国立国語研究所学術情報リポジトリ

『日本語話し言葉コーパス（CSJ）』模擬講演における節頭フィラーの特徴

メタデータ	言語: Japanese 出版者: 国立国語研究所 公開日: 2019-02-14 キーワード: 日本語話し言葉コーパス(CSJ) 作成者: 渡辺, 美知子, 是松, 優作, KOREMATSU, Yusaku メールアドレス: 所属:
URL	https://doi.org/10.15084/00001692

『日本語話し言葉コーパス（CSJ）』模擬講演における節頭フィラーの特徴

渡辺 美知子（国立国語研究所）†
是松 優作（東京大学）

Features of Clause-Initial Filled Pauses in Simulated Public Speaking of "The Corpus of Spontaneous Japanese (CSJ)"

Michiko Watanabe (National Institute for Japanese Language and Linguistics)
Yusaku Korematsu (The University of Tokyo)

要旨

文境界，節境界にはフィラーがよく観察される。これは，文や節などの談話の切れ目では，その先の発話内容や表現を考えるのに時間が必要なためと考えられる。本研究では，文頭，節頭でのフィラーの使用に影響する要因を，節頭フィラー全体ならびに頻度の高い「エー」「アノ」「マー」について CSJ コア中の模擬講演を対象に調べた。考察対象とした要因は，①話者の性別，②年齢，③学歴，④講演経験，⑤直前の境界の種類，⑥節中語数，⑦節頭の接続詞の有無，の 7 つである。節頭フィラーの出現確率との間に有意な関連が見られたのは要因①，⑤，⑥のみであった。フィラーの出現確率は男性話者の方が女性話者よりも高く，強い節境界の方が文境界よりも高かった。頻度の高い 3 種類のフィラーを個別に見ると，エーのみ，文境界の方が強い節境界よりも出現確率が高かった。エーは，アノ，マーよりも，深い談話境界で用いられる傾向のあることが明らかになった。

１．はじめに

フィラー，繰り返し，音節の引き伸ばし，言い誤りなどの言い淀みは，自発発話に頻繁に観察される。このような現象は朗読音声には稀なことから，その場で発話内容を考えながら次々と生成していく自発発話に固有の現象と考えられる。文を書く場合は，自分のペースで遂行しながら書くことができる。表現が適切でない場合は消して修正することができる。けれども話し言葉では，一旦発せられた音声は消すことができないし，先を考えるための間が長すぎると，話し相手や聞き手に不審に思われる可能性がある。フィラー，繰り返し，引き伸ばしなどの言い淀みは，不自然に長いポーズを避け，表現や内容を模索中であることを相手に伝えるための一方策と考えることができる。本研究では，これらの言い淀みの中で頻度の高いフィラーの働きに着目する。

認知的，言語的な制約から，メッセージはなんらかのまとまりを持った単位として生成される必要がある。広く受け入れられている発話生成モデルでは，生成プロセスに，発話内容の生成，言語形式の選択，調音という３つのレベルを想定している（Levelt 1989）。文や節や句などの統語的単位は，あるまとまりを持ったメッセージの単位としてもとらえることができる。内容が大きくシフトする談話境界や文境界，節境界ではフィラーの頻度の高いことが報告されている（Swerts 1998, Shriberg 1994）。これは，談話の切れ目や文頭，節頭では，次に何をどのように話すかという，ある程度長いスパンの発話内容の生成が行われるためではないかと考えられる。もしそうであれば，次に伝えようとするメッセージ

† watanabem@ninjal.ac.jp

の内容が豊富なほど，統語的に複雑になり，表現に必要な語彙も増えるため，生成に時間がかかり，フィラーが出現する確率も上がると考えられる。渡辺・外山（2017）では，後続節中の語数が多いほど，節頭のフィラーの出現確率が上がるという仮説（複雑さ仮説）を『日本語話し言葉コーパス（CSJ）』(2016)を用いて検証し，仮説を支持する結果が得られた。また，内容的に切れ目の深い境界ほど，次に何をどう話すかの自由度が高いため，後続発話開始に時間がかかり，フィラーの出現確率が上がるという仮説（境界仮説）を，文境界，深い節境界，浅い節境界の 3 種類の境界を対象に検証した。フィラーの出現確率は，文境界 ＜ 弱境界 ＜ 強境界の順になり，弱境界と強境界の間では支持されたが，文境界と節境界間では支持されなかった。渡辺・外山では英語コーパスとの比較のために，20 代～30 代前半の，大卒以上の学歴を持つ男女各 10 名のみを調査対象とした。そこで，本研究では調査対象を広げ，より多様な話者を対象に，節境界におけるフィラーの使用に影響する要因の分析を行った。フィラーの使用に影響する要因として分析対象としたのは以下の 7 要因である。

