

DaSiC2023 WORKSHOP

言語データとその「鏡」： 機械学習モデルを用いた 言い誤りと失語症例の 分析

URL : <https://project-ccap.github.io/2023DaSiC/>

言語学とデータサイエンスに関する
ワークショップ・
参加無料

日時

2023年

12月23日〔土〕

13:30 ~ 17:00



登壇者

寺尾康	(静岡県立大学・司会進行)
浅川伸一	(東京女子大学)
上間清司	(武蔵野大学)
大門正太郎	(クラーク病院)
高倉祐樹	(北海道大学)
立場文音	(JCHO熊本総合病院)
橋本幸成	(目白大学)
吉原将大	(東北大学)

概要

健常者は日常の発話でついうっかり、また失語症患者は主に脳の疾患により言い誤り(錯語)を表出することが知られています。今回のイベントでは、こうした言語データを機械学習モデルと神経科学といういわば2枚の「鏡」の前に置いた時、そこに映し出されるのはどのような景色、振る舞いかを実演を交えて示します。はたしてそれは機械学習モデルの貢献か研究者の願望か。言語学者、機械学習の専門家、言語聴覚士という登壇者それぞれの3つの視座から、実際の健常者の言い誤りや失語症患者の錯語の実際のデータを供覧しつつ、それらのデータが機械学習モデルではどのように説明されるのか、から議論していきます。

場所

筑波大学
筑波キャンパス
第1エリア
1D201教室

主催

**Linguistics and Data Science
in Collaboration (DaSiC)**

小野雄一 (筑波大学)
島田雅晴 (筑波大学)
長野明子 (静岡県立大学)

WEB

プログラムと共に、近日公開の予定。お問い合わせは島田雅晴まで (以下メールアドレス)
shimada.masaharu.fu@u.tsukuba.ac.jp

今日はこんな話をします 1

～「鏡」とは？～

昨今の LLM, 生成 AI は,
ハリー・ポッターに登場するみぞの鏡 (Mirror of Erised: Desireの逆綴り) かもしれない。

「心の奥底にある, 最も切実な願望以上のものは何も示してくれない鏡」

使う人の願望によって異なる成果を産む「分身」

健常者の **言い誤り**

失語症の **言い誤り**



機械学習

神経科学

今日はこんな話をします 2



【セッション1】 13:30～14:10 趣旨説明

データの紹介1：健常者の言い間違い

データの紹介2：失語症者の錯語 失語症の概説、音韻性錯語、意味性錯語

【セッション2】 14:10-14:50 機械学習からみた言語モデルの鏡

【セッション3】 15:00-16:15 言語認知モデルの鏡と機械学習の鏡との接点

①認知モデルの説明1：健常者

②認知モデルの説明2：失語症者

③機械学習モデルの説明(ファインチューニング)

【セッション4】 16:25～17:40 実演と全体討論

モデルのデモンストレーション ①百人一首 ②語彙の産出

全体討論

【セッション1】

DaSiC2023 WORKSHOP

言語データとその「鏡」：
機械学習モデルを用いた
言い誤りと失語症例の
分析

言語学とデータサイエンスに関する
ワークショップ・
参加無料

日時

2023年
12月23日〔土〕
13:30 ~ 17:00

データの紹介1：健常者の言い間違い

寺尾 康（静岡県立大学）

データの紹介2：失語症者の錯語

高倉祐樹（北海道大学）
立場文音（JCHO熊本総合病院）
大門正太郎（クラーク病院）

言い誤りって何？ 1

健常な言語能力/運用力を持つ話者は母語をスラスラ話します。

- 意図した通りのことを
- 意識のコントロールなく、オートマティックに
言えている成果

時に、そこからズレた発話をしてしまうことがあります。

ちょっと分類してみましょう。

言い誤りって何？ 2

発話の分類

意識による コントロールなし (オートマティック)		意識による コントロールあり
思った通り	母語スラスラ (規範からの逸脱を含む)	外国語カタコト
思った通り ではない	言い誤り	自己訂正

