『日本語話し言葉コーパス』の形態論情報の概要　ver.1.0

小椋秀樹　山口昌也　西川賢哉　石塚京子　木村睦子

独立行政法人国立国語研究所

目　次

凡　例

各単位の境界については，以下の記号で示す。

文節の境界	……	‖	例：‖国立国語研究所の‖
長単位の境界	……	┃	例：┃国立国語研究所┃
短単位の境界	……	｜	例：｜国立｜国語｜研究｜所｜
最小単位の境界	……	／	例：／国／立／国／語／研／究／所／

注目している単位が分かりにくい場合は，その単位に下線を施すことがある。
また，切らないことを示す場合には「 = 」(例：西が=丘）を用いる。

はじめに

本冊子は，『日本語話し言葉コーパス』（以下ＣＳＪと呼ぶ）のコアにおける形態論情報について解説するものである。

ＣＳＪの形態論情報は，以下の七つの情報からなる。

単位境界　代表形　代表表記　品詞　活用の種類　活用形　その他の情報

以下，各情報について基準等の概略を述べることとする。

なお，ＤＶＤに収録した短単位・長単位データベースの仕様については，「短単位・長単位データマニュアル」（ファイル名：wdb.pdf）を参照されたい。

１　長単位・短単位の設計

ＣＳＪの単位を設計するに当たっては，まずＣＳＪを使ってどのような国語研究を行うのかということを考え，その目的に適した単位を設計することとした。

今回，ＣＳＪを使った国語研究として我々が考えたのは，次の二つである。

(1) ＣＳＪから用例を採集し，話し言葉の語い・語法の研究を行う。
(2) ＣＳＪにおける品詞の分布など計量研究を行う。

もちろんＣＳＪを使った国語の研究は，この二つに限られるものではない。しかし様々な研究を想定し，それらすべてに適した単位を設計することは不可能に近い。そこで，我々はひとまず上記の二つに絞って，それに適した単位を設計することとした。

まず，(1)の目的のためには，合成語を構成要素に分割したような短い単位が求められる。しかし構成要素に分割すると言っても細かく分割しすぎると，文字列検索とほとんど同じになり，不要な用例を検索してしまう可能性が出てくる。また，取り出した単位の意味が文脈から離れすぎるという問題もある。例えば，「至る所」を構成要素に分割すると，「至る」と「所」の２単位に分割できる。しかし，「至る所」はこれ全体で「どこもかしこも」という意味を表しているものであり，「至る」と「所」とに分割して，「至る」を取り出しても，動詞「至る」が本来持っている「行き着く」「到着する」という意味は，ほとんど希薄化してしまっている。そのため，動詞「至る」を検索した際に，「至る所」の構成要素として用いられた「至る」の例が検索されるような単位は，余り望ましいものとは言えないであろう。つまり，(1)の目的のために短い単位が求められるとは言っても，構成要素にすべて分割してしまうような単位では問題があるということになる。

次に(2)の目的のためには，ＣＳＪの資料的な性格を反映するような単位であることが求められる。一般に単位を短くするほど，基本的な語を取り出すことになり，当該資料の性格を反映するような特徴語などは取り出せなくなる。したがって，その資料の特徴語などを取り出せるような，長い単位が必要になる。

このことについて，「言語」という語を例に少し説明しておく。「言語」は，幾つかの学会講演に用例が見られる語であるが，その用いられ方 ── 特にどのような語と結合するか ── については，学会による差異が見られる。例えば，A01学会（工学系学会（音

声関係))・A02学会（人文系学会（日本語関係））での「言語」の例を見てみると，まずA01学会では,「言語」が単独で用いられた例のほか,「言語刺激」「言語情報」「言語的」「言語モデル」などのように合成語の構成要素として使われた例がある。一方，A02学会では,「言語」が単独で用いられた例のほか「言語外」「言語学」「言語作品」「言語体系」などの合成語の構成要素としての例がある。ここで注意したいのは，A01学会にある「言語モデル」「言語刺激」「言語情報」といった語はA02学会には出てこず，一方，A02学会にある「言語体系」「言語作品」「言語外」という語はA01学会に出てこないということである。つまり「言語モデル」「言語刺激」「言語情報」はA01学会を特徴付ける語であり,「言語体系」「言語作品」「言語外」はA02学会を特徴付ける語なのである。このような各分野に特徴的な語を把握するためには,「言語モデル」を「言語」と「モデル」とに,「言語体系」を「言語」と「体系」とに分割するのではなく，全体で一つとして扱うような単位が必要となる。

なお，(1)(2)いずれの目的のためにも，不統一のない単位とすることが必要である。同じ種類の単語が異なる分割のされ方をしていては，効率的な検索ができない。また計量的な研究では，計量される対象である単位が等質であることが求められるので，不統一のない単位にすることが重要である。

以上のことから，ＣＳＪではまず単位を一つに限定せず，長短二つの単位を採用することとした。また，今回は全く新たに単位を設計するのではなく，国立国語研究所がこれまでに行った語い調査の調査単位の中から，先述の目的に適した単位を採用し，必要に応じて拡張等を行うこととした。

その結果，長い単位（以下，長単位と呼ぶ）については，テレビ放送の語い調査で採用された長い単位[注1]を基にして設計を行った。一方，短い単位（以下，短単位と呼ぶ）については，現代雑誌九十種の語い調査で用いられたβ単位[注2]を採用した。

２　長単位・短単位の認定基準

２．１　長単位の認定基準

２．１．１　文節の認定

長単位の認定に当たっては，まず文節の認定を行う。この文節は，テレビ放送の語い調査で用いられた長い単位を基に若干の修正を加えたものである。

付属語には，複合辞も含めた。複合辞は，現代日本語の研究や日本語教育ではよく取り上げられるものである。国立国語研究所『現代語複合辞用例集』（2001年）では，助詞的複合辞・８３語，助動詞的複合辞・４２語を取り上げ，用例を示すとともに解説を加えている。このように現代日本語の研究等では，多くの複合辞が認定されているところではあるが，ＣＳＪでは，それらすべてを複合辞として認定することはしなかった。それは，複合辞の認定には意味の問題が絡んでくるため，その認定自体がかなり難しいということによる。ＣＳＪでは，表１・表２（19～21ぺ）に掲げた助詞相当句・助動詞相当句のみを複合辞として認定した。なお，複合辞については敬語形式のものをどのように扱うかが問題となるが，ＣＳＪでは敬語形式になっているものも複合辞として認定した。

