



Uniwersytet im. Adama Mickiewicza w
Poznaniu

Adam Mickiewicz University

<https://researchportal.amu.edu.pl>

Publikacja / Publication	Generative Transformers versus dictionaries, with particular focus on English phrasal verbs, Lew Robert
Adres publikacji w Repozytorium URL / Publication address in Repository	https://researchportal.amu.edu.pl/info/report/UAMfaafebc72d7741e0aaabd36219bb5717/
Data opublikowania w Repozytorium / Deposited in Repository on	Jul 13, 2024
Rodzaj licencji / Type of licence	Attribution (CC BY)
Cytuj tę wersję / Cite this version	Lew Robert: Generative Transformers versus dictionaries, with particular focus on English phrasal verbs, 2024, PhrasaLex III Workshop on Phraseological Approaches to Lexicography 2024, presentation



ADAM MICKIEWICZ UNIVERSITY IN POZNAŃ
Faculty of English

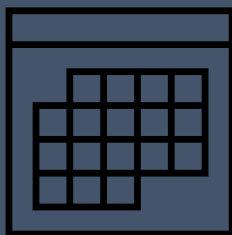
Generative Transformers versus dictionaries, with particular focus on English phrasal verbs

Robert Lew

PhrasaLex III, Innsbruck / online
2024-07-11, 16:00-16:30

Writing entries is so much work... and time...

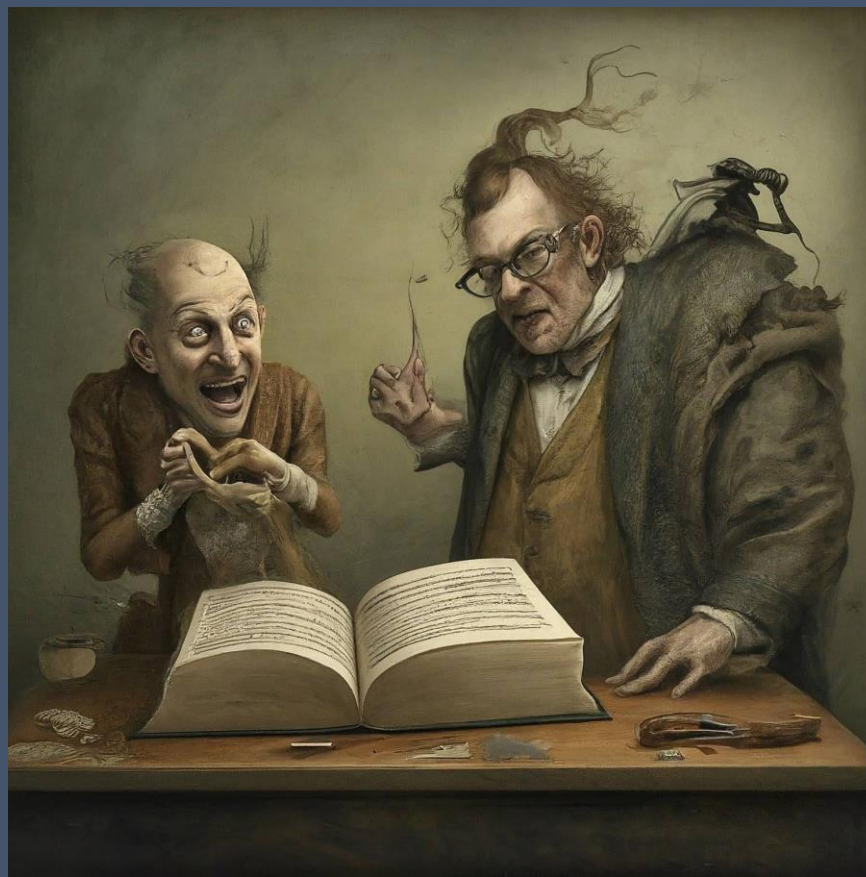
- Writing dictionary entries is ‘harmless drudgery’ (Samuel Johnson 1755)
- Very time-consuming and laborious
- Some prominent lexicographic projects:
 - **71** years: *The Oxford English Dictionary*, 1857-1928 (Mugglestone 2008)
 - **123** years: *Das Deutsche Wörterbuch*, 1838-1961 (Haß-Zumkehr 2011)
 - **144** years: *Woordenboek der Nederlandsche Taal (WNT)*, 1864-1998 (De Schryver 2005, Van Sterkenburg 1984)



A definition of lexicographer (?)

lexicographer /lɛksɪ'kɒgrəfə/ Noun ▶ A rare species with unique personal qualities and special skills from years of training.

- Or is it...?



Can computers do the work?

- De Schryver & Joffe (2023): “The end of lexicography, welcome to the machine: On how ChatGPT can already take over all of the dictionary maker's tasks”
 - AI-generated illustrations in the Phrase-based Active Dictionary (PAD, Linda Prossliner’s yesterday’s talk
- Earlier: Grefenstette (1998) : “Will there be lexicographers in the year 3000?”
- Rundell (2011): “Will there still be dictionaries in 2020?”
- Two separate questions: *(how much) do we need lexicographers* vs. *(how much) do we need dictionaries*
- In other words: *can computers do the work* vs. *does this work need to be done at all*

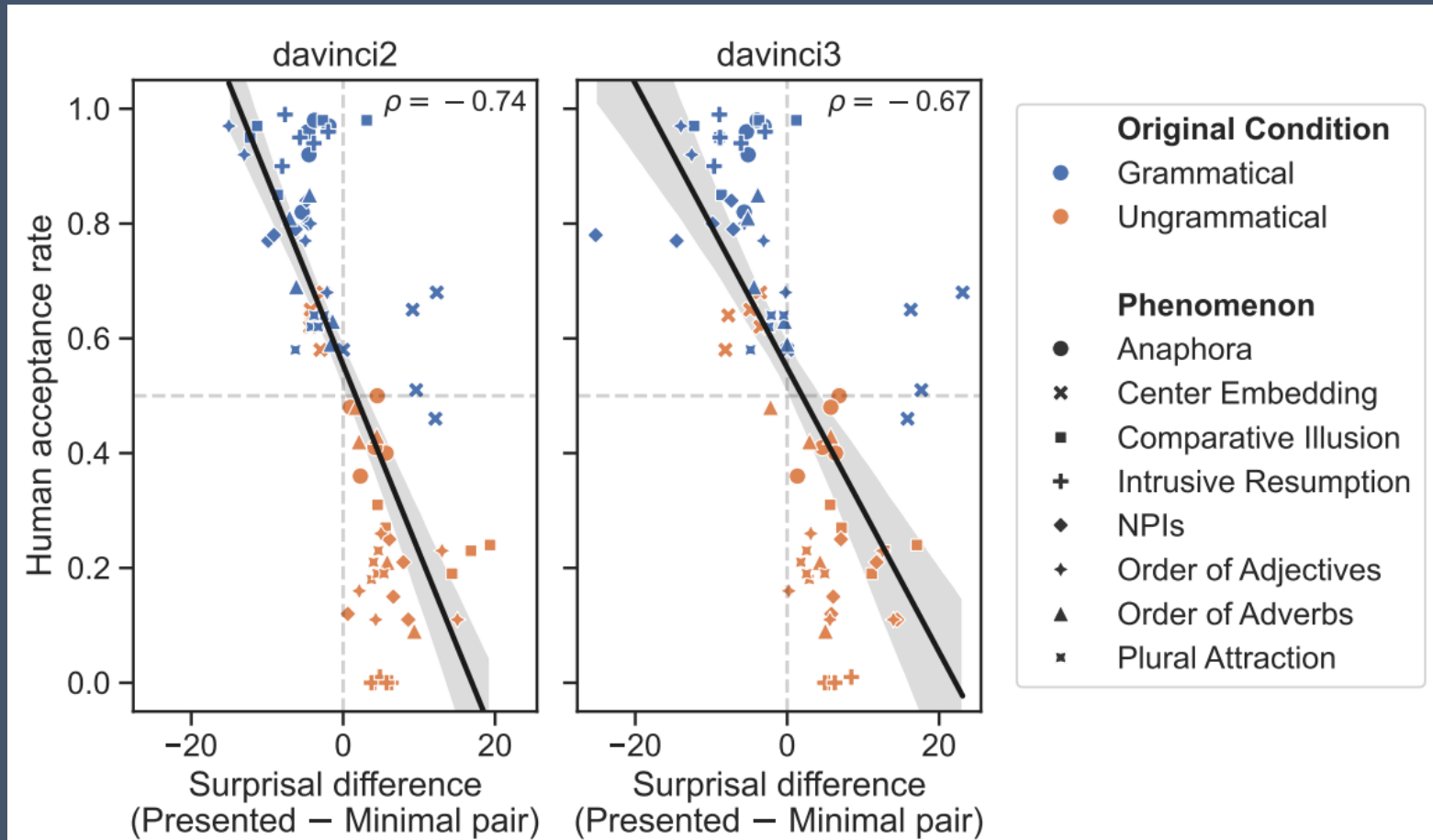
Large Language Models

- LLM and GPT technologies are not new (*Deep Learning* around ~ 2010)
- GPT = Generative Pre-Trained Transformer
- ‘You shall know a word by the company it keeps’ (Firth 1957: 11)
- ‘You shall predict a word from the company it keeps’ (GPT)
- GPT predicts the most natural continuation from conditional probability matrices
- It doesn’t always go with the most probable choice (*temperature: norms and exploitations?*)
- GPT explosion: critical mass in data volumes (*scale*), advances in training technology, releasing a publicly available interface
- + attention mechanisms, reinforcement learning, ? chatbot layer