言い誤りって何？ 3 簡単な定義

「 **A** と言おうとして思わず **B** と言ってしまった」

▼意図的なウケ狙いではなく、読み間違いでも障害でもない

▼もし、そこに規則性があれば…

それは意識の及ばないレベルでのことばの
部品やその組み立て作業を反映している

発話研究の重要なデータとみなされてきた

言い誤りって何？ 4

実例と「基本要素」

あらゆる言語レベルに様々なタイプの誤りが生じる

「打点さらの原」(打点3の原)

daten san no hara

- ターゲットとエラー、そのソース
- レベル
- タイプ
- 距離と方向
- その他

言い誤りの実例

ジャカンカップ (ジャパンカップ)
音韻レベル 代用 予測

時間で言うと、深夜に時間するのは (電話)
語彙レベル 代用 保続

ながせばはない (話せば長い)
形態素レベル 交換

いどろり (いろどり)
音韻レベル 交換

ばまき (パジャマ + ねまき)
語彙レベル 混成

青木となかじ の組 (なかじま)
音韻レベル 欠落

言い誤りの全体的傾向

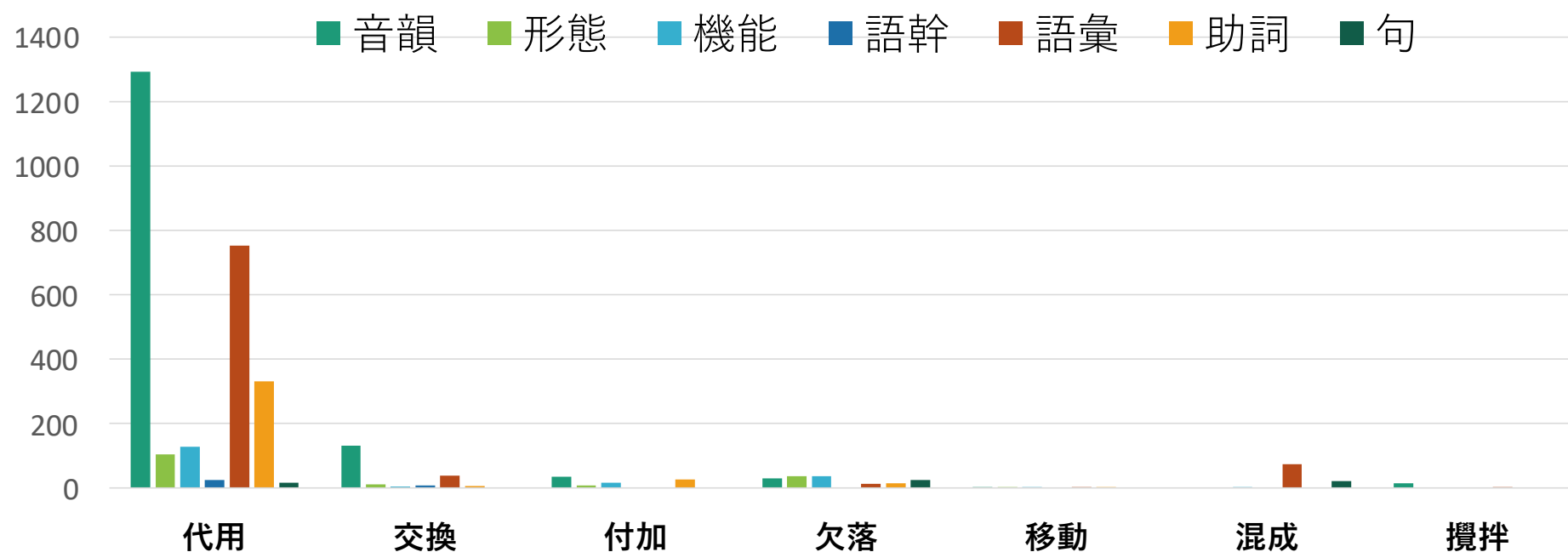


図1. uSedにみる日本語言い誤りのレベルとタイプの分布

本日注目する言い誤り 語彙代用

- a. エジプトにのぼった（ピラミッド）
- b. かたことの言葉（カタカナ）
- c. それってサーフィン？（サイフォン）
- d. こんな日は風がほしい（車）＜風が強い野外で＞
- e. 麻雀飲んだり、酒飲んだり（麻雀したり）
- f. 常磐線の中で電車吸ってる人がいた（タバコ）