以下，文節の認定で問題となる点について説明しておく。

○　以下に挙げるものについては，その内部に文節の切れ目があっても切らない。

①　固有名のうち，次に挙げるもの。

〔人名（芸名・しこ名・あだ名などをふくむ)〕

【例】　‖源（みなもとの）=頼朝‖　　‖千代の=富士‖

〔国名〕

【例】　‖グレートブリテン=及び=北アイルランド連合王国‖

〔行政区画名〕

【例】　‖西が=丘‖　　‖お茶の=水‖

〔地形名〕

【例】　‖塔の=岳‖

〔場所名〕

【例】　‖丸の=内線‖　　‖虎の=門交差点‖

〔建造物名〕

【例】　‖五重の=塔‖

〔組織名（社名・会議・委員会など）及びそれに関連する肩書〕

【例】　‖国立少年自然の=家‖

〔歴史的できごとの名称〕

※　戦争・革命・事件などで，日本史・世界史の教科書において，慣用的に一定の名で呼ばれるもののみとする。

【例】　‖関ヶ原の=戦い‖　　‖蛤御門の=変‖　　‖明治十四年の=政変‖

〔祝日〕

※　『国民の祝日に関する法律』に定められたもの。

【例】　‖建国記念の=日‖　　‖子供の=日‖

②　「―の～」「―が～」の体言句のうち，以下に挙げるもの。

〔「―の～」の体言句〕

‖麻の=葉‖　　‖味の=素‖　　‖有りの=儘‖　　‖絵の=具‖　　‖男の=子‖

‖思いの=丈‖　‖思いの=外‖　‖女の=子‖　‖髪の=毛‖　‖上の=句‖
‖気の=毒‖　‖木の=芽‖　‖木の=下‖　‖下の=句‖　‖茶の=間‖
‖念の=為‖　‖日の=出‖　‖目の=当たり‖　‖身の=上‖　‖身の=程‖
‖身の=回り‖　‖目の=敵‖　‖山の=手‖　‖世の=中‖

〔「一が～」の体言句〕

‖万が=一‖

③　分数の読み上げ

【例】

‖三分の=二‖　‖後続単語種類数分の=先行単語頻度‖

④　動植物名

【例】

‖タツノ=オトシゴ‖　‖サキシマスオウノ=キ‖

○　並列及び同格の関係にある語は互いに切り離す。

【例】

‖安心‖確実な‖方法‖　‖塩‖こしょうを‖かける‖
‖機関誌‖計量国語学‖

並列及び同格の関係にある体言連続のうち，並列された体言全体に係る体言・接辞がある場合は切らない。また並立された体言全体を受ける体言・接辞・形式的な意味の「する」「できる」「なさる」「いたす」がある場合は切らない。

【例】

‖平成=九年=十年‖
‖関東=東北=地方‖　‖機関誌=計量国語学=発行‖　‖観察=整理=する‖

○　体言連続の一部分が連体修飾語を受けている場合，その部分の後で切る。
ただし，「以降」「間」「ごと」「自体」「達」が付いた場合は切らない。

【例】

‖項構造の‖曖昧性‖解消‖
‖文章の‖途中=以降‖　‖住んでる‖人=達‖

○　体言及び副詞に形式的な意味の「いたす」「する」「できる」「なさる」が直接続く場合，体言及び副詞と「いたす」「する」「できる」「なさる」との間は切らない。

【例】

‖許容=する‖　‖演出=できる‖　‖体験=なさる‖
‖きらきら=する‖　‖きちんと=する‖

ただし，前の体言が連体修飾を受けている場合は用言部分を切り離す。

【例】

‖面白い‖説明‖する‖人‖

○　「お（ご）＋動詞連用形（名詞）＋する・くださる・いただく・なさる・いたす・ねがう・もうしあげる・あそばす」は全体で一続きとする。

【例】

‖御会いする‖　‖御与えください‖　‖御電話なさる‖　‖御登場願う‖

○　数量を表す要素を含む自立語は，以下のように処理する。

① 前の要素に関する順序・番号を直後の要素が表している場合，両者を切り離さない。

【例】

‖昭和十三年=八月=八日‖　　‖朝=八時‖　　‖予稿集=八十七ページ‖

‖入所=二十年目‖　　‖野村=一九九四‖

② 上記の規則に該当しない場合，数量を表す要素とその直前の要素とを切り離す。

【例】

‖果汁‖百パーセント‖※　　‖バニラエッセンス‖少々‖※

‖山の手線‖京浜東北線‖二本‖※　　‖一箱‖三万‖　　‖週‖二通‖

‖一学年‖上‖　　‖十年以上‖前‖　　‖延べ‖百二十九文‖

※印を付けた形式については，数量を表す要素と前の要素とを受ける体言がある場合は，切り離さない。

【例】

‖果汁=百パーセント=オレンジジュース‖

○ ２文節以上からなる形式全体を受ける，若しくはそれに係る接辞及び体言的な形式は，その前後で切る。

【例】

‖円形劇場とか‖水路‖等‖　　‖への‖字‖型‖　　‖神の‖国‖発言‖

なお，ここで述べた文節は，長単位の認定を行うために規定するものであり，各作業者が作業上必要な概念として持っておくという性質のものである。したがって，この文節の境界はＸＭＬファイル，短単位・長単位データベースのいずれにも示されていない。

また，転記テキストにおける改行基準としての文節とは細部において一致しないところがある。転記テキストにおける文節の詳細については，bunsetsu.pdf を参照されたい。

２．１．２　長単位の認定

長単位は，２．１．１で規定した文節から助詞・助動詞を切り出し，自立語に相当する部分を１単位，付属語に相当する部分を１単位とするような単位である。

以下，長単位の認定の際に問題となる点について説明しておく。

○ 付属語について

①長単位で認定する付属語には，表１・表２に示した複合辞を含む。

【例】

｜データベース化する｜という｜仕事｜を｜やっ｜ており｜ます｜

② 形容動詞及び形容動詞活用型の助動詞(そうだ・みたいだ・ようだ)の活用語尾は助動詞として扱う。

【例】

｜統一的｜な｜視点｜で｜切り｜ましょ｜う｜

｜涙｜が｜出｜そう｜に｜なる｜　　｜エンジニア｜な｜んだ｜そう｜です｜

｜駅員さん｜が｜いる｜みたい｜だ｜　　｜使える｜よう｜に｜し｜たい｜

③ 文節の認定規則で，その内部に文節の切れ目があっても切らないものとして挙げた固有名・「―の～」「―が～」の体言句・分数の読み上げ・動植物名の内部にある助詞・助動詞は切り出さない。

