

令和元年6月18日現在

機関番号：62618

研究種目：基盤研究(C)（一般）

研究期間：2015～2018

課題番号：15K02553

研究課題名（和文）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

研究課題名（英文）A contrastive study of speech disfluencies: How does the complexity of the following constituents affect the occurrence of filled pauses?

研究代表者

渡辺 美知子 (Watanabe, Michiko)

大学共同利用機関法人人間文化研究機構国立国語研究所・コーパス開発センター・プロジェクト非常勤研究員

研究者番号：60470027

交付決定額（研究期間全体）：（直接経費） 3,300,000円

研究成果の概要（和文）：日常的発話で頻繁に観察される、「アノー」や「エート」などのフィラーの生起頻度や出現環境を、発話設定の類似した、日本語と英語の話し言葉コーパスを用いて調べた。そして、フィラーの使用に関連していると考えられる要因を両言語間で比較した。どちらの言語においても、直前の境界の深さ、直後の節の長さが関係していることが明らかになった。一方、日本語では性別の影響が観察されたが、英語では観察されなかった。フィラーを引き起こす要因には、言語によって異なる側面のあることが示された。

研究成果の学術的意義や社会的意義

日常発話に頻出するフィラーの頻度とそれを引き起こす要因を、類似した日英語コーパスを用いて調べることによって、両言語間の発話生成プロセスの異同に光を当てた点に本研究の特色がある。境界の深さや後続節長とフィラーの出現確率との関連は英語の方で大きかった。次の発話を考えているときにフィラーが発せられることを考えると、英語話者の方が、節の開始前に、節レベルで次の発話内容を考える傾向の強いことが本研究によって示唆された。

研究成果の概要（英文）：This study investigated factors related to the occurrence of fillers in English and Japanese, using parallel spontaneous speech corpora. The results show that the clause-initial filler probability is related to the immediately preceding boundary depth and the immediately following clause length in both languages. However, a gender effect is observed only in Japanese. The results indicate that factors triggering fillers partly depend on the language.

研究分野：言語学

キーワード：フィラー 自発発話コーパス 日英語対照 言い淀み 独話

1. 研究開始当初の背景

言い淀み (disfluency, 非流暢性とも呼ばれる) とは、フィラー、くり返し、音節の引き伸ばしなどのように、自発発話に頻繁に観察される、一見、余剰な現象のことである。これらは、言語学の分野では、長年、研究に値するものとは考えられていなかった。しかし、近年、言い淀みは自発発話に内在的な現象で、発話生成プロセスを知る手掛かりとなるもの、さらには、円滑なコミュニケーションに必要な不可欠なものという認識が広がりつつある。

2. 研究の目的

本研究では、既存の日本語話し言葉コーパスと、それに類似した英語コーパスを用いて、言い淀みの日英語対照研究を行う。特に、後続要素の複雑さと、言い淀みの出現確率との関係を調べることによって、統語的複雑さが発話生成上の認知的負荷にどのように影響しているかを両言語で調べる。これによって、発話生成モデルの精緻化を目指す。本研究は日英語の話し言葉の特徴の一端を実証的に明らかにしようとするものであり、外国語としての日本語教育や英語教育にも貢献する。

3. 研究の方法

- (1) 対照研究を可能にするために、既存の日本語コーパスと類似した設定で収録された英語のコーパスを構築する。
- (2) それら 2 つのコーパスを用いて、言い淀み、特に、フィラーの頻度や持続時間を比較する。
- (3) 頻度の高い、節頭のフィラーの出現確率に影響すると思われる要因を調べる。
- (4) 影響する要因やその大きさを 2 言語間で比較することによって、発話生成プロセスの言語普遍性と固有性について検討する。

4. 研究成果

(1) “The Corpus of Oral Presentations in English (COPE)” の構築

『日本語話し言葉コーパス (CSJ)』中の模擬講演と比較可能な英語コーパス COPE を構築することが本研究の課題の 1 つであった。プロジェクト開始時点で、音声収録、書き起こし、言い淀み情報の付与、定型動詞節境界の付与は行われていたが、本プロジェクトにおいて以下の言語情報を追加した。これによって、CSJ とのより詳細な比較が可能となった。

- ① フィラーとポーズ (50ms 以上の無音区間) の開始、終了時間
- ② 単語境界の時間情報
- ③ Basic clause 境界ラベル
- ④ より客観的な文境界ラベル
- ⑤ 語の品詞情報

例 1 に、節境界情報、言い淀み情報を付与した COPE の書き起こしテキストの一部を示す。/cb/, /ce/ は 等位節の開始と終了、<cb>, <ce> は従属副詞節の開始と終了、(ncb), (nce) は名詞節の開始と終了を意味する。タグ内に接続詞も含めてある。(F uh) はフィラー “uh”, (F um) はフィラー “um” を示す。フィラーはこの 2 種類に限った。

