

誤りやすい同音異義語の収集

新納 浩幸

茨城大学 工学部 システム工学科

〒 316-8511 茨城県日立市中成沢町 4-12-1

Tel. 0294-38-5220, E-mail shinnou@lily.dse.ibaraki.ac.jp

本論文では、誤りやすい同音異義語に内在するヒューリスティクスを提案し、既存の辞典を用いて、そのヒューリスティクスの妥当性を示す。また、そのヒューリスティクスを利用して、既存の辞書から誤りやすい同音異義語を自動収集する。ここでは、通常同音異義語と片方が実在しない単語となるような同音異義語を収集する。後者の収集の評価として、1ヶ月分の新聞記事からそのような同音異義語の誤り検出を行い、収集した効果を確認した。読みを考慮した収集を行うことを今後の課題とする。

Collection of homophone apt to be mis-spelled carelessly

Hiroyuki Shinnou

Ibaraki University.

Dept. of Systems Engineering

4-12-1, Nakanarusawa-cho, Hitachi-shi, Ibaraki-ken 316-8511, Japan

Tel. 0294-38-5220, E-mail shinnou@lily.dse.ibaraki.ac.jp

In this paper we present a heuristics on homophone apt to be mis-spelled carelessly, and show the appropriateness of this heuristics by homophone dictionaries. Further, by using this heuristics, we automatically extract homophone apt to be mis-spelled carelessly from the dictionary. We collect two types of homophone. One is the general homophone, and another is the homophone which is a set with two elements. First element is a real word, and second element is the mis-spelled word for the first one. To evaluate the collection of the latter homophone, we detect mis-spelled homophone in the newspaper one month articles. As the result, we could confirm the effectiveness for the collection of homophone. In future, we will study the collection of homophone with the consideration of the pronunciation of Kanji characters.

1 はじめに

文書中の書き誤りを検出、修正する校正支援システムが有益であることは明らかである。そして、校正支援システムのひとつの機能として同音異義語の書き誤りの検出、修正がある。ここでは同音異義語の書き誤りの検出、修正を行う上で、まず最初に設定しておかなければならない、書き誤りやすい同音異義語の組の収集について述べる。

同音異義語を単純に「平仮名表記が同じだが、漢字表記が異なる単語の組」と定義した場合、その収集は一見容易そうに思える。なぜなら辞書は平仮名文字によってソートされており、同じ平仮名表記を持つ単語は並んで出現するので、それらを取り出せばよい。しかしそのような単純な収集では以下の2点で不都合がある。

1つは、同音異義語の書き誤りの検出、修正を行う上で、上記のような単純に収集された同音異義語の組では問題を無用に困難にする点である。例えば、「かいとう」という平仮名表記を持つ単語は少なくとも以下の14種類が存在する。

{ 垣内, 会党, 会頭, 回答, 灰陶, 快刀, 戒刀, 怪盗, 械闘, 開冬, 解党, 解凍, 解答, 解糖 }

ある文章中に「回答」という単語が現れた場合、その「回答」という単語は正しいか、あるいは他の13種類のどれであるかを判定することが、誤り検出、修正の基本的手順である。もちろん、この一般的なアプローチで十分だが、実際は14種類それぞれの表記になるという判定のための規則を作らなければならず、これは自動学習を考慮したとしても比較的困難な作業である。ひとつの対処方法として、同じ平仮名表記を持ったとしても、書き誤りやすいものだけを対象にすることが考えられる。この対処方法は単純だが、作成すべき規則が減る点、実際に誤りやすい単語についてだけ検出が行われるため、適合率が良くなる点などから効果的である。また修正を行わずに、誤りの可能性のある単語だけを指摘する単純な校正支援システムでは、誤りやすい同音異義語の組を収集しておくことは重要なポイントとなる。

2点目の不都合は、上記のように単純に収集した同音異義語の組では実在しない単語表記という最も単純な誤りに対処できないことである。例えば、「画竜点睛」という表記がある。これは「がりょうてんせい」という平仮名表記を持ち、「画竜点睛」の書き誤りである。「画竜点睛」と