【例】

｜西=が=丘周辺｜　　｜油絵=の=具｜　　｜万=が=一｜

｜三分=の=二｜　　｜後続単語種類数分=の=先行単語頻度｜

｜サキシマスオウ=ノ=キ｜

ただし，分数の読み上げにおいて，分母・分子に当たる要素のいずれかが２文節以上の場合は，以下のように分割する。

【例】

｜標本化周波数｜分｜の｜帯域幅｜(A パイ;π)｜という｜

○　並列及び同格の関係にある語は互いに切り離す。

【例】

｜安心｜確実｜な｜方法｜　　｜塩｜こしょう｜を｜かける｜

｜機関誌｜計量国語学｜

並列及び同格の関係にある体言連続のうち，並列された体言全体に係る体言・接辞がある場合は切らない。また並立された体言全体を受ける体言・接辞・形式的な意味の「する」「できる」「なさる」「いたす」がある場合は切らない。

【例】

｜平成=九年=十年｜

｜関東=東北=地方｜　　｜機関誌=計量国語学=発行｜　　｜観察=整理=する｜

○　体言連続の一部分が連体修飾語を受けている場合，その部分の後で切る。
ただし，「以降」「間（かん）」「ごと」「自体」「達」が付いた場合は切らない。

【例】

｜項構造｜の｜曖昧性｜解消｜

｜文章｜の｜途中=以降｜　　｜住ん｜でる｜人=達｜

○　体言及び副詞に形式的な意味の「いたす」「する」「できる」「なさる」が直接続く場合，体言及び副詞と「いたす」「する」「できる」「なさる」との間は切らない。

【例】

｜許容=する｜　　｜演出=できる｜　　｜体験=なさる｜

｜きらきら=する｜　　｜きちんと=する｜

ただし，前の体言が連体修飾を受けている場合は用言部分を切り離す。

【例】

｜面白い｜説明｜する｜人｜

○　「お（ご）＋動詞連用形（名詞）＋する・くださる・いただく・なさる・いたす・ねがう・もうしあげる・あそばす」は全体で一続きとする。

【例】

｜御会いする｜　　｜御与えください｜　　｜御電話なさる｜　　｜御登場願う｜

○　数量を表す要素を含む自立語は，以下のように処理する。

①　前の要素に関する順序・番号を直後の要素が表している場合，両者を切り離さない。

【例】

｜昭和十三年=八月=八日｜　　｜朝=八時｜　　｜予稿集=八十七ページ｜

｜入所=二十年目｜　　｜野村=一九九四｜

② 上記の規則に該当しない場合，数量を表す要素とその直前の要素とを切り離す。

【例】

｜果汁｜百パーセント｜※　　｜バニラエッセンス｜少々｜※

｜山の手線｜京浜東北線｜二本｜※　　｜一箱｜三万｜　　｜週｜二通｜

｜一学年｜上｜　　｜十年以上｜前｜　　｜延べ｜百二十九文｜

※印を付けた形式については，数量を表す要素と前の要素とを受ける体言がある場合は，切り離さない。

【例】

｜果汁=百パーセント=オレンジジュース｜

○ ２文節以上からなる形式全体を受ける，若しくはそれに係る接辞及び体言的な形式は，その前後で切る。

【例】

｜円形劇場｜とか｜水路｜等｜　　｜へ｜の｜字｜型｜　　｜神｜の｜国｜発言｜

○ 言い直し・言い換えについては，以下のように処理する。

① 語の一部を述べたところで，語全体を言い直している場合。

【例】

｜益岡田窪氏｜の｜基本日本語｜基礎日本語文法｜

｜太平洋開戦｜太平洋戦争開戦｜の｜年｜に｜

｜高原農家｜高原野菜農家｜で｜働い｜ている｜

② 前に述べた語の一部のみを直後で言い直している場合。

【例】

｜阪倉篤義さん｜篤義先生｜の｜　　｜国語｜について｜つき｜まし｜て｜

③ 前に述べた語全体を言い換え，若しくは言い直している場合。

【例】

｜向こう｜で｜教育機関｜教育事業｜始め｜たい｜という｜こと｜で｜

｜一つ｜は｜勿論｜装着性｜ウェアラブル｜という｜こと｜です｜ね｜

④ １長単位の内部に言い直しがある場合。

【例】

｜国立=日本語=国語研究所｜で｜　　｜国語｜につい=つき=まして｜は｜

ところで，ＣＳＪの転記ファイルには各種のタグが用いられているため，単位解析の際に，これらのタグをどのように処理するかが問題となる。これについては，タグを単独で切り出すことはせず，以下のように直前・直後の単位に含めることとした。

【例】

｜ひとえ｜に｜(A エヌ;N)グラム｜を｜

｜(0 春｜過ぎ｜て｜夏｜来｜に｜けら｜し)｜

｜(M 小さい)｜の｜(A エス;S)｜の｜部分｜

以上の規定によって，長単位を認定した結果を次に示す。

【例】

｜(F えー)｜パラ言語情報｜という｜こと｜な｜んです｜が｜(F あ)｜簡単｜に｜最初｜

に｜<雑音>｜<雑音>｜(F えー)｜復習｜を｜し｜ておき｜たい｜と｜思い｜ます｜(F ま)｜(F あのー)｜こう｜やっ｜て｜(D あっ)｜話し｜ており｜ます｜と｜それ｜は｜勿論｜(F あの)｜言語的情報｜を｜伝える｜という｜こと｜が｜

※　フィラー・言いよどみの単位認定については，「２．３ 話し言葉特有の現象の単位認定」を参照。

２．２　短単位の認定基準

２．２．１　最小単位の認定

短単位の認定に当たっては，まず最小単位を認定する。最小単位は，現代語において意味を持つ最小の単位であり，和語・漢語・外来語・記号・固有名(人名・地名)の種類ごとに次のように認定される。

【例】

和　語　：　／話し／言葉／　　／お／話し／し／ます／
　　／大／雨／が／降っ／た／の／で／
漢　語　：　／国／語／　　／研／究／
外来語　：　／データー／ベース／　　／ネット／ワーク／
記　号　：　／図／Ａ／　　／ＮＨＫ／
人　名　：　／星野／仙一／
　　※ 姓と名それぞれが１最小単位。
地　名　：　／大阪／府／豊中／市／待兼山町／
　　／六甲／山／　／琵琶／湖／
　　※ 地形名の名を表す部分は１最小単位。

「だが」「では」などの助詞・助動詞から転化した接続詞も「／だ／が／」「／で／は／」のように分割する。また「ていく」「について」などの複合辞も「／て／いく／」「／に／つい／て／」のように最小単位を認定する。また接続助詞「ので」や副助詞「とか」のような複数の助詞・助動詞が結合してできた助詞についても，「／の／で／」「／と／か／」のように最小単位を認定する。

