

# 日本語表記の変換に関する実験的研究

国立国語研究所 中野 洋

## 1 目的

日本語表記を計算機によって変換するプログラムを作成する。

## 2 問題の所在

それらしくない日本語が存在する。

ワープロで書かれた文章、機械翻訳で作られた文章、日本語を母語としない人が書いた文章(注1)によく見られる。これまでは、機械が作るのだからとか日本語が外国語なのだからと我慢していた。しかし、日常生活のどこでも出会うようになると気になってしょうがない。何が原因なのだろう。どうすればよいのだろう。

ここでは問題を表記に限って、機械で直す方法を考える。

機械化にかかる前に日本語の表記について考えよう。ここでの問題は、それらしい日本語表記とは何かである。つまり日本語らしい日本語表記とは何か。

日本語は4つの文字体系を持つ言語である。したがって、漢字だけでも平仮名だけでも片仮名だけでも、またアルファベットだけでも書くことができる。もちろん、現在は、漢字仮名混じり文で書いている。これらの書き方はすべて正しい。正しければ日本語らしい日本語か。もちろんそうではない。

では、それらしくない日本語表記とは何か。

明らかに誤った日本語表記がある。ひらがな、片仮名、漢字、ローマ字のそれぞれの文字の書き方を誤ることがある。不注意や知らないために起こる書き間違いが多い。しかし、ここで問題にしているのは無い字を書いてしまう間違いである。ところが、このような間違いは機械を使えば起こらない。注意すれば人間にも起こらない。したがって、このことは問題としない。

ある語に用いる文字を誤ることがある。たとえば、「学習」を「学秀」とか「学修」と書けばそれは誤りである。これも機械の中の辞書が完備すれば機械が自動的にチェックし修正できるようになるだろう。

ところで、上の例で「学習」を「学しゅう」と表記した場合、これは間違いではない。しかし、この表記は普通の日本人が用いる表記ではない。小学生の、それも1年か2年生の表記だと思いに違いない。

このように、間違いではないが普通ではない表記がある。

ここで重要なのは「普通」の意味である。

小学4年生としては普通の表記、日本語学習者としては普通の表記、年賀状、社内文書、公文書としては普通の表記、近頃の若い女性としては普通の表記などと普通の表記がいろいろある。そしてこれら複数の文章の表記を語ごとに集計するといろいろな表記が存在することがわかる。これを広い意味での表記のゆれとよぶ。

ある文章には期待されるそれ用の表記がある。表記がそれ以外のものだとして「あれおかしいな、誰が書いたのかな、教養がないな」など思ったりする。極端な場合には商談が壊れたり、信頼関係が壊れたりすることもある。

さて、一般の文章の表記を決める要因は何だろうか。

## 3 表記選択に関わる要因

さて、表記を決める要因には、言語的要因、表記の位相に関わる要因、心理的要因、社会的要因などが考えられる。それぞれの要因が以下のように表記の決定に関わっていると思われる。もちろんこれは仮説であって、それが本当かどうかは調査によって確かめるべきだろう。その

結果は今後の研究に待つことにする(注2)。

### 3.1 言語的要因

そのために作られた文字体系を使わなければ発音された通りに表記することはできない。したがってその言語に用意されている文字および用字法のフィルターを通して表記することになる。そのフィルターにはどのようなものがあるのかを現代日本語について考える。

3.1.1 文字に関わる要因 通常行われている日本語の文字の使い方には以下のような傾向が認められ、文章作成での表記の指針となっている(例：中野(1979))。以下、文字体系ごとにしよう上の問題をあげる。

【平仮名】には、仮名遣いの問題がある。「4つがな」の使い方、助詞「は」「へ」「を」の表し方、長音、促音・拗音の表記が問題となっている。

また、助詞・助動詞、副詞・接続詞・感動詞や用言の語尾の部分は平仮名で書くことを原則とする。常用漢字で書けない部分など他の文章体系を使わない場合は平仮名で書く。

【片仮名】は、外来語、擬声語・擬態語、外国人・外国の地名、外国語音、不明語は片仮名で書くことを原則とする。動植物名は片仮名で書くことが多い。強調したい語に片仮名を用いることがある。

【英文字】は、略語や外国語をそのまま表すのに用いる。ローマ字表記では、訓令式とヘボン式などの特定音節、分かち書き、長音の表記の方法、外来語を外国語表記するか日本語表記するかなどの問題がある。

