

スマートフォンにおける日本語入力の実状と課題

小町 守
奈良先端科学技術大学院大学
komachi@is.naist.jp

木田 泰夫
Apple Inc.
kida@apple.com

1 はじめに

近年携帯電話の利便性と PDA の高機能性の双方を兼ね備えたスマートフォンが爆発的に普及している。スマートフォンではメールの読み書きやウェブのブラウジングなど、さまざまな場面で文字を入力する。これまでデスクトップにおける日本語入力は盛んに研究されてきたが、ハードウェア的な制約が強いスマートフォン環境における日本語入力は未だ発展途上である。また、iPhone/iPad に代表されるマルチタッチスクリーンなど、これまで存在しなかった入力デバイスの登場は、新たな日本語入力の可能性を開いたと言える。

そこで本発表では、タッチスクリーンを備えたスマートフォンにおける日本語入力の現状と課題について議論する。まず 2 節で 2011 年現在の状況について述べ、次に 3 節でスマートフォン特有の問題を検討する。最後の 4 節でデスクトップも含めた今後の日本語入力の方針を示す。

2 スマートフォンの日本語入力の現状

デスクトップではことえり (Mac) や MS-IME (Windows), Google 日本語入力 (Mac/Windows) などの日本語入力が用いられている。これらの手法は主にキーボードからの入力を画面に表示し、ユーザが入力した読みに対応する変換候補の中から選択する、という入力方法である。

一方、携帯電話やスマートフォンではキーボードに比べて入力速度が遅いため、先頭の数字文字をタイプすることによる予測入力の手法が広く受け入れられている。デスクトップの日本語入力にも予測入力をサポートするシステムはあるが、スマートフォンではデフォルトで予測入力が有効であり、常に複数の予測候補が画面に表示されている点が特徴である。

そこで、本研究では日本語入力のタスクを 1 かな漢字変換と 2 予測入力の 2 つのタスク^{*1}に分け、スマー



図1 フリック入力 (左 2 枚) および確定時予測 (右)。
フリック入力中は入力時予測も表示されている。

トフォンにおけるそれぞれのタスクの現状と、ハードウェア的な制約について述べる。

2.1 スマートフォンにおけるかな漢字変換

スマートフォンは従来の携帯電話の延長線上で語られることが多いが、かな漢字変換にとってもっとも大きな違いは、マルチタッチスクリーンの存在である。従来の携帯電話はハードウェア的に用意された 10 個の数字キーと数個の記号キーを用いて入力するが、スマートフォンでは画面上に自由にキーが配置されたソフトウェアキーボードによって入力する。

ひとつ例を挙げよう。図 1 は iPhone におけるソフトウェアキーボードを用いた入力の手順を示している。携帯電話では、数字キーのそれぞれに 50 音の 1 行が割り当てられており、たとえば数字の 9 の位置にあるキーを押すと順に「ら→り→る→れ→ろ→ら→…」という順に入力文字が遷移する (マルチタップ入力)。

一方、iPhone では携帯電話と同じようなマルチタップ入力も使うことが可能であるが、数字の 9 を押すと図 1 の左図のように 4 方向に各段が 4 方向に表示され、そのまま指を各方向にスライドさせることで、中央図のようにそれぞれの段の文字を 1 アクションで入力することができる (フリック入力)。フリック入力は熟練することによりマルチタップ入力よりも高速に入力できるため、iPhone の日本語入力以外にも iPhone では

^{*1} 統計的かな漢字変換 [6] の枠組みに従うと、日本語入力は $\hat{P}(x|y) = \arg \max_x P(x)P(y|x)$ ただし x はかな漢字混じり文、 y は出力に対する読み、と定式化でき、日本語入力モデル

$P(y|x)$ がかな漢字変換と予測入力とで違うことに相当する。

ATOK Pad for iPhone ^{*2} が、Android では Simeji ^{*3} がフリック入力をサポートしており、スマートフォンにおける入力の名詞として使われることもある。^{*4}

また、iOS 4 から通常の 10 キーでは * に当たる位置に顔文字キーが実装された (右図)。このキーは、もともと図 1 の左の画面のように濁音・半濁音・促音・拗音を入力するためのキーであったが、iOS 4 からは何も入力がないときに表示されるようになり、タップすることで (^_^) のような顔文字の選択ウィンドウを開くことができる。これは入力の冒頭では濁音や半濁音を入力することがないという日本語の性質から、役割を持たないキーに新しい機能を持たせた例であり、これはソフトウェアキーボードならではの機能拡張である。

一方、スマートフォンとデスクトップにおけるかな漢字変換の違いは、デスクトップが主にキーボードからの入力による操作を前提としているため、3 節で述べる「文節」に基づく変換を行なうが、スマートフォンや携帯電話では予測入力を組み合わせて文頭から順次確定していく入力スタイルが主流なため、「文節」に対する操作を持たない、といった点が挙げられる。

