

# 手書き漢字認識システムを用いた自律型漢字学習支援

Kanji Learning Supports for Learner Autonomy Using Handwriting Recognition System

岩手大学工学部

三輪 譲二

miwa@cis.iwate-u.ac.jp

キーワード：日本語教育、漢字学習、自律学習、連合学習、筆順

## 1. まえがき

従来の漢字教育では、漢和辞典などを用いて、画数や部首を用いて検索する方法が指導されてきた。この方法では、画数の数え方や、部首の種類などを理解する必要があり、あらかじめ、基礎的な漢字教育を受けている必要があり、非漢字圏の初心者にとっては、漢字学習の障壁となっていた。

本報告で提案する自律型漢字学習法では、Web型手書き漢字認識システムに、漢字の形を自動認識させることにより、漢字の意味や読みを連合的に検索表示することができ、初心者でも教師に頼らないで、いつでも、どこでも、手軽に、繰り返し、漢字を自律学習することができる。また、漢字の筆順誤りの場合は、その誤り結果が表示されるので、正しい、筆順情報を表示することにより、学習し直すことができ、教師がいない環境でも、学習できる特色がある。さらに、手書き入力した漢字は、データベースに自動的に記録されているため、後で、誤りの多い漢字を特に重点的に学習でき、自律学習の意欲を高めることができる。

本報告では、2008年12月24日からの無料公開運用実験により取得した約5万個の漢字データから、「げんき I」の教科書に掲載の144個の漢字について、漢字入門の学習者が陥りやすい筆順誤りの傾向解析の結果を考察し、効果的な漢字筆順教育と新しい自律型漢字学習のための指針について述べる。

## 2. 多段自律漢字学習

語学学習において、言語学習において、学習者オートノミーが着目されている。この学習者オートノミーの定義は、Holec(1981)によれば、「自分自身の学習を管理する能力」(the ability to take charge of one's own learning)とされている。また、Dickinson(1987)によれば、self-instructionに、semi-autonomyとautonomyの二つの段階があると言われている。このことから、本研究では、下図に示すように、自律学習のアドバイザーの多段階の関わり度合いに応じて、多段的な自律学習を支援する日本語学習支援システムを開発して、運用評価する研究を実施する。

自律学習 アドバイザー (自律度)	社会全般	チューター 友人	教師
学習資源	インターネット情報等		教科書情報等
多段自律学習	高段階自律学習	中段階自律学習	低段階自律学習
学習形態	生涯教育 ⇔		学校教育

図1 多段自律学習システム e-Manabix の概念図

本研究では、多段自律学習を、学びのエアロビクス、すなわち、好氣的自律学習から、文字表記で e-Manabix、音声表記で [e:manabiccs] (e-Learning using Multistage Autonomy Networks for Associative Brain and Intelligence in Cloud Computing Systems) と呼ぶ。

## 2. Web 版手書き漢字認識システム

### 2. 1 システム構成

いつでも、どこでも、だれにでも、手軽に、繰り返しユビキタス環境で手書き認識を利用して多段自律型漢字学習をすることを考慮して、Web 版の手書き認識システムを開発し、2008年12月24日から、地球規模の運用実験を開始した。本システムは、Ajax (Asynchronous JavaScript and XML) により、1画毎に認識結果が表示されるため、対話性を高め、学習者に気付きを産みやすい構成となっている。表1に手書き認識システムの仕様を示す。

表1 手書き漢字認識システムの仕様

認識対象漢字数	約 2,300 漢字 (新常用漢字 2,136 含む)
漢字認識エンジン	Web アプリケーション (JSP)
画データ転送	1画毎の非同期処理
認識処理時間	約 30 ミリ秒
最大認識候補漢字数	5 漢字
誤り表示	筆順誤り表示マーク有り

### 2. 2 手書き漢字認識のための点画規則

漢字教育において、点画の標準化と記述化が有用であり、これまで、様々な検討<sup>1,2)</sup>がされている。本研究では、2008年2月15日に制定された Unicode CJK Strokes の 36 個の点画の中から教育に重要な 15 個組の点画を選んだ。表1に、手書き漢字認識<sup>3,4)</sup>のための漢字点画規則分類表を示す。

