

情動的知能資源としての リアクション絵文字

砂岡和子

早稲田大学 政治経済学術院

言語処理学会第30回年次大会併設WS1 (NLP2024)

2024年3月15日

報告の内容

- ▶ 1 研究目的
- ▶ 2 対象データと分析方法
- ▶ 3 リアク字の統計結果
- ▶ 4 AI 対 ヒトの絵文字翻訳
- ▶ 5 まとめ
- ▶ 参考文献

概要

- ▶ 大規模言語モデルの言語資源は、ヒトが書いたり話した言語テキストを中核に構築が進む。本発表は、昨年第29回年次大会(NLP2023)併設の同名ワークショップにおいて、Zoomと併用のプラットフォーム「Slack」上で使用されたリアクション絵文字（リアク字と略）に注目し、その言語テキストと伴奏する情動的知能および社会的知能の認知プロセスについて初歩的分析結果を報告する。
- ▶ 上記「Slack」で使用されたリアク文字は100種類以上、回数は延べ1500回を超え、参与者間の活発な情動的・社会的インタラクションを観察できる。
- ▶ 分析では、各コミュニケーション用途に応じリアク文字を5種に分類し、それぞれの出現頻度と言語テキストとの相関を調べた。
- ▶ 今後、「汎用性ある智能」を構成するには、言語テキスト以外に、リアク文字のような新しい言語パターンと表現を言語処理の俎上に載せ、言語資源の源泉に加えることで、より自然で柔軟なインタラクションを生成してゆくことが期待される。

研究目的

- ▶ 第二言語習得（L2中国語）とその実践応用
- ▶ 日本の大学L2授業の課題
- ▶ 大規模・初級クラス，教員主導，インタラクション方略の不足
- ▶ 発言に消極的な学習者
- ▶ ⇒ 授業活動振り返りによる教育方法の改善
- ▶ ⇒ ZoomChatやSlackなど新しい言語表現ツールの授業活用
- ▶ L2発話は複数言語コードのCode Switching，多種文字&記号が混在
- ▶ ⇒ 授業発話分析ツールの精度向上
- ▶ AIツールはMultilingual Interactions認識が苦手
- ▶ より自然で柔軟なインタラクションある言語資源の処理促進
- ▶ テキスト転写の精度向上が課題

分析対象データ

- ▶ NLP2023併設WS 1， Zoom併用のSlack投稿中のリアク字
- ▶ 本研究の分析データ
- ▶ WS開催日3月17日分のみ（9:41:08-17:55:27）． 約8時間余
Slackサイトオープンは3月4日～3月19日， 3月31日にクローズ.
- ▶ SlackのLogはダウンロードツールを使用せず， 手動で複製
2023年5月30日， NLP2023主催者に問い合わせ
- ▶ Slack上の画像や発言者などの情報はワードファイルに直接コピー不可，
- ▶ リアク字の送信時間も非表示
- ▶ 手動と目視でExcel入力
- ▶ リアク字はテキスト文字として入力 例「:arigato:」=9文字
- ▶ リアク字送信時間は， テキストの送信時間として記録.

参考；北京語言大学Zoom Chat の特徴

- ▶ ソースデータ：2020年7月以来，北京語言大学出版社主催“全球中文教学オンライン連続公開講座（BLCU Seminar）
- ▶ Zoom Text Chat 6回分，計3h，12628 threadsを分析対象
- ▶ 参加者は中国国内が8割，在米を中心に世界の中国語教員と学生が2割.
- ▶ 毎回の聴講者数は200-800名，使用言語は中国語+英語やタイ語など.
- ▶ 参加者の交流権限はTextChatのみ，カメラとマイク使用禁止，ハンドル名可
- ▶ 分析結果：Participantsの相互依存的なInformation exchangeが，初見同士の参加者が協調して会議を運営するWe-mode への媒介役を果たしている．大量のThanks & complimentsは，参加者が共同ゴールの知覚を促す機能がある.
- ▶ 日本のオンライン授業時，学生の積極参与を促すヒント（砂岡，2022）