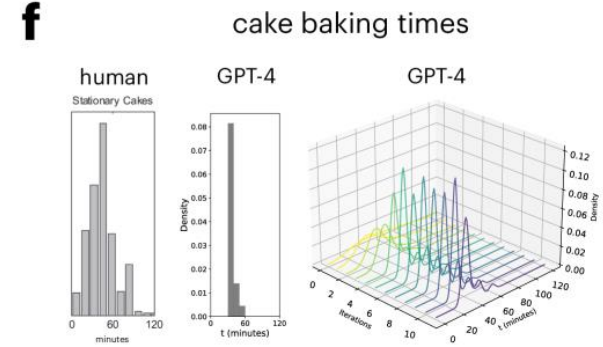
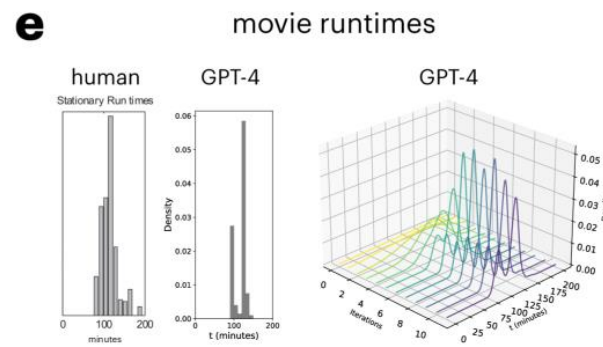
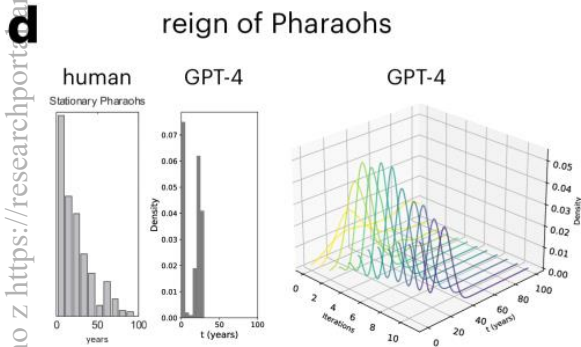
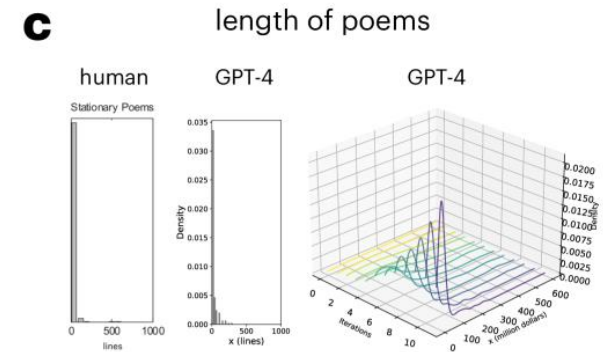
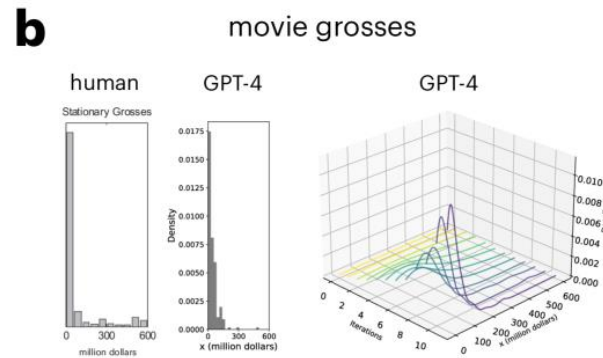
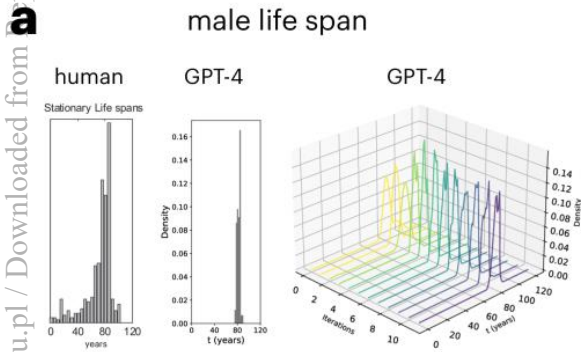
Language Models gain fine insights into 'grammar' by merely digesting input

- Mastery of novel utterances and judgments of grammaticality taken as evidence of 'innate linguistic knowledge' (Chomsky 1957, 1965, 1986)
- Recent evidence from Machine Learning: Language Models can learn complex grammar, even from data volumes corresponding to what a human child might normally be exposed to (BabyLM)
- They exhibit emergent learning of 'grammatical abstractions' (Baroni 2022, Piantadosi 2023)
- Language Models are 'stochastic parrots'? In a sense, yes, but so are humans
- c.f. the human Tolerance Principle (Yang 2016)
 - 'Children form productive rules when it is more computationally efficient' (Schuler, Yang, Newport 2021)
 - number of exceptions $e < n / \log(n)$

Hu (2024): Language models align with human judgments on key grammatical constructions



Zhu & Griffiths (2024). Eliciting the Priors of Large Language Models using Iterated In-Context Learning



Marjeh et al. (2023). Large language models predict human sensory judgments across six modalities

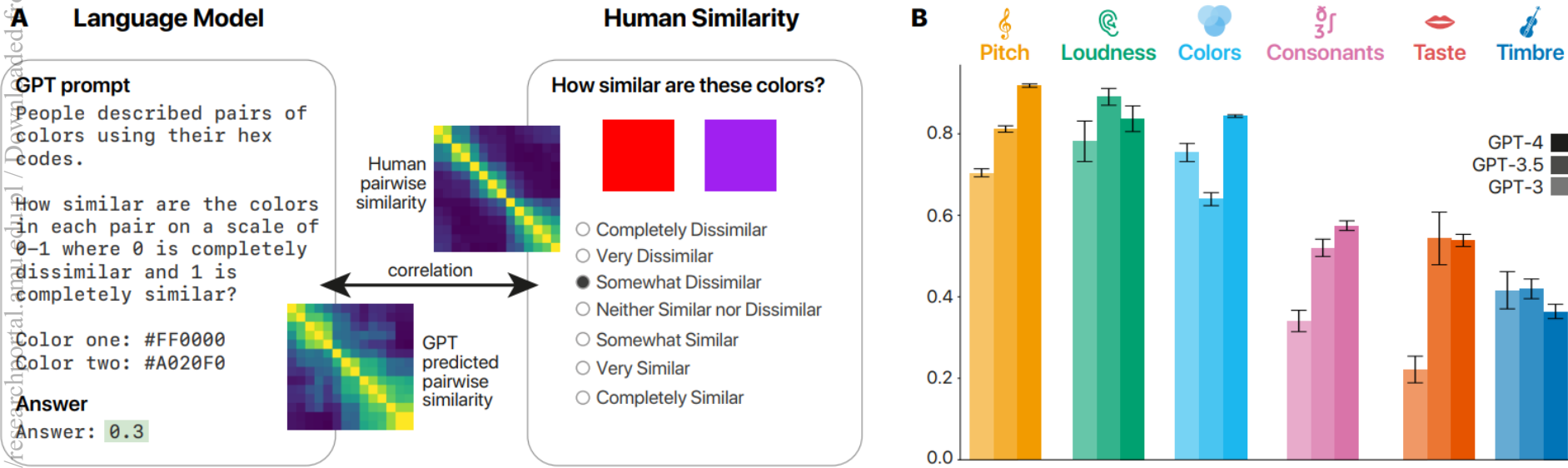
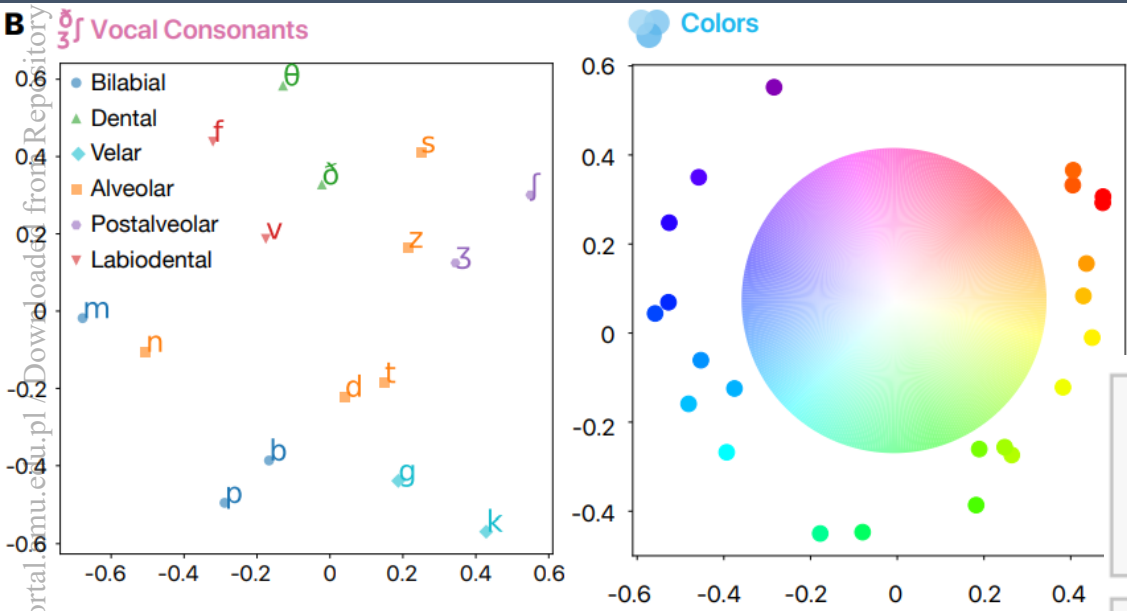


