

データ作成方針

文責：COJADS班
初版：2020年3月19日

1. データの構成

音声、TextGrid、方言テキスト、標準語テキストの4つのデータからなります。

方言テキストと標準語テキストはホームページで提供します。音声とTextGridはUSBメモリでの配布を予定しています。

1.1 収録データについて

『日本語諸方言コーパス（Corpus of Japanese Dialects：COJADS）』は、日本各地の方言の談話音声を大量に集めた、日本で初めての諸方言コーパスです。談話音声は、文化庁が 1977～1985 年に行なった「各地方言収集緊急調査」の方言談話の収録データを使用しています。この収録データは文化庁から国立国語研究所に移管されています。元のデータは、日本全国 47 都道府県の 200 地点あまりにおける、約 4,000 時間の方言談話の録音テープからなり、一部は『全国方言談話データベース日本のふるさとことば集成』（2002-2008、国書刊行会）として刊行されています。

2019年度は『ふるさとことば集成』に含まれる47都道府県のデータ、合わせて約24時間分を公開しました。以降、随時拡張します。

1.2 メタデータについて

話者属性：ID・生年・年齢・性別

談話属性：ジャンル（自然談話、場面設定）、話題（「年中行事」「ご祝儀のこと」等）

その他：収録場所、収録年月日

1.3 データの形式について

ファイルの形式

音声データ：無圧縮 .wavファイル

テキストデータ：CSV ファイル、TextGridファイル

ファイル名

『ふるさとことば集成』を含む「各地方言収集緊急調査」の各談話ファイルは、以下の規則によって命名・管理されています。

《都道府県名_都道府県番号_地点ID_ジャンル番号+ファイル番号（-分割番号）》

県番号は数字、地点IDはアルファベット、ジャンル番号は「0: 自然談話」「1: 場面設定」のいずれか、ファイル番号と分割番号は数字となっています。ジャンル番号とファイル番号（-分割番号）はアンダーバーなしの続きとしています。

たとえば「大阪_27_d_002-2」は「大阪府(27)河内長野市(d)自然談話(0)のファイル番号02を分割したうちの2」、「宮城_04_c_103-2a」は「宮城県(04)仙台市(c)場面設定(1)のファイル番号03を分割したうちの2a」を指します。

ただし、『ふるさとことば集成（書籍版）』のデータは地点ID以下を099としており、「茨城_08_c_099」のようになります。

ファイルの詳細

音声ファイルおよびTextGridファイル

TextGridでは話者ごとに層を設定しています。複数の話者による音声の重なりは、それぞれの層に分割されます。知覚可能なポーズ（およそ0.2秒）を目安にboundaryを挿入しており、boundaryで挟まれた発話区間をテキストの転記単位としました。

方言テキスト／標準語テキスト

エクセルファイルで管理しており、対外的にはCSVで配布します。

テキストは音声を転記した方言テキストと、その訳である標準語テキストから成ります。

転記単位に基づいて行を区切り、各行に開始時刻と終了時刻を記しています。

発話内容と時刻の他、話者ID、メタ情報、注を含みます。

記号類の使用は基本的に『ふるさとことば集成』に依拠しますが、そこから追加、削除しています。また、用途を変更して使用しているものもあります。各種記号、タグの詳細は「2. 方言テキスト」「3. 標準語テキスト」を参照。最終頁に一覧を掲載しています。

変更していないもの： { } 「 」 [] 。 (句点) | | (注意書き)

追加したもの： F、D、Z、O、R、Gの各種タグ

削除したもの： 読点 疑問符？ イントネーション↑↓ ()

() によるあいづちはかっこを外し、一部はフィラーとした

変更したもの： [] は用途を変更した

[=] は<A=B>とした

言い間違い×はDタグで音形を生かした

1.4 プライバシーへの配慮

調査協力者に関して、直接的、間接的を問わず害悪（誹謗、風説等）となる可能性のある情報を非公開としました。

非公開対象

非公開対象とした情報は、話者やその関係者の個人名、屋号、一部の地名や団体名です。

また内容上非公開とした箇所もあります。

非公開手段

該当部分は方言テキスト・標準語テキスト上でX1, X2,... などと、なるべく同一性を保持したまま、記号で伏字化します。発話参加者を指すことが明らかである場合は、対応する話者記号（A, B など）で置き換えます。

方言テキスト	標準語テキスト	地点
アソコデ ヒーテタラ X1ノ オ バサンカ° ネー。	あそこで 挽いていたら (R: X1))の おばさんがねえ。	山梨_19_a_099
Bサン カサネーデ クンナイ ヨ。	(R: B)さん 貸さないで くださ いよ。	埼玉_11_a_099

また該当部分の音声は規制音でマスクします。この作業の都合上、標準語テキストでは該当部分にプライバシータグ（R）を付しています。

例外1：個人名そのものが話題となっており、メタ的言及を含めると秘匿しきれない場合がありますが、これは誹謗中傷を含まない場合に限り、許容して公開します。

例外2：いわゆる歴史上の人物、公人、話者の知己というよりは、調査地域に広く知られている人物名および団体名については、明らかな誹謗中傷を含まない場合に限り、プライバシー情報とはみなさず公開します。

2. 方言テキスト

本節では方言テキスト・データについて述べる。
中納言では文字列検索のみ対応。

2.1 書き起こし、表記

書き起こしは基本的に『ふるさとことば集成』本文および「各地方言収集緊急調査」報告手稿によるが、いくつかの改変を加えている。

	例	地点
長音や促音は、音声的な現象であっても原則として音声通りに表記する。	ソヤケド <u>コレニ</u> <u>アノニ</u> ベンジョニ <u>アノニ</u> (けど これ あの 便所に あの)	京都_26_d_099
ガ行鼻音は、カ°、キ°、ク°、...のように「°」を用いる。	カナラズ <u>ログカ°ズワ</u> ハー ソー ヤ スندا モンダモナ。 (必ず 六月は もう そう 休んだ も のだからな。)	岩手_03_c_099
前鼻音にはnガ、nダ、...のように「n」（全角のn）を用いる。	ソノ ハッタンノ ナカオ <u>ホリnダシ</u> <u>エ</u> 。 (その 八反の 中を 掘り出して。)	香川_37_e_099
合拗音[kwa]、[gwa]、...にはクワ、グワ、...のように小書きのワを用いる。	<u>シグワツ</u> マ <u>サングワツ</u> ユキガ キ エーヤニ ナート (四月 まあ 三月 [で] 雪が 消える ように なると)	島根_32_a_099
口蓋化した[je]、[je]はシェ、ジェと表す。中間的なものはセ、ゼかシェ、ジェかのいずれかで表す。	マー <u>フーシェンナラ</u> <u>フーシェン</u> オモ チャオ コーテ ヤッテ (まあ 風船なら 風船 おもちゃを 買って あげて)	兵庫_28_e_001-2
[ti]、[di]はティ、ディ、[tu]、[du]はトゥ、ドゥと表す。摩擦の弱いチ、ツも同様に表記する。	イエン マワリコージカラ アンター ショーガトウン <u>グワンジトウン</u> ヒワ (家の 持ち回りで あなた 正月の 元 日の 日は)	大分_44_c_099
[ɸa]、[ɸi]、[ɸe]はファ、フィ、フェのように表す。両唇の摩擦を伴うハ行も同様に表記する。	エー ゴドア ネアガタ。 シコーシ <u>デ</u> <u>ファレnバ</u> ケnダニ エデナヤ。 (よい ことが なかった。 少し 出る と ツツガムシ [が] いてね。)	秋田_05_e_099
半母音を伴う[je]、[wo]はイエ、ウオと表す。	イヤ ソリガナ <u>イエイヨニ</u> トッチョ リヤルカリ ソンゲシテ ハヨー オク ル。 (いや それがね 栄養 [を] とって い らっしゃるから そんなに 早く 起き る。)	宮崎_45_e_099
[æ]、[ɛ]のようなアとエの中間音はアエまたはエアと表す。	ケサ ミズー ミーン イタガ モー ウ チモ <u>ワリアエニ</u> サビシーンジャ。 (今朝 水を見に 行ったが もう 私 の家も 割合 寂しいのだ。)	岡山_33_d_099

以上については、『ふるさとことば集成』の方針とほぼ同じだが、下の点は『ふるさとことば集成』から変更、追加したものである。

方言特有の音

事例	例	地点
南琉球などの琉球諸語に見られる	ウヌ <u>ソーガツィ°ヌ</u> <u>キィ°スイ°</u>	沖縄_47_c_099

中舌母音には「イ°」「イ°」を用いる。ただし、東北などの本土方言では、「イ」か「ウ」のいずれかで表記する。	バーンナ <u>マーツィ°</u> <u>カニーヤ</u> シヤ (その 正月が 来る ときには 待ちかねてね もう)	
唇歯音[v]には「ヴ」を用いて「ヴァ」、「ヴィ」、「ヴェ」のように表記する。ただし、[f]は[ɸ]と同様に「ファ」、「フィ」、「フェ」と表記する。	バガドウ ウパータ イジューキィ° ティッスィ° <u>ツヴァガ</u> ウパータ イジューキィ° ティー (私が たくさん もらったと する あなたが たくさん もらったと)	沖縄_47_c_099
撥音「ン」と対立する、成節的両唇鼻音[m]には「◆」を用いる。	ツヴァタガミヤー <u>ミドウ◆キョーダイヌ</u> キヤーマイ アリューティー (あなたたちは 女兄弟たちも あると)	沖縄_47_c_099
無声化した撥音には「°」を用いて「ン°」と表す。	チソー チュワハリマン° ネー。 (「ちそお」 と言われますねえ。)	—
ダ行鼻音(子音全体が鼻音化したダ行子音)には、ガ行鼻音に倣い「°」を用いて「タ°」、「テ°」、「ト°」と表す。	ソレオ イマタ° ニ アノー ズニ ノッテ アノ マー ヤイヨル ワケヤナ (それを いまだに あの 図に 乗って あの まあ やっている わけだね)	兵庫_28_e_001-2

明らかな誤り・抜けの訂正

音声データと対照し、明らかに書き起こしのミスであろうと考えられる箇所に関しては、周囲とだいたい同じような方針で文字化した。また、書き起こしから漏れている箇所に関しては、微小なものに関しては文字化したが、数十秒以上にわたる中略については文字化せずそのままとした。

頻度の低い表記について

方言音声は多様であるから、カタカナの可能性を最大限に発揮して気持ちを表現したいところではあるが、検索の便宜、および沖縄などカタカナの特殊な書き分けを行う方言の便宜のために、なるべく簡潔に統一された表記法を用いる。そのため原典から変更した箇所がある。

小書き：ネエ～ネエ～ネーといった同母音連続はなるべく長音（ネー）に統一する。感嘆符も同様。異母音連続はソエ→ソエのようになるべく小書きを避ける。

四つ仮名：当該方言で四つ仮名の区別（九州の多くの地点、高知では区別をする）がない場合は、連濁によるものも含め、ザ行の「ジ」、「ズ」に表記を統一する。

記号類

使用する約物は句点（。）のみ用いる。読点（、）や疑問符（？）などは用いない。

『ふるさとことば集成』所収のテキストについては、注釈記号〔 〕が入っているが、中納言では現状不可視になっている。

句点について

文字化については、ポーズがあつて、意味的にひとつのまとまりを持つ文と考えられる個所に句点「。」を打った。句点は以下の項目を確認して付す。

1. 発話のターンの交替から判断する。
2. 文が閉じられているかで判断する（目安：終助詞、動詞・形容詞の終止形、命令形や、体言止め、言いさし等）。

次の手順で判断する。

1. 発話のターンの交替から判断する。
[ターンが代わっている] → 句点を付す（下記例①を参照）

- [ターンが代わっていない(一人の発話が長い)] →形態的特徴で判断
2. 文が閉じられているかで判断する。※チェック担当者の解釈に委ねる。
 (目安: 終助詞、動詞・形容詞の終止形、命令形や、体言止め、言いさし等)
 [文が閉じられている] →句点を付す。
 [文が閉じられていない] →句点を付さない。

①発話のターンの交替

話者	方言テキスト	標準語テキスト
B	トーモンカイ アリスタデバー。	東門会 [が] ありましたではないか。
A	アー トーモンカイ。	ああ 東門会。
B	ネサー オランドモ アノー ログネンバリ アエサ エッタンデセー。	ねえ 俺達も あのー 六年ばかり あれに 行ったんですよ。

②文が閉じられているところ (目安: 終助詞、動詞・形容詞の終止形、命令形や、体言止め、言いさし等)

環境	方言テキスト	標準語テキスト
終助詞の直後	ジューゴセン モラーテネー。	十五銭 もらってね。
終止形	ヤマー コエチャー コー ユク。	山を 越えては こう 行く。
命令形	フトツ ツクテ ミロ。	一つ 作って みろ。
体言止め	ソレワ ヒカ° ンノ チューニジ。	それは 彼岸の 中日。

2.2 記号およびタグ

分節音として認識困難なときは、星印(*)を用いて該当する箇所を示す。記号の個数はモーラ数と一致させるのが好ましいが、難しければ時間長を勘案して適当に定める。対応する標準語訳にも同様に付す。

方言テキスト	標準語テキスト	地点
マズ ニガイモ サンカイモ ン マガラ ツ オジデ コー オジ ダリ ナニカ * * * * *。	まず 二回も 三回も 馬から ツ 落ちて こう 落ちたり 何か * * * * * * *。	福島_07_c_001
ジダイカ° * * *カラ アレダ ケンド。	時代が * * *から あれだけど。	栃木_09_b_099
ワカエー シューニ * * * ケ ンキューカエー。	若い 人たちに * * * 研究会。	静岡_22_a_099

{ } : 非言語的要素タグ

{咳}、{笑}など、発話区間に入っている非言語音については波括弧{ }で表示する。標準語訳にも同様に付ける。

方言テキスト	標準語テキスト	地点
ナンダカ ムカ° ムチューデー {笑} ハナシ スルヨナ コト モ。 {笑}。	何だか 無我夢中で {笑} 話 [を] するような ことも。 {笑}。	栃木_09_b_099
アスグノ オ ムスコ X2 ツン ダッタгентモ。 {息を吸い込 む音}	あそこの オ 息子 [は] X2 とい うんだっけれども。 {息を吸い込 む音}	宮城_04_c_010
ハーハーハーテ イッテ {笑} {手打ち}	ホラホラホラと 言って {笑} {手打ち}	熊本_43_b_001