なお，ここで述べた最小単位は，短単位の認定を行うために規定するものであり，各作業者が作業上必要な概念として持っておくという性質のものである。したがって，この最小単位の境界は，ＸＭＬファイル，短単位・長単位データベースのいずれにも示されていない。

２．２．２　短単位の認定

２．２．１で規定された最小単位を表３（９ぺ）のように分類する。

各分類について少し説明しておく。

付属要素とは，接頭辞・接尾辞のことである。ただしすべての接頭辞・接尾辞が付属要素として扱われるわけではない。ＣＳＪに出現したものの中から，造語力が高いなど特に注目されるものを「付属要素一覧」（表４・表５（21～23ぺ））というリストに掲げ，そのリストに掲げられたもののみを付属要素として扱うこととした。

数には，一・十・百・千などの数詞のほか，「数十」「何百」「幾千」の「数」「何」「幾」も含めることとした。また数詞のうち，数え進むことができないと考えられるもの ── 例えば「一応」の「一」や「百科」の「百」など ── については，一般に分類した。

助詞・助動詞には，助詞・助動詞から転化した接続詞「だが」「では」なども含めた。「だが」「では」は先に示したように「／だ／が／」「／で／は／」と最小単位が認定されるが，その「だ」「が」「で」「は」をそれぞれ助詞・助動詞に分類したということである。形容動詞及び形容動詞活用型の助動詞(そうだ・みたいだ・ようだ)の活用語尾も助動詞に分類した。

表３　最小単位の分類

分　類	語　例
一　般	和　語　：　山　川　白い　話す　言葉　……
	漢　語　：　社　会　用　研　究　所　……
	外来語　：　オレンジ　ボックス　アルゴリズム　……
付属要素	接頭的要素　：　相　御(お)　各　御(ご)　……
	接尾的要素　：　合う　致す　っぽい　性　的　……
記　号	Ａ　Ｂ　ω　イ　ロ　ア　ＮＨＫ　ＪＲ　……
数	一　二　十　百　千　幾　数　何　……
人名・地名	星野　仙一　大阪　六甲　……
助詞・助動詞	う　た　です　ます　か　から　て　も　……

短単位の認定基準は，上記の各分類ごとに適用すべき規則が定められている。その規則のうち，短単位認定の基本原則に当たるのが，一般の最小単位に適用される以下の規則である。

○　一般に分類した最小単位２個の１次結合は１短単位とする。

【例】

｜母親｜　　｜食べ歩く｜　　｜音声｜　　｜レーザープリンター｜

｜無口｜　　｜オレンジ色｜

この結合に当たっては長単位を超えないという制約を設けている。これによって，長単位の下位に短単位が位置付けられるという階層構造を持つことになる。

一般に分類した最小単位であっても，それ単独で１短単位になるものや３最小単位以上の結合であっても全体で１短単位とするものがある。それを以下に示す。

○　１最小単位を１短単位とするもの。

①　最小単位が三つ以上並列した場合の各最小単位。

【例】

｜衣｜食｜住｜　　｜松｜竹｜梅｜　　｜都｜道｜府｜県｜

②　重複形の擬音語・擬態語で，重複が奇数回の場合の，その重複されている要素。

【例】

｜ぐる｜ぐる｜ぐる｜っと｜回る｜　　｜ちょこ｜ちょこ｜ちょこ｜動く｜

※　偶数回の繰り返しの場合は，原則に従う。

【例】

｜ぐるぐる｜っと｜回る｜　　｜ぐるぐる｜ぐるぐる｜っと｜回る｜

③　類概念を表す部分と名を表す部分とが結合してできた固有名詞のうち，類概念を表す部分と名を表す部分とがともに１最小単位の場合の，それぞれの部分。

【例】

｜さくら｜屋｜　　｜リクルート｜社｜　　｜ハーバード｜大｜
｜のぞみ｜号｜　　｜むらさき｜会｜　　｜キリスト｜教｜
｜タイムズ｜紙｜　　｜キャノン｜カメラ｜

※　名を表す部分が１字の漢語で，類概念を表す部分が１最小単位である場合は，その一次結合体を１短単位とする。

【例】

｜仏教｜　　｜儒教｜　　｜阪大｜

④　外来語の最小単位うち英語の接続詞・前置詞・冠詞に当たるもの。

【例】

｜アウト｜オブ｜ドメイン｜　　｜ショアーズ｜アット｜ワイコロア｜
｜基本｜レフト｜トゥー｜ライト｜構造｜　　｜コール｜フォー｜ペーパー｜

⑤　外来語の最小単位２個の１次結合体が１１拍以上になる場合の各最小単位。

【例】

｜インサーション｜ペナルティー｜　　｜スペクトル｜パラメーター｜

⑥　外国語。

【例】

｜イッツァ｜ペン｜

⑦　短単位認定基準によって得られた短単位に，前又は後ろから結合した最小単位。

【例】

｜内閣｜府｜　　｜副｜大統領｜　　｜光ファイバー｜網｜　　｜自衛｜隊｜
｜国立｜国語｜研究｜所｜

⑧　単独で文節を構成する最小単位。

【例】

｜やっぱり｜これ｜も｜ひと｜つ｜の｜　　｜親｜と｜面談｜する｜
｜オレンジ｜を｜食べる｜

○　３最小単位以上の結合であっても全体で１短単位とするもの。

①　三つ以上の最小単位からなる組織名等の略称。

【例】

｜日経連｜　　｜通総研｜

②　切る位置が明確でないもの，あるいは切った場合と一まとめにした場合とで意味にずれがあるもの。

【例】

｜大統領｜　　｜不可解｜　　｜明後日｜　　｜殺風景｜
｜輸出入｜　　｜国内外｜　　｜町村長｜　　｜原水爆｜　　｜市町村長｜
｜大袈裟｜　　｜大雑把｜　　｜大丈夫｜　　｜一辺倒｜
｜十文字｜　　｜二枚目｜　　｜十八番｜

③　「－の～」「－が～」の体言句のうち以下に挙げたもの。

〔「－の～」の体言句〕

｜麻の葉｜　　｜味の素｜　　｜有りの儘｜　　｜絵の具｜　　｜男の子｜
｜思いの丈｜　　｜思いの外｜　　｜女の子｜　　｜髪の毛｜　　｜上の句｜

｜気の毒｜　　｜木の芽｜　　｜木の下｜　　｜下の句｜　　｜茶の間｜
｜念の為｜　　｜日の出｜　　｜目の当たり｜　　｜身の上｜　　｜身の程｜
｜身の回り｜　　｜目の敵｜　　｜山の手｜　　｜世の中｜