Fig. 1. A. Schematic of the LLM-based and human similarity judgment elicitation paradigms. **B.** Correlations between models and human data across six perceptual modalities, namely, pitch, loudness, colors, consonants, taste, and timbre (Pearson r ; 95% CIs).

Marjeh et al. (2023). Large language models predict human sensory judgments across six modalities



I gave this similarity rating because the frequencies of the two notes are in a **1:2 ratio**, which means they are an **octave apart** and share a **strong harmonic relationship**, making them more similar than other pairs with less harmonic connection.



Both 'm' and 'n' are **nasal consonants**, which means they are produced with the airflow passing through the nose. However, they have different places of articulation, with 'm' being **bilabial** (produced with both lips) and 'n' being **alveolar** (produced with the tongue touching the alveolar ridge). This makes them somewhat similar but not identical in sound.



I gave this similarity rating because the two colors are on **opposite sides** of the **color spectrum**, with one being a bright red (**#FF0000**) and the other being a bright blue (**#00b2ff**). They have very different **hue values**, making them **visually dissimilar**.

AI-augmented lexicography

- Grefenstette (1998) : “Will there be lexicographers in the year 3000?”
- Rundell (2011): “Will there still be dictionaries in 2020?”
- De Schryver & Joffe (2023): “The end of lexicography, welcome to the machine: On how ChatGPT can already take over all of the dictionary maker's tasks”
- Two separate questions: ***(how much) do we need lexicographers vs. (how much) do we need dictionaries***
- In other words: ***can computers do the work vs. does this work need to be done at all***

ChatGPT as a COBUILD lexicographer

[Robert Lew](#) 

Humanities and Social Sciences Communications **10**, Article number: 704 (2023) | [Cite this article](#)

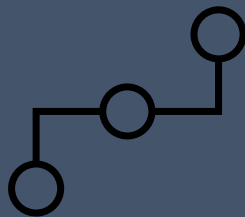
[Metrics](#)

Study 1 (in honour of Patrick) ChatGPT writes COBUILD Definitions

How well will ChatGPT emulate the performance of a human lexicographer in compiling COBUILD-like entries?

Why was COBUILD (1987) special?

- Widely seen as innovative
- Based on a corpus of digital texts as evidence of use and meaning
- Reliance on authentic examples
- A view of language that unifies structure and meaning
- Senses as associations between patterns of usage and meaning
- As a consequence, *contextual* definition (aka *FSD*, *Full-Sentence Definition*)



COBUILD-style definition

- Compare two definitions of one sense of the verb APPROVE, classical/analytical and COBUILD-style:

APPROVE: to think that someone or something is good, right, or suitable

If you **approve of** someone or something, you like and admire them.

- Two clauses:
 1. illustrates the usage pattern in the sense defined
 2. provides a paraphrase of meaning
- COBUILD breaks with the substitutability orthodoxy, introduces a more conversational defining style

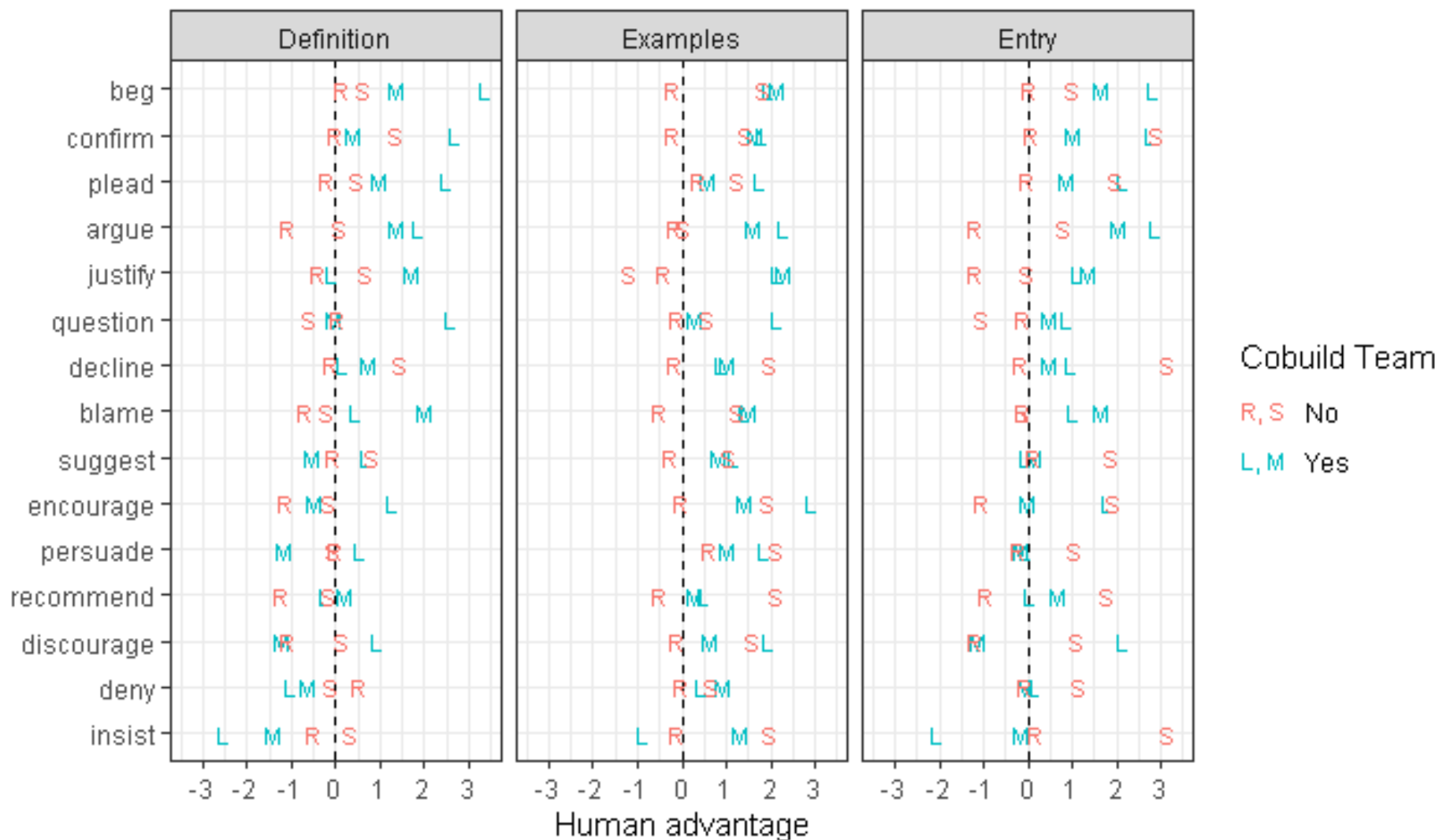
Design

- 15 verbs of communication, B1+ (English Vocabulary Profile, Capel 2015)
- for each verb, one entry from Cobuild online, one generated with ChatGPT Plus (omitted syntactic codes, phonemic transcriptions, inflected forms, all easy for ChatGPT)
- Prompting:
 - An introductory prompt setting out the purpose
 - Two model entries presented from COBUILD Online: APPROVE and ASSASSINATE, with some comments
 - Then just the 15 headwords, one by one
- *few-shot* learning (Brown et al. 2020), via interactive text interface

