

TR-IT-0003

音声言語データベースのための
日本語形態素情報と表記の体系
Japanese morphological information and orthography
for ATR spoken language database

浦谷 則好 竹沢 寿幸 田代 敏久 衛藤 純司†
Noriyoshi URATANI Toshiyuki TAKEZAWA Toshihisa TASHIRO Junji ETOH†

1993. 7. 13

内容梗概

ATR 自動翻訳電話研究所で開発された言語データベースと、音声翻訳システム ASURA で使われた音声認識用文法、日本語解析用文法は、それぞれ別の目的のために独立に開発されてきたので、日本語形態素においてさまざまな差異(大きく分けて、品詞の設定の差異、語の分割の差異、表記の揺れの3種類)が存在していた。ATR 音声翻訳通信研究所で新たに構築する音声言語データベースでは、できる限り、その差異を減らし、音声言語統合処理の研究を円滑に進められるようにしたい。本報告書では、そのために検討した日本語形態素と表記の体系について述べる。

ATR 音声翻訳通信研究所
ATR Interpreting Telecommunications Research Laboratories
†日本アイアール株式会社

© 株式会社 エイ・ティ・アール音声翻訳通信研究所
© 1993 by ATR Interpreting Telecommunications Research Laboratories

目次

1	まえがき	1
2	名詞	1
2.1	固有名詞	1
2.2	サ変名詞	3
2.3	形容名詞	3
2.4	普通名詞	4
2.5	数詞	5
2.6	代名詞	6
2.7	名詞その他	6
3	動詞	7
3.1	本動詞	7
3.2	補助動詞	9
3.3	動詞その他	10
4	形容詞	11
5	副詞	12
6	連体詞	13
7	接続詞	14
8	感動詞	16
9	助動詞	16
10	助詞	18
11	接尾辞	20
A	日本語の表記	21
A.1	「出現形」と「標準形」	21
A.2	「出現形」の表記基準	21
A.2.1	漢字とかなの使い分け	21
A.2.2	送りがな	23
A.2.3	外来語・外国語固有名詞	24
A.2.4	その他	24
A.3	「標準形」の表記	25
	参考文献	26

1 まえがき

ATR 自動翻訳電話研究所で開発された言語データベースと、音声翻訳システム ASURA で使われた音声認識用文法・構文解析用文法は、それぞれ別の目的のために独立に開発されてきたので、日本語形態素の認定の仕方にさまざまな差異(大きく分けて、品詞の設定の差異、語の分割の差異、表記の揺れの3種類)があった。そのため、音声言語の統合的な処理を進めるうえで支障をきたしていた。そこで、ATR 音声翻訳通信研究所で新たに音声言語データベースを構築するにあたり、日本語形態素の体系を見直すことにした。

その際、次のような原則に従った。

- できるかぎり音声翻訳システム ASURA で用いた音声認識用文法や構文解析用文法との整合性を取る。
- できるかぎり表層で判断できるようにし、作業者による揺れの少ない基準とする。
- これまでの形態素認定を不都合がないかぎり踏襲して、従来との整合性を確保する。

本報告書では、そのようにして検討した日本語形態素と表記の体系について述べる。

2 名詞

2.1 固有名詞

固有名詞とは、同じ種類に属する他のものから区別するために、そのものだけに付けた名前で、「人名」「地名」「団体名」「役職名」「書名」「商品名」など、さまざまなものがある。また、世界中に指すものが一つしかなかったら、それは固有名詞である。(ただし、固有名詞だから世界中に一つしかない、と言っているのではない。例えば「ヒルトン・ホテル」は世界中にいくつもある。)

- 「人名」 鈴木、真弓、レオポルド・ブルーム
- 「地名」 東京、アイルランド、ヨーロッパ
- 「団体名」 ATR、日本航空
- 「役職名」 データ通信部長、田中教授
- 「書名」 日本語形態素解析マニュアル、フィネガンズ・ウェイク
- 「商品名」 ネスカフェ、マッキントッシュ

姓・名、および住所表示の際の地名は、他の固有名詞と区別して、『人名』『住所名』という品詞名を与える。姓と名からなる人名は、姓・名に短単位分割する。「都」「府」「県」「市」「区」は接尾語とし、「東京都/渋谷区」のような住所表示は短単位分割する。

- 鈴木 / 真弓
人名 人名
- ジェイムズ / ジョイス
人名 人名

- 東京 / 都 / 渋谷 / 区
住所名 接尾 住所名 接尾

***注**

☆『人名』という品詞は指示対象が人である場合にのみ使われるので、例えば、複合固有名詞の中に人名が含まれていても、『人名』とはしない。

- 山本厩舎 → 固有名詞

***注**

☆地名であっても住所表示に使われない場合は、『固有名詞』という品詞名を与える。

- 次の会議は 東京 で行われます。
固名
- 東北 地方は雨でしょう。
固名
- 石川 国体で優勝した。
固名
- 日本海 でミサイル実験が行なわれた。
固名

人名と住所を除いて固有名詞は長単位分割を基本とする。『日本語形態素マニュアル』(p. 10)の「固有名詞で切れるもの」として列挙してあるもののうち、次のものは長単位分割する。