2.2 スマートフォンにおける予測入力

一般的に、予測入力には 2 種類ある。1 入力時予測と 2 確定時予測の 2 つである。入力時予測は図 1 の左図のように、2 文字「いわ」を入力中に表示される予測で、前方一致で候補が提示される。他方、確定時予測は右図のように、「どうぞよろしく」を確定したあと「お願いします」が表示されるといった予測で、前者と違って未確定の入力が存在しない、という違いがある。

アルゴリズム的観点からは、前者は未確定の入力文字列によって構築されるラティスのうち、末尾のノードの読みが部分的に与えられている、という状況における探索になる。携帯電話に比べ、スマートフォンは長文の未確定入力を一気に変換する入力スタイルのユーザも少なくないため、複数の文節に渡る入力も受け付ける必要がある。そのため、ラティスを構築する際に必要なメモリ容量に注意する必要がある。

データ構造的観点からは、後者は確定済みの入力から後続の文字列を予測するタスクであり、読みつぎの単語 n -gram を用いて予測することができる。しかしながら、スマートフォンにおいてはストレージ容量に制限があるため、いかにコンパクトに n -gram を格納できるか、といったことが問題となる。また、予測入力は文字入力の度に起動されるため、格納した n -gram の共

起情報にアクセスする際の処理速度も重要である。

2.3 スマートフォンにおけるハードウェア的な制約

以上述べたように、スマートフォンにおける日本語入力が携帯電話と異なる点はマルチタッチをサポートしたソフトウェアキーボードにあるが、デスクトップと一番異なる点は、ハードウェアによる制約である。比較のため、現在入手可能な代表的スマートフォンと、それと同スペックのデスクトップを、表 1 に掲げた^{*5}。

表から分かるように、もっとも大きな制約は使用できるメモリによる制限である。iPhone 3GS および HTC Magic (Google Dev Phone 2) と同程度のメモリを搭載している iBook G3 (初期) の発売日を比較すると、デスクトップで言えばおよそ 10 年前のスペックのハードウェアで動作するように日本語入力を開発しなければならない、ということである。

CPU の処理能力も現在のデスクトップと比較すると貧弱なので、現在のデスクトップにおける日本語入力で使われている手法^{*6}をそのまま用いることができない可能性がある。また、使用できるストレージも 10GB 前後であるため、辞書サイズにも制約がある。

3 スマートフォンの日本語入力の課題

3.1 スマートフォンに最適な入力単位

従来かな漢字変換における「文節」は学校文法における「文節」とは異なる、という議論がなされてきた^{*7}。

たとえば、「自然言語処理の教科書の書評を書いた」をことえりで変換すると

自然 | 言語 | 処理の | ブログを | 書いた

となるが、学校文法ではこれは

自然言語処理の | ブログを | 書いた

となり、かな漢字変換における「文節」は必ずしも学校文法における「文節」と一致しない。

つまり、学校文法における「文節」は自立語連続と付属語連続によって構成される単位であるが、かな漢字変換における「文節」は 1 つの自立語と付属語連続によって構成される単位 [5] となっているため、かな漢字変換における「文節」のほうが学校文法における「文節」より細かく分割される傾向がある。

なぜこのような違いが生まれるのであろうか？

学校文法の「文節」は統語的に決定されるが、かな漢

^{*2} http://www.justsystems.com/jp/products/atokpad_iphone/

^{*3} <http://www.adamrocker.com/blog/302/simeji.html>

^{*4} 携帯電話で慣れたマルチタッチ入力を使う人や、qwerty ソフトウェアキーボードによるローマ字入力を使う人も多い。

^{*5} スペックの出典は英語版 Wikipedia による。

^{*6} 統計的かな漢字変換はヒューリスティックとして用いられてきたコスト最小法 [8] の一つと見なすことができるが、N 文節最長一致法 [5] など、かな漢字変換の過渡期にはさまざまな手法が提案されていた。

^{*7} <http://chasen.org/~taku/blog/archives/2007/07/ime.html>

表1 スマートフォンとデスクトップのスペックと発売日の比較

モデル	iPhone 3GS	iPhone 4	iBook G3 Early	iBook G3 Late	HTC Magic	Nexus One
CPU	ARM 600MHz	ARM 1GHz	PPC 366MHz	PPC 900MHz	ARM 528MHz	ARM 1GHz
メモリ	256MB	512MB	320MB(max)	640MB(max)	288MB	512MB
容量	8GB	32GB	6GB	40GB	16GB	32GB
画面	480x320	960x640	800x600	1024x768	480x320	800x480
OS	iPhone OS 3.0	iOS 4	Mac OS 8.6	Mac OS 10.2.4	Android 1.6	Android 2.1
発売日	2009/06/19	2010/06/24	1999/07/21	2003/04/22	2009/07/28	2010/01/05