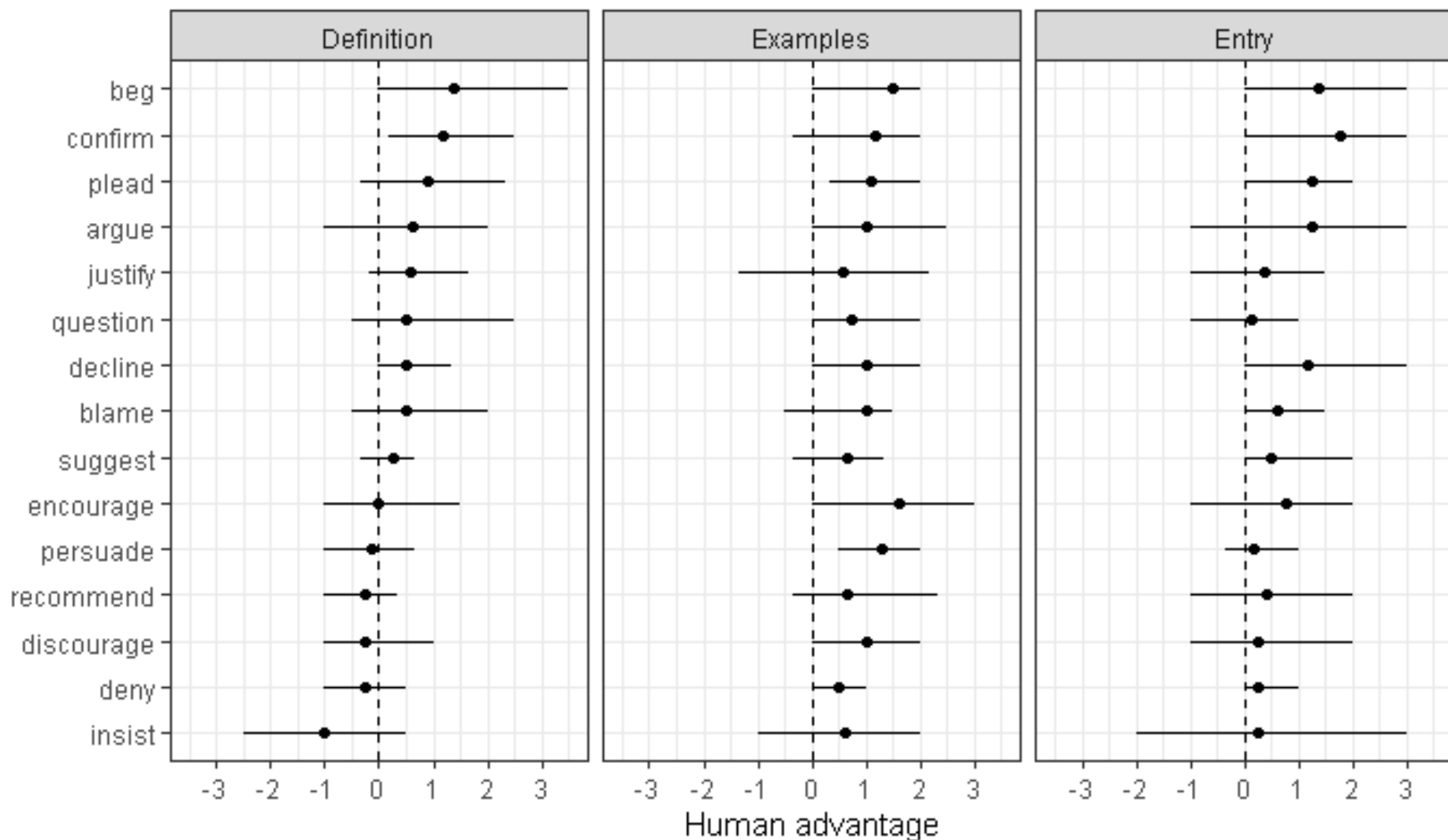
Evaluation

- Four experts on English definition and lexicography (big thanks to them!)
 - Two original COBUILDers: Liz Potter and Michael Rundell
 - Two experts on defining models for learners of English: Reinhard Heuberger and Sylwia Wojciechowska
- Entries randomized, no indication whether human or AI-generated
- Experts asked to rate the quality of:
 - sense definition
 - sense exemplification
 - entry as a whole
- A five-point scale: *Bad, Wanting, Passable, Good, Great* (= 1...5)
- Open-ended comments (optional)

Human advantage by Headword, Element and Rater



Human advantage by Entry and Element (mean + range)



Open-ended comments: ChatGPT entries



Example prompt 1: APPROVE

approve [VERB]

1 If you approve of someone or something, you have a favourable opinion of them or it.

I approve of his plans for the future.

She approves of her children's friends.

They approve of the new law.

2 If you approve a plan, proposal, or idea, you give your official permission or endorsement to it.

The board of directors **approved** the proposal.

The government **approved** the construction of a new highway.

The school **approved** the students' request for a field trip.

3 If you approve a payment, loan, or expense, you authorize it to be made.

The bank **approved** the loan for the new business.

The manager **approved** the expense for the company trip.

The committee **approved** the payment for the consultant.

Examples: fine-tuning

- Open-ended comments were fed back to ChatGPT to see if it could improve on the quality of examples (for the entry with the lowest ratings for examples of all: ENCOURAGE)
- This was done using a newer model GPT 4, version of May 12th, 2023



Original ChatGPT entry

encourage [VERB]

1. If you encourage someone to do something, you give them support, confidence, or motivation to do it.

- *The coach encouraged the team to do their best.*
- *The teacher encouraged the student to study harder.*
- *The parent encouraged the child to try again.*

2. If something encourages you, it gives you support, confidence, or motivation to continue or to do something.

- *The good grade encouraged the student.*
- *The sunny weather encouraged the hikers.*
- *The success encouraged the entrepreneur.*

Fine-tuned ChatGPT entry

encourage [VERB]

1. If you encourage someone to do something, you give them support, confidence, or motivation to do it.

- *In a gentle voice, Mary encouraged her shy little brother to speak up in class.*
- *Despite the initial failure, Rob's friends encouraged him to try starting his own business again.*
- *"Never stop believing in yourself," the coach said, encouraging the team before the big match.*

2. If something encourages you, it gives you support, confidence, or motivation to continue or to do something.

- *Finding her first novel on the bestseller list greatly encouraged the young author to continue writing.*
- *With the sun shining brightly, the hikers felt encouraged and decided to take the more challenging trail.*
- *His early success as an entrepreneur encouraged him, instilling the belief that he could turn any idea into a successful business.*

Fine-tuned entry

- The fine-tuned entry sent back to Experts asking for informal feedback
- All Experts agreed that the new examples were a big improvement and in fact very good



Post-lexicography

- Grefenstette (1998) : “Will there be lexicographers in the year 3000?”
- Rundell (2011): “Will there still be dictionaries in 2020?”
- De Schryver & Joffe (2023): “The end of lexicography, welcome to the machine: On how ChatGPT can already take over all of the dictionary maker's tasks”
- Two separate questions: *(how much) do we need lexicographers vs. (how much) do we need dictionaries*
- In other words: *can computers do the work vs. does this work need to be done at all*

A Learners' Dictionary Versus ChatGPT in Receptive and Productive Lexical Tasks

Bartosz Ptasznik, Sascha Wolfer, Robert Lew ✉

International Journal of Lexicography, ecae011, <https://doi.org/10.1093/ijl/ecae011>

Published: 04 July 2024

Study 2: Ptasznik, Wolfer & Lew

LDOCE (mobile) versus ChatGPT 3.5

Reception (comprehension) and Production

The challenge of English phrasal verbs

Test item	Target structure
turn	turn over something
try	try out for something
take	take against somebody/something
see	see something out
run	run up against somebody/something
play	play up to somebody
pick	pick at something
patch	patch somebody/something up
make	make off with something
kick	kick in
hit	hit somebody up for something
give	give out
fix	fix somebody up with somebody
draw	draw yourself up (to your full height)
do	do something over
demand	demand that somebody (should) do something
conceive	conceive of somebody as something
carry	carry something off
bump	bump somebody off
break	break somebody of something

The challenge of English phrasal verbs

Osoba badana (proszę wpisać numer): _____

Rodzaj narzędzia: Słownik / ChatGPT

Instrukcja: Zapisz, ile czasu zajęło Ci sprawdzanie w słowniku / ChatGPT, a następnie przetłumacz podane zdanie na język angielski, wykorzystując podany wyraz. Możesz zmienić formę podanego wyrazu.

Przykład:

Zdanie do przetłumaczenia: Rodzice Michaliny wiedzą, że jakiegokolwiek jedzenie zawierające pszenicę wywołuje u niej wysypkę.

Wykorzystaj podany wyraz: **bring**

Czas: 1 min 15 s Odpowiedź: Michalina's parents know that any foods containing wheat bring her out in a rash.