- 天理教(固名詞)
- キリスト教(固名詞)
- 仏教(固名詞)

***注**

☆普通名詞と固有名詞の区別は大変難しい。『日本語形態素マニュアル』でも、例えば、「温州蜜柑」は固有名詞の項にあるのに、「夕張メロン」は普通名詞の項にある。固有名詞の判断基準は「唯一性」であるが、商品名のように指示対象が複数個存在するものもあって、唯一性そのものをどう考えるかも問題である。そこで、判断に迷うものは長単位分割して『普通名詞』とする。

- 温州みかん、夕張メロン → 迷ったら『普通名詞』

2.2 サ変名詞

「する」を付けてサ変動詞となる名詞は、文中のどこに出現しても、動作性を保っているかぎり『サ変名詞』とする。

- 私は自動 翻訳 電話 の 研究 に 従事 しています。
 サ名 サ名 サ名 サ名
- 事務局から 催促 の 電話 が来ました。
 サ名 サ名

動作性を失って具体的な物を指すときは『普通名詞』とする。

- 電話 が故障して連絡がつきませんでした。
 普名
- ファックス の使い方がわかりません。
 普名

動作性を保っているかどうか迷うような場合は『サ変名詞』とする。

- 電話 でお知らせします。
 サ名

*解説

☆従来は「する」の前に助詞がくる時はすべて『普通名詞』としていたが、これ以後、動作性の有無で区別する。

- 電話をする → 電話 / を / する
 サ名 格助 本動詞

2.3 形容名詞

助動詞の「だ」「です」「たる」をつけて形容動詞となるもの。いわゆる形容動詞自体は形容名詞プラス助動詞「だ」とする。

文中のどこに出現しても、人・事物の性質・状態を表すかぎりは『形容名詞』とする。

- とても きれい な人ですね。
 形名
- 努力が 無駄 になる。
 形名

性質・状態を表すものではなく、抽象的なものを指す場合は『普通名詞』とする。格助詞の「が」「を」や、連体助詞の「の」が続く場合、および連体修飾される場合が多い。

- 役所のすることには 無駄 が多い。
普名
- それは 緊急 を要する事柄です。
普名
- 幸福 の黄色いハンカチ
普名
- 家族に知らせる 必要 はありません。
普名

性質・状態を表すか、抽象的なものを表すか、迷うような場合は『形容名詞』とする。

- 子供の成長が 楽しみ になる。
形名

***注意**

☆「いろいろ」のように、単独で連用修飾する場合は『副詞』とする。従来は『普通名詞』としていた。

- いろいろ アレンジしてさしあげるとか。
副詞

***注意**

☆『大辞林』にエントリーされてない形容名詞は、コメントを入れる。従来は『新明解国語辞典』を基準にしていた。

2.4 普通名詞

固有名詞は長単位分割するのを基本とするが、普通名詞の場合は(複合語は)短単位分割するのを基本とする。その際の認定基準は『大辞林』の見出し語とする。(派生語は採用しない。)

- 国際会議 → 国際 / 会議
普名 普名
- 中華料理 → 中華 / 料理
固名 普名
- 日本人 → 日本 / 人
固名 接尾

● 日本語 → 日本 / 語
固名 接尾

● 発展途上国 → 発展 / 途上 / 国
サ名 普名 接尾

○ 宅急便 → 固名詞

*注

☆複合名詞に含まれるサ変名詞は、従来『普通名詞』としていたが、動作性を保っているかぎり『サ変名詞』とする。

● 参加申込書 → 参加 / 申込 / 書
サ名 サ名 接尾

『日本語形態素解析マニュアル』では、「北半球」と「南半球」「夏休み」と「冬休み」の例を挙げて、『新明解』（今後は『大辞林』になる）の切り方の不統一をそのまま反映させているが、これも固有名詞・普通名詞の切り方の原則に従うことにし、『大辞林』の見出し語にある方にそろえる。