① 話者の性別:先行研究では男性話者の方がフィラーの使用率は高い（渡辺・外山 2017）。
② 講演時（1999 年～2002 年）の年齢：年齢がフィラーの使用頻度に影響する可能性があると考えた。
③ 学歴：学歴がフィラーの使用頻度に影響する可能性があると考えた。
④ 講演経験：フィラーはリハーサルによって減少することが報告されている（Goldman-Eisler 1968）。講演経験の違いがフィラーの使用に影響する可能性があると考えた。
⑤ 直前の境界の種類：境界が深いほど，フィラーの使用確率は上がると予測した。
⑥ 節中語数：節中の語数が多いほど，節頭でのフィラーの使用確率は上がると予測した。
⑦ 節頭の接続詞の有無：「で，エー」のような，接続詞とフィラーの共起は珍しくない。接続詞の使用とフィラー使用との間に何らかの関係がある可能性を考えた。

次に，フィラーの中でも高頻度の，「エー」，「アノ」，「マー」の使用に影響する要因の分析を同様に行い，それぞれに固有の特徴や働きがあるかどうかについて考察した。

２．方法

２.１　分析資料

『日本語話し言葉コーパス（CSJ）』(2016) コア中の模擬講演 107 講演を分析対象とした。

２.２　手続き

1 で述べた 7 要因に以下のような水準を設定した。

① 話者の性別：講演者の男女の内訳は，女性 54 講演，男性 53 講演である。
② 講演時（1999 年～2002 年）の年齢：20 代から 60 代前半まで 5 歳区切りで 9 グループ。人数の内訳は表 1 の通りである。
③ 学歴：中学または高校卒，大学学部卒，修士以上の 3 グループ。人数の内訳は表 2 の通りである。
④ 講演経験：今回が初めて，1~5 回，6 回～10 回，11 回～20 回，21 回以上の 5 グループ。人数の内訳は表 3 の通りである。
⑤ 直前の境界の種類：直前が文境界（絶対境界），強い節境界（強境界），弱い節境界（弱境界）の 3 カテゴリー。境界タイプの分類は CSJ の分類に寄った（国立国語研究所 2006:

p.270)。

⑥ 節中語数：フィラーや語断片は節中語数には含まれていない。また，節頭の接続詞も，2 つの節を繋ぐものと考え，節中語数には含めなかった。

⑦ 節頭の接続詞の有無：節頭に接続詞があるかないかの 2 カテゴリー。

そして，これらの要因を説明変数とし，節頭でフィラーが用いられる確率を目的変数として，ロジスティック回帰分析混合モデルによって推定した。フィラーの使用頻度や用い方には個人差が大きいため, 話者を変量効果として扱った。分析には, R version 3. 5. 0 上で, lme4 パッケージ内の glmer と lmerTest を用いた。

表 1 話者の年代内訳

年齢	人数
20to24	15
25to29	16
30to34	21
35to39	15
40to44	20
50to54	4
55to59	6
60to64	3
65to69	7
計	107

