International Journal Of Lexicography, 2023, **36**, 355–387
https://doi.org/10.1093/ijl/ecad021
Advance access publication 3 October 2023
Review Article

OXFORD

Generative AI and Lexicography: The Current State of the Art Using ChatGPT

Gilles-Maurice de Schryver[1,2,]

[1]Ghent University, UGent Centre for Bantu Studies (BantUGent) & Language and Translation Technology Team (LT[3]), Belgium
(gillesmaurice.deschryver@UGent.be)
[2]University of Pretoria, Department of African Languages, South Africa

Abstract

In this article, all ten papers and talks that have been devoted to the use of ChatGPT in lexicography so far are critically analysed, their results tabulated and cross-compared, from which the leading trends are determined. Extrapolating from the trendlines, a single short but robust new prompt is fine-tuned with which articles from different word classes are generated fully-automatically for a dictionary which compares favourably to the best practice in dictionary compilation. The conclusion is that a new age, that of the successful application of generative AI in lexicography, has dawned.

Keywords: Lexicography; Artificial Intelligence (AI); Large Language Model (LLM); Generative AI; Generative Pre-Trained Transformer (GPT); Chatbot; ChatGPT; Bard; Bing Chat; Claude; Stable Diffusion; Prompt; Bias; Hallucination; Black Box; Non-Deterministic Output; Memorisation; COBUILD; Full-sentence defining style; Authentic-like example sentences

Dedicated to Patrick Hanks

Prologue

> The first thing Patrick said to me was 'I want to die.' This was 1983. He had just arrived to bring some sense to the intellectually exciting but organizationally-challenged COBUILD project […] — and had presumably just been looking at his in-tray. (Michael Rundell, quoted in de Schryver 2010a: 8)
>
> Less than a year before the end of the project, [Patrick Hanks] judged (to everyone's consternation) that the entries that had been drafted so far were riddled with unacceptable vagueness about the relationship between definitions and definienda, and that this would confuse learners. As a result […] all the definitions were rewritten during the final editing phase in the now familiar COBUILD style of 'full-sentence definitions' […] (de Schryver 2010a: 9)

1. Position statement on the use of ChatGPT in lexicography

On 30 November 2022, ChatGPT, an artificial intelligence (AI) chatbot was released to the public. Barely a month later, it already had 100 million users. Since then, it has looked as if simply everyone, in every field, has been wondering if and how this technology could be used in (and to the benefit of) their discipline. In education, universities scrambled to police

its use, from outright bans at one extreme of the spectrum, to detailed suggestions on how to imbed it into teaching and research on the other. Some people were angry; many more were exhilarated. Some people dismissed it; many more feared its power and announced the end of humanity. In linguistics, luminaries such as Noam Chomsky have concluded that 'we can only laugh or cry at [the] popularity' of large language models (LLMs) such as ChatGPT (Chomsky et al. 2023).[1] Closer to lexicography, colleagues like Piek Vossen have derided ChatGPT as 'a waste of time' (Vossen 2022).[2] The best way to judge whether ChatGPT is of any use in one's own field is to give it a serious try, and this is exactly what nearly two dozen of us have done over the past half year. My own stance may be seen in the following position statement, proclaimed on 28 June 2023 at the eLex 2023 conference, during a round table on large language models and AI in lexicography:

> In a Tokyo talk last February, I summarised my position on the use of ChatGPT in lexicography with 'The end of lexicography, welcome to the machine'. I still stand by this claim.
>
> In reply to Michael Rundell's rebuttal at last week's ASIALEX conference in Seoul — in which he concluded that 'ChatGPT does not herald the end of lexicography' — I can now add that:
>
> (1) I believe that ChatGPT makes *dictionaries* redundant;
> (2) I believe that ChatGPT makes *lexicographers* redundant; and
> (3) I believe that ChatGPT makes the current *post-editing lexicographic tools* redundant.

I say this, and I believe this, not because it is true, but because the mere existence of ChatGPT gives us the illusion that this is possible.

> (1) If dictionaries were not already redundant in the era of mere search engines, they certainly are in today's age of AI chatbots. [Think: 'Bing Chat'.]
> (2) As it stands, the dictionary writing system TLex (aka TshwaneLex) now has an OpenAI section, where users may enter their 'OpenAI secret key' for any of the OpenAI functionality to work. Using either built-in default prompts or their own custom prompts, entire dictionaries may now be compiled — literally overnight, and in beautiful structured XML — without any further intervention, making lexicographers redundant.
> (3) Such a single prompt instruction is certainly an improvement over the current semi-automated tasks of corpus building, corpus annotation, headword-list creation, word-sense divisions [Think: Word Sketches and word embeddings], the pinpointing of salient collocations, the creation of definitions and/or translations, the selection of corpus-derived examples [Think: GDEX], the addition of related words, etc. — all of which need human intervention.
>
> Of course, this brings us to the quality of such a fully-automated 'product'. Well, in an age of 'invisible lexicography', where users treat their smartphones as black boxes that can do everything and anything, the quality may not matter anymore. If users performed AI-enhanced searches, all they wanted was an answer, and they got that.
>
> Further, if publishers — some may perhaps prefer to call them rogue publishers — want to release a dictionary without any human intervention, given it is now possible, they will do it, it will happen (and it already has happened).
>
> Therefore, if we still want to meet in two years from now at a, dare I say, hypothetical eLex conference, we will need to start taking ChatGPT more seriously, and treat it as a fully-fledged lexicographer. If we still want to try convincing the general public that humans are actually better than LLMs, we will have to backtrack and insist that we went back to the art and craft of compiling dictionaries, doing much more manually than is actually the case.

(First position statement in de Schryver et al. 2023)

Now, why do we believe that this time is different? We should certainly be careful in predicting the future. For example, no one doubts that the future of lexicography is digital, so when Macmillan announced on 5 November 2012 — as one of the first major publishers of reference works — that their dictionaries would 'no longer appear as physical books' as of 2013 (Rundell 2012c), it all made perfect sense.[3] But what to say now that the *Macmillan English Dictionary*, *Macmillan English Thesaurus* and *Macmillan Dictionary Blog* websites have all closed down on 30 June 2023 (Anon. 2023)? That made for a very short-lived future indeed: at ten years, 'Digital-Only Macmillan' didn't even turn a teenager — and now there is nothing left.[4] Likewise, over the past few years, we have been treated to the hypes and busts of blockchains in general, and cryptocurrencies in particular, the virtual-reality metaverse, and the blip of non-fungible tokens.

Therefore, to assess where we stand with a generative AI tool like ChatGPT in lexicography, we will first briefly introduce the tool (Section 2), then critically assess all lexicographic studies that have made use of it so far (Section 3), followed by a discussion as well as a new study (Section 4), and conclusions on the way forward (Section 5).

2. ChatGPT: what it is and how it works

ChatGPT is a language model developed by OpenAI that uses deep learning to generate human-like responses to natural language queries.[5] Specifically, ChatGPT is based on the 'Generative Pre-trained Transformer' (GPT) architecture, which is a type of neural network that can learn patterns in large datasets and generate new text based on that learning. ChatGPT has been trained on massive amounts of text data, specifically the 'Common Crawl' dataset, which includes over a trillion words from a wide variety of sources on the Internet. The model has been trained on multiple versions of this dataset, starting with smaller versions (such as GPT-2, which was trained on 40GB of data) and progressing to larger versions (such as GPT-3, which was trained on 175 billion parameters, or roughly 700GB of data). The training process involves exposing the model to vast amounts of text data and using it to predict the next word or sequence of words in the text. The model is then fine-tuned on specific tasks, such as language translation, question-answering, or chatbot-style dialogue, to improve its performance on those tasks. Overall, the size of the training data used to train ChatGPT is one of the factors that makes it so powerful and capable of generating high-quality responses to a wide variety of natural language queries.

The main reason for Piek Vossen's derision of ChatGPT (see Section 1) seems to be encapsulated in his claim that 'ChatGPT is not a revolutionary new system or a different approach. It just builds on previous versions and is somewhat better' (Vossen 2022). True, 'deep learning' has been around since the 2010s, and indeed software development is incremental (as is all progress), but the huge leap that ChatGPT represents is that it put the technology in the hands of everyone with an Internet connection without the need to understand anything about the underlying model or the need to interact with the underlying GPT engine as a programmer would do. One may simply 'converse' ('chat') with the engine via ChatGPT, and a *conversation* is, after all, the most natural thing humans do.[6]

A conversation with ChatGPT occurs using natural language; and the plain texts and queries a user types in are known as 'prompts'. Explanations on how ChatGPT works abound in the literature and online; a good starting point is OpenAI's own introduction (Schulman et al. 2022), another one was published in *The Economist* (2023c). What follows is based on the latter source. All the words of a prompt are first split into chunks of characters which commonly occur together, these are known as 'tokens' and can be plain words, affixes and punctuation. When ChatGPT was launched in November 2022, GPT-3.5 could process up to 4,097 tokens at a time, the current GPT-4 processes up to 32,768 tokens at a go (where 1,000 tokens is about 750 words). Each token of a prompt is then placed into

a 'meaning space', so a word such as 'promise' ends up embedded into a thesaurus-like list with 'aptitude, talent, potentiality, ability, potential, capability, *promise*, capacity'. In a third step, the LLM makes connections between the tokens of a prompt, deploying its 'attention network'. In practice, tokens are converted into 'numbers', while 'weights' are used to encode the structure of the language. The fourth step is the 'completion':

> At this point, for each of the tokens in the model's vocabulary, the attention network has produced a probability of that token being the most appropriate one to use next in the sentence it is generating. […] The LLM generates a word and then feeds the result back into itself. The first word is generated based on the prompt alone. The second word is generated by including the first word in the response, then the third word by including the first two generated words, and so on. This process — called autoregression — repeats until the LLM has finished. (The Economist 2023c)

Now, surprisingly, what one gets is not entirely predictable, in that the same prompts will give different (but similar) outputs. LLMs are thus 'non-deterministic'. In general, the bigger the models (more training data, and longer prompts) the better the output. A good illustration is that GPT-4 passed the American Uniform Bar Examination with flying colours, in the 90th percentile!, while GPT-3.5 had failed it (Arredondo 2023). Bigger is not always better, however, as it becomes increasingly more difficult for large models to counter 'undesirable *representational biases* — harmful biases resulting from stereotyping that propagate negative generalizations involving gender, race, religion, and other social constructs' (Liang et al. 2021). LLMs are also prone to another big problem: 'hallucinations', whereby the models generate texts that are factually incorrect, and they do so — and that is the real issue — with a straight face.

ChatGPT, built by OpenAI (with massive funding from Microsoft) using their GPT-3/GPT-4 LLMs, is not the only generative AI chatbot. Two others are Bard, built by Google (part of Alphabet) using their LaMDA LLM, and Claude, built by Anthropic (being former OpenAI alumni, in partnership with Google) using their AnthropicLM. And of course generative AI is not limited to outputting text, as images, audio, video and any other media (like 3D) or combinations thereof may be generated following mere prompts. Most of the 'fake' images going around, for instance, have been produced with Stable Diffusion, Midjourney or DALL-E. And then you have setups in which chatbots are used to control yet other components, such as HuggingGPT, 'a framework that leverages LLMs (e.g., ChatGPT) to connect various AI models in machine learning communities (e.g., Hugging Face) to solve AI tasks' (Shen et al. 2023), or TaskMatrix.AI (Liang et al. 2023), 'a chatbot that can interact with music services, e-commerce sites, online games and other online resources' (The Economist 2023b). But let us now come back to earth, and look at the potential of an LLM like ChatGPT with lexicographic purposes in mind.

3. ChatGPT and lexicography: The first six months

Lexicographers have always been at the forefront of dreaming up ways to harness the latest technology in order to compile better dictionaries — see de Schryver (2003) for the buzz 20 years ago, and de Schryver (2024) for an update written just prior to the release of ChatGPT. Given that an LLM is inherently more language-oriented than any (language) technology that came before, the hopes and expectations of the lexicographic community were understandably sky-high. These range from a team that didn't wait to see the full and final *proof* of its usefulness, and quickly went ahead to write the necessary code so that their dictionary writing system (DWS) could interact with OpenAI's GPT-3 (de Schryver and Joffe 2023),[7] to a meticulous statistical study of the quality of COBUILD-style full-sentence

definitions as offered by ChatGPT (Lew forthcoming). An overview of all ten studies so far is presented in Table 1, followed by annotated summaries of each.

