

## The Role of Chatbots, Google and Online Dictionaries in Meaning Comprehension and Retention

ANNA DZIEMIANKO

Faculty of English

Adam Mickiewicz University, Poznań, Poland

danna@amu.edu.pl

### ABSTRACT

*Today, language learners can turn to generative AI applications in situations of linguistic deficit, rather than search the web or consult dictionaries. The aim of the study is to find out if meaning comprehension and retention are affected by whether a chatbot, Google or a monolingual learners' dictionary is consulted. Five online tools are investigated: ChatGPT-3.5, Microsoft Copilot, the Google English dictionary, the Longman Dictionary of Contemporary English (LDOCE) and the Collins COBUILD Advanced Learner's Dictionary (COBUILD). In an online experiment, 128 upper-intermediate learners of English explained the meaning of 25 English words based on reference to the five tools. Meaning retention was tested in two post-tests: immediate and delayed. The results indicate that the tool significantly affects meaning comprehension as well as immediate and delayed retention. ChatGPT-3.5 and COBUILD, followed by Copilot, prove the most beneficial for decoding meaning and remembering it immediately afterwards. Google's English dictionary and LDOCE are the least helpful for these purposes. However, AI-generated explanations do not stick in memory for long. The best delayed retention results were obtained after reference to dedicated reference works: COBUILD and LDOCE, the latter being on a par with the Google dictionary. An analysis of the explanatory content offered by the tools suggests possible reasons for the observed regularities. The study implies that AI has not outshone dictionaries yet. While today bots successfully explain meaning and foster its immediate retention, they are not on a par with dictionaries as long-term learning aids.*

*Keywords: online dictionaries; Google; generative artificial intelligence; meaning comprehension; retention*

### INTRODUCTION

The digital revolution in lexicography is said to be complete inasmuch as online dictionaries have virtually replaced paper ones (Ptasznik, 2022). However, it can be argued that the revolution is taking on a new shape as online dictionaries are being superseded by web search engines and, increasingly often, generative AI applications. In the pre-AI era, a pan-European survey into dictionary use found a general web search engine to be the first port of call in situations of lexical deficit when reference to a dictionary gives no immediate help (Kosem et al., 2019, p. 111). Google, with 91% of the search engine market share worldwide in July 2024 (Statcounter), unfailingly heads up any rankings of search engines. It may be expected that Google is *the* search engine that is still often used to address lexical issues, the increasing popularity of AI notwithstanding. While chatbots are created to converse with, a search engine indexes web pages to help users find information on their own. Importantly, Google fulfils searchers' needs directly on the Google website, rather than by clicking through to websites (e.g., via short snippets of text meant to quickly answer a specific query). When Google is queried about lexical meaning, it snips out the content of *Google's English dictionary* provided by Oxford Languages and gives it at the top of search results (cf. Nichols, 2023). The dictionary becomes available through the *define* (or similar) operator in Google Search, or by simply searching a word. Unlike LDOCE and

COBUILD, it is not targeted at language learners; the content comes from native speakers' dictionaries: the *Oxford Dictionary of English* and the *Oxford Thesaurus of English*. Considering users' long-standing gravitation towards googling the meaning of unknown words, it might be worthwhile to compare the usefulness of Google dictionary snippets with online learners' dictionaries and AI-driven tools.

As for the latter, OpenAI's ChatGPT represented a revolutionary change in human interaction with digital tools, mainly due to the sequential user interface (1D). It is argued that sequential access is more intuitive for humans. It is simply more natural to ask what a word means (and get an answer next) than access a book or a web browser to look up the word and search for the right meaning on one's own, as is the case with 2D flat things (paper or computer/phone screens (Tavast, 2023, as cited in de Schryver, 2023). Besides, ChatGPT can create natural-sounding, linguistically correct, cohesive and coherent responses.