- 北半球 → 普名詞
- 南半球 → 普名詞
- 夏休み → 普名詞
- 冬休み → 普名詞

2.5 数詞

数詞は短単位分割を原則とする。

● 三万五千元 → 三 / 万 / 五 / 千 / 円
数詞 接尾 数詞 接尾 接尾

● 三七一四 → 三 / 七 / 一 / 四
数詞 数詞 数詞 数詞

もちろん、数字としての意味が希薄になっているものは、全体として『普通名詞』とする。

- 七五三 → 普通名詞
- 九九 → 普通名詞

月・日、および、時刻の言い方は、数が限られているので、構文解析や翻訳を考慮して例外的に一語とし、『日時』という品詞名を与える。

● 四月(しがつ) → 四月
日時

○ 四月 (よつき) → 四 / 月
数詞 接尾

● 二十日 (はつか) → 二十日
日時

○ 二十日 (にじゅうにち) → 二 / 十 / 日
数詞 数詞 接尾

● 今は五時三十分です → 五時 / 三十分
日時 日時

○ 五時間三十分かかる → 五 / 時間 / 三 / 十 / 分
数詞 接尾 数詞 数詞 接尾

2.6 代名詞

普通の代名詞と疑問代名詞の区別はせず、すべて『代名詞』という品詞を与える。

*注

☆「いつ」「いくら」は、助詞を伴って名詞的に使われることもあれば、助詞を伴わずに単独で副詞的に使われることもある。

● いつ がよろしいですか。

代名詞

● いつ おいでになりますか。

副詞

● いくら で売ってくれますか。

代名詞

● いくら かかりましたか。

副詞

『日本語形態素解析マニュアル』「13.4 疑問詞の扱い」では、「いつ」は「代名詞」のみ、「いくら」は副詞のみとなっている。しかし、構文解析用文法を考慮して、これらは多品詞語とみなす。

2.7 名詞その他

住所は、前述したように「都」「府」「県」「市」「区」を接尾語として、「大阪市」「東区」などは短単位分割する。

● 大阪市東区一丁目二番地 →
大阪 / 市 / 東 / 区 / 一 / 丁目 / 二 / 番地
地名 接尾 地名 接尾 数詞 接尾 数詞 接尾

3 動詞

3.1 本動詞

1. 原則として、語幹と活用語尾に分割する。ただし、不変化部分を語幹とみなすことを基本とする。上一段・下一段動詞の区別は廃して、単に「一段動詞」とする。

● 送る	→	送	/	る
		本動・五段・語幹		活語尾
● 見る	→	見	/	る
		本動・一段・語幹		活語尾
● 教える	→	教え	/	る
		本動・一段・語幹		活語尾

一段動詞の未然形・連用形には活用語尾がなく、語幹と同じ形が用いられる。本来は活用語尾をゼロ記号(ϵ)とすべきだろうが(教え/ ϵ /た)、 ϵ を設定するのは現在の形態素付与では困難なので、この部分を省いて、次のように形態素解析する。

● 教えた	→	教え	/	た
		本動詞・一段・語幹		助動詞

カ変動詞・サ変動詞は語幹がなく、すべて活用形で扱う。

● 来ない	→	来	/	な	/	い
		本動詞・カ変・未然				
来ます	→	来	/	ます		
		本動詞・カ変・連用				
来る	→	来る				
		本動詞・カ変・終止				
来る	→	来る				
		本動詞・カ変・連体				
来れば	→	来れ	/	ば		
		本動詞・カ変・仮定				
来い	→	来い				
		本動詞・カ変・命令				

● しない	→	し	/	な	/	い
		本動詞・サ変・未然				
される	→	さ	/	れる		
		本動詞・サ変・未然				
します	→	し	/	ます		
		本動詞・サ変・連用				
する	→	する				
		本動詞・サ変・終止				
する	→	する				
		本動詞・サ変・連体				
すれば	→	すれ	/	ば		
		本動詞・サ変・假定				
せよ	→	せよ				
		本動詞・サ変・命令				
しろ	→	しろ				
		本動詞・サ変・命令				

2. 慣用表現は長単位で一語とする。ここで慣用表現とは、動詞が本来の実質的な意味を失って、名詞の表す行為・過程・状態・現象などのなんらかの側面を特徴づけているにすぎないような表現を言う。全体の意味を個々の語から構成できないものである。

● 腹が立つ	→	腹が立	/	つ
		本動詞・五段・語幹		活語尾・五段・終止
● 目をかける	→	目をかけ	/	る
		本動詞・一段・語幹		活語尾・一段・終止
● 手に入れる	→	手に入れ	/	る
		本動詞・五段・語幹		活語尾・五段・終止

