公開

002

TR-IT-0053

音声言語データベースにおける 日本語形態素解析マニュアルの補遺

The supplementary volume of
"Users Manual for Japanese Morphological analysis"
in the ATR Spoken Language Database

浦谷 則好　田代 敏久　森田 千帆 *
Noriyoshi URATANI　Toshihisa TASHIRO　Chiho MORITA

1994.5

内容概要

先に日本語形態素解析マニュアルをテクニカルレポート（TR-IT-0009）の形で報告しているが、これはその補遺である。

ATR 音声翻訳通信研究所
ATR Interpreting Telecommunications Research Laboratories

© (株)ATR 音声翻訳通信研究所　1994
© 1994 by ATR Interpreting Telecommunications Research Laboratories

* 株式会社 東洋情報システム
TOYO Information Systems CO., LTD.

目次

1 名詞 1
1.1 普通名詞 2
1.2 数詞 4
1.3 住所名 5

2 副詞 6

3 感動詞 9

4 助動詞 10

5 助詞 12
5.1 格助詞 12
5.2 副助詞 13
5.3 連体助詞 14
5.4 引用助詞 15

6 接頭辞 16

7 接尾辞 19

8 その他 24

はじめに

ＡＴＲ音声翻訳通信研究所では音声言語データベースを構築中であり、すでにそのための日本語形態素解析マニュアルもテクニカルレポート（TR-IT-0009）の形で報告している。これは音声言語データベースのみならず今後予定されている言語データベースや音声データベースにおける形態素情報の基準として位置付けられている。また、おそらくＡＴＲ音声翻訳通信研究所で今後作成される言語処理プログラムにおいても基準として使われることが予想されている。しかしながら、先のマニュアルでは形態素の長さや品詞の判定に不十分のところがあった。この報告書はそれを補完するものである。

なお、形態素辞書の構築を現在進めており、完成後は形態素の長さの認定をそれに基づいて行なうため、作業者による判断基準の揺れが少なくなると期待している。

1　名詞

名詞類に認定においては、いくつか追加条項があったので、ここで改めて詳しく立項していくことにする。

音声言語データベースにおける日本語形態素解析マニュアル（P．1）

自立語。活用しない。格助詞「ガ」を付けて文の主語になる。
固有名詞、サ変名詞、形容名詞、普通名詞、数詞、代名詞、人名、住所名、日時から構成されている。

↓

名詞の定義（新）

- 「普通名詞（サ変名詞）＋普通名詞（サ変名詞）」か1語とするかは「大辞林」に立項されているかどうかを大前提とする。
- 単独で意味をなす長さで一語と認定し、切った結果元の意味が変わってしまう場合は分割できない。
- 「単独で意味をなす長さ」とは、それで主語 and/or 目的語になれる。
  （「が」ないし「を」が接続する）ことを意味する。
- 普通名詞（または固有名詞）＋接尾辞で「大辞林」の見出しにあるものは1語の普通名詞（または固有名詞）とする。
- ある語がａ＋ｂに分割できそうな時、普通名詞（サ変名詞）であることが確かでない方が（普通名詞やサ変名詞でなく）接頭辞（接尾辞）でないかぎり1語とする。＜例：来場所、開催地＞

切った結果元の意味が変わってしまうものとしては、”英語”がある。
”英語”は”英”と”語”に切っても、それぞれ意味をなしているように見えるが、「独が英を攻めた」の「英」と”英語”の「英」は同一とは考えられるが、「語は文字からなる」の「語」と”英語”の「語」は同一とは考えられない。
前の「語」の語義は「単語」であり、後の「語」の語義は「言語」である。
「言語」の意味では「語」は造語成分としてしか（つまり単独では）使用できない。
したがって、”英語”は”英”と”語”に分割できず、1語である。

1.1　普通名詞

音声言語データベースにおける日本語形態素解析マニュアル（P．6）

広義には、同一種類の事物に通じて用いられる名称。狭義には、そのうち「サ変名詞」「形容名詞」を除いたもの。ここでは狭義の意味で用いる。
固有名詞及びサ変名詞は長単位分割することを基本とするが、普通名詞の場合は（複合語は）短単位分割することを基本とする。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）