この表を漢字学習者に示すことの意味は、図2の基本点画に示すように、組数1の場合 (ハネなどの特別な場合を除いて)、→↘↓↖の5個の点画は

存在するが、←↑∧の3個の点画は存在しないことを示している。この知識を持つことにより、手書き漢字認識結果を容易に得ることができ、漢字の検索もより容易にすることができる。このように、学習者が、認識システムに対して、協調的利用者(Co-operated Users)であることが望まれる。

また、漢字指導で用いられる、永字八法では、側(ソク、点)、勒(ロク、横画)、努(ド、縦画)、趯(テキ、はね)、策(サク、右上がりの横画)、掠(リヤク、左はらい)、啄(タク、短い左はらい)、磔(タク、右はらい)の八法があるが、手書き漢字認識では、表1の名称記号S, N, HZ, SZのよ

表2 手書き漢字認識のための漢字点画規則簡素化表

(注:点画の実際の形は、機種依存のため、Unicode表を参照してください。)

名称	分類名:概略形	点画	CJK Strokes	組数	漢字例
H	横:→	一	U+31D0	1	一二三上下
S	縦:↓	丨	U+31D1	1	十上下中
SG	縦鉤:↓	丨	U+31DA	2→1	小水求事打
WG	彎鉤: ↓	丨	U+31C1	3→1	家逐狐
P	左下斜め下:↙	ノ	U+31D2	1	人千毛行利
N	右下斜め下:↘	㇇	U+31CF	1	八人入大木天
D	点:↘	㇇	U+31D4	1	主永丸丹
XG	斜鉤: ↘	㇇	U+31C2	2→1	戊成我
T	右上斜め上:↗	㇆	U+31C0	1	次治地虫
HZ	横折:→↓	𠃍	U+31D5	2	口日四田品
SG	横折鉤:→↓	𠃍	U+31C6	3→-2	月用青羽包鳥
HG	フ:→↙	フ	U+31C7	2	又今努
PZ	ム:↙→	ム	U+31DC	2	公広弘
PD	く:↙↘	く	U+31DB	2	女巡災
ST	レ:→↗	レ	U+31D9	2	良食民恨狼
SZ	縦折:↓→	𠃎	U+31C4	2	山亡区医断
SWG	縦折鉤:↓→	𠃎	U+31DF	3→2	己乱礼
BXG	扁斜鉤:↓→	𠃎	U+31C3	3→2	心必泌忘
SZZ	縦折折:↓→↓	𠃏	U+31DE	3	吳
SZWG	縦折折鉤:↓→↓	𠃏	U+31DE	4→3	弓号強
HZW	横折折:→↓→	𠃐	U+31CD	3	投
HZWG	横折折鉤:→↓→	𠃐	U+31C8	4→3	九几風氣
HXWG	横斜彎鉤:→↙→	乙	U+31E0	4→3	乙乞
HZZP	横折折折:→↙→↙	及	U+31CB	4	及建
HZZZG	横折折折鉤:→↙→↙	乃	U+31E1	5→4	乃

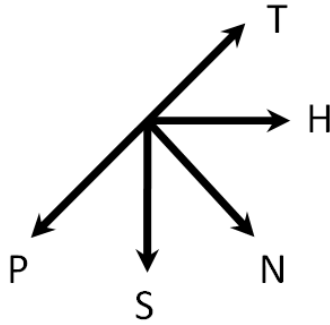


図2 基本点画とその名称

に、ハネなどの有無は考慮しない。この点は、国語教育における従来の漢字教育と、非常に大きな相違がある。しかし、外国人のための日本語教育における漢字教育では、漢字が弁別でき、容易に漢字の検索ができれば良いので、習得が容易になり、また、手書き漢字認識を用いた漢字検索も容易にすることができる。逆に、このことは情報損失したことになり、例えば、手篇と獣篇は方向が違うだけで、形は同一になるという欠点となるので、注意が必要である。