Test:

1. Zdanie do przetłumaczenia: Ktoś włamał się do sklepu i zwinął kilka telewizorów.


Wykorzystaj podany wyraz: **make**

Czas: _____ Odpowiedź: _____

The challenge of English phrasal verbs

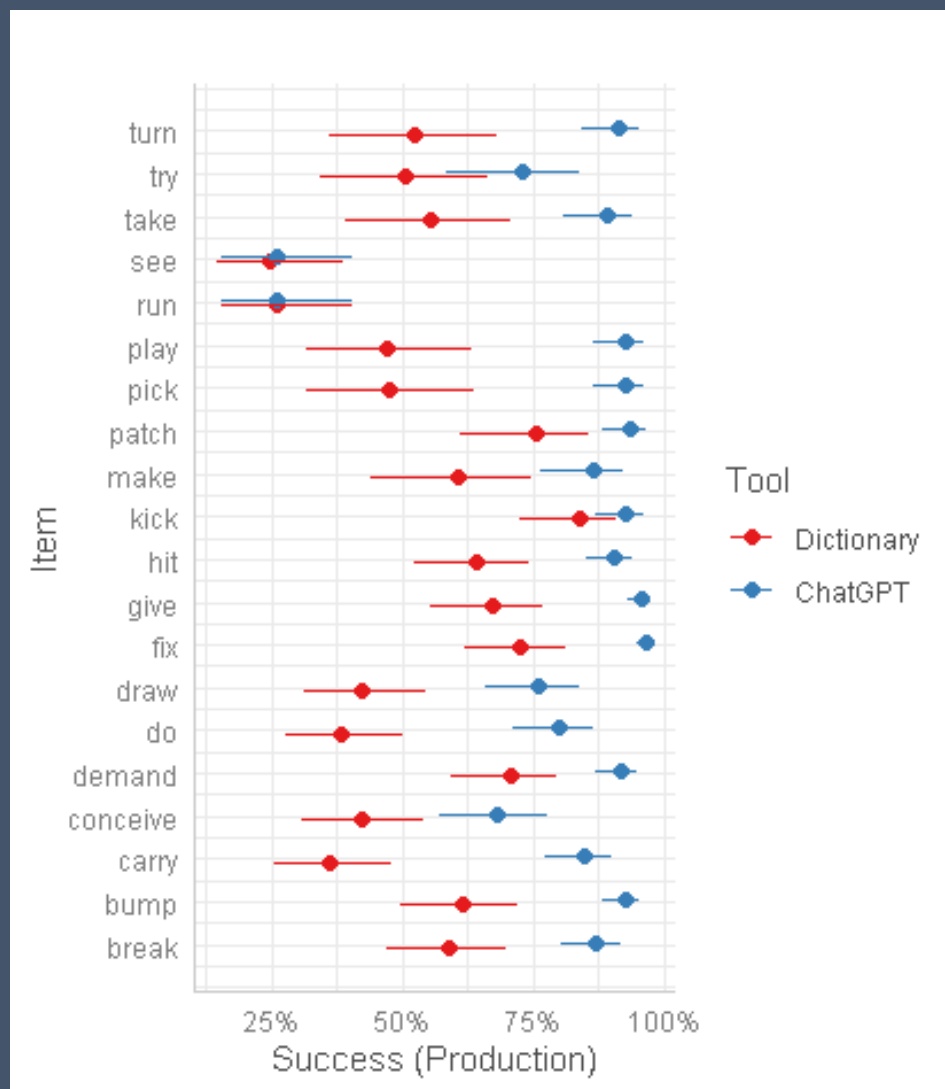
Translate into English the sentence "Ktoś włamał się do sklepu i zwinął kilka telewizorów" using the word "make".

Hmmm, interesting challenge. Here goes: "Someone broke into the store and made off with several televisions." Did I get it right? 🤓

