

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



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

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

時間と言うと、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

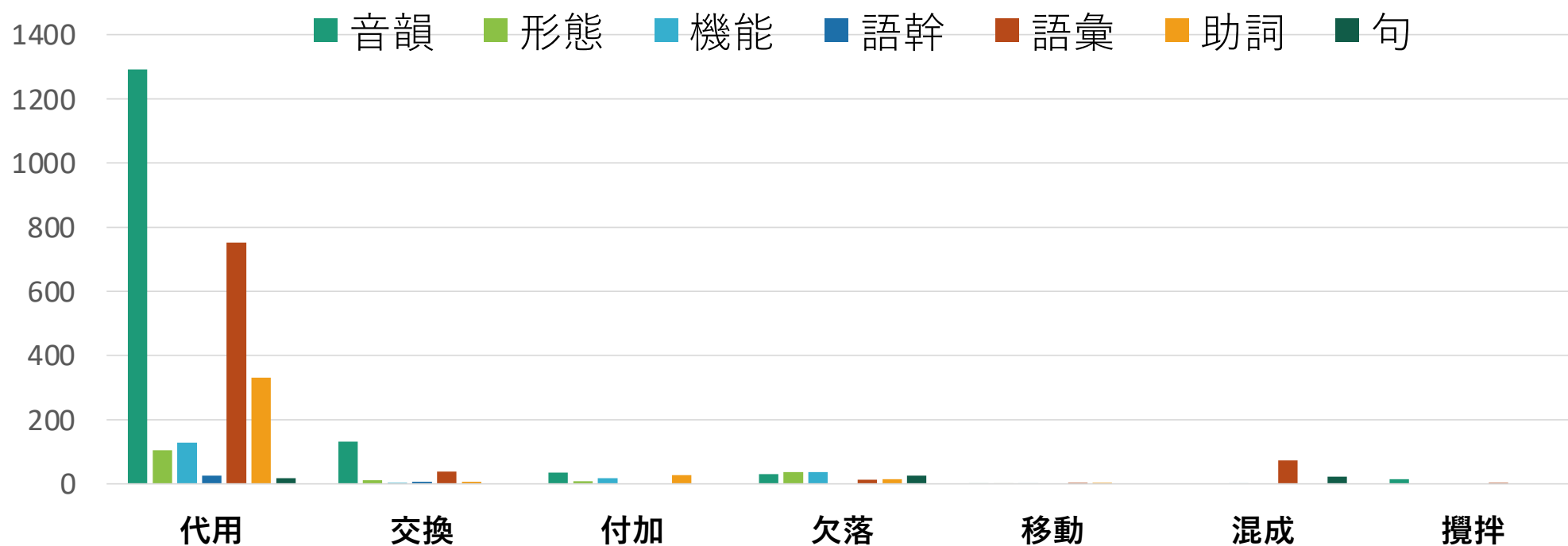


図 1. uSedにみる日本語言い誤りのレベルとタイプの分布

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）

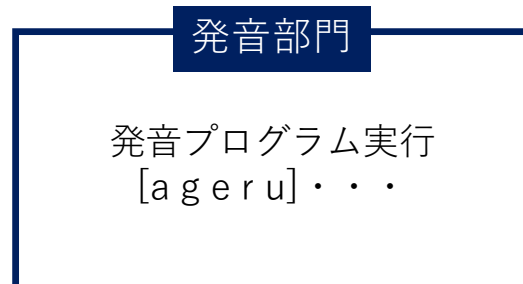
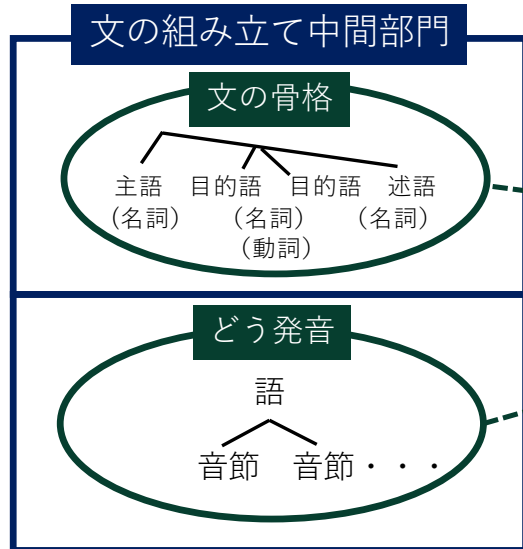
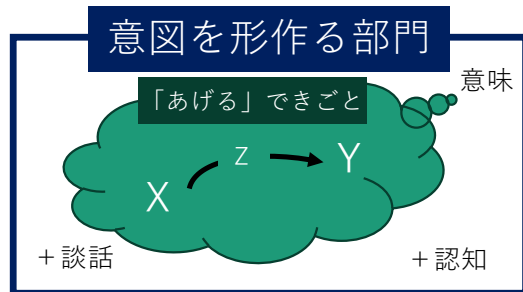
せんざいのせっけんであります （せんでん）