表 2 話者の学歴内訳

学歴	人数
中学卒	1
高校卒	37
学部卒	59
修士以上	10
計	107

表 3 話者の講演経験内訳

講演経験	人数
初めて	66
1～5回	23
6～10回	4
11回～20回	1
21回以上	6
不明	7
計	107

３．結果

3.1　節頭フィラー全体

結果を表 4 に，フィラーの出現確率推定値をプロットした回帰曲線を図 1 に示す。R 言語では，予測変数が因子の場合，変数名を文字順にソートして先頭に来る変数を基準変数とし, その回帰係数値（estimate）を 0 とおくことが慣習となっている。性別要因の「女性」，学歴要因の「中学・高校卒」，境界の種類要因の「弱境界」節頭接続詞要因の「なし」の回帰係数値が 0 となっているのはこのためである。オッズ比はその変数の効果の大きさを表し，値が 1 から離れているほど効果は大きい。7 要因のうち，有意な効果が見られたのは，性別，直前の境界の種類，節中語数の 3 要因だった。即ち，フィラーの使用確率は，女性話者よりも男性話者の方が高く，弱い節境界よりも文境界，文境界よりも強い節境界で高かった。また，節中語数の増加に伴ってフィラーの使用確率も上昇していた。有意な効果のあった要因は, 同じコーパスの 20 代～30 代前半の話者 20 名を対象として調べた，渡辺・外山（2017）の結果と一致している。ただし，境界に関しては，渡辺・外山では，文境界の出現確率が最も低かったのに対し，本研究では弱境界が最も低かった。一方，話者の年齢，学歴，講演経験，節頭接続詞の有無の影響は観察されなかった。

3.2　フィラーの種類別分析

前節と同様の分析を，頻度の高いフィラータイプ「エー」,「アノ」,「マー」について個別に行った。

エー

全ての変数を投入するとモデルが収束しないため，予備的分析から，影響が小さいと思われる，年齢と講演経験の 2 要因を除き，残り 5 要因を説明変数として分析した。分析結

表 4 節頭におけるフィラーの出現確率を推定するロジスティクス回帰分析結果

Variable	Estimate	Std. Error	z value	Pr(>|z|)		オッズ比
(Intercept)	-1.790	0.302	-5.931	3.01e-09	***	0.167
性別						
女性	0					
男性	0.560	0.173	3.238	0.001	**	1.750
年齢	0.004	0.006	0.617	0.538		1.004
学歴						
中学・高校卒	0					
大学学部卒	0.114	0.186	0.613	0.540		1.121
修士以上	0.853	0.513	1.662	0.097	.	2.347
講演経験	-0.019	0.015	-1.215	0.225		0.982
境界の種類						
弱境界	0					
強境界	0.526	0.044	11.881	< 2e-16	***	1.692
文境界	0.464	0.049	9.553	< 2e-16	***	1.590
節中語数	0.012	0.002	4.768	0.000	***	1.012
節頭接続詞						
なし	0					
あり	-0.084	0.053	-1.587	0.113		0.919