リアク字分析方法と手順

- ▶ 中国北京語言大学Chat分析結果と比較，同分類枠用いる（砂岡，2022）
- ▶ WSのAgentを3分類
- ▶ A) **Participants**（参加者）
- ▶ B) **Organizer**（組織者） ←WSプログラム情報による
- ▶ C) **Guest**（招待講演者） ←同上
- ▶ コミュニケーション用途に応じ，リアク字を5種に分類
- ▶ ① **Question**（質問）
- ▶ ② **Request**（要求）
- ▶ ③ **Information exchange**（情報交換）
- ▶ ④ **Thanks & compliments**（感謝）
- ▶ ⑤ **Response**（回答）
- ▶ それぞれの出現頻度と言語テキストとの相関を調査

リアク字統計

- ▶ 「Slack」でのリアク字出現回数は延べ1500回超, 文字数は12000以上
- ▶ 参加者間の活発な情動的・社会的インタラクションを観察できる.

	リアク文字回数(A)	テキスト回数(B)	A+B
Participants	1147	246	1393
Guest	331	72	403
Organizer	75	104	179
total	1553	422	1975
	リアク文字数(C)	テキスト数(D)	C+D
Participants	9279	18211	27490
Guest	2671	7660	10331
Organizer	475	7517	7992
total	12425	33388	45813

リアク字統計

▶ 出現リアク字は110種類

Category	単出回数	リアクション数
:+1:	72	178
両目	88	159
kininaru/kininaru-	63	149
chiken	24	77
naruhodo	34	69
doukan	20	62
おじぎ/男女別	37	57
ii-shiteki	15	48
chatgpt/chatgpt-	12	35
祈る	35	32
sorena	15	32
バンザイ	29	25
拍手	17	21

レアなリアク字 (一部)	単出回数	リアクション数
tashikani-ka-ni-	2	2
ii-hanashi	2	2
ii-shigoto	2	2
考え中	2	2
キリッ	2	2
困ったあ	2	2
muzukashii-ne	2	2
myou-dana	2	2
nazo	2	2
atsu	1	1
ii-shitumon	1	1
ii-teian	1	1
ワクワクハート	3	3
chimdondon	2	2
www	2	2
uoo	2	2
waiwai2	2	2
zawa_zawa	2	2
wakuwaku	1	1
thumbsup_all	1	1
unige-sabira	1	1
meow-attention-party	1	1
denshiba	2	2
gengogaku	2	2
annotation	1	1
ラマ	3	3
閉じた本	2	2
トロフィー	2	2
にふえでびる	2	2
お金の袋	1	1
ぐる目顔	1	1
マイク	1	1

実際のリアク字

Category	単出回数	リアクション数
:+1:	72	178
両目	88	159
kininaru/ kininaru- chiken	63	149
naruhodo	24	77
doukan	34	69
おじぎ/男 女別	37	57
ii- shiteki	15	48
chatgpt/c hatgpt-	12	35
祈る	35	32
sorena	15	32
バンザイ	29	25
拍手	17	21



なるほど

いい指摘

レアなリアク字 (一部)	単出回数
zawa_zawa	2
キラッ	2
myou-dana	2
ラマ	3
にふえでびる	2

ざわ...
ざわ...



妙だな...