***注**

☆何を慣用表現と取るかは人によって判断が分かれるだろう。そこで、以下のテストをして慣用表現か否かを判断する。

(a) 名詞と動詞の間に他の語句が入りにくいものは慣用表現。

腹が立つ: × 腹が猛烈に立つ → 慣用表現
煙が立つ: ○ 煙がもうもうと立つ → 慣用表現ではない

(b) 名詞が自由に連体修飾を受けることができないものは慣用表現。

目をかける: × 優しい目をかける → 慣用表現
電話をかける: ○ 急ぎの電話をかける → 慣用表現ではない

(c) 名詞と動詞をいれかえて連体修飾の構造にできないものは慣用表現。

手に入れる: × 食べ物を入れた手 → 慣用表現
箱に入れる: ○ 食べ物を入れた箱 → 慣用表現ではない

「する」は実質的な意味が希薄な動詞であるが、原則として慣用表現とは取らない。

楽しみにする → 楽しみ / に / する
気にする → 気 / に / する

3.2 補助動詞

1. 本動詞と同様、原則として語幹と活用語尾に分割する。

- 載 / っ / て / い / る
補動・一段・語幹 活語尾・一段・終止
- 送 / っ / て / ください
補動・特殊・命令
- 読み / 始め / た
補動・一段・語幹
- 考え / も / し / ない
補動・サ変・未然
- お / 書き / に / な / る
補動・五段・語幹 活語尾・五段・終止

- おいでくださる → おいで / くださる
動詞・連用 補動詞

3. 基本形2の {お、ご} + {名詞、動詞・連用形} + {だ、です} の形式で例文の「支払い」「考え」「持ち」を『動詞・連用形』としている。しかし、名詞的な意味を持つか動詞的な意味を持つかで判断すべきで、それは文脈によって決まる。例えば、連体修飾されているなら、名詞的と判断できる。

- いくらお 支払い でございますか → お / 支払 / い
動詞的 接頭 本動・連用 活語尾

参加料のお 支払い でございますか → お / 支払い
名詞的 接頭 普名

- 不可能だとお 考え であれば → お / 考え
動詞的 接頭 本動・連用

それがあなたのお 考え であれば → お / 考え
名詞的 接頭 普名

- 登録用紙をお 持ち でしょうか → お / 持 / ち
動詞的 接頭 本動・連用 活語尾

4. 「その他」の項で挙げられている例文のうち、次のものは疑義がある。

- 当日直接お 支払い ということで

「当日」「直接」という副詞が「お支払い」を修飾しているので、『本動詞・連用形』の方がいい。文自体は「お支払い」のあとの「する」が省略されていると考えられる。

- 会期中の催し等についてのお 問い合わせ ですが

「…についての」によって連体修飾されているので、『普通名詞』の方がいい。

- 通知申し上げます

「申し上げます」は「言う」という本来の意味を持たず、「する」の敬語表現であるので、「通知」は『サ変名詞』「申し上げる」は『補助動詞・語幹 + 活用語尾』である。

4 形容詞

本動詞と同様、原則として語幹と活用語尾に分割する。

- 新しい → 新し / い
形容語幹 語尾・形容・終止

- 良く → 良 / く
形容語幹 語尾・形容・連用

5 副詞

1. 「に」のつく副詞の項で、「とうに」は「とう／に」と分割しているが、この「と(疾)う・く」は古語であって、現代語では単独で使われないので、「とうに」と一語にする。

また、「即座に」も「即座(普通名詞)／に」と分割しているが、「即座」には普通名詞としての用法がないので、「即座に」と一語にする。

一般に、語源はどうかであれ、現代語において単独で使われない語を含む表現は、全体で一語とする方がよいと思う。その方が、音声認識においても構文解析においても都合がよい。

2. 「こう」「そう」「ああ」「どう」などに「して(も)」「やって(も)」が後続するもので、本来の意味を失っているものは副詞とする。

- こうして 座っていると体が楽だ
副詞

- ことは こう / し / て ください
副詞 本動詞 接助

- どうして 来なかったのですか
副詞

- 会議は どうしても 成功させなければならない
副詞

「そこまで」は、文脈によって『副詞』の場合と『代名詞 + 格助詞』の場合がある。

- いくらなんでも そこまで やることはないだろう
副詞

- そこ / まで 行くには一時間かかります
代名詞 格助

以上のほかにも結合が密で分割しても意味のない表現は全体で副詞とする。

- 念のため → 副詞
- もうすぐ → 副詞
- 案に相違して → 副詞

***注**

☆『大辞林』などではもちろん分割している。しかし、全体の意味を要素から構成することはできないので、一語登録の方がよい。ただし、次のような場合は分割する。

○ 駅は もう / すぐ そこだ
副詞 副詞

***注**

☆複合副詞には次のようなものがある。

いくら何でも、いざという時、生まれながら、うやむやのうちに、おおよそのところ、国を挙げて、心ゆくまで、ご多分にもれず、事と次第によっては、どうかすると、取るものも取りあえず、耳をそろえて、もの見事に、