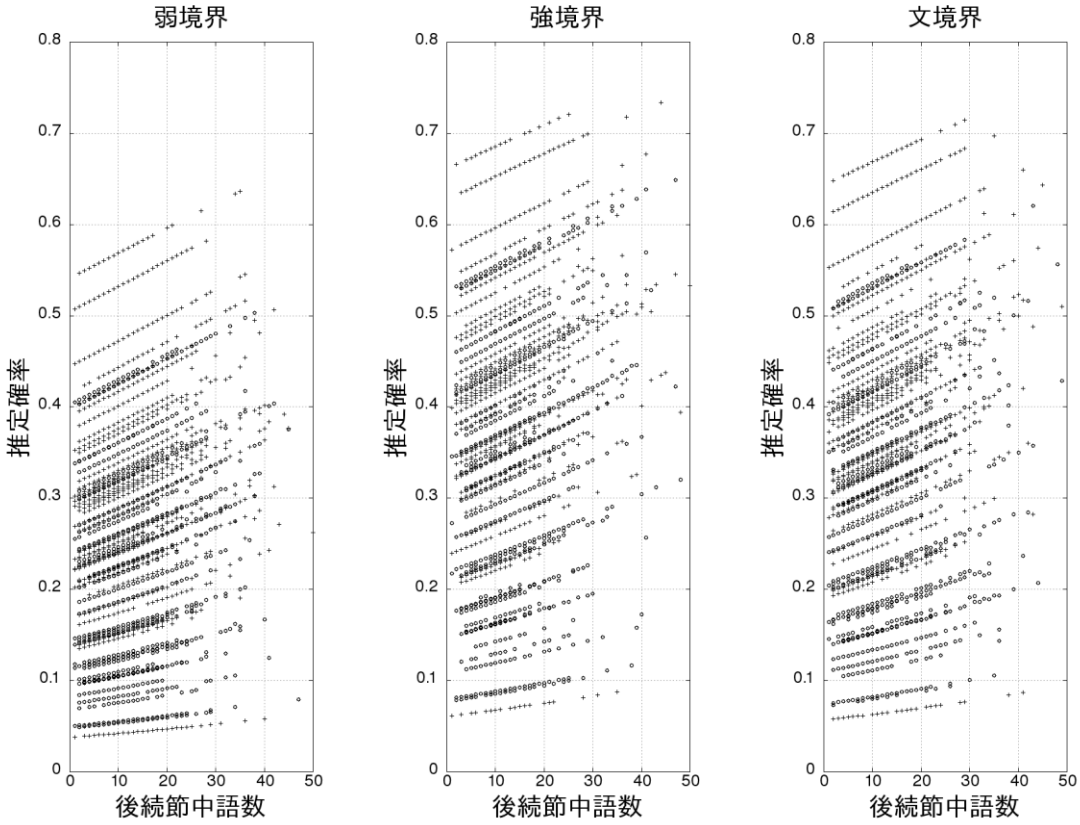

図 1 節頭におけるフィラーの出現確率推定値を示す回帰曲線（境界タイプ別）

果を表 5 に示す。分析対象とした 5 要因全ての効果が有意だった。フィラー全体の傾向同様，エーの使用確率は，女性話者よりも男性話者の方が高かった。節中語数の増加に伴いエーの使用確率が上昇する点もフィラー全体の傾向と一致していた。直前の境界の種類の影響も有意だったが，効果の大きさの順序がフィラー全体の傾向とは異なっていた。効果の大きさは，フィラー全体では，弱境界＜文境界＜強境界 であったのに対し，エーに関しては，弱境界＜強境界＜文境界 であった。つまり，エーの出現確率は境界の深さに対応しており，深い境界ほど高かった。エーは他のフィラーよりも深い談話境界で用いられる傾向のあることが示唆された。

表 4 節頭におけるフィラーの出現確率を推定するロジスティクス回帰分析結果

学歴と節頭接続詞の有無の効果はフィラー全体では現れなかったが，エーの使用確率に関しては有意だった。即ち，大卒の話者の方が中学または高校卒の話者よりも節頭のエーの使用確率は高かった。また，節頭接続詞の使用との間に関係があり，接続詞があるとエーの使用確率は下がる傾向があった。この結果は，「で」などの節頭接続詞とエーは相補的な役割を果たしている可能性のあることを示唆している。即ち，接続詞がフィラー同様，時間稼ぎの働きをしている可能性がある。

表 5 節頭におけるエーの出現確率を推定するロジスティクス回帰分析結果

Variable	Estimate	Std. Error	z value	Pr(>|z|)	オッズ比
(Intercept)	-4.224	0.332	-12.72	< 2e-16 ***	0.015
性別					
女性	0				
男性	1.073	0.327	3.279	0.001 **	2.923
学歴					
中学・高校卒	0				
大学学部卒	0.866	0.359	2.413	0.016 *	2.378
修士以上	0.692	0.639	1.084	0.278	1.998
境界の種類					
弱境界	0				
強境界	0.265	0.070	3.806	0.000 ***	1.304
文境界	0.357	0.071	5.017	0.000 ***	1.429
節中語数	0.013	0.004	3.526	0.000 ***	1.013
節頭接続詞					
なし	0				
あり	-0.207	0.084	-2.465	0.014 *	0.813

アノ

全ての変数を投入するとモデルが収束しないため，予備的分析から，影響が小さいと思われる，学歴と節頭接続詞の有無の 2 要因を除き，残り 5 要因を説明変数として分析した。分析結果を表 6 に示す。フィラー全体で見られた性別の効果が，節頭のアノに関しては有意ではなかった。フィラー全体で見ると，男性話者の方がその使用頻度は高い。しかし，ことアノに関しては，女性話者も男性話者に劣らず頻繁に使用していることになる。一方，フィラー全体では観察されなかった，話者の年齢と講演経験の効果が有意だった。年齢が上がるほど，節頭のアノの使用確率も上昇する傾向があった。一方，講演経験が増えると，節頭でのアノの使用確率は低下した。アノは，話者特性の影響を受けやすいフィラーと考えられる。境界の種類の効果はフィラー全体の傾向と一致しており，弱境界＜文境界＜強境界 の順であった。節中語数の効果もフィラー全体の傾向同様有意で，語数が多いほどアノの出現確率は上昇した。