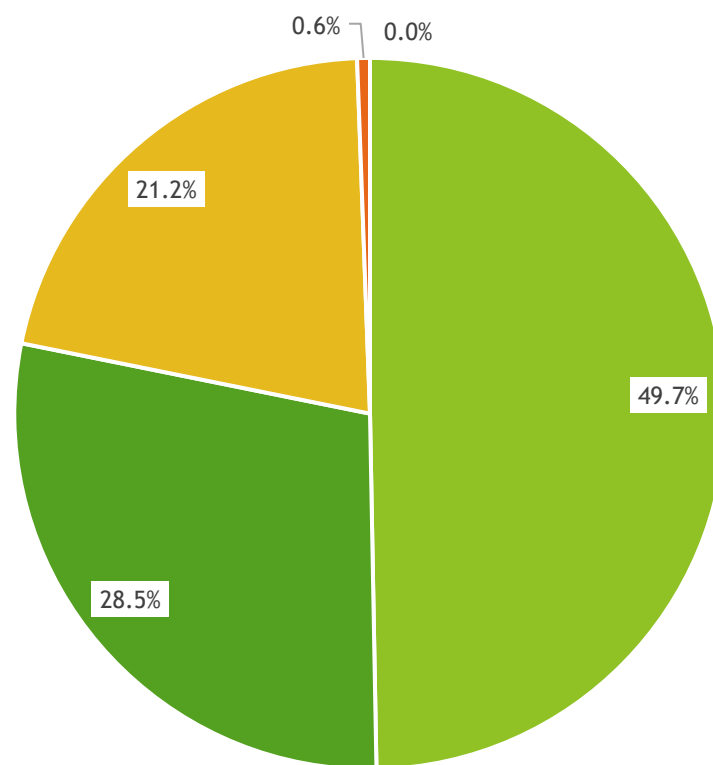
にふえ〜
で〜びる

リアク字統計

Topic	リアク字Topic別頻度
response	49.7%
thanks & compliments	28.5%
information exchange	21.2%
question	0.6%
request	0.0%

