

外国人にやさしい日本語辞書のみだしの表記と配列

東京外国語大学留学生日本語教育センター 土屋順一
tsuchiya@jlc.tufs.ac.jp

1 はじめに

ポーランドのことをポーランド語では Polski とかき、その発音はつづりどおりに英語の知識で発音すればだいたいちかいのだが、ワレサ元大統領の名前は Wałęsa とかき、その発音は、[vawensa] にちかい。このように、ポーランド語では、半母音化した l とそうでない l が別の音韻として区別され、別の文字でかかれるが、ブルガリア語では、人によって、また音韻環境によって、l が半母音化して発音される。ブルガリア人のすきな луканка [lukanka] (ソーセージ) が、[wukanka] のように発音されるのをよくきいた。また、ブルガリア人の友人と英語ではなしていたとき、「夏やすみに wife helper をする」というので、家政婦のようなことをするのかとおもって、よくきいてみたら life helper (saver) だった、ということがあった。

ところで、ブルガリアにすんで数か月たったある日、数字の「0」(ゼロ)を辞書でしらべようとして、нува [nuva] のところをみたらでていない。「あれ？」とおもって нуля [nula] のところをみたらでていた。その時点では、ブルガリア語の l が半母音化して発音されることがある、ということを経験的にしっていたにもかかわらず、なぜ私はまちがえてしまったのだろうか。その理由を内省してみると、つぎのように分析される。

- 1) ブルガリア語のききとり能力がまだ十分でない。l が半母音化して発音されるといっても、摩擦音であるブルガリア語の w とはちがうはずで、そのちがいをききわけられなかった。
- 2) 数字の「0」は、電話番号をつたえるときなどにきく機会、はなす機会がおおい単語だが、いつも「0」とかかれるので、そのつづりを目にすることはほとんどない。
- 3) ヨーロッパ語の知識がとぼしい。英語の nil、ロシア語の ноль [nol]、ラテン語の nihil などの知識があれば、たとえ l の音がききとれなくても、数字の「0」のつづりが l をもっているということを知ることができるが、それができなかった。

この状況は、漢字のよみ方がなかなかおぼえられない外国人の状況によくにている。それは、たとえば、日本人による発話の長母音と短母音の発音の区別がはっきりしていない、という現象とともに、外国人側につきのような理由があることがかんがえられる。

- 1) 日本語のききとり能力がまだ十分でない。特に母語に長母音と短母音の区別がない者は、ききわけがむずかしい。
- 2) 漢字語彙はいつも漢字でかかれるので、そのつづりを目にすることはほとんどない。
- 3) ほかの語彙や、おなじつくりをもつ漢字の知識がとぼしいので、つづりを類推することができない。たとえば、「揚」のよみ方が「ヨ」か「ヨー」かわからなくても、「湯」「陽」「場」などのうちどれかをしっかりおぼえていれば、長母音であることを類推できるが、それができない。

アラビア数字は10個しかないからいいが、日本語の表記システムは、1945個の数字をつかっているようなものなので、外国人が日本語の辞書のみだしをさがすために時間を無駄にすることも、私の体験とくらべて、ずいぶんおおいはずである。

2 トルコ語研究者がひきやすいトルコ語の辞書の配列

ふるいトルコ語を研究する者の間では有名な "An Etymological Dictionary of Pre-Thirteenth-Century Turkish" という辞書がある。この辞書に収録された語彙は、歴史的・地理的におおきくひろがった原典からとられている上、その原典がアラビア文字やウイグル文字でかかれているため、音韻が特定しにくい。たとえば、アラビア文字、ウイグル文字では、円唇母音と平唇母音は区別して表記されるが、それ以上は区別されず、a と e、o と u はおなじ表記になる。また、アラビア文字では、

母音が表記されないこともおおい。その場合、転写のしかたによって、もともと同根の語がまったく別のみだし語としてあつかわれる可能性がある。それをさけるために、母音は前舌・後舌にわけ、それ以上は区別しない、破裂音・破擦音の有声・無声を区別しない、音節数ごとにわけなどの特殊なみだし配列規則にしたがっている。たとえば、「BRG二音節・後舌母音」というみだし語群のところには、barak, barig, burig, burki, birkig, burkig, barkin, burkan, bo:rgu:y というみだし語がならんでいるのである。

3 外国人による日本語キーボードの誤入力

筆者が外国人による日本語キーボード入力の過程を分析した結果、つぎのような傾向があきらかになった(土屋 2000)。

- 1) 外国人は、日本語の母音の長短の区別、促音の有無の区別、破裂音の有声・無声の区別など、調音のタイミングに関する音韻事項を誤記憶することがおおい。その一方で、母音、摩擦音・破擦音など調音法に関する音韻事項は比較的正確に記憶されている。
- 2) 音韻の誤記憶は在日期间がながくなくても改善されない。
- 3) 日常的によくつかうような語彙の音韻も正確に記憶されていない場合がおおい。
- 4) 和語よりも漢語の誤記憶がおおい。

たとえば、「授業」という語を日本語でキーボード入力しようとした外国人(平均在日歴3年の留学生)141名のうち62名がただしく入力できなかった(表1)。

「授業」のように、日常的につかっている語彙の正確な音韻でさえ、これだけあやふやに記憶されているのだから、ききとって意味がわからないことばの意味を辞書でしらべる、という場合には、何度も何度も辞書をひきなおす、という作業のために相当の時間が浪費されているはずである。