3. 標準語テキスト

中納言では文字列検索および短単位検索に対応しており、間接的に方言テキストを検索することができる。検索機能に最大限に供するため、標準語テキストの文体はなるべく統一する。

COJADS において、標準語テキストは方言テキストの逐語訳であることを目指しているが、一方で方言テキストを検索するための媒介としての役割を担わせていることになる。方言テキストへの忠実性と検索効率は両立させるのが難しい。なるべく方言の形態素ひとつひとつを生かそうとすれば、一般に標準語テキストの多様性が高まるが、検索でのヒット率は低下するからである。本節では、こうした要求をなるべく満足するための方針を記載する。

3.1 表記

方言テキストと同様に、使用する記号は句点（。）のみで、読点（、）や音調を表す疑問符（？）、矢印（↑・↓）は用いない。句点は、文の終止箇所、および発話のターンが交替している箇所、を目安に付加するが、厳密な規則は立てにくく、作業者の裁量によってよい。ただし句点の個数および位置は、方言テキストと標準語テキストとで一致させる。

アラビア数字は使わず、漢数字とする（例：12歳→十二歳、昭和35年→昭和三十五年）。固有名詞や個人名の記号化（1.4節）で例外的に許容される。

臨時的な言語形式に現れる引き延ばし音は長音符「ー」で表記し、母音を重ねることはなるべく避ける（例：おおい→おーい、ふうん→ふーん、へえ→へー）。

方言テキストに強調を表す特殊拍があっても、標準語テキストには反映させない。また談話停滞を示す語中の長音化を標準語テキストに反映させる必要はない。ただし、フィラーの場合はこの限りではない。

方言テキスト	標準語テキスト	地点
ナユー チーツト ニノー ヨ ケー ニノーチ イキャ ノマー ルシ	苗を <u>ちよつと</u> （×ちよーつと） ニ ノー 多く 担って 行けば めりこ むし	大分_44_b_019
ハイタマンマ オモテー トビダ シテ <u>ヨーレワ</u> ヤラレタネ。	履いたまま 表へ 飛び出して <u>これ</u> は（×こーれは） やられたね。	東京_13_a_099
<u>ナーガーナーガー</u> コノネ ナ デハラナクテネー。	<u>なかなか</u> このね ナ 出かけられな くてね。	岩手_03_c_099

文法的意味を担う要素は原則として、ひらがなで表記する。それ以外の漢字・カナの揺れに特に統一的な方針はない（検索上は語彙素としてある程度吸収される）。

方言テキスト	標準語テキスト	地点
カンジンナ <u>トキニ</u> キラレルン ジャ。	肝心な <u>とき</u> に 切られるんだ。	香川_37_e_099
スーテーセー うレー デュンニ ミヂラシー クとウエン。	来ていたね それ [は] 本当に 珍 しい <u>こと</u> だ。	沖縄_47_a_099

もともと『ふるさとことば集成』所収のテキストに付されていた引用記号（「 」）および注意書き記号（| |）は、あえて消すことはせず残してあるが、新たな追加はしない。

方言テキスト	標準語テキスト	地点
キューシャメン オリル トゴ エタテ スタサ エテガラ コエ ダベー。	急斜面 [を] 下りる ところ [は] いいけれど 下に 行ってから <u>こ</u> <u>れ</u> ぶつかるしぐさをしながら だろ う。	青森_02_e_099
Dに アッ シカクイ ヤツ	Dに あ 四角い やつか。	京都_26_d_099

力。		
----	--	--

これに加え、標準語テキストにはいくつかのタグが付されている。対訳上の不完全さを補うためのタグを3.2.2節、特定の言語表現を標識するタグを3.3節で後述する。なお COJADS モニター版には、これらのタグを検索する機能は実装されていない。

3.2 内容

3.2.1 基本方針

標準語テキストは、**方言テキストの逐語訳**を志向する。標準語訳として多少不自然であっても、逐語訳の方針をなるべく優先する。

可能な限り、方言テキストで用いられている語と同源の語を選ぶ。言い回しに揺れがある場合は、書き言葉として規範的なほうを優先し、縮約、省略などのくだけた文体とみなされる表現はなるべく避ける（例：って→と、しちゃう→してしまう）。

意味的に対応する同源語が標準語にない場合は、単に意識とする。とくに当該語彙が名詞である場合は、方言語彙の語形そのままに活かして後述の意識タグを用いて補足する方法をとってもよい。

方言テキスト	標準語テキスト	地点
モチゴメ コー <u>カシテ</u> アゲタンノ。	餅米〔を〕 こう <u>水に漬けて</u> 上げであるの。	京都_26_d_099
エー ヤキガ デテナー。	いい <u>土の層</u> が 出てね。	山口_35_e_099

方言テキストに現れる助詞が標準語と食い違うとき、それが言い間違いでなく、文法的に正しいと考えられる場合は、単に意識とする。ゴト、サカイのような、標準語訳に対応する語形のないものは、助詞の新作として適当な格助詞に置き換えて訳す。一方、動詞テ形＋カラなど助詞の拡張とみなされるものは、後述のZタグで拡張部分を標示する。ただし、オバは単に「を」とうつす。

方言テキスト	標準語テキスト	地点
オンナショーデネー。 マタ ガイナ <u>ショワ</u> オッタワネ。	女の人でねえ。 また 丈夫な <u>人</u> が いましたよ。	新潟_15_d_099
オクサंगा <u>ユタカヤマニ</u> スイチョットンジャラ。	奥さんが <u>豊山</u> を 好いていたんだよ。	大分_44_b_011
ヨソン シトオバ ギョーサン モゾガッ トコジャジナー。	よその <u>人</u> を 非常に かわいがるところだがね。	宮崎_45_c_099

標準語と方言で付属語の自然な順序が異なる場合があるが、これは適宜、標準語にとって自然な表現として訳し、特にタグ等によって明示しなくてもよい。助動詞の個数のずれも同様。

方言テキスト	標準語テキスト	地点
イスコニ <u>ハエマシタサーナケン</u>	いい具合に <u>生えた</u> そうですから	島根_32_a_099
イマノ ヨリーオ ソロエテワ ヤネタタキ <u>シマショッタ</u> 。	今の <縫り＝縫り縄>を そろえては やに叩き〔を〕 <u>しておりました</u> 。	島根_32_a_099
<u>オイシカリヨッタ</u> ヨー。	<u>おいしかった</u> よ。	大分_44_b_014
アキグチ ナッたら マツリ ッチュガ <u>アリヨッタ</u> ネー。	秋口〔に〕 なったら 祭り というのが <u>あった</u> ね。	福岡_40_a_099
ナンデモ オラ ホラ イドゴダン <u>イダッターガンダーカ</u> 。	なんでも 私〔は〕 ほら 親類だの <u>いたもんだが</u>	宮城_04_c_099
ミワー タベタカ タベンカ ソ	身は 食べたか 食べないか そんな	大分_44_b_014

ギナ コタ オボエンケド	ことは <u>覚えていないけど</u>	
ジェニモーケノ ホーガ サキー タチマスダケンネー。	お金もうけの ほうが 先に <u>立ちま す</u> からね。	島根_32_a_099

準体助詞として属格と異なる形式をもつ方言では、属格助詞や「その」などの一部連体詞と準体助詞とを重ねて用いることがある。この際、準体助詞を「もの」などの形式名詞で訳す。

方言テキスト	標準語テキスト	地点
ニタンブノ タンボ <u>ジブンノカ ニ</u> シリヤー。	二反歩の 田んぼ〔を〕 <u>自分のも のに</u> すれば。	新潟_15_d_099
スッド コッド <u>コノナオ</u> コー シテ ハジッコモ ヨロ ヨッテ	そうすると 今度 <u>このものを</u> こ う して 端っこも ヨロ 燃って	山形_06_c_099
ヨガ <u>カイゲンノター</u> リレッ ショガ アツチャライニー	よい <u>海軍のものは</u> 履歴書が あ るじゃないね	鹿児島_46_a_099
ホーシチェ ソン <u>ヨソノンナ</u> ムシガ クチェ シモータ チュケンド	そして その <u>他家のものは</u> 虫が 食って しまった というけれど	大分_44_b_002

3.2.2 特例

さまざまな理由で上の基本方針を満足できない場合がある。状況に応じて以下のようなタグを用いて標準語テキストを標識する。

意味不明記号 〃〃〃

方言が分節音としては明瞭に聞き取れるが、意味がとれない場合は、標準語テキストの対応する部分を全角スラッシュ（〃）で表記する。全角スラッシュの個数は任意。言いさし、言いよどみについては「言いよどみタグ（D）」を参照。

方言テキスト	標準語テキスト	地点
X11ドンノ ウシ ッチューテー ナー <u>ダンテ</u> イマノ	X11さんの 牛 と言ってなあ 〃 〃 今の	大阪 _27_d_002-3
ドテッ <u>***テンヤテ</u> 。	どてっ 〃〃〃〃〃〃。	大阪 _27_d_002-2
ア ゾーリ ツクルヤラ <u>タノン</u> ナカッタワネー。	あ 草履〔を〕 作ったりして 〃 〃 なかったわね。	石川_17_a_099

意訳タグ <A=B>

意訳タグは、標準語テキストの一部分（A）を山かっこで括り、等号をはさんだ右辺（B）に補足説明を記入するタグである。左辺Aは文節の一部を切り取る格好でも構わない。また右辺Bにはスペースを入れない。AとBは置き換えても標準語テキストの意味がそのまま通るように調整する。意訳タグは以下のような場面で用いる。

方言テキストは、いちおう逐語訳することができるが、そのまま標準語として読むと文意が異なる、あるいは誤読の可能性がある場合、意訳タグで自然な標準語テキストをつける。とくに「ヨーセン」系の不可能表現は以下のように、意訳タグとZタグを組合せて示す。

方言テキスト	標準語テキスト	地点
バサマ バサマ オメ <u>クネ</u> エ ゴスナッテ	「おばあさん おばあさん あなた <食べるに いいですか=食べら れますか>」と	青森_02_e_099
ワタシラ ヨー センサカイニ	私たち〔は〕 <(Z:AP) し ない=できない>から	和歌山_30_d_099

エシャー ナントモ シェネア モンダ。	医者は なんと <u>＜せられない＝ できない＞</u> ものだ。	秋田_05_e_099
コー イッタゴティ シテカラ タ チッター タブル	こう 煎ったように <u>＜してから＝ すれば＞</u> タ 少しは 食べる	大分_44_b_014
ショーバイニ ツグッテ ダシテ イル シトダケ アンダカラ。	商売に 作って 出して いる 人 だって <u>＜ある＝いる＞</u> んだから。	茨城_08_c_099
ソノママ イレテ タテンノカ キカネーガラ ワカンネ。	そのまま 入れて 立てるのか <u>＜ 聞かないから＝聞かなかったから＞</u> わからない。	茨城_08_c_099
アン トキ シジミガ シツカリ アッテヨー。	あの とき シジミが <u>＜しツカリ ＝いっぱい＞</u> あってね。	埼玉_11_d_099
ワた ミャーラ いりーラ いぢ てィち チューちビけー タッチ マッスグヨー。	私たち〔の〕 庭から 西から <u>＜ 出てきて＝蛇が出てきて＞</u> 尾 <u>＜で だけ＝だけで＞</u> 立って まっすぐ ね。	沖縄_47_a_099

標準語に対応する語彙はあるが、談話中で符牒のように用いられている語彙について、意識タグで平易な言い方を添える。あるいは、標準語としても不自然ではないが、文脈が不足しており、そのままでは（特にコーパスの検索結果として断片的に見た場合に）理解困難であると思われる場合に、適当な箇所文脈を追加する。また、数詞周りの元号や助数詞を補足する。

方言テキスト	標準語テキスト	地点
サ サンギョー ツーチ トリシ マリジャート ユー。	サ <u>＜産業＝群馬県産業試験場＞</u> ツーチ 取り締まりではと いう。	群馬_10_c_099
ンー アノー ココンチノ タ ネー ハズレタカラ	んー あのー この家の 種〔は〕 <u>＜外れた＝繭が採れなかった＞</u> から	群馬_10_c_099
オラネ ツリッ ツリ ッデ モ ンジャ ネーヤ。	俺はね ツリッ 釣り という <u>＜も のでは＝ちゃんとしたものでは＞</u> な いや。	栃木_09_b_099
ヨニン。 ソレニ	<u>＜四人＝四人来ていた＞</u> 。 それに	香川_37_e_099
ハー。 コトシガ ゴジューサン ネンジャーキナ	はい。 今年が <u>＜五十三年＝昭和五 十三年＞</u> だからね	大分_44_c_099

代名詞の指示対象や「ナニ スル」のような代用表現が本来指示したかったものを明示する必要はないが、テキストの理解に資するようであれば、しておいてもよい。

方言テキスト	標準語テキスト	地点
ナ。 ソナイナ モン ヤイヤイ イワン コトワ イワン。	ね。 <u>＜そんな もの＝生活が苦しい などと＞</u> やいやい いわない こと は いわない。	滋賀_25_e_099
ナン シタラ ソラー オイシ カッタ。	<u>＜あれ したら＝作ったら＞</u> それは おいしかった。	滋賀_25_e_099

標準語にない語彙について、意識タグを用いて方言語彙の音形を活かしてもよい。昔語りでは同一指示対象について複数の呼称を用いることもしばしばあるので、単に意識とするのか、〇タグを付すのか、以下のように意識タグを挟むかは、状況により判断すればよい。

方言テキスト	標準語テキスト	地点
コー アバーガ コーデ アスン デッダカラ。	こう <u>＜アバー＝漁網の縁＞</u> が こう やって 遊んでいるんだから。	千葉_12_c_099
モー アター ネコロンデカラ ヘソテンデ ニョーテモ ラク	もう あとは 寝ころんで（Z : A P） <u>＜へそ天＝仰向け＞</u> で 寝てい	岡山_33_d_099

ジャ ユーン。	ても 大丈夫だ [と] 言う。	
---------	-----------------	--

方言テキストは部分的に聞き取り困難、あるいは言いさしているが、文脈から意味がじゅうぶんに推測できるとき、意識タグで補足してもよい。

方言テキスト	標準語テキスト	地点
リョーホーデ スンジャ** イケガ コメーンジャ ソリヤー。	両方で <u>くするのだ**=するのだけれど</u> 池が 小さいのだ それは。	岡山_33_d_099

