

国立国語研究所学術情報リポジトリ

『日本語日常会話コーパス』（2018年度版）の調整 頻度を活用した分析

メタデータ	言語: Japanese 出版者: 公開日: 2021-03-26 キーワード (Ja): キーワード (En): Chunagon 作成者: 柏野, 和佳子, 小磯, 花絵, Kashino, Wakako, Koiso, Hanae メールアドレス: 所属:
URL	https://doi.org/10.15084/00003266

This work is licensed under a Creative Commons
Attribution-NonCommercial-ShareAlike 3.0
International License.



『日本語日常会話コーパス』 (2018年度版) の調整頻度を活用した分析

柏野 和佳子 小磯 花絵



「度数・頻度」「粗頻度」

- 度数・頻度：コーパスから取り出された生の値。
- 粗頻度 raw frequency：コーパスから観測された頻度情報。rawには、「生の、未加工の」といった意味があり、コーパスから取り出された状態のまま、調整を加えていないことをふまえ、このように呼ぶ。
- 粗頻度の特性：元となったコーパスサイズの情報が含まれていない。「10回」という粗頻度が1万語コーパスから得られた10回なのか、100万語コーパスから得られた10回なのか示されていない。



サイズの異なるコーパスから取り出された粗頻度は互いに比較できない。

「調整頻度」

- 調整頻度 adjusted frequency : 異なるコーパスから得られた頻度の相互比較が可能になるよう、粗頻度を一定の基準によって調整した値のこと。
- 調整頻度の特性:
 - たとえば、500万語コーパスから得られた10回と、2,000万語コーパスから得られた30回を、そのまま比較することに意味がない。
 - 共通の物差しを用意して、個々の頻度をそれにあわせて調整する必要がある。
 - 最も単純な方法は、度数÷総度数 で求めること。
 - 0.000002回($10 \div 500万$)と、0.0000015回($30 \div 2,000万$)

※あまりに値が小さく、扱いがたい。

→ 単純な相対頻度に変えて、調整頻度を用いる。

「PMW」

- 調整頻度の基準値にはさまざまな値を使用できるが、値をある程度大きくして読み取りやすくするために、コーパス研究では、通例、
「100万語あたり調整頻度」
が使用される。
- PMW per million words: 100万語あたり調整頻度は、英語の頭文字をとり、PMWと表記する。
 - 0.000002回 ($10 \div 500万$) \times 100万 = 2回
 - 0.0000015回 ($30 \div 2,000万$) \times 100万 = 1.5回
- ただし、常にPMWが適当なわけではない。100万語未満のコーパスには不適とする立場がある。
 - 調整しようとするすべてのコーパスサイズの下限值を超えない範囲で最も大きな10の乗数(1万、10万、100万、1,000万…)を選ぶのが原則。

『日本語日常会話コーパス』 (CEJC) 現在構築中, モニター版を公開中

『日本語日常会話コーパス』は, 国立国語研究所の機関拠点型基幹研究プロジェクト「大規模日常会話コーパスに基づく話し言葉の多角的研究」(プロジェクトリーダー: 小磯花絵) の研究成果です。

<http://pj.ninjal.ac.jp/conversation/>

「大規模日常会話コーパスに基づく話し言葉の多角的研究」(約10分半)

<https://youtu.be/HksF4QtbrI8>

- 一般の協力者に依頼し, 日常に生じる自然な会話200時間を記録
- 映像・音声を含めて収録・公開
- 多様な場面・話者を対象

『日本語日常会話コーパス』語数表①(2018年度版)

年齢	男性	女性	計
10代以下	24,044	5,270	29,314
20代	60,269	41,198	101,467
30代	37,201	64,415	101,616
40代	44,391	96,987	141,378
50代	32,051	81,729	113,780
60代	23,476	38,910	62,386
70代以上	37,978	17,592	55,570
不明	773	3,043	3,816
合計	260,183	349,144	609,327

形式名	会話数	会話時間	語数
雑談	84	36.0	434,541
用談相談	23	11.3	122,017
会議会合	9	3.0	52,769
合計	116	50	609,327

場面名	会話数	会話時間	語数
自宅	28	12.7	122,727
職場	8	2.3	32,723
学校	8	3.8	53,207
施設	48	20.6	261,987
室内	15	6.9	92,495
屋外	5	1.9	20,957
交通機関	4	2.1	25,231
合計	116	50	609,327

<https://www2.ninjal.ac.jp/conversation/cejc-monitor/cejc-wc.html>

『日本語日常会話コーパス』語数表②(2018年度版)

話者間の関係性	語数(記号等除外・全て)
サービス場面関係	11650
家族	162578
家族・親戚	62614
仕事関係	24146
仕事関係・同僚	2426
仕事関係・友人知人	7895
先生生徒	14481
先生生徒・友人知人	8581
先生生徒・友人知人・同僚	9832
同僚	26706
友人知人	234916
友人知人・家族	21287
友人知人・家族・初対面	6382
友人知人・仕事関係	4137
友人知人・仕事関係・初対面	2717
友人知人・先生生徒・仕事関係	8979
総計	609327

話者間の関係性	語数(記号等除外・全て)
家族	162578
家族・親戚	62614
友人知人	278418
同僚	26706
先生生徒	32894
仕事関係	34467
接客	11650
総計	609327

<https://www2.ninjal.ac.jp/conversation/cejc-monitor/cejc-wc.html>

【分析①】

助動詞「です」の語形の出現傾向

助動詞「です/ます」の縮約形

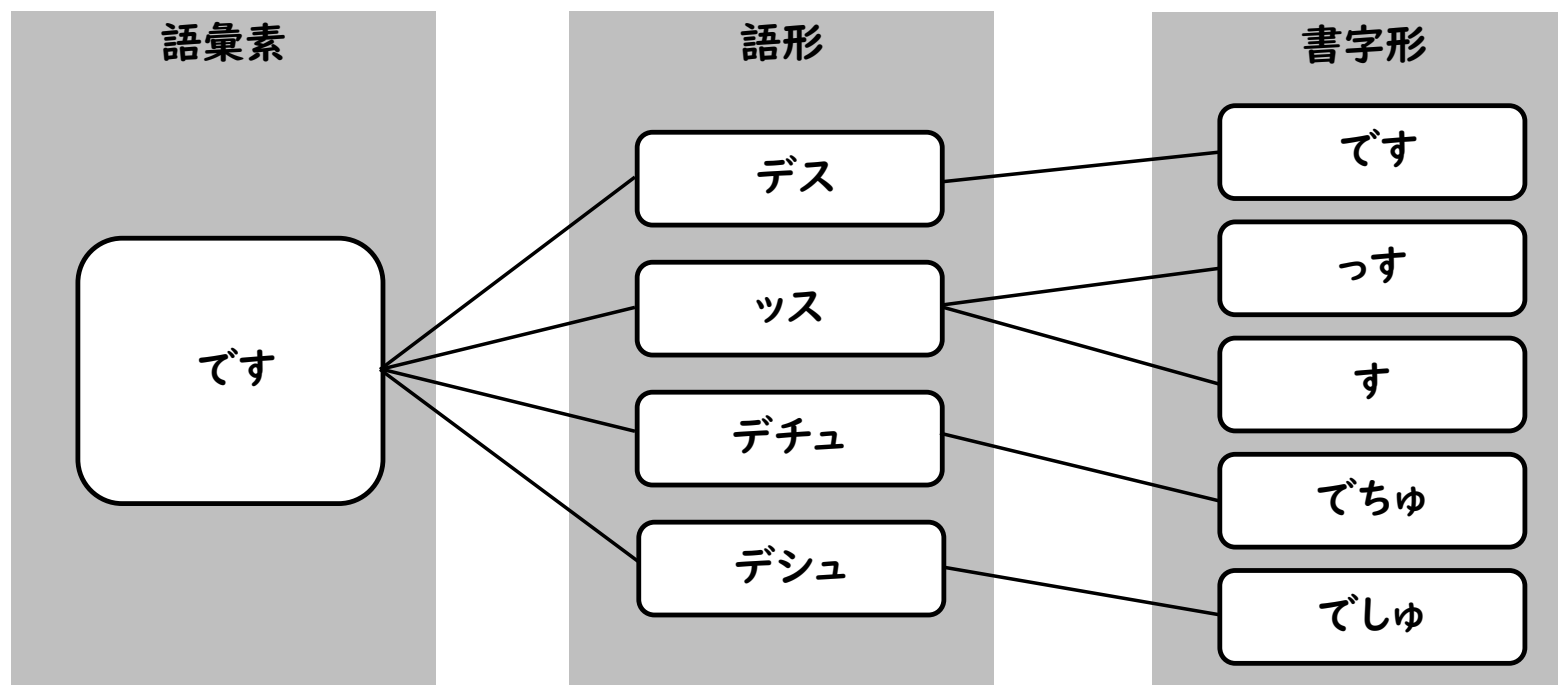
例：早いです（形容詞＋です） → 早いっす
ご飯です（名詞＋です） → ご飯す
やります（動詞＋ます） → やるっす