The year 2023 was particularly transformative for search. Search engine algorithms were considerably improved, user experience enhanced, and AI integrated into apps and work environments (Davis, 2024). AI integration materialised, for example, in Microsoft Copilot, a conversational chatbot integrated into Microsoft 365, which can be accessed on any device. The chatbot, built on multimodal large language models (LLM) GPT-4 and text-to-image model DALL-E3, can answer complex questions, summarise information, provide inspiration and generate text. It remains to be seen if it becomes a serious competitor to the popular free version of OpenAI's ChatGPT (at the time of the study: ChatGPT-3.5).

## LITERATURE REVIEW

There is already a body of literature on whether generative AI can assist (or replace) lexicographers (e.g., de Schryver, 2023; Jakubiček & Rundell, 2023; Lew, 2023; Rundell, 2023). de Schryver (2023), for example, enthuses that AI can already take over dictionary makers' tasks and successfully supersede both lexicographers and post-editing tools. de Schryver and Joffe (2023) demonstrate that ChatGPT can be useful at various stages of dictionary making from building and analysing corpora, through dictionary writing, to publishing and even marketing the final product. Rundell (2023, p. 9), by contrast, holds that ChatGPT does not herald the end of the discipline. When employed to compile entries, it has problems with sense identification, grammatical information or example sentences. However, definitions generated by ChatGPT are found relevant and accessible, even if slightly lengthy (Jakubiček & Rundell, 2023, p. 526). Lew (2023) shows that with proper training, definitions in the COBUILD style (which incorporate the word being defined) may well be written by a generative transformer.

The conclusions about leveraging ChatGPT in dictionary-making follow from output evaluation by experts, who use "their practical lexicographic experience as a reference, ...[and] then reflect on the tool's performance" (Rees & Lew, 2024, p. 52). Unfortunately, little is known about the relative usefulness of AI-generated explanations and dictionary content for language learners. Recently, Rees and Lew (2024) tested the usefulness of ChatGPT and dictionary definitions from the *Macmillan English Dictionary* (MED) in a lexically-oriented multiple-choice reading task, where the comprehension of new words was investigated. A total of 43 L2 users of English at B2 to C1 proficiency levels (in CEFR) at a Spanish university took part in the experiment. They read a text and answered multiple-choice questions about the meaning of nine target items. One third of the participants consulted MED definitions, another third – definitions

generated by ChatGPT, and the rest had no access to dictionary explanations. No difference in comprehension was observed between learners with MED definitions and those produced by ChatGPT.

However, Rees and Lew (2024) did not explore learning; they just compared the usefulness of MED and ChatGPT definitions for decoding meaning. Besides, MED definitions were isolated from dictionary entries, where examples and synonyms normally convey semantic information. ChatGPT users may not see definitions alone, either, when conversing with the bot. Even the prompt “explain X”, employed in the study by Rees and Lew (2024), may generate output including examples, etymology or usage notes. Thus, a user study is needed where definitions would not be extracted from their typical context, and which would also explore retention effects (both immediate and delayed).

A recent study by Ptasznik et al. (2024) meets the first requirement. To evaluate the effectiveness of ChatGPT and LDOCE in language reception and production, participants accessed full LDOCE entries on their phones and were allowed to interact with ChatGPT as they wished on desktop computers. A sample of 223 upper-intermediate to advanced (B2-C1 in CEFR) university students was involved. In the receptive task, they were requested to understand 20 English sentences and translate them into Polish, underlining low-frequency nouns. In the productive one, 20 Polish sentences had to be translated into English using the given highly polysemous English verbs. About half of the participants did both tasks with the help of a mobile web app version of LDOCE, while the other half used the free version of ChatGPT 3.5. The results reveal significantly better performance on both tasks among ChatGPT users. Unfortunately, neither lexical retention nor long-term learning was measured; only real-time performance was measured. Also, interaction with ChatGPT and its output was not controlled. This makes it virtually impossible to analyse the content of the responses generated by AI and reflect on its effect.