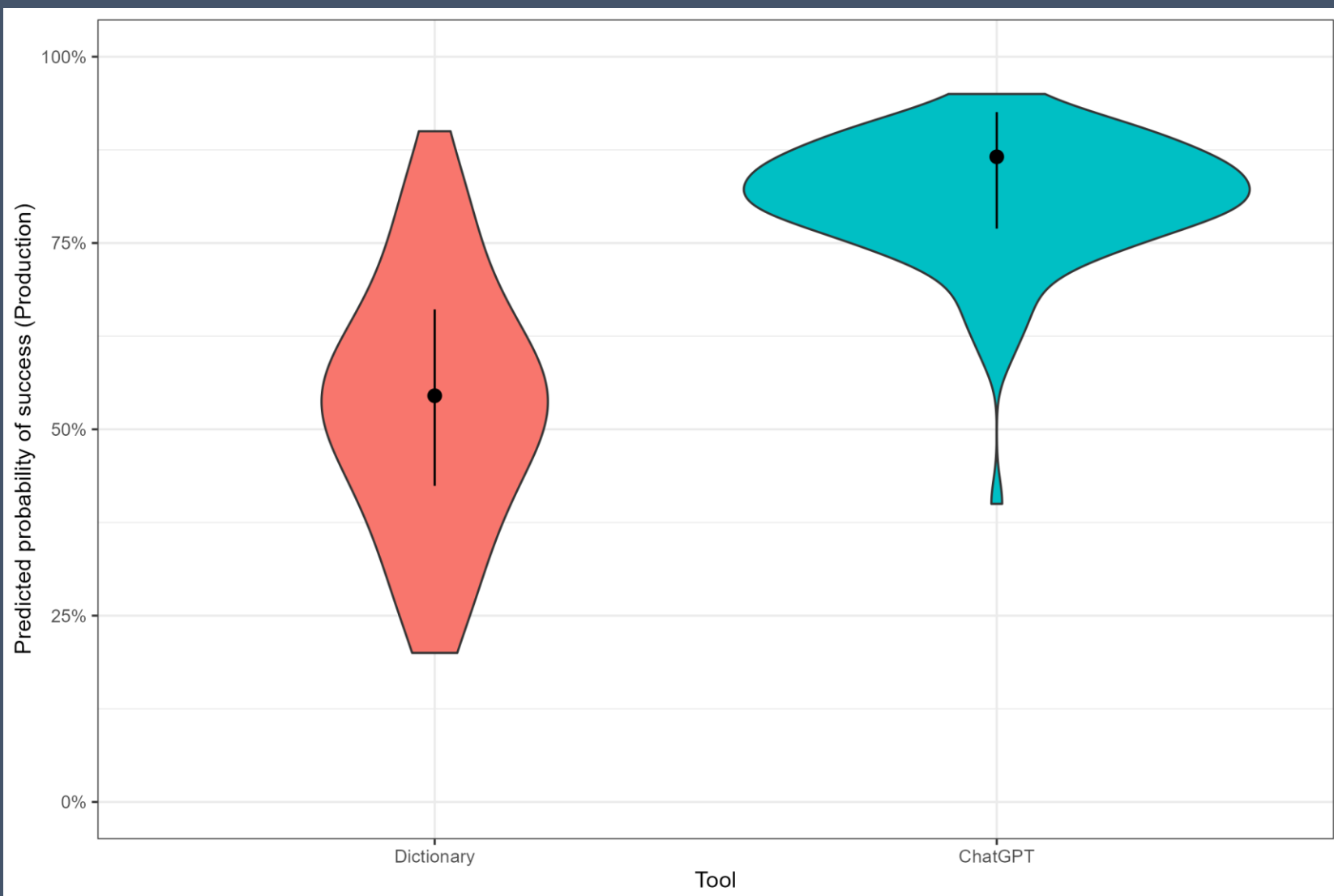
Pi, your personal AI 

Producing phrasal verbs

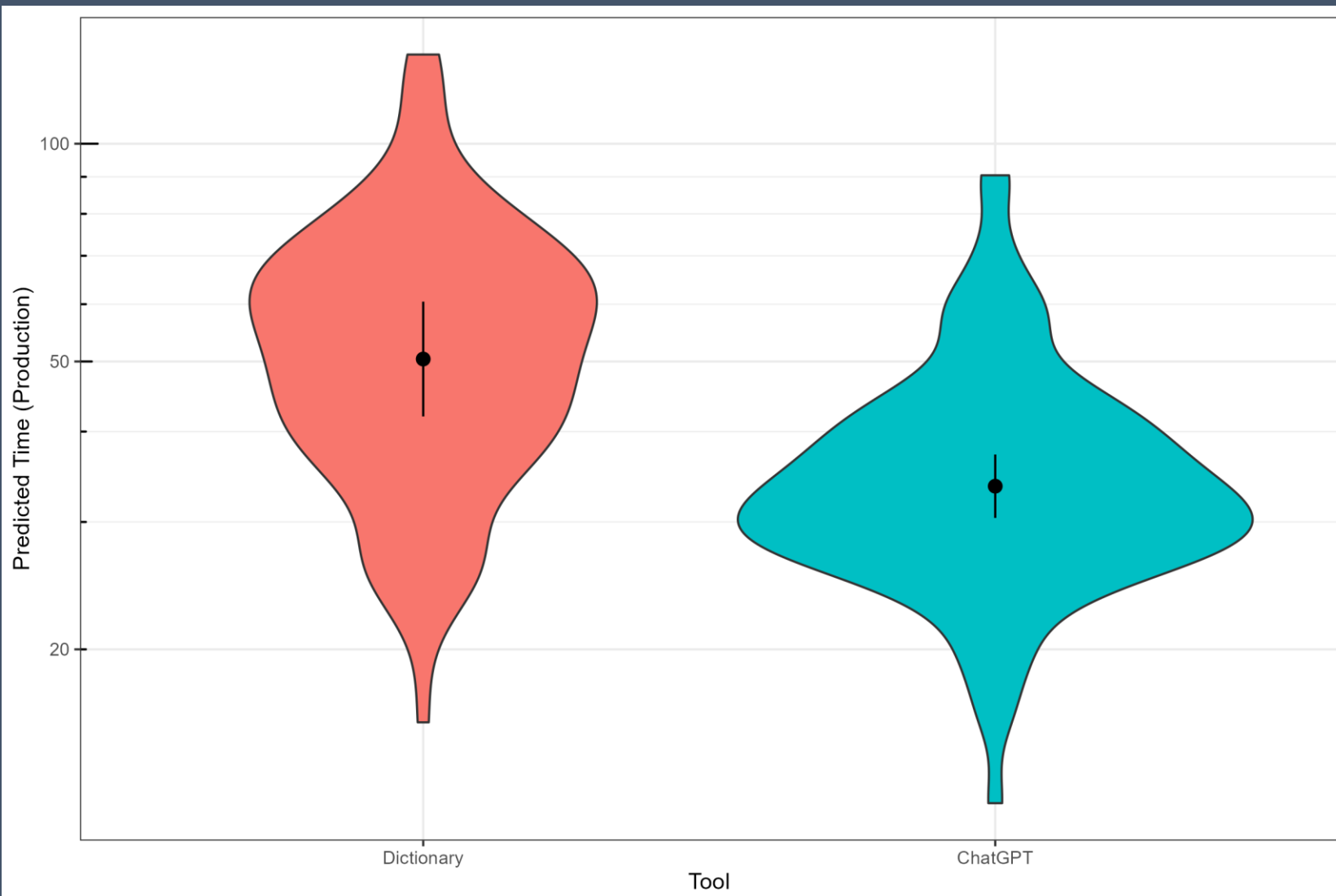
Test item	Target structure
turn	turn over something
try	try out for something
take	take against somebody/something
see	see something out
run	run up against somebody/something
play	play up to somebody
pick	pick at something
patch	patch somebody/something up
make	make off with something
kick	kick in
hit	hit somebody up for something
give	give out
fix	fix somebody up with somebody
draw	draw yourself up (to your full height)
do	do something over
demand	demand that somebody (should) do something
conceive	conceive of somebody as something
carry	carry something off
bump	bump somebody off
break	break somebody of something



Predicted success: production LDOCE versus ChatGPT

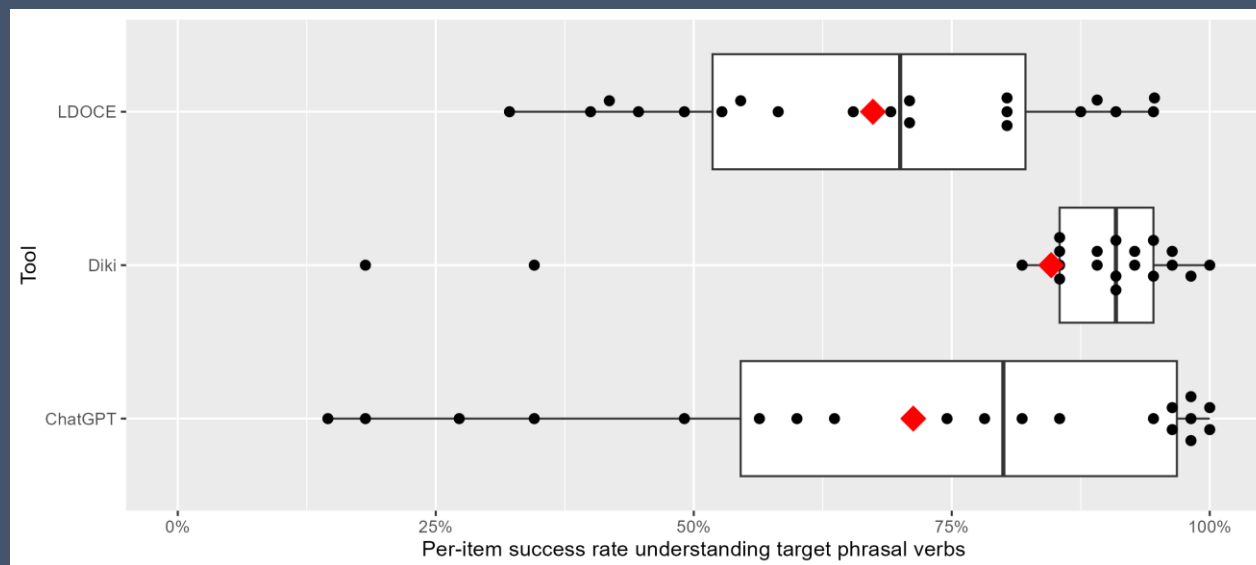
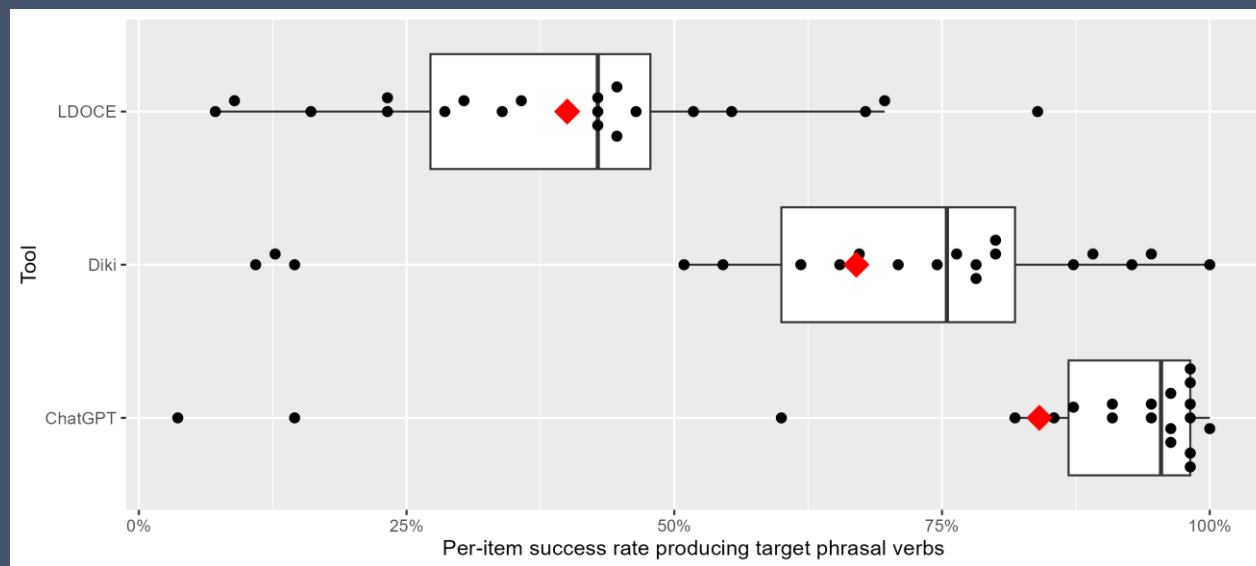