※ 20代・30代の若い男性を中心に使われている。
女性はあまり用いない。60歳以上では男女ともに使用は見られない。

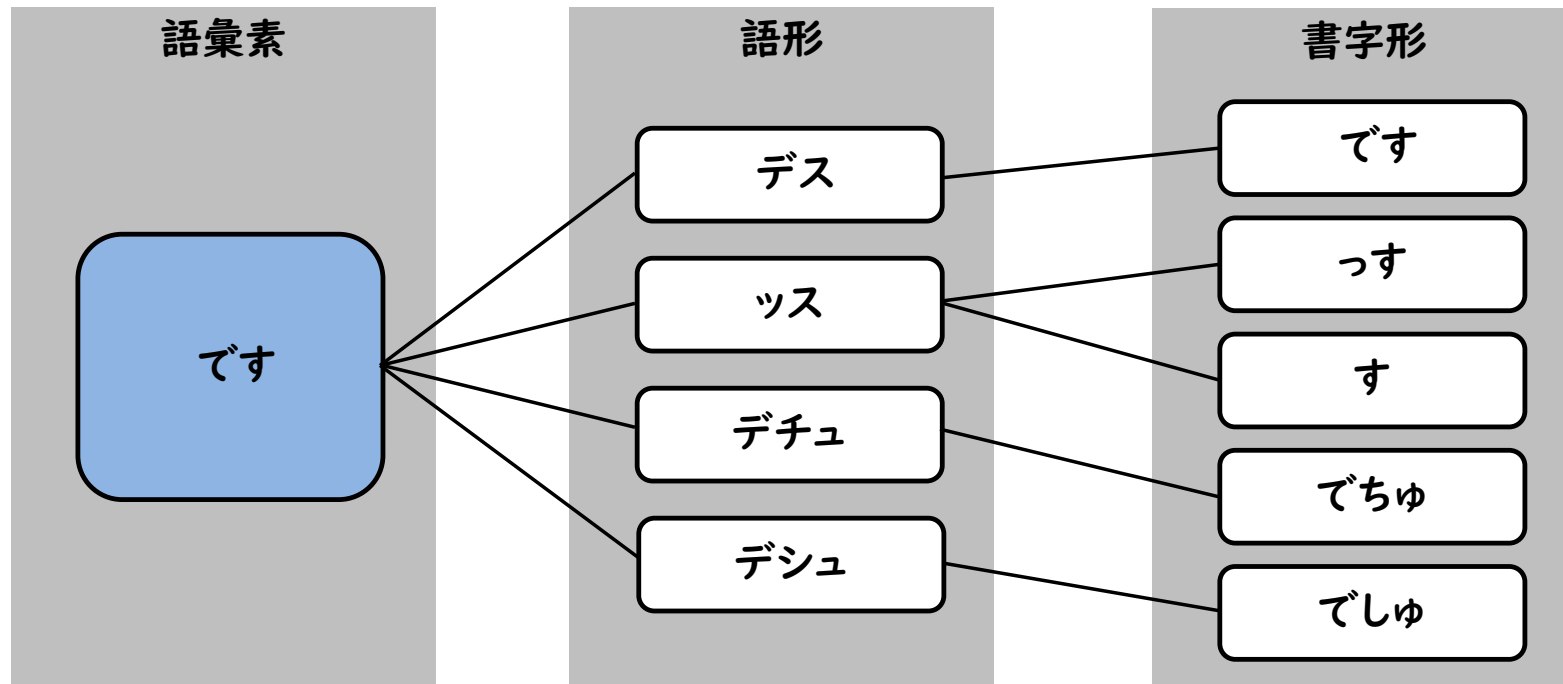
尾崎喜光（2002）「新しい丁寧語「（っ）す」」『男性のことば・職場編』 pp.89-98, ひつじ書房