字変換のために定義されたものではない。従って、複数の同音異表記語が含まれる文節においては、表示できる上位 n 個の候補の中に正解が含まれないことがある。逆に、形態素単位で変換するスタイルにおいては、機能語や接辞などもそれぞれ独立に候補が表示されるため、自明な変換候補の確認を強いられる。

かな漢字変換の「文節」は、これら2つの極端の間で、ユーザの入力のストレス軽減のために産まれたものであろう。すると、スマートフォンにおける入力単位は、表示できる候補エリアがデスクトップより小さいため、従来のかな漢字変換の「文節」の妥当性は再検討する必要がある。

3.2 かな漢字変換と予測入力の評価尺度

かな漢字変換において、従来「このIMEは精度が高い」「新しいバージョンは精度が悪くなった」と漠然と言われてきているが、果たして日本語入力の「精度」とはなんであるか、ということも、議論の余地がある。

森ら[6]は文単位で変換して得られる1-best解と正解との最長共通部分列に基づく再現率と適合率により、かな漢字変換の精度を比較した。日本語は漢語由来の同音異表記語が多いが、それ以外表記の曖昧性のある箇所は少なく、大部分はほとんど曖昧性なく同定できる、といった特徴を考慮した尺度で、かな漢字変換の体感的な精度をより反映していると考えられる。

またGaoら[1]は文字誤り率(CER: character error rate)を用いてかな漢字変換の精度を評価した。CERはOCR文字認識でも用いられる尺度であるが、単語分割の曖昧性がある日本語の場合、機械翻訳の評価で用いられる単語誤り率(WER: word error rate)より頑健に評価でき^{*8}、性能を客観的に測るために優れている。

一方、予測入力においては、言語モデリングの分野ではパープレキシティが一般的に用いられているが、予測入力では入力に要した操作数も評価するため、文字単位打鍵数(KSPC: keystrokes per character)のような尺度が提案されている。KSPCを用いることで、予測

入力の精度改善のための評価の自動化ができる。^{*9}

しかしながら、一般ユーザが使用するアプリケーションにおいて、重要な評価尺度は精度だけではない。

たとえば、上記の再現率・適合率やCERは主に1-best解の評価を目的として使われるものであるが、スマートフォンにおいては、画面上に1-best解以外の候補も表示される。しかしながら、仮に1-best解が正解の候補であり、2番目以降の解に正解が含まれないとしても、2番目以降の解に「日本語としてはありえない」ような候補が提示されていた場合^{*10}、ユーザの精度に対する印象は大きく下がる。KSPCのような予測入力に適した自動評価尺度であってもこのような問題を捉えることはできないため、「いかにユーザにありえない候補を見せないか」といったことを評価の際に考慮しなければならない。

また、情報検索においては、検索の反応速度がユーザエクスペリエンス(UX: user experience)に影響を与えることが知られている[3]。日本語入力も同様に、予測・変換の速度がUXに影響を与え、しばしば変換精度よりもユーザにとってインパクトが強い。特にスマートフォンにおいてはCPUの処理速度がデスクトップより遅く、かつ辞書のサイズにも制約があるため、予測入力用の辞書を空間効率よく圧縮することに加え、展開時の速度とのバランスも検討する必要がある。

3.3 誤りを含む入力・曖昧な入力

スマートフォンは入力領域が狭いため、頻繁にタイプライプミスが起こる。また、キー入力にかかる時間がデスクトップより長い場合、入力の訂正にかかる時間も長くなる。そのため、入力誤り訂正はスマートフォンにおいて重要な機能である。

^{*9} 予測入力では最尤の候補を1つだけ提示する場合と、確信度以上の上位 n 個の候補を提示する場合とあり、 n の値はスクリーン上の表示領域などのデバイスの制約を受ける、という特徴がある。

^{*10} スマートフォンは携帯電話と同じくかな漢字変換の候補が予測入力の候補と混在して常時表示されるが、スマートフォンでは直接変換候補を1タッチで選択できるため、表示される候補間の順位が携帯電話やデスクトップほど重要ではない。

^{*8} 同じ目的で機械翻訳で広く使われている単語単位のBLEU[2]を文字単位で用いることもできる。

また、マルチタップやフリック入力の場合、(半)濁音や拗音・促音の入力に1操作かかるため、たとえば「かかく」で「科学」の変換も候補に入れることで、入力にかかる操作数を減らすことができる。

もっと極端には、T9^{*11}や[4]のように子音だけで入力することもできる。たとえば「かあかあ」で「こうかい」つまり「公開、後悔、更改、航海」などを入力する方式である。