Overall, what is still missing from the field is an experimental study into the relative usefulness of AI assistants (not only ChatGPT) and dedicated reference works (online dictionaries) for language learning. It would also be recommendable to create a design in which the content provided by the tools could be analysed to explain the results. It is perfectly feasible for online dictionaries, whose output is usually frozen, and repeating a given query yields the same response. However, AI-generated output is non-deterministic. As it is produced ‘on the fly’, with nothing analysed and synthesised in advance, the same prompts may yield different answers (Jakubiček & Rundell, 2023, p. 511). Obviously, it is even more challenging to analyse the content supplied by AI tools when participants are allowed to interact with them as they wish.

Expert lexicographers criticised AI-generated entries for redundancy and repetitiveness. Definitions are considered wordy and longer than desirable, and examples strike as repetitive and lacking diversity (Jakubiček & Rundell, 2023, p. 516; Lew, 2023, p. 6). Redundancy, which occurs when the same information is presented simultaneously in different forms (Kalyuga & Sweller, 2021, p. 212), is usually considered detrimental to learning. In cognitive load theory (Sweller, 2005), it increases extraneous cognitive load, generated by the need to process the needlessly duplicated information. The increase imposes heavy demands on the limited resources of working memory and wastes them on mental activities unrelated to learning goals. That is why, according to the redundancy principle (Mayer, 2009), redundant information should best be omitted (Kalyuga & Sweller, 2021, p. 212). No wonder, then, that duplicating information is traditionally advised against in dictionaries (Lew & Tomczak, 2022, p. 81).

On the other hand, repetition is considered a route to understanding. Put differently, “it is with multiple exposures that understanding takes place” (Larsen-Freeman, 2012, p. 198). Also, the more repetition, the better recall; repetition is important for working memory and paves the way for information transfer to long-term memory (Larsen-Freeman, 2012, p. 199). However, in complexity theory, language is dynamic; it grows and organises itself from the bottom up. For complexity theorists, what matters is iteration, or inexact replication, a transformative and approximative process which highlights mutability (Larsen-Freeman, 2012, pp. 201-202). Iteration is particularly important for making and negotiating meaning. When content is not repeated in exactly the same way, options are created which give language users choice in how they want to make meaning, which is conducive to learning (Larsen-Freeman, 2012, pp. 206-207). This implies that there might also be some mileage in what seems to be redundant or repetitive/iterative.

## AIM

The study aims to determine if online tools affect meaning comprehension and retention. Free versions of five tools are investigated: LDOCE, COBUILD, the Google dictionary, ChatGPT-3.5 and Microsoft Copilot. Two learners’ dictionaries were selected because of significant differences in defining meaning. LDOCE employs mostly the classical (analytical) defining style, where the genus term (hyperonym) is followed by differentia specifica (a set of distinguishing features), e.g., *remiss: careless because you did not do something that you ought to have done*. COBUILD, in turn, features contextual (full-sentence) definitions, modelled on conversational explanations in natural language. Most typically, they include the word being defined (definiendum) in its usual context, e.g., *remiss: If someone is remiss, they are careless about doing things which ought to be done* (cf. Hanks, 1987). Long before the AI era, the defining format (full-sentence vs. classical) was found not to affect comprehension (Cumming et al., 1994). It is not known if, 30 years later, it still makes no difference to the next user generation. Google, as noted above, tends to be queried about language issues, including those not resolved by dictionaries. ChatGPT-3.5, a versatile conversational interface handling various topics, is a major tool in AI-enhanced computer-mediated communication. Copilot, in turn, which leverages the latest LLM, is claimed to have a wider knowledge scope and be more powerful and accurate than ChatGPT-3.5 and is said to be “the only way to use GPT-4 for free at this time” (Diaz, 2023).

The following research questions are asked:

1. Does meaning comprehension depend on whether ChatGPT-3.5, Copilot, Google, LDOCE or COBUILD is consulted?
2. Do these online tools affect immediate and delayed retention of meaning?