漢字認識のための方向辞書では、例えば、漢字「二」は「H H」、「十」は「H S」、「木」は「H S P N」、「六」は「S H P N」、「女」は「P D P H」となる。実際の漢字認識では、方向辞書のほかに、交差するクロスの有無などの構造辞書を利用する。

### 2. 3 点画に関する注意点

点画における注意点の第一として、日本語や繁体字の「過」の3-4画は、「H-S」の2画であるが、簡体字の「過」は、HSの1画となり、筆順の順序を変えることで、このように減画でき、筆順と連続卓画による経済則があることを理解することができる。

また、2番目としてさらに、サンズイや手篇の3画目はTであるが、Pと誤る場合が多い。このような場合、3画目から4画目の間に、空画(Null Stroke)があると考えれば、経済原則に則していることが理解でき、正しい筆順指導の指針として利用することができる。

さらに、教育フォントと明朝フォントに関する注意点として、以下について特に注意させる必要がある。

1. 折り方に関する例 (衣、去、玄)
2. 点画の組合せ方に関する例 (人、家、北)
3. 「筆押さえ」等に関する例 (入、八、芝、史)
4. 曲直に関する例 (子、手、了)
5. その他 (心)

## 2. 4 漢字認識方法

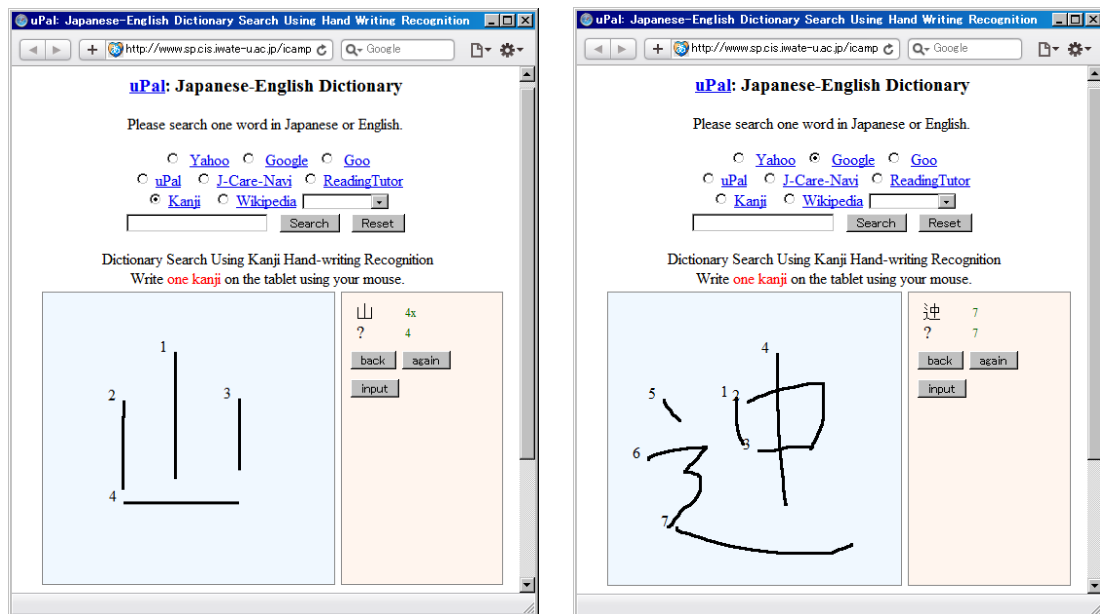
漢字の認識は、構造辞書で予備選択を行ったのち、方向辞書で距離を計算し、その値が小さい順に最大 5 個の候補が表示される。なお、距離差が一定値以下の場合だけ表示するので、候補数が 5 個に達しない場合がある。

例えば、漢字「六」と「木」の方向辞書は、それぞれ、「S H P N」、「H S P N」となる。これより、この 2 つだけの漢字認識の場合、筆順の誤りが無ければ、方向特徴だけで漢字を正しく認識できる。しかし、筆順を誤った場合は、同一になる場合が生じ、正しく認識することができない。このため、構造辞書を利用して、H と S が交差していなければ「六」、交差していれば「木」の漢字が予備選択される。