省略タグ []

標準語で丁寧に発話するときには存在する要素が、方言には存在しないような場合は、**省略タグ** [] によってその要素を明示する。ただし、この省略タグを使用するのは格助詞など、書き言葉の規範文法上、なければ非文法的とみなされるものや、一部の助詞に限り、単なる意味的な補填には用いない。方言文法としては「助詞の省略」とはいえない（無助詞格）場合も、対応する標準語テキストで何らかの助詞が必要ならば [] で明示される。

方言テキスト	標準語テキスト	地点
モー オモリ シテ モラワナ シヨナカッテンノ	もう お守り [を] して もらわないと <u>しかた [が]</u> なかったんだね	和歌山_30_d_099

縮約と判断される場合は、省略タグをつけない。

方言テキスト	標準語テキスト	地点
カウヤー ドサーツト。	<u>川に</u> ドサツト。	熊本_43_c_099
オリヤ アリヤ ウマイト オモーケドモ。	<u>私は あれは</u> おいしいと 思うけれども。	愛知_23_d_099

ゼロタグ (Z)

標準語にうまく対応する要素がないものについては、**Zタグ**でその存在を明示する。Zタグは対応する要素が訳出上存在しないことを示すタグである。Zタグとしては以下に示す、一部の文法要素のみを認めている。

副助詞 (AP)

助詞拡張に参加する語には様々な由来が認められるが、一律で副助詞と判断し (Z : AP) を置く。また「ヨーセン」系の不可能表現については、(Z : AP) と意識タグを用いて例のように記述する。沖縄の係助詞ドゥは標準語に訳出しにくく、しばしばこれで示される。

方言テキスト	標準語テキスト	地点
アノ ニギヤメス <u>ステカーヌ</u>	あの 握り飯 [を] <u>して (Z : AP)</u>	島根_32_a_099
タウエ タテ ムカシャー <u>キョードーデカラ</u> ヤリヨッタヨナ。	田植え といったって 昔は <u>共同で (Z : AP)</u> やっていたよね。	山口_35_e_099
イッペン <u>ネテカッテ</u> <u>トカエテ</u> <u>カエッテ</u> ツカワシャイ。	一度 <u>煮て (Z : AP)</u> <u>溶かして (Z : AP)</u> 使いなさい。	富山_16_d_099
ホンデ ゴハンヤッテ <u>ユートサーガ</u> モー オツソロシヤー。	それで ごはん <u>と いうと (Z : AP)</u> もう こわい。	岐阜_21_d_002
ソー <u>ストツシャガナ</u> ホラーモ サケノ ハイルクン	そう <u>すると (Z : AP)</u> ね ほら もう 酒が 入るから	熊本_43_b_002

ワタシラ ヨー センサカイニ	私たち [は] < (Z: AP) し ない=できない>から	和歌山_30_d_099
ヤーち ムル マチャー** ち ユミティ ムラネーティル フサ ギーテル。	家に みんな / / / ** 清め て 村で (Z: AP) お供えしてい た。	沖縄_47_a_099
ビキドウ◆ヌキヤーヤ ユードウ ツィ° キ ウキィ° サーイ	男たちは よく (Z: AP) つい て いたね	沖縄_47_c_099

指小辞 (DM)

方言テキスト	標準語テキスト	地点
** ナンダガサ ウリコ° エコ アレー エー ゲンマエパンノ ホヤホヤダテ	** なにか 売り声 (Z: DM) [が] あれは ええ 「玄米パンの ほやほやだ」と	青森_02_e_099
サワツコニワ イタモンネー。 モット。	沢 (Z: DM) には いたものねえ。 もっと。	栃木_09_b_099
ワーガモー フォーマイ イッカ ネンナ	豚 (Z: DM) を 食べるのも 一年 には	沖縄_47_c_099

終助詞 (SP)

終助詞については、ネイティブチェックで適当な対訳が得られればそれでよいが、これを経ても標準語テキストに終助詞が欠けている場合は (Z: SP) を置く。

方言テキスト	標準語テキスト	地点
ンダ。 スワnダノ ナガダガラ クッドワ。	そうだ。 諏訪だの 中田から くる と (Z: SP)。	宮城_04_c_010
オレタチダッテワネ メーデ。	俺達だって (Z: SP) ね ほら。	山梨_19_a_099
ソーヤッタナモ。	そうだった (Z: SP)。	岐阜_21_c_099
ミンナ イッテ サー キタノイ ケガ ハシャクゾヨト	みんな 行って さあ 北ノ池が 王 上がるぞ (Z: SP) と	愛知_23_d_099
クマデデ ヨシチャー パーット タン ナカデ ヤイテ シマウ ケードガ。	熊手で 寄せては パーッと 田の 中で 焼いて しまうけれど (Z: S P)。	岡山_33_d_099
ホージャロゾイナー。	そうだろう (Z: SP) ね。	愛媛_38_c_099
アリヤー ズルイ モンチャ ナ カッタゼヨネー。	あれは たやすい ものでは なかつ たよ (Z: SP) ね。	高知_39_b_099

接頭辞 (PR)

標準語ではつけると不自然になるような動詞接頭辞の使用に用いる。

方言テキスト	標準語テキスト	地点
ソナ モノー オツコシャー ショーチシネー チュンデ	そんな 物を (Z: PR) 壊せば 承知しない というので	群馬_10_e_002
フルシキズツミ カツコシラエテ	風呂敷包み [を] (Z: PR) こし らえて	新潟_15_d_099
ソーツオ コー ヤッテ ウロキ メテッテ	そいつを こう やって (Z: P R) 決めていって	岐阜 _21_d_001-1
ミズ ヒンノーダモン。	水 [を] (Z: PR) 飲んでしまっ たもの。	熊本 _43_b_004-1

敬語形式 (HON)

標準語ではつけると不自然になるような敬語形式の使用に用いる。

方言テキスト	標準語テキスト	地点
オホートー	(Z : HON) ほうとう	山梨_19_a_099
ノム ミズダケ イガワサン モ ライ イキヨラシタ。	飲む 水だけ 堀井戸 (Z : HON) [に] もらい [に] 行っておられ た。	熊本_43_b_001

助動詞 (A U X)

方言テキスト	標準語テキスト	地点
ナンダッテ ヨイジェーデ カッ タツケカ。	なんだって 共同で 買った (Z : A U X) が	福島_07_c_099

3.3 標準語テキストを修飾するタグ類

標準語には研究目的での利用を想定して、さまざまなタグが付けられている。ただし、COJADS モニター版にはタグ検索の機能が実装されていない。

人称タグ

方言形には実にさまざまな形の人称表現が現れる。これを将来に備えてタグ付けしておくためのタグ。標準語訳を () に入れて人称と単複を表すタグをつける。縮約している場合はほぐしてタグ付けする。人称は全角数字、単数は S G、複数は P L とする。沖縄には双数 D U もみられる。

方言テキスト	標準語テキスト	地点
マダ アタシガ オモー トーリ ニ ヤッチョラン。	まだ (1 S G : 私) が 思う とお りに やっていない。	山口_35_e_099
ヒトツモ タペラレンデッテ ホ ンデ アンタト ウラト ハンブ ンワケ ショーカ。	ひとつも 食べられないからと それ で (2 S G : あなた) と (1 S G : 私) と 半分分け しょうか。	福井_18_b_099
イマノ ヒトラチャ ソレ オラ チャホド ンマラト タベトラン モンヤチャ。	今の 人たちは それ (1 P L : 私 たち) ほど おいしく 食べていない ものだよ。	富山_16_d_099
マー オラガ コロワ ソンナ ロクネンシェーデ シューガク リョコワ ナカッタゾヤ	まあ (1 P L : 私たち) の 頃は そんな 六年生で 修学旅行は な かったですよ	三重_24_c_099
ムカシワ シンドー シタチ エ ライ モノヨ オンシャー。	昔は 苦勞を したと たいへんな ものよ (2 S G : お前) は。	高知_39_b_099
オヤ ナンノ ヨージダヤート オモテ イッタラ ワレオバ ヨメ ニ クレ チャンダンカ。	おや 何の 用事だろうかと 思って 行ったら (2 S G : おまえ) を 嫁に くれ というんだが	新潟_15_d_099
ヤッタチガ イッテカラ オラ イッデーッテ イイヤッタッゲ デー	「(2 P L : あなたたち) が 入っ てから (1 S G : 私) は 入るから」 と 言いなさったそうさ	熊本_43_c_099
シーシバ ワッてイグンナー コ ズカイ いッセンチュン デン ネー バヨー。	そうすると (1 D U : 私たち二人) とも こづかい 一銭たりとも お金 [が] ない わけね。	沖縄_47_a_099

類似複数と思われる場合は以下のように処理する。

方言テキスト	標準語テキスト	地点
--------	---------	----

アレ <u>ワシラニ</u> ユワシタラ ナー。	あれ [は] <u>(1SG:私)</u> などに 言わせたらねえ。	兵庫_28_d_099
<u>オレラモ</u> マイーニチ マイト ルハカイニ	<u>(1SG:私)</u> なども 毎日 参って いるから ※現状「(1SG:私) も」となっていますが、修正します	奈良_29_c_005

一般名詞の呼びかけ的使用、代名詞的使用には人称タグをつけない。意識タグにより談話参加者であるものを明示してもよい。

方言テキスト	標準語テキスト	地点
<u>バーサンギャン</u> マンマエジャツ デノー。	<u><おばあさん=C>のうちの</u> <真ん 前=真ん前が線路>だからね。	熊本_43_c_099

人称代名詞と同形の語であっても、人称代名詞として用いられていない場合は人称タグをつけない。「ワレ」や「テメエ」などの再帰代名詞としての用法、「アンタ」のフィラーとしての用法など。後者の場合はかわりにフィラータグ (F) をつける。

方言テキスト	標準語テキスト	地点
ホタラ <u>ワカ</u> ワルイ コト タナエ アゲテ オラカ ワル イヨーニ ヤッパリ オモトラ イ。	そしたら <u>自分が</u> 悪い こと [は] 棚へ 上げて (1SG:私) が 悪いように やはり 思っているよ。	香川_37_e_099
クギミタイノガ <u>アンター</u> ズーット ナロージョルダケジャ ローガエ	釘みたいなのが <u>(F:あなた)</u> ずっと 並んでいるだけだろうがね	大分_44_b_019

人称代名詞と名詞が一体化している場合

含まれている形態素が当該方言の人称代名詞の形態素と同一である場合は人称タグを付与するが、異なる場合は人称タグをつけない。特に「～の家」によく見られる。

方言テキスト	標準語テキスト	地点
<u>アンダエ</u> フルエテ ユーゲント ナンダエ ナンノッシャナテ	「 <u>(2SG:あなた)</u> [の] 家 [は] 古いって いうけれど 何代 [に] なるのですか」なんて	宮城_04_c_001
<u>オリカタン</u> シングサ モー ホ ント ジュータンヌ シータゴツ デケチオル	<u>(1SG:俺)</u> [の] 家の しんぐさ [は] もう ほんと 絨毯を 敷い たように できている	大分_44_b_015

言語片タグ (D)

話者がある言葉を発音しようとして失敗した (噛んだ) 場合や、意図的に途中で打ち切って別の表現に言い換えたりした場合に生じる意味をなさない言語片をDタグで標識する¹。Dタグの範囲はなるべく小さくすることとし、文節中、語中に言い間違い音が挿入されている場合はそこだけDタグをつける。文中で不完全な立場にある文節においても、語が完全な形を保っていれば、Dタグを付さず訳出する。

方言テキスト	標準語テキスト	地点
モ ジェンジェン タキモンノ モー <u>クサ</u> アノ ボローボロ	もう 全然 薪が もう <u>(D:ク サ)</u> あの ぼろぼろ [に] なり	佐賀_41_c_099

¹ 原典のひとつである『ふるさとことば集成』の標準語訳では、こうした言いよどみはバツ印 (××) により表現されているが、COJADSでは音形を活かしている。

イチナッパッテンナタ ツカワン モン。	ますけれどね 使わないもの。	
デ キタノイケノ ミズガ ハ シャイアデ シマウト。	で 北ノ池の 水が 乾い (D: ア) て しまうと。	愛知_23_d_099
マー イク イクベート オモー ダーヨ。	まあ 行く 行こうと 思う [ん] だよ。	東京_13_b_105c
ミチノリ ミチノリン。	道のり 道のりが。	神奈川_22_a_099
ム ムスヨ フタバヤマン ムス コグ ナンカ モー ソーリヨニ ナッチョル チュ	(D: ム) 息子 双葉山の 息子 が なにか もう 僧侶に なって いる という	大分_44_b_011

助詞を言い間違えており、かつ正しい助詞を言い直していない場合は、以下のようにする。

方言テキスト	標準語テキスト	地点
トーバカーガ カーチョーマスガ ネ。	十あまり (D: ガ) [を] 飼ってい ますけどね。	島根_32_a_099
コゲナ フーノ フクレチナ。	こんな ふう (D: ノ) [に] ふく らんでね。	福岡_40_a_003

節末、文末などで言葉尻が切れてしまったとき、相手の発話にさえぎられて発話がとまってしまったときなど、意識タグを組合わせて文意を補っても構わない。

方言テキスト	標準語テキスト	地点
ソノ アドーサ タ タデダエ ン。	その あとに (D: タ) <建てた よう=建てたようだ>。	秋田_05_e_099

フィラータグ (F)

「それを取り去っても伝達する文、談話の命題内容に変化を及ぼさないもの」(中島 2011²) をフィラーの定義として採用し、フィラー的機能を持つものに対しては、標準語訳に F タグを付ける。

方言テキスト	標準語テキスト	地点
マズ ムイガワ ホレ タナバダ デンダシ。	まず 六日は (F: ほれ) セタ のわけだし。	山形_06_c_099
オガト ソーダン シテ イヤ モッテ キタラ	おかあさんと 相談 して (F: いや) 持って きたら	福島_07_c_099
エー キョア ドーモ アンタカ タ ゴクロサンデス。	(F: えー) 今日は どうも あ なたがた ご苦労さます。	石川_17_a_099
アノ ソヤケド センソーチュー ニ アノー アノー ナニ マメ クレヨッタ。	(F: あの) そうだけど 戦争中 に (F: あの) (F: あの) (F: なに) 豆 [を] くれた。	滋賀_25_e_099
イシノ ダイショーオ マー ミ セル ワケダ。	石の 大小を (F: まあ) 見せ る わけだ。	新潟_15_d_099
ウーン ヤッパー イッシューニ ヤッパー ジューナンガグレー アガリヨッタキタナー。	(F: うーん) (F: やはり) 一つの漉し台に (F: やはり) 十何荷ぐらい 塩がとれていたから ねえ。	大分_44_b_004
アキタヤ アオ ホンシューノ マタ キタノ ホーワ マタ	秋田や アオ 本州の (F: ま た) 北の 方は (F: また)	岐阜_21_d_001-2