3.1. The end of lexicography, welcome to the machine (de Schryver and Joffe 2023)

Over the course of one week in February/March 2023, we gave five lectures on lexicography in Tokyo, Japan. The first one, presented together with the creator of TLex, David Joffe, was devoted to ChatGPT (de Schryver and Joffe 2023). In this lecture we start by pointing out the difference between using a search engine and conversing with ChatGPT when in need of lexical or encyclopaedic information. Using the famous Tokyo broadcasting tower

Table 1: The first ten studies on the use of ChatGPT in lexicography

Date	Title	Who?	Occasion	Available?
27 February 2023	The end of lexicography, welcome to the machine: On how ChatGPT can already take over all of the dictionary maker's tasks	de Schryver & Joffe	The Tokyo Lectures on Lexicography	YouTube
31 May - 3 June 2023	Defin-o-bots: Challenging A.I. to create usable dictionary content	Barrett	DSNA 2023	—
22-24 June 2023	Automating the creation of dictionaries: Are we nearly there?	Rundell	ASIALEX 2023	Proceedings
22-24 June 2023	The return on investment of AI in lexicography	McKean & Fitzgerald	ASIALEX 2023	Proceedings
27-29 June 2023	The end of lexicography? Can ChatGPT outperform current tools for post-editing lexicography?	Jakubíček & Rundell	eLex 2023	Proceedings & YouTube
27-29 June 2023	Exploring the capabilities of ChatGPT for lexicographical purposes: A comparison with OALD within the microstructural framework	Phoodai & Rikk	eLex 2023	Proceedings & YouTube
27-29 June 2023	Definition extraction for Slovene: Patterns, transformer classifiers and ChatGPT	Tran, Podpečan, Jemec Tomazin & Pollak	eLex 2023	Proceedings & YouTube
27-29 June 2023	Round table on 'Large language models and AI in lexicography'	de Schryver, Rundell, Tavast, Rychlý, Kokol (panellists) & Krek (moderator)	eLex 2023	YouTube
27-29 June 2023	Invisible lexicographers, AI, and the future of the dictionary	Nichols	eLex 2023	YouTube
1 July 2023* [*latest version; first version 12 June 2023]	ChatGPT as a COBUILD lexicographer	Lew	Humanities and Social Sciences Communications	Pre-print

'Skytree' as an illustration, we immediately conclude that ChatGPT is far more superior: it gives better answers, in an intuitive way, and the output may be regenerated until one is satisfied. One could have stopped there: Exit Lexicographer, Enter Machine. However, we went back to the basics of lexicography, having ChatGPT generate full dictionary entries for both real as well as invented words and concepts, with and without supporting context. For the invented material (such as for 'Internet-transfer-based epigenetic inheritance'), ChatGPT clearly hallucinates = invents 'facts'. In a next series of tests, we let ChatGPT chat with itself, also recursively down to a number of levels. This then morphs into tests to have ChatGPT write programming code, first for casual aspects, then for single dictionary entries, then for a number of randomly-generated entries, using an XML template, with ChatGPT chatting with itself, to end with increasingly better-formatted dictionary material in structured XML, as well as JSON, YAML, CSV, TEI XML, and specifically for terminology even TMX and TBX, and finally XLIFF which is used in translation/localisation. Armed with all this knowledge of what ChatGPT can already do, and with a feeling of which prompts return relatively good results, AI functionality was added to TLex using OpenAI's GPT-3 API.

This new functionality is then demonstrated live by David Joffe, and exactly one hour into the lecture a public online dictionary (at https://dictionaryq.com/GPT-dictionary/) is updated in real time with the new entries generated by the AI a few minutes prior. Figure 1 shows the new OpenAI section in TLex: Apart from the default prompts, especially important is the fact that one can additionally set up one's own custom prompts (which is any prompt of your own, making the tool a lot more powerful, to for instance cover languages other than English, to focus on LSP aspects, etc.), and the option to allow 'batch processing' (feed it a list of headwords and have it automatically create articles for all of them).

In a discussion section, brief consideration is given to the authorship of such (future) AI-generated dictionaries, colloquially stated as: 'Bob or ChatGPT? Or both? Or should the full history be shown if/when the ChatGPT output is amended by Bob?'[8] Other questions pondered include *'Should "allow integrated AI functionality" be a privilege in user management in TLex?'* and *'Might editors want to block lexicographers from using it?'*

In addition to the use of ChatGPT for the actual compilation of dictionary articles, the talk also spent some time on the use of ChatGPT during the pre-compilation phase, namely to generate mini starter corpora, which could be especially useful for LSP lexicography. Lastly, with regard to the post-compilation phase, examples were given of how ChatGPT can be brought in during publishing, sales (including marketing) and support, and even metalexicography given that ChatGPT summarises texts, has opinions, checks grammar/

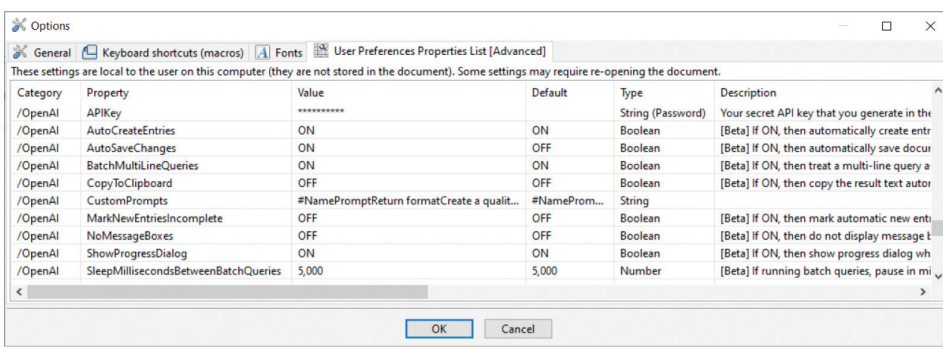

Figure 1: OpenAI section in TLex, where users may enter their own 'OpenAI secret key' for any of the OpenAI functionality to work.

spelling, does keyword extraction, … all aspects metalexicographers need,[9] and it can write on a meta-level well enough to fool most of our colleagues already.[10]

In conclusion we suggest to accept that AI has matured, and to use it. Rather than moan about the dangers of AI (which are real!), one should jump on the wagon and use the technology, to free up time for us humans to do more useful things. If we don't, future human lexicographers will be forced to stop claiming anything digital was involved, will have to stop claiming that their dictionaries are based on 'real language' — as seen in corpora, and processed with corpus query software, and synthesised by in-the-flesh-lexicographers as started by COBUILD in the 1980s (Sinclair 1987b). Rather, they would have to insist that their dictionaries were produced the good-old-fashioned way, i.e. 'manually'. It is better to let 'the machine' do the bulk of the work, with human intervention only at the vetting stage. In short, augment the production rate and quality of lexicographic products with AI.

3.2. Defin-o-bots (Barrett 2023)

At the DSNA 2023 conference, Grant Barrett (2023) gave a report on his attempts to use ChatGPT to write dictionary definitions, give etymologies and compose examples, and Stable Diffusion to generate (dictionary) illustrations. Disclosure: Prior to his talk, Barret had watched the recording of our Tokyo lecture on ChatGPT (see Section 3.1), and he also refers to it in his wrap-up. In a recent e-mail he stated: 'while we did cover some similar ground, I think our approaches are complementary' (Barrett, personal communication, 15 July 2023). Indeed, his experiments with generating images using AI are intriguing, especially so that no other study in lexicography so far did go there.

Rather than worry about or laugh at ChatGPT, Barrett suggests to simply view it as another tool. He also contends that, for the best results, there should be human input at the beginning (to arrive at the optimal prompt via trial and error) and at the end (to clean up the output). His first attempts involve the writing of definitions for 'mirusvirus', a term that entered the English language in April 2023. He does not get anywhere with ChatGPT, which is not surprising given that the data collection for GPT-4 ended in September 2021. ChatGPT offers to define 'virus' instead, and this leads to the highly unexpected realisation that no less than four dictionaries are plagiarised in the process. In discussing this outcome, Barrett offered: 'I wonder about dictionary definitions, which are prone to appearing all over the Internet, especially when licensed, and how much they "poison the well"' (Barrett, personal communication, 15 July 2023). We agree, as it seems GPT-4 took phrases like 'a submicroscopic infectious agent that is unable to grow or reproduce outside a host cell' wholesale, as it occurs very many times (here 104 times according to Google, after drilling down and actually going to the last page, including 'similar' pages) exactly as such on the Internet. This must push the predict-next-word model used in the LLM to 'memorisation' and thus to output the repetitive input verbatim.[11] In the end, Barrett resorted to Google's Bard to obtain a set of decent definitions for 'mirusvirus', stratified for college, unabridged and scientific dictionaries. Bard also generated acceptable (though wordy) etymologies for 'mirusvirus', as well as for the homograph and homophone 'loo' (in the sense of 'strong, dry summer wind in India', a loan from Hindi), although the quotations in support seem to be hallucinations. The etymologies offered for a third test word, however, the slang and slightly offensive 'shithousery', were all useless, whether generated by ChatGPT, Bard or even Bing Chat (basically Microsoft's search engine Bing combined with OpenAI's GPT engines). Further fine-tuning of the prompt in ChatGPT resulted in a tolerable etymology here, with lots of caveating and hedging. When it comes to the generation of example sentences, the results for 'shithousery' using ChatGPT, and 'loo' using Bard are poor (only about three out of ten are proper exemplifications with enough context), while the examples for 'monkey' seem rather good using ChatGPT, but very bad with Bard.

Throughout the talk, the various test words are illustrated with AI-generated images.[12] One of the most fascinating moves is when Barrett uses ChatGPT to generate good image

prompts; prompts which he then feeds into Stable Diffusion to generate images. In a section on intellectual property rights, Barrett laments the fact that tools to detect whether something was AI-generated or not, seem not to work very well. In conclusion, he would like to see a situation where 'our existing careful corpora [are] being used as the text sources for custom-built models for generating example sentences and definitions using the generative AI techniques, and then subjected to the usual lexicographical scrutiny'. He also fears the day when LLM output found on the Internet will be fed into future LLMs as input, 'leading to feedback loops, inbreeding, and monoculture'.

3.3. Automating the creation of dictionaries (Rundell 2023)

During the opening keynote lecture at the ASIALEX 2023 conference, Michael Rundell (2023) contrasted the current state of the art in dictionary compilation with his first impressions on the use of ChatGPT for lexicographic purposes. Disclosure: At various places during his talk, Rundell refers to and discusses our Tokyo lecture on ChatGPT (see Section 3.1). The title of his talk only differs in its sub-title compared to an earlier study: While he still wondered in 2011 'Where will it all end?' (Rundell and Kilgarriff 2011), he now wonders 'Are we nearly there?' (Rundell 2023). His conclusion is that for the time being (read: with the current state of LLMs like ChatGPT) we are not yet there.

He begins his argument by looking at the state of the field in 2011 (see also Rundell 2012b), noting that the post-lexicographic phase of publication, as well as the pre-lexicographic phase of corpus building and headword-list development, had both already been substantially automated. With regard to the central task of dictionary compilation, he reviews lexical profiling software (Word Sketches offered neat summaries; Kilgarriff and Tugwell (2001)), word sense disambiguation (not yet automated; Kilgarriff and Rychlý (2010)), example selection (the GDEX algorithm provided good examples; Kilgarriff et al. (2008)), labelling (only modestly automated), and the production of definitions (not automated). Moving to the next decade (2011-2022) he notes that 'the corpora used by lexicographers grew by an order of magnitude or more', that labels 'continued to resist easy automation', and that a new requirement appeared, namely, the need 'for the dictionary to be always up to date' (only partially achieved). Continued improvements to the Sketch Engine (Kilgarriff et al. 2004; Kilgarriff et al. 2014), allowed for a lot of automation during the compilation of two Slovene dictionaries (Kosem et al. 2014), and even more so for the three multilingual Lao, Tagalog and Urdu (SLs) to Korean cum English (TLs) dictionaries (Baisa et al. 2019; Jakubíček et al. 2021). The latter project is considered to be the 'first attempt at full-scale "post-editing lexicography" on a major project' and as such the state of the art in lexicography.

Using a number of test words, Rundell then looks at how ChatGPT handles meanings and definitions. He notes that the simple technical term 'carbon cycle' is well-defined, as is the adjective 'remiss'. For the polysemous 'party' the results are poor, and for the even more polysemous 'overwhelm' it is weak. When asked for a definition of 'presentation' it even starts with 'according to the Merriam-Webster dictionary …' while what follows is not in M-W — a clear (and highly annoying) hallucination. In a next series of prompts, he focuses on examples and grammar. The examples for 'fair', 'party', and 'command' are consistently bad. Looking at the output for 'aside' and 'haunt' he notes that ChatGPT mixes up the parts of speech.

In a concluding section, Rundell asks three important questions. For the first, *'Can ChatGPT successfully answer users' lexical queries?'*, he answers in the affirmative, but observes that existing resources can do this as well and even 'outperform ChatGPT in terms of simplicity and reliability'. We disagree with the 'simplicity' part, but given that we need 'trust' and 'curation' we definitely agree with the 'reliability' part. For his second question, *'Can ChatGPT generate good dictionaries with minimal human input?'*, he answers with a straightforward no. This is debatable as there are already several aspects of the dictionary-making process for which ChatGPT gives surprisingly good results that need no further

editing. With regard to his last question, *'Can ChatGPT outperform existing technologies in creating a draft dictionary for post-editing?'*, he points out that 'it would be hazardous in the extreme to rely on any large language model which did not allow access to the underlying data on which its output is based'. When post-editing lexicographic material with the currently available tools, it is always possible to see the 'proof' in the underlying (and linked) corpus data. Rundell judges this to be of paramount importance, and precisely because today's LLMs are black boxes, with their output non-deterministic at that, he concludes that ChatGPT, for the time being, does not herald 'the end of lexicography'.