Predicted time (production) LDOCE versus ChatGPT



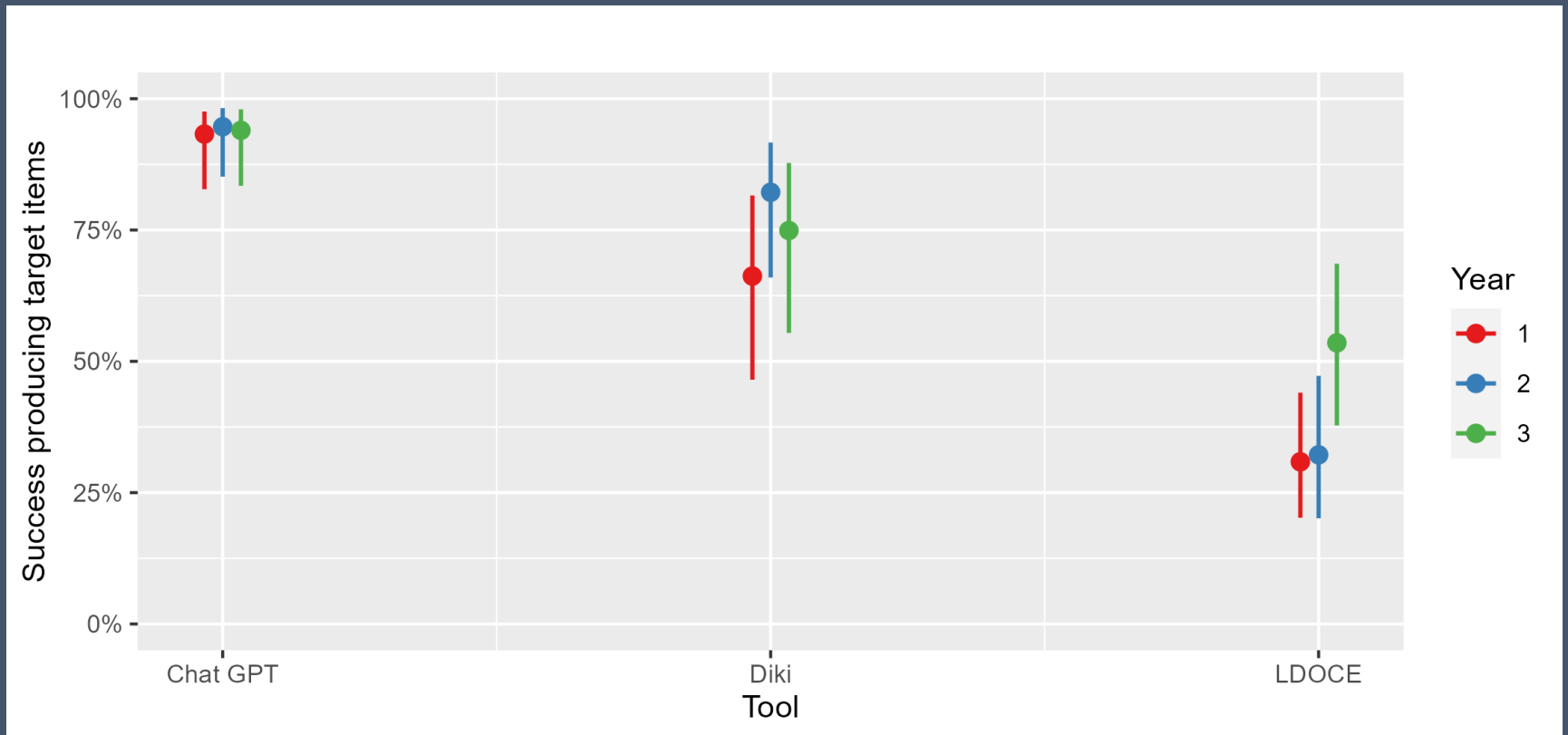
Study 3: Lew, Wolfer & Ptasznik (under review)

LDOCE vs Diki.pl vs ChatGPT 3.5



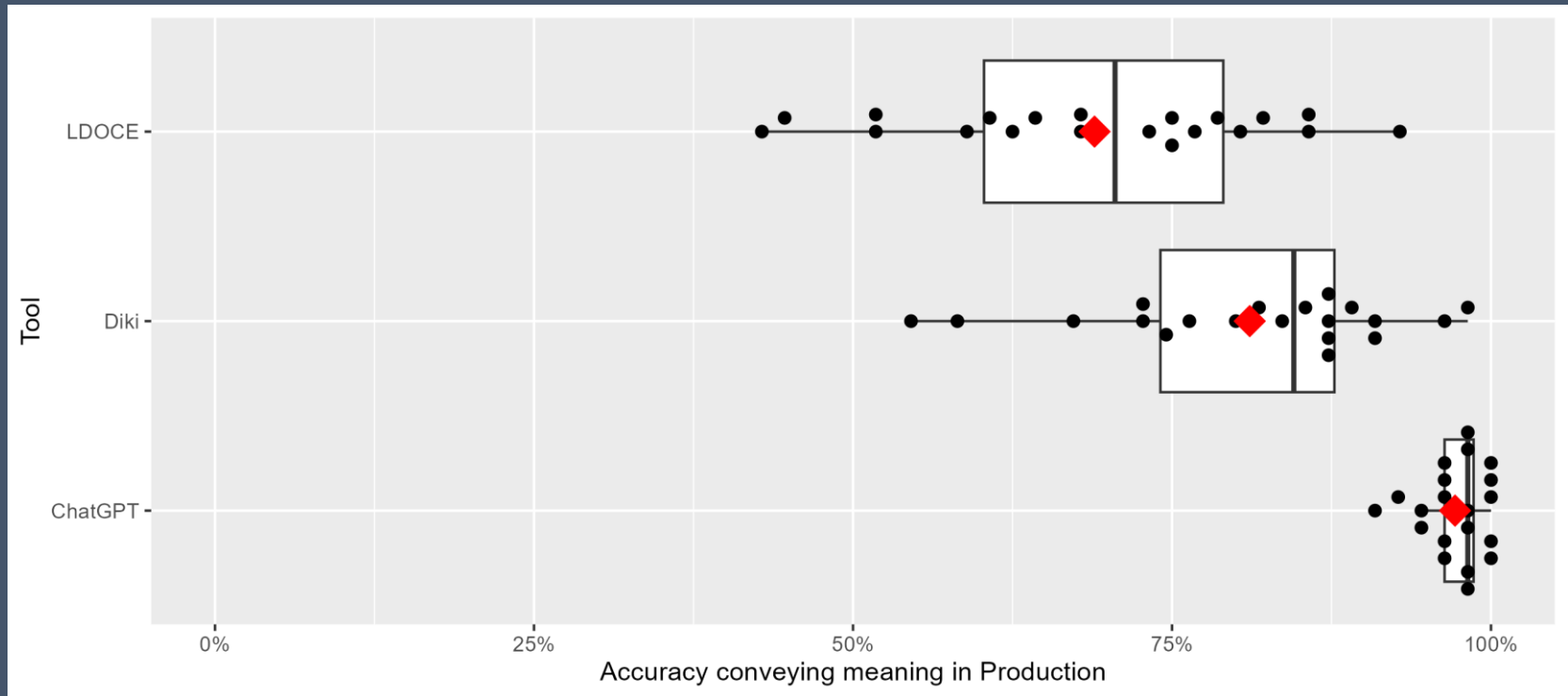
Study 3: Ptasznik, Wolfer & Lew (under review)