〔「ーが～」の体言句〕
｜万が一｜

以下，一般以外の最小単位に対する短単位認定規則を示す。

○　記号，人名・地名，付属要素，助詞・助動詞は，１最小単位を１短単位とする。
【例】
記　　号　：　｜図｜A｜　　｜NHK｜
人　　名　：　｜星野｜仙一｜
地　　名　：　｜大阪｜府｜豊中｜市｜待兼山町｜
　　｜六甲｜山｜　　｜琵琶｜湖｜
付属要素　：　｜お｜母｜さん｜　　｜見｜にくい｜
助詞・助動詞　：　｜単位｜に｜切り｜ましょ｜う｜　　｜｜が｜良い｜の｜
　　｜それ｜に｜つい｜て｜　　｜とても｜きれい｜だ｜

○　数は，ほかの最小単位と結合させない。数どうしの結合については，結合の次数にかかわらず，一・十・百・千のとなえを取る けたごとに１短単位とする。「万」「億」「兆」などの最小単位は，それだけで１短単位とする。小数部分は，１最小単位を１短単位とする。
【例】
｜十｜二｜月｜二十｜三｜日｜　　｜七百｜万｜語｜　　｜五｜分｜の｜二｜
｜何十｜倍｜　　｜一二｜年｜前｜　　｜二三十｜回｜

以上の規定によって，短単位を認定した結果を次に示す。
【例】
｜(F えー)｜パラ｜言語｜情報｜と｜いう｜こと｜な｜ん｜です｜が｜(F あ)｜簡単｜に｜最初｜に｜<雑音>｜<雑音>｜(F えー)｜復習｜を｜し｜て｜おき｜たい｜と｜思い｜ます｜(F ま)｜(F あのー)｜こう｜やっ｜て｜(D あっ)｜話し｜て｜おり｜ます｜と｜それ｜は｜勿論｜(F あの)｜言語｜的｜情報｜を｜伝える｜と｜いう｜こと｜が｜

※　フィラー・言いよどみの単位認定については，「２．３ 話し言葉特有の現象の単位認定」を参照。

２．３　話し言葉特有の現象の単位認定

話し言葉には，書き言葉にはない様々な現象が見られる。このうち，単位認定の際に問題となる現象として，次のような融合・省略・フィラー・言いよどみという現象がある。

融　　合　：　そりゃ　　面白きゃ　　食べりゃ　　じゃ　　てる
省　　略　：　やんだっけ　　そうっす
フィラー　：　(F えー)　　(F あのね)　　(F んーと)
言いよどみ　：　(D こ)ここから

融合を処理する方法としては，まず元の語形に戻した上で，単位認定するという方法がある。例えば，「面白きゃ」を「面白ければ」，「じゃ」を「では」に戻した上で単位認定するというものである。この方法は，過去の国語研究所の語い調査でとられたものでもある。このような処理は，基礎語の選定等を目的とした語い調査においては，妥当なものといえよう。しかし，話し言葉コーパスにおける処理方法としては，話し言葉の特徴である融合という現象を分からなくするという点で問題がある。またＣＳＪでは融合現象が多く見られることが予想されるため，すべて元の形に戻していたのでは，作業が煩雑になるという問題もある。そこで，ＣＳＪでは，融合を元の形に戻さずに単位認定をすることとした。例えば，「面白きゃ」「じゃ」「てる」は，長単位・短単位ともにそれぞれ１単位となる。

省略についても，元の形に戻すことなく，可能な範囲で単位分割した。例えば，「やんだっけ」は，「や」を「やる」の活用語尾が省略された形，「ん」を準体助詞「の」の撥音化したものと考え，長単位では「┃や┃んだ┃っけ┃」，短単位では「|や|ん|だ|っけ|」と分割する。

フィラーについては，「(F あの)」「(F えーと)」のようにＦタグが付されているので，長単位・短単位ともに，そのＦタグが付された範囲を１単位とした。ただし，以下のように助詞・助動詞を含む場合，長単位ではＦタグが付されている範囲全体で１単位としたが，短単位では，助詞・助動詞を切り出した。

【例】

長単位　：　┃(F あのですね)┃　　┃(F あのね)┃

短単位　：　｜(F あの｜です｜ね)｜　　｜(F あの｜ね)｜

またフィラーが，単位の中に現れる場合がある。例えば，以下のような例である。

【例】

味わうことが(Fえー)でき<u>ま(F えー)せ</u>ん

ここでも<u>メタ(F あ)言語行動表現</u>てものを手掛かりに

「ま(F えー)せ」は，長単位・短単位いずれにおいても１単位となる助動詞「ます」の未然形の中にフィラーが現れたもので，「メタ(F あ)言語行動表現」は１長単位となる「メタ言語表現行動」の中にフィラーが現れたものである。このような場合は，長単位・短単位いずれにおいても，フィラーを無視して単位認定を行うこととした。つまり，上の二つの例は，次のように単位がされることになる。

【例】

｜味わう｜こと｜が｜(Fえー)｜でき｜<u>ま(F えー)せ</u>｜ん｜

※　長単位も短単位と同様の単位認定となる。

┃ここ┃で┃も┃メタ<u>(F あ)</u>言語行動表現┃て┃もの┃を┃手掛かり┃に┃

※　短単位については，「メタ言語表現行動」が「｜メタ｜言語｜表現｜行動｜」と分割されるので，ここで問題としている１単位の中に現れるフィラーには当たらない。２．１．２に示した認定規則により，「｜メタ｜(F あ)｜言語｜行動｜表現｜」と分割される。

言いよどみについても，フィラーと同様にタグの付された範囲全体を１長単位又は１短単位とした。また，１長単位又は１短単位の中に現れる言いよどみについても，フィラーと同様に無視して単位認定を行った。

【例】

┃それ┃を┃<u>利用(D す)する</u>┃の┃も┃

｜それ｜と｜ポライト｜<u>ノン(D プロ)ポライト</u>｜と｜いう｜風｜に｜

なお，言いよどみのうち，数詞・助詞・助動詞・接頭辞・接尾辞の言いよどみ（Ｄ２タグを付したもの）については，通常の助詞・助動詞・接頭辞・接尾辞と同様に単位認定を行った。