↓

普通名詞の定義（新）

ここでの普通名詞とは、1ページの「名詞の定義（新）」で定義された名詞から固有名詞、サ変名詞、形容名詞を除いたものである。また、それ以外にも、以下の普通名詞を認める。

※ 大辞林の見出し語で、名詞の品詞がついているもの。

※ 動詞から転成したもの。

例 お近付き、申し込み

※ 接頭辞に接尾辞が直接ついているもの。

例 何回　両人

※ 「1字の普通名詞（固有名詞）＋接尾辞」となるものでよく使われるものは合わせて1つの普通名詞とする。

例 会費　会員　男性　人類　年末　船便

＜他の品詞と紛らわしい普通名詞の例＞

- 電話番号 （サ変名詞＋普通名詞 → 普通名詞（「大辞林」エントリーされている））
- 「第三者」は「者」に「～する人」という意味がないので、「者」を接尾語と認定できない。したがってこれは一語で普通名詞とする。
- プロフェッサー・マーフィー様にご講演（タイトル名は普通名詞とする）
- 写真とフィルムは別添えでお送り致します。
  （別（普通名詞）／添え（普通名詞）／で（格助詞）とする）
- 彼らにとっても、国内外の人と話すのは、（1語で普通名詞）
- わが社が開催致します。（1語で普通名詞）
  ＜「貴社」、「弊社」も1語で「大辞林」にある。＞

- リスト作りの件ですが、（1語で普通名詞）
- 会議終了後、資料をお送り下さい。（接尾辞ではなく普通名詞）
- 他の宿泊地の情報にかんしましては、ここにお送り下さい。（一語で普通名詞）
- 挨拶文が、まだ届いていません。（一語で普通名詞）
- 「しだい」が活用語の連用形につく時は、接続助詞となる。
  「しだい」の後ろに格助詞類の助詞が付くと、普通名詞となる。
  - ファックスを送り次第、返事をください。（接続助詞）
  - ～しだいの(普通名詞)
  - ～しだいが(普通名詞)
- 普通名詞につく「度」は切らずに前の普通名詞につけたまま一語の普通名詞とする。

  例 来年度の予算は減りそうです。

1.2　数詞

数詞は部屋番号などを読み上げる際に「０」を「まる」と発声することがあるので、新たにそれに対応した読み仮名をふれるように数詞の読み方の表を変更する。

音声言語データベースにおける日本語形態素解析マニュアル（P．８）

	零	一	二	三	四	五	六	七	八	九	十
1漢語	れい	いち	に	さん	し	ご	ろく	しち	はち	きゅう	じゅう
2和語		ひ	ふ	み	よ（ん）	いつ	む	なな	や	ここの	と（お）
3欧米語	ぜろ										

↓

数詞の読み方（新）

	○	零	一	二	三	四	五	六	七	八	九	十
1漢語		レイ	イチ	ニ	サン	シ	ゴ	ロク	シチ	ハチ	キュウ	ジュウ
2和語			ヒ	フ	ミ	ヨ（ン）	イツ	ム	ナナ	ヤ	ココノ	ト（オ）
3欧米語		ゼロ										
4その他	マル											

1.3　住所名

音声言語データベースにおける日本語形態素解析マニュアル（P．１２）

住所表示に使われる地名。短単位分割を基準とし、「都」「府」「県」「市」「区」などは、接尾辞とする。

住所名の認定（新）

住所名とは行政単位の「県」「市」「郡」「町」「区」などを接尾とし、その前の地名のことを指している。したがって、行政単位でなくて、「町」がついているものは分割できない。
<例>（大阪市北区）茶屋町（六丁目）→「茶屋町」は全体で住所名。

※ 例えば、東京の場合、（港区）赤坂（一丁目）となるが、この「赤坂」と「茶屋町」は同列に扱われるべきものである。

（行政単位が有るかどうかで判断する）

郡の後の町は → 切る
市の後の町は → 切らない

2　副詞

音声言語データベースにおける日本語形態素解析マニュアル（P．２３）

1. 「副詞」の正規表現について
「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。今後かな表記を採用するものが出てきた場合はマニュアルに追加していくこととする。