3.4. The return on investment of AI in lexicography (McKean and Fitzgerald 2023)

During the second keynote lecture at the ASIALEX 2023 conference, Erin McKean's main contribution to the discussion of ChatGPT in lexicography revolved around a series of concerns she raised (McKean and Fitzgerald 2023). Before doing so, and just as the three previous studies had done (Sections 3.1, 3.2 and 3.3), she begins by reviewing ChatGPT's performance, guided by the assumption that: 'Since much modern lexicography is based on investigation and analysis of large-scale corpora similar to the corpora used to train LLMs, we hypothesize that LLMs could be used for typical lexicographic tasks'. Her first three tests, however, all lead to disappointing results: Provided with an incomplete macrostructure, ChatGPT performs poorly on suggesting missing headwords; Asked to give phrases with 'cut' ChatGPT also offers material that does not even include the target word ('Like a hot knife through butter', in addition to 'Cut and dried', 'Cut corners', …); and Tasked with giving forms for 'alter' ChatGPT fails to order the forms alphabetically (alteration *n*., alterable *adj*., alterably *adv*., alterationist *n*., altercate *v*., altercation *n*., alternating *v*.). In contrast, when fed with one up to two corpus lines for words like 'cheapfake', 'lash tab', 'psychobiome', 'claxonomy', 'booksona', etc. ChatGPT provides good definitions. Actually, even though only words without a Wiktionary definition were chosen, it is clear to us that more than just the information from the provided corpus lines is used to generate the definitions; so in addition to the prompt, information in the LLM itself is also recycled. The task to rewrite (simplify) definitions (from adult to child level) is less impressive, as the result is often clunky or awkward. Moving to exemplification, asking ChatGPT for exact first citations is reliably unreliable, while a prompt that contains both a headword *and* a definition returns reasonable examples that include typical collocates. Lastly, labelling ('vulgar', 'formal', 'derogatory', …) is more useful than anticipated, while the production of IPA pronunciation from respelled forms is wildly inconsistent.

McKean's first concern is an environmental one: She calculates that if ChatGPT were used for the quarterly OED updates, an extra 1,000 litres of water would be required each year (to cool the data centre which hosts the LLM). Her second set of concerns starts with bias: 'English-language text from the open web is not produced equally by all English speakers […] Using LLMs to produce dictionary text practically ensures that the language of non-white, non-male, disabled, queer, and poor people will be underrepresented in dictionaries, and that definitions created by generative AI are at risk of perpetuating harmful stereotypes'. She is also concerned about confabulation (inaccurate facts) and hallucination (invented facts), transparency (we do not know which texts went into the LLM), and reproducibility (results are non-deterministic). Thirdly, there are the industry concerns: 'By using LLMs [we] risk losing lexicography as a profession entirely'. Fourth, she has concerns about IP and copyright: 'memorisation' may lead to plagiarism (see Section 3.2), and the following opinion will surely lead to a legal quagmire: 'While some prompts may be sufficiently creative to be protected by copyright, that does not mean that material generated from a copyrightable prompt is itself copyrightable' (U.S. Copyright Office 2023). Fifth, there is the regulatory compliance: ChatGPT is available in the EU today, but may be pulled if regulations become unfavourable (Reuters 2023).[13] Six, there are concerns about

consumer confidence: while several dictionary brands are known for their 'authority' and 'reliability', there are no norms and customs yet for LLM output. In conclusion, McKean feels it is premature to go all-in on ChatGPT, rather, she suggests the community develops an evaluation taskset for dictionaries first, and also advocates the creation of lexicography-specific LLMs.

3.5. The end of lexicography? (Jakubíček and Rundell 2023)

At the eLex 2023 conference Miloš Jakubíček and Michael Rundell (2023) presented their observations after having fed 99 headwords from the DANTE sample list (Atkins et al. 2010) to ChatGPT.[14, 15] They begin by noting that the automation of word-list creation, example extraction and collocation selection developed alongside corpus development (Jakubíček et al. 2013). Left implicit, LLMs could then represent the ensuing step towards full automation of the lexicographic process. In a next section they discuss some technical aspects of ChatGPT and OpenAI's GPT-based models, paying particular attention to the concept of 'token' in LLMs (being used for subword tokenisation), and the non-deterministic nature of LLMs (same question, different answers). Even though the models are proprietary, and even though the link between (academic) publications and the actual products is often unclear, they list seven properties which they deem 'durable'. (1) Language modelling, not reasoning: 'The model does not perform any *logical* inference in terms of formal reasoning, nor does it build any kind of knowledge base of facts'. (2) Both the training and the inference are non-deterministic. (3) 'Once the model is trained, it is static' (cf. the frequent 'admissions' by the current version of ChatGPT that it 'knows' nothing after September 2021). (4) '[T]he model does not keep references to training sources', so if entire sentences happen to be cited verbatim, that is random and unintentional. (5) The length of prompts and their responses is limited, currently to about 500 – 1,000 words. (6) Over 90% of the training data is in English; prompts in any other language are answered 'via' English, and are thus patterned onto the (linguistic) characteristics and structures of the English language. (7) There is a need to fine-tune prompts, also between model versions, which is an obstacle to formalisation.

Reporting on their experiments, they point out that ChatGPT does not do well in terms of word-sense division, coming across as a 'splitter' (as opposed to a 'lumper' in lexicographic parlance). For 'climate', for instance, it lists six senses, where all of them should be lumped under a single meaning, while a second meaning was left out altogether. Next, they find that ChatGPT tends to define words like 'command *n*.' or 'efficient' more than adequately, though at times a key component is missing, such as at 'garden' and 'beach'. In terms of grammatical information, (in)transitivity is not always handled well and some parts of speech are wrongly assigned. When it comes to labelling, the results are good: 'half-caste' is marked as *offensive*, 'betimes' is identified as *archaic*. Lastly, for example sentences, it is 'as if they have been made up by a rather unimaginative human editor [and] predominantly follow the formula "3rd person subject with simple past verb"'.

In a concluding section, Jakubíček and Rundell list arguments in favour of ChatGPT: it is easy to use, it continues to be developed, it is affordable, it is multilingual, and the API can readily be integrated into a DWS. Furthermore, when a human converses with ChatGPT they can also challenge the bot in an attempt to improve the lexicographic response. And, they note, there is even one aspect for which ChatGPT exceeds the current state of the art, namely for synonym/antonym classification. That, of course, should not surprise, as it is part and parcel of the inner workings of LLMs (see Section 2, par. 2, step 2: 'meaning space'). They also discuss arguments against ChatGPT. First, and with a nod to Kilgarriff (2007), '*GPTology is bad science*': 'ChatGPT is using unknown data sources, with non-deterministic (and very likely soon-to-be-personalized) responses'. Second, ChatGPT reproduces bad lexicographic practice from an earlier age, such as 'lexicographese' (e.g. 'the act or state of X', 'characterised by Y', etc.), likely because

the free reference works that ended up in the LLM are those for which the copyright lapsed rather than contemporary ones. Third, 'ChatGPT and GPT-like models do not make back-linking evidence possible'. In conclusion, they are convinced that lexicographers are still needed to compile dictionaries, and would rather see the fine-tuning of LLMs for lexicography.

3.6. ChatGPT vs. the OALD (Phoodai and Rikk 2023)

At the eLex 2023 conference Chayanon Phoodai and Richárd Rikk (2023) set out to compare, for a selection of 50 top-frequent English headwords,[16] the types of microstructural slots ChatGPT could provide information for vs. the inclusion or not of such slots in the *Oxford Advanced Learner's Dictionary of Current English, Tenth Edition* (OALD-10, Hornby et al. 2020). Given that they are only interested in knowing whether or not ChatGPT can fill a slot, without taking the quality of what it produces into account, the outcome of a comparison with a real-world reference work like the OALD-10, onto which space-constraints are enforced, is wholly predictable: Of course ChatGPT will come out the winner! In their study, ChatGPT fills the slots 68% of the time, the OALD-10 just 57% of the times. But what does this outcome mean?[17] The most polite way to summarise this effort — which has serious methodological problems and even contains unexplained abbreviations[18] — is to say that this is much ado about nothing. To add insult to injury — the majority of us not being fans of Wiegand's so-called theorising (Rundell 2012a) — the point of departure is a selection of Wiegand's (1989) 'more than 200 classes of functional text segments that serve as structural indicators within the dictionary microstructure' (Phoodai and Rikk 2023: 338).

3.7. Definition extraction and ChatGPT (Tran, Podpečan, Jemec Tomazin and Pollak 2023)

Rather than write a definition from scratch, a popular method employed by terminographers, is to have software do the job by querying an LSP corpus. Over the years, very many rule-based (starting in the 1990s), machine-learning (starting in the noughties), and neural-network (starting in the 2010s) solutions have been developed for that goal. The computational task is to automatically identify definitions from unstructured running text, for which terminographers favour output with a high precision over high recall, or thus fewer definition-candidates for which the probability that they are also definitions is high (rather than the reverse: more candidates with a lower probability). With the arrival of ChatGPT, an obvious urge is to see whether this LLM can also help with that task, and this is exactly what Tran et al. (2023) looked into at the eLex 2023 conference. In their paper they evaluate three types of approaches to definition extraction in Slovene: a baseline pattern-based classifier (which is rule-based), four deep-learning transformers (so neural networks), and prompts submitted to ChatGPT. For the latter, sentences in Slovene that were either definitions or not had to be judged by ChatGPT, which was requested to reply with only 'Definition' or 'Non Definition'. The results show that when the LSP corpora only contain a limited number of well-structured definitions with clear linguistic signals, the baseline pattern-based classifier outperforms the other methods, but for LSP corpora with fewer clues, deep-learning transformers and ChatGPT prompting are more effective. When comparing the latter two, transformers lead to higher precision, ChatGPT to higher recall. So, while not yet there, this is clearly a very encouraging result.

3.8. Round table on ChatGPT (de Schryver, Rundell, Tavast, Rychlý, Kokol (panellists) & Krek (moderator) 2023)

In the wake of the release of ChatGPT, it was felt that a round table on large language models and AI in lexicography simply had to be organised at eLex 2023 — the most recent instalment of the biennial conference series which focuses solely on the digital aspects

of our field. Organised by Iztok Kosem and Jelena Kallas, and moderated by Simon Krek (SK), five experts were invited to participate. Each of the panellists was asked to state their position first. The moderator then framed the ensuing discussion around five questions/topics.[19]

3.8.1. Position statements

Our position statement was reproduced verbatim in Section 1 already; the next four are summarised here.

Michael Rundell (MR) is critical, arguing that with the current tools for dictionary compilation, we can always go back to the corpus, which is not possible with ChatGPT. For him, this is a deal-breaker: if LLMs remain a black box, this is not good; we really need access back to the source.

Arvi Tavast (AT) stresses that we need to distinguish between the engine (i.e. the large language model), and the chat interface. Engines are not new (GPT-3.5 is years old), we just became aware of them thanks to the fact that a chat interface (i.e. ChatGPT) was made available to the general public. He is surprised by the development so far, and expects the developments 'next week' to be as surprising. He then introduces some interesting terminology regarding user-interface styles in general. While 2D ('flat things') has been around for a long time (paper, computer displays, phone screens), 3D (augmented reality) was until recently the latest fad. However, he argues that we now entered the era of 1D (sequential user interfaces), and that sequential access is more intuitive for humans. It is simply more natural to 'just ask' what something means, rather than to have to open a book (or website) and to have to start looking up (or searching for) a word, to then seek the proper meaning, etc.

Pavel Rychlý (PR) argues that LLMs including ChatGPT are really just that, 'language models': they merely generate the most probable continuation of a prompt, no more, no less. For him, a good analogy is 'dictating' (you cannot go back, you dictate word by word). This leads to his position that LLMs cannot currently be used in lexicography, because even very good lexicographers cannot dictate without some thinking, consulting corpora, and so on. He is convinced that LLMs will never generate a dictionary entry from scratch. On the other hand, he believes LLMs may be used for small tasks, as we do now in post-editing lexicography, especially if the LLMs are trained/fine-tuned for that task. Lastly, he believes that that fine-tuning will eventually become available for lexicography.

Marko Kokol (MK) trusts that lexicographers are not going to be replaced: when an LLM is wrong, it is very confidently wrong. But, he adds, 'the part that scares me is that we do not really understand some of the emerging behaviour; while it is true that the model only predicts the next word, we do not really know why it sometimes sounds like it is making logical decisions (as it is not really designed to do that)' — so we might end up being replaced after all. He has high hopes that with open-source models (such as Meta's LLaMA), it will in future be possible to train specific LLMs that help us perform lexicographic tasks.

3.8.2. LexGPT

SK: 'How should the lexicographic community deal with or react to the arrival of LLMs and resulting applications? Does it make sense to think about a specialised lexicographic 'LexGPT' model and/or applications? If yes, based on what? Which data, which technology?'

MK: Regarding the data, we should recall that ChatGPT has been trained primarily on English corpora, so it is very important to start creating the data resources for under-resourced languages (like Slovene, with 2.5 million speakers) to train adequate modelling.

AT: The models need to speak our languages — in general, not just for lexicography. But know that people will require answers to questions about language, rather than requiring

a dictionary. Hence, these technologies should not be used as part of the lexicographic process, but as part of providing information to the users. Reformulated, it should not be lexicographers using these models, but users.

MR: But we already do that; for a quick fix we go to DeepL or Google; so ChatGPT looks more cumbersome to do the same. Yet we should not look at it in binary terms: Does ChatGPT replace all the things that we do now, or not? It is better to lower the expectations: Figure out what it does really well, and what it doesn't, and incorporate ChatGPT into the tools that we use at the moment.

G-M: Remember that ChatGPT is a general-purpose tool; it does everything for everyone, it passes exams on the weirdest topics, we are just this tiny slice of the population who work on dictionaries and we are already impressed. So, should we make a LexGPT? Of course! Feed it with good dictionaries rather than the freebies currently online. What is it good at right now? COBUILD-style definitions! Why? Well, it's natural language.

3.8.3. Specialised models, (retrained) smaller models, etc.

SK: 'What should lexicographers be attentive to as far as work with LLMs is concerned, which important developments? (e.g. specialised models, (retrained) smaller models, etc.)'