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

基本的な流れ
↓
↓
↓



↓
↓
↓

↓
↓
↓

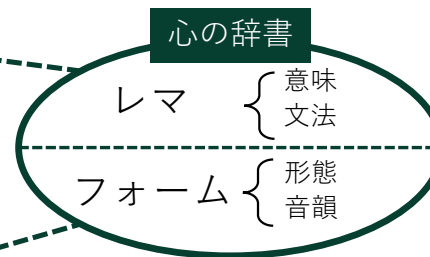


図 発話の主要部門
(Levelt(1989)に基づく)

「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	枠文脈型	ｲﾝﾃﾙ・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例中的一致数)

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例中的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>

【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）

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

言語を支える神経基盤から診る失語症

北海道大学大学院保健科学研究院
言語聴覚士 PhD 高倉祐樹

失語症とは？

大脳の損傷に由来する

→心因ではない

一旦獲得された

→発達の問題ではない

言語記号の操作能力の低下ないし消失

→難聴や発声発語器官の損傷によるものではない

(山鳥、1985)

はじめに：神経心理学的アプローチについて

「心理過程の破壊の諸相を研究することで、人間の心理現象の構造を知り、患者の治療に貢献せんとするもの」(山鳥, 1985)

- ① 脳損傷者の認知機能障害（発話を含む）を詳細に検討し
損傷部位と認知機能障害の関係を明らかにする
→ 「脳」を中心とした視点
- ② 脳損傷者の認知機能障害の現れ方から、認知機能の
構造・メカニズムを推測する方法
→ 「認知機能のメカニズム」を中心とした視点

(大槻, 2019)

二重乖離の原理について (Teuber, 1955)

