

発話単位ラベリングマニュアル
version 2.1

Japanese Discourse Research Initiative

2017年1月16日

Utterance-Unit Labeling Manual
Japanese Discourse Research Initiative

Copyright © 2010–2017 Japanese Discourse Research Initiative. All rights reserved.

Version 2.1	16 January 2017
Version 2.0	30 April 2014
Version 1.4	20 March 2014
Version 1.3	31 March 2011
Version 1.2	21 November 2010
Version 1.1	8 November 2010
Version 1.0	10 October 2010

目次

はじめに	1
1 作業要領	2
2 「長い発話単位」認定規則	3
2.1 概要	3
2.2 X境界	3
2.3 F境界	4
2.4 R境界	5
2.5 L境界	6
2.6 「長い発話単位」境界認定手続きのまとめ	15
3 「短い発話単位」認定規則	16
3.1 概要	16
3.2 x境界	16
3.3 f境界	17
3.4 s境界	17
3.5 X-JToBIとの比較	19
3.6 「短い発話単位」境界認定手続きのまとめ	19
付録A 応答系・感情表出系・呼びかけ系感動詞一覧	20
付録B 接続詞（相当表現）一覧	21
付録C 変更履歴	22

はじめに

対話研究においては、談話の流れや他者との相互作用の中で発話が果たす役割を明らかにすることがとりわけ重要であり、そのためには、語より上位の発話のレベルに関して単位や機能を認定する必要がある。しかし、書き言葉における句点のように明示的な文・発話の区切りを持たない話し言葉において、発話単位の認定は自明ではない。対話における発話単位の認定基準を策定することは、対話研究の発展にとって重要な要件である。

本マニュアルは、対話における発話単位の認定手続きについてまとめたものである。対話研究には、(1) 日常場面での発話産出・理解の仕組みを探求する認知志向の研究や、(2) 複数の参加者の間の情報交換・相互行為などの仕組みを探求するコミュニケーション志向の研究など、様々な目的のものがある。これらの目的に応じてふさわしい発話単位は異なるであろう。本マニュアルでは、おもに上記2つの研究目的を想定して、それぞれ「短い発話単位」「長い発話単位」と呼ばれる、2種類の異なる粒度の発話単位を用いる。もちろん、これらの発話単位は他の研究目的にも使えるものと考えている。

本マニュアルを用いて行なった研究を公表する場合は、以下のいずれかの文献を引用すること。

- Japanese Discourse Research Initiative. (2017). 発話単位ラベリングマニュアル *version 2.1*.
- Den, Y., Koiso, H., Maruyama, T., Maekawa, K., Takanashi, K., Enomoto, M., & Yoshida, N. (2010). Two-level annotation of utterance-units in Japanese dialogs: An empirically emerged scheme. In *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC 2010)*, 1483–1486. Valletta, Malta.

本マニュアルの発話単位の策定に際しては、日本学術振興会科学研究費補助金基盤研究 (B) 「対話における発話単位とその機能の認定に関する研究」(2008～2010 年度、代表：伝康晴)、日本学術振興会科学研究費補助金基盤研究 (B) 「発話単位アノテーションに基づく対話の認知・伝達融合モデルの構築」(2011～2013 年度、代表：伝康晴) から助成を受けた。本マニュアルはその成果の一部である。本マニュアルの執筆は、伝康晴 (千葉大学)・丸山岳彦 (国立国語研究所)・小磯花絵 (国立国語研究所) が担当した。

1 作業要領

本作業では、対話中の各話者の発話を「長い発話単位 (Long Utterance-Unit, LUU)」「短い発話単位 (Short Utterance-Unit, SUU)」と呼ばれる単位に分割する。これらの単位は、

長い発話単位 話し手と聞き手が行為や情報を交換する際の基本単位に相当し、統語的・談話的・相互行為的なまとまり

短い発話単位 話し手が一時に産出する情報のチャンクに概ね相当し、音響的・韻律的なまとまり

に対応する。これらの単位は、以下に述べる認定基準に則って認定される。以下、これらの単位の認定手続きについて説明する。

作業対象となるデータは、「転記単位」に分割されており、出現時間順に並べられているとする。これらは発話単位の切れ目を必ずしも反映したものではなく、発話単位の認定においては、2つ以上の転記単位を連結して1つにまとめたり、ある転記単位の途中で分割したりする必要がある*1。

■**転記単位の末尾で分割する場合** 分割位置に‘/’ (半角) の記号を記す。

[chiba0632:76.6650-79.0743]
A: カフェコンパーナとかあの辺がそうなるかな/

■**転記単位の途中で分割する場合** 分割位置に‘/’ (半角) の記号を記す。

[chiba0632:262.3220-264.8656]
B: そうですね/別々にみんな行きましたけどね/

*1 本マニュアルの事例は『千葉大学3人会話コーパス』『日本語話し言葉コーパス』から取られた事例である。事例冒頭の [chiba0632:76.6650-79.0743] などは出典を示す。事例は、休止ごとに「転記単位」に区切られており、以下の転記記号が用いられている。なお、対象となるデータは必ずしもこの基準で転記されている必要はない。

[重複発話の開始位置
:	非語彙的な音の引き伸ばし
%	非語彙的な音のつまり
-	語の中断
(F_あの)	フィラー
(I_うん)	応答系・感情表出系感動詞
(T_チョー ちょっと)	言いさし (意図された語が同定可能)
(D_スハ)	言いさし (意図された語が不明)
(W_ジュオー 授業)	言い誤りや非標準的な発音
<声>	聞き取れないか、言語音と見なせない音声