【例】

長単位：┃実験三┃(D2 の)┃として┃は┃　　┃(D2 未)┃未観測┃だっ┃た┃
┃六十(D2 二)二パーセント┃の┃

短単位：｜実験｜三｜(D2 の)｜と｜し｜て｜は｜　　｜(D2 未)｜未｜観測｜だっ｜た｜
｜六十｜(D2 二)｜二｜パーセント｜の｜

３　付加情報

第２節に示した単位認定基準によって認定された長短２種類の単位に対して，以下のような品詞情報を付与する。

代表形　　代表表記　　品詞　　活用の種類　　活用形　　その他の情報

以下，各情報について概略を説明する。

３．１　代表形・代表表記

代表形とは，同一語の活用変化・音の転化・揺れ・省略・融合等によって生じた異形態を一まとめにし，そのまとめられた語群に対して付与する情報で，辞書の見出し語に相当するものである。ＣＳＪでは片仮名で表記した。

代表表記は，代表形に対して付与する漢字等の国語の表記である。漢字表記を付与することにより，同語形異語の区別が可能となった。

３．２　品詞情報

品詞・活用の種類・活用形・その他の情報の詳細については，表６（１５ぺ）を参照されたい。以下，品詞・活用の種類・活用形及びその他の情報について補足しておく。

品詞は，長単位・短単位ともに共通で１５種類となっている。このうち，形状詞というのは，いわゆる形容動詞の語幹のことである。長単位・短単位ともに，形容動詞は「｜きれい｜だ｜」「｜新鮮｜だ｜」のように活用語尾が分割されるので，その語幹に当たる部分に付与する品詞として形状詞を用意した。活用語尾については，断定の助動詞「だ」として扱うので，助動詞という品詞が付与されることになる。

活用の種類・活用形は，手動で単位解析を行ったコアと自動で単位解析を行った非コアとで若干異なる。例えば，カ行五段活用は，コアではカ行五段としか情報付与されないが，非コアでは連用形の音便形がイ音便になる「書く」などと，促音便になる「行く」などとを分け，前者をカ行五段１とし，後者をカ行五段２としている。またその連用形についても，非コアでは，連用形を非音便形と音便形とに分け，前者を連用形１とし，後者を連用形２としている。このように，非コアの活用の種類・活用形の方がコアのそれよりも細分化されているのである。

その他の情報は，その他の情報１～３の３種類に分かれる。

その他の情報１は，名詞・助詞に対して付与する情報である。名詞のうち人名・地名には「固有名詞」，数詞には「数詞」という情報を付与する。助詞には「格助詞」「準体助詞」「接続助詞」「係助詞」「副助詞」「終助詞」という助詞の下位区分を付与する。

その他の情報２は，音便等の語形変化に関する情報である。促音便等の「音便」という情報のほかに，融合又は省略という現象の見られる単位に付与する「融合」「省略」がある。この「融合」「省略」については，「３．３　話し言葉特有の現象に対する情報付与」を参照されたい。音便には，「知らない」が「知んない」となるような一般に活用形の一つとして認められていない音便がある。これについては，「撥音便Ａ」のように末尾にＡという記号を付け一般の音便と区別している。つまり，「知んない」の「知ん」には「撥音便Ａ」という情報を付与する。

その他の情報３には，転記でMタグ付与されたものに付ける「メタ」，転記でＤ２タグを付与されたものに付ける「言いよどみ」，長単位で認定される複合辞に付ける「連語」がある。

表６　ＣＳＪの品詞情報

品詞	活用の種類	活　　用　　形	その他の情報		
			1	2	3
名　詞			固有名詞 数詞	融合	メタ
代名詞				融合	メタ
形状詞				融合	メタ
連体詞				融合	メタ
副　詞				融合	メタ
接続詞				融合	メタ
感動詞					メタ
動　詞	ｘ行五段 ｘ行上一段 ｘ行下一段 カ行変格 サ行変格 ザ行変格 文語ｘ行四段 文語ｘ行上二段 文語ｘ行下二段 文語カ行変格 文語サ行変格 文語ナ行変格 文語ラ行変格	未然形･連用形･終止形･連体形･仮定形･命令形･語幹 未然形･連用形･終止形･連体形･仮定形･命令形･語幹 未然形･連用形･終止形･連体形･仮定形･命令形･語幹 未然形･連用形･終止形･連体形･仮定形･命令形･語幹 未然形･連用形･終止形･連体形･仮定形･命令形･語幹 未然形･連用形･終止形･連体形･仮定形･命令形･語幹 未然形･連用形･終止形･連体形･已然形･命令形･語幹 未然形･連用形･終止形･連体形･已然形･命令形･語幹 未然形･連用形･終止形･連体形･已然形･命令形･語幹 未然形･連用形･終止形･連体形･已然形･命令形･語幹 未然形･連用形･終止形･連体形･已然形･命令形･語幹 未然形･連用形･終止形･連体形･已然形･命令形･語幹 未然形･連用形･終止形･連体形･已然形･命令形･語幹		音便 音便Ａ 融合 省略	メタ
形容詞	形容詞型	未然形･連用形･終止形･連体形･仮定形･命令形･語幹		音便 音便Ａ 融合	メタ
	文語形容詞型１ 文語形容詞型２ 文語形容詞型３	未然形･連用形･終止形･連体形･已然形･命令形･語幹 未然形･連用形･終止形･連体形･已然形･命令形･語幹 未然形･連用形･終止形･連体形･已然形･命令形･語幹		音便	メタ
助動詞		未然形･連用形･終止形･連体形･仮定形･命令形･語幹		音便 音便Ａ 融合 省略	言いよどみ メタ 連語
	文語	未然形･連用形･終止形･連体形･已然形･命令形･語幹		音便 融合 省略	言いよどみ メタ
助　詞			格助詞 準体助詞 接続助詞 係助詞 副助詞 終助詞	融合	言いよどみ メタ 連語
接頭辞					言いよどみ メタ
接尾辞	〔無活用の接尾辞〕				言いよどみ メタ
	〔動詞性接尾辞〕 ｘ行五段 ｘ行上一段 ｘ行下一段	未然形･連用形･終止形･連体形･仮定形･命令形･語幹 未然形･連用形･終止形･連体形･仮定形･命令形･語幹 未然形･連用形･終止形･連体形･仮定形･命令形･語幹		音便 音便Ａ 融合 省略	言いよどみ メタ
	〔形容詞性接尾辞〕 形容詞型	未然形･連用形･終止形･連体形･仮定形･命令形･語幹		音便 音便Ａ 融合	言いよどみ メタ
	文語形容詞型１ 文語形容詞型２	未然形･連用形･終止形･連体形･已然形･命令形･語幹 未然形･連用形･終止形･連体形･已然形･命令形･語幹		音便 融合 省略	言いよどみ メタ
記号					メタ
言いよどみ					