いう単語は実在しないために、辞書見出しから単純に作成した同音異義語では、{画竜点睛, 画竜点睛}のような組を得ることができない。このように実在しない単語を想定した場合、同音異義語の数は莫大な数になるため、それらをすべて記憶しておくことは現実的ではない。ここでの対処も、1点目の不都合の対処と同じように、誤りやすい同音異義語の組だけを収集すれば良いだろう。

ここで、誤りやすい同音異義語を以下のように2つのタイプにわける。

タイプ1

文字列の集合

$$\{w_1, w_1, \dots, w_n\}$$

において、各 w_i は同じ平仮名表記を持ち、各 w_i は実在する単語である。

タイプ2

文字列の集合

$$\{w_1, w_2\}$$

において、 w_1, w_2 は同じ平仮名表記を持ち、 w_1 は実在する単語である。また、 w_2 は実在しない単語であり、 w_1 の書き間違いである。

本論文ではタイプ1、タイプ2の同音異義語の組を自動的に収集することについて述べる。概略、誤りやすい同音異義語に内在するヒューリスティクスを提案し、そのヒューリスティクスと同音異義語辞典及び国語辞書を用いて、自動収集を行う。収集したタイプ2の同音異義語を利用して、誤り検出、修正の実験を行い、その効果を確認した。

2 誤りやすい同音異義語のヒューリスティクス

まず本論文で提案する誤りやすい同音異義語に内在するヒューリスティクスを以下に示す。

H1 α と β が同音異義語 (同じ平仮名表記を持つ) であり、 α (正しい表記) は β (誤った表記) に書き誤りやすいとすれば、以下の性質がある。

h1-1 α は実在する単語である.

h1-2 α と β の文字列長は等しい.

h1-3 α と β は, ある位置の漢字 1 文字だけが異なっている.

今 H 1 を満たすような α と β を以下の式で表現することにする.

$$\alpha \rightarrow \beta$$

タイプ 2 の同音異義語に対しては, H 1 の条件がそのまま使える. つまり,

H 2 $\{\alpha, \beta\}$ がタイプ 2 の同音異義語であれば, $\alpha \rightarrow \beta$ が成立する.

例えば, タイプ 2 の同音異義語「がりょうてんせい」 α = 画竜点睛, β = 画竜点晴 を考える. まず α は実在する単語であり (h1-1), α と β はともに文字列長は 4 で等しく (h1-2), 4 文字目の文字“睛”と“晴”だけが, α と β の違いである (h1-3). よって,

画竜点睛 \rightarrow 画竜点晴

が成立し, H 2 が成立している.

タイプ 1 の同音異義語に対するヒューリスティックスを述べる前に, 次の式を定義する.

各 k ($= 0 \sim n - 1$) に対して,

$$\alpha_k \rightarrow \alpha_{k+1}$$

が成立するとき,

$$\alpha_0 \Rightarrow \alpha_n$$

と表記することにする.

H 3 $\{\alpha_0, \alpha_1, \dots, \alpha_n\}$ がタイプ 1 の同音異義語であれば, 任意の i と j に対して, $\alpha_i \Rightarrow \alpha_j$ が成立する.

例えば, タイプ 1 の同音異義語「かくりつ」 $\{\text{確率, 確立}\}$ の場合, 確率 \rightarrow 確立 および, 確立 \rightarrow 確率 が成立しているので, H 3 を満たすことは明らかである. また, タイプ 1 の同音異義語「しゅうけつ」 $\{\text{終決, 終結, 集結}\}$ の場合, 終決 \rightarrow 終結 \rightarrow 集結 および, 集結 \rightarrow 終結 \rightarrow 終決 が成立しているので, H 3 を満たすことは明らかである.

3 既存辞典からの確認

ここでは, H 3, H 2 がタイプ 1, タイプ 2 の同音異義語に対して成立していることを, 既存の辞典 A [1] と辞典 B [2] を用いて確認する.

辞典 A はタイプ 1 の同音異義語 1,195 組が収録されている. この中で H 3 を満たすものは, 1,099 組 (92.0%) であった. 付表 1 に H 3 を満たさなかったタイプ 1 の同音異義語 96 組を示す.

辞典 B の第 1 章「うっかり書き間違いやすい語」には, タイプ 2 の同音異義語 804 組が収録されている. この中で H 2 を満たすものは, 758 組 (94.3%) であった. 付表 2 に H 2 を満たさなかったタイプ 2 の同音異義語 47 組を示す.

