

DaSiC2023 WORKSHOP

言語学とデータサイエンスに関する  
ワークショップ

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

# 今日はこんな話をします 1

まず、このワークショップ、ちょっと変わったタイトルですね。「鏡」とは次のような言葉からきています。

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

健常者の言い誤り、失語症者の症例を機械学習の「鏡」、神経科学の「鏡」に映した時に見えるものは何か、という問いからきています。

# 今日はこんな話をします 2

## 【セッション1】趣旨説明およびデータ紹介 13:30-14:10

- ・ 趣旨説明
- ・ データ紹介1: 健常者の言い誤り
- ・ データ紹介2: 失語症の錯語

## 【セッション2】機械学習からみた言語モデルの鏡 14:10-14:50

- ・ 講演

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

- ・ 認知モデルの説明1: 健常者
- ・ 認知モデルの説明2: 失語例
- ・ 機械学習モデルの説明

## ・ 【セッション4】実演と全体討論 16:25-17:40

- ・ モデルのデモンストレーション
- ・ 全体討論

# 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。それは、
  - 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と言えます。
- でも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 無し  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く
- 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と考えられる。→

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

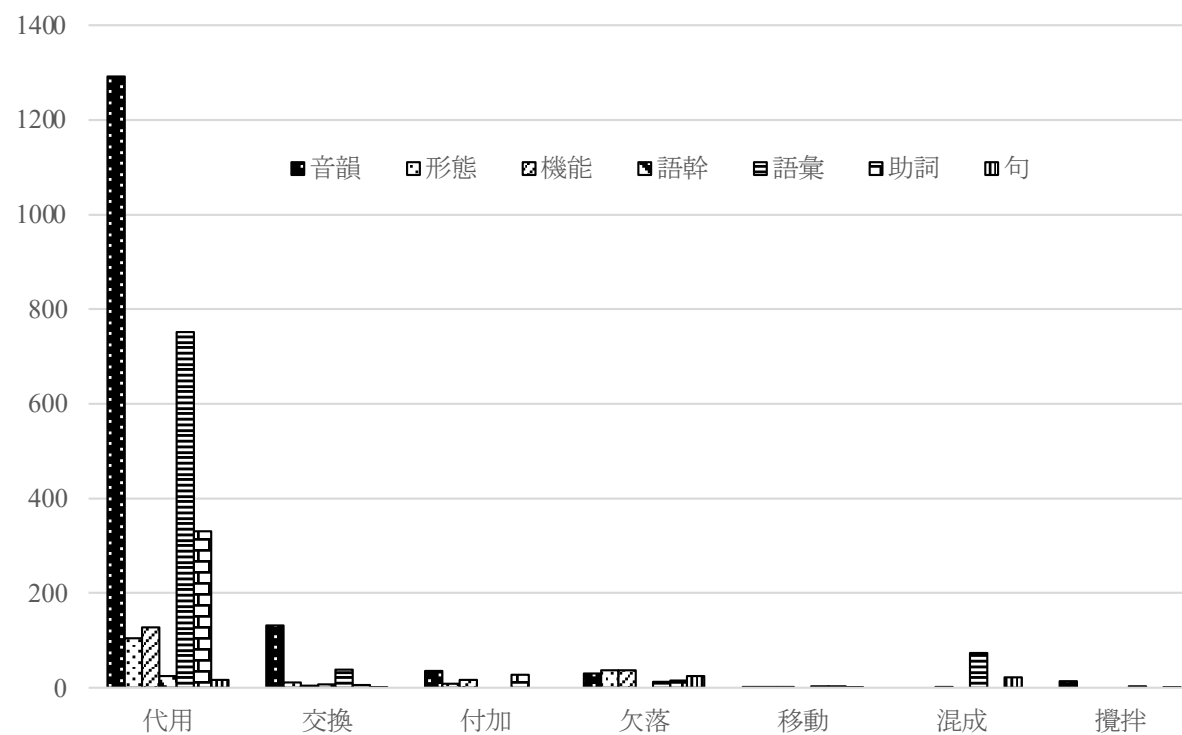


図 1. uSedにみる日本語言い誤りのレベルとタイプの分布

## 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...

そこには

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

が関わっているらしい。どう相互に関係する？

## 語彙代用の「タテ」と「ヨコ」

タテ：その語彙スロットに収まることができる語  
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「タテ」と「ヨコ」

ヨコ：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）

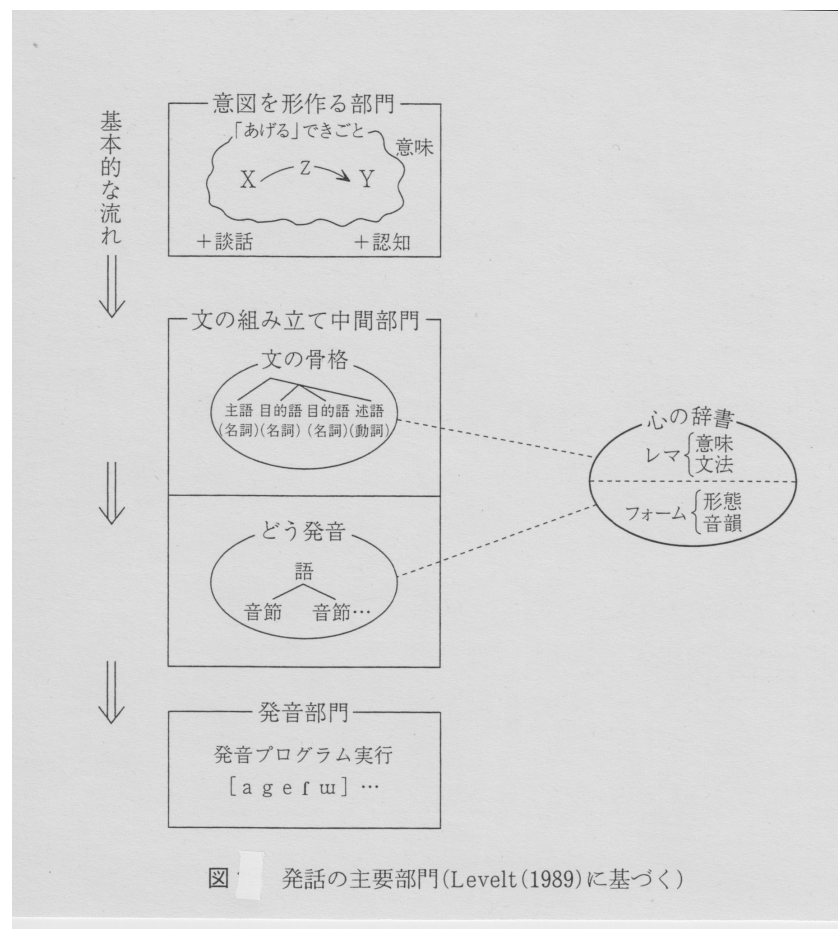
### 間接(意味文脈型)

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）

# タテとヨコはどこで？ 発話の基本図

認知モデル  
では



## 課題: タテ と ヨコはどこで? どう?

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?

タテ型もヨコ型も同じように予想できるのか?

# おまけ 語彙代用の分析

表 1. 語彙代用のタイプ分析結果 (752例中的一致数)

\*全体から複雑語の実例数を引いた577例中的一致数を示している

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



# おまけ 語彙代用のタイプ

表 2. 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	文脈・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