

思考時間に基づくユーザインタフェース

7E-1

椎尾一郎

日本アイ・ビー・エム株式会社 東京基礎研究所

はじめに

連文節かな漢字変換方式や自動変換方式が、日本語入力方式として、一般的になってきた。これら的方式では、利用者が文節区切りを変換キーなどで指定する必要がなく、効率のよい日本語入力が期待できる。しかし、文節分割の過程で文法的には正しい多数の候補が出現し、この中から利用者の意図した文節区切りを選び出すのは、困難である。例えば、

「きょうはいしゃへいく」

という文は、「今日は医者へ行く」とも、「今日歯医者へ行く」とも取れる。しかし、いずれも文法上も意味上も正しいので、従来の文法的な方法⁽¹⁾⁽²⁾を用いて選び出すことはできない。そこで、打鍵以外の付加的な情報を用いて、利用者の意図に合った文節分割を行うユーザインタフェースが望まれる。

一方、文章をオンラインで入力する場合、利用者の思考のために、文節などの意味のある場所で、キー入力が停滯すると考えられる。このことに着目すると、キー入力の時間間隔（打鍵間隔）により区切りを識別するユーザインタフェースを考えることができる。たとえば、さきの例で、「きょう」と「は」、「は」と「いしゃへいく」の打鍵間隔を比較して、長い方を文節区切りとするような方式である。

本報告では、日本語ワードプロセッサ利用者の打鍵時間データから、文章の文節区切りで入力が停滯することを示しその時間を測定した。また、打鍵間隔を利用して、文節分割の精度を改善するユーザインタフェースについて検討した。

思考時間の測定

打鍵の測定実験

思考時間の測定のために、次のような実験で得られた打鍵時間データを用いた。

被験者 女性2名。いずれも英文タイプは経験があるが日本語ワードプロセッサは未経験。

実験期間 8日間。1セッション3時間の測定。

打鍵内容 用意された原稿の複写または産出による作文（平均600文字）。

入力方式 ローマ字入力。

測定方法 キーボード割り込み時に打鍵と時間を記録する打鍵データ収集プログラム⁽³⁾を利用。時間測定の単位は10ms。

実験で得られた原データから、句読点または変換キーで区切られた複数文節の打鍵列を抜き出して、以下の分析を行った。

文節区切りでの打鍵の停滞

図1は、被験者1が複数文節からなる文を入力したときの打鍵間隔の一例を示す。

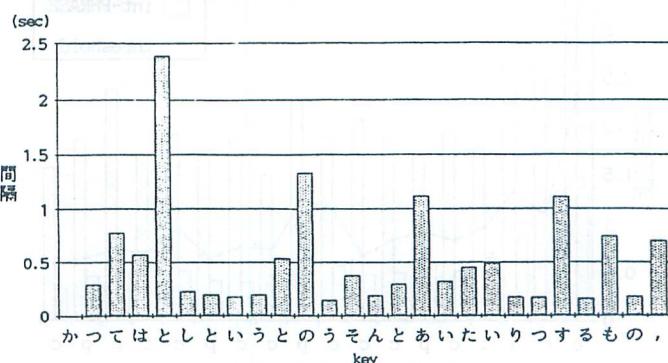


図1: 複数文節の打鍵の例

これは、

「かつては都市」と「農村」と相対立するもの、」

という文を入力したときの打鍵間隔で、文節区切りの場所で打鍵間隔が大きくなっている。

図2は、被験者1が640字の原稿を複写して入力したときの文節区切り(int-PHRASE)と、それ以外の文字の間(int-LETTER)での打鍵間隔をヒストグラムにしたものである。

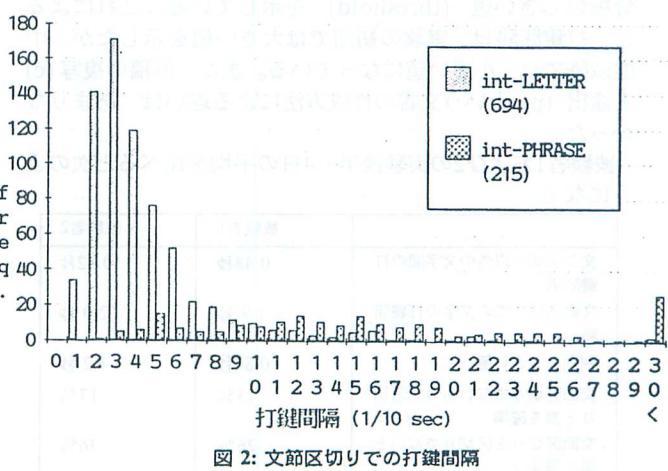


図2: 文節区切りでの打鍵間隔

文節区切り以外の文字間の打鍵間隔は、平均 0.45秒、標準偏差 0.35秒であったのにたいして、文節区切りでの文字の間での打鍵間隔は、平均 1.7秒、標準偏差 1.4秒であった。それぞれの分布をガウス分布と仮定すると、判別分析によりある打鍵間隔の文字が文節区切りの文字であるか否かを判別することができる。図 2 の場合、判別のしきい値は 0.71秒であり、また、このしきい値より長い打鍵を文節区切り、短い打鍵を文節区切りではないと判別したとき、文節区切りでないものを区切りと誤る確率は 11%、文節区切りを区切りでないと誤る確率は 18% である。