辞典 A と辞典 B を利用することで, H 3, H 2 がタイプ 1, タイプ 2 の同音異義語に対しておおむね成立していることが確認できる.

4 自動収集

タイプ 1, タイプ 2 の同音異義語に対して, H 3, H 2 が成立していることを前章で確認した. ここでは, 逆に, H 3, H 2 を利用してタイプ 1, タイプ 2 の同音異義語を自動収集することを試みる.

4.1 タイプ 1 の同音異義語

まずタイプ 1 の同音異義語であるが, これは実在する単語だけが対象なので, 辞書¹から同音異義語を収集し, 各同音異義語の組を \rightarrow の関係からグループ化すれば良い. グループ化された集合の要素が 2 以上であれば, それをタイプ 1 の同音異義語として収集する. ただし平仮名表記が平仮名 1 文字の場合には, 同音異義語としないことにした. この処理を行うと, 当初, 単純に辞書見出しから収集した同音異義語 16,469 組 (平均要素数 3.469) は, 同音異義語 12,401 組 (平均要素数 2.851) に減少した.

4.2 タイプ 2 の同音異義語

タイプ 2 の同音異義語は, H 1 を満たす. つまり, ある位置のある漢字 1 文字が誤りやすい. ここでは, この誤りやすい漢字の組, 及び漢字の位置に傾向があると考え. 例えば, $\{\text{任命, 任名}\}$ はタイプ 2 の同音異義語であるが, 異なる漢字の組は“命”(○)と“名”(×)であ

¹ここでは第 4 版広辞苑 [3] を利用した.

り、その位置は2文字中の2番目である。よって、「■命」（■はある漢字1文字）という単語があり、しかも「■名」という単語が実在しなければ、{■命, ■名}をタイプ2の同音異義語として収集する。{■命, ■名}という収集の元になる規則が与えられれば、収集は辞書を利用することで容易に実現できる。ちなみにこの例の規則からは、以下の82組のタイプ2の同音異義語が収集される。

{復命, 復名}	{運命, 運名}	{恩命, 恩名}	{渴命, 渴名}	{急命, 急名}	{教命, 教名}	{愚命, 愚名}	{景命, 景名}	{巖命, 巖名}	{懇命, 懇名}	{死命, 死名}	{主命, 主名}	{峻命, 峻名}	{続命, 続名}	{身命, 身名}	{絶命, 絶名}	{待命, 待名}	{長命, 長名}	{定命, 定名}	{電命, 電名}	{任命, 任名}	{幕命, 幕名}	{非命, 非名}	{報命, 報名}	{明命, 明名}	{来命, 来名}	{領命, 領名}	{露命, 露名}
{違命, 違名}	{延命, 延名}	{革命, 革名}	{貴命, 貴名}	{救命, 救名}	{矯命, 矯名}	{君命, 君名}	{傾命, 傾名}	{抗命, 抗名}	{冊命, 冊名}	{使命, 使名}	{宿命, 宿名}	{詔命, 詔名}	{辱命, 辱名}	{身命, 身名}	{存命, 存名}	{短命, 短名}	{朝命, 朝名}	{帝命, 帝名}	{特命, 特名}	{捍命, 捍名}	{反命, 反名}	{復命, 復名}	{奉命, 奉名}	{奔命, 奔名}	{余命, 余名}	{立命, 立名}	{曆命, 曆名}
{遺命, 遺名}	{王命, 王名}	{活命, 活名}	{九命, 九名}	{窮命, 窮名}	{欽命, 欽名}	{敬命, 敬名}	{懸命, 懸名}	{顧命, 顧名}	{策命, 策名}	{辞命, 辞名}	{受命, 受名}	{上命, 上名}	{助命, 助名}	{性命, 性名}	{台命, 台名}	{致命, 致名}	{勅命, 勅名}	{天命, 天命}	{内命, 内名}	{薄命, 薄名}	{藩命, 藩名}	{奉命, 奉名}	{奔命, 奔名}	{余命, 余名}	{立命, 立名}	{曆命, 曆名}	

問題は {■命, ■名} のような展開の元になる規則をどのように得るかであるが、ここでは辞典Bを再び利用する。辞典B中のタイプ2の同音異義語の中で758組はH2を満たしている。この758組から758個の規則を作ることができる。ただし単語の中に平仮名が入っている場合は、上記のような展開は妥当性が薄いと思われるので、漢字文字だけで構成されているものを対象にする。結果、561個の規則を得ることができた。