あらかじめ複合副詞を列挙するのはむずかしい。迷う場合は、『大辞林』の連語欄を参照すること。また、新たに複合副詞が出現した場合は、『マニュアル』に追加する。

3. 会話に特有な言い回しで、「すみませんが」「恐れ入りますが」「失礼ですが」のように、文頭にきて文全体を修飾する表現がある。これらは、本来の意味をとどめているものも失っているものもあるが、すべて副詞として一語登録する。

- すみませんが、登録料を至急お支払いください
- 恐れ入りますが、田中先生をお願いします
- 失礼ですが、登録料はもうお支払いでしょうか
- 申し訳ありませんが、登録はもうしめきりました

6 連体詞

1. 「このような」「どのような」は、『日本語形態素解析マニュアル』では、「この(連体詞) / ような(助動詞)」「どの(連体詞) / ような(助動詞)」と分割している。しかし、連体詞が助動詞に係るような規則は構文解析の観点からは好ましくない。

『大辞林』では「このよう」「どのよう」で『形容動詞・語幹』としているが、これも「…はこのようだ」「…はどのようですか」とはあまり言いそうもないので気持ちは悪い。そこで、長単位で『連体詞』とする。

- このような ケースでは
連体詞
- どのような ご用件でしょうか
連体詞

***注**

☆したがって、「このように」「どのように」は全体で『副詞』となる。また、「このくらい」「どのくらい」も全体で『副詞』とする。

- このように してください
副詞
- どのくらい かかりますか
副詞

2. 「こう、そう、ああ、どう」に「いう、いった、した」などがつくものは『日本語形態素解析マニュアル』では分割しているが、すべて一語で『連体詞』とするように改める。

- こうした 親切が時にはあだになる。
連体詞
- ほかには どういう 事が
連体詞
- ほかには どういった 事が
連体詞

3. 『日本語形態素マニュアル』では「このまま」「この間」を『名詞』としている。確かに、名詞として用いられることもあるが、格助詞を伴わずに副詞的に用いられることの方が多。

この他にも、名詞の中には格助詞を伴わずに副詞的に用いられる語がある。そのような場合、構文解析文法では、名詞単独で述語に係るような規則を作る方法と、多品詞語として名詞にも副詞にも登録するという方法がある。曖昧性を排除するという点では後者の方が望ましい。

- このまま 仕事を続ける。
副詞
- このまま で結構です。
名詞
- この間 お尋ねした件ですが。
副詞
- この間 のことなんです。
名詞

7 接続詞

1. 「だ」「です」が文頭にきて全体として接続詞的な働きをする表現は、一語で『接続詞』とする。これは丁寧体でも同じ。

- 値段は安い。 けども、品質が良くない。
接続詞
- 彼にはアリバイがある。 だとすると、犯人は誰だろう？
接続詞
- 応募者は二千人もいます。 ですけれども、参加できるのは千人です。
接続詞
- 賛成多数でございます。 でございますから、本件は可決されました。
接続詞
- 参加資格はありません。 ですからね、どなたでも参加できます。
接続詞

2. 「こう」「そう(そ)」に関しても、以下の形で文頭に現れるものは『接続詞』とする。

- こうして、彼は大統領になった。
接続詞
- そしたら、こうしましょう。
接続詞
- そうしたらね(そうしましたらね)、明日は休みにしましょう。
接続詞

***解説**

☆『日本語形態素解析マニュアル』では、「そう/し/たら」「そう/し/まし/たら」と分割している。また、「でありますから」「でございますから」などの丁寧体も分割することになっている。これを改めて、長単位分割して一語で接続詞とする。

3. 引用の助詞「と」を含むもので、文頭にきて接続詞の働きをするものがある。これも全体を一語として『接続詞』とする。

- というのは、会議で発表したいのです。
接続詞
- なぜかと言いますと、ホテルの予約がとれなかったのです。
接続詞
- だからと言って、割引はできません。
接続詞

***解説**

☆これらを分割すると、構文的に例外的な規則を作らなければならなくなるうえに、意味的にも全体の意味を構成することはできない。そこで、奇異に見えるかもしれないが、以上のようにする。

3. 文末のムードを表す助動詞・補助動詞は、分割せずに全体を一語とし、『助動詞』とする。許可・禁止・当為・提案・意志・概言などの意味がある。

許可

- 送っ てもいい → てもい / い
助動形容語幹 活語尾形容終止

禁止

- 送っ てはいけない → てはいけな / い
助動形容語幹 語尾形容終止
- 送っ てはいけません → てはいけません
助動特殊終止
- 送っ てはだめだ → てはだめだ
助動特殊終止

当為

- 送ら なければならぬ → なければならぬ / い
助動形容語幹 語尾形容終止
- 送ら なければなりません → なければなりません
助動特殊終止
- 送ら なければだめだ → なければだめだ
助動特殊終止
- 送ら なければだめです → なければだめです
助動特殊終止