留学生の授業でディスカッションをした際に、提題者の学生が「『同性愛は罪悪か』ということについて、はなしあいたいとおもいます」といったところ、学生たちはすぐに電子辞書をひきはじめたのだが、そのうち何人かは「座薬」のところをひいていた、ということがあった。

表1：基礎語彙の誤入力率

入力語彙	誤入力者数	入力者数	誤入力率	入力語彙	誤入力者数	入力者数	誤入力率
北海道	15	48	31%	最初	12	34	35%
東京	31	149	21%	留学	16	63	25%
京都	26	81	32%	生活	11	56	20%
神戸	15	27	56%	授業	62	141	44%
九州	6	31	19%	文法	11	33	33%

4 外国人がひきやすい日本語の辞書の配列

キーボード入力過程調査であきらかになった外国人がまちがえやすい音韻事項にもとづいて、"An Etymological Dictionary of Pre-Thirteenth-Century Turkish"の形式の日本語辞書をかりにつくるとすると、その配列規則はつぎのようになる。

- 1) 収録語彙は漢語だけとする。
ex. 「受領」は収録するが、「受け取り」は収録しない。
- 2) 長母音と短母音を区別しない。
ex. 「保守」と「報酬」はおなじみだし語群にはいる。
- 3) 促音の有無を区別しない。
ex. 「主張」と「出張」はおなじみだし語群にはいる。

4) 破裂音の有声・無声を区別しない。

ex. 「退学」と「大学」はおなじみだし語群にはいる。

5) 摩擦音の有声・無声を区別する。

ex. 「賞金」と「常勤」は別のみだし語群にはいる。

6) 母音を区別する。

ex. 「鼓動」と「駆動」は別のみだし語群にはいる。

7) 漢字数を区別する。

ex. 「核」と「学」はおなじみだし語群にはいるが、「家具」と「学区」はそれとは別のおなじみだし語群にはいる。

三省堂「例解新国語辞典」の収録語の一部を、この規則にもとづいてならべかえると、つぎのようになる。(2)は漢字二字であることをあらわす。

・ KOKO(2)

個個 (KOKO)	後光 (GOKOO)	豪語 (GOOGO)	皓皓 (KOOKOO)
古語 (KOGO)	国庫 (KOKKO)	口腔 (KOOKOO)	煌々 (KOOKOO)
午後 (GOGO)	国交 (KOKKOO)	後攻 (KOOKOO)	皇后 (KOOGOO)
虎口 (KOKOO)	公庫 (KOOKO)	後項 (KOOKOO)	囂々 (GOOGOO)
枯槁 (KOKOO)	江湖 (KOOKO)	高校 (KOOKOO)	轟々 (GOOGOO)
孤高 (KOKOO)	口語 (KOOGO)	航行 (KOOKOO)	
古豪 (KOGOO)	交互 (KOOGO)	孝行 (KOOKOO)	

・ KOKOKAKU(3)

考古学 (KOOKOGAKU)

・ KOKOKU(2)

故国 (KOKOKU)	刻刻 (KOKKOKU)	抗告 (KOOKOKU)
五穀 (GOKOKU)	広告 (KOOKOKU)	興国 (KOOKOKU)

・ KOKOSE(3)

光合成 (KOOGOOSEE)

・ KOKOTAI(3)

口語体 (KOOGOTAI)

・ KOKOTU(2)

硬骨 (KOOKOTU) 恍惚 (KOOKOTU)

・ KOKOBUN(3)

口語文 (KOOGOBUN)

・ KOKOMOKU(3)

子項目 (KOOKOOMOKU)

5 おわりに

日本人が辞書をひくのは、ただしいひらがなのつづりと、ただしい意味がわかっているけれど、ただしい漢字がわからない、という場合がおおい。たとえば「スイセンジョー」の「セン」はどういう字だったかなあ、という時である。でも、最近はワープロをつかう人がおおくなって、このようなつかい方はへっていこう。

一方、外国人が日本語の辞書をひくのは、あやふやなひらがなのつづりだけがわかっていて(場合によっては漢字もわかっていて)、ただしい意味がわからない、という場合がおおい。意味をしりたいから辞書をひく、というのは、辞書本来のつかい方である。意味をしりたいのに、なかなかただしいみだし語にたどりつくことができない外国人に必要とされるのは、はやく辞書をひけるように、ただしいつづりをおぼえる訓練をすることではなく、ただしいつづりをしらなくても、はやくひけるよ

うな辞書をつくることなのである。

この研究は、科学研究費補助金基盤研究(C)(2)「外国人学習者の日本語作文キーボード入力過程の分析とデータベースの作成」(平成12~14年度 代表者:土屋順一 課題番号:12680300)の成果の一部である。

参考文献

Clauson, G. "An Etymological Dictionary of Pre-Thirteenth-Century Turkish"
Oxford University Press, 1972

土屋順一:『外国人学習者の日本語ワープロ誤入力の分析と外国人用漢字変換辞書の開発』

文部省科学研究費補助金基盤研究(C)(2)研究成果報告書 課題番号09680296、2000

土屋順一・土屋千尋・杉田幸代:「外国人による日本語キーボード入力を支援する漢字変換辞書」
『2000年度日本語教育学会秋季大会予稿集』pp.265-266、2000