## 3. 手書き漢字認識を用いた自律学習

### 3. 1 単語の検索利用

手書き漢字認識を用いた辞書検索では、非漢字圏の学習者が部首や画数が不明でも、辞書を簡単に引けるという特色を持っている。図 3 は、漢字認識の表示例であり、図 3(a)の例では、筆順誤りがあっても正しく認識されており、誤筆順マーク X が表示されており、学習者に気付きを生まれやすくする仕組みが組み込まれている。また、図 3(b)の例では、「込」を認識させ、漢字検索をすることにより「とても」という意味を見つけ出すことができる。



(a) 筆順誤りの正認識例

(b) 未知漢字の検索例

図 3 漢字検索の例 (山、込)

### 3. 2 筆記宿題への利用

漢字ノートへの筆記の宿題は、漢字教育で良く用いられている方法である。しかし、この漢字ノートでは、画の採点はできても、筆順の採点を行

うことは困難な場合が多い。このため、本システムにログインして、漢字筆記の宿題を行えば、システムが、自動的に採点してくれ、また、誤った場合には、どの画の筆記を誤ったかを、後でチェックすることが可能である。図4は、漢字筆記の宿題の実施例であり、何個の漢字の筆記をしたかは一目良善である量善とする良善である。また、難問の「千」の漢字が赤字で表示されてとなっていることから、この漢字を筆記誤りを犯したことが即座に確認することがわかる。この漢字をクリックすると、図5のストロークが表示されるため、1画目の画の方向を誤ったことを確認することができる。

The screenshot shows a web browser window displaying the 'Scores for Kanji Handwriting Recognition' page. The page includes a search bar and a table with the following data:

No	Kanji	Recognition	No. of Strokes	No. of Rec.	Rank	Uncomplete	Time (JST)
1	三	二	2	1	1	0	08/29 19:33
2	三	三	3	1	1	0	08/29 19:33
3	四	四	5	1	1	0	08/29 19:34
4	五	五	4	1	1	1	08/29 19:34
5	六	六	4	1	1	1	08/29 19:34
6	七	七	2	1	1	0	08/29 19:34
7	八	八	2	1	1	0	08/29 19:35
8	九	九	2	1	1	1	08/29 19:35
9	十	十	2	1	1	0	08/29 19:36
10	百	百	6	1	1	0	08/29 19:38
11	千	千	3	1	1	0	08/29 19:38
12	万	万	3	2	1	0	08/29 19:38
13	円	円	4	1	1	0	08/29 19:38
14	時	時	10	1	1	0	08/29 19:39
15	?	千	3	1	0	0	08/29 19:40

Below the table, there is a detailed view of the character '千' (thousand). The character is shown with three numbered strokes: 1 (top horizontal), 2 (middle horizontal), and 3 (vertical). The character is displayed in red in the original image, indicating an error. The interface includes 'Clear' and 'Search' buttons and a 'Close Window' link.

図4 漢字認識履歴の表示例

#### 4. システム評価

本システムは、2008年12月24日から地球規模の運用実験を開始した。開始から約18ヶ月間の月毎の漢字入力数を、図5に示す。1月と2月や8

月と9月の利用が多いことから、学期初めによく利用されていることが分かる。しかし、学期末になるに従って、利用数が減少していることから、より使い易くするなど改善が必要である。

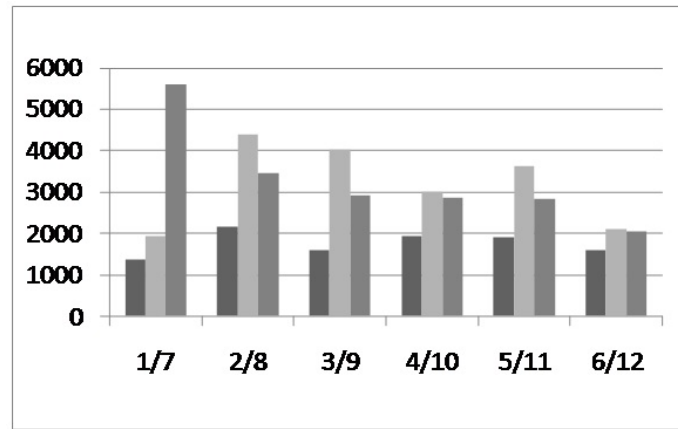


図5 手書き漢字認識システム月別利用数(09/1-6, 09/7-12, 10/1-6)  
約5万漢字のなかで、2画以上異なるものが30.2%あった。残りの69.2%のうち、正しく認識された割合は92.7%であり、誤りは7.3%であった。また、表3から、第1位に認識された漢字は、88.3%の正解率があることが分かる。