とあるので、以下に最新の副詞のリストを挙げておく。

＜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　だんだん　ちょうど　ちょっと　ともに　なかなか　なるほど　ほとんど　まず　まだ　まったく　むしろ　もし　もちろん　よろしく　いつも　すべて　よく　たいそう　また　どうにか　とても　いつ　そうとう　かなり　そんなに　みんな　いつ　いつでも　いつまで　いかが　あと　どれほど　うまく　いつでも　やはり　ぜひとも　ぜひに　いくらか　とうに

＜漢字表記をとるもの＞

以前　一応　一番　一斉　因縁　お気の毒ですが　必ず　結構　最低　再度　先ほど　早速　至急（に）　失礼ですが　十分　少々　少し　既に　全然　大変　多少　例えば　直接　当然　特に　何分　何とか　何とも　何ら　後ほど　初めて　早くも　普通　別に　前もって　万が一　約　一切　確か　一度　幾分　改めて　昔ながら　極力　比較的　随分　心より　最近　本当　別途　逐一　先頃　共に　優に　常に　随時　主に　誠に　今後とも　常時　おって　今後とも　つい最近　早速に　実際　大分　即座に

音声言語データベースにおける日本語形態素解析マニュアル（P．２４）

4. 複合副詞について

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。

・

・

・

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。

こちらも、新たなものが現れた場合はマニュアルに追加するということなので、以下に最新のリストを挙げておく。

＜複合副詞の一覧＞

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　もう少し　ものの見事に　念のため　もうすぐ　案に相違して　もうちょっと　どうしても　すくなくもとも　もしかすると　お気の毒ですが　できるだけ

音声言語データベースにおける日本語形態素解析マニュアル（P．２４）

5. 特有な表現

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。

- すいませんが、登録料を至急お支払下さい。
- 恐れ入りますが、田中先生お願いします。
- 失礼ですが、登録料はもうお支払でしょうか。
- 申し訳ありませんが、登録はもう締切ました。

＜追加事項＞

- 特有な表現における副詞については以上のようになっているが、「お尋ねしたいのですが この道はどこに続いていますかの「お尋ねしたいのですが」を副詞（文修飾の副詞）と認定するのは問

題無いが、「ちょっとお尋ねしたいのですが。」の「お尋ねしたいのですが」を副詞（文修飾の副詞）と認定するのは問題である。なぜならこの文では「お尋ねしたい」は主節の述部であるからである。

したがって「お（接頭辞）／尋ね（本動詞）／し（補助動詞）／た（助動詞）／い（語尾）」とる。

同じように「お尋ねしたいのですが。」と「。」で終っている場合も「お／尋ね／し／た／い／の／で／す／が／。」とする。

- 「どうぞ」の品詞の判定は、「どうぞ」が後続する節や文にかかっている場合は副詞とし、後続する節や文にかかっていない場合は感動詞とする。

  （単なる挨拶の場合は、もちろん感動詞とする）

- 「感動詞＋接続助詞」の形で出てくるものは副詞として認める。リストにないもので、このようなパターンで現れた語はリストに追加していくこととする。

  例 会議では申し訳ございませんが、協賛団体の〜。

  （この場合、「申し訳ございませんが」は文頭ではないが、「会議では」が「申し訳」にかかっていないので、単なる挿入と判断して良い。）

3　感動詞

＜追加事項＞

- 以下の例の下線部分を感動詞として認める。

  - それは どうもありがとうございます。
  - それは どうも。
    （どうもで終っている時は、感動詞とする）
  - おかげさまで、順調です。
  - では、 これで。

4　助動詞

音声言語データベースにおける日本語形態素解析マニュアル（P．３１）

2. ムードを表す助動詞について

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。

ムードを表す助動詞を長単位の助動詞とし、長単位の助動詞と認められたものについてはマニュアルに追加していくものとする。

以下に現在の最新の長単位の助動詞のリストを挙げておく。

＜長単位の助動詞一覧＞

- 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　たらけっこうです　たらけっこうだ　ていい　てよい　てもよい　てよろしい　はけっこうだ　ばよろしい　でもかまいません