これら規則を元にして、先の手順からタイプ2の同音異義語を収集した。その結果、10,026組のタイプ2の同音異義語を収集することがで

きた。

5 実験

収集したタイプ1の同音異義語についての評価は考察で述べる。ここでは収集したタイプ2の同音異義語についての評価として、新聞記事からタイプ2の同音異義語の誤り検出、修正の実験について述べる。使用した新聞記事は毎日新聞の94年度版CD-ROMの1ヶ月分(約8.1万文)である。この新聞記事から収集したタイプ2の同音異義語の実在しない単語を含む文を取り出し、その単語を実在する方の単語に修正する。ただし、実在しない単語を含むかどうかを判断する場合には、形態素解析を利用して単語切りを誤らないようにする。例えば、収集したタイプ2の同音異義語には、{処理, 所理}があるが、形態素解析を利用せずに文字列の一致だけをみる場合、「○○研究所理事長」という文章中の“所理”を“処理”の誤りであると誤検出してしまう。形態素解析を利用して“究”と“所”の間や“理”と“事”の間に単語切りが生じないことを利用すれば、このような誤検出を避けられる。

実験の結果、223個所を誤りの可能性ありとして提示した。そのうち以下の3種(4個所)は実際に誤りであると思われる。

- 【警報機】が鳴っている～(→警報器)
- 圧縮空気を使う【噴霧機】で～(→噴霧器)
- 訴訟の【口答弁論】が十三日午後、～(→口頭弁論)

これら3種の同音異義語は辞典Bには記載されておらず、規則を展開して収集した効果があった。

また以下の7種(8個所)は判定がやや難しい。誤りではないかもしれないが、警告することに意味はあろう。これらの同音異義語も辞典Bには記載されておらず、規則を展開して収集できたものである。

- 【価値感】薄い商品券(→価値観)
- 地域【情報紙】などで(→情報誌)
- 最大部数を誇る【大衆紙】(→大衆誌)
- アラゴンの【大公】ドン・ペドロ(→太公)
- 世阿弥の「目前【心後】」(→身後)

- 内藤【産政】局長 (→ 産制)
- のう胞【性線】維症 (→ 性腺)

残りの211個所は誤りではなかった。つまり誤検出であった。この原因は以下の通りである。

原因1 実在しない単語が句である (15種類, 171個所)

例えば、収集された同音異義語として{月機, 月期}がある。確かに、“月機”²という単語は実在しないが、“月期”という単語は実在しないが、“3月期決算報告”というように“月”+“期”として、“月期”が現れる。“月期”のような比較的句になることが予想しやすいものもあるが、以下の例のように、句になるのが予想しにくいものもある。

- 上智【大初】の首相となった細川首相 (→ 太初)
- 中華スープの【素大】サジ3 (→ 粗大)

原因2 実在しない単語が固有名詞 (あるいはその一部) である (20種類, 35個所)

例えば、収集された同音異義語として{精子, 清子}がある。“清子”という単語は辞書に記載されていないため、実在しないと判断されるが、固有名詞として一般に利用される単語である。また固有名詞部分は、未知語になる場合もあり、その場合、形態素解析で単語切りが誤る。そのための誤検出も幾つか存在した。

原因3 実在しない単語が造語である (4種類, 4個所)

以下の4つである。「真体」「魚歌」「爆裂! 異種格闘技TV」「真学塾」それぞれ、「心体」「漁歌」「爆裂! 異種格闘技TV」「神学塾」の誤りであると、誤検出している。

原因4 形態素解析の誤り (1種類, 1個所)

形態素解析で「罪悪感」を“罪”+“悪感”と区切っている。“悪感”を“悪寒”の間違いだと誤検出している。

原因1, 2ともに抽出された同音異義語を検査することで、基本的に対処は容易である。原因1, 2が誤検出の97.6%であり、これらへの対処が誤検出を大幅に減らす。

²“月機”の平仮名表記は「つきばた」であり、{月機, 月期}は同音異義語ではない。この問題は考察で述べる。

6 考察

6.1 収集したタイプ1の同音異義語の評価

収集したタイプ1の同音異義語は、当初用意していた同音異義語を分割、縮小したものである。例えば、最初に例に出した「かいとう」の同音異義語では、14種類から以下の2組に分割、縮小している。