次に、文節区切りでの打鍵の停滞の様子を他の文書の打鍵時間データについて調べた結果を図 3 に示す。

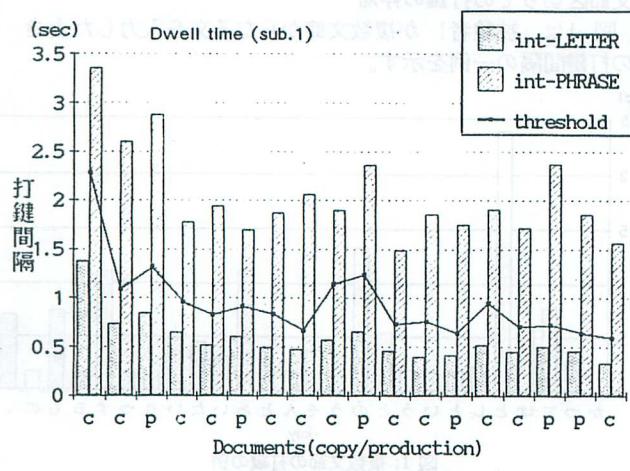


図 3: 打鍵間隔 (被験者 1)

図 3 は、被験者 1 が実験期間の 2 日目以降に作成した 18 の文書の打鍵時間の様子である。横軸には、文書を作成した順番にならべ、作成方法により、原稿の複写 (c) か産出 (p) のマークをつけた。グラフは、文節区切り以外の文字間の打鍵間隔の平均 (inter-LETTER) と、文節区切りでの文字の間での打鍵間隔の平均 (inter-PHRASE) および判別分析のしきい値 (threshold) を示している。これによると、打鍵間隔は、実験の初期では大きい値を示したが、中盤以降では、小さい値になっている。また、原稿の複写 (c) と産出 (p) という文書の作成方法による違いは、あまりなかった。

被験者 1 および 2 の実験後半の 4 日の平均を比べると次のようになる。

	被験者 1	被験者 2
文節区切り以外の文字間の打鍵間隔	0.48秒	0.82秒
文節区切りでの文字の打鍵間隔	1.9 秒	2.0 秒
判別のしきい値	0.80秒	1.2 秒
文節区切りでないものを区切りと誤る確率	13%	17%
文節区切りを区切りでないと誤る確率	26%	36%

被験者 2 は、被験者 1 より打鍵間隔が大きい。個人差は、文書の内容や作成方法の差よりも、大きいようであった。

文節分割への応用と問題点

打鍵間隔から判別する方式

前節の分析から、利用者の打鍵間隔から統計的な手段を用いて、20% 未満の誤判別確率で文節区切りでない文字を判別できることが分かった。一方、文節区切り打鍵の誤判別確率は、20-40% であった。以上のことから、利用者の打鍵は、文法上の文節区切りの場所で、いつも停滞するわけではないが、停滞するとすれば、文節区切りの場所であると結論づけることができる。そこで打鍵間隔から、統計的な手段を用いて文節区切りを判別するユーザインターフェースの実現は可能であるといえる。しかし、文節区切り文字を判別する打鍵間隔のしきい値には個人差や習熟度による差があるので、しきい値の決定にはこれらを考慮する必要がある。

文法的手段と併用する方式

複数の文節からなる文を文節に分割する方法として、二文節最長一致法⁽¹⁾、文節数最小法⁽²⁾などの文法的手段をもとにした方法が実用化されている。文節区切りは、文法的に正しいと考えられる候補の中から、文節長や文節数の値を評価して選ばれる。このような文法的手段は、文字列情報だけをもとに処理されていて、オンラインで日本語入力するシステムで得られる打鍵間隔の情報が無駄になっている。そこで、文節区切りで打鍵が停滞する性質を候補の選択に利用すれば、より利用者の意図に合った文節分割が実現できるであろう。

今後の方針

今後、従来の文法的な方式に基づいたかな漢字変換方式に、打鍵間隔の情報を取り入れたかな漢字変換の試作と、その評価を予定している。

[謝辞] 打鍵時間データを提供して頂いた当研究所の原田氏、甲氏、加藤氏に感謝します。

文献

1. 牧野寛、木澤誠: べた書き文の分かち書きと仮名漢字変換. 情報処理学会論文誌 20, No.4 (1979) pp. 337-345.
2. 吉村賢治、日高達、吉田将: 文節数最小法を用いた日本語文の形態素解析. 情報処理学会論文誌 24, No.1 (1983) pp. 40-46.
3. 粕川正充、木村泉: パーソナルコンピュータ用打鍵データ収集プログラムとその応用. 情報処理学会第30回全国大会 3G-8 (1985) pp. 1645-1646.