- 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　てはいけない

- 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　ばいい　ほうがよい　ほうがよろしい

- 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

ところだ　ところです　ばかりだ　ばかりです　てはおりません　てはございません　つもりだ　ている　ておりません　てございません

＜追加事項＞

- 長単位で１語の助動詞とするべきか分割して名詞＋αとすべきか迷う語については、前に用言類がある場合は助動詞とし、それ以外は分割して名詞＋αとする。

＜例＞ 助動詞か名詞＋αか？

言った<u>ほうがよい</u>と思うよ → 助動詞

早めの<u>方がよい</u>と思います → 普通名詞（ほう）＋格助詞（が）＋形容詞・語幹（よ）＋語尾（い）

「～ません」という形の長単位の助動詞は語尾をとらずに１語とし、活用形を「なし」とする。

- 「て（接続助詞）＋補助動詞」の形で出てくるものは長単位の助動詞として認める。リストにないもので、このようなパターンで現れた語はリストに追加していくこととする。

- 接続助詞か助動詞か迷うものとして「なら」があるが、これは「なら」の前に「の」があるかどうかで判断することとする。

  - される<u>なら</u>～。（一語で接続助詞）
  - されるの<u>なら</u>～。（助動詞「だ」の仮定形）

5　助詞

5.1　格助詞

音声言語データベースにおける日本語形態素解析マニュアル（P．３５）

＊注
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。

以下に最新の格助詞の一覧を挙げておく。

＜格助詞一覧＞

から　が　って　で　と　に　の　へ　まで　より　を　にて

＜複合格助詞一覧＞

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　にしろ　にせよ　でもって　をもって　をして　にかけて　にて　に際し　のために　のため　の際　の際に

＜複合格助詞（丁寧体）一覧＞

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　に当たりまして　に際しまして　にわたりまして　にかけまして　を通しまして　を通じまして　と比べまして　といたしまして　にかけまして

5.2　副助詞

＜副助詞「でも」と格助詞「で」＋係助詞「も」の認定基準＞

- 「だけ」に後続する場合は副助詞。

  ＜例＞ ひとめだけ でも

- 不特定を表す語に後続する場合は副助詞。

  ＜例＞ 何 でも、どこ でも、誰と でも。

- 軽い例示の場合は副助詞。

  ＜例＞ お茶 でも いかが

- 極端な例を挙げて他を類推させる場合は副助詞とする。

  ＜例＞ 子供 でも わかることです。

- 副助詞「でも」と、格助詞「で」＋係助詞「も」との差異

  ※「でさえも」と言い換えても意味が不変な時には副助詞とする。
  これに対して「ドル でも いいですか」や「小切手 でも お支払いできますか」の「でも」は「でさえも」の意味が付加されているとは考え難い。むしろ「ドル で いいですか」や「小切手 で お支払いできますか」の「で」に「も」（係助詞）が付いていると思う方が自然であると思われる。一般に道具や場所、手段に接続する場合には「で」＋「も」と認定する。（迷う場合には「も」をとっても意味が通じれば格助詞「で」＋係助詞「も」と認定する方を優先する。）

5.3 連体助詞

＜追加事項＞

- 前の語（名詞類）が後の語を連体修飾する時の間の「の」や「という」などをこう認定する。したがって、連体修飾されるもの（後の語）も名詞類でなければならない。

- つまり、連体助詞は名詞（類）（一部の副詞「たくさん」なども含む）と名詞（類）をつなぐ品詞である。

  したがって連体助詞の後は（大分類の）名詞か準体助詞か接頭辞しか考えられない。

  「大坂城の近くという と．．．」のばあい、「という」の後の「と」は接続助詞なので、「という」は連体助詞ではない。この場合の「という」は「と（格助詞）／いう（本動詞）」となる。

  「それでも と言う のでしたら．．．」のばあい、「という」の後の「の」は準体助詞なので、良さそうだが、「それでも」が「の」を修飾しているわけではない。（「それでも」は接続詞）。したがって、「と（引用助詞）／言う（本動詞）」となる。（ここで「と」が引用助詞となるのは、この文は「それでも（私は構わない）とでも言うのでしたら．．．」の（）が省略されてると考えられるからで、引用助詞と判断するのが難しければ、格助詞でも構わない）

  ようするに、「と 言う」の「と」が何かを引用している場合、（この場合引用している節が省略されていることもある）引用助詞になるので、「という」の連体助詞は成立しないことになる。