提案

- 送れ ばいい → ばい / い
助動形容語幹 活語尾形容終止

意志

- 送る つもりです → つもりです
助動特殊終止

概言

- 送る かもしれない → かもしれな / い
助動形容語幹 活語尾形容終止
- 送る かもしれません → かもしれません
助動特殊終止

様相

- 送った ところだ → ところだ
助動特殊終止

***注**

☆「ムード」を表す文末表現には次のようなものがある。

許可: てもよろしい、たっていい、てもかまわない、てもかまいません、て(も)さしつかえない、て(も)さしつかえありません、て(も)けっこうだ、て(も)けっこうです

禁止: てはならない、てはなりません、てはだめです、たらだめだ、たらだめです、ないでください

当為: なければいけない、なければいけません、なくてはいけない、なくてはいけません、なくてはならない、なくてはなりません、ねばならない、ねばなりません、ないといけない、ないといけません、なくてはだめだ、なくてはだめです、べきだ、べきです、ざるをえない、ざるをえません、よりほか(は)ない、よりほか(は)ありません、ものだ、ものです、ことだ、ことです、ほうがいい、わけにはいかない、わけにはいきません、にはおよばない、にはおよびません

提案: ばよい、といい、とよい、たらいい、たらよい

概言: かもわからない、かもわかりません、にちがいない、にちがいありません、はずだ、はずです、とのことだ、とのことです、はずがない、はずがありません

様相: ところです、ばかりだ、ばかりです

以上の他にも、新しくムード表現として認めるものは、そのつど『マニュアル』に追加する。

10 助詞

1. 以下のものを複合格助詞として認める。

「として」「について」「につき」「において」「にとって」「をもって」「に関して」「に対して」「によって」「により」「をにおいて」「に当たって」「に際して」「にわたって」「にかけて」「を通して」「を通じて」「と共に」「と一緒に」「と比べて」「に比べて」「のために」

これらの丁寧体も分割せずに、全体を『格助詞』とする。

「としまして」「につきまして」「におきまして」「にとりまして」「をもちまして」「に関しまして」「に対しまして」「によりまして」「をおきまして」「に当たりますして」「に際しまして」「にわたりますして」「にかけまして」「を通しまして」「を通じまして」「と比べまして」「に比べまして」

以上の他に、新たに『格助詞』としたい連語が出現したときは、そのつど『マニュアル』に追加する。

***解説**

☆『日本語形態素解析マニュアル』では丁寧体は分割することになっているが、固定的に使われるものなので全体を一語とする。

2. 属格を表す「の」を新たに『連体助詞』とし、主格を表す「の」と区別する。

- 太郎の電話 → の
連体助詞
- 太郎の書いた手紙 → の
格助詞

また、次のような連語を『連体助詞』として認める。

「における」「に関する」「による」「に対する」「にわたる」「という」
「としての」「についての」「にとつての」「に当たつての」「に際しての」
「のための」

3. 「って」「とか」を『引用助詞』として認める。

- 出席したいって 言っていました。
- 鈴木さんはロンドンにいる とか 聞いています。

4. 接続助詞の「じゃ」の項の例文を次のものと替える。

- いくら金をためても死ん じゃ 何もならない → 接続助詞

***解説**

☆『日本語形態素解析マニュアル』で『接続助詞』の「じゃ」の項に挙げている例文はまちがいである。「じゃ」には助動詞の場合と格助詞の場合と接続助詞の場合とがあり、例文の「じゃ」は助動詞である。

- わかんないん じゃ ないかって → 助動詞
- やかましくて、ここ じゃ 話もできない → 格助詞

5. いくつかの語で構成される句が、全体として接続助詞のような働きをすることがある。
このような表現は、全体を一語として『接続助詞』とする。

時間的關係

- 家に帰る やいなや
- 時がたつ につれて

順接条件

- もしその説が正しい とすれば
- 知らなかった ものですから

逆接条件

- 反対があった にもかかわらず
- たとえ知って いたとしても

***注**

☆次の連語を『接続助詞』と認める。

時間的關係: とたんに、あげく(に)、ところ(が)、まま(で)、にしたがい、にしたがって

順接条件: とすると、としたら、ものなら、もので、ものですから、だけに、だけあって、せいで、おかげで

逆接条件: としても、としたって、にしても、にしたって、にしろ、にせよ、からといって、といっても、ところで、くせに、わりに、どころか、のに

以上の他にも、新たに『接続助詞』としたい連語が出現したときは、そのつど『マニュアル』に追加する。

11 接尾辞

活用する接尾辞は、原則として語幹と活用語尾に分割する。

● 働きやすい → 働 / き / やす / い
本動詞 活語尾 接尾・語幹 活語尾・終止

● 暑がる → 暑 / が / る
形容詞 接尾・語幹 活語尾・終止

A 日本語の表記

A.1 「出現形」と「標準形」

「出現形」とは、模擬会話を漢字かな交じり文に書き起こすときの形である。「標準形」とは、文字どおりその語の標準的な形である。「出現形」は、音声認識結果の文字化の規範となるべきもので、後の言語処理(形態素解析や構文解析)の入力となるはずのものである。一方、「標準形」はデータベース上で検索や統計処理を行なう際に用いられる。