表 6 節頭におけるアノの出現確率を推定するロジスティクス回帰分析結果

Variable	Estimate	Std. Error	z value	Pr(>|z|)	オッズ比
(Intercept)	-3.995	0.527	-7.574	0.000 ***	0.018
性別					
女性	0				
男性	-0.372	0.343	-1.086	0.277	0.689
年齢	0.024	0.012	2.033	0.042 *	1.025
講演経験	-0.078	0.032	-2.435	0.015 *	0.925
境界の種類					
弱境界	0				
強境界	0.355	0.072	4.93	0.000 ***	1.426
文境界	0.179	0.077	2.336	0.020 *	1.196
節中語数	0.011	0.004	2.462	0.014 *	1.011

マー

7 要因を説明変数とした，節頭のマーのロジスティック回帰分析結果を表 7 に示す。最も顕著な特徴は，フィラー全体で観察された節中語数の効果がマーでは観察されなかったことである。語数が，概念を言語化するプロセスの認知的負荷を反映するものと考えると，マーの使用は言語化の負荷とは別の要因に起因していると考えられ，他のフィラーとは働きの異なることが推測される。

一方，フィラー全体では観察されなかった節頭接続詞の効果が有意だった。エーでも接続詞の効果は有意だったが，エーでは接続詞があると使用確率は低下するのに対し，マーでは上昇した。

表 7 節頭におけるマーの出現確率を推定するロジスティクス回帰分析結果

Variable	Estimate	Std. Error	z value	Pr(>|z|)	オッズ比
(Intercept)	-3.856	0.483	-7.983	0.000 ***	0.021
性別					
女性	0				
男性	1.511	0.274	5.515	0.000 ***	4.532
年齢	0.004	0.010	0.411	0.681	1.004
学歴					
中学・高校卒	0				
大学学部卒	-0.143	0.293	-0.489	0.625	0.867
修士以上	-0.044	0.784	-0.056	0.955	0.957
講演経験	-0.026	0.024	-1.11	0.267	0.974
境界の種類					
弱境界	0				
強境界	0.567	0.066	8.645	< 2e-16 ***	1.763
文境界	0.458	0.074	6.223	0.000 ***	1.581
節中語数	-0.003	0.004	-0.849	0.396	0.997
節頭接続詞					
なし	0				
あり	0.154	0.074	2.092	0.036 *	1.167

他の効果はフィラー全体の傾向と一致していた。性別に関しては，女性話者よりも男性話者の方が使用確率は有意に高かった。オッズ比から，性別の効果は他のフィラーに比べて大きいことがわかる。マーは男性に好まれるフィラーと言えそうである。境界の種類の効果もフィラー全体の傾向と同じで，弱境界＜文境界＜強境界 の順であった。話者の年齢，学歴，講演経験の効果は有意ではなかった。

４．考察

①話者の性別，②年齢，③学歴，④講演経験，⑤直前の境界の種類，⑥節中語数，⑦節頭の接続詞の有無，の 7 要因を説明変数として，節頭フィラーの出現確率をロジスティック回帰分析混合モデルによって推定した。これら 7 要因のうち，①～④は話者属性に関するもので，⑤～⑦は産出された言語の特性に関するものである。話者属性に関しては，フィラー全体では性別の効果のみが有意だった。フィラーを種類別に見ても，アノ以外は男性話者の使用確率が高かった。CSJ と類似した英語コーパスを用いた研究では，節頭フィラーの使用確率に有意な男女差はなかった（Watanabe & Korematsu 2017）。日本語における男女差の原因が生得的なものなのか，文化的なものなのかの追及は今後の課題である。話者の年齢の効果はアノ以外なかった。しかし，表 1 からわかるように，今回用いたコーパスでは 50 代以上の話者の人数が，107 名中 20 名と少ない。さらに，70 代以上の話者は皆無である。年齢層を広げ，年代のバランスを取ることによって，アノ以外にも年齢の効果が表れる可能性はある。学歴の効果はエーにおいてのみ有意で，大卒の話者の使用確率が中学，高校卒の話者よりも高かった。フィラー全体で見ると，修士以上の学歴の話者の使用確率が中学，高校卒の話者よりも高い傾向があった。学歴の高い話者の方が節頭のフィラーの使用確率が高いのかどうか，今後，節中のフィラーの使用確率も考慮して分析を進める必要がある。講演経験はアノにおいてのみ有意で，経験が増えると節頭のアノは減少する傾向があった。文頭，節頭でのアノの多用は，講演に不慣れな印象を聞き手に与えるかもしれない。