語彙代用が教えてくれるもの

「心の辞書」の引き間違い…

- ・ 意味的要因
- ・ 文法的要因（品詞など）
- ・ 音韻的要因
- ・ 文脈的要因

どう相互に関係する？

語彙代用 の「たて」と「よこ」

「たて」

その語彙スロットに収まることのできる

・
・
語同士の競争が起こした代用

[ジュース]

パンと コーヒー で朝食を

[スープ]

[紅茶]

・

・

語彙代用 の「たて」と「よこ」

「よこ」

ターゲット語彙の周囲にある語彙、言語情報が

直接・間接に邪魔をして起きた代用

直接（予測/保続）

ロードレースはマラソンで言えば、陸上で言えばマラソン…

今日は雲が多くて雲がふる（雪）

間接（意味文脈型）

せんざいのせっけんであります（せんでん）

すごく気に入っている好きだ（曲）

「たて」と「よこ」はどこで？ 発話の基本図

言語産出の認知モデルでは

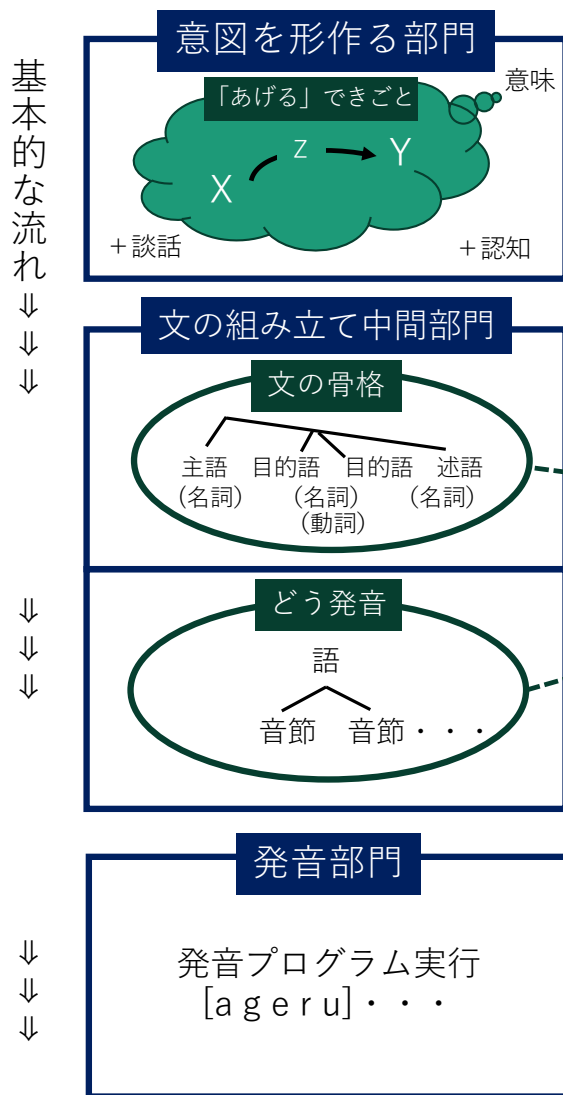


図 発話の主要部門
(Levelt(1989)に基づく)

「たて」と「よこ」はどこで？ シミュレーションでは・・・

「**たて**」型も「**よこ**」型も同じように起こるのか？
「交通整理のお巡りさん」は必要か？

両タイプのシミュレーションにおいて、
機械学習モデルはどのようにふるまうのか？

特に、間接的な「**よこ**」型をうまく再現することはできるのか？

おまけ 言い誤り雑感

- 言い誤りをしない人はいません
- 意図した形と誤った形が同時に手に入る現実的なデータ
- 仮説の検証という方向もありますが、データから何が言えるのか、
うアプローチが可能です。

とい

- 様々な意味で「許されるもの」です。
 - 誤りが起こっても話は通じます、そして多くは笑いに通じます。
 - 機械にこれができるでしょうか？
- こういう時代ですから
 - AIに「食わせるエサ」として、あるいはAIの「審査員」として