👉 尾崎（2002）の指摘のとおり若い男性が使用している？

語彙素「です」の階層構造

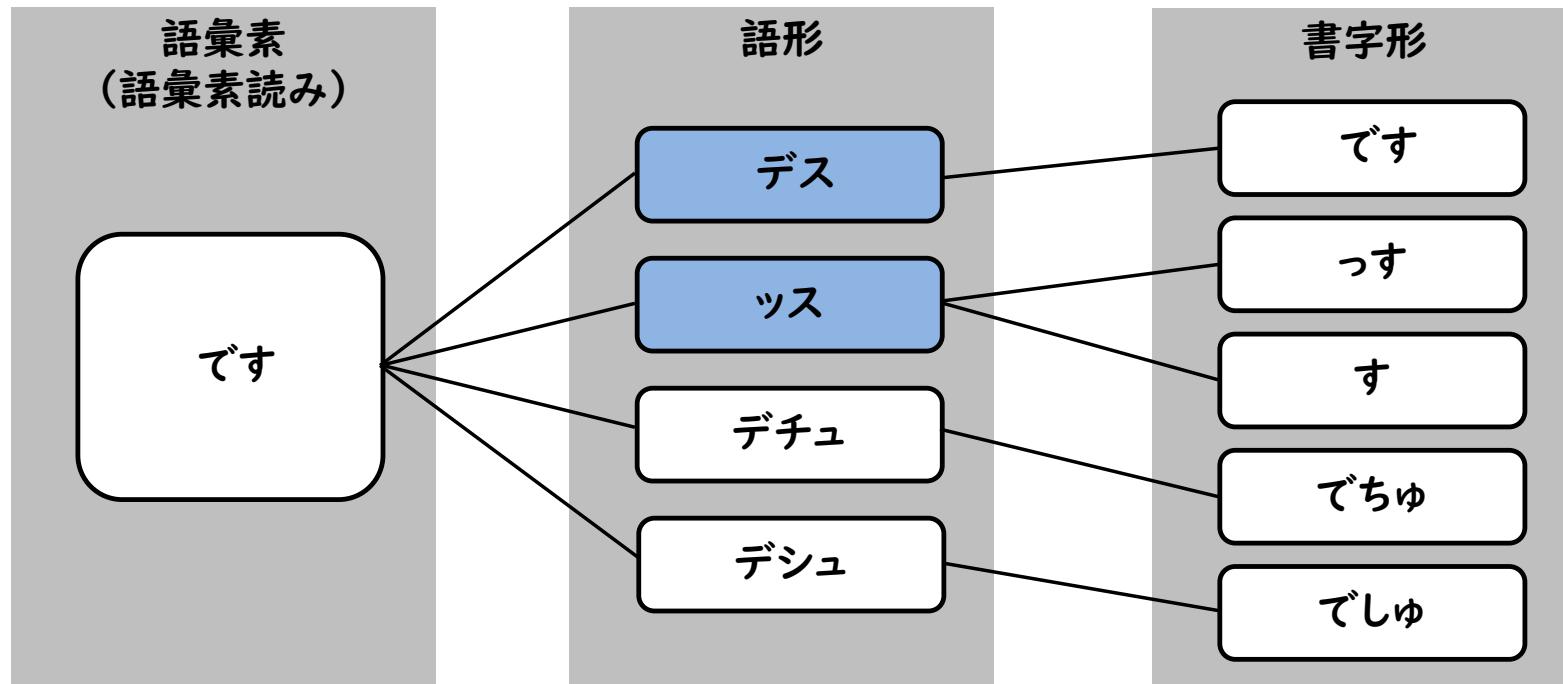


語彙素「です」の階層構造



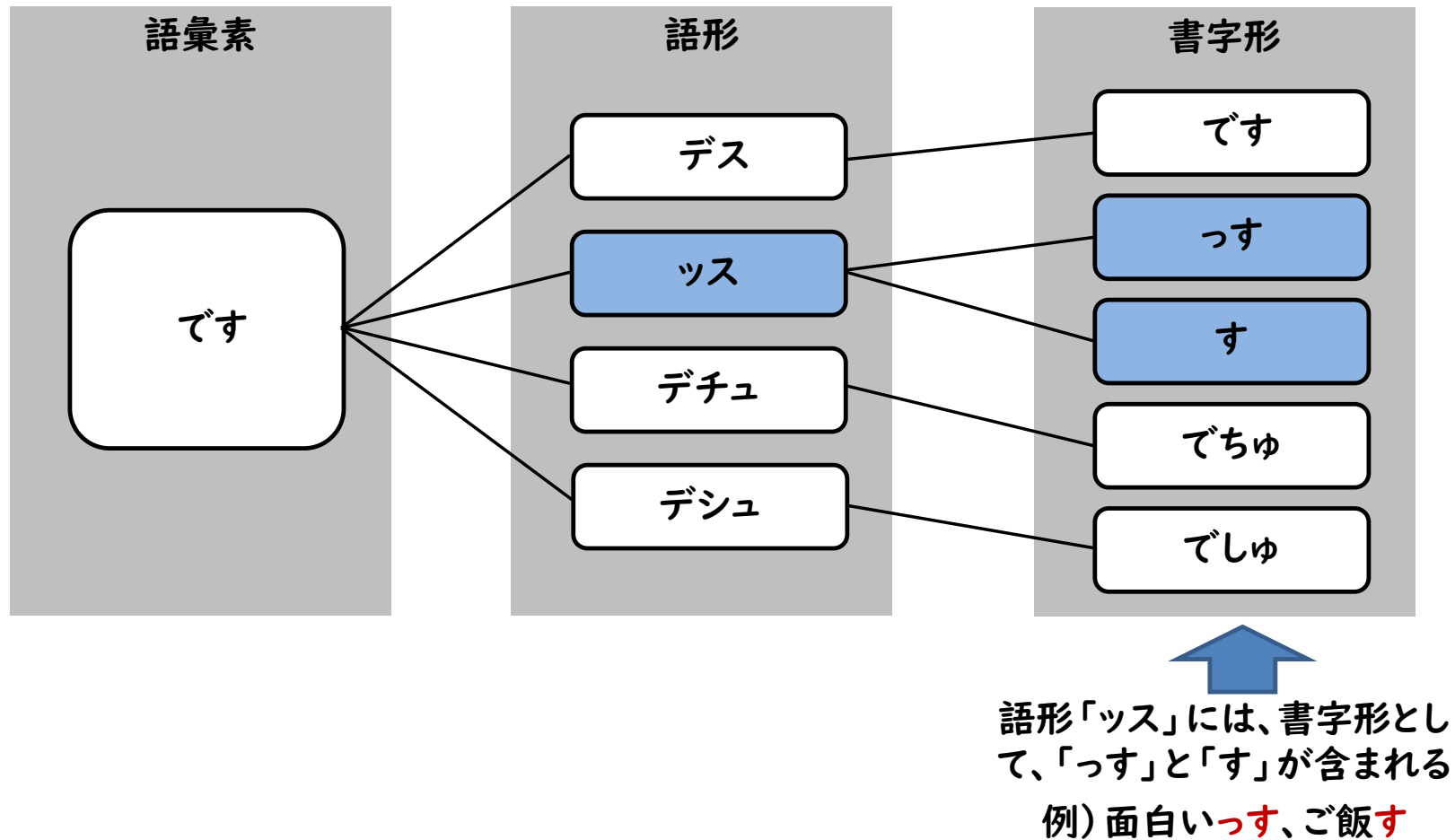
①語彙素「です」で検索

語彙素「です」の階層構造



②検索結果から、4つの語形のうち、
今回対象とする「デス」と「ッス」の頻
度をもとに、各語形の出現率を計算

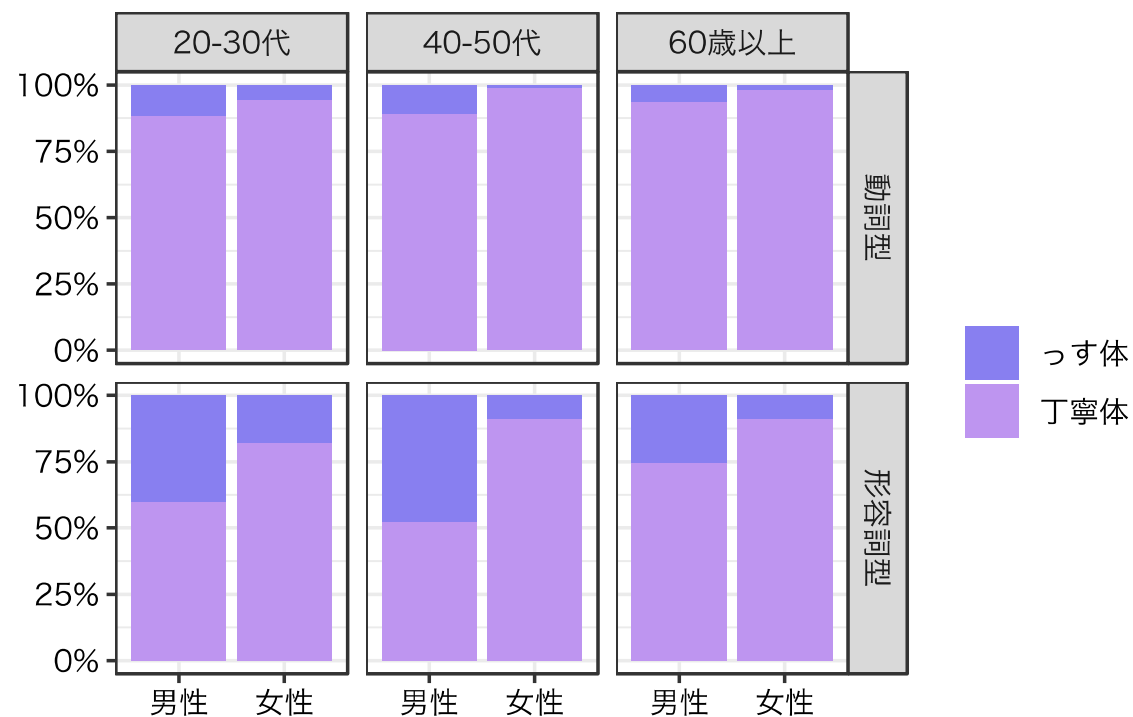
語彙素「です」の階層構造



形容詞+です	早いです	→	早いっす
名詞+です	ご飯です	→	ご飯す
動詞+ます	やります	→	やるっす

形容詞+です	早いです	→	早いっす
名詞+です	ご飯です	→	ご飯す
動詞+ます	やります	→	やるっす

形容詞+です	早いです	→	早いっす	限定
名詞+です	ご飯です	→	ご飯す	
動詞+ます	やります	→	やるっす	



【分析①】

形容詞（終止形） or 名詞 +「です/(っ)す」（終止形）のパターンに着目して分析

ステップ1 検索条件 前方共起条件：品詞「形容詞」、活用形「終止形」
キー： 語彙素「です」、品詞「助動詞」、活用形「終止形」


前方共起1		キーから	1	語	<input type="checkbox"/> キーと結合して表示	この条件をキーに	この共起条件を削除
品詞	の	大分類	が	形容詞			
AND	活用形	の	大分類	が	終止形		短単位の条件の追加

キー	--	1	語	<input type="checkbox"/> キーの条件を指定しない			
語彙素	が	です					
AND	品詞	の	大分類	が	助動詞		
AND	活用形	の	大分類	が	終止形		短単位の条件の追加

【分析①】

ステップ2 検索履歴から検索条件式を編集

```
キー: (語彙素="です" AND 品詞 LIKE "助動詞%" AND 活用形 LIKE "終止形%") AND  
前方共起: (品詞 LIKE "形容詞%" AND 活用形 LIKE "終止形%")  
ON I WORDS FROM キー  
WITH OPTIONS tglKugiri="|" AND tglBunKugiri="#" AND  
limitToSelfSentence="I" AND tglWords="20" AND  
unit="I" AND encoding="UTF-16LE" AND endOfLine="CRLF"
```

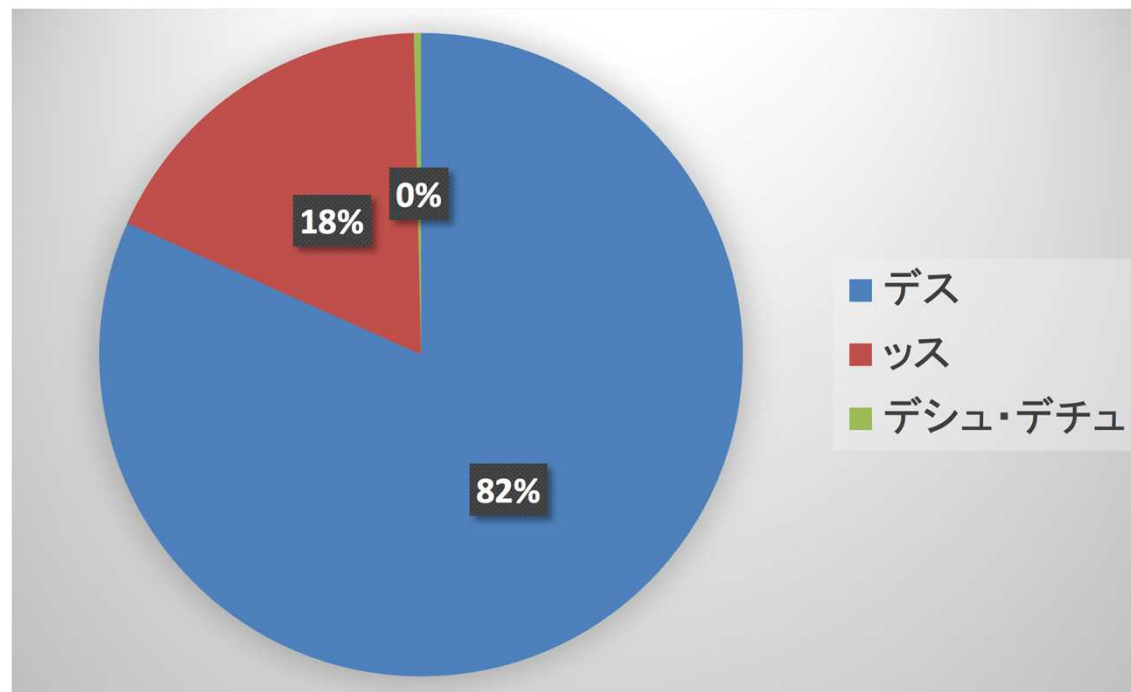


```
キー: (語彙素="です" AND 品詞 LIKE "助動詞%" AND 活用形 LIKE "終止形%") AND  
前方共起: ( (品詞 LIKE "形容詞%" AND 活用形 LIKE "終止形%") OR  
品詞 LIKE "名詞%" )  
ON I WORDS FROM キー  
WITH OPTIONS tglKugiri="|" AND tglBunKugiri="#" AND  
limitToSelfSentence="I" AND tglWords="20" AND  
unit="I" AND encoding="UTF-16LE" AND endOfLine="CRLF"
```