■転記単位を連結する場合 先行する転記単位の末尾に‘+’（半角）の記号を記す。

```
[chiba0632:81.2428-84.2909]
C: ウィンナーコーヒーって+
C: クリームがくるくるって乗ってるやつですか/
```

すべての行末には、‘+’か‘/’のいずれかの記号を記す。1つの行に‘/’を2つ以上用いてもかまわないし、‘/’を用いた行の末尾に‘+’を用いてもかまわない。

「長い発話単位」「短い発話単位」のいずれの認定作業においても、対象テキストの音声聞くことが重要である。ELAN^{*2}や Excel play マクロ^{*3}が作業環境として利用できる。

2 「長い発話単位」認定規則

2.1 概要

「長い発話単位」の境界は、以下の認定規則に則って認定する。

■「長い発話単位」認定規則 単語境界ごとに以下の規則をこの順に適用する。

1. 直前の境界以降が非言語音（笑い・息・咳など）なら、‘/X’でマークする。
2. 直前の境界以降が言いさしや言い止めで休止が後続するなら、‘/F’でマークする。
3. 直前の境界以降が応答系・感情表出系・呼びかけ系感動詞なら、‘/R’でマークする。
4. 当該境界が統語的・談話的・相互行為的な境界なら、‘/L’でマークする。
5. 上記のいずれにも該当しなければ、次の単語境界へ進む。

この規則により、単位境界を認定すると同時に境界の種類を記す。これらは、転記テキスト中に‘/X’、‘/F’、‘/R’、‘/L’のラベルを付与することで行なう。以下、境界の種類ごとに詳細を述べる。

2.2 X 境界

直前の境界以降が非言語音（笑い・息・咳など）の場合、その直後に境界を認定し、‘/X’のラベルを付与する。

```
[chiba0232:15.5831-16.4675]
B: またかよ/L
B: <笑>/X ←笑い
```

^{*2} <https://tla.mpi.nl/tools/tla-tools/elan/>

^{*3} <http://www.jdri.org/open-data/> より入手可能

ただし、非言語音の前後で発話内容が連続している場合は、単位境界としない。

[chiba0932:210.2705-214.5371]

B: でも牛乳が+

B: <笑>+ ←前後で発話内容が連続しているので境界としない

B: 牛乳がどれぐらいの頻度で+

B: 補充されてんのか不明/L

2.3 F 境界

直前の境界以降が発話の言いさしや言い止め（フィラーは除く）で休止が後続する場合、その直後に境界を認定し、‘/F’ のラベルを付与する。

[chiba0632:349.0430-351.3359]

C: [でも三月+

B: [(D_カ)/F ←言いさし

C: 箱根/L

B: 箱根/L

[chiba0232:150.1013-153.6165]

A: なん[かあんまり/F ←言い止め

B: [(I_え)/Rきょうの発表はどうだったの/L

B: 何分ぐらい発表[したの/L

A: [発表って何が/L

ただし、以下に該当する場合は単位境界とはしない。

■言いさしや言い止めの後、休止を置かず発話が続いている場合

[chiba0232:191.2408-192.5371]

C: (D_ナ)家電はやらない感じ/L

↑休止がないので境界としない

■言いかけた部分（名詞句など）が直後で繰り返されたり、言い換えられたりしている場合

[chiba0232:59.9754-62.0868]

A: (T_カン-|完熟)+ ←直後で繰り返されているので境界としない

A: (T_カン-|完熟)+ ←直後で繰り返されているので境界としない

A: 完熟か:/L

[chiba0732:259.6934-260.5750]

C: (T_コード-|子供)+ ←直後で「小さい」に言い換えているので境界としない

C: 小さい時/L

■言い淀みの前後で発話内容が連続している場合

[chiba0932:206.6316-208.7707]

A: なんか(D_プ)+ ←前後で発話内容が連続しているので境界としない

A: カプチーノっぽくなるよね/Lなんかね/L

なお、発話がほとんど終端部に達しており、末尾のみが打ち切られている、あるいは、小声や笑いのため転記されていない場合は、F境界とはせず、後述のL境界とする。

[chiba1232:156.6086-159.1905]

A: (I_うん)/Rあんま+

A: 人格形成はないの[かも [しれ-/L ←発話の終端部での打ち切り

B: [(I_あ[-:)/R

C: [<笑>/X

2.4 R境界

「うん」「はい」などの応答系感動詞や「あっ」「えっ」「へー」「ふーん」などの感情表出系感動詞や「ねえ」「おい」などの呼びかけ系感動詞（付録A）はその直後を単位境界とし、‘/R’のラベルを付与する。

[chiba0932:112.4189-113.7930]

A: さすがだよ/L

C: (I_うん)/R ←応答系感動詞

[chiba0832:378.5018-379.9459]

B: そう/L落と[したの:/L

C: [(I_え)/Rどうなんの/L落とすと/L
↑感情表出系感動詞

[chiba0732:551.5133-553.555]

B: 噂の新宿で/L(I_はい)/R

C: [(I_ええ)/R

A: [ね/R新宿の何/L

↑呼びかけ系感動詞

同一の感動詞が繰り返されているものは、1つ1つ切り離さず、全体を1つの単位とする。

[chiba0932:78.6610-80.1150]

B: 暇だったから[一人で:/L

A: [(I_うん)(I_うん)(I_うん)/R ←1つ1つ切り離さない

2.5 L 境界

統語的・談話的・相互行為的な切れ目はその直後を単位境界とし、‘/L’のラベルを付与する。統語的境界、談話的境界、相互行為的境界か否かをこの順に判断し、該当すれば単位境界とする。