表3 全漢字の認識率(18カ月間、約35,000漢字)

	All	1	2	3	4	5
All	92.7%	69.6%	14.3%	5.0%	2.1%	1.7%
1	88.3%	69.6%	12.5%	3.7%	1.5%	1.0%

表4にげんきIの漢字の筆順誤りの多い順に示す。最も筆順誤りの生じた漢字は「右」であり86%であった。また、最も誤りの少ない漢字は、「員少雪読読」であり、0%であった。

表4 漢字筆順誤り率(18カ月間、げんきI、143漢字(一除く))

誤率	漢字
89-80%	右近画
79-70%	紙有道北所連市
69-60%	国働母何火好出院男入年
59-50%	神住青町来度学曜每思女色旅
49-40%	水田立子万別家山歌病生上私売赤
39-30%	七千聞昔円金西長友手十校文書新食
29-20%	四半仕午作車社早六牛持知土左雨気語時次中電木
19-10%	九先百話五朝買外起強古口大本飲正会今言小後三父日月使八
9-0%	勉帰高東映間京川夜下二白分名休行事自前天南人明見元員少雪読

## 5. むすび

本報告で提案する自律型漢字学習法では、Web 型手書き漢字認識システムに、漢字の形を自動認識させることにより、漢字の意味や読みを連合的に検索表示することができ、初心者でも教師に頼らないで、いつでも、どこでも、手軽に、繰り返し、漢字を自律学習することができた。また、漢字の筆順誤りの場合は、その誤り結果が表示されるので、正しい、筆順情報を表示することにより、学習し直すことができた。さらに、手書き入力した漢字は、データベースに自動的に記録されているため、後で、誤りの多い漢字を特に重点的に学習でき、自律学習の意欲を高めることもできた。

なお、Web 版手書き認識システムや連合型漢字学習システムは、以下の URL で無料公開中である。

<http://www.sp.cis.iwate-u.ac.jp/icampus/u/udic.jsp>

## 謝辞

本研究の一部は、平成 22-24 年度科学研究費補助金・基盤研究 (B), 22320091, 「デジタル世代の多段自律ネットワーク型日本語学習支援システムの地球規模の運用評価」によった。

## 参考文献

- 1) 外田、龍岡、前田、齋木、迎、押木: 中学生を対象とした学年別漢字配当表所収全字種の筆順調査結果のパターン別分析、書写書道教育研究、第 17 号, pp.15-24, 2003
  - 2) ヴォロビヨワ・ガリーナ/Vorobeva, Galina: 「非漢字圏の漢字教育の効率化を目指す漢字のスタンダード化について/About standardization of kanji aimed to increase efficiency of kanji teaching in the non-kanji using countries, 第 14 回ヨーロッパ日本語教育シンポジウム, ベルリン (Sep. 4, 2009).
  - 3) 三輪譲二: ユビキタス環境における漢字教育支援システムの評価 /An Evaluation of a Computer-Assisted Kanji Learning System in a Ubiquitous Environment, 電子情報通信学会, 教育工学研究会 (ET) 技術報告, Vol. 108, No. 470, ET2008-127, pp.199-204, 香川大学 (Mar. 7, 2009).
  - 4) 三輪譲二: A Pedagogical Method of Kanji Writing for Independent Learning Using a Web-based Handwriting Recognition System/ Web 版手書き認識システムを用いた独習型漢字書字教授法, ヨーロッパ日本語教師会, 第 14 回ヨーロッパ日本語教育シンポジウム, ベルリン (Sep. 4, 2009).
-