SK: For instance, the Swedes wanted to make their own LLM for Swedish, but found out that they do not have enough data, and then wondered whether they could include neighbouring languages that are similar. The minimum needed is 300 billion tokens and 40 billion parameters 'for Nordic languages'.

PR: For lexicographic purposes, you do not need such big models; tens of billions of words are enough, and these exist for many languages. Smaller models which are specialised for the task at hand also do not need to be that big, and do not need to run on supercomputers but may run on computers that can be rented for just a few hundred to a thousand dollars. The development of these language models is quite fast, especially in the open-source area, where they are even farther than at commercial companies like Google and OpenAI. In a few years from now, we will be able to train such models using much smaller resources than now; it will be easy to make a Lex-model.

AT: And then there is also 'fine-tuning', meaning that we do not need to build models from scratch, we can fine-tune the foundation models.

MK: Indeed, LLaMA as well as some quantised models can already be fine-tuned with consumer hardware (like NVIDIA 4090), but in our experience it is very hard to use for languages that have not been included in the original training. In other words, fine-tuning is fine when the model has seen the language during its training phase; if not at all, then the results can be pretty random.

AT: The original LLaMA from Meta has been trained on the Slovene Wikipedia, which is relatively small, and also on the Estonian Wikipedia, but Open LLaMA has only seen English, so it is a bit of a challenge.

3.8.4. Languages

SK: 'Languages other than English: Are LLMs and applications useful for smaller languages? How big is the usual digital divide between English >< German, French, Spanish, Italian >< all the rest?'

AT: We need to get rid of three things: copyright, personal data, and confidentiality, so that we can include everything in our corpora, and that will help our languages to survive. In other words, as much text as possible must be made available to the large commercial companies.

MK: Do not forget about the 'alphabet challenge', as extra characters raise the complexity of training the models.

G-M: From the point of view of the generated output by ChatGPT, there are no copyright issues (you cannot see who wrote what, it's a machine), and there are no privacy issues (there is no one involved, it's a black box). All you need to do is to build your billion-word corpora.

How? Let everyone speak at home, put recorders there, let it be transcribed automatically, and you have your billions of words (for Zulu, and Swahili, and all the other languages).

SK: Imagine you need a corpus, you order it from a model, and then use it for a lexicographic description. Or you want a new book by Saint Thomas Aquinas, you order it. Is that blasphemy?

MR: I would just like the system to be a lot more discriminating. Take 'cookie' in Urban Dictionary, for which there are dozens of different so-called meanings (most borderline pornographic). The system scraped all of that off Urban Dictionary, and presented all of it as if these are all real senses. There is something deeply wrong with that, especially in comparison with all the systems we have now, even quite automated ones.

SK: Imagine you have a completely under-resourced language, but from the data that you get across languages, you actually can produce a lot of data which is semi-OK for your language. Is that a good thing or a bad thing?

G-M: I can confirm that we did some tests, and the output for very basic words in some of the Bantu languages is OK, but once you want to do something beyond Primary 1, it starts to struggle, and very quickly it mixes up languages even. That is because it actually goes via English and then remaps it to the other languages. So it comes back to adding more data, for example in the way I suggested.

3.8.5. The near future

SK: 'What do you think can be expected in the near future, which technologies will prevail? To what extent is it possible to predict what will happen next?'

MK: We are at the top of a hype cycle. As with the metaverse, we are going to distil what is actually useful, and what isn't. I think we are going to see a lot of specialised models appearing.

PR: In one or two years from now, there will be many more models we can play with (maybe not for Bantu languages, but for most European languages), to use and fine-tune.

AT: I can confidently predict one thing. There will be things that will happen in the near future, which will be significant, and which we cannot predict now.

MR: It's such a rapidly moving target, isn't it? Yes, we cannot predict.

G-M: No, it's very easy to predict. There will be three types of dictionaries: (1) those produced fully automatically (and I know that these already exist, because our software is used for it; they are being sold, and no one knows that the machine made them); (2) those produced by more scrupulous publishers who will say, well we got help from ChatGPT and we modified the material; and (3) those for which dictionary makers will say, you know, this is compiled manually, we did not use computers, no corpora, I sat down with my team and for 20 years we wrote this by hand, so this is human, and this will then also sell because it was compiled by actual people and not a machine anymore.

3.8.6. Open LLMs in Europe

SK: 'What do you expect will happen with (open) LLMs in the EU/ Europe? EU vs. USA vs. China. (Also connected with this: EU AI Act, Google Bard not available in the EU, etc.)'

SK: The European Commission will invest to produce an open LLM, which will include all (regional) EU languages.

Miloš Jakubíček [from the floor]: The problem of the EU is that it's just throwing money all over the place, in an attempt to chase the US; it's not going to happen.

3.9. Invisible lexicographers, AI, and the future of the dictionary (Nichols 2023)

During the closing keynote lecture at the eLex 2023 conference, Wendalyn Nichols (2023) considered the arrival of OpenAI's ChatGPT with entirely different eyes, namely those 'of a vanishing breed, which is the head of an in-house dictionary team at a publishing company', adding that she consequently 'know[s] a little bit about survival and reinvention, and wresting

opportunities from threats'.[20, 21] The latter announces the bulk of her talk, which revolves around a SWOT analysis of ChatGPT relative to a learner's thesaurus. In order to collect arguments for SWOT's four quadrants (Strengths – Weaknesses – Opportunities – Threats), she looks at a number of conversations with ChatGPT and compares the information therein with the lexicographic data in the *Cambridge Thesaurus* (McIntosh 2023). Nichols recognises three main Strengths of ChatGPT: it is easy to use; it can provide the meaning of a word as it is used in a sentence; and it can synthesise vast amounts of data on demand. The first three Weaknesses of ChatGPT she lists are: it hallucinates; answers are too complex (for typical users); and one requires skill in writing prompts. She also claims that ChatGPT makes syntactic errors, and that it cannot easily deal with synonymy. We could not see those syntactic errors in the examples provided, so perhaps 'poor syntax' is what was meant. As to the complaint about synonymy, the goal was actually to *explain* detailed differences of near-synonymy; but this was not made explicit in the various prompts used. The last Weakness is that ChatGPT is only as accurate and current as the data it is trained on. Next, and from a publisher's point of view, Nichols sees four Threats. The first is that ChatGPT is free to use (which the *Cambridge Thesaurus* is as well, but supported by ads). The second is that answers by ChatGPT sound utterly plausible, and confidently so, but are often untrue: 'so it is a material threat to content producers, because we know that people largely do not bring their critical thinking skills to the active consuming of information on the Web'. The third Threat revolves around the age-old user need for context-sensitive answers whereas dictionaries only provide general context-free information (Varantola 2002: 33). The fact that ChatGPT can provide the meaning of a (polysemous) word as it is used in a sentence (the second Strength above), is thus simultaneously also an existential Threat to publishers: here is the Holy Grail of lexicography, effortlessly 'solved' by ChatGPT. The last publisher Threat, with reference to Barrett's finding of (occasional) wholesale copying (see Section 3.2), is that ChatGPT violates copyright law: 'someone cannot by law re-use content that may be legitimately licensed for the site where it originated, but isn't licensed for re-use'.

This then leads to the important question: 'How do we manage the *strengths* and exploit the *weaknesses* to mitigate the *threats* and make *opportunities* for ourselves?' Seeing that ChatGPT can provide the meaning of a word as it is used in a sentence (Strength) but also that its answers are too complex (Weakness), leads to the first Opportunity: the wish to develop a contextual search feature trained on *our* content. The fact that ChatGPT is only as accurate and current as the data it is trained on (Weakness), leads to the Opportunity to account for neologisms and a community effort by dictionary brands (which have active followings on social media) to 'lead the public conversation about language change'. The last area of Opportunity is to provide better answers for what bots are currently not good at, which leads to the suggestion to improve contextual advice as users write in real time, and the suggestion that 'AI can help users to curate their own experiences of online dictionaries, so that the sites are no longer generic, but truly "my dictionary"'.

Throughout her talk, Nichols also raises a number of concerns. Three stand out. First, the bots are generating content faster than regulatory agencies and legislators can keep up with.[22] Second, bots like ChatGPT are unable to make value judgements; they cannot analyse data and draw conclusions. Third: 'Publishers with storied brand names have the advantage of an association in users' minds with authority of content' (and by implication: ChatGPT lacks this authority). In conclusion, Nichols would like to see a situation where 'we ourselves can train AI on *our* good content, to retrieve information in imaginative new ways to improve our customer's experience'.

3.10. ChatGPT as a COBUILD lexicographer (Lew forthcoming)

In a preprint of a study which first became available on 12 June 2023, Robert Lew (forthcoming) offers a meticulous analysis of the accuracy and quality of COBUILD-style entries for English monolingual learner's dictionaries as produced by ChatGPT.[23] With a sample of fifteen verbs of communication,[24] his focus is on definitions, example sentences, and the entries as a

whole. Using long and detailed prompts, he first collected ChatGPT outputs, as well as the corresponding entries from the actual COBUILD dictionary (COBUILD 2023), and presented this material in a blinded randomised way to four human experts for evaluation, two of whom had worked on the original COBUILD dictionary and two of whom had not but were experts in English pedagogical lexicography. None of them was informed that half the material had been generated by AI; they only learned about this at the debriefing stage.

Starting with the first COBUILD dictionary (Sinclair et al. 1987), dictionaries in this series are well known for a number of revolutionary aspects (Sinclair 1987b), chief among them their full-sentence defining style which consists of two clauses, the first illustrating the usage pattern, the second a paraphrase of meaning (Hanks 1987; Barnbrook 2002). They were also the first learner's dictionaries for which all examples were taken verbatim from a corpus rather than being invented (Fox 1987). And in terms of overall and article structure they moved grammatical and other aspects to an Extra Column on the right-hand side of each page (Sinclair 1987a). When COBUILD morphed from paper to digital, that Extra Column disappeared (also in the print versions), but the grammatical and other information that had been separated out and made super-explicit, was kept in dedicated slots.

Lew's assumptions were that the conversational format of the full-sentence defining style should be easily emulated by ChatGPT (which is, after all, designed to converse), while for the example sentences the probabilistic nature of LLMs should 'result in natural word choices that are authentic-like'. Lew did not formulate a hypothesis on the use of ChatGPT for the entry as a whole, and as a matter of fact the syntax codes and lists of inflected forms were left out from his test entries. The evaluation of this third aspect, then, boils down to a combination of the evaluations of the quality of the definitions and examples (together with, in all likelihood, an evaluation of sense division and entry organisation).

The four lexicographers were asked to rate each sense definition and each sense example — whether human-generated or AI-generated (but that was thus unbeknown to them) — using a five-point scale. They could also leave open-ended comments on all three aspects (definitions, examples, entries). By means of statistical analysis using R, Lew proceeds with a very convincing investigation. Amongst others, he shows that 'for six out of the fifteen headwords, the sense definitions proposed by ChatGPT are at least as good as those written by human COBUILD lexicographers'. The AI-generated examples are found to be less satisfying, but Lew then hypothesises that this may be due to sub-optimal training, so he rephrases his prompt (which by now approaches the length of a page). ChatGPT's revised example sentences are indeed 'more elaborate and varied, use less boilerplate structures, and exhibit more variation in grammatical tense'. All in all, then, these are undeniably ground-breaking findings: ChatGPT definitions (of the COBUILD-style) are 'practically indistinguishable in quality from those written by highly trained human lexicographers' and with prompt engineering ChatGPT is 'capable of generating example sentences that are both authentic-sounding and accessible'.

4. Discussion, with particular reference to COBUILD

Even though (at just over half a year) the field of applying generative AI to lexicography is still young, ten dedicated studies and presentations have already been produced, involving nearly twenty different colleagues, whose work and thinking either exclusively or mainly revolves around ChatGPT. Interest has come especially from Europeans (12 colleagues), followed by North Americans (4), Asians (2) and an African (1). As the review of the papers and talks has shown, the topics covered are varied, and together encompass the entire process from pre-lexicography, over lexicography, to post-lexicography. Americans seem to be very interested in the monetary aspects, as well as copyright issues (Barrett; McKean & Fitzgerald; Nichols), Europeans tend to focus on dictionary quality (Rundell; Lew), while the application builders cover both extremes: the true believers (Joffe & team) versus the more cautious ones (Jakubíček & team).

With regard to lexicography proper, all the major macro- and microstructural components have been given attention: headwords, pronunciations, parts of speech, word senses, definitions, translations, examples, labels, synonyms and antonyms, etymologies, as well as illustrations. Underrepresented are proper multilingual studies as well as those in languages other than English.[25] Most attention went to definitions and examples. Overall, the studies found that the definitions generated by ChatGPT are good to very good; the results for examples range from poor to rather good following extra prompting. All of this is summarised in Table 2, which includes a new study in the last column (to be discussed further down, in the present Section 4).[26]

The lines in Table 2 that contain the most positive assessments so far on the use of ChatGPT in lexicography are the ones for definitions. And within that block, ChatGPT's emulation of the type of definitions found in COBUILD monolingual English learner's dictionaries seems exceptional (see Section 3.10). Given this outcome, it is highly pertinent to devote a further analysis to this aspect. The now famous COBUILD full-sentence defining style actually came about *accidentally* (Hanks, personal communication, 2009-2010).[27] After four years of filling 'the database'[28] with useful facts about how the English language is really *used* based on the material seen in the Birmingham Corpus (7.3 million running words at the start of the project, 20 million near the end), and with deadlines looming, the managing editor, Patrick Hanks, came up with a cunning plan to streamline the style of all definitions, which up to that point had been entered into the database in a variety of ways. Having meticulously studied corpus data, Hanks came to realise that 'All statements made about word meaning are statements about word use' (Hanks 1987: 135). A core goal of Hanks was to get that 'use' captured into the dictionary, over and above the mere listing of usage as seen in multiple example sentences (obviously taken from the corpus; cf. Fox (1987)) and the novel grammar slots (as found in COBUILD's Extra Column; cf. Sinclair (1987a)). Hanks's genius insight was to insist including word usage into the definition itself, and to put it up front at that. In his words: 'In general, then, the first part of each Cobuild explanation shows the use, while the second part explains the meaning'. This, of course, is easier said than done, so Hanks went back to the corpus data, and manually devised a 'large range of explanatory strategies' for that first part of each definition (or 'explanation' in Hanks's terminology). Re-reading Hanks (1987), these explanatory strategies must have numbered at least a dozen, as shown in (1) for different parts of speech.