これらの処理は間違った候補を提示してしまう可能性があるが、パーソナライズやトピック推定によって高精度に曖昧性解消が可能である場合、タイプ数の削減に効果的である。KSPCなどによる自動評価も、誤りを含んだ入力を用意して評価することで、より体感的な精度を反映したものになると考えられる。

4 これからの日本語入力に向けて

2節ではスマートフォンにおける日本語入力の現状と課題について議論した。本節では、スマートフォンに限らず日本語入力全般の課題について議論する。

4.1 開発・メンテナンスコストの削減

携帯端末における日本語入力は、デスクトップと異なる発展を遂げてきたものの、開発にかけられる人的資源もかぎられており、開発とメンテナンスに必要なコストを削減することが課題である。

スマートフォンにおける日本語入力は、デスクトップにおける入力と異なる点もあるが、類似する点も多いため、共通部分をくり出すことによって効率的な開発が可能である。たとえば言語モデルを共通化することによって、かな漢字変換と予測入力と異なる部分に焦点を当てることができる。

日本語入力の開発を行なうことができるエンジニアの確保も課題であり、少ない人数・人手でメンテナンスできる体勢を作ることが重要である。

4.2 日本語入力の規範性

日本語を入力する際に電子機器を用いることが一般的になり、日本語入力の候補になにを挙げるべきか、といった問題が再燃している。

古くは『電脳辞書の国語学』[9]に触発されてATOK監修委員会が設立された経緯は『電脳日本語論』[10]に詳しいが、「ATOKの言葉狩り」と揶揄されることもあるように、過度な自主規制となる恐れがある。

また、Google日本語入力^{*12}はウェブ上の単語の頻度に基づく日本語入力システムであるが、必ずしもウェブ上における頻度がユーザの期待と一致しないため、「ビジネスには使えない」といった批判や思わぬ予測候補が提示されたりする弊害がある。^{*13}

頻度主義は統計的自然言語処理の得意とするところであるが、特にUXに大きく関わる予測入力においては、規範性や単語の親密度[7]を考慮する必要がある。

4.3 入力のパーソナライズ

従来のかな漢字変換では、過去に変換した履歴を利用して入力のパーソナライズを行なうことが広く行なわれてきた。^{*14}

ユーザによって入力方式や確定のタイミング、主に使うアプリケーションなど組み合わせは多岐に渡るため、デフォルトの設定でそれなりに使えるシステムを提供しなければいけない反面、少量のユーザの入力でシステムが適応することが求められている。

しかしながら、現状では日本語入力のパーソナライズに決定的な手法は提案されておらず、ヒューリスティックな方法がなされている。今後は、推薦システムにおけるコンテンツベースの手法と協調フィルタリングの手法を組み合わせるなど、集合知を活用する方向性が考えられる。

参考文献

- [1] Jainfeng Gao, Hisami Suzuki, and Yang Wen. Exploiting headword dependency and predictive clustering for language modeling. In *Proc. of EMNLP*, pp. 248–256, 2002.
- [2] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In *Proc. of ACL*, pp. 311–318, 2002.
- [3] Eric Schurman and Jake Brutlag. Performance related changes and their user impact. Velocity 09, 2009.
- [4] Kumiko Tanaka-Ishii, Yusuke Inutsuka, and Masato Takeichi. Japanese input system with digits –can japanese be input only with consonants?–. In *Proc. of HLT*, pp. 211–218, 2001.
- [5] 牧野寛, 木澤誠. べた書き文の分かち書きと仮名漢字変換: 二文節最長一致法による分かち書き. 情報処理学会論文誌, Vol. 20, No. 4, pp. 337–345, 1979.
- [6] 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.
- [7] 天野成昭, 小林哲生. 基本語データベース: 語義別単語親密度. 学習研究社, 2008.
- [8] 久光徹, 新田義彦. 接続コスト最小法による日本語形態素解析. 情報処理学会第42回全国大会予稿集, 第3巻, pp. 1–2, 1991.
- [9] 箭内敏夫. 電脳辞書の国語学 — ワープロ日本語変換の徹底検証. おうふう, 1994.
- [10] 篠原歩. 電脳日本語論. 作品社, 2003.

^{*13} スマートフォンにおける日本語入力も同じ問題を抱えている。たとえば、iPhone OS 3.xでは「おは」と打つと「おはヨーグルト」が予測候補に登場するという現象が知られていた。

^{*14} 一般的にこの機能は「学習」と呼ばれることが多いが、予測入力を含めた日本語入力は、過去の入力に基づく分野適応も含めたより広い概念なので、ここではパーソナライズと呼ぶ。

^{*11} <http://www.t9.com/jp/>

^{*12} <http://www.google.co.jp/intl/ja/ime/>