NLP参加者のリアク字使用目的
「回答」 > 「感謝」 > 「情報交換」

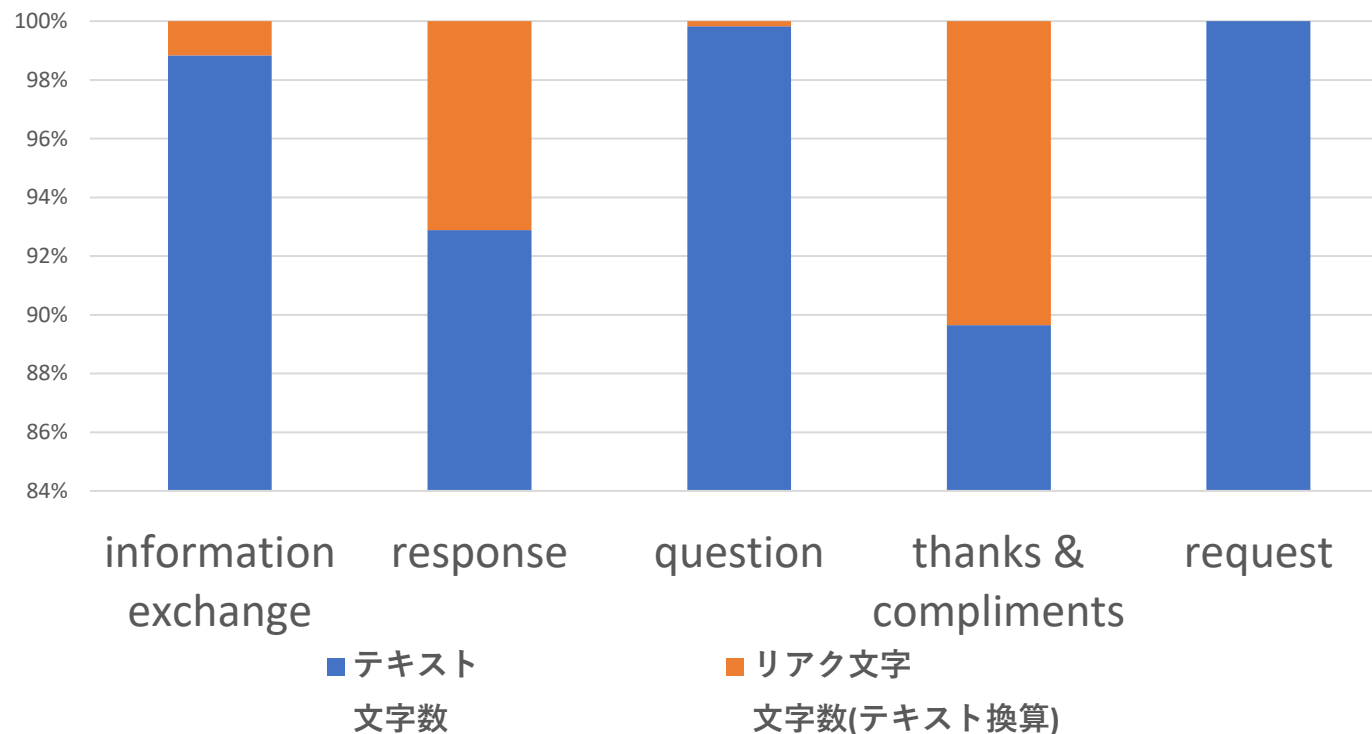
リアク字 Topic別頻度



■ response ■ thanks & compliments ■ information exchange ■ question ■ request

リアク字統計

テキストVSリアク文字数



リアク字使用比率が高いのは「感謝」 > 「回答」 > 「情報交換」
絵文字のほうが「感謝」を表現し易い？

リアク字統計

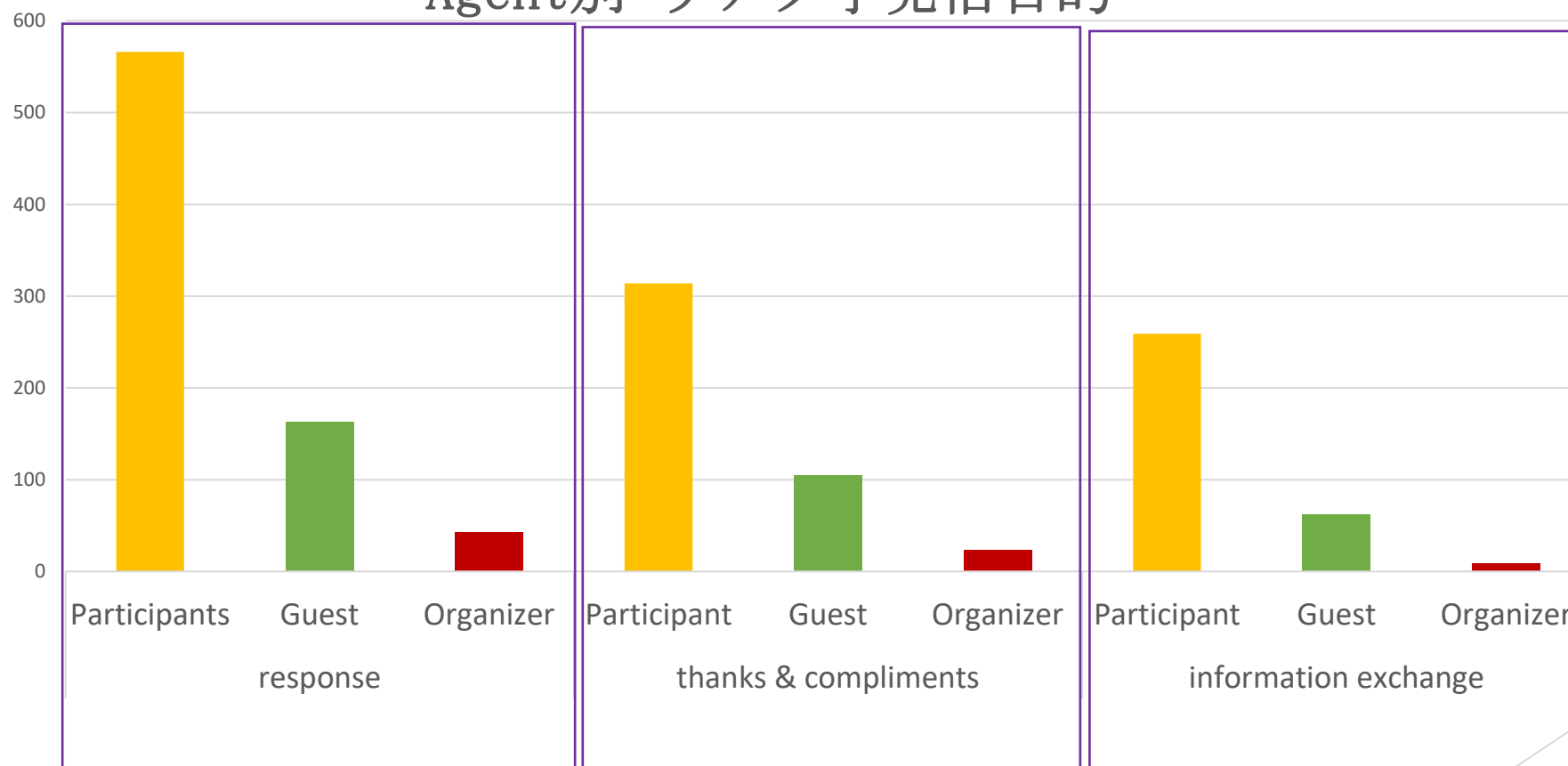
(10件以下は除く)