- 電話番号を呼ぶ場合の「五四の七六三二」などの「の」も連体助詞とする。

  ※ 例えば、”支払いという事で．．．”は”支払い（サ変名詞／という（連体助詞）／事（普通名詞）／で（格助詞又は接続助詞）．．．”ということになる。

  また「事」は「形式名詞」であるが、そういう品詞は設定していないので「普通名詞」とする。なお、助動詞「ことだ」は”～に話すことだ”のように動詞に接続する。

  サ変名詞に接続する「する」、名詞類に接続する「だ」「です」などの特殊なものを除いて一般に助動詞や補助動詞は用言類（動詞、形容詞、補助動詞、助動詞）に接続する。

- 「の」には、連体助詞、準体助詞、格助詞、、並列助詞、終助詞の５つがあることに注意。

  - 私の本（連体助詞）
  - 本を読んだのは（準体助詞）
  - 私の書いた本（格助詞）
  - よろこんだ の、喜ばない の といったら（並立助詞）
  - 君がやったの。（終助詞）

- 「の」

  格助詞の「の」と連体助詞の「の」との見分け方は、

  『「が」や「を」と入れ換えられるものは格助詞と考える』

  とする。

5.4　引用助詞

＜追加事項＞

- 節をとる場合だけ引用助詞とする。（名詞類をとるときは格助詞とする）

  すぐ帰る と 言っていました。　→　引用助詞
  武蔵 と 呼んでいた　→　格助詞

  ※ 節に準体助詞の「の」または形式名詞が付属した後の「と」も引用助詞とする。

  例 そういうの と 判断した。
  例 そういうこと と わかっていても。
  例 ～すること と 命ぜられた。

- 「～と思う」の「と」は無条件で引用助詞とする。（「と（引用助詞）／思（本動詞）／う（語尾）」）

  （「～」の部分は節か節の省略形としか考えられない）

- 「～かどうか」も引用助詞とする。

  例 行って戴ける かどうか 教えて下さい。

- 「か否か」も同様に引用助詞とする。

  例 できる か否か、教えて下さい。

- 「か」を含む助詞について

  - どんな構成になっている のか 分らない。（「のか」で引用助詞）
  - その資料で十分 かと 存じます。（「かと」で引用助詞（用言・形容名詞に接続する場合））
  - その置物は猫 かと 存じます。（「かと」は副助詞（体言に接続する場合））
  - 彼の話は本当 かも 分らずに、（「かも」は副助詞）

- 「か」が「かどうか」「のか」などに言い替えることができる時、「か」は終助詞ではなく引用助詞とする。

  例 どの分科会に参加する か、決めたいのです。（引用助詞）

- 「終助詞＋格助詞」は一語で引用助詞とする。

  例 自分の家に泊まらない かと、申しでて

6　接頭辞

音声言語データベースにおける日本語形態素解析マニュアル（P．４５）

認定基準
他の基本となる構成要素に前接して構成する造語成分

（例）お－、ご－〔御〕、ま－〔真〕、新－、大－、総－、不－、無－、とり－〔取〕、ひき－〔引〕

1. ｛お、ご｝のような接頭辞は辞書にひとまとまりでエントリーしていても切って意味がとおるものは切っている。

● お金 → お（接頭） ／ 金（普名）

● 不明瞭な点 → 不（接頭） ／ 明瞭（形名） ／ な（助動）

2. 切ると意味がなくなるものは、一つで名詞とする。

● 小雨（名詞）が降ってきた

↓

接頭辞の定義（新）

接頭時は一覧に挙げられたもの以外は認めない。
ただし、接頭辞に直接接尾辞がつくものは接頭接尾を合わせて１語の普通名詞とする。

＜例＞ 何人、何回、全線、数センチ。

＜接頭辞一覧＞

名詞類に先行するもの		数詞に先行するもの		用言類に先行するもの	
御	住所	御	一通	真	黒い
お	名前	お	一人	お	聞かせする
各	講師	第	三回	ぶっ	倒す
不	参加	幾	千年		
非	公式	各	四名		
未	記入	全	五巻		
全	予算	数	万円		
真	犯人	約	四百人		
大	ホール	何	千人		
中	規模	ほぼ[1]	六合		
小	家族	およそ[1]	七人		
毎	日曜	おおむね[1]	六割		
最	先端	大体[1]	九本		
再	入会				
高	画質				
低	次元				
同	機種				
他	業種				
諸	制度				
貴	学会				
正	会員				
準	会員				
副	首相				
純	文学				
新	製品				
現	時点				
丸	裸				
素	裸				
長	時間				
短	時間				
助	教授				
旧	体勢				
本	会議				
超	ＬＳＩ				
総	収入				
悪	趣味				
重	工業				
軽	犯罪				