前方共起の条件で、形容詞（終止形）に加えて名詞を **OR条件** を使って追加

【分析①】

各語形の出現率

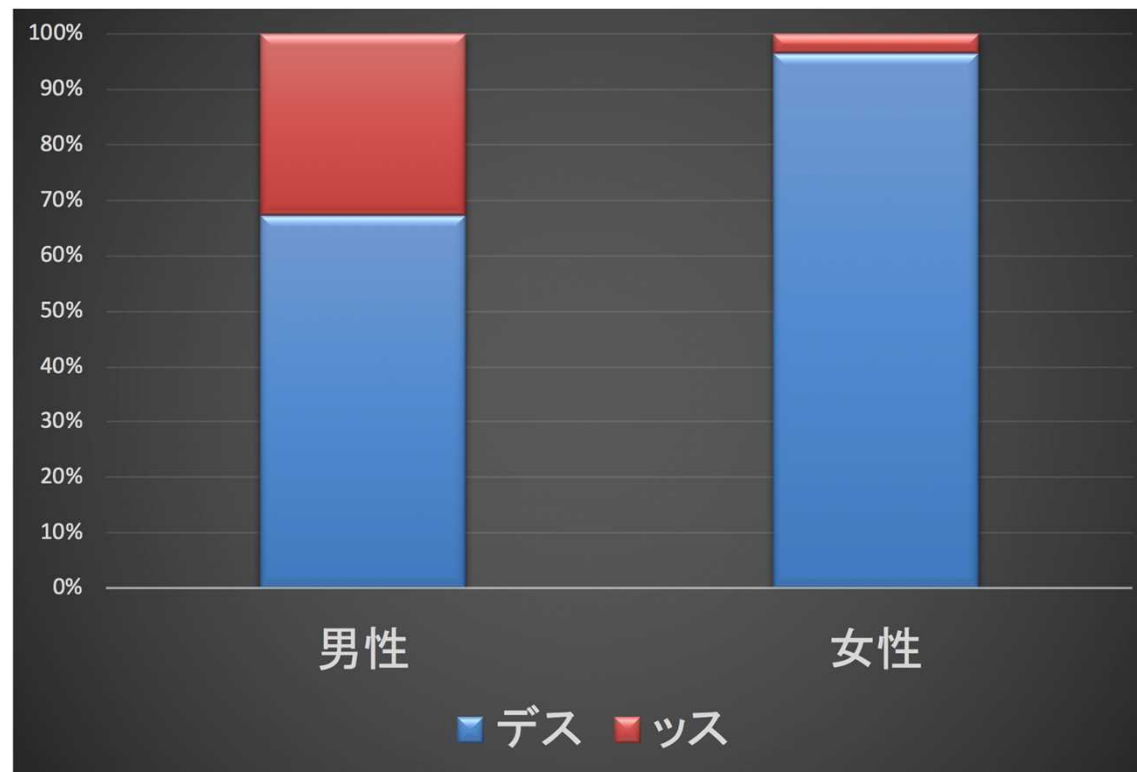


「っす」が全体の2割弱と少なからず見られる

※「でしゅ」「でちゅ」は幼児語 …… このあとの分析からは除外

【分析①】

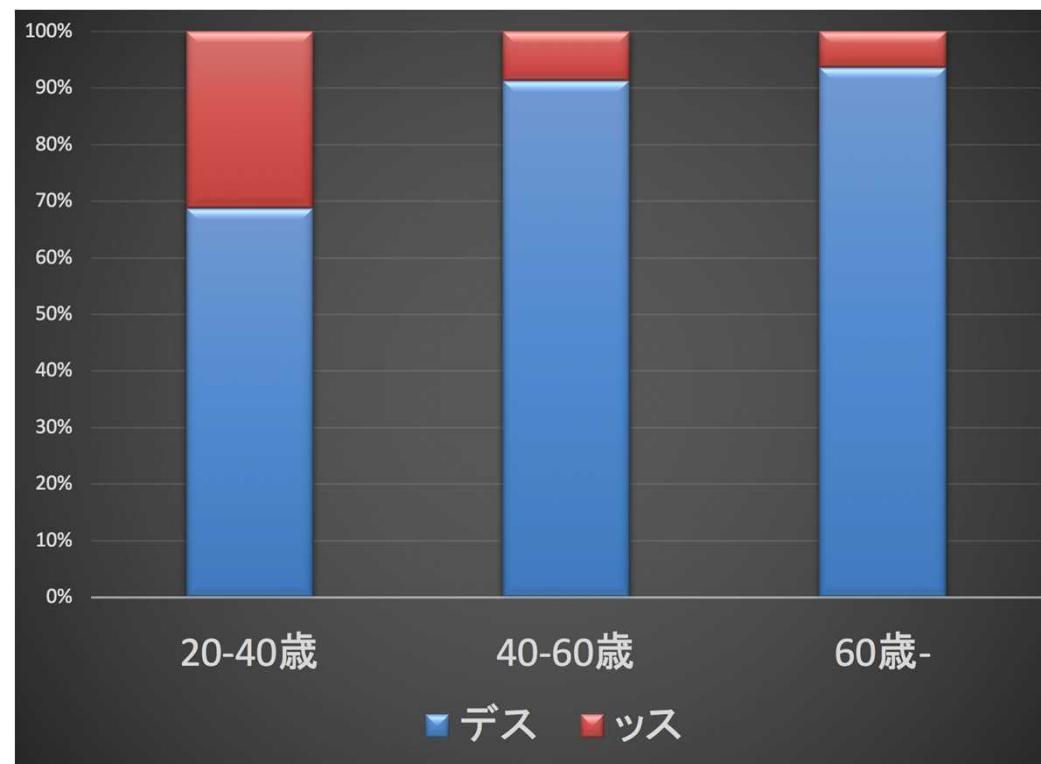
性別との関係



男性の方が女性よりも「っす」を多く用いる

【分析①】

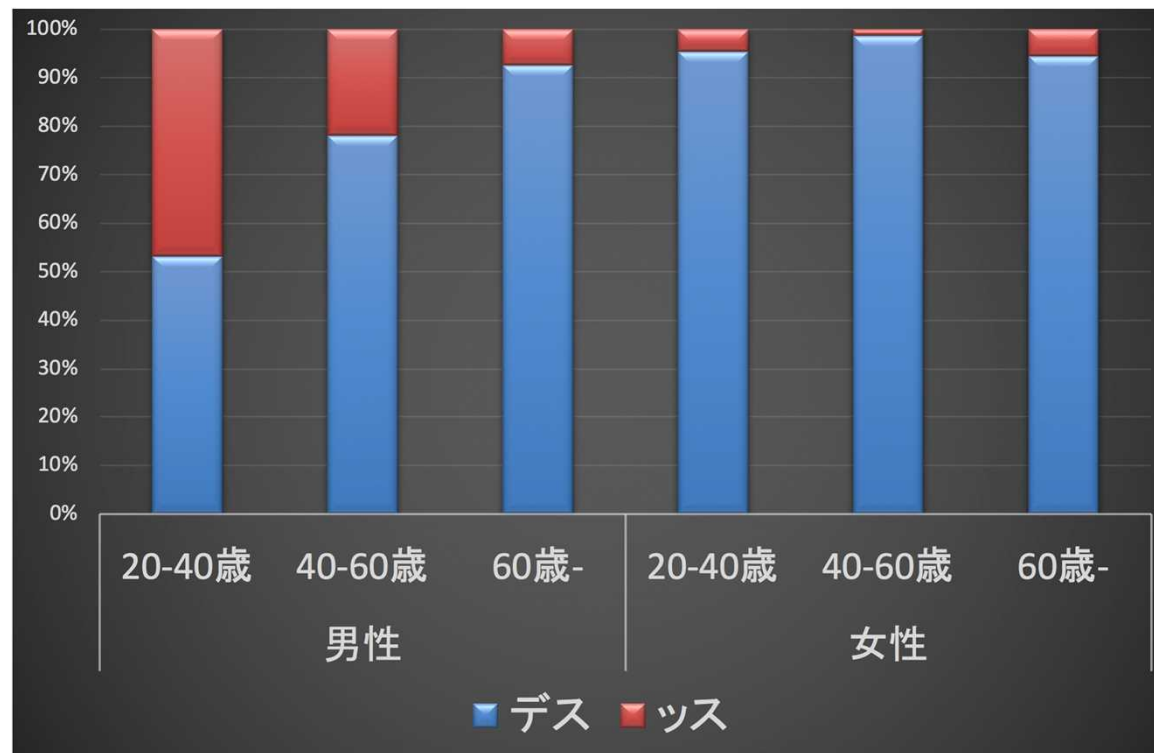
年齢との関係



20～40歳の若手の方が「っす」を多く用いる

【分析①】

性別・年齢との関係



男性は若い世代ほど「っす」を用いるという傾向が顕著に見られる。
女性はこの世代もあまり用いない。

【分析②】

縮約の出現傾向

接続助詞「て/で」+ 動詞「いる」 → てる/でる

接続助詞「て/で」+ 動詞「おく」 → とく/どく

接続助詞「て/で」+ 動詞「しまう」 → ちゃう/ちゃう

接続助詞「て/で」+ 係助詞「は」 → ちゃ/ぢゃ

準体助詞「の」+ 助動詞「だ」終止形

助動詞「で」+ 係助詞「は」 → 「じゃ」

動詞假定形 + 接続助詞「ば」 → 融合

[こそあ]れ+係助詞「は」 → [こそあ]りゃ

(例: 見ている → 見てる)