病巣A

症候aあり

症候bなし

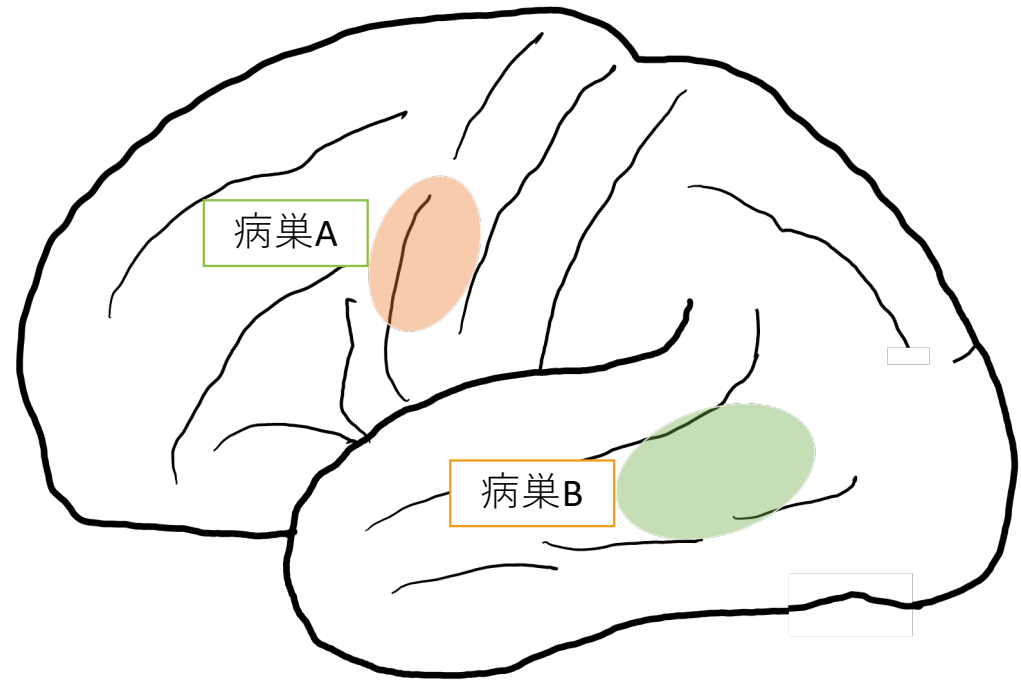
例) 病巣Aでは発話の障害が出現するが、
単語理解の障害は出現しない

病巣B

症候aなし

症候bあり

例) 病巣では発話の障害が出現しないが、
単語理解の障害は出現する



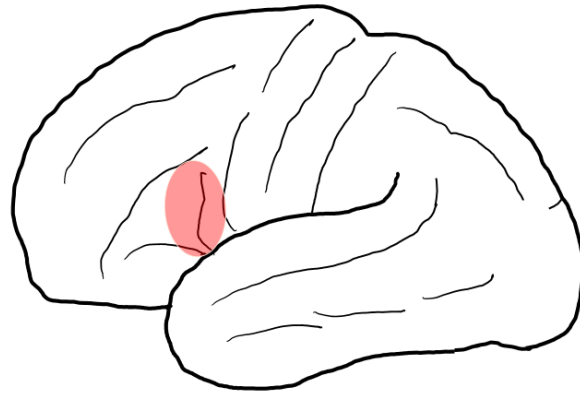
このような2つの事象が生じた場合、脳部位Aは症候aに
脳部位Bは症候bに関係するという推測が可能となる

「ことばの障害」に関する神経心理学的研究の端緒

【ブローカによる歴史的症例の報告(1861)】

「彼は、人が彼にいうことはなんでも分かった」

「彼にどんな質問を向けても答えはつねにtan, tan」



本例の障害は「構音言語の障害（**aphemie**）」と名づけられ
死後剖検の結果から「第三前頭回（ブローカ領域）」が
障害の発現に関与している脳部位であると考察

Pierre Paul Broca

Wikipediaより引用

[http://ihm.nlm.nih.gov
/images/B03578](http://ihm.nlm.nih.gov/images/B03578)

「構音言語（発話）の障害」をもたらす病巣
＝ブローカ領域と考えられていたが・・・

「ブローカ領域は、ことばの情報を音声の情報に変え、口唇、舌、声帯などの発声・発話器官につながった運動神経に必要な指令を出す役割をもつものと考えられる。すなわち、**ブローカ領域は脳における話しことばの出口**である。したがって、ブローカ領域が傷ついた場合、まるでことばを失ったように**話す機能だけが著しく低下**する」（日本音響学会編：「音のなんでも小事典」（1996）より引用）



画像診断の普及とその進歩により
今日では「発話の障害」の病巣は
「左中心前回」と訂正されている

（大槻, 2005; 松田ら, 2005; Itabashiら, 2016）

今日のブローカ領域についてのコンセンサス

ブローカ領域の限局損傷

- | | |
|--------------|----|
| ・発話の障害 | なし |
| ・ことばの思い出しにくさ | あり |
| ・文法理解の障害 | あり |

中心前回の限局損傷

- | | |
|--------------|----|
| ・発話の障害 | あり |
| ・ことばの思い出しにくさ | なし |
| ・文法理解の障害 | なし |