名詞類に先行するもの		数詞に先行するもの		用言類に先行するもの	
初	対面				
前	市長				
元	議員				
無	理解				
両	外人				
反	体制				
何	種類				

＊ 名詞類に先行する接頭辞が固有名詞に接続している場合は、接頭辞と固有名詞に分割せず、一語で固有名詞とする。

ただし、以下のような場合は接頭辞と固有名詞に分割する。

例 各ＡＴＲ職員 → 各（接頭辞）／ＡＴＲ（固有名詞）／職員（普通名詞）

[1]数字を修飾している場合のみとする。

7　接尾辞

音声言語データベースにおける日本語形態素解析マニュアル（P．４６）

認定基準

他の基本となる構成要素に後接して単語を構成する造語成分。活用する接尾辞と、活用しない接尾辞がある。活用する接尾辞には活用の種類を入れる。

1. 活用しないもの

・

・

2. 活用するもの

原則として語幹と活用語尾に分割する。

・

・

3. 接尾辞がついて他の品詞を構成するもの大辞林に見出し語もしくは、派生語でエントリーされていれば他品詞を構成するが、エントリーされていない時は切る。

・

・

4. 「箇」の正規表現について接尾辞として使う「箇」は正規表現をひらがなとする。

・

・

↓

接尾辞の定義（新）

接尾辞は一覧に挙げられたもの以外（貨幣などの単位は除く）は認めない。

<接尾辞一覧>

名詞類に付くもの		数詞に付くもの		用言類に付くもの		接尾辞に付くもの	
鈴木	様[8]	三	日[1]	静か	さ	九日	目
鈴木	さん[8]	三	円	うれし	げ	三箇	月
鈴木	氏[8]	三	者	遅	過ぎ（る）	四食	分
鈴木	ちゃん[8]	三	部	暑	がる	三人	とも
鈴木	殿[8]	三	店	ほめっ	放し	四枚	中
鈴木	君[8]	三	面	言いたい	放題	参加者	用
参加	者	三	号			六語	内
金銭	面	三	代（目）			六語	外
弁慶	号[8]	（第）三	便			二年	生
ホテル	代	三	点			七年	型
往復	便[8]	三	誌			四時	着
注意	点	三	室			四時	発
普及	版[8]	三	版			三日	後
都会	式[8]	三	式			三日	前
国際	線[82]	三	線			三日	間
研究	棟	三	棟			一日	当たり
最終	回	三	車			一週間	置き
ゴルフ	コース[8]	三	回			五日	弱
絵画	展	三	コース			二人	連
東京	行き[8]	三	通り			三階	建
保護	色<しょく>[3]	三	行			五割	引き
国際	級[8]	三	品			五割	増し
実施	案	三	紙			五割	減
佐藤	あて[8]	三	級			四人	足らず
先生	方	三	案			四人	あまり
両方	とも	三	年			四人	強
期間	中	三	か（月）			（3）号	線
交換	用[8]	三	食			3か	所
ホテル	内[8]	三	枚			講演者	名
高速	型	三	時間			一人	頭
私	ども	三	泊				付き
参加	費	三	人				様
可能	性	三	分				
自分	たち	三	名				
学生	ら	三	つ				
晩秋	頃	三	語				
手続	上	三	月				
比較	的[4]	三	週				

名詞類に付くもの		数詞に付くもの		用言類に付くもの		接尾辞に付くもの	
朝食	付き	三	度				
機器	類	三	階				
世紀	末	三	件				
入金	済み	三	番				
消防	団	三	割				
簡単	化[4]	三	歳				
団体	別	三	国				
東洋	風[8]	三	台				
中華	街	三	層				
実物	大	三	本				
君主	制	三	冊				
本革	製[8]	三	口				
範囲	外	三	次				
世間	並み[8]	三	通				
合格	圏	三	センチ				
銀河	系[8]	三	秒				
街道	沿い[8]	三	パーセント				
自己	流[8]	三	丁目				
指導	力	三	丁				
会計	係	三	組				
製造	業	三	社				
国家主義	寄り[8]	三	間[5]				
主流	派[8]	三	期				
暴走	族[8]	三	色				
日照	権	三	手				
ユリ	科	三	番地				
伏し目	がち	三	メーター				
支店	長	三	字				
参加	料	三	個				
政府	側[8]	三	桁				
入会	金	三	遍				
救済	策	三	重				
観念	論	三	周				
秘書	課	三	戸				
ロケット	形[6]	三	列				
営業	部[7]	三	倍				
大阪	発[8]	三	島				
ニューヨーク	着[8]	三	段				
青山	通り[8]	三	掛け				
		三	球				
		三	勝				