３．３　話し言葉特有の現象に対する情報付与

３．３．１　代表形・代表表記の付与

話し言葉特有の現象のうち，省略については可能な範囲で単位分割していくので，代表形・代表表記の付与に当たって問題となることはない。一方，融合については，本来複数の単位に分割されるものを一まとめにして扱っているため，どのように代表形・代表表記を付与するかということが問題となる。

これについては，以下のように処理することとした。

① 活用語の融合 …… 活用語の終止形を代表形として付与する。代表表記についても同様。

【例】

［出現形］	［代表形］	［代表表記］
面白きゃ	オモシロイ	面白い
（行か）なきゃ	ナイ	ない
（加味さ）れりゃ	レル	れる
じゃ	ダ	だ

② 非活用語の融合 …… 元の形を代表形として付与する。

【例】

［出現形］	［代表形］	［代表表記］
こた	コトハ	事は
こら	コレハ	此れは
そりゃ	ソレハ	其れは
ちゃ	テハ	ては
じゃ	デハ	では

３．３．２　品詞情報の付与

活用語の融合については，活用語の品詞を付与する。また，活用形については，元の形に戻した場合の活用形を付与する。またその他の情報２には融合という情報を付与する。

【例】

［出現形］	［代表形］	［品詞］	［活用形］	［その他２］
面白きゃ	オモシロイ	面白い	仮定形	融合
（行か）なきゃ	ナイ	ない	仮定形	融合
（加味さ）れりゃ	レル	れる	仮定形	融合
じゃ	ダ	だ	連用形	融合

活用語の融合のうち，以下に挙げるものは助動詞として扱った。

「てる」（「ている」の融合）　「てらっしゃる」（「ていらっしゃる」の融合）
「てく」（「ていく」の融合）　「とく」（「ておく」の融合）
「とる」（「ておる」の融合）　「ちまう・ちゃう」（「てしまう」の融合）
「たげる」（「てあげる」の融合）　「たる」（「てやる」の融合）
「つう(っつう・(っ)ちゅう)・ってえ)」（「という」の融合）

これらは，例えば以下のように規則的に活用するものとしてとらえることが可能だからである。

【例】

［未然形］［連用形］［終止形］［連体形］［仮定形］［命令形］

てる：	て	て	てる	てる	てれ	てろ
てく：	てか／てこ	てっ	てく	てく	てけ	てけ

非活用語の融合については，先行語の品詞を付与する。また，その他の情報２として融合という情報を付与する。

【例】

［出現形］	［代表形］	［品詞］	［その他２］
こた	コトハ	名詞	融合
こら	コレハ	代名詞	融合
そりゃ	ソレハ	代名詞	融合
ちゃ	テハ	助詞	融合
じゃ	デハ	助詞	融合

省略については，その他の情報として省略という情報を付与する。それ以外は，通常の単位に対する情報付与と同様である。

【例】

［出現形］	［代表形］	［品詞］	［その他２］
や（んだっけ）	ヤル	動詞	省略

終わりに

以上，ＣＳＪにおける長短二つの単位の認定基準及び付加情報について概略を述べた。

ＣＳＪでは，用例採集・計量研究という二つの研究目的を設定した上で，用例採集のための短単位，計量研究のための長単位というようにその目的に応じて２種類の単位を設計した。

しかし設計に当たって，単語とは何かという議論を避けているという批判もあろう。これは，従来の国語研究所の語い調査でも常に問題となっていることである。この点については，今回の成果を基に考えていく必要がある。

なお，最後に形態論情報を付与する際の音声情報の利用について補足しておく。話し言葉の単位や品詞の認定においては，イントネーション・ポーズなどの情報が重要な役目を果たすことがある。ＣＳＪの単位認定及び品詞情報の付与に当たっても，必要に応じてイントネーション等の情報を利用している。

注

（１）　国立国語研究所報告１１２『テレビ放送の語彙調査Ｉ』(1995年，秀英出版)

（２）　国立国語研究所報告２１『現代雑誌九十種の用語用字(1)』(1962年，秀英出版)

表１　ＣＳＪで認定した複合辞（助詞相当句）

見出し	異形態		連体修飾型	
	連用形	丁寧形	普通形	丁寧形
でもって				
にあたって	にあたり	にあたりまして		
にあって		にありまして		
に至る				
において		におきまして	における	におけます
に応じて		に応じまして	に応じた	
に関して	に関し	に関しまして	に関する	
に比べて	に比べ	に比べまして		
に際して				
に従って	に従い		に従った	
に対して	に対し	に対しまして	に対する	に対します
について	につき	につきまして		
につれて	につれ	につれまして		
にとって		にとりまして		
にとっては				
に伴って	に伴い		に伴う	
に基づいて	に基づき	に基づきまして	に基づく	
			に基づいた	
によると		によりますと		
によって	により	によりまして	による	によります
によっては				
にわたって	にわたり	にわたりまして	にわたる	
として		としまして		
		といたしまして		
を通じて		を通じまして		
を通して				
をもって				
をもとにして	をもとに	をもとにしまして	をもとにした	
		をもとにいたしまして		
をめぐる				
という			という	
			ていう	
			っていう	
といった			といった	
			ていった	
			っていった	

表２　ＣＳＪで認定した複合辞（助動詞相当句）

種　類	見　出　し	丁　寧　形	異　形　態
肯定・否定			
（肯定）	である		
	でございます		
	のだ	のです	んだ
		のである	んです
		のでございます	んである
			んでございます
（否定）	でない		
	ではない		じゃない
		ではありません	じゃありません
		ではございません	じゃございません
	のではない		のでない
			のじゃない
			んではない
			んじゃない
		のではありません	
許可・依頼・勧誘	てもいい		ていい
			たっていい
		てもよろしい	
	てほしい		
禁止・当然・義務	てはいけない		ちゃいけない
		てはいけません	ちゃいけません
	てはならない		てはならぬ
			ちゃならない
			ちゃならぬ
	ないといけない		ないといけぬ
		ないといけません	
	なければいけない		なきゃいけない
			なけりゃいけない
		なければいけません	なきゃいけません
	なければならない		なきゃならない
			なきゃならぬ
		なければなりません	なけりゃならない
			なきゃなりません
	なくてはいけない		なくちゃいけない
		なくてはいけません	
	なくてはならない		なくちゃならない
	ねばいけない		ねばいけぬ
	ねばならない	ねばなりません	ねばならぬ
	ざるを得ない		ざる得ない
		ざるを得ません	
推量	かもしれない		かもしんない
		かもしれません	
	かもわからない		かもわかんない
			かもわからぬ
		かもわかりません	