² 中島悦子 (2011) 『自然談話の文法』東京：おうふう。

ホーゲン アッテ <u>マタ</u> ゼンゼン チガーデナ。	方言[が] あって (F:また) 全く 違うからな。	
ソレガラ <u>コンド</u> アノ トギノ ギンジュグワ ケンドーカ。ヘドエ アノー ケーゴデ アッタヨ。	それから (F:今度) あの ときの 義塾は 剣道が ひどい (F:あの) 稽古で あったよ。	青森_02_e_099
ア。ンデ <u>マ_nズ</u> アカ。ラ シェデ イダダギス <u>マ_nズ</u> 。	はあ。で [は] (F:まず) 上がらせて いただきます (F:まず)。	宮城_04_c_107
ソレガナ オマハン コントキ ヨバレテワ シカラレタ モンジャロ。	それがね (F:お前さん) 何度も 呼ばれては 叱られた ものだよ。	岐阜_21_d_002
ムかーシヤ <u>やー</u> わー ヤちー シグニちメーカラ スグイ なー ソーガち うりー ヤン ミーち マちカンてィシチャー	昔は (F:おまえ) 豚[を] 焼いて 四五日前から すぐに もう 正月[を] その // // // 待ちかねてね	沖縄_47_a_099

オノマトペタグ (G)

方言に見られるオノマトペは、そのまま標準語として許容される場合は何も標識を付さないでよい。標準語に存在しない類のオノマトペは、Gタグで囲って方言形を示す。可能な場合は意訳タグ<A=B>で自然な標準語を付す(この際、無理にオノマトペで示す必要はない)。

方言テキスト	標準語テキスト	地点
ヒョロ ^ー ット シタ オトコジ	ひょろ ^ー と した 男で	大分_44_b_011
オラ カッター キケーワ ビンボービンボー ^ー ッテ ナル ヤズ。	私[が] 買った 機械は (G:ビンボービンボー)と 鳴る やつ。	福島_07_c_099
アノ ナカヤマニ オットーネ。ソノ オンチャンカ ^ー チンバコチンバコ シテ	あの 中山に いたよね。その おじさんが (G:チンバコチンバコ) して	高知_39_b_099
ムガス エダケーダオンネヤ。ジジド サッデサー。	昔 [は] 痛いくらいだものねえ。 <(G:ジジ) =ぐいぐい>と されてねえ。	宮城_04_c_010
ジブンバカイ ボトーボト サレンモン ^ー 。	自分だけ <(G:ボトーボト) =もたもた> できないものね。	佐賀_41_c_099
ヤッパリ アサイドバック ^ー ダバ ヒョータラクッタラ ^ー デー デバエーガノ。	やっぱり 麻糸 ^ー だけだと <(G:ヒョータラクッタラ) =腰がない> といえば よいのかな。	山形_06_c_099

固有名タグ (O)

外国語、古語、書名、作品名、場所名、その他の固有名詞、音や言葉に関するメタ的な引用などにはOタグを付す。小型辞典、現代書き言葉UniDicに掲載のないものを中心とする。

歴史上の人物や公人の名前には、RタグではなくOタグを付す。名字+個人名など、名前の複数の部分を含む場合はスペースを入れずに示す。敬称はOタグに入れない。地名+一般名詞(地名を冠した施設名など)は全体で一語とみてOタグに入れる。同定できない人名や地名は、ひらがなで書けばよい。

方言テキスト	標準語テキスト	地点
イマー ホトンド <u>ビリュ^ーシ</u> <u>ビョ^ー</u> ツツ ^ー ビョーキワ ミア ミカケネーヤネ。	今は ほとんど (O:微粒子病) という 病気は ミア 見かけないよね。	群馬_10_c_099

オレン トコアタリ イマジヤ アレダヨ ヨセマツダヨ	私の ところあたり 今では あれだ よ (O: 寄せ松) だよ	神奈川 _14_d_099
ソヤ ソヤ ドーキュー アンタ モ ホン モロタヤロ。 アノ フルサトノオモイデバナシ。	そうだ そうだ 同級 あなたも 本 [を] もらっただろう。 あの (O: ふるさとの思い出話)。	兵庫_28_d_099
アノー ニタノ ノーキョーヤ ネー。	あの (O: 仁多) の 農協だね。	島根_32_a_099
アリヤー マー コーヤマデ コ ニャーオエマー。	あれは まあ (O: こうや) まで 来なければだめだろう。	岡山_33_d_099
ホイテ オーサカノ ビーケーエ ホーソーニ イタンヤラ	そして 大阪の (O: BK) へ 放 送に 行ったんだ	和歌山 _30_d_099
ホンデ イマ マタ ナンダンナ チヨノフジカナー アシ。	そして 今 また あれですね (O : 千代の富士) かな 足 [を] 。	奈良_29_c_099
セーブケ° キノ ジョンウエイン ト オナジヨーナ コトー ヤッ テタロート オモーンダヨ。	西部劇の (O: ジョンウエイン) と 同じような ことを やっていただ ろうと 思うんだよ。	東京_13_a_099
コノコ° ロ モー アノ ゴイサ レマセ チュナ コト ユー オ バーサン モー シンデ シモー タネー。	この頃 もう あの 「(O: ごいさ れませ)」 というような こと [を] 言う お婆さん [は] もう 死んで しまったねえ。	和歌山 _30_d_099
ヨッチョイク° ルマッテ ユー クル アノー クルマヤカ° アッタ。	(O: よっちょいぐるま) と いう クル あの 水車小屋が あった。	山梨_19_a_099

4. 文節認定規則

「日本語諸方言コーパス」では、方言テキストと標準語テキストは**全角スペース**で分割されている。これは0.2秒を目安とした時間長単位と、文節という構造的単位に依拠している。

COJADS における文節は、方言テキストにおいて認定される単位であって、有意味でありつつ自然に区切れる発話の最小単位、原則として1個の自立語と0個以上の付属語を含む単位と規定されるが、付属語のみからなる文節も存在する(4.3節)。

標準語テキストのスペース区切りは方言テキストのそれに従属させる。したがって、方言テキストと標準語テキストはスペースの個数が一致し、またスペースによって意味的に対応する単位に分かち書きされる。文境界は文節区切りとなる。とくに句点(。)直後にスペースを入れ忘れないようにする。

方言テキストを文節に分かつ作業自体は、標準語の感覚で大部分がうまく行くが、話し言葉であること、また方言談話であることから、例外的な処理をほどこす必要がある場合がある。まず、意味上は2つの文節に分けられるものの、文節境界が融合して切断できない場合があるが、この場合は(切れないので)切らない。標準語形は融合していない形で示す。

方言テキスト	標準語テキスト	地点
アツバッテ ソコデ ノミクイナ ンターバ	集まって そこで <u>飲み食いなんて</u> <u>いえば</u>	北海道_01_b_099
ハマン カモー タックルゴトウ ナッタ。	浜の 釜を 炊き [に] くるように なった。	大分_44_b_011

また、表層的には文節に切れるが、慣習化しており複数の文節でひとつの意味単位をなすものや、いくつかの複合辞については、切らずに示す(4.5節)。

これとは別の原理として、発話区間の優越性に基づく例外がある。COJADS のテキストデータは機械的処理によって「発話区間」に分割されているが(1節)、言いよどみなどにより文節を構成する語同士が異なる発話区間に分かれてしまったときは、これをあえてつなげることをしない。つまりスペースの挿入に関しては、文節認定よりも発話区間の分割が優先する。以下の例は「」の位置に0.2秒以上の言いよどみがあり、前後で異なる発話区間に分かれている。

方言テキスト	標準語テキスト	地点
メヤグ カゲル「ヨンタ イヌデ ネーイア。	迷惑[を] かける「ような 犬で [は] ないよ。	青森_02_e_116
オーミソカーノ「ニ ススハライ ヤッチャウヨーナ ダンダン	大晦日「に 煤払い[を] やって しまうような だんだん	東京_13_a_099

4.1節以降は語類に基づいて、例を挙げながらプラクティカルな文節認定方針を記述する。

4.1 自立語のみの文節

4.1.1 連用修飾成分

副詞など。

方言テキスト	標準語テキスト	地点
ナンボ「ステモ マゲルンダド。	いくら「しても 負けるんだと。	青森_02_e_099
ソレ「ハッチリ「ワガッタノ ソ レワ。	それ[が] 「はつきり「わかったの それは。	福島_07_b_099
ネーサン「エライ「アワタダシイ ネー	姉さん「えらく「あわただしいね	新潟_15_d_099
ホッデ「ヤッパ「タノシカッタタ	だから「やはり「楽しかったよ 子供	熊本_43_b_099

イ コドモガ	が	
--------	---	--

連用中止法。

方言テキスト	標準語テキスト	地点
ニワノ ホーエ デ _レ スイシャゴ ヤノ ホーエ イキ _レ シテ。	庭の ほうへ 出 _レ 水車小屋の ほうへ 行き [など] _レ して。	島根_32_a_099
トッショリーバガリデ ナク _レ ワカイ シトタチモ	年寄りばかりで [は] <u>なく</u> _レ 若い 人たちも	北海道_01_b_099
ウシバックージャー ネ _レ ヤサ イモノモ ツクラッシャーケン。	牛ばかりでは <u>なく</u> _レ 野菜ものも 作られるから。	島根_32_a_099
エンリョ ナグ _レ ヤスデモ イヤー。	遠慮 <u>なく</u> _レ 休んでも いいよ。	青森_02_e_003
マー ミンナ オダヤガニ _レ オダ ヤガニデ マダ イジネン ノバ シ イジネン ノバシ	まあ みんな おだやかに _レ おだや かにで また 一年 延ばし 一年 延ばし	茨城_08_a_011
アンタノ オヤデ アリ _レ	あなたの 親で <u>あり</u> _レ	兵庫_28_e_001-2
アキーニャ モー アンタ カタ ンナギ モー シゴツ チュ シ ゴトウア アラ シェズ	秋には もう あなた 軒なみに もう 仕事 という 仕事は <u>あり</u> は <u>せず</u> _レ	大分_44_b_011

4.1.2 連体修飾成分

用言連体形、形連体詞などは独立の文節をなし、被修飾要素とは切り離す。

方言テキスト	標準語テキスト	地点
ソヤケン ブラクガクゲーカイ スン トキャ ソン タンニンワ ショウシー _レ コッチャナ	だから 部落学芸会 [を] <u>する</u> と きは その 担任は <u>忙しい</u> _レ こと だったね	大分_44_c_099
ケッキョク コワイ _レ クワ	結局 <u>堅い</u> _レ 桑	群馬_10_c_099
ソリヤー オーイ _レ トキニア ヒャクニンモ	それは <u>多い</u> _レ ときには 百人も	愛知_23_d_099
ユキガ フ _レ ズブンホド ト マッテネー。	雪が <u>降る</u> _レ 頃だけ 泊まってね。	島根_32_a_099
オーリャ コマッタ _レ モンダケド モ。 ソリャ アノー。	それは <u>困った</u> _レ ものだけれども。 それは あの。	愛知_23_d_099
ノサギサ クル _レ メァニ。	野崎へ <u>来る</u> _レ 前に。	岩手_03_c_099
アル _レ テードナー。	<u>ある</u> _レ 程度ね。	山口_35_e_099
ア オハギー ボタモチミタイ _レ モンワヨ。	ア おはぎ <u>ぼたもちみたいな</u> _レ もの はね。	滋賀_25_e_099

4.1.3 述語

用言終止形。

方言テキスト	標準語テキスト	地点
ナンツ エワレデモ _レ エ _レ	何と 言われても _レ いい _レ	宮城_04_c_011-1
ドーモ アレギリワノ ムラビト ワ イチバン _レ ウレシー _レ	どうも あれだけはね 村人は 一 番 _レ うれしい _レ	愛媛_38_c_099
デ タショノ シンルイダケア _レ トマル _レ	で よその 親類だけは _レ 泊まる _レ	福井_18_b_099
ヒツツケテ _レ サラク _レ	くつつけて _レ 歩き回る _レ	長崎_42_b_099

用言命令形。

方言テキスト	標準語テキスト	地点
ツナケ° ネ _n デ オイデ ミロー ┘	つながないで おいて <u>みろ</u> ┘	青森_02_e_003
アゲナ コトヤナンカ クライデ シタドモネー ソノ <u>デー</u> ┘イー コトデ。	あのような ことなど [は] 嫌いで したけどね その <u>出る</u> [と] ┘いう ことで。	島根_32_a_099

4.1.4 体言

無助詞格や助詞の省略された名詞。標準語訳では適宜、省略タグ [] で助詞を補う。省略と縮約の違いについては3.2節を参照。

方言テキスト	標準語テキスト	地点
ハー オサンカ° ニチ┘スキ° ッ ト ハジマッタンダヨ。	もう お三が日 [が] ┘過ぎると 始 まったのだよ。	栃木_09_b_099
ハナシガ <u>ヨコミチ</u> ┘ソビチャッ タガナ。	話が <u>横道</u> [に] ┘それてしまった じゃないか。	愛知_23_d_099

呼びかけとして用いられる体言。

方言テキスト	標準語テキスト	地点
<u>オトサン</u> ┘オッカサンノ イジメ ナンナエ。	<u>お父さん</u> ┘お母さんを いじめなさる なよ。	大分_44_c_099

4.1.5 接続詞

ソ系列も分析しない。4.5.1節「指示詞」も参照のこと。

方言テキスト	標準語テキスト	地点
<u>ソシテ</u> ┘コンナ ムシ オルカ° ヤ ユーテ。	<u>そして</u> ┘こんな 虫 [が] いるの だ [と] 言って。	富山_16_d_099
<u>ダケン</u> ┘ソン トキノ オーミズ ジャ ナカモン。	<u>だから</u> ┘その ときの 大水では ないですよ。	熊本_43_b_002
<u>シェヤンド</u> ┘カンガエテ ミタラ ソレオ アキラサマニ メン ドーテ ユエランダナ ココロデ	<u>そうだけれど</u> ┘考えて みたら そ れを あからさまに 体裁が悪くて 言えないだろうね 心で	兵庫 _28_e_001-2