2.5.1 統語的境界

■明示的な文末表現 発話が「～です」「～ます」「～だよ」「～なんですか」のような文末表現で終了することがある。文末表現は、一まとまりの発話が終了したことを形式的に明示する要素であり、述語句（動詞句・形容詞句・名詞＋助動詞など）の形式によって、話し手の意志や推量、聞き手への質問や命令などの態度が示される。これらの直後は統語的境界として単位分割する。

文末表現にはさまざまな形式がある。代表的な文末表現を表1に挙げる。

表1 明示的な文末表現

動詞（＋助動詞）（＋終助詞）	「泊まるんですか」「どこにあったんだろ」「決めてるもん」 「寝なきゃいけないじゃないですか」
形容詞/形容動詞（＋助動詞）（＋終助詞）	「ちょっと怖い」「すごく親切だった」「怪しいんだよね」
名詞（＋助詞類）＋だ（＋助動詞）（＋終助詞）	「技術者でしたね」「する訳ですよ」「男の方ですか」 「あれですよ」「おばあちゃんばかりでしょ」
副詞＋だ（＋助動詞）（＋終助詞）	「そうですね」「そうなんだ」「どうでしょうね」 「どうなんだろ」
引用の形式（＋終助詞）	「何様のつもりなんだと」「機会がなくて行かなかったと」

- 「だ」は「です」「である」「でした」「でしょう」などを含む。
- 終助詞は、以下のもの、およびその連鎖を含む。
「ね」「よ」「か」「な」「の」「さ」「じゃん」「っけ」「もん」「かしら」「わ」「や」「ん」「ぜ」「ぞ」「け」

[chiba0632:73.4943-76.0360]

A: よっぽど日本人はウイナーコーヒーって言うんでしょね/L ←動詞+助動詞+終助詞

[chiba0932:172.0191-174.0174]

A: そうそう/Lあのコーヒーメーカーすごいよね/L ←形容詞+終助詞

[chiba0632:140.6621-142.2585]

C: デメルってウイーンでしたっけ/L ←名詞+だ+助動詞+終助詞

[chiba0232:150.2820-151.5822]

B: え/Rきょうの発表はどうだったの/L ←副詞+だ+助動詞+終助詞

[chiba0632:62.8916-66.1102]

A: でもウイナーコーヒーってほんとはウイナーコーヒー
じゃないんです(D_ス)ってね/L ←引用の形式+終助詞

ただし、明示的な文末表現に見えても、引用節や連体節の末尾であれば、単位境界とはしない。

[chiba0632:69.4449-72.8392]

B: で:どうしようって言ってたらウエイトレスさんが来て+
↑引用節の末尾

B: ウイナコーヒー/L

[chiba0632:53.1466-54.8265]

B: 全部歩いて行ける距離にあったから
↑連体節の末尾

■独立度の高い従属節の終端境界 明示的な文末表現以外にも、統語的な境界と見なすべき箇所がある。その一つは、接続助詞が・ケレドモ・ケレド・ケドモ・ケド・シで終わる従属節の終端境界である。これらの従属節は主節からの独立度が高く（主節への従属度が低く）、統語的に大きな切れ目を構成することから、その直後を統語的境界として単位分割する。代表的な例を表2に挙げる。

表2 独立度の高い従属節の終端境界

が（+終助詞）	「思うんですが」「ファゴットがあるんですが」
け（れ）ど（も）（+終助詞）	「話すんだけど」「違うんですけど」「と思うんだけどね」
し（+終助詞）	「あんまりうれしくないし」「一切出なかったし」

[D01F0030:64.1920-65.7360]

R: もしかしたら顧問だったかもしれないんですが/L ←接続助詞ガ

[chiba0632:231.6788-233.5472]

B: 日本人は大丈夫って言うけどね/L ←接続助詞ケド+終助詞

[D01F0030:351.8160-357.3690]

R: (F_と)クラシック:の曲をアレンジして+

R: ブラスバンド用に作ったものもありますし:/L ←接続助詞シ

■独立度の低い従属節の終端境界 明示的な文末表現以外の統語的境界の二つ目は、ガ・ケレドモ・ケレド・ケドモ・ケド・シ以外の形で表される従属節の直後が、一まとまりの発話の切れ目になっている場合である。これらの従属節は形態・統語的には大きな切れ目は構成しないが、(1) 同じ話者による応答系・感情表出系・呼びかけ系感動詞(付録A)や接続詞(相当表現)(付録B)が後続する場合、(2)直後で話者交替(あいづちを除く)が起こった場合は、大きな切れ目があると見なし、統語的境界として単位分割する。このような従属節の代表的な例を表3に挙げる。

表3 独立度の低い従属節の終端境界

から	「海じゃねえから」	ので/んで	「夜怖くなかったんで」
たら	「ウイーンとかだったら:」	助動詞デ	「怒鳴り合い:みたいな感じで」
て	「名前があるらしくて」	っていう	「きれられてもな:っていう」
とか	「泊まってったりとかね」	みたいな	「いつのだろうみたいな」
ながら	「言われながらね:」	ような	「言えないような」

[chiba1232:121.6627-125.9139]

B: 仕事のことになると:や[たら:+

A: [(I_うん)/R

B: うるさくて:/L ←同じ話者による感動詞が後続

A: (I_あ-:)/R

B: (I_うん)/R ←同じ話者による感動詞

[chiba1232:75.5676-80.2849]

B: それで相当俺はてんぱって[:/L ←同じ話者による接続詞が後続

A: [(I_うん)/R

B: けどあんまフォローにもきてくれないから:+

↑同じ話者による接続詞

[chiba0932:218.6195-223.1283]