## METHODS

### INSTRUMENTS

To achieve the aims of the study, an online (Moodle-based) experiment was conducted. It consisted of a pretest, a main test and two retention tests (immediate and delayed), all based on 25 difficult words from de Schryver (2023). Prior to the study, 21 students of English (B2 in CEFR), who represented the same population as participants in the main test (see the next section), read the

paper and took down the words which they did not know. Out of them, 25 most frequently listed ones were selected: *augment, blip, boilerplate, circumspect, clunky, distil, emulate, fodder, fungible, luminary, meticulous, mitigate, morph, noughties, ostensibly, pan out, quagmire, regurgitate, remiss, rogue, sentient, tadpole, tweak, vet, wrest*.

de Schryver's (2023) sentences featuring the words were employed. Some of them were only shortened, but heavily context-dependent ones were adapted so that they could function independently. To prevent contextual guessing and encourage the use of the investigated tools, the sentences could not reveal the meaning of the test items. To illustrate, de Schryver's (2023, p. 363) sentence: *The task to rewrite (simplify) definitions (from adult to child level) is less impressive, as the result is often clunky or awkward*, which makes it possible to infer the meaning of *clunky* from the context. Its modification, *the result is often clunky*, does not.

In the pretest, the sentences were listed, and the target items were underlined. The participants explained the words in writing by providing L1 (Polish) equivalents without reference to any tools. The aim was to test the participants' lexical knowledge prior to the study. While the risk of the target items being familiar to them was largely reduced by the exploratory stage involving 21 students (who did not take part in the experiment), it was not eliminated. The pretest made it possible to identify such cases (which did not exceed 0.2%) and exclude them from further analysis.

The immediate and delayed post-tests were the same as the pretest, with the participants supplying L1 equivalents from memory immediately after the treatment and two weeks later, respectively. The aim was to test how much they had learnt from the tools.

In the main test, the participants saw the same sentences with underlined words and provided L1 equivalents based on reference to screenshots from the five online tools investigated in the study. They were expected to read the screenshots rather than access the tools on their own to prevent individual differences in prompts, interfaces or browsers. The screenshots also saved logging in or setting up a user account for those who had not registered before to converse with ChatGPT-3.5 or Copilot. Additionally, they made it possible to carefully control the participants' internet activity and reduce confusion or intimidation among those who might not have been fully conversant with the tools. Importantly, they also enabled an analysis of online explanatory content and its effect on the obtained results, which would be a serious challenge if users had interacted with bots or searched the web at will.

To elicit responses from the chatbots, zero-shot prompts were provided in the general interactive text interface. The zero-shot setting "only uses a natural language description or invocation of the task to be performed"; no demonstrations or examples are offered to better steer the model to successfully carry out the task (Brown et al., 2020, p. 2). The zero-shot prompts, the same for all target items, consisted of the instruction *explain* followed by a target item (e.g., *the verb X*). The same prompts were used to search Google, from which only the snippets of *Google's English dictionary* were screenshotted for the study.

The participants were not informed which tools they were consulting. Any explicit information which could betray the sources was removed. Noise (like advertisements) was eliminated, too. Only the content visible on the screen immediately after entering the prompt was captured in the screenshots. No cross-references were followed, nor were expand buttons clicked to access information collapsed in parent nodes. In this way, the typical behaviour of a learner who wants to find meaning fast and is unwilling to click was mimicked. It was ascertained that all the sources provided explanations of *the* senses in which the words were used in the test.



To assign the tools to the test items, a counterbalanced design was employed; each five-word item was paired with a different tool. Tool assignment was rotated across the words in five test versions (A-E, see Table 1 and the Appendix).

TABLE 1. Assignment of the tools to target items in the main test

Word sets of 5 / Test version	A	B	C	D	E
Set 1	Copilot	GPT-3.5	Google	LDOCE	COBUILD
Set 2	GPT-3.5	Google	LDOCE	COBUILD	Copilot
Set 3	Google	LDOCE	COBUILD	Copilot	GPT-3.5
Set 4	LDOCE	COBUILD	Copilot	GPT-3.5	Google
Set 5	COBUILD	Copilot	GPT-3.5	Google	LDOCE

A participant dealt with one test version, and thus with all five tools in equal measure. In this way, the potential effect of learner differences was reduced. Also, each word was accompanied by screenshots from all the tools, which minimised the item effect.