例 1: COPE の抜粋

/cb1 so/ so definitely (F uh) one of my most memorable moments in life was (F um) (ncb when) when me and my family went on our panama canal cruise (nce when) /ce1 so/ /cb1/ (F um) my grandpa had wanted (ncb to) to do this for a really long time (nce to) /ce1/ <cb and so when> and so (F uh) finally when all of us kids were at a good age <ce and so when> /cb1/ (F um) we decided (ncb to) to finally go (nce to) /ce1/ /cb1 so/ so it was really cool /ce1 so/

(2) フィラーとポーズの持続時間と頻度

COPE と同世代の話者 (20 代~30 代前半) によるスピーチ 20 を CSJ から抽出し、フィラーとポーズの持続時間が講演全長に占める割合を比較した。図 1 に結果を示す。フィラーの割合に有意差はなかったが、ポーズ (50ms 以上の無音区間) の割合の差は有意だった。日本語話者の方が間の多いスピーチをしていることが明らかになった。また、100 語あたりのフィラーの頻度は、日本語 6 回、英語 4 回で、日本語の方が有意に高かった (渡辺・外山, 2017)。

(3) 直前の境界の深さ、後続節の長さ、話者の性別が節頭フィラーの出現確率に及ぼす影響

日英語、各 20 講演を対象に、これらの要因が節頭フィラーの出現確率に及ぼす影響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。フィラーの使用頻度には個人差が大きいため、話者を変量効果とした。英語の節境界の深さの認定は、3 名のラベラーに文境界 (談話の切れ目) をマークさせ、マークした人数によって節境界の深さを 4 段階に定めた。日本語の節境界の深さの認定は、CSJ 中の節境界ラベルを利用し、弱境界 (浅い節境界)、強境界 (深い節境界)、絶対境界 (文境界) の 3 段階とした。節の長さは、フィラーや他の言い淀みを含まない語数を指標とした。英語については、節中の引用節数 (たとえば、She said, “it doesn’t

count.”の””内の節数)も要因に加えた。

分析の結果、どちらの言語においても境界要因と節長要因は有意だった。英語では引用節数要因も負の方向に有意だった。性別要因は日本語においてのみ有意で、男性話者の方がフィラーの出現確率は高かった。次に、データの9割を用いてモデルを作成し、そのモデルで残り1割の節の節頭フィラーの出現確率を予測した。図2に日本語の予測結果、図3に英語の予測結果を境界のタイプ別に示す。●は節頭にフィラーのあったケース、+はなかったケースである。