A: なんかね怪しいんだよね:/L

A: [いつのだろうみたい[な/L ←直後に話者交替

C: [怪しい/L

B: [<笑>/X

C: (I_あー)(I_あー)(I_あー)(I_あー)(I_あー)/Rそれ怖いね/L

これらの従属節が出現しても、上述の条件を満たさない場合は、単位境界とはしない。

[chiba0632:165.1925-183.4243]

C: まだ+

C: (D_夕)+

C: (F_あの)OLさんみたく:+

C: 高収入じゃないから[:+ ←条件を満たさないので境界としない

B: [(D_ン)/F

C: 別にお買い物目当てじゃなくて+ ←条件を満たさないので境界としない

B: (I_あーん)/R

C: 結構+

C: 見るほう中心に+

C: いつもしてて:/Lで最終日は-(W_ンチ|日)だけ(D_フ)+

↑条件を満たすので境界とする

C: なんか+

C: フリーっていう旅を選ぶんですけど:/L

■非明示的な文末表現・語彙的なあいづち表現 話し言葉では、明示的な文末表現以外の箇所が発話が完了したり、語彙的なあいづち表現のみで発話が構成されたりすることがある。また、ほぼ終端部に達した文末表現が途中で打ち切られる場合もある(2.3項参照)。これらの直後も統語的な境界とする。代表的な例を表4に挙げる。

表4 非明示的な文末表現・語彙的なあいづち表現

名詞(+格助詞等)(+終助詞)	「箱根」「やり過ぎ」「どの辺」「病院か」「夜は」「しらふが」「外ではね」
副詞(+格助詞等)(+終助詞)	「ちょっとね」「なんかね」
体言止め	「ほんとは違う名前」「新しいバイト先」「そいつはほんとだめ」
語彙的なあいづち表現	「そうそう」「そうか」「なるほど」「まあね」「ね:」

[chiba0632:291.8560-293.7470]

A: (I_うん)/Rマケドニアです/L

B: マケドニア/L ←名詞

[chiba1132:80.9579-85.8670]

C: なんか+

C: (F_あの:)ズボンの補正ってあるじゃん/L

B: (I_あ[-](I_あー)(I_(W_ア|あー))/R

C: [(F_あの:)/F

A: 裾直[し [ね/L ←名詞+終助詞

C: [裾直[し/L ←名詞

B: [うん/R

[chiba0932:374.7500-376.7130]

A: やばいよ/Lそれは/L

C: ちょっとね:/L ←副詞+終助詞

[chiba0632:306.0463-311.6320]

B: マケドニアって+

B: どの辺/L ←体言止め

A: ウイーンよりもっと下/Lもう旧ユーゴ:/L
↑体言止め ↑体言止め

[chiba1232:320.9560-322.2711]

A: そうそう/Lがんばってるらしいよ/L

↑語彙的なあいづち表現

[chiba1232:156.6086-159.1905]

A: (I_うん)/Rあんま+

A: 人格形成はないの[かも [しれ-/L ←発話の終端部での打ち切り

B: [(I_あ[-:)/R

C: [<笑>/X

■倒置要素の直前・直後 話し言葉では、本来あるべき語順が逆転した「倒置」の構造が生じることがある。たとえば、「あれはなぜかだめだね」が、「なぜかだめだねあれは」という語順になるような場合である。このような倒置要素の直前・直後の箇所は統語的境界として単位分割する。

[chiba0932:350.7177-353.3076]

倒置要素直前↓ ↓倒置要素直後

A: わたしはだめ/Lなんか/L

C: (I_へえ[:)/R

A: [なぜかだめだね/Lあれは/L

倒置要素直前↑ ↑倒置要素直後

■引用節・連体節の中にある境界の扱い 上記までに示した統語的境界が、大局的に見ると、引用節・連体節の内部に生じていることがある。このような場合は単位境界とはしない。

[chiba1132:213.5703-214.6725]
B: でしょうかどうでしょうかなみたいな/L
↑引用節中の文末表現

[D01F0030:142.7040-150.2270]
R: わいわいやれるけど目標のない+
↑連体節中のケド節
R: これといった目標のない部活(W_ツテュー|っていう)風な方針がない(W_テュー|という)
風に思ったんだと思います/L

■挿入構造等の扱い 発話が続いている途中で、別の発話が挿入されることがある。挿入された発話が統語的境界で終わるものであっても、その前後で発話内容が続いている場合には、単位境界とはしない。

[chiba0132:153.2780-156.5020]
C: あんまなんて言うのかな一般的な話に持ってけなくて:個別:な+
↑挿入構造末尾の文末表現

挿入構造に似た現象として、一度発話した内容を言い直す場合に、文末表現が現れる場合がある。このような文末表現の直後も単位境界とはしない。

[chiba0532:559.3180-564.6230]
C: 服着て+
C: 臭いかな+ ←言い直しに関わる文末表現
C: 臭いじゃないわ+ ←言い直しに関わる文末表現
C: なんか+
C: 汚いかな:とか思ったりするけど/L

さらに、発話の途中で自己の発話内容を確認し、直後に元の発話に復帰する場合がある。このような文末表現の直後も単位境界とはしない。

[chiba1232:160.6270-164.6590]
A: でもね:+
A: バイト先でもう+
A: 十年目ぐらいかななる人がいんの/Lバイトで/L
↑自己確認に伴う文末表現

■繰り返される単一文節の扱い 単一の文節が休止を置かずに繰り返される場合がある。それぞれの文節末が上述の条件にあてはまる場合でも、1つ1つ切り離さず、全体で一つの単位とする。