(1) VERBS
— 'If you …' (with variants 'When something …' and 'If something …')
— 'To [verb] means …'
— 'When you …' (with variant 'If someone …')
NOUNS
— 'A [noun] is …'
— 'If you … [noun] to …'
FUNCTION WORDS
— 'You use …'
— 'If you do …'
PREDICATIVE ADJECTIVES
— 'Something that is …'
— 'Someone who is …'
ATTRIBUTIVE ADJECTIVES
—'[Adj] means …'
FIGURATIVE or METAPHORIC EXPRESSIONS
— 'If you say that …'
— 'If you call someone a …'
IDIOMATIC EXPRESSIONS
— 'If you say … [idiom], you mean …'

Table 2: ChatGPT in lexicography: The studies so far

	de Schryver & Joffe	Barrett	Rundell	McKean & Fitzgerald	Jakubíček & Rundell	Tran et al.	Round Table	Nichols	Lew	de Schryver
Section	§ 3.1	§ 3.2	§ 3.3	§ 3.4	§ 3.5	§ 3.7	§ 3.8	§ 3.9	§ 3.10	§ 4
ChatGPT for pre-lexicography										
Corpus building	possible for LSP						SK			
Definition extraction from corpus						promising for LSP				
Direct connection with a DWS, XML import	implemented in TLex									
ChatGPT for lexicography										
Suggest missing headwords				poor						
Suggest common phrases				poor						
Suggest related forms				poor						
Provide frequency information										good
Provide IPA pronunciation				poor						good
Provide part of speech			poor		so-so					good
Provide morphological forms										good
Provide word-sense division					poor					good
Provide word meaning as used in sentence								good		
Write definition for monosemous headword	good		good		good				good	good
Write definitions for polysemous headword	so-so		poor						good	good
Simplify definition				so-so						
Write definition, given corpus line(s)				good						

Table 2. Continued

	de Schryver & Joffe	Barrett	Rundell	McKean & Fitzgerald	Jakubíček & Rundell	Tran et al.	Round Table	Nichols	Lew	de Schryver
(Write definition using Bard)		good								
Provide translation equivalents	good									
Give first citation				hallucinates						
(Give first citation using Bard)		hallucinates								
Generate example		poor	poor		poor				so-so	good
Generate humorous example	good									
Generate example, given headword + definition				good						
(Generate example using Bard)		poor								
Translate example	good									
Quality of syntax								so-so		good
Label headword				good	good					good
Synonym/antonym classification					good					good
Deal with near-synonymy								poor		good
(Give etymology using Bard)		so-so								
Summarise a dictionary article										good
Compile non-English article					poor					
(Generate image using Stable Diffusion)		so-so								
ChatGPT for post-lexicography										
Publishing	useful									
Sales & Support	useful									
Metalexicography	useful									

Round Table contributor (see Section 3.8): SK = Simon Krek

For verbs, for instance, 'If you …' 'was established as the type strategy for verbs with human subjects' (Hanks 1987: 125), the traditional infinitive citation form 'To [verb] means …' was used for the 'absence of selection preferences on the subject' (Hanks 1987: 126), and 'When you …' was kept to convey certain social attitudes. While Hanks's discovery that 'lexical selection preferences are associated with particular syntactic structures' (Hanks 1987: 127) could be exploited to design explanatory strategies for verbs, this was not the case for nouns, for which he fell back on 'evidence from the surrounding context' (Hanks 1987: 127) as well as the use of prepositions. For function words, he realised that 'a discussion' was the most user-friendly approach. And so on for the other parts of speech. Or as summarised by Hanks (1987: 130): 'All Cobuild explanations, then, may be read as motivated variations on a few basic explanatory themes, principally 'An X is …', 'X means Y', and 'You use the word X like this". Hanks then went to the Cobuild programmer, Jeremy Clear, taking with him the various stylistic principles he had developed, and asked to combine these with the grammatical information already in the database, in order to basically start re-writing all COBUILD definitions in a uniform way. In the words of Clear himself:

> One interesting feature of the extract program was that it made an attempt to convert the definition recorded in the database into the prose style of definition which is to be found in the Cobuild dictionary […] The computer program used the syntax information associated with each sense category to generate an appropriate phraseology for the formulaic beginning of each definition. If, for example, the database entry for NOMINATE contains the information that this verb takes an object and the object is usually a person, then the computer employs the formula: 'If you nominate someone, you …' Of course, in many cases the computer would phrase the definition incorrectly, but it was agreed that overall there was a significant saving of manual effort in making these editorial modifications. (Clear 1987: 59)

If we now return to Lew's study (see Section 3.10), we understand better why his results are so impressive. In practical lexicography, for many centuries humans have recycled much more directly than today's transformer systems, and substantial progress in lexicography only came when they started to act more like 'traditional' computers.[29] Under the direction of Patrick Hanks, that computer was brought in to work on COBUILD definitions. If true, ChatGPT should not only be able to mimic the COBUILD style for definitions at verbs (and verbs of communication at that), but be able to write definitions in the COBUILD style at any part of speech. To test this, we asked ChatGPT to generate 'full' COBUILD entries (including frequency information and morphological forms) for all the parts of speech for which Hanks described explanatory strategies. Using one and the same prompt of merely eight lines of text, ChatGPT indeed managed to offer striking COBUILD-style dictionary articles for the *verb* 'recommend', the *nouns* 'tadpole' and 'bank', the *conjunction* 'and', the *adjectives* 'capacious' and 'floating', the *metaphorical meanings* of 'crumpet', and the *idiomatic expression* 'cut corners'. For good measure, the *interjection* 'goodbye' was added, generated using the same prompt plus one extra sentence. That extra sentence was added in order to obtain even 'fuller' dictionary articles, now complete with pronunciation as well as lists of synonyms and antonyms. The prompts as well as all dictionary material generated by ChatGPT are shown in the Addendum.

In general, and as one may conclude from the Addendum, the overall article structure is excellent, with large paragraph blocks for each part of speech, and numbered sense blocks within that. Each part-of-speech block starts with frequency information that comes across as convincing; and each sense block consistently contains morphological information, a COBUILD-style definition, as well as rich and varied example sentences. Sense division, too, seems more than good. A novelty that ChatGPT added throughout is a 'Note'-slot, in which even more context is given and/or in which the entire dictionary article is summarised in a

very accessible way. Let us now look at ChatGPT as a *full* COBUILD lexicographer as seen in the Addendum, compared to what we find in the actual COBUILD (2023) dictionary.

In order to illustrate the verb, and to link up with Lew's (forthcoming) study, we picked the verb of communication 'recommend'. ChatGPT claims that this verb is commonly used; it receives four (out of a maximum of five) diamonds in COBUILD. ChatGPT lists five senses; COBUILD lists three. One could say that ChatGPT is the splitter here, as sense 4 could be lumped with sense 1 (1: *'If you recommend someone or something, …'*; 4: *'If someone is recommended for a job or position, …'*), and sense 2 with sense 3 (3: *'If something is recommended, …'*; 2: *'If a doctor recommends a particular treatment or course of action, …'*). Reformulated, ChatGPT's senses 4 and 2 are sub-senses of its senses 1 and 3 respectively. The last senses are related, but worded differently by ChatGPT (5: *'If a product or service is recommended, …'*) vs. COBUILD (3: *'If something or someone has a particular quality to recommend them, …'*). The number of example sentences provided by ChatGPT range from four to two for each sense, and indeed correctly illustrate each definition, providing enough context. The Note is intriguing, as it literally maps meaning onto use: *'The word 'recommend' can also be used in other senses, such as when giving advice or suggestions in a non-formal context. However, the above definitions cover its most common and widely accepted usages'*. Lastly, of the explanatory strategies for verbs proposed by Hanks, we recognise 'If you …' and its variant 'If something …', as well as 'If someone …'. So far so good, as these results confirm the findings of Lew (forthcoming), with an improvement on the level of the example sentences — likely the result of slightly more tweaking of the prompt in that respect.

The first noun that was submitted to ChatGPT, 'tadpole', was chosen thus because Hanks (1987: 129) had pointed out that a 'clear preference in the evidence for the plural led the Cobuild editors to choose to begin their explanation: **Tadpoles** are …'. However, ChatGPT sticks to the prototypical strategy for nouns 'A [noun] is …' when defining it, but *does* use only the plural 'tadpoles' in each of the three example sentences. ChatGPT says that 'tadpole' is moderately used; COBUILD has no diamonds. Also, compared to COBUILD, ChatGPT adds a second informal sense (referring to a young and inexperienced person or thing). The second noun chosen, 'bank', a classic since Chomsky (1957: 95) and often used by Hanks, has multiple nominal and verbal uses. ChatGPT presents the information in two paragraphs, first a noun bloc, then a verb block. For the noun, ChatGPT correctly differentiates between the financial institution, the side of a river, and the raised area of land, illustrating these with three appropriate example sentences each. The explanatory strategies used are 'A [noun] refers to …' and 'A [noun] means …'. For the verb bloc, 'bank a vehicle' and 'bank a curve' are mentioned, as well as the phrasal verb 'bank on'. Compared to COBUILD, the nominal 'blood bank' is missing, as are the verbal 'bank money' and the phrase 'break the bank'. Two more explanatory strategies for verbs as predicted by Hanks are also seen here: 'When you …' and 'To [verb] means …'. Lastly, the frequency information is comparable: frequently used for the noun and moderately used for the verb according to ChatGPT; four diamonds overall in COBUILD.

The conjunction 'and' was chosen for its bewildering polysemy. ChatGPT lists six 'senses', each illustrated with three example sentences, COBUILD has twice as many sense blocks. This suggests that ChatGPT is the lumper this time, which is indeed apparent from the first definition: '"And" is used to connect words, phrases, or clauses that are similar or related, indicating addition or continuation'. At the same time, ChatGPT lists one (obvious) use which is (surprisingly) not mentioned among the ones in COBUILD, that of enumerating, such as in 'eggs, milk, *and* bread'. The overall treatment is satisfactory, as only infrequent COBUILD uses are missing from ChatGPT's effort. The explanatory strategy employed here by ChatGPT is the same for all definitions, being a variant of 'You use …', namely '[conjunction] is used to …'. The frequency information again compares favourably: very frequently used according to ChatGPT; five diamonds in COBUILD.

While the adjective 'capacious' only receives three lines in COBUILD, ChatGPT generated an extended treatment with no less than four senses, including three example sentences at each. In addition to the literal use (having a large capacity), ChatGPT also offers uses that are figurative (holding a vast amount of knowledge), metaphorical (feelings that are profound) and even abstract (being inclusive). Although the predicative encoding 'Something that is …' could have been used to define capacious, ChatGPT opted to use the simpler '[Adj] means …'-strategy normally reserved for adjectives which only or typically occur in attributive position. (But then again, ten of the twelve example sentences for 'capacious' illustrate attributive uses.) ChatGPT says that capacious is moderately used; COBUILD has no diamonds. The one important aspect that is missing from ChatGPT's treatment is that it does not point out that capacious is a 'formal' word. (That said, labels were not explicitly requested in our short prompt.) Similar outcomes may be observed for the attributive adjective 'floating'. While this entry is missing from COBUILD (where 'floating voter' was entered instead), ChatGPT generated an entry for it with no less than five senses, including three example sentences each. The first use is literal (something that is buoyant or suspended), the second a field-specific collocate ('floating currency exchange'), the next two are figurative (not attached or fixed to the ground; unanchored, uncommitted, transient), and the last is another collocate ('floating employee'). In terms of defining style, in addition to '[Adj] means …', also '[Adj] refers to …' and '[Adj] describes …' are employed.

Moving to metaphorical uses, and hoping to see how ChatGPT deals with what Hanks (1987: 134) calls the COBUILD 'displacement strategy', exemplified by him for 'bitch' as in 'If you call a woman a **bitch**, you mean …', resulted in ChatGPT's safety breaks being applied. All we learn is that 'bitch' can be 'used as a noun to refer to a woman in a derogatory way'; ChatGPT refuses to generate a proper dictionary article for it. Instead, and inspired by the discussion in van der Meer (1996: 429), we opted for 'crumpet'. Next to the literal meaning (a small, round and soft bread or cake), ChatGPT offers two extended uses, one labelled British informal slang (an attractive woman), the other labelled British informal language (something desirable). The wording of the second sense is probably not enough of a warning against its rude connotations; nor is the label 'figurative' included, although this is in line with: 'The word figurative itself need not be used, as there are other and more direct ways to suggest that a word has metaphorical applications' (van der Meer 1996: 427). The third sense, however, *is* referred to as crumpet being 'used metaphorically'. The defining style used here by ChatGPT, then, is simply 'X is used metaphorically to …'. Note, in passing, that here the three examples at sense 1 are exceptionally well-formed. Lastly, ChatGPT says that this headword is infrequently used; in COBUILD it has no diamonds.