4.1.6 感動詞、応答詞、フィラー、言語片

感動詞、応答詞、フィラーは単独で文節をなす。言語片は前後の文節に編入されたり、あるいはどの文節にも編入されず、宙ぶらりんの状態で存在したりする。フィラー、言語片については第3節も参照。

方言テキスト	標準語テキスト	地点
<u>イー</u> ┘ヨツカ _n ドデー	<u>えー</u> ┘四つ角で	青森_02_e_099
マイバン┘ <u>ソノ</u> ┘チャント。	毎晩┘ <u>その</u> ┘ちゃんと。	愛知_23_d_099
フトンノ センタクヤ┘ <u>アノ</u> ┘ワ ラガ	布団の 洗濯や┘ <u>あの</u> ┘自分などが	愛媛_38_c_099
<u>ハイ</u> ┘ナンドキデモ。	<u>はい</u> ┘いつでも。	石川_17_a_099
<u>ヘー</u> ┘コリヤ コン ワケー モ	<u>ほう</u> ┘これは この 若い 者は	大分_44_c_005

ンナ		
----	--	--

4.2 自立語＋１つ以上の付属語からなる文節

4.2.1 助詞

格助詞。標準語訳の方針については3.1節などを参照。

方言テキスト	標準語テキスト	地点
ソノ トキノ <u>クチョーワ</u> ヨケリヤーネー。	あの その ときの <u>区長が</u> よければね。	新潟_15_d_099
オマイノ イッテカラ ヨケナイナー。 チャバ キヤトッデーツテ イオッタタイ。	<u>あなたが</u> 入ってから 休んでおいでよ。 茶を 沸かしておくからと言っていたものだよ。	熊本_43_c_099
オクサंगा <u>ユタカヤマニ</u> スイチョットンジャラ。	奥さんが <u>豊山を</u> 好いていたんだよ。	大分_44_b_011
ソノ ビンボナ <u>エサ</u> イッテモハー。 オクダサニ イッテデスネ。	その 貧乏な <u>家へ</u> 行っても はあ。 <u>奥田へ</u> 行ってますね。	岩手_03_c_099 福岡_40_a_099
<u>ソエゴト</u> キレコデ コー クルンデー	<u>それを</u> 布で こう くるんで	青森_02_e_099
カイトヨー モー カイト カイト <u>カキカラ</u> カイト	書いてね もう 書いて 書いて <u>書きに</u> 書いて	岐阜_21_c_099
モー <u>サンダイカラ</u> ナンギュシヨッタジャナ。	もう <u>三代にわたって</u> 苦勞を していたよね。	山口_35_e_099
ヨーモリモッテ カーテ クレテネー	<u>傘を</u> 買って くれてね	新潟_15_d_099
キヤクホンヌ イー イー キニナッチ ジブンガ ツクッチ	脚本を 良い 良い 気に なって 自分が 作って	大分_44_c_099
ソレカ° ヤスインダー。 ヒヤグメ <u>ニエンニジュツシェンニ</u> レダ	それが 安いんだ。 百匁 <u>二円三十銭で</u> 売れた。	福島_07_c_099

係助詞。

方言テキスト	標準語テキスト	地点
タイガイ イッショニ アソンデル ヤツモ <u>ヨラ</u> シマセンヤロ。	たいてい 一緒に 遊んでいる 仲間も <u>寄りは</u> しませんでしょう。	京都_26_d_099
アッ コマー トキャ ソー メダタン ンー <u>オトコジカッタン</u> ジャガタナー。	あっ 小さい ときは そう 目立たない ううん <u>男でこそあったのがねえ</u> 。	大分_44_b_011
アガ ミヨ クチガ コエチオルキ コギナ モンナ <u>クヤー</u> 「シェン チュチ オヤッサंगा...	(2SG: お前) [が] 見ろ 口がこえているから こんな ものは <u>食べや</u> 「しない」 といって 夫が...	大分_44_b_014
アリー <u>ナルチコ</u> イヨッタンガ アンタ フタ ユタカヤマニ ナッタンジャキタナー。	あれに <u>なるとこそ</u> 言っていたのが あなた フタ 豊山に なったんだからねえ。	大分_44_b_011
カナラズイ° <u>ウイ°ドウ</u> スイ°ダラ。	必ず <u>いは</u> するでしょう。	沖縄_47_c_099

副助詞。

方言テキスト	標準語テキスト	地点
ネングオ ハカッター ノコル ヤツ <u>チョイトシカ</u> アラヘ ン。	年貢を 計ったら 残る もの [は] <u>ちょっとしか</u> ない。	滋賀_25_e_099
ドーコエモ イガネーデ コノ アタマノ <u>ウエデベ</u> ナッテン ミテーナンテ	どこへも 行かないで この 頭の <u>上でばかり</u> 鳴っているみたいなんて	埼玉_11_d_099
ホンヌ <u>ショードクホドワネ</u> イマンゴ シェワイテ シェナ イケンダケン。	本当に <u>消毒だけはね</u> 最近 [は] 一生懸命に しないと いけないか ら。	島根_32_a_099
アノ <u>ジシンダケノ</u> ジシンナ マー マイスミト スレバ ハジ メテジャ ネンカネー。	あの <u>地震ほどの</u> 地震は 今までに まあ 黛と すれば 初めてでは ないのかねえ。	埼玉_11_d_099
ソノ <u>アラリヤカテヨ</u> アオイ ヤツヤラヤ キーロイ ヤツヤ ラー。	その <u>あられだってね</u> 青い ものや ら 黄色い ものやら。	滋賀_25_e_099
<u>キョーラム</u> ココ デテタデ ショー。	<u>今日なども</u> ここ [に] 出ていたで しょう。	福井_18_b_099
<u>クビンジョー</u> コーナ	<u>首ばかり</u> こうね	大分_44_c_099

接続助詞。

方言テキスト	標準語テキスト	地点
ウシノ <u>コヤ=ナンカイ</u> ナカナカ デキマシエンダケンネー。	牛の <u>小屋なんて</u> なかなか できま せんからね。	島根_32_a_099
ダンゴト ユーナ コト <u>ユータ カテ</u> イマノ ヒト シラッサラ ヘンワ。	団子と いうような こと [を] <u>いっても</u> 今の 人 [は] ご存知 じゃないよ。	滋賀_25_e_099
<u>アゲタナリデワ</u> アカンネンナ。	<u>炊きつめた</u> だけでは だめなんだね。	滋賀_25_e_099
ミテ <u>トーリヨン=シェン=カ=デ</u> [笑]	見て <u>覚えているから</u> [笑]	長崎_42_b_099

以下のものは元の語彙的意味を喪失したとして、独立した文節として扱わず、切らない。

方言テキスト	標準語テキスト	地点
キノコ <u>ニシロ</u> ソー ユー ナリ モン <u>ニシロ</u> 。	茸にしる そう いう なりのものに しろ。	栃木_09_b_099
<u>フタリデモツテネ</u> ツツテネー。	<u>二人でもってね</u> つり上げてね。	新潟_15_d_099

一方、以下のものは切り離す。類似の事案として、4.5.5節「繰り返し」も参照のこと。

方言テキスト	標準語テキスト	地点
マ イマノ ヒトワ シアゲ <u>ユーカ</u> <u>ドーカ</u> シランケドナー	まあ 今の 人は 仕上げ [と] い うか <u>どうか</u> 知らないけどね。	兵庫_28_d_099
<u>バンヤ</u> <u>ナンカニ</u> コナーシドモ サーナンツ コトワ。	<u>晩や</u> <u>何かに</u> 脱穀など [を] する なんていう ことは。	島根_32_a_099

4.2.2 属格助詞、準体助詞 属格助詞

方言テキスト	標準語テキスト	地点
コクゴノ _レ シェンシェーダッタガ。	国語の _レ 先生だったが。	島根_32_a_099
オラエノ _レ ター マダ	私 [の] 家の _レ 田 [が] また	福島_07_c_099
アノ ソラマメノナー カンカラカンオヨー。	あの <u>そら豆のねえ</u> かたいのをねえ。	滋賀_25_e_009

独立属格

方言テキスト	標準語テキスト	地点
アー <u>キンタロサンノ</u> モ アッタノ _レ ソー ユワ。	ああ <u>金太郎さんの</u> も あったねえ そう 言えば。	和歌山_30_d_099
ハルノ サンブンオ ズーットマクケー。 <u>アンタガタナー</u> エー イロー シトルガ。	春の 蚕糞を ずっと まくから。「 <u>あなたたちの</u> は よい 色を しているよ。	岡山_33_d_099

準体助詞は全国的にかなりバリエーションがある。ゼロ準体は省略タグで明示する。コトなど形式名詞との区別も考慮する必要がある。形式名詞なら切るべきである。4.5.2節「形式名詞」も参照。

方言テキスト	標準語テキスト	地点
キデ <u>ツクッタノ</u> オ クミヨッタ _レ ンジャガ。	木で <u>作ったの</u> を 組んでいたのだが。	広島_34_c_099
ワタシワ ソノ <u>ミズイロノン</u> コーテ モーテナー	私) は その <u>水色の</u> [を] 買って もらってねえ	兵庫_28_d_099
ウゴイデ テマエ ジブンラチアハダケニ <u>ツクッタカ</u> 。オ。	ウゴイデ 自分 自分たちが 畑に <u>作ったの</u> を。	富山_16_d_099
コマカトバ アーノ ノゾイテ エー <u>フトカトバ</u> リッパ <u>フトカトノ</u> リッパカトオ ソノー	小さいのを あの 除いて えー <u>大きいのを</u> 立派 <u>大きいので</u> <u>立派なの</u> を その	長崎_42_b_099
ソレデ ニルダケド	それで 煮る [の] だけれど	山梨_19_a_099
アノ ネンニ イッカイワ ソノ <u>イセブネテオ</u> ダイテ	あの 年に 一回は その <u>伊勢舟と</u> いう [の] を 出して	愛知_23_d_099

4.2.3 終助詞、助動詞

COJADS のデータでは方言テキストに形態素分析を施すわけではないため、終助詞や助動詞の厳密な区分けは行わない。文節区切りにおいては、単にどちらも付属語として扱われる。終助詞については、3.2節の (Z : S P) も参照のこと。

方言テキスト	標準語テキスト	地点
コーケーシャワ <u>ナイラカナモ</u> 。	後継者は <u>ないでしょうか</u> (Z : S P) (Z : S P) = <u>ね</u> 。	岐阜_21_c_099
コレオ タベナカ _ラ <u>ヒケシッテ</u> イマノ オバサンカ _ラ ネ	「これを 食べながら <u>挽きなさいよ</u> 」と 今の おばさんがね	山梨_19_a_099
タイガイ イッショニ アソンデル ヤツモ ヨラ <u>シマセンヤ</u> ロ。	たいてい 一緒に 遊んでいる 仲間も 寄りには <u>しませんでしょう</u> 。	京都_26_d_099
イネコギ キカイガ <u>デキタゲナ</u> ナンツッテネ。	稲扱き [の] 機械が <u>できたそう</u> だ などと言ってね。	島根_32_a_099
アノ ガッコーガ スル チュート イウユル センシュ _ミ タイナ	あの 学校が する というと いわゆる <u>選手みたい</u> な ように なるね	大分_44_c_099

ゴト ナルワナ		
---------	--	--

モノについては、これで節が終止するときはこれを終助詞とみなすが、さらに別の終助詞が後接するときはこれを形式名詞とみなし、独立の文節をなすものとする。

沖縄の以下のコピュラ動詞については、独立の文節をなすものとする。

方言テキスト	標準語テキスト	地点
ヤッぱり クシユくイ ルヌ いミ _レ ンてーヤー	やっぱり <腰休め=骨休め> っていう 意味 _レ だったんだね	沖縄_47_c_099
ヤッぱり ムルー セイネン _レ ン _レ パートゥ	やっぱり みんな 青年 _レ なのだ _レ から	沖縄_47_c_099

ゴザル系

存在動詞としての本動詞用法は文節をなす。丁寧の～デ_レゴザイマスも規則通りに切る。ただし縮約したもの、～ザンス、～デ=ゲス、～デ=ガス、～デ=ガンスなどは自立性を失っているとみて切らない。またオイシュウ=ゴザイマスのような形容詞のテ形でない連用接続は切らない。

方言テキスト	標準語テキスト	地点
ソー ユー ミジアンナエーノ カダー _レ ゴザンナクテ トッテ モー。	そう いう 道案内の 方が _レ いらっしやなくて [は] とても。	宮城_04_c_099
ンデ ヨッポド メアノ ゴツテ _レ ゴザイスペ。	では よっぽど 前の ことで _レ ございましょう。	宮城_04_c_099
ダケン ダイブ ムカシノ モンデ _レ ゴジェマズワ	だから だいぶん 昔の もので _レ ございますよ	島根_32_a_099
ソーザンショネー。 オモチワ	そうでございましょうねえ。 お餅は	京都_26_d_099
ワダシ ナ コーリヤマモズヤデ ガスー ツワッダノ。	わたし [は] ナ 郡山餅屋でございます と言われたの。	宮城_04_c_010
ユーリンヤラネー カイチャラ ナツカシゴザイマスネ。	有隣だとかねえ 開智だとか なつかしゅうございますね。	京都_26_d_099
ヒトツツモ ゴワヘンデ アンナ モンワー。 エー。 メズラ シーゴワス エー。	ひとつも ございせんよ あんなものは。 ええ。 珍しゅうございます ええ。	大阪_27_b_099
モト ホラ マチノ ミチノ セボーゴザシタローガ。	もと ほら 町の 道が 狭うございましたでしょう。	長崎_42_b_099

4.2.4 付属語の連接

既に多くの用例に出てきている通り、切らずにまとめて扱う。

方言テキスト	標準語テキスト	地点
トンヤノー タテモンデ ノコッテルノア アー シンタクノ モノオキ=ダケ=ダンベ=ヤ ハー。	問屋の 建物で 残っているのは ああ 分家の 物置=だけ=だらう=やもう。	埼玉_11_d_099