Agent別リアク文字使用目的		リアク文字回数
Participants	response	566
	thanks & compliments	314
	information exchange	259
Guest	response	163
	thanks & compliments	105
	information exchange	62
Organizer	response	43
	thanks & compliments	23

リアク字使用の主体は「**参加者**」，「**組織者**」と「**ゲスト**」は少ない

リアク字統計

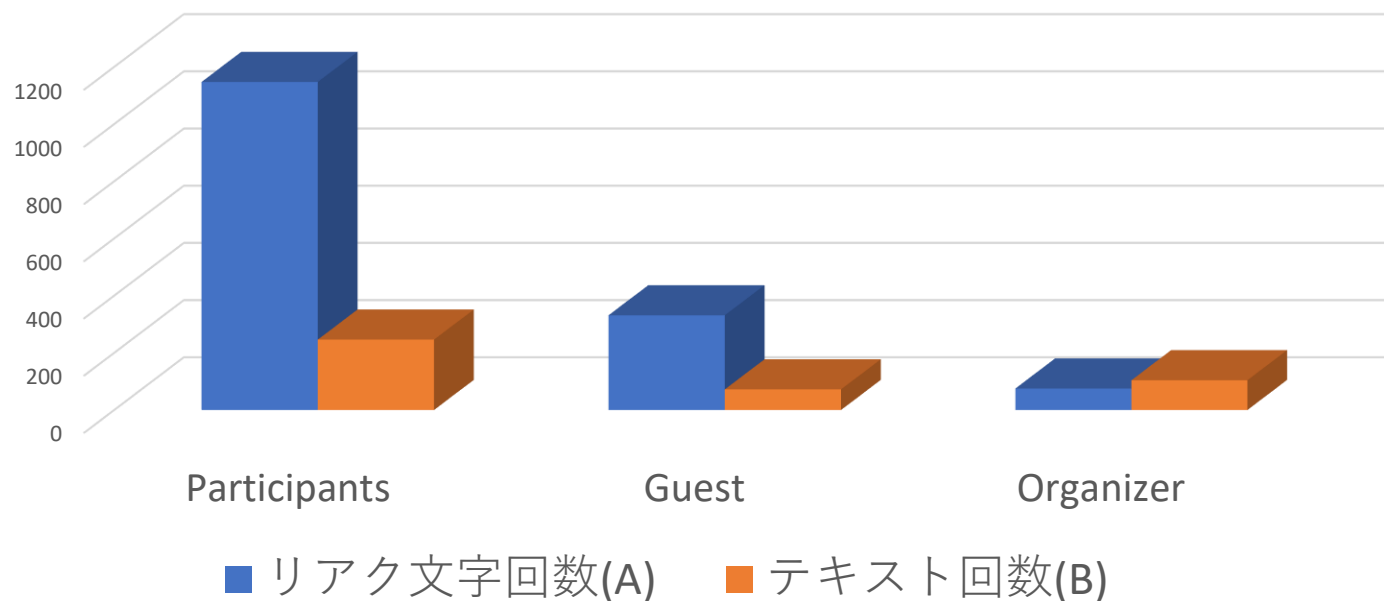
Agent別 リアク字発信目的 (10件以下は除く)