The last category deals with idiomatic expressions, for which 'cutting corners' was chosen. ChatGPT's dictionary article is highly intriguing, as it consists of three sequentially numbered blocks, a first for the idiomatic expression 'cutting corners' (which it also correctly labels as such), a second for the verb 'cut', and a third for the plural noun 'corners'. Rather than Hanks's suggested 'If you say … [idiom], you mean …', ChatGPT uses 'The phrase [idiom] is an idiomatic expression used to …'. (In COBUILD itself, this idiom is found as sense 14 (out of 16) under 'corner', and the defining strategy used is 'If you [idiom], you …'.) As to the verb 'cut' in block 2, the intriguing part is that ChatGPT does not give a full treatment (which would be rather counterproductive), but instead localises the definition for the idiom at hand: 'As a verb, "cut" means to make an incision or divide something by using a sharp tool, but in this context, it is used figuratively to represent the act of taking shortcuts or reducing efforts'. The three examples that follow then attempt to use both the verb 'cut' and the idiom 'cut corners', e.g. 'The company faced financial constraints and had to *cut* its expenses, often *cutting corners* to stay afloat'. Likewise, the definition of 'corners' in block 3, is localised: 'In this context, "corners" refers to the areas or aspects of a task or project where shortcuts or compromises are made', which is then again followed by example sentences. No doubt, the resulting dictionary article is excessively long, but it is truly

as if ChatGPT is thinking out of the box here. It invites one to go back to the initial prompt and wonder how ChatGPT manages to produce such convincing lexicographic material, over and over.

True, prompt engineering is known to be of paramount importance when using an LLM like ChatGPT. Given the more than satisfactory output across the various word categories so far — most of which indeed use Hanks's various explanatory strategies for defining, and successfully emulate all other COBUILD aspects — we took the same prompt, to which we added just one sentence, now also requesting pronunciations in IPA format, as well as lists of synonyms and antonyms where relevant. This prompt was applied once, for a headword with a part of speech not yet covered, namely the interjection 'goodbye'. The dictionary article generated by ChatGPT and seen in the Addendum is, in two words, simply stunning, and once again strengthens our conviction that ChatGPT is already a more than respectable lexicographer (see the last column in Table 2). As a matter of fact, the final prompt that was created can be fed 'as is' to a dictionary writing system like TLex, and with it (draft) quality articles for an entire dictionary may be compiled fully automatically in batch mode, and assigned to ChatGPT as the (co-)author. As of this day, therefore, one would be remiss not to want to make use of ChatGPT for dictionary compilation.

5. Conclusions and Further musings

We have seen that, with the right prompts, an LLM like ChatGPT can already be brought in to either compile a dictionary on its own, or, somewhat more safely, to speed up dictionary compilation by providing quality draft material which human lexicographers then assess and improve upon. We have also seen that all colleagues have voiced one or more concerns in exploiting this technology for lexicography. These concerns have been brought together in Table 3, together with the (main) final assessments and suggestions by those colleagues.

One of the concerns most often heard is that LLMs are prone to hallucination (mentioned six times), followed by concerns that, on the one hand, the *input* is a black box while, on the other hand, the *output* is non-deterministic (both mentioned five times). Similarly frequently stressed is the current English bias, and the need to become adept at composing generic prompts that can do the job at scale (both also mentioned six times). All of these could be reason enough to dismiss the use of ChatGPT in lexicography out of hand. However, when we look at the section 'Final word on ChatGPT' in Table 3, we see that its full adoption is already suggested six times, versus seven for 'not yet' and only one vote against its use in its current form. Clearly, lexicographers at large would do well to continue their experiments in harnessing the beast known as ChatGPT in particular as well as LLMs in general. This brings us to the main suggestions to bring LLMs under control for lexicography in the short term — the last line in Table 3. There are five mentions of LexGPT, meaning the goal to develop an LLM specifically for lexicography. Two shortcuts are also suggested to get there, namely, to either fine-tune a foundation model, or to depart from an open-source LLM. The more circumspect suggestion is to limit the use of LLMs to well-defined sub-tasks in the lexicographic process. To return to ChatGPT, one colleague prefers a pragmatic approach whereby one simply uses the tool 'as is, but selectively' for dictionary making, or prefers to see it as just one of many tools for post-editing in lexicography. Three colleagues, finally, advocate using ChatGPT 'as is' to augment the production rate and quality of lexicographic products.

Missing from everything that has been said above is 'the dictionary user'. First, assuming that there will still be users in the future, it is a bit surprising to note that only professional lexicographers have judged and commented on ChatGPT output to date. Surely, the proof of the pudding would lie in knowing whether actual users are also satisfied with material generated by ChatGPT. Future studies should thus include human

Table 3: ChatGPT in lexicography: Concerns

	de Schryver & Joffe	Barrett	Rundell	McKean & Fitzgerald	Jakubíček & Rundell	Tran et al.	Round Table	Nichols	Lew	de Schryver
Section	§ 3.1	§ 3.2	§ 3.3	§ 3.4	§ 3.5	§ 3.7	§ 3.8	§ 3.9	§ 3.10	§ 4
ChatGPT concerns										
Dangerous	+									
Authorship (Bob or ChatGPT?)	+									+
Privileged use?	+									
Environmental worries				+						
Societal bias				+						
English bias	+				+		G-M, AT, MK, SK			
Confabulation				+						
Hallucination	+	+	+	+			MK	+		
Answers too complex								+		
Transparency (what's in the LLM?)	+		+	+	+		MR			
Output non-deterministic	+		+	+	+		MK			
Indiscriminate							MR			
IP and copyright; plagiarism		+		+			AT	+		
Regulation				+				+		
Brand status (trust; curation)			+	+				+		
LLM output used as LLM input		+								
Modelling, not reasoning					+		PR	+		
Static (no data beyond a certain date)					+			+		
Length of prompts and responses is limited					+					
Need to fine-tune prompts					+	+	G-M	+	+	+
End of dictionaries							G-M, AT			
End of lexicographers				+			G-M			

Table 3. Continued

	de Schryver & Joffe	Barrett	Rundell	McKean & Fitzgerald	Jakubíček & Rundell	Tran et al.	Round Table	Nichols	Lew	de Schryver
End of post-editing lexicographic tools							G-M			+
Final word on ChatGPT										
Full adoption	yes	not yet	not yet	not yet	not yet	yes	↓	not yet	yes	yes
						G-M: yes; MR: not yet; AT: yes; PR: no; MK: not yet				
Main suggestion	AI-augmented	LexGPT	post-edit	LexGPT	LexGPT	classify > generate	↓	LexGPT	AI-augmented	AI-augmented
						G-M: LexGPT; MR: selective use; AT: fine-tune foundation; PR: for sub-tasks; MK: open-source LLM				

Round Table contributors (see Section 3.8): G-M = Gilles-Maurice de Schryver, MR = Michael Rundell, AT = Arvi Tavast, PR = Pavel Rychlý, MK = Marko Kokol, SK = Simon Krek

users who evaluate 'ChatGPT the Lexicographer'. Second, we all seem to have naively assumed that dictionaries as we know them will still be needed in the future, so have checked whether ChatGPT can perform the various steps in producing them. However, it is our contention that dictionaries of the future will at best be subsumed within, at worst be gobbled up by, other digital tools. While there is no doubt that there will still be a need for lexical knowledge in the future, it is thus not clear at all whether it will need to be synthesised and presented as is the case in current reference works. If that is so, we should have asked ChatGPT to produce the type of lexical analysis that will be needed in the future, rather than mimic the past. At that point ChatGPT would really constitute a disruptive technology in a disruptive landscape, rather than merely being employed as a time and money-saving device by which dictionaries are getting only incrementally better. Third, ChatGPT generates answers for users 'on the fly': nothing whatsoever is analysed and synthesised in advance. So why would lexicographers want to go back in time, and stop that time by 'freezing' and 'capturing' output, to then either return it as in traditional dictionaries (point 1) or have it used as invisible lexicography (point 2)? Why assume that lexicographers need to prepare anything at all, and why not simply admit that the only thing which future users will need is ChatGPT itself with whom they chat?

The answers to these three questions about the dictionary user of the future have important consequences. First, if there will be no future dictionary user, understood as a person who narrowly seeks lexical information in a dedicated reference work, then there is indeed no point in asking non-lexicographers to 'evaluate' ChatGPT output for a hypothetical dictionary that will never exist. Second, if future dictionaries will disappear from view, we can safely bin all existing dictionary use(r) research, as none of it will be relevant going forward in a world in which users chat with tools sequentially, rather than consult dictionaries compiled by lexicographers. Third, lexicographers could ask themselves whether they are still going to produce dictionaries for any humans or just as fodder for LLMs. And if the latter, why would lexicographers take the trouble pre-processing lexical facts at all? Surely, LLMs could work out these facts for themselves, starting from the raw data.

These three questions are ostensibly about the user, but in reality bring us back to our opening position statement, where we claimed that dictionaries, lexicographers and post-editing lexicographic tools have all become redundant in an age of LLMs like ChatGPT. But we also said that we do not really stand by that claim. Here's why. Just as the arrival of the camera did not lead to fewer paintings, just as the invention of the calculator did not lead to the mass-firing of mathematicians, and just as the introduction of smartphones did not stop people making calls, writing e-mails, browsing the Web, and listening to music; if anything, as a result of these revolutionary technologies, paintings became more original, mathematicians performed ever more complex calculations, and smartphones ended up venturing into group chats, ride-hailing, short video, casual gaming, and mobile payments.[30] The same may be expected for the use of ChatGPT in lexicography. Just give it time, as we do not know yet how this revolution will pan out.

Epilogue

When asked at the most recent eLex conference whether he knew that the left parts of all COBUILD definitions had initially been written by a computer program, COBUILD lexicographer Michael Rundell replied that he had no recollection of that, and certainly not that it was done to bring order to the project with the chief aim to complete the first COBUILD dictionary in a timely manner (Rundell, personal communication, 28 June 2023). This is important information. If the COBUILD lexicographers accepted the draft material with which they were presented without questioning it, and worked through editing that without further

ado, then it proves that Hanks was right, and indeed, that a computer could help in writing dictionary definitions. 'Hanks was right' looks innocent, but it has profound implications. In COBUILD, the different senses (or 'meaning potentials' in Hanks's terminology) are grouped into paragraphs, and 'The division of entries into paragraphs is based mainly on the distinctions that can be made in the wording of the first part of each explanation' (Hanks 1987: 118). In other words, Hanks was right in that meanings can indeed be mapped onto use (Hanks 2002), and as such he proved Sinclair's axiom that *'Every distinct sense of a word is associated with a distinction in form'* (Moon 1987: 89). The fact that those distinct forms could then be used as starting points to automate the writing of a revolutionary type of dictionary definition, is even more revealing. Simply put, the type of Corpus Pattern Analysis which Hanks undertook when the corpus was first used in lexicography four decades ago, is exactly what ChatGPT is now mimicking. The first ChatGPT engine in lexicography, then, was none other than Patrick Hanks.

Notes

1 In theoretical linguistics, Noam Chomsky is best known — rather ironically — for his work on transformational-generative grammar (TGG), starting with Chomsky (1957).
2 In computational lexicology, Piek Vossen is best known for his work on EuroWordNet, see e.g. Vossen (1998).
3 This announcement generated quite some reactions, especially on the EURALEX mailing list, see https://www.freelists.org/archive/euralex/11-2012.
4 Thank goodness that we still have our Macmillan paper dictionaries from the noughties!
5 Regarding the concept 'open' in the name OpenAI Elon Musk pointed out: 'OpenAI was created as an open source (which is why I named it 'Open' AI), non-profit company to serve as a counterweight to Google, but now it has become a closed source, maximum-profit company effectively controlled by Microsoft. Not what I intended at all' (Tweet, 17 February 2023, https://twitter.com/elonmusk/status/1626516035863212034). On 12 July 2023, Elon Musk announced that he has now formed a new AI startup: xAI (https://x.ai/).
6 Or as Miloš Jakubíček worded it in an e-mail, referring to the GPT-3 engine of the time: 'there is substantial fine-tuning performed (for the 'chat' task) on top of GPT-3 that makes it perform so well. To make it clear: ChatGPT is GPT-3 based roughly like Bentley is based on a 12-cylinder fuel engine. Obviously access to a 12-cylinder does not give you a Bentley' (Jakubíček, personal communication, 27 February 2023).
7 More precisely, the model used at the time was text-davinci-003. For an overview of the various OpenAI models, see https://platform.openai.com/docs/models/overview.
8 We write 'colloquially', as technically what was demonstrated should have been 'Bob or text-davinci-003?', cf. endnote 7, as hinted at by Miloš Jakubíček during question time. This observation became moot just two days later, however, as on 1 March 2023 OpenAI sent out an e-mail stating: 'The new Chat API calls gpt-3.5-turbo, the same model used in the ChatGPT product'.
9 See for a longer list of applications with ChatGPT: https://platform.openai.com/examples.
10 Actually, and to make this point, in the present article the first paragraph of Section 2 was written as such by ChatGPT itself. ChatGPT was prompted with 'What is ChatGPT, and on how much data was it trained?' The same section was also used in the Tokyo lecture on ChatGPT. Since then, entire scientific articles (except for the reference sections) have indeed been written, adjudicated *and published*, as is for instance the case for the punningly titled 'Chatting and cheating' by Cotton et al. (2023), with an intriguing acknowledgments section, published online on 13 March 2023, but revealed to be GPT-4 output on 19 March 2023 (Fazackerley 2023).
11 That 'memorisation' is indeed the likely cause, is also suggested by this research finding: 'We find that OpenAI models have memorized a wide collection of copyrighted materials, and that the degree of memorization is tied to the frequency with which passages of those books appear on the web' (Chang et al. 2023).
12 Immediately following the Tokyo lecture on ChatGPT (see Section 3.1), David Joffe had also prompted Stable Diffusion to generate a number of images illustrating the future of lexicography. Two of these were inserted into the banner at https://tshwanedje.com/, and are reproduced here, showing competing

visions of the future of our field: one where the lexicographer is still in control but gets help from generative AI, the other where generative AI does all the lexicographic work:

13 Conversely, Google's Bard only became available in the EU as late as 13 July 2023, having been 'held up after the main data regulator in the bloc raised privacy concerns' (BBC 2023).