4.3 付属語のみの文節

4節の文節の定義からは外れるが、付属語のみで文節として成立する例。

方言テキスト	標準語テキスト	地点
<A>マタズイ シヨッタモンネー。 <C>アシカラ チノ デヨッタモン ナー。 ナター。	<A>股擦れ していたものね。 <C>足から 血が 出ていたものね。 でしょう。	佐賀_41_c_099
<A>オケニ ミズニ ツケトイ テー。 <D>ネ。 カメニ コーネ。 ズーット。 <A>ズーット コー ワケテ。	<A>おけに 水に つけておいて。 <D>ね。 瓶に こうね。 ずーっ と。 <A>ずうっと こう 分けて。	京都_26_d_099

逐語訳では標準語として不自然な場合は、3節の方針に拠って適宜処置する。

方言テキスト	標準語テキスト	地点
ソン コロニヤ イーヤマニヤ キチャオランチャッタヂェ シタ ンバッ チャラセンチャッタ。	その 頃には 飯山には 来ていな かったから 知らないけれども < では=そうでは>なかった。	鹿児島 _46_a_099

なお、一文節の中にフィラーが貫入した場合は次のようにスペース処理する。言い間違いとは異なり、フィラーの前後にスペースを入れることに注意（4.1.6節）。

方言テキスト	標準語テキスト	地点
ソー ドコマデー アノー デ モー エー ッチュー コト ア レシマヘンワナ。	んー どこまで <u>あの</u> で もう い い という こと [が] ありませ んよね。	奈良_29_c_099

4.4 複合語の認定

自立語と自立語が連続していても、複合語をなしているのであれば文節に切らない。

4.4.1 複合名詞

属格助詞の脱落を認めるのであれば文節に切るが、複合語とみなすのであれば切らない。連濁やアクセント等、複合語と認められる証拠があればよいが、判然としない場合は作業者の裁量による。標準語訳として不自然な場合は、これに〔の〕を補ってよい。

方言テキスト	標準語テキスト	地点
ゴカ [°] ツセックワ アマリ ヤラ ンナー。	五月〔の〕節句は あまり やらな いな。	北海道_01_b_099
アンマリ ハヤク ムロ イレッ ト アメ フッたら カワミズ デたら ムロエモ ハイッテ ク ルシテ。	あんまり 早く 室〔に〕 入れる と 雨〔が〕 降ったら 川の水 〔が〕 出たら 室へも 入って くるしで。	北海道_01_b_099
アサクサカリ ドッテ マー	朝草刈り といって まあ	福島_07_c_099
キワダツカワドガデー アノ ニ ト ホレ ユデットー カワ デル ズワツチェ	キワダの皮とかで あの ニ ト ほら ゆでると 皮〔が〕 出る と言われて	福島_07_c_099
イマナンド テレビミパッカ シ テラレンノー ホニ。	今など テレビ見ばかり してい られるの 本当に。	福島_07_c_099
アノー ネンジンバ コマコー ワカイ トコ トッテ コー シ	あの 人参の葉〔を〕 細かく 若 い ところ〔を〕 取って こう	福井_18_b_099

タリネー。	したりね。	
ンーン。 <u>カンショーオヤジーサ</u> <u>ンヤッタケンナター。</u> アンタン シッタゴト。	うん。 <u>短気なおとうさん</u> でしたか らね。 あなたが 知っているとお り。	佐賀_41_c_099
アー マタ ン クチョーモ <u>オーダクチョー</u> ジャキタナー	ああ また ン 区長も <u>大字区長</u> だからねえ	大分_44_b_019
<u>ソーガチヂン</u> フシルンリ いー バ ソーサちー。	<u>正月</u> <u>〔の〕</u> <u>着物</u> <u>〔を〕</u> 着せると 言えば <うれしがって=うれし がっていた>。	沖縄_47_a_099

特に、前部要素が1人称代名詞である場合や、後部要素が場所・時間をあらわす名詞である場合に、しばしば属格助詞が現れない。

方言テキスト	標準語テキスト	地点
<u>アンタ</u> <u>オタクヘンモ</u> <u>ヨメクサ</u> <u>ン</u> <u>デンサロモン。</u> <u>アンタンガ</u> <u>タヘンモ。</u>	あなた <u>〔の〕</u> お宅なども お嫁さ ん <u>〔は〕</u> 出られるでしょう。 <u>あ</u> <u>なたの家</u> なども。	佐賀_41_c_099
<u>ワた</u> <u>ブスー</u> <u>ノーギョー</u> <u>ネッ</u> <u>シン</u> ヤてン バー エとウ	私たち <u>〔の〕</u> 祖父 <u>〔は〕</u> 農業 <u>〔に〕</u> 熱心 だった わけ だか ら	沖縄_47_a_099
<u>オークラヘンニ</u> <u>アンダケ</u> イ シャガ アリヤーネー。	<u>大蔵あたり</u> に あれだけ 医者が あればね。	福岡_40_a_099
<u>ヒロツパノー</u> <u>イデトコロノ</u> ホーワ	広々した所の <u>井手</u> <u>〔の〕</u> <u>所の</u> ほ うは	山口_35_e_099
<u>ガツコムケナ</u> <u>モンダガラ</u>	<u>学校</u> <u>〔の〕</u> <u>向かい</u> な ものだから	北海道_01_b_099
<u>ショーガツコー</u> =ズダイデモ ナン ダイダネー。	<u>小学校時代</u> でも なんだね。	島根_32_a_099

4.4.2 複合動詞

動詞連用形+動詞からなる複合動詞を認定する。複合語の判断基準は標準語に準じてよい。敬意や相に関する補助動詞には縮約したものも多い。

方言テキスト	標準語テキスト	地点
イッソ ハー ネモトカラ <u>ヒキ</u> <u>タオセーテ。</u>	いっそ もう 根元から <u>引き倒せ</u> と。	山口_35_e_099
ソナ <u>ゴド</u> <u>イッテ</u> <u>オイゲー</u> <u>ス</u> <u>ワゲニ</u> <u>イカ</u> <u>ネンダヨ。</u>	そんな こと <u>〔を〕</u> 言って <u>追い返</u> <u>す</u> わけに <u>〔は〕</u> いかないんだよ。	茨城_08_a_011
モ イッキ <u>イーオヂェツ</u> <u>コ</u> <u>セツ</u> <u>ヒッチャユツ</u> <u>トゴイオ</u>	もう 一機 <u>射ち落として</u> こうして 落ちる ところを	鹿児島 _46_a_099
ヨビニ <u>キハリマンネン。</u> キ=ト クナハレヤ ユテ。	呼びに <u>来られるんですよ。</u> 来てく ださいよ <u>〔と〕</u> 言って。	大阪_27_b_099
カイコー シテ ヘトヘトニ ナ ンジャー カイ <u>カイアゲトン</u> <u>ノー</u> <u>コンダー</u> <u>ムギノ。</u>	養蚕を して ヘトヘトに なんだ カイ <u>飼い終わっているのを</u> 今度は 麦の。	岡山_33_d_099
モー ハコー <u>ツミカサネテ</u> ネ ガ <u>デハナエタラ</u>	もう 箱を <u>積み重ねて</u> 根が <u>出始</u> <u>めたら</u>	山口_35_e_099
トチト ユー モノワ ヒトリカ 。 <u>モチツメル</u> <u>モノチャ</u> ナ イ。	土地と いう ものは 一人が <u>持ち</u> <u>続ける</u> ものではない。	高知_39_b_099
ワったー いー あとウラ うー てィ ユナガとウナー <u>うーたッ</u>	私たち <u>〔は〕</u> おまえたち <u>〔の〕</u> 後 ろから 追って 夜通し <u>追い続けて</u>	沖縄_47_c_099

ちゅたん パーてー。	いた わけね。	
------------	---------	--

方言テキスト	標準語テキスト	地点
コノ フキンノ カクメーシガ オネガイ＝モーシテ。	この 付近の 各名士が お願い申し て。	埼玉_11_d_099
クチョサンカ° アンデ クレッ テ ユー＝マシタ ナーン ホンナ コト キレーニ デキンワニ ツタラ	区長さんが 編んで くれと おっ しゃった 少しも そんな こと [は] 上手に できないよ と言っ たら	石川_17_a_099

「二字漢語＋スル」「外来語＋スル」のような構造は文節に分ける。この際、標準語訳であえて格助詞「を」を補って示す必要はない。

方言テキスト	標準語テキスト	地点
ミナ ヤッパ シンケンニ シン ジンシヨッタキナ。	みんな やっぱり 真剣に 信心し ていたからね。	大分_44_c_099
フン ヤスンドルケーナー キューコーシトルケー。	うん 休んでいるからな 休耕し ているから。	岡山_33_d_099
ホイデ ワシ ホレカラ アンシ ンシトンジャ。	それで 私 それから 安心し ているのだ。	徳島_36_b_099
オラヤチャ イマノ ショーガッ コーオ ソツギョーステカラ	私たちは 今の 小学校を 卒業し てから	島根_32_a_099

4.5 例外的な規定

区切りに関して特に基本方針からの例外が多いものについて、具体的に記載する。

4.5.1 指示詞

コー・ソー・アー・ドーは縮約していなければ、少々意味が慣習化していても切る。

方言テキスト	標準語テキスト	地点
ワタシャー ライネンワ ソー ユー フーニ ナリヤー エーガ ト オモーテナー。	私は 来年は そういふ ふうに なれば いいがと 思ってね。	山口_35_e_099
マタ コドモカ° コーシテ タベテ クレルノモ イー ウレ シー コトダナ。	また 子供が こうして 食べて くれるのも いい うれしい ことだ ね。	神奈川 _14_e_099

縮約しているものは切らない。指示詞自体が音形を保っていても、これを文節としてくりだすことで（ソーイのように）文節未満の言語片が生じる場合は切らない。

方言テキスト	標準語テキスト	地点
ソーイ コトダッタカイナ	そういう ことだったかな	長野_20_e_099
アシガ コーヨーニ ダンダンボ ト ヘーリョータンジャ。	足が このように ダンダンボと 入っていたのだ。	岡山_33_d_099
ホシタラ ガッコーノ ニシノ	そうしたら 学校の 西の	群馬_10_c_099

「コノゴロ」「ソノゴ」「コノクライ」「ソノママ」のように一語化しているものは切らない。

方言テキスト	標準語テキスト	地点
コノゴロ オミウケ シマヘンケド	<u>この頃</u> お見受け しませんけど	大阪_27_b_099
ソノコ [○] ナンクウエモ ミ ⁿ ジ アタノモヨ。	<u>その後</u> 何回も 水害 [が] あった けれどね。	秋田_05_e_099
ホスダローヨ。 <u>ソノママデワ</u> ダメダカラ。	干すだろうよ。 <u>そのままでは</u> だめ だから。	茨城_08_c_099
コノッサンダチノ ユワシャルカ [○] キートッタカチャ ナン キ ノドクデ キートレン モンジャ チャ。	<u>この人たちの</u> 言われるの [を] 聞 いていると どうも 気の毒で 聞い ておられない ものだね。	富山_16_d_099
<u>アノコニ</u> コネアイ コネアイ シタランカト ユーテ	あの子に こう こう してやらない かと いって	兵庫_28_e_099

4.5.2 形式名詞

形式名詞は原則的に名詞と同様に扱うが、付属語化していると判断したものについては、文節に切らずつなげる方針をとる。その度合いにより、以下の2つのパターンで処理している。

モノ、ハズ、ワケ、フーなどは、自立度が高いとみて、格助詞に接続するもの、用言連体形に接続するもの、いずれも文節に区切る。

方言テキスト	標準語テキスト	地点
ナントカ ナラン <u>モン</u> ダーナン テ オモイヨッタ。	どうにか <u>ならない</u> <u>もの</u> かななどと 思っていました。	島根_32_a_099
マー アレワ マー ユーガ ヨー ホッタ <u>コト</u> イナ。	まあ あれは まあ 言うが よく 掘った <u>もの</u> よね。	山口_35_e_099
オイズナノ バアイワ アワセル <u>ワケ</u> カ。	追い綱の 場合は 合わせる <u>わけ</u> か。	山形_06_c_099
ミンナ タネムギデ ダス <u>ワケ</u> ノガ ヒトーツモ ウカラネーデ サー	全部 種麦として 出す <u>はず</u> のが ひとつも 合格しないでさあ	埼玉_11_d_099
アー ユー <u>フー</u> ナ エンターガ デキルヨーニ ナッタンダ ノー。	ああ いう <u>ふう</u> な 堰堤が 出来る ように なったんだねえ。	群馬_10_e_005

ただし以下のゴドは格助詞とみる。

方言テキスト	標準語テキスト	地点
<u>ソエゴト</u> キレコデ コー クル ンデー	<u>それを</u> 布で こう くるんで	青森_02_b_099

ヨー、ゴツなどは、格助詞に接続するものは文節に区切るが、用言連体形に接続するものは区切らない。

方言テキスト	標準語テキスト	地点
タキモンオ タクシ マ <u>ムカシ</u> <u>ノ</u> <u>ヨーニ</u> タクサンワ タカン	焚き物を 焚くし まあ <u>昔の</u> <u>よう</u> <u>に</u> たくさんは 焚かない	長崎_42_b_099
エツパエ ノム <u>ヨーニ</u> コシャ エッカラ	一杯 飲む <u>ように</u> 料理するから	宮城 _04_c_102-1
イマン タウエン <u>ゴツ</u>	今の 田植えの <u>ように</u>	大分_44_b_019
ナカッタゴタル。 キオクガ ネー ガクゲーカイ チューノワ	なかった <u>ようだ</u> 。 記憶が ない 学 芸会 というのは	大分_44_c_099

4.5.3 引用助詞類

ト・ッテ・テ・チなどの引用助詞も原則として、直前要素とともに文節をなすとする。引用標識の標準語訳としては可能な限り「と」を用い、「って」を避ける。

方言テキスト	標準語テキスト	地点
ンニャ ネゴツナッタト ユーノ ワ	いや なくなった <u>と</u> いうのは	宮崎_45_c_099
カンザキノ ドコントガ ウマカ ケン <u>テ</u> ユーテナンター。	神埼の どのが おいしいから <u>と</u> 言ってますね。	佐賀_41_c_099
ヨーコ° <u>ッテ</u> ユーノカ ヤリカ タノネー コトバカ° ミンナ チカ° ウノ。	用語 <u>と</u> いうのか やり方のねえ こ とばが みんな 違うの。	東京_13_a_099
カゼ フケヘン <u>テ</u> 。	風 [が] 吹かない <u>と</u> 。	徳島_36_b_099
コラ ホンナラ ヒツパッチョ キャ イーワ <u>チ</u>	これを それなら 引いておけば い いよ」 <u>と</u>	大分_44_b_002