[chiba0532:61.8473-63.1422]

A: 席がボツ[クスんなってるやつがあるの/L

C: [あるあるあるある/L ←1つ1つ切り離さない

2.5.2 談話的境界

■文脈上の切れ目 統語的境界でない位置であっても、ひと続きの話題がそこで収束したり、そこから別の話題に転じたりするなど、文脈上の大きな切れ目があると判断できる場合がある。そのような場合、当該の箇所を談話的境界として単位分割する。

[chiba0132:227.2980-239.0655]

A: (F_その)どうでもいいことを:す[ごい+

C: [(I_うん)/R

A: 話されると:聞いているほうがすごいいらするっていうのあ[って:/L[わたしがドイツ語の+

↑文脈上の切れ目

B:

[(I_う [ん)/R

C:

[(I_うん)/R

A: 予備校に行ってたときに:(F_その)/L ←文脈上の切れ目

A: でいい先生なんだけどさ/L

■話題導入表現 新たな話題の冒頭で、話題を導入する表現が出現する場合がある。典型的には「～のは」などの形で話題が提示され、その後ひと続きの説明が続くような場合である。このような要素の直後は談話的境界として単位分割する。次の例では、「松屋でバイトしてて思うのは」という形で話題を導入した後、その内容についての説明が長く続いている。

[chiba1132:162.4930-173.3620]

B: そうそう/Lなんか松屋:でバイトしてて思うのは[:/L ←話題導入表現

A: [(I_うん)/R

B: きれやすい+

B: なんか酔ってる客って俺きれやすいと思ってたんだよ/L

[庄やの [時点では/L

A: [(I_あ-[:)(I_はあ)(I_はあ)(I_はあ)/R

C: [(I_あ-)(I_あ-)(I_あ-)(I_あ-)/R

B: 逆だな/L

A: しら[ふが/F

B: [酔ってる客の方[が+

A: [<笑>/X

A: (I_う[ん)(I_うん)/R

B: [きれない/L

■直後がまとめ表現 逆に、ひと続きの説明が続いた後、それまでの説明をまとめるための表現が出現する場合がある。ここで、まとめ表現の直前が統語的境界でない場合、談話的境界として単位分割する。次の例では、話者Cが店員として振る舞いがよくなかったことを説明する発話が続き、最後に「反省した一幕が」と説明をまとめている。

[chiba1132:507.1360-521.1099]
 C: ちょっとこっちも気まずいな:と[思っ/L
 B: [笑>/X
 ((中略))
 C: それも気まずいな:と/L
 B: [(I_あー)/R
 A: <笑>/X
 A: まあな[フィリピン人いじめちゃいけねえよ/L[や[っば/L
 C: [(D_と)/F
 C: [(I_うん)/R
 C: [笑>/X
 B: [(I_うん)/R
 A: <笑>/X
 A: [笑>/X
 C: [(I_あー)ちょっとやりすぎたな:と思っ/L ←直後がまとめ表現
 A: (I_へえ)/R[そんな貴重な経験が/L
 C: [反省した一幕が/L ←まとめ表現

2.5.3 相互行為的境界

■述語末への付け足し 述語の後で聞き手の反応を窺い、さらに助動詞・終助詞などを付け足すことがある。統語的には全体でつながりの発話を構成するが、この場合には述語末でいったん相互行為的境界として単位分割する。

[chiba0833:278.7000-281.0260]
 B: てかスローラブ/L ←聞き手の反応を窺っている
 B: [だよね/L ←述語末への付け足し
 C: [(I_う:ん)(I_うん)/R ←聞き手が反応

最初の述語末で必ずしも聞き手が反応しないこともある。助動詞・終助詞などの付け足しが聞き手の反応を促すためになされることを考えると、むしろ聞き手の反応がない場合のほうが多いかもしれない。聞き手の反応を窺っているか否かは、休止の長さや音調などから判断する。

[chiba1231:318.8530-320.9410]
 A: 長いんだよ/L
 (0.9) ←聞き手の反応を窺う長い休止
 A: ね/L ←述語末への付け足し

■発話途中の働き掛け 発話途中で述語以外の箇所を疑問調などにして聞き手の反応を窺うことがある。統語的には前後がつながって一つながりの発話を構成するが、この場合には当該箇所ですいたん相互行為的境界として単位分割する。

[chiba0931:104.8530-111.5490]

C: (I_うん)(I_うん)(I_うん)(I_うん)(I_うん)(I_うん)/Rあ/Rでもねちょっと関連したこと/L昨日ね:カイロ+

C: プラク[ティック/L ←聞き手の反応を窺っている

B: [はいはいはい[はい/R

A: [うん/R

C: に行っちゃった/L ←統語的につながる後続要素

■発話途中の応答 発話途中で自己発話を中断し、相手発話に応答することがある。そのような応答の前後は相互行為的境界として単位分割する。応答の前はしばしば言い止め（F境界）になる。

[chiba0632:76.6650-80.6229]

A: カフェコンパーナとかあの[辺がそうなるかな/L

B: [(I_うん)/Rなんか:/Fそう/L ←Aの「そうなるかな」に反応

B: 名前が違う/L

ただし、発話途中の応答表現（「うん」「そう」など）が相手発話に対する応答としてではなく、自己の発言に対する確認などの意味で使われている場合は、単位境界とはしない。

[D01F0023:383.6836-389.8434]