14 In the original project there were '100 sample dictionary entries' (Atkins et al. 2010: 550), while the Addendum to the eLex paper (Jakubíček and Rundell 2023: 523) lists 101 — but one notices two headwords each for 'grave' and 'might' there. In any case, nothing statistical is done with the sample.

15 Both presenters, but this should be clear from the foregoing, had seen (and again commented on) our Tokyo lecture on ChatGPT (see Section 3.1).

16 Ten each in five different word classes; *nouns*: time, year, people, way, man, day, thing, work, child, government; *verbs*: be, have, do, say, go, make, get, see, know, take; *adjectives*: other, good, new, more, first, many, such, last, own, same; *adverbs*: not, so, then, as, well, now, just, also, only, very; *prepositions*: of, in, to, for, on, that, with, at, by, as.

17 With regard to the syntactic-semantic level, for instance, we learn that ChatGPT will often not even include the headword for phrasemes, provide collocations instead of idioms, and omit the headwords in proverbs, but the slots for phrasemes, idioms and proverbs are all given 100% scores for ChatGPT (Phoodai and Rikk 2023: 347-348). At that rate, and with enough tweaking of the prompts, ChatGPT could have filled every single slot, at all times, except, perhaps, for first citations + their correct reference (but the OALD-10 does not do that either), and cross-references (but the latter is a mediostructural slot, so should not even have been discussed).

18 Seemingly, the unexplained abbreviation AUSA refers to pronunciation, RA to spelling, WAA to part of speech, ABED to meaning, and ETYA to etymology, in all likelihood based on German (!) abbreviations as found in the *Wörterbuch zur Lexikographie und Wörterbuchforschung* (Wiegand et al. 2010-2019). Deep down in that dictionary's entry for *'Etymologieangabe'*, for example, one finds 'EtyA' as abbreviation. 'EtyA' and the other abbreviations have not been entered as lemmata into Wiegand's five-volume dictionary, though, so one has to guess: AusA =? *Ausspracheangabe*, RA =? *Rechtschreibangabe*, WAA =? *Wortartangabe*, ABed =? *Angabe zur Bedeutung*.

19 The round table was concluded by taking questions from the participants. This Q & A section has not been summarised here, though some of the points raised have been recycled in the concluding section of this article.

20 From her talk it is clear that Nichols was unfortunately fooled into thinking that OpenAI's ChatGPT is 'open source'; for which, see Endnote 5.

21 Responding quickly to the round table on ChatGPT from the day before (see Section 3.8), Nichols 'interacted' with some of the points made by Arvi Tavast (on LLMs now put into the hands of the public), Pavel Rychlý (on ChatGPT not being sentient), and Gilles-Maurice de Schryver (on wholesale copying) who then returned to it during question time. She also borrowed and discussed the '(mirus) virus' slide from Barrett's talk at the DSNA 2023 (see Section 3.2).

22 Nichols specifically refers to the by now infamous case of two New York lawyers who were sanctioned and fined for having submitted a court brief written up by ChatGPT. Amongst others, ChatGPT had made up six fictitious case citations (which thus do not exist), but it had nonetheless attributed these

cases to real judges (Merken 2023). It is surprising that lawyers took this shortcut, as it is well known that ChatGPT references and bibliographies are typically invented and not real, as the LLM simply combines familiar names, years, titles, etc. as seen during its training.

23 Although less relevant here, but in the interest of full disclosure, Lew had also watched our Tokyo lecture on ChatGPT (see Section 3.1), and refers to it in his study. Further, Lew had been given access to an early version of the text of Michael Rundell's ASIALEX 2023 keynote (see Section 3.3), and vice versa: Rundell had access to Lew's preprint, which he quotes in his keynote.

24 These are: insist, deny, discourage, recommend, persuade, encourage, suggest, blame, decline, question, justify, argue, plead, confirm, beg. Verbs of communication tend to have syntactic patterns as well as pragmatic uses that are non-trivial, so they are a lexicographic challenge.

25 Jakubíček and Rundell (2023: 512-513) convincingly illustrate this with monolingual Czech dictionary articles for the noun *stát* 'country' as generated by ChatGPT (using both GPT-3.5 and GPT-4): the articles are clearly back-translated from English, and consequently include material that does not exist in Czech.

26 Phoodai and Rikk (2023) has not been included any further, as it does not contain any proper evaluations (see Section 3.6).

27 The information divulged here comes from personal communication with Patrick Hanks, obtained while doing the research for his Festschrift (de Schryver 2010b). While this information was eventually not fully 'revealed' in the introductory chapter to that Festschrift (de Schryver 2010a), we feel the time is right to do so now.

28 COBUILD literally stands for 'the Collins Birmingham University International Language *Database*' (emphasis added).

29 This sentence has in turn been recycled from an e-mail received from Robert Lew on 2 June 2023. His was in reply to our suggestion the day before that tools like ChatGPT will become ever more human, and this rather sooner than later, for the simple reason that we are convinced that the brains of humans do not work very differently from the 'brains' of an LLM. We too simply regurgitate material based on everything that we have seen before. See e.g. O'Connor (2023).

30 These three points are inspired by arguments made, respectively, by Orin Hargraves at the DSNA 2021 conference (Hargraves 2021), Marko Kokol during the Round Table at the eLex 2023 conference (Q & A section in de Schryver et al. 2023), and an article on Apple's Vision Pro in *The Economist* (2023a).

References

A. Dictionaries

COBUILD. 2023. *Collins Cobuild Advanced Learner's Dictionary*. https://www.collinsdictionary.com/dictionary/english

Hornby, Albert Sydney, Diana Lea, Jennifer Bradbery, Victoria Bull, Leonie Hey, Stacey Bateman, Kallah Pridgeon and Gary Leicester. 2020. *Oxford Advanced Learner's Dictionary of Current English, Tenth Edition*. Oxford: Oxford University Press.

McIntosh, Colin. 2023. *Cambridge Thesaurus: Synonyms, Antonyms and Examples*. Cambridge: Cambridge University Press & Assessment. https://dictionary.cambridge.org/thesaurus/

Sinclair, John McH., Patrick Hanks, Gwyneth Fox, Rosamund Moon and Penny Stock. 1987. *Collins Cobuild English Language Dictionary*. London: William Collins Sons & Co.

Wiegand, Herbert E., Michael Beißwenger, Rufus H. Gouws, Matthias Kammerer, Angelika Storrer and Werner Wolski (eds). 2010-2019. *Wörterbuch Zur Lexikographie Und Wörterbuchforschung/ Dictionary of Lexicography and Dictionary Research. Mit Englischen Übersetzungen Der Umtexte Und Definitionen Sowie Äquivalenten in Neun Sprachen/ With English Translations of the Outer Texts and Definitions as Well as Equivalents in Nine Languages*. Berlin: Walter de Gruyter.

B. Other literature

Anon. 2023. 'Macmillan Dictionary to Close after 14 Years of Online Excellence.' Macmillan Education 30 June 2023. https://macmillaneducation.secure.force.com/help/bg_FAQArticle?id=kA04H0000005kwWSAQ

Arredondo, Pablo. 2023. 'GPT-4 Passes the Bar Exam: What That Means for Artificial Intelligence Tools in the Legal Profession.' *Stanford Law School* 19 April 2023. https://law.stanford.edu/2023/04/19/gpt-4-passes-the-bar-exam-what-that-means-for-artificial-intelligence-tools-in-the-legal-industry/

Atkins, B.T. Sue, Adam Kilgarriff and Michael Rundell. 2010. 'Database of Analysed Texts of English (DANTE): The NEID Database Project' In Dykstra, Anne and Tanneke Schoonheim (eds), *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6-10 July 2010)*. Leeuwarden/Ljouwert: Fryske Akademy, 549–556.

Baisa, Vít, Marek Blahuš, Michal Cukr, Ondřej Herman, Miloš Jakubíček, Vojtěch Kovář, Marek Medveď, Michal Měchura, Pavel Rychlý and Vít Suchomel. 2019. 'Automating Dictionary Production: A Tagalog-English-Korean Dictionary from Scratch' In Kosem, Iztok, Tanara Zingano Kuhn, Margarita Correia, José Pedro Ferreira, Maarten Jansen, Isabel Pereira, Jelena Kallas, Miloš Jakubíček, Simon Krek and Carole Tiberius (eds), *Electronic Lexicography in the 21st Century (eLex 2019): Smart Lexicography. Conference Proceedings. Sintra, Portugal, 1-3 October 2019*. Brno: Lexical Computing, 805–818.

Barnbrook, Geoff. 2002. *Defining Language: A Local Grammar of Definition Sentences. (Studies in Corpus Linguistics 11.)*. Amsterdam: John Benjamins.

Barrett, Grant. 2023. 'Defin-O-Bots: Challenging A.I. to Create Usable Dictionary Content.' Paper presented at the 24th Biennial Conference of the Dictionary Society of North America. Boulder, CO, USA, 31 May – 3 June 2023.

BBC. 2023. 'Bard: Google's ChatGPT Rival Launches in Europe and Brazil.' *BBC* 13 July 2023. https://www.bbc.com/news/technology-66187283

Chang, Kent K., Mackenzie Cramer, Sandeep Soni and David Bamman. 2023. 'Speak, Memory: An Archaeology of Books Known to ChatGPT/GPT-4.' arXiv. https://arxiv.org/pdf/2305.00118.pdf

Chomsky, Noam. 1957. *Syntactic Structures. (Janua Linguarum.)*. The Hague: Mouton.

Chomsky, Noam, Ian Roberts and Jeffrey Watumull. 2023. 'AI Unravelled: The False Promise of ChatGPT.' *New York Times* 8 March 2023. https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html

Clear, Jeremy. 1987. 'Computing' In Sinclair, John McH. (ed.), *Looking Up: An Account of the Cobuild Project in Lexical Computing and the Development of the Collins Cobuild English Language Dictionary*. London: Collins ELT, 41–61.

Cotton, Debby R.E., Peter A. Cotton and J. Reuben Shipway. 2023. 'Chatting and Cheating: Ensuring Academic Integrity in the Era of ChatGPT.' *Innovations in Education and Teaching International*. https://www.tandfonline.com/doi/full/10.1080/14703297.2023.2190148

de Schryver, Gilles-Maurice. 2003. 'Lexicographers' Dreams in the Electronic-Dictionary Age.' *International Journal of Lexicography* 16.2: 143–199.

de Schryver, Gilles-Maurice. 2010a. 'Getting to the Bottom of How Language Works' In de Schryver, Gilles-Maurice (ed.), *A Way with Words: Recent Advances in Lexical Theory and Analysis. A Festschrift for Patrick Hanks. (Linguistics Series 2.)*. Kampala: Menha Publishers, 3–34.

de Schryver, Gilles-Maurice (**ed.**). 2010b. *A Way with Words: Recent Advances in Lexical Theory and Analysis. A Festschrift for Patrick Hanks. (Linguistics Series 2.)*. Kampala: Menha Publishers.

de Schryver, Gilles-Maurice. 2024. 'The Future of the Dictionary' In Finegan, Edward and Michael Adams (eds), *The Cambridge Handbook of the Dictionary*. Cambridge: Cambridge University Press.

de Schryver, Gilles-Maurice and David Joffe. 2023. 'The End of Lexicography, Welcome to the Machine: On How ChatGPT Can Already Take over All of the Dictionary Maker's Tasks.' Paper presented at the 20th CODH Seminar. Center for Open Data in the Humanities, Research Organization of Information and Systems, National Institute of Informatics, Tokyo, Japan, 27 February 2023. https://youtu.be/watch?v=mEorw0yefAs

de Schryver, Gilles-Maurice, Michael Rundell, Arvi Tavast, Pavel Rychlý, Marko Kokol (panellists) and Simon Krek (moderator). 2023. 'Round Table on "Large Language Models and AI in Lexicography"'. Panel at the 8th Electronic Lexicography in the 21st Century Conference. Brno, Czech Republic, 27-29 June 2023. https://www.youtube.com/watch?v=dNkksTDYa_s

Fazackerley, Anna. 2023. 'AI Makes Plagiarism Harder to Detect, Argue Academics – in Paper Written by Chatbot.' *The Observer* 19 March 2023. https://www.theguardian.com/technology/2023/mar/19/ai-makes-plagiarism-harder-to-detect-argue-academics-in-paper-written-by-chatbot

Fox, Gwyneth. 1987. 'The Case for Examples' In Sinclair, John McH. (ed.), *Looking Up: An Account of the Cobuild Project in Lexical Computing and the Development of the Collins Cobuild English Language Dictionary*. London: Collins ELT, 137–149.