「発話の障害」と「ことばの思い出しにくさ」「文法理解の問題」が機能的にも解剖学的にも分離されること（二重乖離が成立すること）が示されている (相馬,2003)

話しことばに関する症候・病巣と、症候から推察される機能単位

【症候】

言語レベル

- 言葉が思い出せない
= 喚語障害
- 別の音に言い誤る
= 音韻性錯語
(例：りんご→りんも)

【病巣】

左下前頭回or
左角回or
左下側頭回後部
およびこれらの皮質下

左上側頭回～左縁上回
～左中心後回
およびこれらの皮質下



【機能単位】

意味処理
語の処理



音韻処理



発話運動
プログラミング

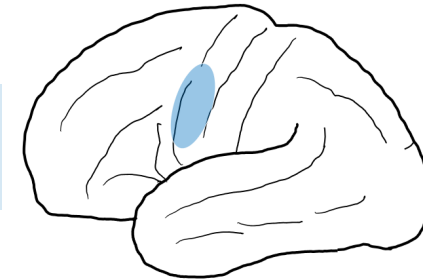


発話実現

発話レベル

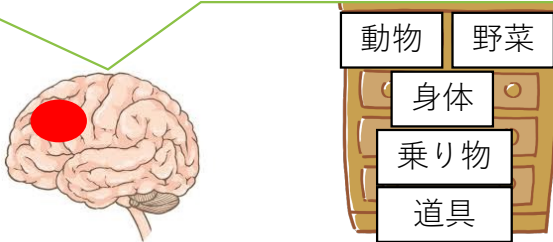
- 構音が不明瞭になる
- 音のつながりが悪くなる
- 抑揚が乏しくなる
- 不自然なアクセントに変化する
= 発語失行

左中心前回
およびその皮質下

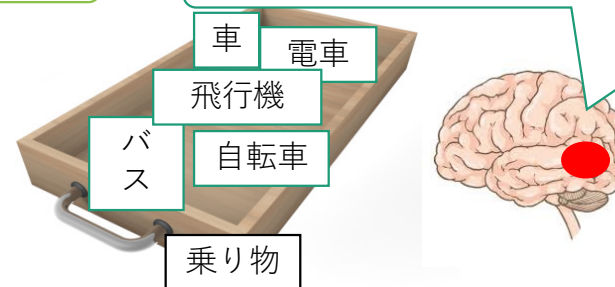


喚語のための2つの処理過程

語彙の蓄えられた引き出し（カテゴリー）を選ぶ



引き出しの中から該当の語を選ぶ



(大槻, 2007)

意味的関連なし



「ひざ・・・あ、ちゃうちゃう
ちゃう・・・さし・・・筆
差しじゃなくて・・・」



「えーと・・・・・・・・ちり
とりじゃなく、んー
と・・・えー」

意味的関連あり



「こ、あの、汽車、あの、でん
わ・・・でんわね・・・あのそ
こに・・・じんしゃ、その電
車にのって・・・電車だよね」



「刃・・・包丁・・・あの・・・
か、かわ・・・ほれ・・・あ
の・・・なわ・・・とかの
包丁切るやつなんだけ
ど・・・」

なぜ錯語分析が重要なのか？

脳内の単語処理のしくみを推測できる(大槻, 2003)

- 音の入れ替わりが生じる場合
→「音」を選択し、組み合わせる過程が存在
- 意味的に似通った語への言い誤りが出現する場合
→意味的に類似した単語の中から目標語を
選択する過程が存在

障害された単語処理のしくみを知ることは
より適切な治療的介入への「鍵」となりうる

【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）

失語症者の言い誤り（錯語）

		語			
		非実在語	実在語の合成	実在語	
音	音韻類似なし	新造語	記号素性錯語	語性錯語	
				意味的関連あり	意味的関連なし
				意味性錯語	無関連錯語
	音韻類似あり	音韻性錯語		音意味混合	
混合性錯語			形式性錯語		