講演の言語特性に関する要因では，20 代～30 代前半の話者を対象とした渡辺・外山（2017）同様，境界の種類の効果と節中語数の効果が有意だった。ただし，境界での出現確率は，渡辺・外山では 文境界＜弱境界＜強境界 の順だったのに対し，本研究では 弱境界＜文境界＜強境界 の順だった。この結果の違いは節頭の接続詞の扱いの違いに起因していると考えられる。渡辺・外山では「で，エー」のようにフィラーの前に接続詞がある場合，そのフィラーは節頭のフィラーとは見なさなかったのに対し，本研究では接続詞は節を繋ぐものと考え，このようなケースのフィラーを節頭のフィラーと見なした。文頭には接続詞が用いられることが多いため，文頭のフィラーの確率が上昇し，このような結果になったものと考えられる。このような結果の違いはあるものの，どちらの研究においてもフィラーの出現確率が 文境界＜強境界 である点には変わりがない。文境界は節境界よりも深い境界であるため，続く発話内容生成のための認知的的負荷が大きく，フィラーの出現確率は文境界の方が高いと予測していたが，そのような結果にはならなかった。文境界では節境界よりも一般に長いポーズが置かれる。そのポーズ中に，後続発話内容の生成が可能なため，フィラーの出現が抑えられている可能性がある。

節中の語数はその節によって伝えられるメッセージの豊かさの指標と考えることができる。そして，語数が多いほど内容が複雑で，適切な言語形式の選択には時間がかかると考えられる。節中語数の増加に伴い節頭フィラーの出現確率も増加しており，フィラーが発

話生成のための認知的負荷を反映しているという仮説を支持している。節中語数要因の効果に関しておもしろいのはマーの結果である。マーだけが，この要因の効果が有意ではなかった。つまり，マーの使用には，適切な言語形式選択のための時間確保とは別の要因が働いていることになる。節頭に接続詞があるとマーの出現確率は上昇した。接続詞は談話の切れ目に，副詞のマーは談話を区切ってまとめるような場合によく用いられる。フィラーのマーも，談話標識としてのマーの性質を受け継いでいるのかもしれない。

節頭接続詞の有無要因に関しては，エーとマーにおいて有意な効果があり，接続詞があるとマーの出現確率は上昇し，エーの確率は低下した。節頭接続詞としては「で」の頻度が高く，「で」は「でー」と引き伸ばされることがよくある。「で」はその直後に声門閉鎖が入って母音が続くと「で，エー」となり，「でー」の母音部分とフィラー「エー」との違いは紙一重である。接続詞があるとエーが減少する理由の一つとして，母音の引き伸ばしがエーの代わりをしている可能性が考えられる。

本研究では，節頭におけるフィラー使用全般の特徴だけでなく，頻度の高い３つのフィラータイプの特徴にも光を当てた。今後，講演のサンプル数を増やし，本研究で明らかになった傾向がどこまで一般化できるのかを調べる予定である。

謝　辞

本研究は科研費基盤(C)「後続要素の複雑さが言い淀みの発生に及ぼす影響についての日英語対照研究」（2015～18 年度，課題番号 15K02553）および「日本語と英語のパラレルコーパスを用いた言い淀みの対照言語学的研究」（2018～20 年度，課題番号 18K00559）の助成を受けて行われた。

文　献

Goldman-Eisler, F. (1968) *Psycholinguistics*, London: Academic Press.

国立国語研究所（2006).『日本語話し言葉コーパスの構築法』東京：国立国語研究所. URL: http://pj.ninjal.ac.jp/corpus_center/csj/

Levelt, W., J. M. (1989) *Speaking: From intention to articulation*. The MIT Press: Cambridge, Massachusetts.

Swerts, Marc (1998) “Filled pauses as markers of discourse structure”, *Journal of Pragmatics* 30, 485-496.

Shriberg, Elizabeth (1994) *Preliminaries to a Theory of Speech Disfluencie*, dissertation submitted to UC Berkeley.

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.443.7755&rep=rep1&type=pdf, 2016 年 7 月 14 日アクセス

Watanabe, M. & Korematsu, Y. (2017) “Factors affecting clause-initial filler probability in an English monologue corpus”, *Journal of the Phonetic Society of Japan*, Vol. 21 No. 3, pp.24–32.

渡辺美知子・外山翔平 (2017).「『日本語話し言葉コーパス』と対照可能にデザインされた英語話し言葉コーパスにおけるフィラーの分布の特徴」『国立国語研究所論集』12 号, pp.181-203.