【漢字】は、語の概念を表す部分を漢字で書く。したがって、多くの名詞や用言の語幹部分は漢字で書く。これらは辞書による。ところで、名詞であっても、用言の語幹部分であっても平仮名を使う場合が少なくない。

ひとつは、漢字制限にしたがう場合(常用漢字だけで書くとか、小学校4年生までの学習漢

字だけで書くなどという場合)がある。また、基本語彙や和語は漢字を用いないという場合もある。

【記号】の使い方にも注意が必要である。句読点は、縦書きでは「、」「。」、横書きでは「,」「.」。または「.」を用いる。繰り返し記号である「々」は漢字に、「ゝ」は仮名に用いる。

【算用数字】は、主に横書きに用いる。置き換えられない数は横書きでも漢数字を用いる。縦書きでも強調したい場合は算用数字を用いることがある。桁の大きい数字を表す場合は、「万」「億」「兆」などの単位を漢字で入れる。

以上のようにこれら指針には原則と例外がある。例外は、語ごとに見る他ない。中にはひとつの語でいくつもの表記があることもある。

3.1.2 語に関わる要因 上でも記したように、語種による表記法がある。和語は平仮名、漢語は漢字、外来語は片仮名表記をするというのがそれである(注：昨年の本新プロの加藤秀俊氏の講演)。また、品詞が表記を決定していることがある。上に述べた通りである。さらに、語彙が表記を決定することがある。「くみたて語彙」(中野(1975)：注3)は平仮名で書くことが多いとか、専門語は漢字で書くことが多い、自動車やパソコンの名前、コンピュータ用語は英文字を使うことが多いということがある。

同じ語でも用法により表記を変えることがある。形式名詞や補助用言としての用法の場合には平仮名表記し、原爆都市の場合は「ヒロシマ」と仮名表記するなどはこの例である。

個々の語の表記を決定するには、表記台帳ファイルを用いる。表記台帳には、その語の使用されたあらゆる表記、それぞれの頻度、語種、品詞、意味コード、位相情報、語彙情報を記する。他の表記の要因を決定したうえでこの台帳を引き、どの表記を選ぶかを選択する。

3.1.3 文章に関わる要因 一つの文章の中で

の表記のゆれは避ける。

また、文章のジャンル別に見ると他のジャンルとは異なった表記をとることがある。例えば、社内文書、公文書にはそれぞれ決まった用字法がある。広告やスポーツ新聞の見出しには読み手の関心を引く普通ではない表記が見られる。礼状や挨拶状、公文書、企画書など社内文書、各種の申込書などには決まった文章表記の型がある。

### 3.2 位相が関わる要因

3.2.1 書き手による表記の決定 個々人がそれぞれの表記法を持っている場合がある。それはその人の表記に対する主張によってそうしている場合もあり、その人が意識しない理由である表記法をとる場合もある。その人が育った時代や文化、受けた教育、職業などによら表記の違いが見える。

性による表記の違いがある。女性は、角ばった漢字より、丸みをおびた平仮名を好む傾向がある。昔、漢字を男手、平仮名を女手とよばれていた。例えば、次は、女性が書いた商用文書である。

前略御免くださいませ。

本日16日付のお尋ねのお手紙いただきました。お尋ねについての追加資料を同封いたしました。よろしく願い申し上げます。

かしこ

3.2.2 読み手による表記の決定 我々は、相手にあわせて文章を書くことができる。相手が子供であれば、その子が読める表記、理解できる語を使うだろう。ところで、相手は、年齢、性、職業、書き手との関係などのいろいろな特徴を持っている。そのどの部分にあわせて書くのかを選択しなければならない。

3.2.3 時代 表記は、時代によって変化している。仮名遣いや使用漢字の量、漢字の宛方な

どに違いが見られる。

### 3.3 心理的要因

字を書くのが上手でない、漢字の書き方を間違える、表記の型を知らないなどの理由で文章を書くのが嫌いだという人が少なくない。この人達はその理由がなくなれば文章を書き出す可能性がある。もしそうなれば、もともと他人の評価に敏感な人達だから表記もある種の傾向を持つかも知れない。たとえば、いい内容の文章と思わせるために、あるいは英文字を多く用いるかも知れない。また、表記にあるイメージをこめるために平仮名を多用するかも知れない(注4)。