{ 解答, 回答, 解糖, 解党, 解凍, 会党, 会頭 }
{ 戒刀, 快刀 }

このように分割、縮小することによって、作成すべき同音異義語の判定規則が減る点、実際に誤りやすい単語についてだけ検出が行われるため、適合率が良くなる点などから有用であると考えられる。

一方、縮小したことによる悪影響も考えられる。例えば、先の「かいとう」の例であれば、以下の5種類は削除されている。

{ 垣内, 灰陶, 怪盜, 械闘, 開冬 }

これらの単語が“解答”や“回答”の誤りとして記述されたり、逆に“解答”や“回答”がこれらの単語に書き誤る場合には縮小したことによる悪影響があったことになる。この点を調査することは困難であり、今後の課題とするが、一般的には悪影響はないと予想できる。付表1で示したようなH3に反する同音異義語を随時追加してゆくのが妥当だと思われる。

6.2 表記の利用

同音異義語の誤り検出、修正は基本的には語義選択問題と同等であるが、通常の語義選択問題とは非常に大きな違いがある。それは正解となる語義(正しい表記)が多くの場合、見えている(表記されている)ことである。これは文脈から語義を推定する問題にはない特徴である。同音異義語の誤り検出、修正にはこの特徴を利用すべきだと考える[4]。

本論文で提示したヒューリスティクスH1は、そのために利用できる。通常、同音異義語の組が $\{w_1, w_2, \dots, w_n\}$ であり、ある文書中に w_i が現れた場合、その w_i の部分は、 w_1 から w_n のどれになるかを調べることが、同音異義語の誤り検出、修正の基本的手順である。これは文書中に w_i が現れたという情報を利用してない。なぜなら文章中の w_i の部分が $w_j(j \neq i)$ であっても全く同じ計算が行われるからである。一方、





H1で導入した \rightarrow という関係は、推移律が成立しないために、 $w_i \rightarrow w_j$ が成立しているとは限らず、「 w_i が現れた」という情報を利用していることになる。例えば、「かいとう」の同音異義語として、

{ 解答, 回答, 解糖, 解党, 解凍, 会党, 会頭 }

を考える。文書中に「回答」が現れた場合には、「回答」から \rightarrow の関係を持つ、{ 解答 }だけを修正候補として考えれば良い。同様に「解凍」が現れれば、{ 解答, 解糖, 解党 }が修正候補となる。

6.3 読みの利用

本論文で提案したタイプ2の同音異義語を収集する手順では、収集された同音異義語の組内の各単語が同じ平仮名表記を持つとは限らない。この場合、その組は同音異義語ではないため、誤収集である。この点が不必要に多くの同音異義語を収集している原因の1つとなっている。

例えば、辞典Bの中で{異口同音, 異句同音}というタイプ2の同音異義語があるが、これを利用して{口, 句}という形の組を取り出したとしても、「口」の部分が「く」と読むか「くち」とよむかはわからない。「く」と読めば同音異義語になるが、「くち」と読めば同音異義語にはならない。例えば{一口同音, 一句同音}は正しく展開できているが、{折口信男, 折句信男}は誤っている。

同音異義語の収集では、この点を考慮すべきであったが、タイプ2の同音異義語を生成する手順では生成された単語が辞書に未記載であることを確認するので、収集した同音異義語の組が増えるだけで悪影響はないと考えた。しかし、前章の実験の誤り原因の例として挙げた{月機, 月期}や{精子, 清子}も読みの問題から誤収集した形になっており、実際に悪影響が出ている。本手法を用いた同音異義語の収集では、各漢字に読みを割り当てる作業が必要であった。ただしこの作業は単純ではなく、この点も今後の課題である。

この読みを考慮しなくてはならない点は、タイプ1でも生じている。たとえば、タイプ1の同音異義語として、{告白, 黒白}が収集されている。たしかに“黒白”は「こくはく」とも読むが、このように読む場合は非常に稀であろう。つまりタイプ1の同音異義語の収集に関しては、その読みの頻度のようなものも考慮すべきである。