(例: 取っておく → 取っとく)

(例: 死んでしまう → 死んじゃう)

(例: 書いてはだめ → 書いちゃだめ)

(例: 見るのだけど → 見るんだけど)

(例: 好きではない → 好きじゃない)

(例: 知れば → 知りゃ)

(例: これは → こりゃ)

接続助詞「て/で」 + 動詞「いる」 → 助動詞「てる/でる」

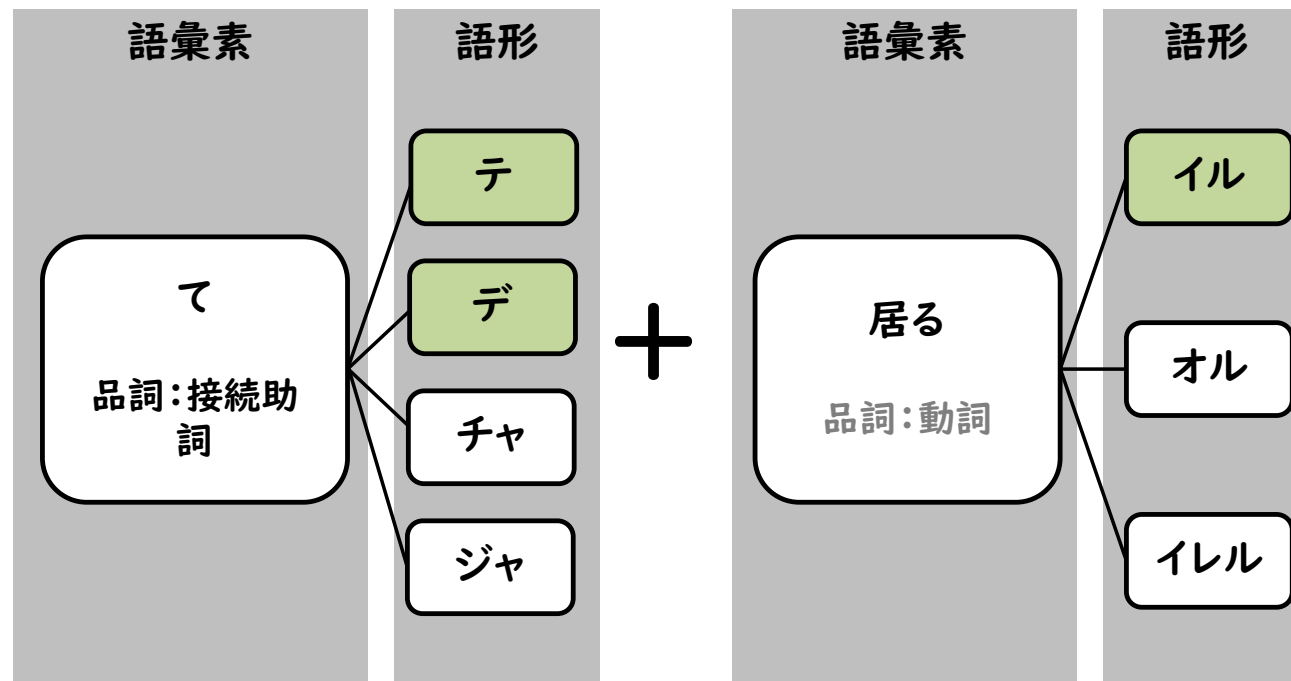
例：見+て+いる → 見+てる
飲ん+で+いる → 飲ん+でる

接続助詞「て/で」 + 動詞「いる」 → 助動詞「てる/でる」

例：見+て+いる → 見+てる
飲ん+で+いる → 飲ん+でる

動詞 +
連用形

例)
見る
飲む



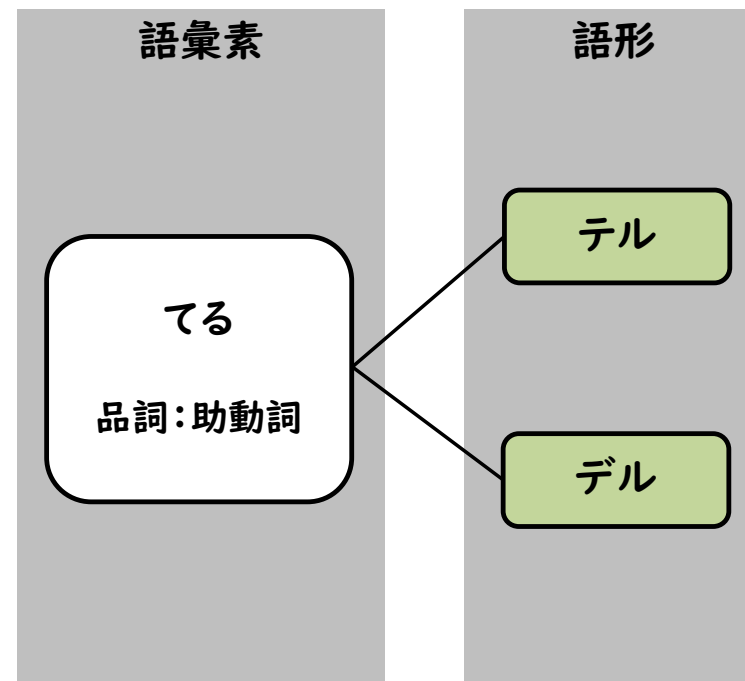
接続助詞「て/で」+ 動詞「いる」 → 助動詞「てる/でる」

例：見+て+いる → 見+てる

飲ん+で+いる → 飲ん+でる

動詞 +
連用形

例)
見る
飲む



【分析②-1】 接続助詞「て/で」+ 動詞「いる」 → てる/でる

ステップ1 検索条件 前方共起: 品詞:動詞

キー: 品詞(中分類)「助詞-接続助詞」、語彙素「て」

後方共起: 語彙素「居る」、語形「テル」

前方共起1		キーから	1	語	<input type="checkbox"/> キーと結合して表示	この条件をキーに	この共起条件を削除
品詞	の	大分類	が	動詞	短単位の条件の追加		
キー							
キー		--	1	語	<input type="checkbox"/> キーの条件を指定しない		
品詞	の	中分類	が	助詞-接続助詞	短単位の条件の追加		
AND	語彙素	が	て	短単位の条件の追加			
後方共起1		キーから	1	語	<input type="checkbox"/> キーと結合して表示	この条件をキーに	この共起条件を削除
語彙素	が	居る	短単位の条件の追加				
AND	語形	が	イル	短単位の条件の追加			

検索結果から
接続助詞「て」
の語形が
「チャ」「ジャ」
があれば除く
※実際はなし

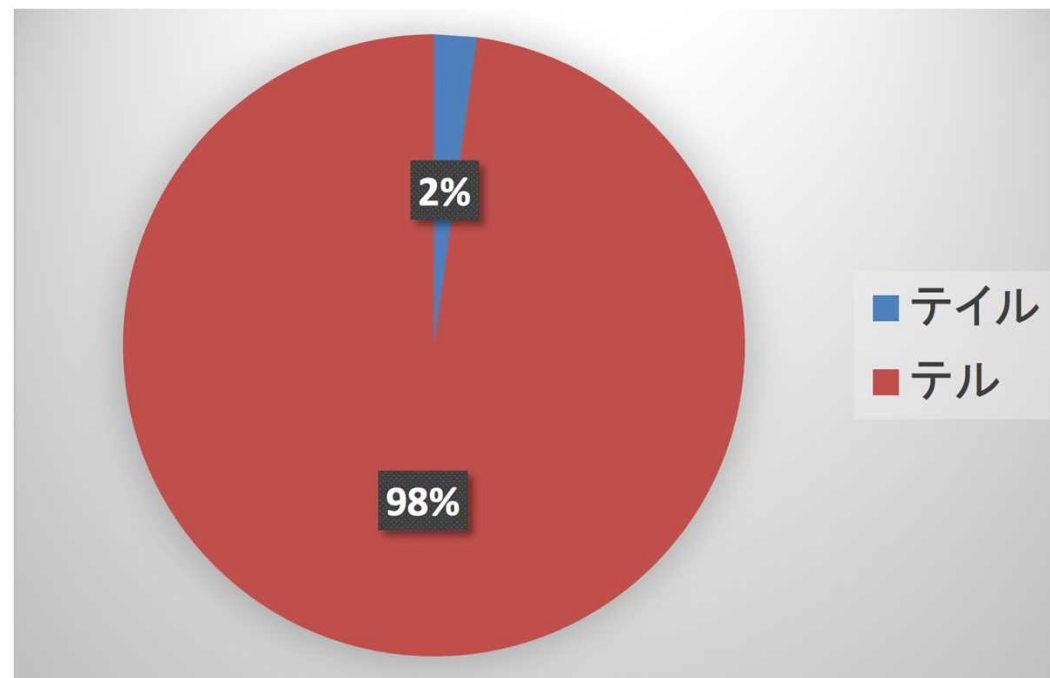
ステップ2 検索条件 前方共起: 品詞:動詞

キー: 品詞「助動詞」、語彙素「てる」 ※縮約形「てる」

前方共起1		キーから	1	語	<input type="checkbox"/> キーと結合して表示	この条件をキーに	この共起条件を削除
品詞	の	大分類	が	動詞	短単位の条件の追加		
キー							
キー		--	1	語	<input type="checkbox"/> キーの条件を指定しない		
品詞	の	大分類	が	助動詞	短単位の条件の追加		
AND	語彙素	が	てる	短単位の条件の追加			