Hanks, Patrick. 1987. 'Definitions and Explanations' In Sinclair, John McH. (ed.), *Looking Up: An Account of the Cobuild Project in Lexical Computing and the Development of the Collins Cobuild English Language Dictionary*. London: Collins ELT, 116–136.

Hanks, Patrick. 2002. 'Mapping Meaning onto Use' In Corréard, Marie-Hélène (ed.), *Lexicography and Natural Language Processing. A Festschrift in Honour of B.T.S. Atkins. s.l.*: Euralex, 156–198.

Hargraves, Orin. 2021. 'Lexicography in a Post-Dictionary World.' Paper presented at the 23rd Biennial Conference of the Dictionary Society of North America. Held virtually, 4 June 2021.

Jakubíček, Miloš, Adam Kilgarriff, Vojtěch Kovář, Pavel Rychlý and Vít Suchomel. 2013. 'The TenTen Corpus Family' In Hardie, Andrew and Robbie Love (eds), *Corpus Linguistics 2013, Abstract Book*. Lancaster: UCREL, 125–127.

Jakubíček, Miloš, Vojtěch Kovář and Pavel Rychlý. 2021. 'Million-Click Dictionary: Tools and Methods for Automatic Dictionary Drafting and Post-Editing' In Gavriilidou, Zoe, Lydia Mitits and Spyros Kiosses (eds), *Euralex XIX, Congress of the European Association for Lexicography, Lexicography for Inclusion, 7-9 September 2021, Virtual, Book of Abstracts*. Komotini: SynMorPhoSe Lab, Democritus University of Thrace, 65–67.

Jakubíček, Miloš and Michael Rundell. 2023. 'The End of Lexicography? Can ChatGPT Outperform Current Tools for Post-Editing Lexicography?' In Medveď, Marek, Michal Měchura, Carole Tiberius, Iztok Kosem, Jelena Kallas and Miloš Jakubíček (eds), *Proceedings of the eLex 2023 Conference: Electronic Lexicography in the 21st Century*. Brno: Lexical Computing, 508–523. https://www.youtube.com/watch?v=8e52vvDpdfQ

Kilgarriff, Adam. 2007. 'Googleology Is Bad Science.' *Computational Linguistics* 33.1: 147–151.

Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý and Vít Suchomel. 2014. 'The Sketch Engine: Ten Years On.' *Lexicography: Journal of ASIALEX* 1.1: 7–36.

Kilgarriff, Adam, Miloš Husák, Katy McAdam, Michael Rundell and Pavel Rychlý. 2008. 'GDEX: Automatically Finding Good Dictionary Examples in a Corpus' In Bernal, Elisenda and Janet DeCesaris (eds), *Proceedings of the XIII Euralex International Congress (Barcelona, 15-19 July 2008) (Sèrie Activitats 20.)*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, 425–432.

Kilgarriff, Adam and Pavel Rychlý. 2010. 'Semi-Automatic Dictionary Drafting' In de Schryver, Gilles-Maurice (ed.), *A Way with Words: Recent Advances in Lexical Theory and Analysis. A Festschrift for Patrick Hanks. (Linguistics Series 2.)*. Kampala: Menha Publishers, 299–312.

Kilgarriff, Adam, Pavel Rychlý, Pavel Smrz and David Tugwell. 2004. 'The Sketch Engine' In Williams, Geoffrey and Sandra Vessier (eds), *Proceedings of the Eleventh Euralex International Congress, Euralex 2004, Lorient, France, July 6-10, 2004*. Lorient: Faculté des Lettres et des Sciences Humaines, Université de Bretagne Sud, 105–115.

Kilgarriff, Adam and David Tugwell. 2001. 'Word Sketch: Extraction and Display of Significant Collocations for Lexicography' *Workshop Proceedings. Collocation: Computational Extraction, Analysis and Exploitation (39th Annual Meeting of the Association for Computational Linguistics and 10th Conference of the European Chapter of the Association for Computational Linguistics, Toulouse, 7 July 2001)*, 32–38.

Kosem, Iztok, Polona Gantar, Nataša Logar and Simon Krek. 2014. 'Automation of Lexicographic Work Using General and Specialized Corpora: Two Case Studies' In Abel, Andrea, Chiara Vettori and Natascia Ralli (eds), *Proceedings of the XVI Euralex International Congress: The User in Focus, 15-19 July 2014, Bolzano/Bozen*. Bolzano/Bozen: EURAC Research, 355–364.

Lew, Robert. forthcoming. 'ChatGPT as a COBUILD Lexicographer.' *Humanities and Social Sciences Communications*.

Liang, Paul Pu, Chiyu Wu, Louis-Philippe Morency and Ruslan Salakhutdinov. 2021. 'Towards Understanding and Mitigating Social Biases in Language Models' *Proceedings of the 38th International Conference on Machine Learning, PMLR 139*. https://arxiv.org/pdf/2106.13219.pdf

Liang, Yaobo, Chenfei Wu, Ting Song, Wenshan Wu, Yan Xia, Yu Liu, Yang Ou, Shuai Lu, Lei Ji, Shaoguang Mao, Yun Wang, Linjun Shou, Ming Gong and Nan Duan. 2023. 'TaskMatrix.AI: Completing Tasks by Connecting Foundation Models with Millions of APIs.' *arXiv*. https://arxiv.org/pdf/2303.16434.pdf

McKean, Erin and Will Fitzgerald. 2023. 'The ROI of AI in Lexicography' *Proceedings of the 16th International Conference of the Asian Association for Lexicography: "Lexicography, Artificial Intelligence, and Dictionary Users"*. Seoul: Yonsei University, 10–20.

Merken, Sara. 2023. 'New York Lawyers Sanctioned for Using Fake ChatGPT Cases in Legal Brief.' *Reuters* 26 June 2023. https://www.reuters.com/legal/new-york-lawyers-sanctioned-using-fake-chatgpt-cases-legal-brief-2023-06-22/

Moon, Rosamund. 1987. 'The Analysis of Meaning' In Sinclair, John McH. (ed.), *Looking Up: An Account of the Cobuild Project in Lexical Computing and the Development of the Collins Cobuild English Language Dictionary.* London: Collins ELT, 86–103.

Nichols, Wendalyn. 2023. 'Invisible Lexicographers, AI, and the Future of the Dictionary' Paper presented at the eLex 2023 Conference: Electronic Lexicography in the 21st Century. Brno, Czech Republic, 27-29 June 2023. https://www.youtube.com/watch?v=xYpwftj_QQI

O'Connor, Brendan H. 2023. 'Are You Part Robot? A Linguistic Anthropologist Explains How Humans Are Like ChatGPT – Both Recycle Language.' *The Conversation* 12 June 2023. https://theconversation.com/are-you-part-robot-a-linguistic-anthropologist-explains-how-humans-are-like-chatgpt-both-recycle-language-203477

Phoodai, Chayanon and Richárd Rikk. 2023. 'Exploring the Capabilities of ChatGPT for Lexicographical Purposes: A Comparison with Oxford Advanced Learner's Dictionary within the Microstructural Framework' In Medveď, Marek, Michal Měchura, Carole Tiberius, Iztok Kosem, Jelena Kallas and Miloš Jakubíček (eds), *Proceedings of the eLex 2023 Conference: Electronic Lexicography in the 21st Century.* Brno: Lexical Computing, 335–365. https://www.youtube.com/watch?v=y7k9olAllbo

Reuters. 2023. 'OpenAI May Leave the EU If Regulations Bite - CEO.' *Reuters* 24 May 2023. https://www.reuters.com/technology/openai-may-leave-eu-if-regulations-bite-ceo-2023-05-24/

Rundell, Michael. 2012a. 'It Works in Practice but Will It Work in Theory? The Uneasy Relationship between Lexicography and Matters Theoretical' In Fjeld, Ruth V. and Julie M. Torjusen (eds), *Proceedings of the 15th Euralex International Congress, 7-11 August, 2012, Oslo*. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo, 47–92.

Rundell, Michael. 2012b. 'The Road to Automated Lexicography: An Editor's Viewpoint' In Granger, Sylviane and Magali Paquot (eds), *Electronic Lexicography*. Oxford: Oxford University Press, 15–30.

Rundell, Michael. 2012c. 'Stop the Press: Dictionary No Longer a Page-Turner.' *Macmillan Education* 5 November 2012. https://www.prnewswire.com/news-releases/stop-the-press-dictionary-no-longer-a-page-turner-177210131.html

Rundell, Michael. 2023. 'Automating the Creation of Dictionaries: Are We Nearly There?' *Proceedings of the 16th International Conference of the Asian Association for Lexicography: "Lexicography, Artificial Intelligence, and Dictionary Users"*. Seoul: Yonsei University, 1–9.

Rundell, Michael and Adam Kilgarriff. 2011. 'Automating the Creation of Dictionaries. Where Will It All End?' In Meunier, Fanny, Sylvie De Cock, Gaëtanelle Gilquin and Magali Paquot (eds), *A Taste for Corpora. In Honour of Sylviane Granger. (Studies in Corpus Linguistics 45.)*. Amsterdam: John Benjamins, 257–281.

Schulman, John *et al*. 2022. 'Introducing ChatGPT.' https://openai.com/blog/chatgpt

Shen, Yongliang, Kaitao Song, Xu Tan, Dongsheng Li, Weiming Lu and Yueting Zhuang. 2023. 'HuggingGPT: Solving AI Tasks with ChatGPT and Its Friends in Hugging Face.' arXiv. https://arxiv.org/pdf/2303.17580.pdf

Sinclair, John McH. 1987a. 'Grammar in the Dictionary' In Sinclair, John McH. (ed.), *Looking Up: An Account of the Cobuild Project in Lexical Computing and the Development of the Collins Cobuild English Language Dictionary.* London: Collins ELT, 104–115.

Sinclair, John McH. (ed.). 1987b. *Looking Up: An Account of the Cobuild Project in Lexical Computing and the Development of the Collins Cobuild English Language Dictionary.* London: Collins ELT.

The Economist. 2023a. 'Apple's Vision Pro Is an Incredible Machine. Now to Find out What It Is For.' *The Economist* 6 June 2023. https://www.economist.com/leaders/2023/06/06/apples-vision-pro-is-an-incredible-machine-now-to-find-out-what-it-is-for

The Economist. 2023b. 'Large Language Models' Ability to Generate Text Also Lets Them Plan and Reason.' *The Economist* 19 April 2023. https://www.economist.com/science-and-technology/2023/04/19/large-language-models-ability-to-generate-text-also-lets-them-plan-and-reason

The Economist. 2023c. 'Large, Creative AI Models Will Transform Lives and Labour Markets.' *The Economist* 22 April 2023. https://www.economist.com/interactive/science-and-technology/2023/04/22/large-creative-ai-models-will-transform-how-we-live-and-work

Tran, Hanh Thi Hong, Vid Podpečan, Mateja Jemec Tomazin and Senja Pollak. 2023. 'Definition Extraction for Slovene: Patterns, Transformer Classifiers and ChatGPT' In Medveď, Marek, Michal Měchura, Carole Tiberius, Iztok Kosem, Jelena Kallas and Miloš Jakubíček (eds), *Proceedings of the eLex 2023 Conference: Electronic Lexicography in the 21st Century.* Brno: Lexical Computing, 19–38. https://www.youtube.com/watch?v=rQC3Rz04b20

U.S. Copyright Office. 2023. 'Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence.' *Federal Register* 16 March 2023. https://www.federalregister.gov/d/2023-05321

van der Meer, Geart. 1996. 'The Treatment of Figurative Meanings in the English Learner's Dictionaries (OALD, LDOCE, CC and CIDE)' In Gellerstam, Martin, Jerker Järborg, Sven-Göran Malmgren, Kerstin Norén, Lena Rogström and Catalina Röjder Papmehl (eds), *Euralex '96 Proceedings I–II, Papers Submitted to the Seventh Euralex International Congress on Lexicography in Göteborg, Sweden*. Gothenburg: Department of Swedish, Göteborg University, 423–429.

Varantola, Krista. 2002. 'Use and Usability of Dictionaries: Common Sense and Context Sensibility?' In Corréard, Marie-Hélène (ed.), *Lexicography and Natural Language Processing. A Festschrift in Honour of B.T.S. Atkins*. s.l.: Euralex, 30–44.

Vossen, Piek (ed.). 1998. *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*. Dordrecht: Kluwer Academic Publishers.

Vossen, Piek. 2022. 'ChatGPT Is a Waste of Time.' *VU Magazine* 22 December 2022. https://vumagazine.nl/professor-piek-vossen-chatgpt-is-a-waste-of-time?lang=en

Wiegand, Herbert E. 1989. 'Der Begriff Der Mikrostruktur: Geschichte, Probleme, Perspektiven' In Hausmann, Franz J., Oskar Reichmann, Herbert E. Wiegand and Ladislav Zgusta (eds), *Wörterbücher/ Dictionaries/ Dictionnaires. Ein Internationales Handbuch Zur Lexikographie/ An International Encyclopedia of Lexicography/ Encyclopédie Internationale de Lexicographie. (Handbücher zur Sprach- und Kommunikationswissenschaft, HSK 5.1)*. Berlin: Walter de Gruyter, 409–462.