### 3.4 社会的要因

3.4.1 国が決めた表記のめやすがある。

表記について個人の主張があっても、公けの文章では表記についての規則を守る場合がある。

3.4.2 日本語表記に関わる電子機器の制約がある。

ディスプレイには縦書きがないので横書きが多くなる。

機械に無い字は使わない。

機械辞書にない表記は面倒なので使わない。

同音語の中から選択する。

いろいろなフォントや文字飾りを使おうとする。

3.4.3 社会が要求する表記がある。

漢字は、真名と呼ばれたように本当の字と考えられ、正式な文章には漢字が多く使われる。これに対し仮名が多い文章は私的な文章だと思われる。

儀礼的な文書の表記に従う。たとえば、商用文書ではフォーマルな表記が好まれる。次の二つの文章は、同じ英語の手紙からの翻訳である。前の文章が漢語を多く用いているので、よりフォーマルな感じを与えている。また先に女性の表記であげた例はより私的な表記と感じられる。

拝復

本日、16日付御照会拝受致しました。  
御照会に付いての追加資料を同封致しました。  
よろしく願い申し上げます。

敬具

拝復

本日16日付のお尋ねのお手紙いただきました。  
お尋ねについての追加資料を同封いたしました。  
よろしく願い申し上げます。

敬具

#### 4 実験方法

表記選択が何によって行われているのか現在のところ決めがたい。しかし、本研究はそれを決めるのが目的ではない。シュミレーションを行うことによって表記選択の要因を探る方法を提案することにある。

したがって、実験プログラムでは各要因ごとに表記変換モジュールを作成する。つぎにそれぞれのモジュールをどのような方法で起動するか、各変換プログラムで生じた表記候補をどのような方法で選択するかを決める。

##### 4.1 実験の目的

書かれた文章の表記をパラメータによって望む表記に変換する。

##### 4.2 方法

1. パラメータを読む。
2. 入力データを仮名、数字、記号に変換する。  
一貫処理(形態素解析)にかける。
3. 指定された変換モジュールを起動する。
4. 得られた複数の表記からどれか一つを選択する。
5. 出力する。
6. 評価する。

##### 4.3 変換モジュール

###### 1) 使用漢字セットの指定

- ・教育漢字(第1学年、第2学年...)で書く
- ・常用漢字音訓表を用い、常用漢字で書く
- ・漢字セットを指定する
- ・国語研究所漢字調査使用率順漢字表
- ・代表音50音順漢字表
- ・音訓順漢字表から選択する)
- ・できるだけ漢字で書く

###### 2) 仮名遣いの指定

- ・現代仮名遣い(内閣告示、昭和61年)にしたがう
- ・歴史的仮名遣いにしたがう

###### 3) 外来語の表記の指定

- ・外来語の表記(内閣告示、平成3年)にしたがう
- ・音節単位の変換表を作りそれに従う
- ・外来語表記台帳に従う

###### 4) ローマ字表記の指定

- ・ローマ字のつづり方第1表(訓令式、内閣告示、昭和29年)を用いる
- ・ローマ字のつづり方第2表を第1表より優先する
- ・ローマ字のつづり方を決めてそれにしたがう

###### 5) 送り仮名の付け方の指定

- ・送り仮名の付け方(内閣告示、昭和56年)の本則、例外にしたがう
- ・同上の許容を優先する
- ・表記台帳からできるだけ多く送る表記にしたがう

###### 6) 数字の指定

- ・横書きでは、算用数字を用いる
- ・縦書きでは、漢数字を用いる

###### 7) 記号の指定

- ・句読点の指定をする。「、。」「、,」「,。」  
「、,」
- ・繰り返し記号「々々」を使う、使わない

###### 8) 語種による表記の指定

- ・和語は平仮名、漢語は漢字、外来語は片仮名で書く

###### 9) 品詞による表記の指定

- ・助詞・助動詞、副詞・接続詞・感動詞は、平仮名で書く
  - ・外国人・外国の地名、外国語音、不明語は片仮名で書く
  - ・擬声語、擬態語は、片仮名で書く
- 10) その他の語彙の指定
- ・基本語彙は、平仮名で書く。
  - ・ことがら語彙は、平仮名で書く
  - ・指定の専門語彙は、漢字で書く
- 11) 書き手の指定
- 12) 読み手の指定
- 13) 国が決めた表記の規則に従うかどうか
- 14) 表記台帳による表記の指定
- ・表記台帳での使用率の高い表記で書く
  - ・表記台帳での平仮名表記で書く
  - ・表記台帳での片仮名表記で書く
- 15) その他