#### PARTICIPANTS AND PROCEDURES

A total of 128 upper-intermediate Polish learners of English (B2 in CEFR), 31 male and 97 female, at Adam Mickiewicz University in Poznań, participated in the experiment (test version A–25 students, B–27, C–24, D–26, E–26). Their level was established on the basis of the exam taken at the end of the academic year and the teaching materials used in class. The experiment was conducted in regular student groups during standard classes held in a computer lab. The students were seated at uniform desktops. First, they were requested to do the pretest and supply L1 equivalents without reference to any sources. Immediately after the pretest, they took the main test. To get assigned to the test versions, a student drew one card with a letter from A to E from a set prepared beforehand. They were then requested to access the test version with the letter they had drawn. In the main test, they provided L1 equivalents of the words underlined in sentences relying only on the supplied screenshots. Right after the main test, they took the immediate post-test, where L1 equivalents had to be given from memory. The same post-test was repeated two weeks later (see Figure 1). At each stage, the sequence of test items was randomised. Throughout the whole experiment, the participants' internet activity was closely monitored. The experiment was self-paced; the participants decided how much time to devote to each test item.<sup>1</sup> However, they had to proceed sequentially through the questions and could not go back or modify any already supplied answers.

<sup>1</sup> On average, they needed 49 minutes 22 seconds to complete the tasks (max: 54 min 27 s, min: 46 min 4 s).

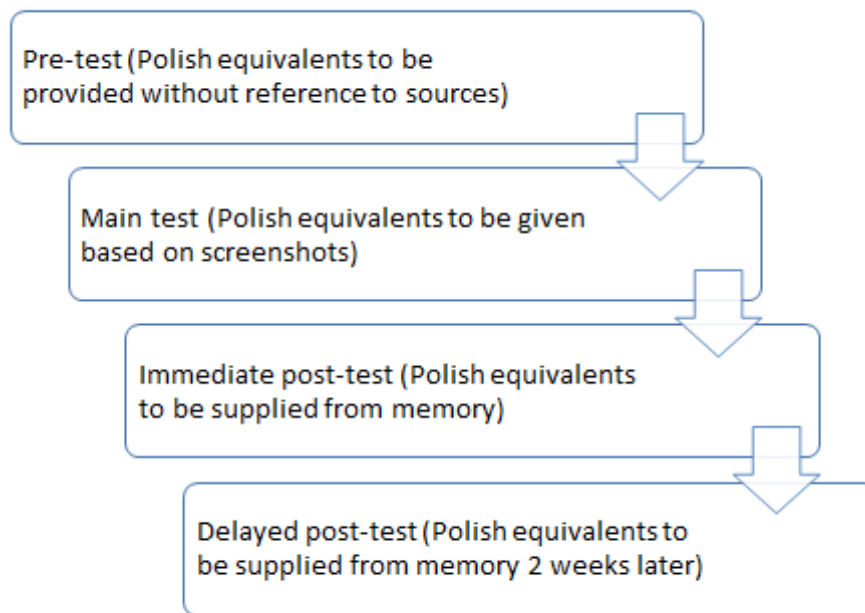


FIGURE 1. A flowchart illustrating the research procedure

## SCORING

The responses were evaluated independently by two raters, native speakers of (*Polish*) proficient in English (C2), who had considerable professional experience in teaching EFL at the university level. The participants' answers scored one point if they met two evaluation criteria. First, they had to describe the meaning of the target words in the sentences shown in the main test. That proved to be an important prerequisite; some participants accounted for senses which did not fit the sentence cues or neglected the part of speech. For example, instead of explaining the meaning of the verb *to vet*, explanations of *a vet* as *a veterinarian* or *a veteran* were given. Second, the answers had to coincide with the equivalents of the test items found in the most prestigious and popular English-Polish dictionaries, both on paper and online.

