

辞書編纂で最も厄介な仕事のひとつが誤登録の判断と処理です。

複数の目で見ることが極めて重要です。[フォーム](#)での誤登録の報告と議論に是非ご協力下さい。

## links

- ・ [辞書登録・削除希望単語入力フォームの、誤登録削除希望一覧](#)
- ・ [SKK-JISYO.wrong.annotated](#)
- ・ [SKK-JISYO.noregist](#)
- ・ [採録基準](#)

## SKK-JISYO.wrong.annotated と SKK-JISYO.noregist

簡潔化のため、以下では SKK-JISYO.wrong.annotated を「wrong」、SKK-JISYO.noregist を「noregist」と呼びます。

- ・ wrong には、明白に「誤登録である」と判断された語が、
- ・ noregist には、誤登録ではないが L 辞書にはない方が良いと判断された語が

それぞれ収録されています。

wrong はユーザ辞書から引くことも視野に入れていますが、noregist はそうではありません。

```
mv .skk-jisyo .skk-jisyo.$$; skkdic-expr2 .skk-jisyo.$$ - SKK-JISYO.wrong.annotated > .skk-jisyo;
rm .skk-jisyo.$$
```

## 原則論

SKK と L 辞書は、日本語を入力するための道具です。

日本語を入力するための道具に必要なことは何か？使用者が入力したいと思う文字を出力することです。

L 辞書は、入力用の辞書であって国語辞典ではありません。求められている性能も、国語辞典とは自ずから異なってくるわけです。

さて、SKK の場合、使用者は「読み」と「送り位置」を SKK に伝えることで、意図した文字を受け取れることを期待します。この時に必要なことは、使用者が期待した文字が得られることであって、入力や出力が辞書的や文法的に正しいことではありません。ですから、世間的にもしくは辞書的に誤用と見做されていることは、それだけでは削除の理由にはならないと考えられます。

## 各論

逆に、SKK で一番あってはならないのは、「入力者の意図とは違う文字が出力されてしまう」ペアであると言えます。

単純な誤登録 = 見出し不一致

とくがわみつくに / 水戸光圈 /  
すきかって n / 好き勝手に /

これは「徳川光圈」が欲しいのに「水戸光圈」が、「好き勝手に」が欲しいのに「好き勝手にに」が出力されてしまうので誤登録として削除。

「とくがわみつくに / 水戸光圈 /」は、十分な annotation が添えられていれば連想辞書のエントリとしては可能かもしれませんが、しかし、通常の変換を望む L 辞書の実用者にとっては困るだけです。

## 送り位置

むこ u / 向 / 向こ ; 送り位置補正 /

「向う」は非正則とされてはいますが、使われている送り方ですし、誤用というわけでもありません。

「むこ u / 向 /」の存在は、"MuKou" で「向こう」を入力したい人の妨げになることはありません。送り位置を使用者が決定するのは SKK の最も大きな特徴であり特長であるので、こうしたペアを非正則というだけの理由で排除すべきではないと思われます。

"MukoU" と打鍵して「向う」が出るのは、その人がそう送ったのです。思うに、手書きしていたとしてもそう送っていたのではないのでしょうか？

「送り位置補正」の候補については送りありエントリの送り仮名を参照。

にが s / 苦 / 逃 ; 非正則 ( に g ) /

競合がある場合、またはあまり一般的でない場合には、annotation が添えられている場合もあります。

はず k / 辱 /

これは「辱める」が現代的な表記ですが、「辱かしめる」という表記も非正則であるというだけで「誤用」ではありません。ただ、一般的とは言えない上、「はず k / 恥 /」とも競合するので、L 辞書に入れておく必要はないだろうと考え (wrong ではなく) noregist に移しています。

- ・ 内閣告示 (S56) 送り仮名の付け方 : 「非正則」とは、これに則さない送り方を言うようです。
- ・ 青空文庫 : 「非正則」な送り方の方がつい最近まで本流だった場合も実に多いのです。
- ・ 送りありエントリの送り仮名

## 仮名遣い・読みの揺らぎ

とう r / 通 ; ; 「とお - る」が正しい /

これは現代仮名遣いでは「とお・る」ですが、発音の実態としては「とおる」と「とうる」に区

別はありません。"TouRu" と入力された時に求められているのが「通る」であることは明らかですし、このペアの存在によって他の語の入力が妨げられるわけでもありませんから、国語辞典は誤表記としていても L 辞書には残しています。