トイウ相当の要素が縮約し、チ・チュ・ズ・ツーなどとなっているものは、直前文節に入れるのではなく、トイウで独立の文節とする。トイウは文頭に生起するなど、独立性が高いと考えられる。

方言テキスト	標準語テキスト	地点
ケンカ ズ モノ スタ ゴドァ ネ	けんか <u>という</u> もの [を] した ことは ない。	岩手_03_c_099
ヒトナンノ コー イネード オ モッテ <u>アルグ</u> ズド ケツツマ ズイダリ ナンカ シテナ	人なんか こう いないと 思って <歩く <u>という</u> と=歩くと> けつま ずいたり なにか してな	福島_07_b_002
ナーシ キタ <u>ツチャ</u> アルカ バカモンメッテ イテ	何しに 来た <u>ということが</u> あるか ばかもめと <u>言</u> って	新潟_15_d_099
ヨー ソノネー フーフ <u>ムスバ</u> <u>レタッテノニネー</u>	よく そのねえ 夫婦 [が] <u>結ばれ</u> <u>たというのにねえ</u>	兵庫 _28_e_001-2
イヤー ヨー サー <u>ツー</u> コト ワ ナイ。	いや よく する <u>という</u> ことは ない。	島根_32_a_099
ジシンカ° ユル <u>ツチュータ</u> ヤマカラ ヘビカ° オリテ ク ルトカ	地震が 起こる <u>といったら</u> 山から 蛇が 下りて くるとか	徳島_36_b_099
クサモ ナニモカモ アツメテイ テ ヤクワ <u>ツチューテナ</u>	草も 何もかも 集めて 焼くよ <u>と</u> <u>いってね</u>	徳島_36_b_099
ナン <u>チューカナ</u> 。	なん <u>という</u> かね。	福岡_40_a_099
オリガ シャクルドー <u>チュチ</u>	おれが 汲み入れるぞ」 <u>と</u> 言 <u>って</u>	大分_44_b_002

トイウ相当の形式が文頭などで接続表現の一部として用いられる場合は、融合していなくてもスペースを入れない。

方言テキスト	標準語テキスト	地点
ヤリヨシ ヤリヨシ チューター ヤラサレタ。 <u>チューノー</u> マー ワタイ ゴハン タベイ トキ コボスサカイヤケド。	やりなさい やりなさい <u>と</u> 言 <u>って</u> やらされた。 <u>という</u> の [は] まあ 私 [は] ご飯 [を] 食べ る とき [に] こぼすからだけ ど。	和歌山_30_d_099
<u>テュー</u> コトワ アツー マクト	<u>という</u> ことは 厚く まくとね。	山口_35_e_099

ナ。		
----	--	--

ソウ、ソモー、トモーなどの語形は場合により「そう言う」「そう思う」あるいは「という」「と思う」などと訳す。

方言テキスト	標準語テキスト	地点
タ _n ビ ケー ッタキヤ	「足袋[を] <u>くれ</u> 」 <u>と言ったら</u>	青森_02_e_003
サゲー イッショー ウ _n ド ウツテクナエ ツワエッデ	酒[を] 一升 ええと 売ってくだ さい <u>と言われて</u>	宮城_04_c_099
ソレ チット ナラーセテ ク ロー <u>ドッテ</u> 。	それ ちょっと 習わせて <u>くれ と 言ッて</u> 。	福島_07_c_099
ジシン カミナリ <u>ジューケンド</u> カミナリモ コエデ。	地震 雷 <u>というけれど</u> 雷も <u>これ で</u> 。	埼玉_11_d_099
デアージニ トリ アノ トリア ツカエ <u>トツタラネ</u>	大事に トリ あの 取り扱え <u>と いったらね</u>	東京_13_b_119
ワレオバ ヨメニ クレ <u>チャン ダンカ</u>	おまえを 嫁に <u>くれ というんだが</u>	新潟_15_d_099
ウチノ オレカ ^o ノ コドモワ ババーダ <u>ソウシ</u> 。	うちの 私の 子どもは 「ババー」 だ <u>と言うし</u> 。	長野_20_e_099
コメン トレナエート コマル <u>トモッテ</u>	米が とれないと 困る <u>と思ッて</u>	静岡_22_a_099
ナカ ^o サカデモ <u>ソツタダヨ</u> 。	長坂でも <u>そう言ッたぞ</u> 。	長野_20_e_099

引用に際して引用助詞ト・ッテを用いない方言（「と」抜けの方言）は、標準語テキスト引用箇所の末尾文節に省略タグで【と】を補う。またイッテ、トイッテ相当の形式が引用標識の機能をもつ方言では、下例のように意識タグ＜A=B＞で標示する。

方言テキスト	標準語テキスト	地点
ソシテ イッたら ツナビキ ス ル モンカ ^o <u>タラン ユーテ</u> 。	そして 行ったら 綱引き する 者 が 足りない【と】 <u>言ッて</u> 。	富山_16_d_099
オーサカベン ンナン ツカウ ナー <u>ユーテ</u> 。	大阪弁 そんなの【は】 使うな 【と】 <u>言ッて</u> 。	大阪_27_b_099
メズラシー シッテ ソーヤ。	珍しい < <u>と</u> 言ッて > <u>と言う＝と言う よ</u> 。	長野_20_e_099
ハヨー シタ ホーガ マンガ エカッタローガナ <u>ユーテ ユー ナー ユータンジャケド</u>	早く した ほうが 運が よかった だろう【と】 < <u>言ッて</u> <u>言うには＝ 言うには</u> > <u>言ッたのだけれど</u>	岡山_33_d_099

ナンテ（＜ナドト）やトカ相当の形式は、文節規則上は例外とせず、ナンテイウやトカイウで単独の文節と扱うことはしない。【と】を補うかは作業者の裁量による。

方言テキスト	標準語テキスト	地点
ムカシワ マー <u>ムロナンテー</u> 、 <u>ユー</u> ゴトカ ^o <u>アッテモ</u>	昔は まあ 室などと <u>いう</u> こと が あっても	北海道_01_b_099
オラン ウチデワ マー <u>カザツ タナンチュー</u> ゴト メギラニ アッタダ。	私の うちでは まあ <u>飾ったなど という</u> ことは // // あった んだ。	北海道_01_b_099
チョット ユクト オチャ ノン デ <u>イカッセヨーナンテッテ</u> ヨ ンデ クレルケド	少し 行くと お茶[を] 飲んで いきなさいよなど <u>言ッて</u> 呼ん で くれるけれど	岐阜_21_c_099
ナベサカエ アガートカ ジョー ヤマエ <u>アガートカ</u> イーヤナ。	鍋坂山へ 登るとか 城山へ 登る とか【と】 <u>いうような</u>	島根_32_a_099

4.5.4 動詞のテ形

動詞テ形は何事もなければ、規則通り以下のように切る。

方言テキスト	標準語テキスト	地点
ハコ イレチャーネー ショイ アルッテ「ドンドンドンドン モッテ」アルッタ。	箱 [に] 入れてはねえ 背負い 歩 いて「どんどんどんどん 持って」歩 いた。	群馬_10_c_099
ワシノ ホーガ オモイタッチ「 カイカク シタンジャケナー。	私の ほうが 思い立って「改革 し たのだからね。	福岡_40_a_099

動詞テ形を含む文節はしばしば後続する文節と縮約して、区切ることができない。

方言テキスト	標準語テキスト	地点
ニンゲンオ ノセテメーッテ ユー フネナケラナ	人間を 乗せてもいいと いう 船 [で] ないと	愛知_23_d_099
ソレ オラ ジーチャ オンネ ユーテカシタ モンダ	それ [を] 私 [の] じいちゃん [は] 私に 言っ <u>て</u> 聞かせた もの だ	新潟_15_d_099
ヤッチャーマスガ マー コイツモ ジェニニ ナーヤン ナーダヤ	やっ <u>て</u> はいますが まあ これも お 金に <u>なるように</u> なるのか	島根_32_a_099

この種の縮約はテ形＋補助動詞に多く見られる。補助動詞の音形が保持されている場合は規則通りにスペースを入れる。いずれの場合も、標準語は縮約していない形とする。

方言テキスト	標準語テキスト	地点
ンマヤガラ ゼンブ ソーコガラ ゼンブ <u>マイテ</u> ヤル。	馬屋から 全部 倉庫から 全部 ま いて <u>やる</u> 。	北海道 _01_b_099
ソサ <u>トマッテ</u> イルンダモノ タベタ アトカラー マタ <u>オサ エテ</u> シマウンニヤ。	そこに <u>泊まって</u> <u>いる</u> の <u>だ</u> もの 食べた 後から また <u>押さえて</u> し ま <u>う</u> の <u>だ</u> 。	山形_06_c_099 福井_18_b_099
<u>ツクッテ</u> <u>ゴシエ</u> イー コトデ ステナー。	<u>作って</u> <u>くれ</u> [と] いう こと <u>で</u> て <u>ね</u> 。	島根_32_a_099
ソシテ オラネ コーモリ <u>サシ テケテ</u> イワン。	そして 私に 傘 [を] <u>さして</u> い <u>け</u> と <u>いう</u> んだ。	新潟_15_d_099
ハヤラト アノ シャバエ デテ ハタラエテラッシャル	早く あの 世の中へ 出て 働いて おられる。	富山_16_d_099
チョット オチャ <u>ノンデカッセ ヨー</u> アー テッテ。	ちょっと お茶 [を] <u>飲んで</u> いき <u>な</u> <u>さいよ</u> ああ <u>と</u> 言 <u>っ</u> て。	岐阜_21_c_099
キモノ <u>コガシテ</u> マウ。	着物 [を] <u>焦が</u> してしま <u>う</u> 。	京都_26_d_099
ヒョージュンゴ <u>ツコーテ</u> ハッタ ト ワタシワ オモイマスナー。	標準語 [を] <u>使</u> っ <u>て</u> お <u>ら</u> れたと 私 は 思 <u>い</u> ます <u>ね</u> え。	大阪_27_b_099
ミズテンデ <u>ヨータレ</u> チャナ コトー ユー。	ミズテンで <u>買</u> っ <u>て</u> や <u>れ</u> <u>と</u> いう <u>よ</u> う な <u>こ</u> とを <u>言</u> う。	大阪_27_b_099
カセー <u>イッチャレ</u> ッテ。	手伝いに <u>行</u> っ <u>て</u> や <u>れ</u> と。	山口_35_e_099

いわゆるテ敬語に例外はない。後続する付属語は切らず、以降はその付属語の規則に従う。

方言テキスト	標準語テキスト	地点
ソナ コメ イッショーデモ ウロト <u>オモテヤッテン</u> ナー オ	そんな 米 [を] 一升でも 売ろう と <u>思</u> っ <u>て</u> いた <u>ん</u> だ <u>ね</u> 親が <u>ね</u> 。	滋賀_25_e_099

ヤガヨ。		
ドーガナカイ マワーテデシタダ ライラネー。	どのくらい 舞っておられたのでしょ うかね。	島根_32_a_099
コノゴラー シェーガ ワラー イレテジャケー	この頃は それが 藁を 入れなさる から	岡山_33_d_099
アンタガ ユーテノ ヨーニ	あなたが おっしゃる ように	山口_35_e_099

4.5.5 繰り返し

同一要素・類似要素の繰り返しは切り離す。ただし、代名詞や指示詞を含む慣用的な表現、オノマトペは切り離さない。

方言テキスト	標準語テキスト	地点
チジャ ナカ° レデ キンカラノ ターサ コー ミナ チジガラ ナンエガラ オジダナナ。	土が 流れて 金川原の 田へ こう 全部 土から なにから 落ちたの はね。	秋田_05_e_099
ナンデモ カンデモ ウチデ。	何でも かんでも 家で。	栃木_09_b_099
ヤン ドナンツツタッテ ヤドサナ ン トマッカマ。 アツツ コッ ツ ドゴサデモ	宿なんていったって 宿になんか 泊 まるか（Z：SP）。 あっち こっ ち どこにでも	福島_07_b_099
ウン フウン チューノカ° ネ ウン	運 不運 というのがね うん	茨城_08_a_099
オギャクサンガ ナンノ カン ノッテ	お客さんが 何の かんのと	宮城_04_c_010
ミンナニ ゲーオ オシエコマ ニャ ナランケ ナニカ カニカ ナー。	みんなに 芸を 教えこまないと な らないし なにか にかかねえ。	大分_44_c_099
カカ° イッカラ ナンデカンデ タガネ ツツード イランニエ ガッタンダ。	蚊が いるから なんで [も] かん で [も] <焚かない というと=焚 かずには> いられなかったんだ。	福島_07_b_099
ダレソリヤ コナイ コナイ イ ヨッタ ダレソリヤ	誰それは こう こう 言っていた 誰それは	兵庫 _28_e_001-2
アノ ドコソコノ エヤ ダマッ トッケンド アノヒトノ アノ フクソーガ イチバン ヨカッタ ナト ダ テ ソレオ	あの どこそこの 家は 黙っている けれど あの人の あの 服装が 一 番 良かったなと ダ テ それを	兵庫 _28_e_001-2

動詞連用形・終止形などの繰り返しも同様に切る。

方言テキスト	標準語テキスト	地点
ホエ ンメ ンメダ。 ホエズ。	それ [を] <埋め 埋め=埋めなが ら>だ。 そいつ。	宮城_04_c_099
ワカ° ムスビ スー スー オ シエ オシエ ムスビ。	自分の 結び [を] <しい しい= しながら> <教え 教え=教えなが ら> 結び [を]。	福島_07_b_099
アレ ナオサニャー ナオサ ニャー ユー ユー ドーシテモ アノ	あれ [を] 直さねば 直さねば [と] <言う 言う=言いながら> どうしても あの	広島_34_c_099
ナマノ マー ドンドンドンドン ハー カル カル コグカラ。	生の まま ドンドンドンドン もう <刈る 刈る=刈りながら> 脱穀 するから。	山口_35_e_099
ソリユー アンター ウマジー	それを あなた 馬で 何度も 行っ	大分_44_b_019