このように「出現形」と「標準形」は用途が異なるので、それに沿った基準が必要である。以下に、「出現形」と「標準形」の表記基準を示す。

A.2 「出現形」の表記基準

A.2.1 漢字とかなの使い分け

漢字がその語の意味を明瞭に表しているものは漢字で、そうでないものは平がなで表記する。当て字や借字は使わない。ただし、慣用的な表記が定まっているものは慣用に従う。(原則として『NHK用字用語辞典』に従う。ただし、「用例集」の中で<>に示された許容表記がある場合は、それを優先する。)

1. 「名詞」「動詞」「形容詞」は漢字がその語の意味を明瞭に表すことが多い。したがって、原則として漢字表記とする。

<例外>

名詞: あらすじ、しかた

動詞: ございます、さしあげる、できる、なくす、なさる

形容詞: ありがたい、おもしろい

接頭語: 「何」で「ナン」と「ナニ」の両方の読みが可能な場合。「御」で「オ」と読む場合、および「オン」「ゴ」の両方の読みが可能な場合。

形式名詞: こと(常にかな表記)、やつ、はず、ところ

「モノ」は抽象物を指すときは「もの」(あきらめたものと見える)、人を指すときは「者」(参加する者は集まれ)、人以外の具体物を指すときは「物」(いい物をあげよう)。

「固有名詞」については、現地(あるいは当人)の表記を原則とする。有名人の表記はNHK(あるいは朝日新聞)の表記と合わせる。一般に、中国名は漢字、韓国を含め外国名はカタカナ表記である。JIS第2水準にもない場合は「カタカナ」で代用する。

2. 「感動詞」「連体詞」「接統詞」「代名詞」「助動詞」「補助動詞」「助詞」は漢字があまり意味を持たないことが多い。したがって、原則としてかな表記とする。

<例外>

感動詞: 分かりました、申し訳ありません

連体詞: 大きな、小さな、当初の

接続詞: 実は

代名詞: 皆

補助動詞: 願う、願える、申し上げる
(「～てほしい」はかな表記)

助詞: に関して、に対して

3. 「数詞」は原則として漢数字表記とする。

<例>

三千二百五十七 (金額などの場合)

六の五四九零 (電話番号などの場合)

(注: 「ゼロ」「レイ」は「零」と表記する。)

4. 「副詞」は名詞起源でその元の意味が残っているものは漢字表記、そうでないものはかな表記とする。(あまり使われない副詞は「名詞起源でその元の意味が残っているもの」が多いと推測される。反対によく使われるものほど、「元の意味」から外れてきているものが多い。)

作業による揺れを防ぐために、以下に現在確定しているものを列挙する。

<かな表記をとるもの>

あいにく あまり あらかじめ いくら いっぱい いろいろ おそ
らく さらに すぐ(に) ぜひ だいたい たくさん ただ たぶ
ん ちょうど ちょっと ともに なるほど ほとんど まず まっ
たく むしろ もし もちろん よろしく

<漢字表記をとるもの>

以前 一応 一番 お気の毒ですが 必ず 最低 再度 先ほど 早
速 至急(に) 失礼ですが 少々 少し 既に 全然 大変 多少
直接 当然 特に 何分 何とか 何と 何ら 後ほど 初めて 別
に 前もって 万が一

5. 「同形異音語」はかな表記とし、「読み」の揺れを防ぐ。

<例>

昨日: さくじつ きのう

今日: きょう こんにち

明日: あす あした みょうにち

私: わたし わたくし

ただし、十分(じゅうぶん/じゅっぶん)や方(かた/ほう)などは、文脈によって判断できる場合、漢字表記とする。一意に決められない可能性がある場合はかな表記とする。

6. 「読み」の情報は音声処理にとって重要なので、漢字表記のもので「読み」が二つ以上あるものについては、必ずく>で読みを付与する。

<例>

研究所: < kenkyuusyo >
< kenkyuujyo >

A.2.2 送りがな

送りがなについては『NHK 用字用語辞典』に準拠する。『NHK 用字用語辞典』にない語のときは『大辞林』に準拠する。

1. 活用のある語は活用語尾を送る。ただし、一段活用の動詞は「-i」「-e」の部分を含んで送る。また、語幹が「し」で終わる形容詞は「し」から送る。

言う
過ぎる
当てる
長い
早い
忙しい
涼しい

2. 活用形によって同形異義になることのあるものは、互いに区別できるように送る。

行(い)った
行(おこ)なった (×行った)