【分析②-1】

「て+いる」の縮約形「てる」



日常会話では縮約形「てる」が多用される

接続助詞「て/で」 + 係助詞「は」 → 接続助詞「ちゃ/じゃ」

例：見＋て＋は → 見＋ちゃ

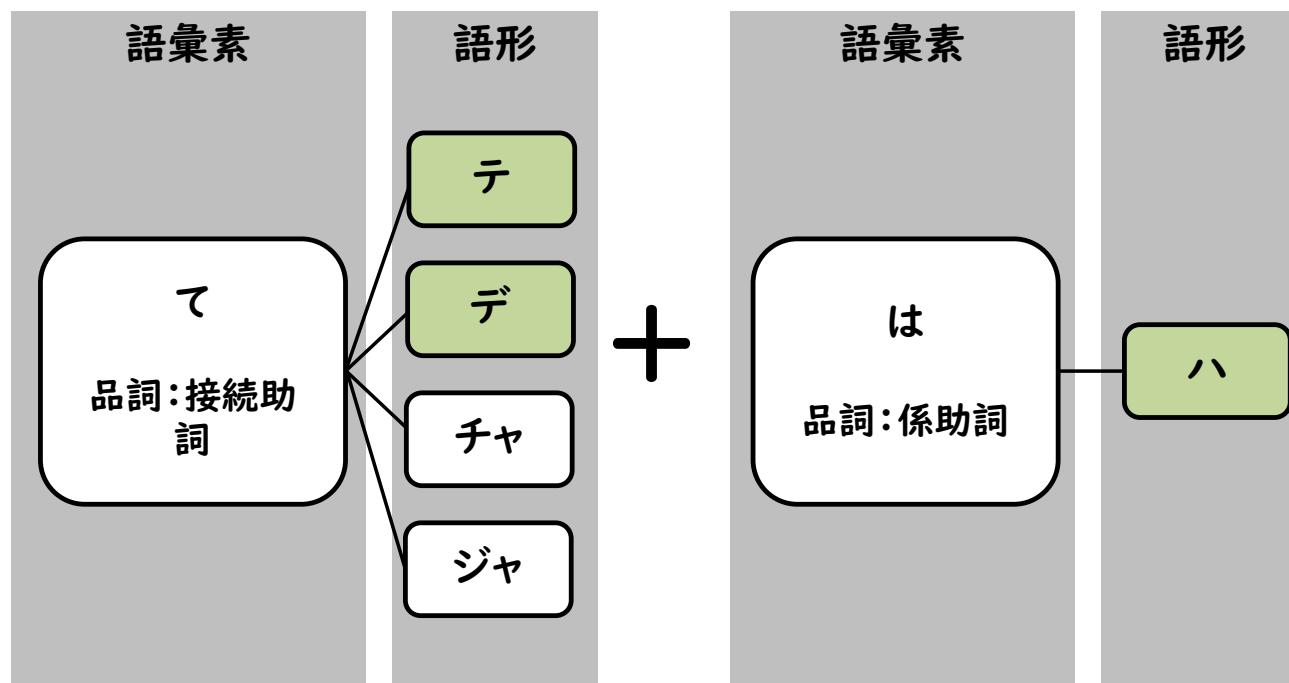
飲ん＋で＋は → 飲ん＋じゃ

接続助詞「て/で」 + 係助詞「は」 → 接続助詞「ちゃ/じゃ」

例：見＋て＋は → 見＋ちゃ
飲ん＋で＋は → 飲ん＋じゃ

動詞＋
連用形

例)
見る
飲む



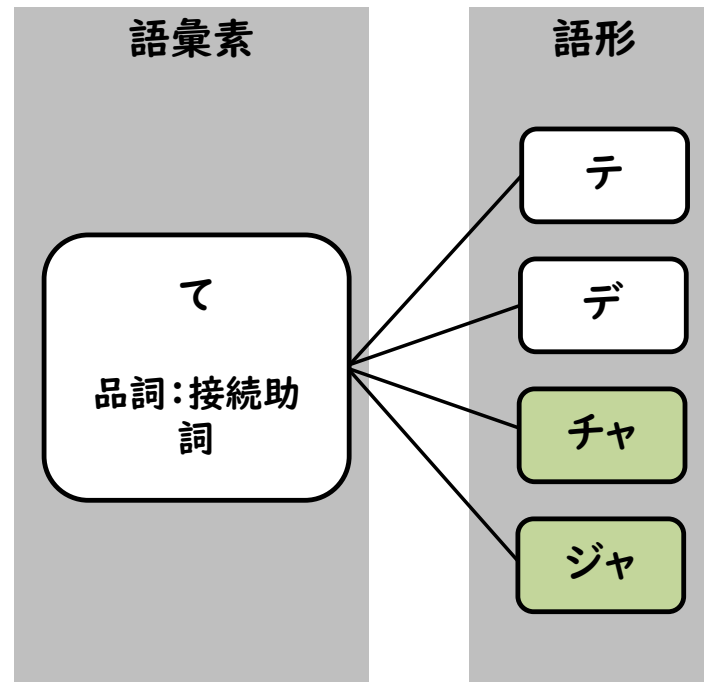
接続助詞「て/で」 + 係助詞「は」 → 接続助詞「ちゃ/じゃ」

例：見＋て＋は → 見＋ちゃ
飲ん＋で＋は → 飲ん＋じゃ

動＋
詞

連用形

例)
見る
飲む



【分析②-2】 接続助詞「て/で」+ 係助詞「は」 → ちゃ/ぢゃ

ステップ1 検索条件 前方共起: 品詞:動詞

キー:品詞(中分類)「助詞-接続助詞」、語彙素「て」

後方共起:品詞(中分類)「助詞-係助詞」、語彙素「は」

i 前方共起1	キーから	1	語	<input type="checkbox"/> キーと結合して表示 i	この条件をキーに	この共起条件を削除
品詞	の	大分類	が	動詞		短単位の条件の追加

i キー	--	1	語	<input type="checkbox"/> キーの条件を指定しない i	
品詞	の	中分類	が	助詞-接続助詞	
AND 語彙素	が	て		短単位の条件の追加	

i 後方共起1	キーから	1	語	<input type="checkbox"/> キーと結合して表示 i	この条件をキーに	この共起条件を削除
品詞	の	中分類	が	助詞-係助詞		
AND 語彙素	が	は		短単位の条件の追加		

【分析②-2】 接続助詞「て/で」+ 係助詞「は」 → ちゃ/ぢゃ

ステップ2a 検索条件 前方共起: 品詞:動詞

キー: 品詞(中分類)「助詞-接続助詞」、語彙素「て」、語形「ちゃ」

前方共起1		キーから	1	語	<input type="checkbox"/> キーと結合して表示	この条件をキーに	この共起条件を削除
品詞	の	大分類	が	動詞	短単位の条件の追加		
キー		--	1	語	<input type="checkbox"/> キーの条件を指定しない		
品詞	の	中分類	が	助詞-接続助詞			
AND	語彙素	が	て				
AND	語形	が	ちゃ	短単位の条件の追加			