R: (F_あの:)<声>+

R: (F_う)<声>後は(I_うん)こっちから見ると喧嘩してるようにも見えるんだ[けど/Lとにかく+
↑相手発話に対する応答ではないので境界としない

L: [(I_うーん)/R

■応答の完結可能点 質問などに対する応答がいったん完結したように見えた後、統語的につながる要素が後続する場合がある。統語的には全体で一つながりの発話を構成するが、この場合は最初の完結可能点ですいたん相互行為的境界として単位分割する。

[chiba1232:288.8730-292.9540]

A: (W_モッ|もう一)回[あんじゃんね/L

B: [一万ちょっ [とね/L

C: [(I_あ)/R一応親に/L ←応答として完結可能

A: (I_あ[-)/R

B: [(I_あ-:)/R

C: [毎回もらってますけど/L ←統語的につながる後続要素

2.6 「長い発話単位」境界認定手続きのまとめ

以上に述べた「長い発話単位」境界認定手続きを図1・表5にまとめる。

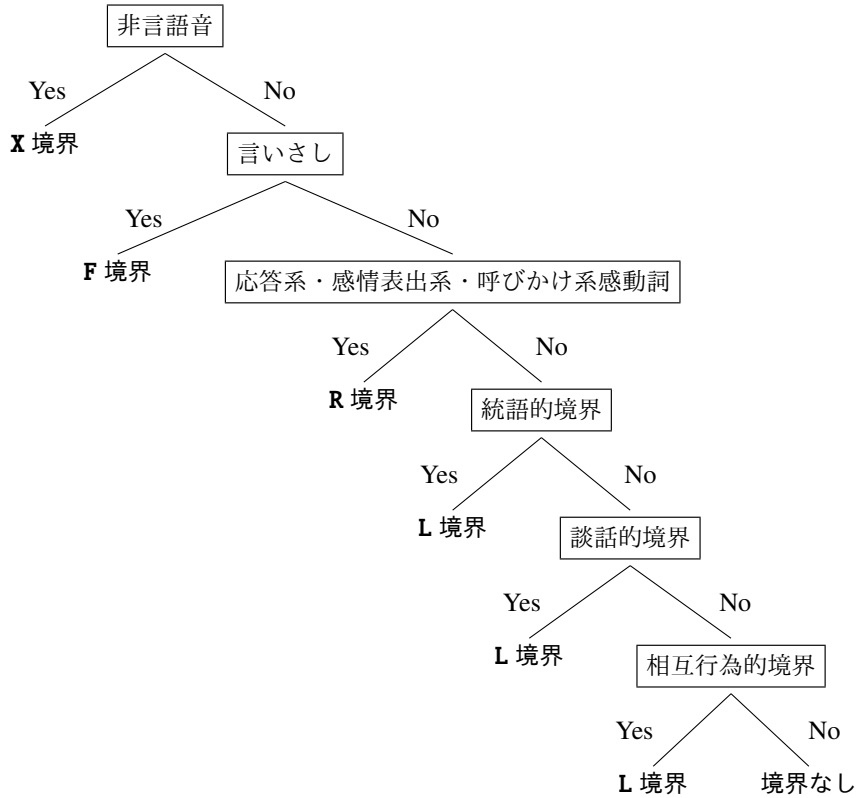


図1 「長い発話単位」境界認定手続き

表5 L境界一覧

統語的境界	明示的な文末表現（表1） 独立度の高い従属節の終端境界（表2） 独立度の低い従属節の終端境界（表3）+感動詞・接続詞・話者交替 非明示的な文末表現・語彙的なあいづち表現（表4） 倒置要素の直前・直後
談話的境界	文脈上の切れ目 話題導入表現 直後がまとめ表現
相互行為的境界	述語末への付け足し 発話途中の働き掛け 発話途中の応答 応答の完結可能点

3 「短い発話単位」認定規則

3.1 概要

「短い発話単位」は「長い発話単位」とは独立に認定する。両者の間には必ずしも階層関係が成立しない。すなわち、「長い発話単位」の境界であっても「短い発話単位」の境界とならないことがある。

「短い発話単位」の境界は、以下の認定規則に則って認定する。

■ 「短い発話単位」認定規則 単語境界ごとに以下の規則をこの順に適用する。

1. 直前の境界以降が非言語音（笑い・息・咳など）なら、‘/x’でマークする。
2. 直前の境界以降が言いさし・言いよどみや応答系・感情表出系感動詞なら、‘/f’でマークする。
3. 当該境界が音響的・韻律的な境界なら、‘/s’でマークする。
4. 上記のいずれにも該当しなければ、次の単語境界へ進む。

この規則により、単位境界を認定すると同時に境界の種類を記す。これらは、転記テキスト中に‘/x’、‘/f’、‘/s’のラベルを付与することで行なう。以下、境界の種類ごとに詳細を述べる*4。

3.2 x 境界

直前の境界以降が非言語音（笑い・息・咳など）の場合、その直後に境界を認定し、‘/x’のラベルを付与する。

```
[chiba0232:15.5831-16.4675]
15.5831 16.0127 B: またかよ/s
16.0127 16.4675 B: <笑>/x ←笑い
```

「長い発話単位」の X 境界（2.2 項）とは異なり、非言語音の前後で発話内容が連続していても、その前後を単位境界とする。

```
[chiba0932:210.2705-214.5371]
210.2705 211.0995 B: でも牛乳が/s ←前後で発話内容が連続していても境界とする
211.3350 211.4150 B: <笑>/x ←前後で発話内容が連続していても境界とする
211.7262 213.0350 B: 牛乳がどれぐらいの頻度で+
213.1059 214.5371 B: 補充されてんのか/s不明/s
```