おまけ 語彙代用の誤りの量的分析

表. 語彙代用のタイプ別 実例数 (N=752**)

タイプ	意味型	混合型	形式型	文脈型	意味文脈型	粹文脈型	行・イム・特殊型
頻度	321	230	59	207	46	24	14

** 重複があるため、頻度の合計数は752より多い

おまけ 語彙代用の分析

表. 語彙代用のタイプ分析 結果 (752例中の一致数)

観点	文法 範疇	文 脈 あり	意 味 関連	語頭 モーラ	語 末 モーラ	モーラ 数	アクセ ント型	音節重 さ
頻度	745	255	577	247	171	359*	489	354

* 全体から複雑語の実例数を引いた577例中の一致数を示している

おまけ 語彙代用の実例 英語

- a. He got hot under the belt. (collar)
- b. a lot of answers. (questions)
- c. magnificent dialect. (dialogue)
- d. not happy in Hawaii. (Illinois)
 <considering Hawaii>

ありがとうございました

参考文献

- Dell, G. (1988) The retrieval of phonological forms in production: Test of predictions from a connectionist model, *Journal of Memory and Language* 27, 124-142.
- Ferreira, F. (2000). TAG-based grammar. In L. R. Wheeldon. (Ed.). *Aspects of language production* (pp.291-330). Psychology Press.
- Fromkin, V.A. (1973) *Speech Errors as Linguistic Evidence*, Mouton.
- Levelt, W.J.M. (1989) *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.
- Roelofs, A. (2000). WEAVER++ and other computational models. In L.R. Wheeldon. (Ed.). *Aspects of language production* (pp. 71-114). Psychology Press.
- Stemberger, J.P. (1983). An interactive activation model of language production. A.W. Ellis (Ed.). *Progress in the Psychology of Language* (pp.143-186). Vol.1, Lawrence Erlbaum Hillsdale.
- 寺尾康 (2002) 『言い誤りはどうして起こる?』 東京：岩波書店.
- 寺尾康 (2022) 言語産出研究における基礎的データとしての言い間違いの価値, 『認知科学』, 第29巻1号, 74-84.

【セッション1】

DaSiC2023 WORKSHOP

言語データとその「鏡」：
機械学習モデルを用いた
言い誤りと失語症例の
分析

言語学とデータサイエンスに関する
ワークショップ・
参加無料

日時

2023年
12月23日〔土〕
13:30 ~ 17:00

データの紹介1：健常者の言い間違い

寺尾 康（静岡県立大学）

データの紹介2：失語症者の錯語

高倉祐樹（北海道大学）
立場文音（JCHO熊本総合病院）
大門正太郎（クラーク病院）

【セッション2】

DaSiC2023 WORKSHOP

言語データとその「鏡」：
機械学習モデルを用い
た言い誤りと失語症例の
分析

言語学とデータサイエンスに関する
ワークショップ・
参加無料

日時

2023年

12月23日〔土〕

13:30 ~ 17:00

講義 「機械学習からみた言語モデルの鏡」

浅川伸一 （東京女子大学）

【セッション3】

DaSiC2023 WORKSHOP

言語データとその「鏡」：
機械学習モデルを用いた
言い誤りと失語症例の
分析

言語学とデータサイエンスに関する
ワークショップ・
参加無料

日時

2023年

12月23日【土】

13:30 ~ 17:00

言語認知モデルの鏡 と 機械学習の鏡 と の接点

①認知モデルの説明1：健常者

寺尾 康 (静岡県立大学)

②認知モデルの説明2：失語症者

上間清司 (武蔵野大学)

橋本幸成 (目白大学)

③機械学習モデルの説明

浅川伸一 (東京女子大学)

【セッション4】

DaSiC2023 WORKSHOP

言語データとその「鏡」：
機械学習モデルを用いた
言い誤りと失語症例の
分析

言語学とデータサイエンスに関する
ワークショップ・
参加無料

日時

2023年
12月23日〔土〕
13:30 ~ 17:00

モデルのデモンストレーション

①百人一首 ②語彙の産出

全体討論

浅川伸一 (東京女子大学)
吉原将大 (東北大学)