目的に拘わらず、「参加者」のリアク字使用がダン突

リアク字統計

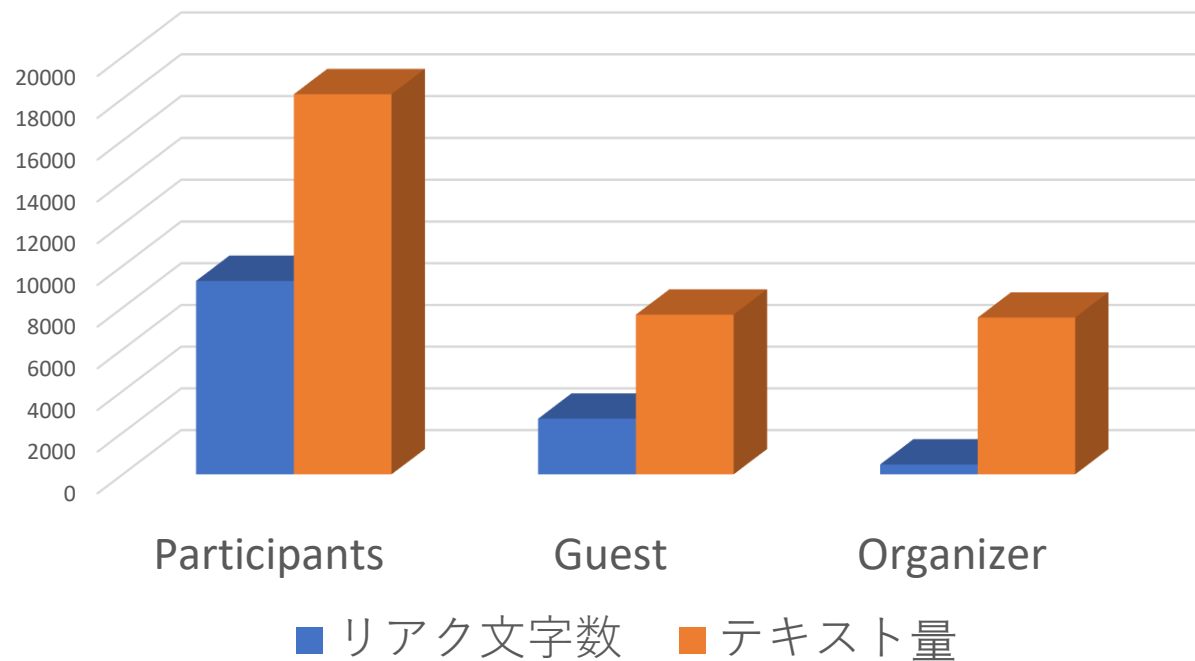
Agent別 リアク字 VS テキスト回数



当然ながら，Slackでは「テキスト」による書き込み回数が，リアク字より多い．例外は「**組織者**」．多忙のため絵文字の投稿が便利？

リアク字統計

Agent別 リアク字 vs テキスト量



すべてのAgentで、「テキスト」の量はリアク字よりはるかに多い

ChatGPTの絵文字翻訳

ChatGPTは絵文字の翻訳ができる（近藤，2024）

和子

You

「ややこしい！」という文を表象する絵文字を全部教えてください



ChatGPT

「ややこしい！」を表現する絵文字を選ぶ場合、その感情や状況の混乱や複雑さを象徴するものを考えます。全てを網羅するのは難しいですが、一般的にこの感情や状態を表すのに適している絵文字のいくつかを紹介します：

- 🌀（目が回る） |
- 💥（頭が爆発する）
- 😕（困惑）
- 🔄（上下逆さ）
- 😡（怒りやイライラ、ここでは混乱や複雑さを感じている状態を表すのに使える）
- ❓❓（疑問、混乱）

これらの絵文字は、「ややこしい！」という感情や状況の混沌とした感覚を象徴的に表現できます。

絵文字・リアク字は**多義性**あり，解釈（翻訳）に揺れ














ChatGPTのリアク字翻訳

和子 You

「毎度のことですがChatGPTすごいです。」という文を、数種類の絵文字に翻訳してください。

 ChatGPT

「毎度のことですがChatGPTすごいです。」という文を数種類の絵文字で表現すると、以下のようになります：

1.    - 「毎度のことですが、ChatGPTは素晴らしいです。」
2.    - 「いつもですが、ChatGPTは驚異的です。」
3.    - 「毎回のように、ChatGPTは驚きを与えてくれます。」
4.     - 「いつも通り、ChatGPTの知能に拍手です。」



ヒトのリアク字反応

末田知彦(ヤフー) 09:44

(毎度のことですが)ChatGPT すごいです。



2 件の返信

[Redacted]

ChatGPT は読みが苦手なようで関西学院大学も先斗町も読めません。



石原祥太郎(日経新聞) 09:45

次回開催地の兵庫でも余部(あまるべ よべ)とかがありますね。曖昧性回避のために、前者の駅名の漢字が「餘部」になっているらしいです。



鈴木正敏(Studio Qusia/東北大) 09:46

母方の実家の近くに、登米市登米町(とめし・とよままち)があります。ややこしい！

たしかに

竹岡邦彦(NEC) 09:48

羽生結弦さんと羽生善治さん問題は混同してそうです(プロンプト次第かもしれないです)