*4 「短い発話単位」の認定では休止が重要な役割を果たすため、本節の事例では発話の開始・終了時間を明記する。

3.3 f 境界

直前の境界以降が言いさし・言いよどみ（フィラー）や応答系・感情表出系感動詞で休止が後続する場合、その直後に境界を認定し、‘/f’ のラベルを付与する。

[chiba0232:227.1639-228.8675]
227.1639 227.4268 B: (T_テレ-|テレビ)/f ←言いさし
227.6070 228.8675 B: NECテレビとかあるよね:/s

[chiba0432:388.6503-390.9760]
388.6503 388.9790 C: (F_あの)/f ←言いよどみ
389.2660 390.9760 C: まあ/s水飲みつつ会議をしてました/s

[chiba0932:112.4189-113.7930]
112.4189 113.1784 A: さすがだよ/s
113.4066 113.7930 C: (I_うん)/f ←応答系感動詞

[chiba0632:80.9277-84.2909]
80.9277 80.9930 C: (I_え)/f ←感情表出系感動詞
81.2428 82.0990 C: ウィンナーコーヒーって/s
82.2068 84.2909 C: クリームがくるくるって乗ってるやつですか/s

「長い発話単位」の F 境界（2.3 項）と同様、休止を置かず発話が続いている場合は、単位境界とはしない。

[chiba0232:191.2408-192.5371]
191.2408 192.5371 C: (D_ナ)家電はやんない感じ/s
↑ 休止がないので境界としない

3.4 s 境界

音響的・韻律的な切れ目はその直後を単位境界とし、‘/s’ のラベルを付与する。音響的境界・韻律的境界か否かをこの順に判断し、該当すれば単位境界とする。

3.4.1 音響的境界

0.1 秒程度以上の休止が後続する箇所は音響的境界とする。

[chiba1132:11.0486-13.4710]

11.0486 11.9610 B: どうですかじゃ:先/s ←休止が後続

12.4771 13.4710 B: ユニクロ班として/s

3.4.2 韻律的境界

韻律的に強い切れ目を感じる箇所は韻律的境界とする。

[chiba0632:96.6962-99.1083]

96.6962 99.1083 B: しかも/sカフェにいるのって/sおじいちゃんおばあちゃん/sばかりでしょ/L

↑ 韻律的に強い切れ目 ↑ 韻律的に強い切れ目 ↑ 韻律的に強い切れ目

韻律的な切れ目は具体的には以下のように判断する。

1. 談話音声には、声の高さが上がって下がるというパターンが繰り返し見られる。たとえば、図2・3の例では、それぞれ「多分／関西の／人だと」「しかも／カフェに／いるのって／おじいちゃん／おばあちゃん／ばかりでしょ」の単位で、声の高さが上がって下がるというパターンが見られる。
2. 次に、それぞれの単位ごとに、一番声の高い位置と一番低い位置を同定する。たとえば、図2の「多分」では、単位冒頭が一番高く、末尾が一番低い音となる。この際、末尾の局所的な上昇や上昇下降部分は無視する。たとえば、図2の「ばかりでしょ」の場合、最後の上昇部分を除いた上で、一番高い音と低い音を同定する。
3. 最後に、それぞれの単位ごとに、一番高い音と一番低い音の幅（以下、ピッチレンジ）に着

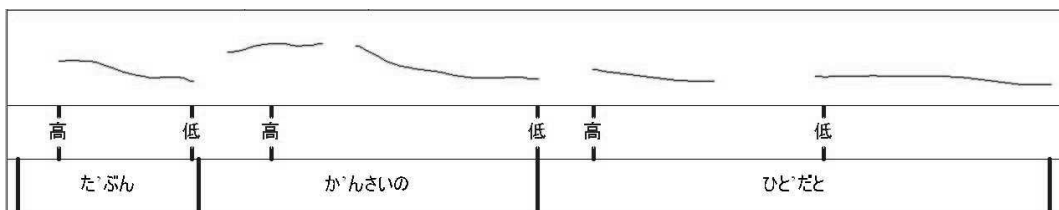


図2 発話例「多分関西の人だと」

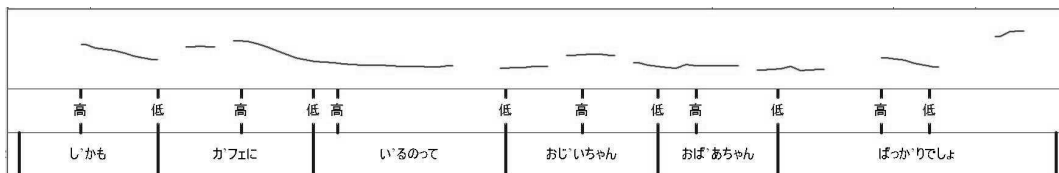


図3 発話例「しかもカフェにいるのっておじいちゃんおばあちゃんばかりでしょ」

目し、隣接する2つの単位のピッチレンジを比較して後続単位のピッチレンジのほうが大きい（あるいは同等程度の）場合、そこに韻律的境界があるとみなす。

図2の発話例では、「多分」のピッチレンジより「関西の」のピッチレンジのほうが大きいため、ここに韻律的境界があるとみなす。一方、「関西の」と「人だと」では後者のピッチレンジのほうが小さいため、ここには韻律的境界がないと判断する。

図3の発話例では、「しかも」より「カフェに」のほうが、「いるのって」より「おじいちゃん」のほうが、「おばあちゃん」より「ぼっかりでしょ」のほうがピッチレンジが大きいので、これらに韻律的境界があるとみなす。

3.5 X-JToBI との比較

韻律ラベリングスキーム X-JToBI には、韻律境界の切れ目の強さに関する情報として **Break Index (BI)** が存在する。短い発話単位と BI との関係は次の通りである。