呼称

復唱

音読

漫画の説明

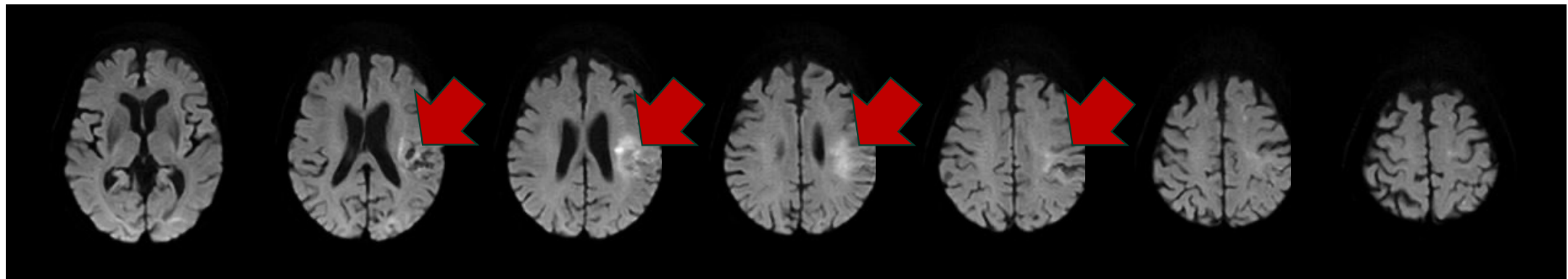
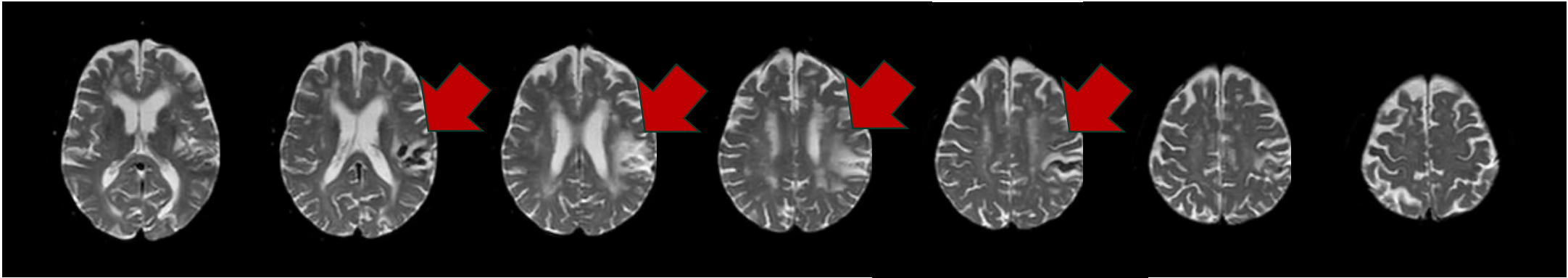
標準失語症検査（SLTA）