In a few cases (less than 0.8%), Polish explanations of the target items rather than equivalents were provided. As they were always translations of the supplied content, they did not score any points. Likewise, meaning explanations in English, given in less than 0.5% of all cases, were copied from the screenshots and were not awarded any points. The raters reached perfect agreement in evaluating the participants' responses ( $r=1.00$ ).

## RESULTS

### EXPERIMENTAL DESIGN AND DATA ANALYSIS

There was one within-group factor, *tool*, with five levels: ChatGPT-3.5, Copilot, Google, LDOCE and COBUILD. First, the effect of *the tool* on three dependent variables (*meaning comprehension*, *immediate retention* and *delayed retention*) was investigated using a repeated-measures MANOVA. Next, to further explore differences in meaning comprehension, immediate and delayed retention generated by the five tools, repeated-measures MANOVAs were computed for each dependent variable in turn. MANOVAs were used instead of univariate repeated-measures

ANOVAs because of the number of levels of the within-group factor (over two), in which case the multivariate approach is recommended. Significant MANOVA results were further analysed with the Bonferroni test. All the analyses were conducted using the Statistica package.

#### MEANING COMPREHENSION, IMMEDIATE AND DELAYED RETENTION

The MANOVA results indicate a statistically significant role of *tool* (Wilks's  $\lambda=0.01$ ,  $F=106.64$ ,  $p<0.001$ , partial  $\eta^2=0.784$ , see Figure 2).

First, meaning comprehension depended on the tool (Wilks's  $\lambda=0.12$ ,  $F=39.23$ ,  $p<0.001$ , partial  $\eta^2=0.414$ ). Comprehension proved the most (and comparably) successful when COBUILD (73.99%) and ChatGPT-3.5 (71.02%) were consulted ( $p=1.00$ ); it was much more successful for these two tools than the remaining three ( $p<0.01$ ). Copilot (61.11%) had a significant advantage over the Google dictionary (53.66%,  $p=0.01$ ) and LDOCE (53.48%,  $p=0.00$ ). The last two brought about comparably low comprehension ( $p=1.00$ ).

Second, the tool affected immediate retention (Wilks's  $\lambda=0.06$ ,  $F=80.63$ ,  $p<0.001$ , partial  $\eta^2=0.503$ ). ChatGPT-3.5 (52.64%) and COBUILD (50.96%) helped to learn meaning most successfully ( $p=1.00$ ); they were significantly more useful for immediate retention than the other tools ( $p<0.01$ ). Over two-fifths (42.68%) of meaning explanations were remembered when Copilot had been consulted. Significantly less ( $p<0.00$ ), about one third of the semantic content, was retained when the Google dictionary (31.84%) and LDOCE (32.55%) had been used, with no difference between them ( $p=1.00$ ).

Third, delayed retention depended on the tool, too (Wilks's  $\lambda=0.10$ ,  $F=46.92$ ,  $p<0.001$ , partial  $\eta^2=0.437$ ). COBUILD (32.10%) and LDOCE (29.13%) helped to remember the meaning of about one third of the words two weeks after exposure ( $p=0.07$ ). Delayed retention following Google dictionary lookup (28.45%) was as successful as after reference to LDOCE ( $p=1.00$ ), but significantly less successful than based on COBUILD ( $p=0.01$ ). The least was remembered when ChatGPT-3.5 (21.43%) and Copilot (18.88%) had been consulted ( $p=0.19$ ). The delayed retention rate (of about one-fifth) in these two cases was much lower than when the other three tools had been used ( $p<0.001$ ).