#### 4.4 選択の方法

いくつかのモジュールを動かすと複数の表記を得ることがある。これら複数の表記のどれをどのような方法で選ぶかが問題である。

モジュール1から7までを用いれば表記の揺れは生じない。しかし、8から15までのモジュールを使うことがある。それゆえ、ゆれが生じる。その揺れは我々の表記活動でのゆれと同じだろうか。

我々あるいは筆者がそれらしいと思う表記は何なのだろうか。それらしいという慣れによるものかも知れない。実際、それぞれの表記は、最初、書き手と読み手および内容によりある文体が選択され、その後は、書き進むごとに無意識に右左と選択されているのかも知れない。

表記の規則に従う表記決定はトップダウンに決定する方法であり、一つ一つの語により決定して行く方法はボトムアップに決定する方法である。

現段階では、実験での選択の方法は決定せず、いろいろな方法が取れるように設計する。

#### 5 この実験の意義

この研究の結果として、日本語表記を研究するための道具「表記変換プログラム」を得ることができる。もし、変換実験が初期の目標を達成すれば、そこで用いた表記の規則と表記選択の方法が必要十分であることを示していると考えられる。

この意味で、本研究は実験的方法による表記研究のひとつである。

#### 6 今後の課題

ある文章を同じ内容の他の文章に書き換える場合、書き換える部分は表記だけにとどまらない。用語や表現も変える必要が生じるだろう。ところで、語の選択を考えてみる。『『分類語彙表』形式による語彙分類表』(中野(1989)の分類番号 1.3122 の第3段落には「手紙」の類義語がある。それを形を変え、右端に語種などの注を示すと次の表ができる。

文,手紙,矢文	和語
信書,書信,書簡	漢語
レター	外来語
恋文,付け文	和語
艶文,艶書	漢語
ラブレター	外来語
寸書,寸楮	漢語：謙讓語
貴書,貴札,貴翰,芳書,芳信,芳墨	漢語：尊敬語

この語群を見れば容易に分かるように、用語の選択は表記の選択や文体の選択と連動している。

表記変換は、文体の変換、語彙選択、表現の変換へ進む。これは中野班の次の研究課題である。

#### 注

注1：「日本語表記とは何か。」は明らかにワープロで作った表記である。現在の機械翻訳でどんな日本語が作られるか、成田(1995)他に

実例がある。外国人が書いた文章がおかしなるのは、母語の影響がある。本研究でも実態調査を予定している。Holden(1989)は日本人が書いた英語を機械で修正するための研究である。

注2：井上道雄(1995)は、研究班3の文字言語研究チーム(賀集班)において日本語表記選択の主観的要因について研究を進めている。本れを機械にのせて表記変換実験を行う。その結果を評価することによって表記選択の仕組みを考察する。

注3：文の内容とは関係が薄く、文を組み立てるためだけに必要な語。形式名詞や補助用言など。

注4：文章表記の調査をすると、「おとこのこ」を平仮名表記をする人が1、2%はいるのが普通である。

堀内雄一、山崎一生(1990):英単語のアルファベット表 90-NL-79,1990.9.21)

Natsuko I. Holden(1989): Interference software for Japanese Writers of English (CCL report, UMIST)

#### 参考文献

井上道雄(1995):表記の「ゆれ」と「多様性」-- 筆記実験による分析 1-- (「国際社会における日本語についての総合的研究」第1回研究会予稿集)

中野洋(1975):文章の語彙構造に関する探索的研究(2):(国立国語研究所言語計量研究部季報 1975-冬号)

中野洋、佐竹秀雄(1979):「表記の技法」,(『文章作法事典』,pp110-112、東京堂出版)

中野洋(1989):『分類語彙表』形式による語彙分類表』,(昭和61年度~昭和63年度科学研究費「大量データの収集と処理の研究」代表者:野村雅昭)

中野洋(1993): 機械翻訳に望む日本語の質 (電子情報通信学会資料)

中挟知延子、島田静雄(1995):外国人のための日本語文章校正システム(第7回テクニカルコミュニケーションシンポジウム論文集 pp15-19、テクニカルコミュニケーター協会)