リアク字使用はことばの翻訳ではなく、**情動表現**
テキストと伴奏、かつ**時差表現**が可能

ヒトのリアク字反応

██████████ ←

こちらでしょうか↓

<https://current.ndl.go.jp/car/44879> ←

青空文庫及びサビエの点字データから作成した振り仮名のデータセット 全国書誌データから作成した振り仮名のデータセット ←

(編集済み) ←

 カレントアウェアネス・ポータル ←

[NDLラボ、振り仮名のデータセット「振り仮名注釈コーパス」\(2種類\)を公開](#) ←

2021年9月21日、国立国会図書館(NDL)のNDLラボが、振り仮名のデータセット「振り仮名注釈コーパス」2種類をGitHubで公開しました。日本語のテキストを分かち書きにした上で、分割した箇所に振り仮名を付与したもので、青空文庫の公開作 ←

2021年9月28日 ←

 1 ←

██████████ ←

ああ、それです。ありがとうございます。 ←

 2 ←

██████████ ←


[@宮田将樹\(ノニー\)](#) 教えていただき、ありがとうございます! ←

 1 ←

ChatGPT ↓

 You

「ああ、それです。ありがとうございます。」という文を、絵文字に翻訳してください。

 ChatGPT

「ああ、それです。ありがとうございます。」という文を絵文字で表現すると、「👍🙌🙏」となります。

リアク字は表現のバリエーションが豊富

まとめ

- ▶ NLP2023WS, 約8時間余のSlack投稿中のリアク字を分析
- ▶ Agent別, コミュニケーション用途別に分類
- ▶ 出現リアク字は110種類, 回数は延べ1500回超, 文字数は12000以上
- ▶ 180回近い頻出字から1-2回まで, バリエーションが豊富
- ▶ リアク字使用は「参加者」主体 > 「組織者」「ゲスト」は極少ない
- ▶ リアク字使用目的は昇順に「回答」 > 「感謝」 > 「情報交換」
- ▶ 回数と量はテキストが多いが, 「組織者」のみリアク字の回数が多い
- ▶ AIも絵文字翻訳できるが, テキストの単純な置き換え
- ▶ リアク字はテキストと伴奏する情動表現手段
- ▶ 参加者間の自然で柔軟なインタラクションを支援
- ▶ だが絵文字・記号が混じるテキスト転写には課題が多い

謝辞

- ▶ 本研究の一部は，JSPS科研費C（21K00773）の助成による．
- ▶ データ整理に当たり，譚翠玲さん（北海道大学国際広報メディア・観光学院院生）の協力を得た．併せて感謝する．

参考文献

- ▶ Gardner, Chloros (2009) Code-switching, Cambridge University Press.
- ▶ ChatGPT-4
- ▶ 近藤泰弘 (2024) . 大規模言語モデルの持つ言語知識とコミュニケーション, 第15回産業日本語研究会・シンポジウム招待講演 (2月20日)
- ▶ 徐勤・砂岡和子 (2024) . 複数言語コードを含む発話転写と話者分離: Whisper+Pyannote.audioによる自動音声認識の高度化, 言語処理学会第30回年次大会 (NLP2024) 予稿集, PP 3149-3154
- ▶ 砂岡和子, 王松, 杉江聡子, 徐勤 (2023) . 中国語授業のCode-Switching—包摂的メンバーシップとL2習得最適化. 日本中国語学会第72回全国大会予稿集, pp. 253-257.
- ▶ 砂岡和子, 徐勤 (2023) . 生成系AIと初級外国語学習者のCode Switching発話「聴取」パフォーマンス, 電子情報通信学会信学技報, 教育工学研究会 (ET) 「教育学習支援システムの開発と運用」予稿集, pp.33-37. IEICE Technical Report ET2023-23 (2023-10)
- ▶ 砂岡和子 (2022) . Zoom Chatに見るWe-Mode-中国語話者のMulti-Agent Interaction, 言語処理学会第28回年次大会, pp. 1916-1920