3. 活用語尾の前に「か」「やか」「らか」を含む形容名詞は、その音節から送る。

暖かだ
穏やかだ
滑らかだ

4. 他の語から派生した語は、漢字の部分の読みを変えないで送りがなをつける。

済ます (← 済む)
確かめる (← 確かだ)
重たい (← 重い)

5. 複合語は、複合語を構成する各語の送り仮名を送る。ただし、特定の領域で慣用が固定しているものは、慣用に従って送り仮名をつけない。

手書き (×手書)
書き方 (×書方)
振り込み (×振込 「振込み」は許容)
肩書 (×肩書き)
銀行振込 (×銀行振り込み ×銀行振込み)

A.2.3 外来語・外国語固有名詞

外来語や外国語固有名詞は『朝日新聞の用語の手引き』に準拠する。

<使えない表記>

ツア:	モーツァルト	→	モーツァルト
ツイ:	ソルジェニーツィン	→	ソルジェニーチン
イエ:	イエルサレム	→	エルサレム
クア:	クアルテット	→	カルテット
クイ:	クインテット	→	クインテット
クエ:	クエート	→	クエート
クオ:	クオーク	→	クオーク
グア:	グアテマラ	→	グアテマラ、ガテマラ
ヴ:	ヴラマンク	→	ブラマンク
ヴァ:	ヴァイオリン	→	バイオリン
ヴィ:	ヴィーナス	→	ビーナス
ヴェ:	ヴェルサイユ	→	ベルサイユ
ヴォ:	ヴォルテール	→	ボルテール
ヴュ:	インタヴュー	→	インタビュー
トゥ:	トゥールーズ	→	ツールーズ
ドゥ:	ヒンドゥー教	→	ヒンズー教

<使ってもよい表記>

ウイ:	ウイスキー
ウエ:	ウエールズ
ウオ:	ウォール街
ツエ:	フィレンツェ
ツオ:	カンツォーネ
テイ:	ティーパーティー
テュ:	チュービンゲン
デイ:	ビルディング
デュ:	プロデューサー
ファ:	ファックス
フィ:	フィルム
フェ:	エッフェル塔
フォ:	フォーク
フュ:	フュージョン
チェ:	チェコ
ジェ:	ジェット

A.2.4 その他

相手の割り込み(相づち等)は { }、問投詞は []、言い直し・言い淀みは () でくくる。

先日ですね、{はい} そちらの [えーと] イタリア・ルネサンス・ツアーに、
(も、も) 申し込んだ者なんですが。

A.3 「標準形」の表記

「同形異音語」に対する配慮を除いて、「出現形」の表記基準と同じとする。すなわち、「私」「方」「今日」「昨日」「明日」「十分」「何」などはすべて漢字表記とする。ただし、「できるかぎり」以下の原則を守ることとする。

- 原則として「出現形」の表記基準に記述がないもの(あるいは曖昧なもの)については、『NHK用字用語辞典』準拠とする。
- ただし、『NHK用字用語辞典』ではひらがな書きが多くなりすぎる傾向があるので、「名詞」「動詞」「形容詞」類ではむしろ、「表記の許容」(< > で示されている)の方を標準表記とする。
- 『NHK用字用語辞典』にない語については前述の基準によるが、漢字を用いる場合、原則として

常用漢字 + 亀(き) 舷(げん) 痕(こん) 挫(ざ) 哨(しょう) 狙(そ)
- 謁 虞 箇 且 遵 但 脹 朕 附 又 濫

とする。このため、「混ぜ書き」となることもやむをえない。

金ぱく 啓もう 愛がん 相づち 論ばく てい鉄 はり紙

ただし、「ひらかな一字+漢字(列)」のような「混ぜ書き」は読み誤りの原因となりやすいので、できるだけ避ける。

う回 → 迂回 う匠 → 鵜匠 禅え → 禅衣
が解 → 瓦解 安ど → 安堵

ただし、この場合、漢字化される方はJIS第1水準内の漢字とする。

参考文献

- [1] 日本放送出版協会: 「新用字用語辞典」 (1986).
- [2] 朝日新聞社: 「朝日新聞の用語の手引き」 (1989).
- [3] 東洋情報システム: 「日本語形態素解析マニュアル」 (1993).
- [4] 益岡隆志・田窪行則: 「基礎日本語文法」, くろしお出版 (1991).
- [5] 森田良行・松木正恵: 「日本語表現文型」, アルク (1989).
- [6] 村木新次郎: 「日本語動詞の諸相」, ひつじ書房 (1991).