When immediate and delayed retention results are juxtaposed, it turns out (quite predictably) that for each tool, immediate retention was better than delayed retention. Interestingly enough, the difference was the smallest for LDOCE and the Google dictionary. In their case, as much as 90% of what was remembered immediately after exposure was retained two weeks later (LDOCE:  $29.13 \times 100 / 32.55 = 89.50$ , Google:  $28.45 \times 100 / 31.84 = 89.37$ ). COBUILD was the second-best long-term learning assistant. For this dictionary, delayed retention (32.10%) constituted over 60% of immediate retention (50.96%;  $32.10 \times 100 / 50.96 = 62.99$ ). By contrast, ChatGPT-3.5 and Copilot helped long-term retention the least, inasmuch as about 40% of what was remembered after consulting them was retained two weeks later (ChatGPT-3.5:  $21.43 \times 100 / 52.64 = 40.71$ ; Copilot:  $18.88 \times 100 / 42.68 = 44.25$ ).



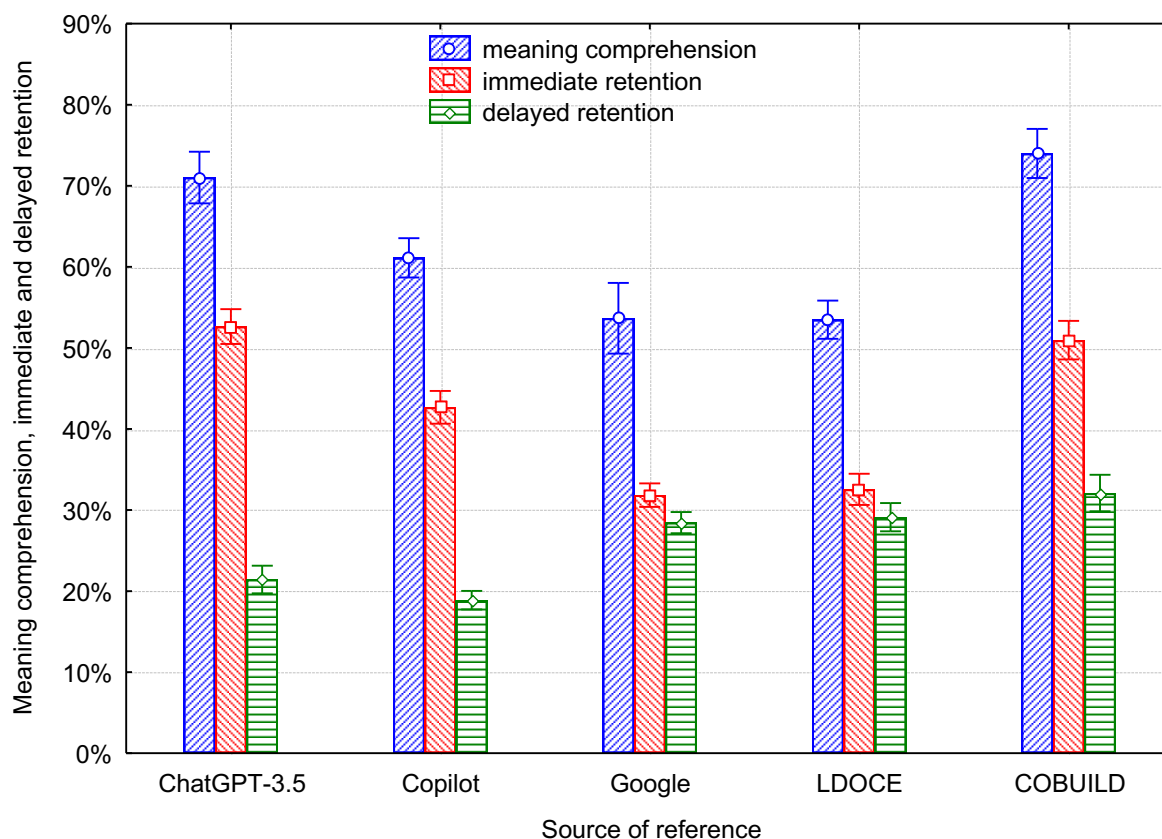


FIGURE 2. Meaning comprehension, immediate and delayed retention by the tool.  
Vertical bars denote 95% confidence intervals

### SUMMARY OF THE RESULTS

Table 2 gives a symbolic presentation of the results. For each dependent variable, numbers from 1 to 3 are assigned to the tools depending on the outcome of statistical analyses, with one standing for the best results and 3 for the worst. The tools which produced statistically comparable results are given the same number (e.g., 1). By the same token, different numbers assigned to the tools mean that the results are significantly different.