6.4 誤りやすいパターン

本論文では、誤りやすいタイプ2の同音異義語を収集する際に、誤りやすい漢字文字とその位置に注目した。これは外観が似ている漢字については機能しているように思われるが、この点だけが誤りやすい単語を特徴づけているわけではない。

例えば、辞典Bでは、{更年期障害, 高年期障害}や{難航, 難行}などがあげられているが、漢字文字だけを見た場合、“更”が“高”あるいは“航”が“行”に書き誤りやすいとは思えない。これは“更年期”という意味から“高年期”、あるいは“難航”という意味から“難行”、という誤りが生じていると思われる。これらは意味的に想起される誤りと考えられる。このように誤りやすいパターンも更に調べる必要があるだろう。

特にワードプロセッサを利用して文書を入力している場合には、実在しない単語が入力されることは少ない。もしもそのような場合があるとすれば、単語の読みを知らずに、その単語を構成している漢字を1つずつ入力することから生じると思われる。例えば、“元帥”の読み「げんすい」を知らずに、“元”と“帥”を各々入力しようとして、“帥”を“師”と誤る場合である。このような誤りのパターンは、単語の読みが難しいという特徴がある。この点からも読みの利用を考えることができる。

7 おわりに

本論文では、まず誤りやすい同音異義語を収集することの意義を述べ、次に誤りやすい同音異義語に内在するヒューリスティクスを提案した。そして、既存の辞典を用いて、そのヒューリスティクスの妥当性を示した。また、そのヒューリスティクスを利用して、既存の辞書から誤りやすい同音異義語を自動収集した。収集したタイプ2の同音異義語の評価として、1ヶ月分の新聞記事からタイプ2の同音異義語の誤り検出、修正を行った。その結果、収集した効果が確認できた。読みを考慮した収集を行うことが今後の課題である。

付表 1: H3 を満たさないタイプ 1 の同音異義語

平仮名表記	同音異義語	平仮名表記	同音異義語
あたる	当たる, 中る	たぶん	他聞, 多分
いとこ	従兄弟, 従姉妹	ためし	例, 試し
いよう	威容, 偉容, 異様	ちゅうしん	中心, 注進, 衷心
うば	乳母, 姥	ていき	定義, 提議
うまい	旨い, 甘い, 美味しい, 巧い, 上手い	てきよう	摘要, 適用
おかみ	女将, 御内儀, 御上	てんかい	展開, 転回
おこす	起こす, 興す, 熾す	とうしょ	当初, 当所, 頭書
おさえる	押さえる, 抑える	とうじょう	登場, 搭乘
おそれ	恐れ, 怖れ, 畏れ, 虞, 憎れ	とうりょう	棟梁, 統領, 頭領
おろす	下ろす, 降ろす, 卸す, 墮ろす	とたん	途端, 塗炭
かいこ	回顧, 懐古	とらえる	捕らえる, 捉える
かいせき	会席, 懐石	なごり	名残, 余波
かかり	係, 掛かり	のほせる	上せる, 逆上せる
かける	駆ける, 駈ける, 翔る	はくちゅう	白昼, 伯仲
かたまり	固まり, 塊	はどこ	再従兄弟, 再従姉妹
かたみ	片身, 形見, 肩身	はね	羽, 羽根
かんき	喚起, 換気	はやる	流行る, 逸る
きかん	気管, 器官	はんえい	反映, 繁榮
きぐう	奇遇, 寄寓	はんこう	反攻, 反抗, 犯行, 版行
きょうい	脅威, 驚異	はんらん	反乱, 汎濫
きょうこう	凶行, 凶荒, 恐慌, 恐惶	ひとり	一人, 独り
きょうせき	行跡, 業績	ひょうしょう	表彰, 標章, 表象
けんぶん	見聞, 検分	ふける	更ける, 深ける, 老ける, 蒸ける, 耽る
こうぎ	公儀, 広義, 講義	ふしん	不信, 不振, 不審, 腐心
こうぎ	剛毅, 蒙気	ふだん	不断, 普段
こうふ	公布, 交付	ふへん	不変, 普遍, 不偏
こま	駒, 駒, 独楽	ふるう	振るう, 奮う, 震う, 揮う
ごま	胡麻, 護摩	へんい	変移, 変異, 変位, 偏倚
さらう	浚う, 漂う, 攪う, 復習う	へんかん	返還, 変換
しい	私意, 思惟, 恣意	ぼいん	母音, 拇印
じぎ	時季, 次期, 時機	ほうえき	防疫, 貿易
しじょう	市場, 誌上	ほうしょう	報奨, 報償, 褒章, 褒賞
しぞう	志操, 思想	ほうよう	包容, 抱擁
したい	肢体, 姿態	ほしょう	保障, 保証, 補償
じたい	自体, 事態	ほそく	補足, 補則, 捕捉
しゅうそく	収束, 集束, 終息	まずい	不味い, 拙い
じゅよう	受容, 需用, 需要	まっしょう	末梢, 抹消
しょうかい	紹介, 照会	むげん	無限, 夢幻
しょうひょう	商標, 証票, 証憑	やもめ	鰥夫, 寡婦
しわけ	任分け, 任訳	ゆうしゅう	有終, 優秀
すいこう	推敲, 推考, 遂行	ゆえん	由縁, 所以
せっせい	節制, 撰生	ようい	用意, 容易
せんゆう	占有, 占有	よだん	予断, 余談
そうさ	搜查, 走査, 操作	らち	埒, 拉致
そうぞう	創造, 想像	らんだ	乱打, 懶惰
そうたい	相對, 総体	るいけい	累計, 類型
そがい	阻害, 疎外	るいさん	累算, 類纂
そくせい	促成, 速成, 即製	わかる	別れる, 分かれる