ステップ2b 検索履歴から検索条件式を編集

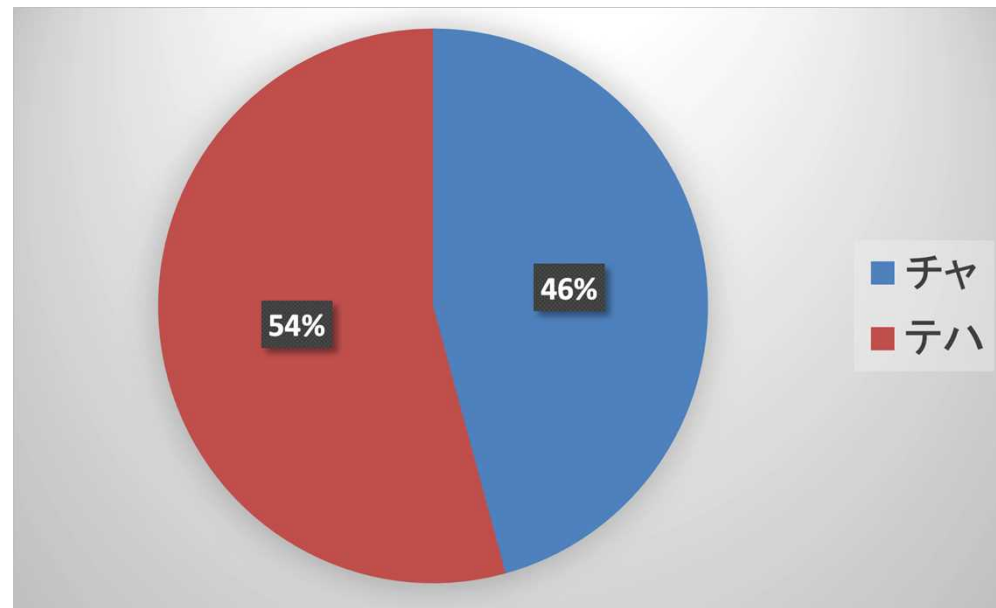
キー: (品詞 LIKE "助詞-接続助詞%" AND 語彙素="て" AND 語形="ちゃ") AND
前方共起: 品詞 LIKE "動詞%" ON 1 WORDS FROM キー



キー: (品詞 LIKE "助詞-接続助詞%" AND 語彙素="て" AND 語形 LIKE "[チジ]ャ") AND
前方共起: 品詞 LIKE "動詞%" ON 1 WORDS FROM キー

【分析②-2】

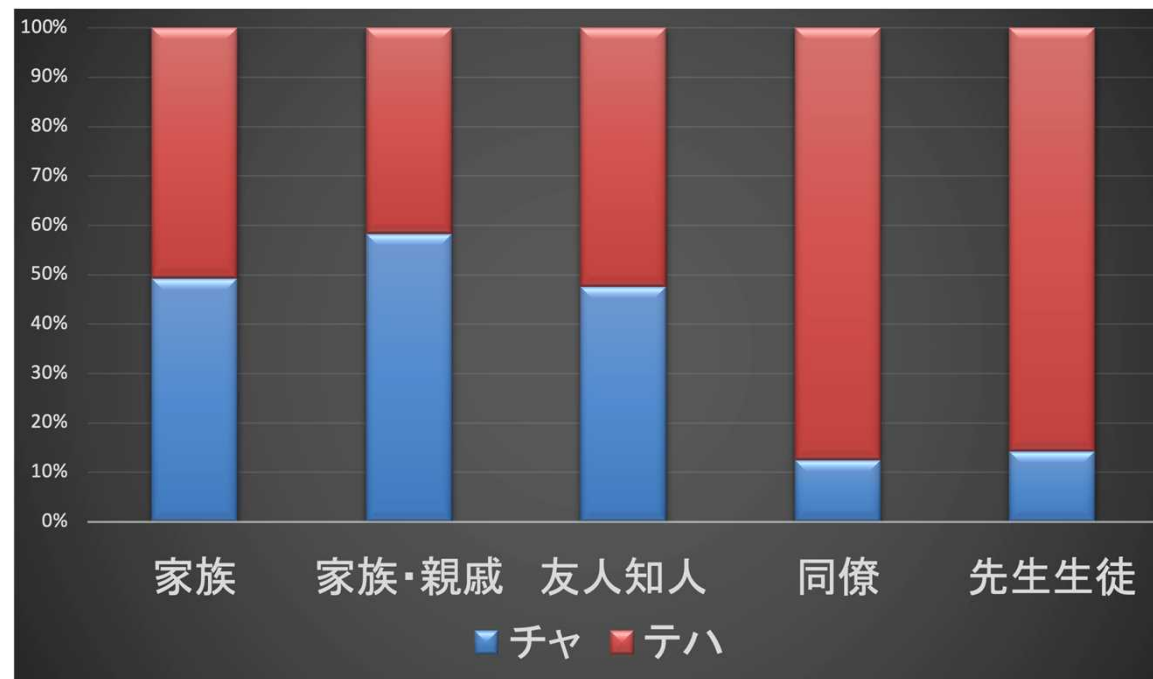
「て/で+は」の縮約形「ちゃ/じゃ」



縮約形「ちゃ/じゃ」と元の形「て/で+は」は同程度に用いられる

【分析② -2】

話者間の関係性



同僚や先生生徒の間では縮約より元の形を用いやすい

動詞仮定形 + 接続助詞「ば」 → 動詞融合形

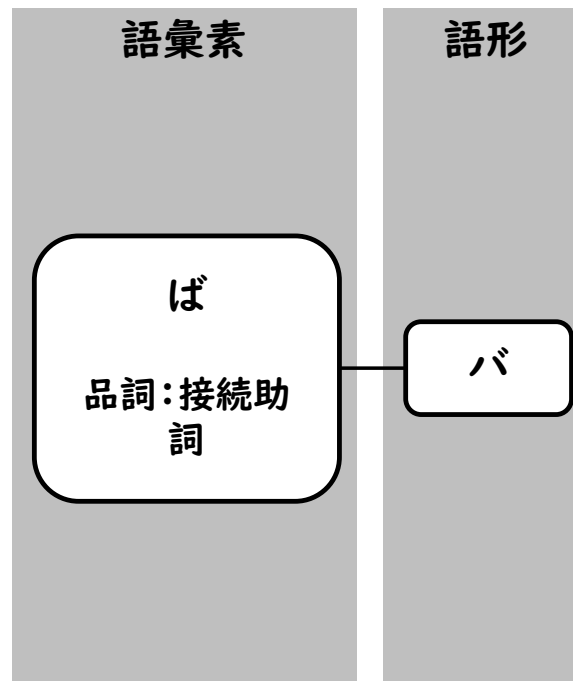
例：見れ + ば → 見りゃ
飲め + ば → 飲みゃ

動詞仮定形 + 接続助詞「ば」 → 動詞融合形

例：見れ + ば → 見りゃ
飲め + ば → 飲みや

動詞
仮定形-一般 +

例)
見れ
飲め



動詞仮定形 + 接続助詞「ば」 → 動詞融合形

例：見れ + ば → 見りや

飲め + ば → 飲みや

動詞 仮定形-融合

例)
見りや
飲みや

【分析②-3】 動詞仮定形 + 接続助詞「ば」 → 融合

ステップ1 検索条件 前方共起: 品詞:動詞、活用形(小分類)「仮定形-一般」
キー:品詞(中分類)「助詞-接続助詞」、語彙素「ば」

i キー	---	1	語	<input type="checkbox"/> キーの条件を指定しない i			
品詞	の	大分類	が	動詞			
AND	活用形	の	小分類	が	仮定形-一般		短単位の条件の追加

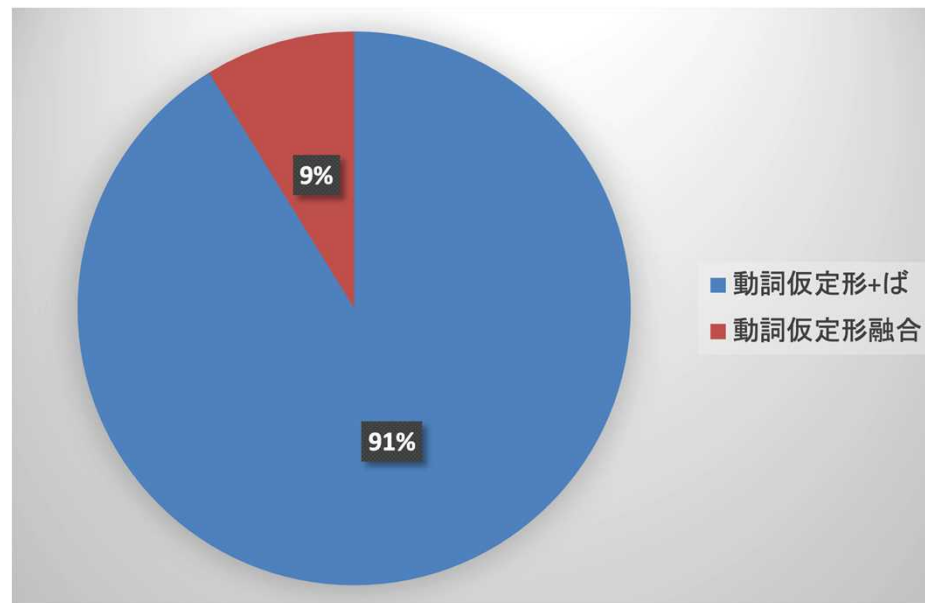
i 後方共起1	キーから	1	語	<input type="checkbox"/> キーと結合して表示 i	この条件をキーに	この共起条件を削除
品詞	の	中分類	が	助詞-接続助詞		
AND	語彙素	が	ば		短単位の条件の追加	

ステップ2 検索条件 キー: 品詞:動詞、活用形(小分類)「仮定形-融合」

i キー	---	1	語	<input type="checkbox"/> キーの条件を指定しない i			
品詞	の	大分類	が	動詞			
AND	活用形	の	小分類	が	仮定形-融合		短単位の条件の追加