症例

- 80歳代右利き女性
- 主訴 話しにくい
- 現病歴

X年11月、右半身の脱力感のため救急搬送。t-PA投与し、右半身の脱力感は改善したが、失語症状は残存。

頭部MRI画像



左縁上回を中心に出血性梗塞を同定

誤りの例① 呼称

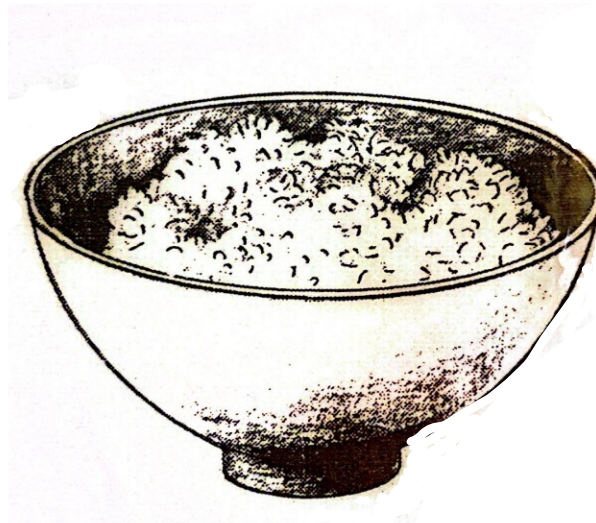
め…め、め、…ねと！ねと！あ、ねこ！



SLTA 呼称 例題：ねこ

誤りの例② 呼称

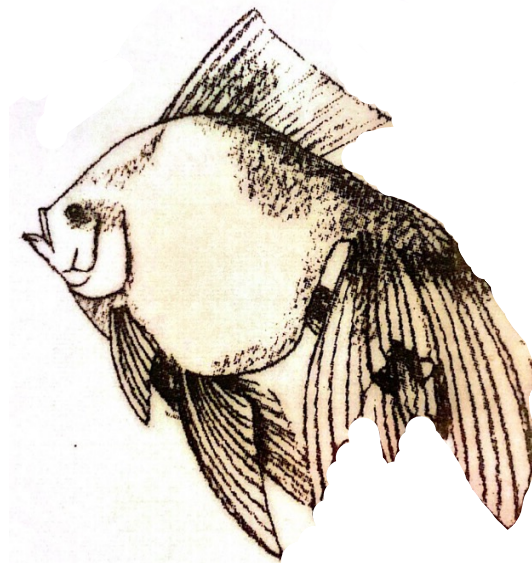
おこ、ちよわんじゃないか、…ろはん // ごはん



SLTA 呼称 刺激：ごはん

誤りの例③ 呼称

り、りんごじゃない…りん、りんぎょ // きんぎょ

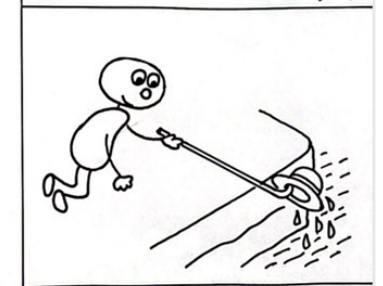
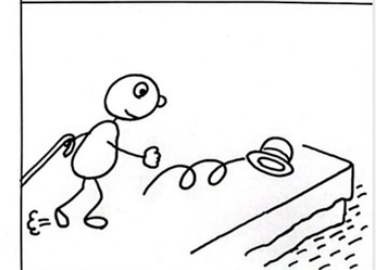
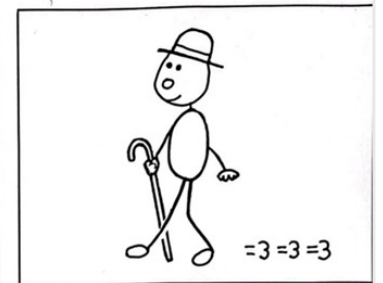
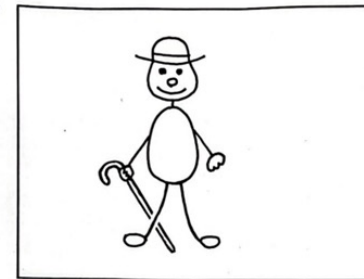


SLTA 呼称 刺激：金魚

誤りの例④ 漫画の説明

男の人が…歩いています…。ステッチ、じゃないか。
ステップ、ステップをこうしてて…

34



誤りの例⑤ 復唱

帽子 → ばーさ

靴下 → こ、きく、くぎ、くき、くぎた // くきした、くぎした

自動車 → にわ、にわしょ、じわしょ // じわーぎよ、しわーしょ

水を飲む → 水をもぐ、水をむぐ

空が青い → ささささが青い

症例のまとめ

呼称や自由会話より、復唱に強い障害を示した

音韻性錯語が主体で意味性錯語が少ない

【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）

失語症



呼称課題



「スプーン」

喚語障害

喚語困難(≡無反応)

錯語

		語			
		非実在語	実在語の合成	実在語	
音	音韻類似なし	新造語	記号索性錯語	語性錯語	
				意味的関連あり	意味的関連なし
				意味性錯語	無関連錯語
	音韻類似あり	音韻性錯語	音意味混合		
				混合性錯語	形式性錯語

