処理学会論文誌, Vol. 40, No. 7, pp. 2946-2953 (1999).