LDOCE vs Diki.pl vs ChatGPT 3.5



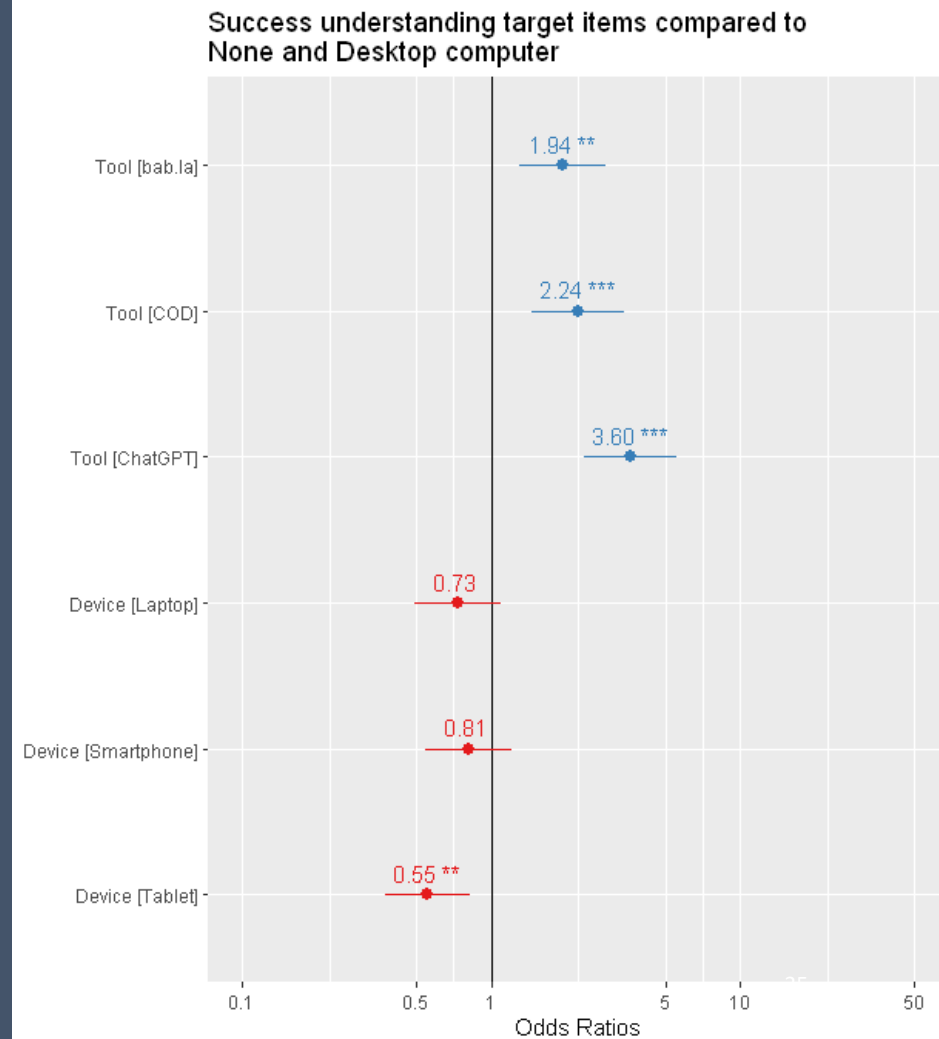
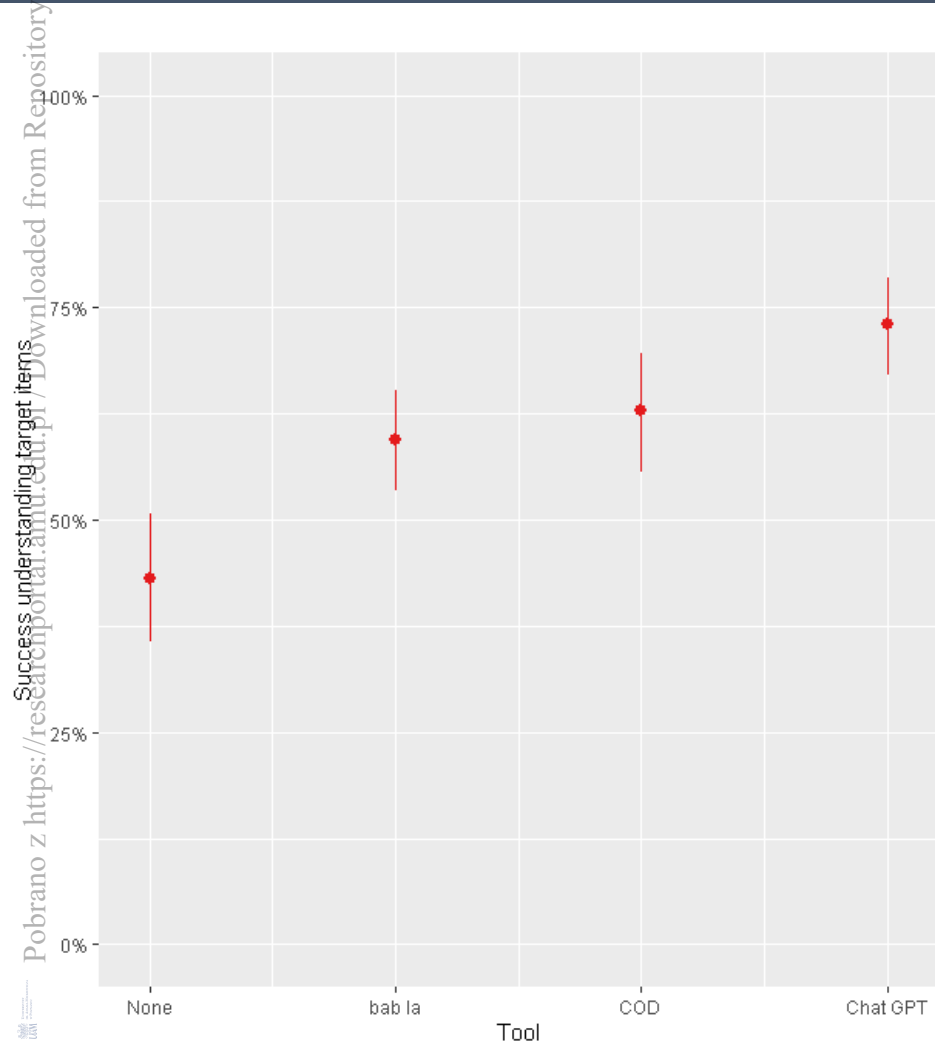
Study 3: Ptasznik, Wolfer & Lew (under review)

LDOCE vs Diki.pl vs ChatGPT 3.5



Study 4: Lew, Wolfer & Ptasznik (in progress)

Control vs bab.la vs Collins vs ChatGPT 3.5



ChatGPT 3.5 on the future of dictionaries

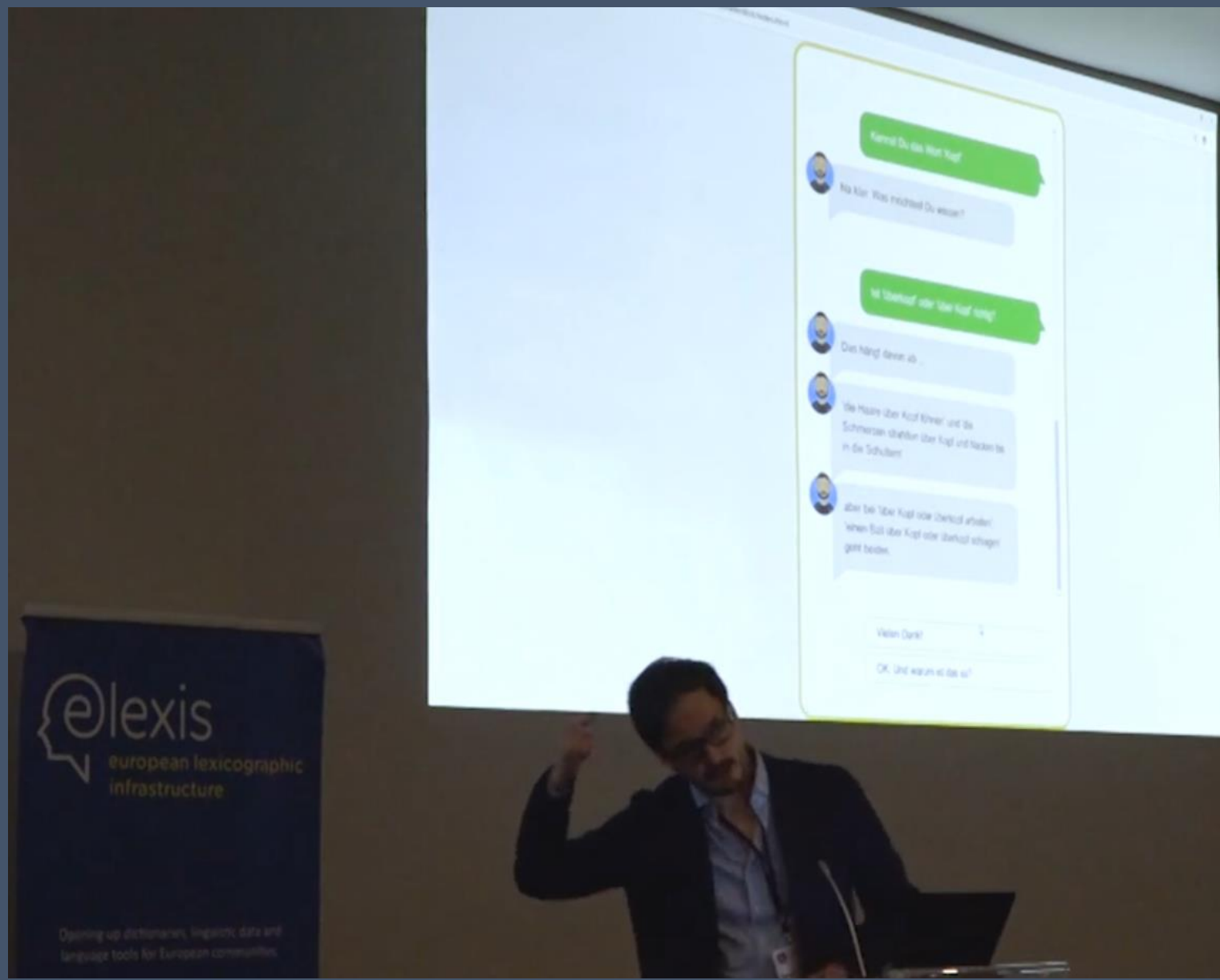


Answer the question "What will dictionaries be like in 2100, if at all?" in about 100 words. Set a high temperature parameter of 0.9.



In the fascinating world of 2100, dictionaries will be transformed into captivating linguistic ecosystems. Powered by advanced AI, these digital marvels will offer immersive language experiences, transcending traditional definitions. Augmented reality will project holographic words into our surroundings, making learning and communication a sensory delight. These dynamic dictionaries will adapt to real-time linguistic shifts, fostering creativity and aiding writers with fresh expressions. They'll decode emotions, cultural connotations, and nuances, enriching cross-cultural understanding. Interactive language games and simulations will make learning enchanting. However, the traditional notion of dictionaries may vanish as AI models integrate seamlessly into conversation, rendering standalone dictionaries obsolete.

Michaelis, Müller-Spitzer & Wolfer 2019





ADAM MICKIEWICZ UNIVERSITY IN POZNAŃ

Faculty of English

Thank you!



wa.amu.edu.pl

Image generated by Aria