表２　ＣＳＪで認定した複合辞（助動詞相当句）（続き）

種　類	見　出　し	丁　寧　形	異　形　態
試行	てみる		
やりもらい	てもらう		
	てもらえる		
	ていただく		
	ていただける		
	てやる		
	てあげる		
	てくれる		
	てくださる		
アスペクト	てある	てございます	
	ている	ていらっしゃる	
	ておる		
	てしまう		
	ておく		
	ていく	てまいる	
	ていける		てける
	てくる	てまいる	

表４　ＣＳＪで認定した付属要素（接頭的要素）

語	備　考
相	※　「相乗り」は除く。
御(オ)	※　次の場合は後の部分と併せて１最小単位とする。 おかげ　おかず　おかま　おさげ　おしゃれ　おたふく　おでき　おとぎ　おなか おにぎり　おふくろ　おまえ　おまけ　おまわり（さん）　おむつ　おもらし　おやつ
御(オン)	※　次の場合は後の部分と併せて１最小単位とする。
各	※　一字漢語と結合したものは除く。
今	※　一字漢語と結合したものは除く。
御(ゴ)	※　次の場合は後の部分と併せて１最小単位とする。 御殿　　御飯　　御免　　御覧
諸	※　一字漢語と結合したものは除く。
全	※　一字漢語と結合したものは除く。
対	※　一字漢語と結合したものは除く。
本	※　一字漢語と結合したものは除く。
御(ミ)	※　次の場合は後の部分と併せて１最小単位とする。 神籤　巫女　神輿　大御

表５　ＣＳＪで認定した付属要素（接尾的要素）

語	備　　考
合う	※　「ともに～する」「たがいに～する」の意のもの
上がり	
致す	
上(うえ)	
得(え)る	※　「…することができる」の意のもの
終える	
遅れる	
終わる	※　「すっかり～する」の意。
化	※　１字漢語と結合したものは除く。
掛かる	※　動作・作用があるものに向けられる意。
がかる	
掛ける	※　「途中でやめる」「～しはじめる」の意。動作や作用をあるもにむけるという意。
方(かた)	※　「しかた（仕方）」は除く。
型(がた)	※　一字漢語及び和語の１最小単位と結合したものは除く。
方(がた)	※　複数を表す。おおよそそのぐらいであることを表す。
難(がた)い	
勝(が)ち	
がてら	
兼ねる	
がましい	
がる	※　助動詞「たがる」は除く。
交わす	※　「たがいに～する」の意。
間(かん)	※　１字漢語と結合したものは除く。
切る	※　「すっかり～し終える」の意。
臭い	※　望ましくない意を強める用法のもの。「かびくさい」「こげくさい」は除く。
下さる	
君(くん)	
気(げ)	
系	※　１字漢語と結合したものは除く。
後(ご)	※　一字漢語と結合したものはのぞく。
ごと	※　「…もいっしょに」の意。
毎(ごと)	※　そのものひとつひとつ、その時その時の意。
熟(こな)す	※　「うまく～する」の意
さ	※　「なさ」「よさ」は除く。ケシ型形容詞語幹に接続する「さ」は除く。
様(さま)	
さん	
時(じ)	※　１字漢語と結合したものは除く。
式	※　形式・方法などの意。１字漢語と結合したものは除く。
染(じ)みる	
中(じゅう)	※　１字漢語と結合したものは除く。
上(じょう)	※　１字漢語と結合したものは除く。
状	※　「～の形」の意。１字漢語と結合したものは除く。
過ぎる	
尽くめ	
為る	※　１字漢語と結合したものは除く。
性	※　１字漢語と結合したものは除く。
そう	※　一般に様態の助動詞「そうだ」の語幹。一般に伝聞の助動詞「そうだ」の語幹。
損なう	
そびれる	
対	※　１字漢語と結合したものは除く。
出す	※　動作を始める意。
達	
給う	
だらけ	
たらしい	
ちゃん	

表５　ＣＳＪで認定した付属要素（接尾的要素）（続き）

語	備　考
中(ちゅう)	※　１字漢語と結合したものは除く。
尽くす	※　「十分に～する」という意。
付き	
っこ	※　「…くらべ」の意。「たがいに…すること」の意。
っこい	
続く	
続ける	
辛(づら)い	
的	※　１字漢語と結合したものは除く。
出来る	
等(とう)	
同士	
通す	※　「ずっとし続ける」という意。
所(ところ)	
殿(どの)	
共(とも)	※　全部の意。
共(ども)	※　へりくだる意味を表すものも含む。
内(ない)	※　１字漢語と結合したものは除く。
乍ら	
為さる	
並(なみ)	※　その類と同じ，あるいは同じ程度であることを表すもの。
形(なり)	※　そのもの相応であるさまの意。「～するまま」「～するに従うさま」の意。
慣れる	
難(にく)い	※　醜悪の意の「みにくい」は除く。
抜く	※　「終わりまでする」という意。
始める	※　その動作をやり出すという意。
果たす	※　「すっかり～し終える」の意。
果てる	※　「すっかり…する」「…し終わる」「完全に…してしまう」の意。
放し	
版	※　１字漢語と結合したものは除く。
風(ふう)	※　様子の意。１字漢語と結合したものは除く。
振(ぶ)り	※　時日の過ぎ去った程度の意のもの。形・姿・様子の意。
分(ぶん)	
ぽい	※　形容詞に接続するものは除く。
ぽっち	
前(まえ)	
捲る	
間違う	
間違える	
周り	
みたい	
向き	
向け	
目	※　順序を示す。中心となる点や場所の意。物の程度の意。
めく	※　擬態語的なものは「めく」を切り出さない。
易い	
良い	
様(よう)	※　一般に助動詞「ようだ」の語幹とされるもの。　方法の意。
用	※　１字漢語と結合したものは除く。
等（ら）	※　複数を表す。事物をおおよそに指す。
らしい	※　助動詞「らしい」は除く。
流	※　１字漢語と結合したものは除く。
類	※　１字漢語と結合したものは除く。
忘れる	
渡る	※　「あたり一面～にする」の意。

科学技術振興調整費による開放的融合研究推進制度
「話し言葉の言語的・パラ言語的構造の解明に基づく『話し言葉工学』の構築」

『日本語話し言葉コーパス』の形態論情報の概要　ver.1.0

平成１６年3月２０日

編集・発行　：　独立行政法人国立国語研究所