- f 境界の認定には、次の BI 情報を用いる。
 - BI = F (韻律的フィラー区間の終端)
 - BI = D (韻律的語断片の終端)
- s 境界の認定には、次の BI 情報を用いる。
 - BI = 3 (イントネーション句境界に相当)

3.6 「短い発話単位」境界認定手続きのまとめ

以上に述べた「短い発話単位」境界認定手続きを図4にまとめる。

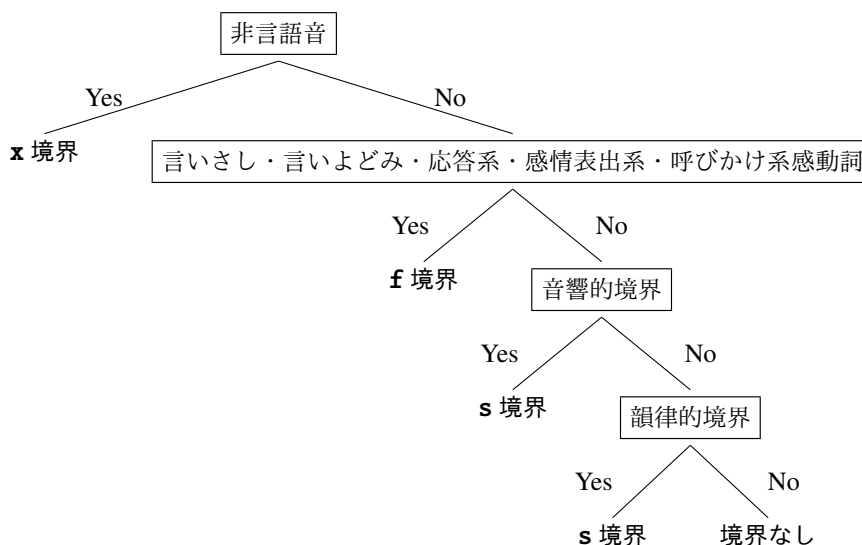


図4 「短い発話単位」境界認定手続き

付録 A 応答系・感情表出系・呼びかけ系感動詞一覧

ああ/あー	うん/うーん	ふむ	まあ/まー ^c
あーん	ええ/えー	ふむふむ	むむ
あい	えっ/え	ふん	もう/もー ^d
あっ/あ	おい/おーい	ふーん	もしもし
あら	おお/おー	ふんふん	やあ
いや	おっ/お	へい/へー ^a	わあ/わー
うーん	ね/ねえ	へえ/へー ^b	わあい/わーい
うお	はあ/はー	ほい	わあん/わーん
うむ	はい	ほう/ほー	
うわ	ふう/ふー	ほほう/ほほー	

^a「へい、さようで」の「へい」(商人言葉)

^b「へえ、あいつがねえ」の「へえ」

^c「まあ、ひどい」の「まあ」

^d「もう、ほんとにすごいんだ」の「もう」

付録 B 接続詞（相当表現）一覧

あと	{ という/っていう/って } ことで
あるいは	{ という/っていう/って } ことは
一方	{ という/っていう/って } ことな { の/ん } で
が	{ という/っていう/って } の { は/も }
かと言って	{ という/っていう/って } 感じで
け（れ）ど（も）	{ という/っていう/って } ように
さて	{ という/っていう/って } 訳で
さらに	とくに
しかし（ながら）	ところが
しかも	ところで
したがって	と言うの { は/も }
（ん）じゃ	と言っても
すなわち	とすると
すると	とは言うものの
そこで	と同時に
そして	なお
それか	なおかつ
{ それ/そい } から	{ な/です } { の/ん } で
{ それ/そい/そん } じゃ	{ な/です } のに
{ それ/そい/そん } で（は/も）	なもんですから
それと	なんですが
それとも	なんですけ（れ）ど（も）
{ だ/です } が	ほいから
{ だ/です } から（こそ）	{ ほい/ほん } じゃ
{ だ/です } け（れ）ど（も）	{ ほい/ほん } で（は/も）
ただ	まず
ただし	また
{ だっ/でし } たら	または
だって	もしくは
ちなみに	もっとも
つまり	ゆえに
（ん）で（は/も）	ようするに
であれば	よって

付録 C 変更履歴

Ver. 2.1

- 「長い発話単位」の R 境界に呼びかけ系感動詞を追加（その他、関連する箇所を修正）
- 「長い発話単位」の統語的境界（L 境界）の「独立度の低い従属節の終端境界」に関する規定から「終助詞（相当表現）を伴う場合」を削除
- 付録 A に呼びかけ系感動詞を追加

Ver. 2.0 一般公開

- 用語変更（「長い/短い単位」→「長い/短い発話単位」）
- 「はじめに」「変更履歴」追加
- 「長い発話単位」修正・追記（「終助詞」の項目を「統語論的境界」に移動）
- 「短い発話単位」修正・追記（境界種類、「X-JToBI との比較」、決定木を追加）
- 事例差し替え
- 敬体から常体に変更

Ver. 1.4

- 用語変更（「語用論的境界」→「相互行為的境界」）
- 「伝達モダリティ」廃止（「発話中の終助詞」以外）

Ver. 1.3

- 「統語的境界」修正・追記

Ver. 1.2

- 「X 境界」追加
- 「語用論境界」項目追加
- 「接続詞一覧」追加
- 事例差し替え

Ver. 1.1

- タイトル変更
- 「語用論境界」項目追加
- 「長い単位」決定木追加

Ver. 1.0 プロジェクト内リリース