参考文献

付表 2: H2 を満たさないタイプ 2 の同音異義語

平仮名表記	正しい表記	誤りの表記
いしんでんしん	以心伝心	意心伝信
いっちょういっせき	一朝一夕	一鳥一石
いみしんちょう	意味深長	意味慎重
うわする	上する	浮わする
おうよう	鷹揚	応容
かいとうらんま	快刀乱麻	怪盗乱魔
かんこつだったい	換骨奪胎	換骨脱体
かんしょうちたい	緩衝地帯	干涉地帯
かんぜんちょうあく	勧善懲悪	完全懲悪
きち	機知	氣智
きとく	奇特	氣篤
ぎまん	欺瞞	偽満
きゅうきゅう	救急	急救
げきりん	逆鱗	激隣
けんびぎょう	顕微鏡	檢微鏡
こうきしゆくせい	綱紀肅正	網規肅正
こうよう	昂揚	高楊
ごたぶん	ご多分	ご他聞
こんすい	昏睡	混酔
さいしょくけんび	才色兼備	才色健美
さんはんきかん	三半規管	三半器官
しこうさくご	試行錯誤	思考錯誤
しこうひん	嗜好品	趣向品
しそうけんご	志操堅固	思想堅固
しゅうしゅう	収集	集収
しゅし	趣旨	主昏
じゅじゅ	授受	受授
じょうじょうしゃく りょう	情状酌量	状情酌量
しょうしんしょうめ い	正真正銘	正真証明
しんしんしょう	心身症	身心症
せいしんせいい	誠心誠意	精神誠意
せっぱ	切羽	接端
そうしそうあい	相思相愛	想思想愛
そうじょうこうか	相乗効果	双乗功果
たいぎめいぶん	大義名分	大儀明文
てんかいず	展開図	転回図
どうひょう	道標	導票
とっさ	咄嗟	突差
なんぎょうくぎょう	難行苦行	難業苦業
はいはい	売買	買売
はつらつ	澆刺	発刺
ふしょうぶしょう	不承不承	不精不精
ふんさい	粉碎	紛細
ほうちく	放逐	傲逐
ほうようりょく	包容力	抱擁力
ゆうしゅうのび	有終の美	優秀の美
ろんこうこうしょう	論功行賞	論行功賞

- [1] 現代言語研究会：“ワープロ時代の同じ読みで意味の違う言葉の辞典”，あすろ出版 (1996).
- [2] 日本語表現研究会：“これは役立つ！間違え言葉の辞典”，PHP 文庫 (1996).
- [3] 新村出 編：“広辞苑 第四版”，岩波書店 (1993).
- [4] 新納浩幸：“複合語判定を優先させた決定リストによる同音異義語判定”，情報処理学会自然言語処理研究会, NL-124, pp. 63-70 (1998).