ただし、

まかりとう r / 罷り通 /

のような派生語まで入れていったのでは際限がないので、こうした仮名遣いの揺らぎを補うペアはある程度基本的な語彙に限定されています。要はバランス感覚ですね。

DDSKK では、こうした仮名遣いの揺らぎを一手に引き受けてくれる優れたコード [skk-correct.el](#) も利用可能です。ただし、DDSKK でしか使えませんし、逆に予想外のものを入力してしまう場合も出て来るので、これをもって辞書側での対応を不要とできるものではないと思われます。

類例：

こうり / 小売 / 小売り / 高利 / 公理 / 行李 / 功利 / 公吏 / 氷； こおり / 氷； こおり /  
はなじ / 鼻血 / 衄 / 衄 /  
はなぢ / 鼻血 /  
かせんしき / 河川敷 /  
かせんじき / 河川敷 /  
じゅっとく / 拾得 / 十徳 /  
じっとく / 拾得 / 十徳 /

・ [内閣訓令 \(S61\) 「現代仮名遣い」の実施について](#)

誤用？新語？

けんけんがくがく / 喧喧譁譁； 喧喧囂囂 (けんけんごうごう) あるいは侃侃諤諤 (かんかんがくがく) の間違い /

このペアの場合、一般的には「喧喧譁譁」という言葉は誤用であるとされていますが、「けんけんがくがく」と入力する使用者は「喧喧囂囂」や「侃侃諤諤」ではなく「喧喧譁譁」が欲しいはずなので、「喧喧譁譁」という表現がある程度行われていることも鑑み、そのまま残しています。[order essay](#) その上で、誤用と見做される場合があることを註釈として表示します。それを見て他の表現に改めるか、それともわかった上でそう入力するかは、辞書編纂者ではなく個々の使用者が決めることである、というのが基本的な立場です。

なお、これらのような「誤用と見做されるが残してある」タイプのペア (のみ) には、「 」を含む [annotation](#) が付与されています。そもそもこうした語は辞書にない方がいいとお考えの方は、この [annotation](#) を持つペアを事前に辞書から削除してしまうことをお勧めします。

```
sed 's@[^/]*;[^/ ]* [^/]*@@@g' SKK-JISYO.L | grep '^|/.*'
```

造語

造語の中でも、特に何か既存の語をもじり、敢えて元の語と似せて作られる種類のものは取り扱いに注意が必要です。

しんぱいしょう / 心配性 / 心配症 /

「心配症」は元は「心配性」の誤記のように思われますが、作品名として使われたとの指摘もあり、また「性質」じゃなくて「ピョーキ」なんだ、として使う人のあることも考えられなくはないので、noregist に移しています。 [write my essay](#)

がっしゅうこく / 合衆国 / 合州国 /

「合州国」は、"United States" のより忠実な訳語を志して作られた造語で、一部では使用されています。

ただ、定着しているとは呼びがたい上に、実際には不注意で「合衆国」と間違えて入れてしまう危険性の方が遥かに高いと考えられるので、これも noregist 扱いとなっています。

めいさく / 名作 / 迷作 /

「迷作」も「名作」から作られた造語で、掲載している国語辞典もあまりないようですが、これは既に市民権を得ていると考えられるので削除対象にはなりません。

最近のものでは

ばいしゅん / 売春 / 買春 ;[ 新語 ] 「売春」をかう側から見た語 /

のような例があります。「買春」などという語は認めないという人も少なくないでしょうが、候補の一つとしてこの語があることでそうした人が受けるデメリットと、ないことでこの語を使いたい人が受けるデメリットとを比較考量して、後者を大として採録しているわけです。

こうした語の線引きは、最終的には利用者個々の言語感覚に基く議論によって形成してゆく他はないでしょう。

誤登録？それとも単にマイナーな言葉？

さくし / 索師 /

という語に削除希望が出ています。確かにあまり見掛けない語です。

ひっしゅう / 必習 /

教科などについては普通「必修」と書くようです。しかし、「必ず習わねばならない」という意味で「必習」と書くことはあってもいいような気がします。これは誤登録なのでしょう。

特に漢語や複合語の場合、それが誤登録と呼べるのかの判定は容易ではありません。大辞典や漢籍を引っ繰り返せば用例があるのかもしれませんが、何かの専門分野では使われる言葉なのかも

しません。

### 何に基準を求めるか？

#### 権威づけ

主要な国語辞典に掲載されていれば恐らくは「正しい」のかもしれませんが、実際にはそれがどれほど大きな国語辞典であっても、使用される語彙の大半は辞典には載っていないものなのです。「載っているから正しい」とは言えても、「載っていないから正しくない」とは言えません。ですから、国語辞典には「削除しない」理由を求めることはできても、「削除する」理由を求めることは出来ない場合がほとんどです。