「フォーク」

「スポン」

「フォーク」

「スプーン」

意味性錯語(semantic paraphasia)

「スプーン」



目標語と意味的関連のある語への誤り

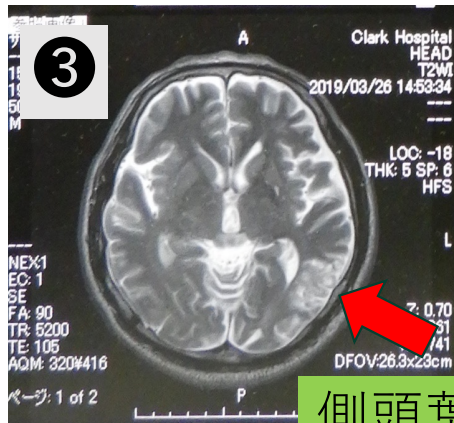
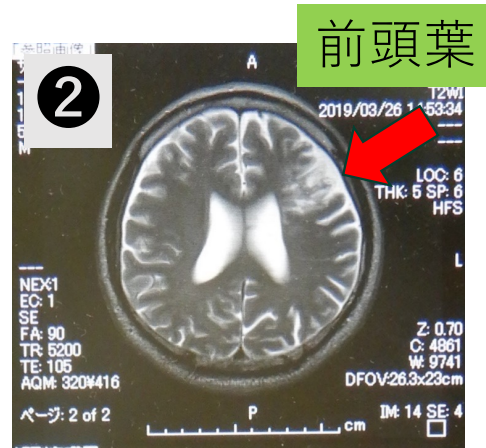
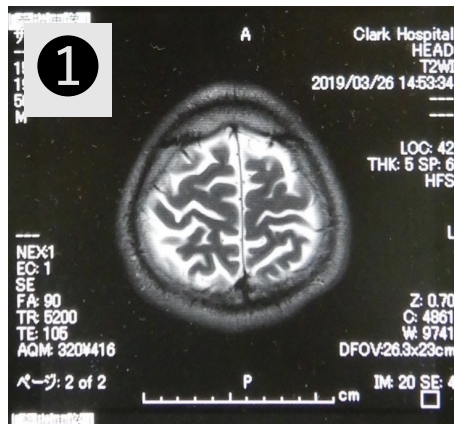
「えんぴつ」
「フォーク」
「食器」
「カレーライス」
「それで食べる」