名詞類に付くもの		数詞に付くもの		用言類に付くもの		接尾辞に付くもの	
		三	錠				
		三	着				
		三	票				
		三	株				
		三	軒				
		三	校				
		三	首				
		三	足				
		三	反				
		三	坪				
		三	党				
		三	頭				
		三	敗				
		三	発				
		三	ページ				

＊ 数詞の後につく接尾辞を接尾につく接尾辞として認める。

例 三千回（三（数詞）／千（接尾辞）／回（接尾辞））

＊ 名詞類に後続する接尾辞が固有名詞に接続している場合は、固有名詞と接尾辞に分割したりせずに、一語で固有名詞とする[8]。

＊ 数詞に付く接尾辞が名詞類（数詞を除く）についたときは接尾と認めない。

例 新聞誌（一語で普通名詞）

- 「部屋割り」は全体で普通名詞。
- 「～次第」は普通名詞（形式名詞）。

- 長さ、重さ、貨幣などの単位はすべて接尾辞とする。
（センチメートル、キログラムなどは切らずに１語とする。）

例 メートル センチメートル キログラム トン フラン 尺 リットル 平方メートル ドル

- 「１字の普通名詞（固有名詞）＋接尾辞」となるものでよく使われるもの（大辞林の見出し語など）は合わせて１つの普通名詞とする。

例 会費 会員 男性 人類 年末 船便

---

[1]意味的に～日間を現している場合のみ。
[2]交通網に限る。
[3]黄色、青色、褐色など色目を現しているものはは切らない。
[4]動詞化等で品詞を変えてしまうもの。
[5]長さの単位の場合
[6]幾何図形以外。
[7]学部名や学科名の場合は全体を一語の普通名詞とする。
[8]固有名詞につく接尾辞になりうる。

<その他例外事項>

- 「十、百、千、万、億」、などの接尾辞は数詞に接続する接尾辞であるが、「十二回」「百円」「数千枚」など、数詞に接続していない場合の品詞は接尾辞とはせずに、数詞とする。

  - 「十」が接尾辞になる場合。

表層	品詞
三	数詞
十	接尾辞
回	接尾辞

  - 「十」が数詞になる場合。

表層	品詞
十	数詞
二	数詞
回	接尾辞

表層	品詞
数	接頭辞
十	数詞
回	接尾辞

- 住所名などに後続する特別な接尾辞は以下のものとする。

住所名	接尾辞
東京	都
大阪	府
神奈川	県
相楽	郡
京都	市
世田谷	区
四条	町[8]
日の出	村

＊「北海道」は一語で住所名とする。

[8]行政単位でなくて「町」がついているものは分割することができないので、「〇〇町」全体で一つの住所名とする。

8　その他

＜追加事項＞

- 何かありましたらお電話 ください。（補助動詞とする）
- 五時 半 でいかがでしょうか。（日時とする）
- 会議の内容とちがうのですね。 いずれにせよ、うまくいきますよ。（接続詞）
- 到着便名が分れば、直接空港へも お迎え／に／上が／り ます。
  （サ変名詞／格助詞／本動詞／語尾）
- 心強いお返事 頂き まして（補助動詞＋語尾 → 本動詞＋語尾）
  ※「心強いお返事」が名詞句なので、「頂く」は本動詞に後続する補助動詞ではなく、本動詞となる。
- ．．．配布する の／か．．．配布する か です
  （「の」＝準体助詞、「か」＝並立助詞）
- 「χ泊できる」の「でき」を補助動詞の語幹とし、この場合の「χ泊」は切らずに一語でサ変名詞とする。
- 「いくら」が「音声言語データベースにおける日本語形態素解析マニュアル」に代名詞としてエントリーされているため、「いくつ」も普通名詞ではなく一語で代名詞とする。