せいりつう / 生理痛 /  
たまいし / 玉石 /  
ぼくたち / 僕達 /

また、送り仮名や仮名遣いなどの問題については、国などの指針も絶対視されるべきものではありません。内閣訓令などは公的な機関では遵守しなければならないのですが、入力用の辞書を編集する時に見るべきなのは、そうした勧告などを受けて現に使われている (i.e. SKK 使用者が入力するであろう) 言語の実態であり、勧告そのものではないということです。

削除希望には出来るだけ詳細な理由が要請されているのは、主にこうした事情によります。

#### 正しくても載せない？

逆に、大きな辞書には載っているというだけでは、必ずしも収録の十分条件にはなりません。

こう / 公 / 項 / 考 / 孝 / 功 / 劫 / 甲 / 香 / 口 / 校 / 行 / 高 / 広 / 硬 / 厚 / 工 / 江 / 港 / 稿 / 後 / 好 / 光  
孔 / 膏 / 綱 / 鋼 / 講 / 侯 / 候 / 抗 / 航 / 斯 / 兮 / 国 / 府 / 向 / 巧 / 降 / 絳 / 工 / 酵 / 哮 / 吼 / 拘 / 恂 / 恂 / 恂  
鉤 / 鈎 / 鈎 / 宏 / 宏 / 法 / 紉 / 紉 / 肱 / 鉅 / 砒 / 砒 / 眩 / 紉 / 弘 / 交 / 効 / 絞 / 郊 / 鮫 / 傲 / 咬 / 效 / 狡 / 皎 / 皎 / 佼  
餃 / 鷄 / 耕 / 畊 / 耗 / 更 / 梗 / 巨 / 構 / 峴 / 荒 / 購 / 荒 / 竊 / 構 / 構 / 窺 / 窺 / 坑 / 坑 / 亢 / 亢 / 頑 / 詭 / 詭  
逅 / 腔 / 倅 / 控 / 控 / 恒 / 恆 / 恆 / 虹 / 扛 / 缸 / 缸 / 缸 / 缸 / 缸 / 鴻 / 桁 / 紉 / 紉 / 攻 / 攷 / 洪 / 哄 / 鬨 / 鬨 / 峴 / 峴 / 峴  
呻 / 狎 / 狎 / 開 / 浩 / 皓 / 皓 / 窖 / 詒 / 靠 / 皇 / 凰 / 惶 / 惶 / 惶 / 惶 / 惶 / 惶 / 惶 / 惶 / 惶 / 隕 / 隕 / 隕 / 隕 / 隕  
眺 / 光 / 統 / 眺 / 昂 / 仰 / 貢 / 禎  
濞 / 闔 / 黃 / 擴 / 廣 / 曠  
神 / 膏 / 昂 / 頁 / 河 / 衝 / 湊 /

のようなエントリで、"Kou" と入れて「尻」や「神」という字を出そうと考える人が果しているのでしょうか。競合の多いエントリでは、「出ないこと」も性能なのではないでしょうか。

・ cf. 「[Re: KOU/YOKO/](#)」@ML

こうした場合には、「正しくても」noregist に移すこともあって良いと思われれます。

#### 誤用か否かわからないけど消す場合

誤用であると断定は出来なくても、まず使われないのが明らかである場合 (誤入力の危険性が大きい場合は特に) は L 辞書から削除して構わないとは思いますが、それを wrong と noregist のどちらとして扱うかは難しい問題です。

wrong に「誤字・誤用でない」語を見つけたらご報告をお願いします。wrong 辞書も終着駅では決してありません。

## 半削除

疑義があるが削除の判断・処理を保留されている語 (‘?’) や、誤りと思われるものを含んではいるが有用性の認められる候補 (‘ ’) については、特定の annotation を付与した上で、エントリの最後部に移してあります。( annotation の項目もご参照ください)

ひょうひょう / 飄々 / 飄飄 / 瓢々 ;?/  
かいづか / 貝塚; 仮名遣い「かいづか」/

登録順を末尾にして警告をつけた候補が通常の入力の妨げになる可能性は低く、半分削除したようなものであろうといったところです。

これらの候補を「半削除」でなく、完全に取り除いて使用したい場合には、

```
% annotation-filter.rb -k -e ' |¥?$' SKK-JISYO.L | skkdic-expr2 > SKK-JISYO.L.customised
```

などとします。SKK-JISYO.L.unannotated 相当の辞書が必要な場合は、-k を外せば他の annotation も同時に除去します。