目標語と反応の間のさまざまな意味関係

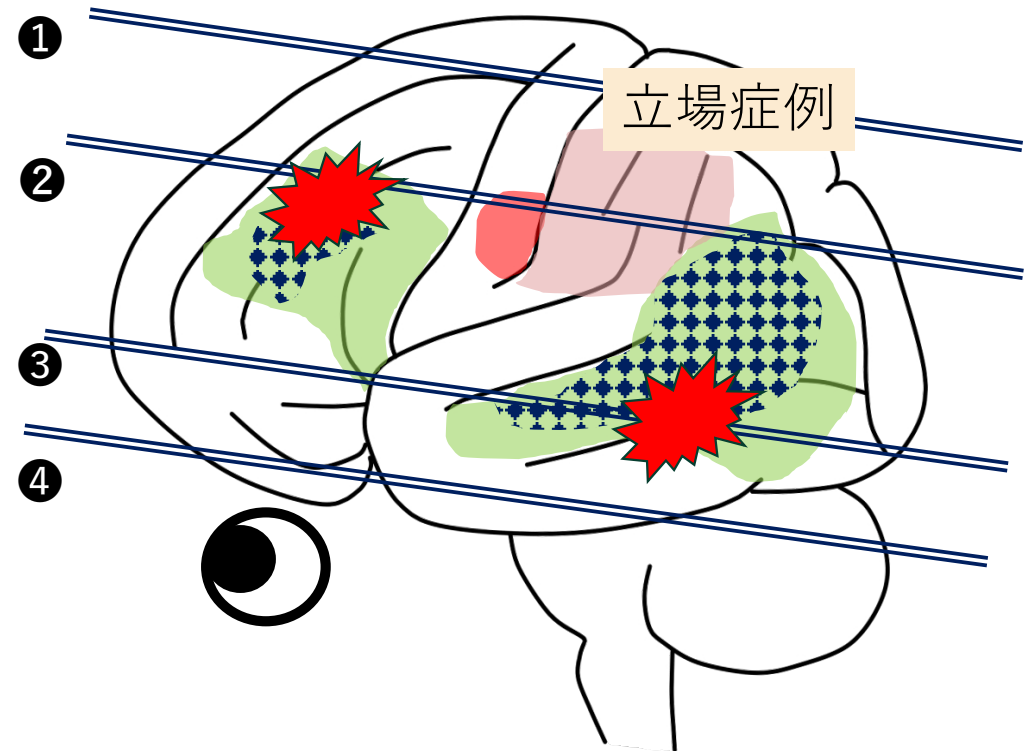
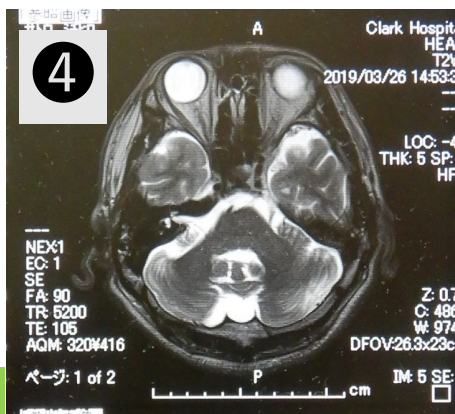
- ・ 意味-視覚性
 - ・ 意味カテゴリー
 - ・ 上位概念
 - ・ 意味連合
 - ・ 機能の記述
- を示す誤りが含まれる。

STは呼称を実施して、失語症者の語彙項目について評価する。

大脳の損傷部位と対応



側頭葉



色がついているのが失語症状を引き起こす脳部位

動画供覧～名詞の呼称～

目標語(絵)	反応
ローラースケート	OK
ボールペン	OK
ソフトクリーム	OK
クリスマスツリー	OK
物干しざお	パントマイム ・・・槍
ゴルフクラブ	ゴルフボール
スノーモービル	も、も、OK、 モビルスーツじゃない？
喉ぼとけ	のどちんこ
救急箱	OK
千羽鶴	折り紙、折り鶴

OK：目標語の表出が観察された場合



2分15秒まで

健常者の言い誤りと失語症者の錯語

	失語症者	健常者
脳の損傷	あり	なし
分析対象	呼称	自発話
音の誤り	あり	あり
意味の誤り	あり	あり
誤りの機序	???	???

CCAPは、この相違点について、
機械学習で、それぞれの誤りを
シミュレーションで予測できないか??
という、試みをしています。

まずは、機械学習の基礎をどうぞ！！

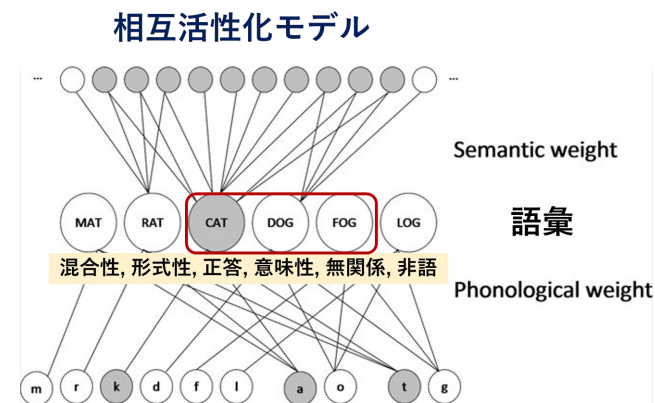


Fig. 1 The semantic-phonological (SP) model architecture
Foygel, D., & Dell, G. S. (2000)