ナンベンモ イタリ キタリ イ タリ キタリ シチェタナー	たり 来たり 行ったり 来たり し てねえ	
----------------------------------	--------------------------	--

語彙化しているものは切らない。

方言テキスト	標準語テキスト	地点
ツグダニニ スンノガ ナン テー。 イワレデッシャ。 <u>イヤ</u> <u>イヤ</u> 。	佃煮に するのか なんて。 言われ てね。 <u>いやはや</u> 。	宮城_04_c_099

以下のような臨時的言語形式を切るべきかどうかは、現状作業者の裁量としている。

方言テキスト	標準語テキスト	地点
<u>ハハハハハ</u> 。 {笑}	<u>ははははあ</u> 。 {笑}	兵庫_28_d_099

4.5.6 複合辞・連辞

ナケレバナライ系

義務・当為を表す形式としてまとめて扱う。センナラン、セネバ、また後部の語彙にもナライ、デキナイ、イケナイ、ダチカン（埒が明かん）など様々な変種があるが、切れそうなものも含めて、全て切らずにおく。

方言テキスト	標準語テキスト	地点
ソゲン コトワ <u>セニヤナラン</u> コトジャー ナカ。	そんな ことは <u>しなければならない</u> ことでは ない。	長崎_42_b_099
マズ クスル トル <u>イカンナラ</u> <u>ンカラ</u>	まず 薬 [を] 取り [に] <u>行かな</u> <u>なければならないから</u>	島根_32_a_099
ユマカダニワ <u>コネワネー</u> ミズ <u>クマネバネアー</u> 。	夕方には <u>こなくてはならない</u> 水 [を] <u>汲まなければならない</u> 。	岩手_03_c_099
ソヤカラ ソレ ジブンデ <u>セン</u> <u>ナラン</u> 。	だから それ [を] 自分で <u>しない</u> <u>といけない</u> 。	京都_26_d_099
オマエ ノボッテカンナンゾッ テ。	おまえ <u>上っていかななければならない</u> ぞと。	新潟_15_d_099
ジキ <u>モドサニヤデケンケナー</u> 。	すぐ <u>返さなければいけないから</u> ね。	福岡_40_a_099
ホンデモ ワカイ ヒトニ オシ エトイテ <u>モラワニヤダチカン</u> 。	それでも 若い 人に 教えておいて <u>もらわねばいけない</u> 。	岐阜_21_c_099
イヌゴト ツナイデ <u>オガネバマ</u> <u>イネナー</u> 。	犬を <u>つないで おかなければいけな</u> <u>いね</u> 。	青森_02_e_116

ただし、間に助詞がある場合は切る。

方言テキスト	標準語テキスト	地点
ツナインデ <u>オガネバダキャ</u> <u>マ</u> <u>イネヤー</u> 。	つないで <u>おかなければ (Z:AP)</u> <u>だめだぞ</u> 。	青森_02_e_116

カモシレナイ系

同様に様々な変種があるが、全て切らずにおく。

方言テキスト	標準語テキスト	地点
マズ ダイイチニ アノー イエ イエデ <u>チガウカシランバツテ</u>	まず 第一に あの 家々で <u>違うか</u> <u>もしれないけれど</u>	長崎_42_b_099

カゼニ アタタント イッショ ニ コー タバニ ナットタン ヤシランノ。	風に あたったのと 一緒に こう 束に なっていたのかもしれないね。	徳島_36_b_099
ソーカモシンネーネ。	そうかもしれないね。	茨城_08_c_099
オシェダガモシェネ n ドモ オ ラァ ッシャネナ。	教えたかもしれないけれども 私は 知らないな。	岩手_03_c_099
ソリヤ マダ カワッ コッチャ イ ワカランサイ。 [笑]。	それは まだ 変わる ことかも わ からないよ。 [笑]。	佐賀_41_c_099
ソーカモワカランノ。 ンー。	そうかもしれないね。 うん。	広島_34_c_099

命令指示表現

命令や指示の機能をもつとみなされる複合辞を認定する。

方言テキスト	標準語テキスト	地点
アゲナ モノオ カケン モンデ ナケニヤ <u>カワンコトナンツツ</u> テネ。	ああいう ものを かけない もので ないと <u>買わないように</u> などと言っ てね。	島根_32_a_099
ハーヨー <u>ハイリヨシ</u> <u>ハイリヨ</u> <u>シ</u> チューテノ。	早く <u>入りなさい</u> <u>入りなさい</u> と 言ってね。	和歌山 _30_d_099

4.5.7 語彙化、慣用表現、固有名

ナク=ナルのような一語化しているものは文節に切らない。対応する標準語の語形が文節に切れたとしても、方言テキストの分綴が優越する。

方言テキスト	標準語テキスト	地点
モー ナンジューネン マエニ ナクナリマシタサカイ	もう 何十年 [も] 前に <u>亡くなり</u> <u>ました</u> から	京都_26_d_099
ナガイキュ シェンデタナー。 ハヨー ノーナツテタナー。	長生きを しなideねえ。 早く <u>亡</u> <u>くなって</u> ねえ。	大分_44_b_011

その他、以下のような例でスペースを入れずにつなげているが、究極的には作業者の裁量による。

方言テキスト	標準語テキスト	地点
アノ ホーガ ジブンデ <u>キニ</u> <u>イッタ</u> カラナー。	あの ほうが <自分=X1自身>で <u>気に入った</u> からね。	山口_35_e_099
ツイチョッテモ <u>マニアワンノ</u> 。	ついていても <u>間に合わないの</u> 。	山口_35_e_099
アトワ カンソーキー カケルン ジャ ユーテ <u>ヘーットモネー</u> ン ジャ	あとは 乾燥機に かけるのだ [と] 言っ て <u>なんともないのだ</u> [と]	岡山_33_d_099
チョットシタ リョコーエネー イッテ	<u>ちょっとした</u> 旅行へね	東京_13_b_133
キノコニシロ ソー ユー ナリ モンニシロ。	茸に <u>しろ</u> そう いう なりのものに <u>しろ</u> 。	栃木_09_b_099
ガイコーニヨッテナー	外交によってね	奈良_29_b_003
フント アン トキニヤ <u>ミルマ</u> <u>ニ</u> ヤケタワイー。	ほんとう あの ときは <u>見る間に</u> 焼けたね。	大分_44_c_005
タネモ ミンナ ダシタカラ ツ グトシ モー ツグハル。	種も みんな 出したから <u>次の年</u> もう <u>次の春</u> 。	群馬_10_c_099

数詞につく元号、「元～」などの接頭辞は独立の文節とは認めない。

方言テキスト	標準語テキスト	地点
タイショージュネンノ イヤ チガウ タイショークューネンノ	大正十年の いや 違う 大正九年の	長崎_42_b_099
オラチノ アノー モトヤケジョ カ° アッタ トコノ	俺[の]家の あの 元夜警所が あった ところの	栃木_09_b_099

語彙化していると判断される「AのB」も表面上の文節に切らない。「三分の一」などの分数表現も同様に切らずにおく。4.4節「複合語」も参照のこと。

方言テキスト	標準語テキスト	地点
コノドモノヒニ ヤルシナ。	子どもの日に やるしね。	北海道 _01_b_099
ウラッチャノマッテ ウラノ ホーニ ヘヤ アッタンダモナ。	裏茶の間って 裏の ほうに 部屋 [が] あったんだもん。	岩手_03_c_099
コ イユ ヨノナガニ ナッタン デー。	こう いう 世の中に なったんで。	宮城_04_c_099
ズゾー タデダエッチャ サエノ カワラサネ。	地藏[を] 建てたじゃない 賽の河 原にね。	宮城_04_c_099
ジブンジブンデ ヒノマルノ ハ ダオ カイダリ	各自で 日の丸の 旗を 書いたり	山形_06_c_099
ワレワレ オンナノコオ カサ カダサネンダッケノ。	我々 女の子を カサ 加えないもの だったね。	山形_06_c_099
ソー ドヨーノウシノヒニネ	そう 土用の丑の日にね	茨城_08_c_099
ヤモノサチワー オーカッタヨ ネー。	山の幸は 多かったよねえ。	栃木_09_b_099
マツノウチワ ダサネーダナ。 ウン。	松の内は 出さないんだね。 うん。	神奈川 _14_d_099
コーイ ハズジャ ナカッタカ° イツノマニガ	こういう はずじゃ なかったが い つのまにか	長野_20_e_099
ボク ソレ オボエテマスデー。 ヒノミ ヒノミ チューテ ア ノー ヒノミヤグラ。	僕 それ[を] 覚えていますよ。 火の見 火の見 といって あの 火 の見櫓。	京都_26_d_099
アンタトコノ サンブンノイチモ ネー ヤ ヤシキジャッタワ エ。	あなた[の] ところの 三分の一も ない ヤ 屋敷だったよ。	宮崎_45_c_099

「腹を_レたてる」「鼻が_レ高い」のような、標準語への逐語訳でそのまま慣用句として通用するものは、規則通りに切ってよい（ただし「水の泡」のような名詞慣用句は上に準ずる）。各地方言独特の言い回しについては、一続きにして意識タグ<A=B>で補足する。

方言テキスト	標準語テキスト	地点
チーット ノートコ トイカッ チェ ミナハリ オーキナ ホネ ジャー。	少し 苗代[が] 遠くて みなさい 大変な 苦労だ。	大分_44_b_019

「よろしくお願いします」など標準語と類似した形式の挨拶表現は切らずにおく。なお、標準語では挨拶として用いない形式については、逐語訳のうえ意識タグ<A=B>で補足する。

方言テキスト	標準語テキスト	地点
ジャ ドーモ アリガトーゴザイ	じゃ どうも ありがとうございます	滋賀_25_e_099

マシタ。	た。	
ソコワ ヨロシューオネガイシマ スッテ アタマ サゲタ ワケヤ	「そこは よろしく願ひします」と 頭[を] 下げた わけだ	兵庫_28_d_099
メーヤグ ステ スマッターネ スー。	＜迷惑 して しまいました＝申し訳 ありませんでした＞。	青森_02_e_119

固有名（動植物名、作品名など）は、字面上切れるようでも、ひとまとまりとして切らない。3.3節の「固有名タグ」も参照のこと。

方言テキスト	標準語テキスト	地点
オダグニ モーコノシ アルン ダッテネーテ ユッテ	お宅に 蒙古の碑[が] あるんだっ てねえ」と 言って	宮城_04_c_099
カイヨーワガマズスアガイジョー ド カガッテンダ	会陽若松市赤井町と 書かれているん だ	福島_07_b_099
モー オチャノミズジョシシハン カ° デキテッカラノ ウタナン ダ。	もう お茶の水女子師範が できてか らの 歌なんだ。	東京_13_a_099
セキカ° ハラノタタカイカ° オ ワッテ	関ヶ原の戦いが 終わって	東京_13_a_099
モチ ツカレタ ムカシワ ヨコ ロミノモチヤ シハリマシタカ。	餅[を] つかれた 昔は 「 <u>試みの 餅</u> 」とか されましたか。	京都_26_d_099
ア ソヤ ソヤ ドーキュー アン タモ ホン モロタヤロ。 ア ノ フルサトノオモイデバナシ。	あ そうだ そうだ 同級 あなたも 本[を] もらっただろう。 あの <u>ふるさとの思い出話</u> 。	兵庫_28_d_099
ユキノワカレダ イーテ。	「 <u>雪の別れ</u> 」だ[と] 言って。	島根_32_a_099

スローガン、歌詞などで、文や節の長さになるものは切る。

方言テキスト	標準語テキスト	地点
カカーデンカニ ヤネノ イシ ツチュンデ。	かかあ天下に 屋根の 石 というの で。	群馬_10_e_099
ナナシャク サーッテ シノ カ ゲオ フマズトカ ナントカ チュ モンデ	七尺 下がって 師の 陰を 踏まず とか 何とか という もので	群馬_10_e_099
カタグルマ ニドメノ チチガ スキニ ナリ。	「肩車 二度目の 父が 好きに な り。」	奈良_29_c_003

タグ一覧

タグ	内容	使用例
。	文の終止箇所、および発話のターンが交替している箇所を目安に付す	方言：ソーデス。 標準語：そうです。
{ }	談話上意味を持つ非言語的要素に付す	【笑】、【咳】など
*	分節音として認識困難な箇所に付す	方言：オチャツケノ * 標準語：お茶漬けの *
///	分節音としては明瞭に聞き取れるが、意味がとれない場合に標準語テキストに付す	方言：タノンナカッタワネー。 標準語：///なかったわね。
「 」	直接引用	方言：マンマ _n ダ _n ジェーッテ 標準語：「ご飯だよ」って
	注意書き	AIに対して
[]	注記。方言形の意味・用法などの注記番号	【1】
[]	方言には存在しないが、標準語では必要と判断される助詞を明示する	方言：ミカン ノセテ 標準語：みかん【を】 乗せて
<A=B>	方言の意識。Bに補足説明を記入する	方言：イマ ユー 標準語：今 <いう=話題にあがった>
(F)	フィラー	(F：ええ)、(F：ほれ)、(F：あなた)
(1SG) 等	人称と単複を表す	(1PL：私たち)
(Z)	方言テキストに対応する標準語テキストがない場合に使用する。 なお、対応する単語や形態素の単位数を示すのにはそれぞれの分類の記号を用いる。	副助詞 (AP) 方言：ハヤグバリ シテ ケサエン。 標準語：早く (Z：AP) して ください。
		終助詞 (SP) 方言：ドーステ エスタバ。 標準語：どうして いたのです (Z：SP)。
		指小辞 (DM) 方言：ツリコサ 標準語：釣り (Z：DM) に
		接頭辞 (PR) 方言：カッコシラエテ 標準語：(Z：PR) こしらえて
		敬語形式 (HON) 方言：オホートー 標準語：(Z：HON) ほうとう
		助動詞 (AUX) 方言：カッタケカ° 標準語：買った (Z：AUX) が
(D)	言い直し・言い淀みに付す	方言：ム ム ムツカシー 標準語：(D：ム) (D：ム) 難しい
(O)	固有名 (外国語、書名、映画名、ドラマ名、固有名詞、場所名、音や言葉に関するメタ的な引用、商品名など)	方言：チヨノフジカナー 標準語：(O：千代の富士) かな
(R)	話者の名前・差別語・誹謗中傷など	(R：X1) さん
(G)	オノマトペ	方言：スッパラスパラッテ 標準語：(G：スッパラスパラ) って