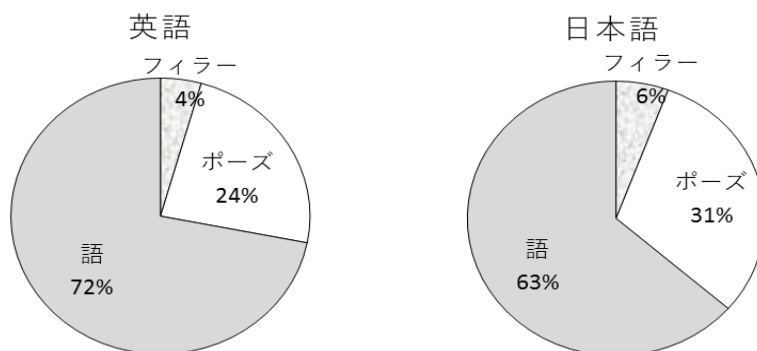


図1 フィラーとポーズの持続時間が講演全長に占める割合

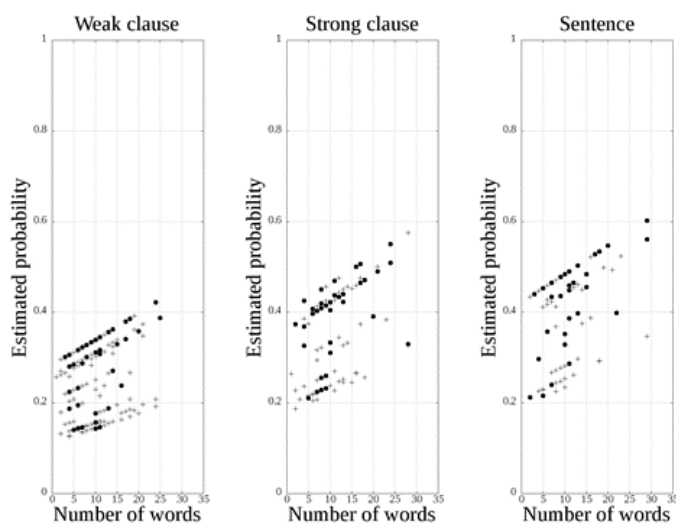


図2 日本語のフィラーの出現確率推定値を示す回帰曲線

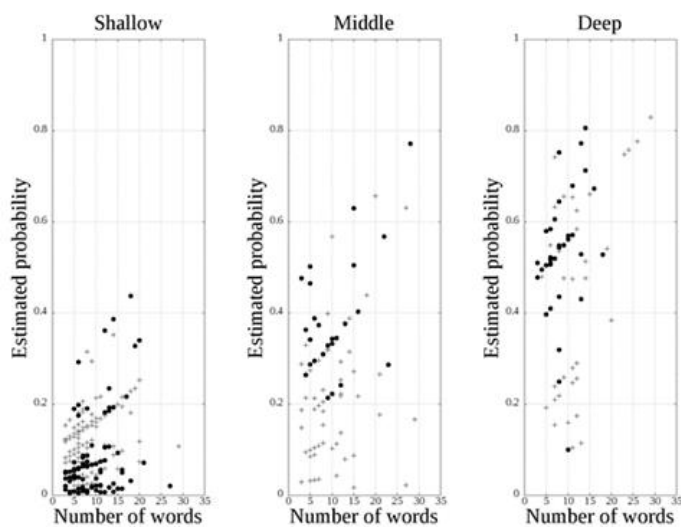


図3 英語のフィラーの出現確率推定値を示す回帰曲線

これらの結果より、節中語数の影響は英語の方で大きいことが推測される。フィラーが、発話生成に時間のかかるときに発せられるものであることを考えると、後続節長との関係が強い英語の方が日本語よりも、節の発話開始前に節レベルで発話内容を考えている傾向の強いことが示唆された。(Watanabe & Korematsu, 2019)

5. 主な発表論文等

[雑誌論文] (計 2 件)

- ① Watanabe, M. & Korematsu, Y. 2017.12 “Factors affecting clause-initial filler probability in an English monologue corpus”, *Journal of the Phonetic Society of Japan*, Vol. 21 No. 3, 24–32. 査読有り. https://doi.org/10.24467/onseikenkyu.21.3_24
- ② 渡辺美知子・外山翔平 2017年1月『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, 181-203. 査読有り. <http://doi.org/10.15084/00000860>

[学会発表] (計 7 件)

- ① Watanabe, M., Korematsu, Y. 2019. Comparison of factors related to clause-initial filler probabilities in English and Japanese, to appear in *Proc. of ICPHS*. 査読有り.
- ② 渡辺美知子・是松優作 2018.9.5『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 言語資源活用ワークショップ発表論文集/Proceedings of Language Resources Workshop, 3, 578-585. 査読有り. <http://doi.org/10.15084/00001692>
- ③ Michiko Watanabe, Shohei Toyama. 2017.02.15. “A cross-linguistic comparison of filled pause frequency at clause boundaries in English and Japanese”, *Proceedings of the International Conference “Fluency & Disfluency Across Languages and Language Varieties”*, 117-119. Louvain-la-Neuve, Belgium. 査読有り.
- ④ Michiko Watanabe, Shohei Toyama 2016.12.9. “Comparison of filled pause durations and frequencies in parallel corpus of English and Japanese”, *Programme of The 16th Australasian International Conference on Speech Science and Technology (SST2016)*, p.42. Parramatta, Australia. 査読有り.
- ⑤ 渡辺美知子 2016.12.3 「英語と日本語のインフォーマルなスピーチにおけるフィラーの頻度の比較」日本音声学会第334回研究例会シンポジウム「フィラーの音声学と言語学：日英中を対象に」
- ⑥ 渡辺美知子 2016.9.10 「日本語学習者が読み上げた文の韻律に視覚的フィードバックを返すシステム構築の試み」『2016年日本語教育国際研究大会予稿集』Bali, Indonesia. 査読有り.
- ⑦ Watanabe, M., Kashiwagi, Y., & Maekawa, K. 2015.8.9 “The relationship between preceding clause type, subsequent clause length and the duration of silent and filled pauses at clause boundaries in Japanese monologues”, *Disfluency in Spontaneous Speech*, DiSS 2015, Session5-4, Edinburgh, UK. 査読有り.

[図書] (計 0 件)

なし

[産業財産権]

該当なし

6. 研究組織

(1) 研究分担者：なし

(2) 研究協力者

研究協力者氏名：ローズ ラルフ

ローマ字氏名：Rose Ralph

所属研究機関名：早稲田大学

部局名：理工学術院

職名：准教授

研究者番号 (8桁)：30404916

※科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。