【分析②-3】

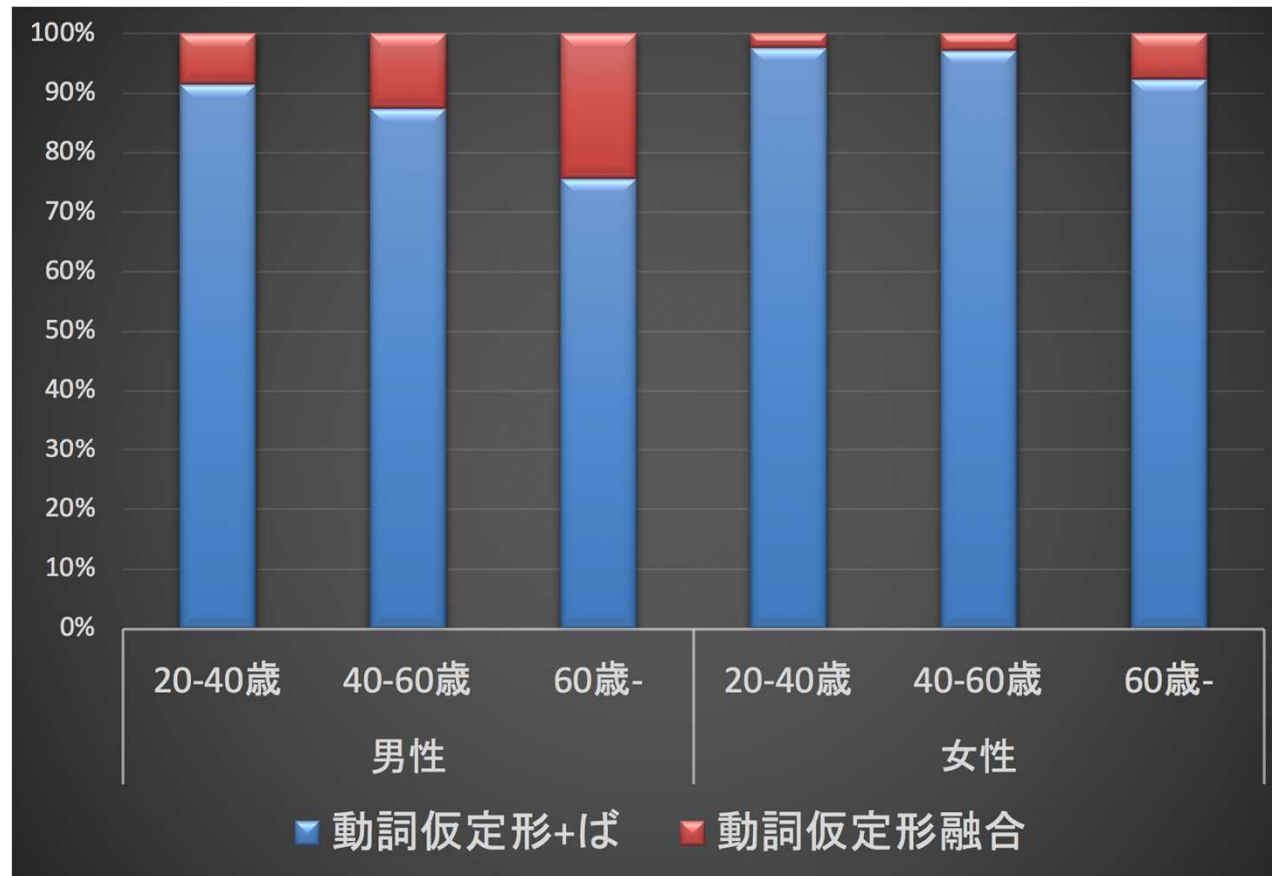
動詞仮定形 + 「ば」の融合形



動詞仮定形 + 「ば」の融合形は全体の約1割程度しか用いられない

【分析②-3】

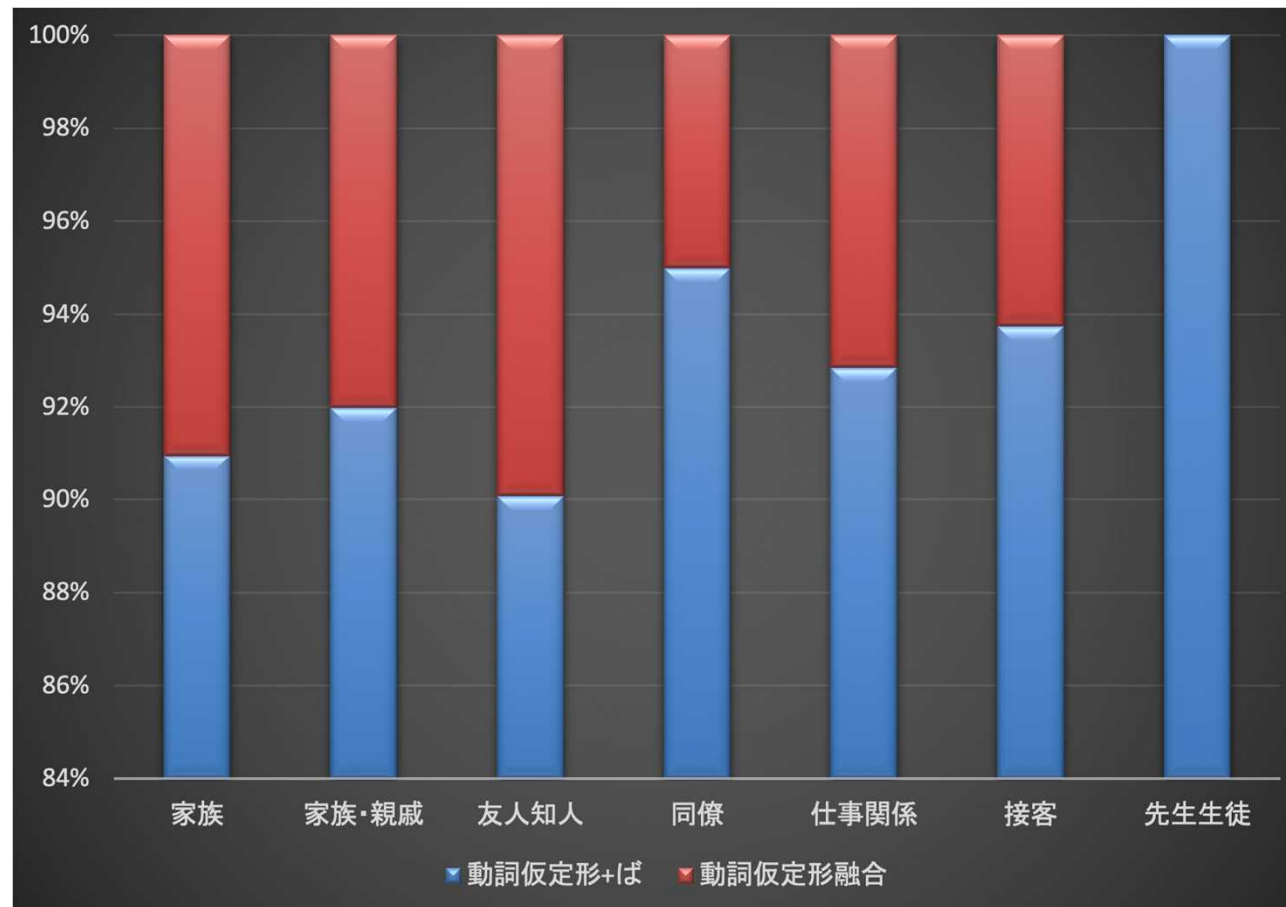
性別・年齢との関係



女性より男性の方が「動詞仮定形+ば」の融合形を用いやすい
年配の世代の方が「動詞仮定形+ば」の融合形を用いやすい

【分析②-3】

話者間の関係性



同僚や先生生徒の間では融合形より元の形「動詞仮定形+ば」を用いやすい

【イベント】国立国語研究所シンポジウム「日常会話コーパス」VI

2021年3月4日(木) 13時から、Webにて開催(参加無料、要参加申し込み)

詳細・申し込み方法:<https://www2.ninjal.ac.jp/conversation/event/sympo6.html>

プログラム

- 13:00-13:15 小磯花絵(国語研究所)
「プロジェクトの進捗報告」
- 13:20-13:50 柏野和佳子(国語研究所)
「『日本語日常会話コーパス』で見直す言葉の用法
ー『岩波国語辞典』第八版でこう改訂したー」
- 13:55-14:25 伝康晴(千葉大学)
「『日本語日常会話コーパス』だからできること：独り言とスマホいじりの分析」
- 14:30-15:00 河野礼実(山梨学院大学)
「『日本語日常会話コーパス』における自称使用」
- 15:05-15:35 居關友里子(国語研究所)
「『日本語日常会話コーパス』における談話行為タグの利用可能性」
- 15:40-16:10 丸山岳彦(専修大学)
「『昭和話し言葉コーパス』の構築(4)：メタデータ的设计とデータの公開」
- 16:15-16:45 田嶋明日香(お茶の水女子大学)・星野祐子(十文字学園女子大学)・
高崎みどり(文教大学)
「『日本語日常会話コーパス』を利用した指示語の用法の識別」
- 16:50-17:20 高田三枝子(愛知学院大学)
「『昭和話し言葉コーパス』における用言のアクセント」