  例 いくつ かの事を教えて頂きたい。
- 格助詞の「の」と連体助詞の「の」見分け方は、「が」や「を」と入れ換えられる「の」は格助詞と考える。
- 「で」が格助詞であるか、助動詞「だ」の連用形であるかの判定は、『「で」を「でも」や「では」に置き換えたとき文として成り立たなくなるならば格助詞ではない（助動詞の可能性が強い）』とする。
- 以下の例では「○」は格助詞としてとり、「×」は（ ）のように分割する。

  ○ 職場 を通じて 知った
  ○ 事務局 といたしまして は
  × 開発者 を中心に 議論する（を／中心／に）
  × 新聞 によりますと 火事は（に／よ／り／ま／す／と）
- 「で」には格助詞、接続助詞、助動詞「だ」の連用形、接続詞の４つがある。

  － 公園 で 遊ぶ（格助詞）
  － 罪を憎ん で 人を憎まず（接続助詞）
  － あっちはホテル で、こっちはレストランです。（助動詞「だ」の連用形）
  － で、どっちに決めたんだ。（接続詞）

- 「それでも」の品詞の判断は、「それでも」の前に「。」が来ている時は接続詞とし、「。」以外のものが前に来ている時は切って「それ（代名詞）／で（格助詞）／も（係助詞）」とする。

  例．．．だけでもいいのですが、それでもよろしいですか。
  （それ（代名詞）／で（格助詞）／も（係助詞））

- 並立助詞か副助詞か悩む「とか」について

  例 なにか、例えばスピーチを要約したものとかですが。（副助詞；並立する相手が明記されてないため）

- 「と」には格助詞、並立助詞、引用助詞、接続詞の４つがあることに注意。

  - どなたかと話をしたい → 格助詞
  - 彼と彼女がやってきて → 並立助詞
  - くださいと言われても → 引用助詞
  - トンネルを抜けると雪国であった → 接続助詞

- 「格助詞＋本動詞＋語尾＋助動詞」は一語で 連体助詞とする。
  コンピュータ技術などに関した書籍（に関する＝連体助詞）

- （同〜の「同」や当事務所の「当」は連体詞とする。

  例 同氏は「センサー」についてお話に〜。

おわりに

この報告書は、「日本語形態素解析マニュアル」[1] に従って、実際に形態素解析作業を実施している中で気がついた疑問点や、判断に迷ったものに関して内部で決定していったものをまとめあげたものである。中にはこの形態素を基に構文木を試作してみて気の付いたものも含まれている。また、「日本語形態素解析マニュアル」では明確でなかった接頭辞、接尾辞に関してはその全てを明示することで認定基準を明らかにすることを志した。

接頭辞、接尾辞の決定に関して討議に参加されたＴＩＳの高橋誠氏、日本ＩＲの衛藤純司氏に感謝する。前述したように、この報告書の内容の多くは実作業から生まれたものである。音声言語データベースの構築において、日本語形態素解析の実作業にあたられているＴＩＳの山田久子さん、手束幸代さん、日本語構文木の作成に当たられている日本ＩＲの坂口明子さんに深謝する。

参考文献

[1] 浦谷則好・田代敏久・山田久子・松本香:「音声言語データベースにおける日本語形態素解析マニュアル」, ＡＴＲテクニカルレポート（ＴＲ－ＩＴ－０００９），（１９９３．９）

[2] 浦谷則好・竹澤寿幸・田代敏久・江藤純司：「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」，ＡＴＲテクニカルレポート（ＴＲ－ＩＴ－０００３），（１９９３．７．１３）

[3] 竹澤寿幸・田代敏久・保坂順子・衛藤純司：「自動翻訳電話研究所の対話データベースと音声翻訳システムＡＳＵＲＡの音声認識用文法・構文解析用文法の間の日本語形態素情報の相違点」，ＡＴＲテクニカルレポート（ＴＲ－ＩＴ－０００６），（１９９３．７．２９）

[4] 篠崎直子・水野康子・小倉健太郎・吉本啓：「形態素情報利用解説書」，ＡＴＲテクニカルレポート（ＴＲ－Ｉ－００７７），（１９８９．３）

[5] 松村明［編］：「大辞林」，三省堂，（１９８８．１１）