TABLE 2. Symbolic representation of the results

Tool	Meaning comprehension	Immediate retention	Delayed retention
ChatGPT-3.5	1	1	3
Copilot	2	2	3
Google	3	3	2
LDOCE	3	3	1, 2
COBUILD	1	1	1

The study gives affirmative answers to both research questions; the tools affect meaning comprehension (Q1) as well as immediate and delayed retention (Q2). In order to understand the meaning and remember it immediately afterwards, COBUILD and ChatGPT-3.5 are the most recommendable, followed by Copilot. Google's English dictionary and LDOCE are the least

helpful. However, in the long run, meaning is remembered best when COBUILD and LDOCE are consulted, the latter being as useful as Google. Both chatbots assist with delayed retention the least and are surpassed by the dedicated reference works. Quite interestingly, in all three respects under discussion (meaning comprehension, immediate and delayed retention), reference to COBUILD remains the most beneficial.

## DISCUSSION

The research confirms the conclusion drawn by Ptasznik et al. (2024) that ChatGPT performs better on reception than LDOCE and reveals that the same holds for Copilot. It also shows that when AI-generated and dictionary definitions are not isolated from the broader explanatory context, the tool does matter for comprehension, in contrast to what Rees and Lew (2024) found. However, the current experiment involved more tools than the previous studies and investigated their potential for promoting learning. Possible reasons for the observed regularities are discussed below.

The best comprehension and immediate retention achieved after reference to COBUILD, ChatGPT-3.5 and Copilot (rather than LDOCE and the Google dictionary) might result from the nature of semantic explanations. As pointed out above, COBUILD (contextual) definitions typically include the word defined, e.g., *tadpole: Tadpoles are small water creatures which grow into frogs or toads*. The corresponding definitions in LDOCE and the Google dictionary were analytical: *a small creature that has a long tail, lives in water, and grows into a frog or toad* (LDOCE), and *the tailed aquatic larva of an amphibian (frog, toad, newt, or salamander), breathing through gills and lacking legs until the later stages of its development* (Google). The chatbots tended to follow the COBUILD strategy and incorporated the word defined in a definition, which showed its regular grammatical characteristics and usage patterns, e.g., *A tadpole is the aquatic larval stage in the life cycle of an amphibian.... Tadpoles typically hatch from eggs laid in water..... As they go through metamorphosis, tadpoles gradually develop into limbs ...* (ChatGPT-3.5). However, some AI-generated definitions might not appear to be plainly the COBUILD strategy where the definiendum is encoded in a phrase that shows its usage, e.g., *tadpole is a noun that means the aquatic larva of an amphibian, such as a frog or a toad, that has a rounded body, a long tail, and gills for breathing. It undergoes a metamorphosis to become an adult* (Copilot). The definiendum itself is discussed here in the abstract (*Tadpole is a noun that means...*), and meaning is described in the form of rules and principles without showing the typical lexico-grammatical context (cf. *The word “augment” is a verb that means to increase the size, value, quantity, or intensity of something. It is often used to describe the process of making sth greater or adding to it in order to enhance its overall quality or characteristics* (ChatGPT-3.5). Such rule-based definitions were sometimes combined with the recognisably COBUILD defining style, where the word looked up is shown in its usual context, e.g., *“Luminary” is used to describe a person who is well-known, distinguished, and outstanding in a particular field or activity [rule-based]. A luminary is often a person of great influence, expertise, or achievement [contextual]. A luminary is someone who is highly regarded and has made significant contributions to their field [contextual]* (ChatGPT-3.5). Importantly, when selection preferences are not relevant (e.g. for function words), COBUILD employs rule-based definitions, too. In fact, “all COBUILD explanations ... can 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, 1987, p. 130). Thus, the

chatbots' definitions, which do not encode the target items in their typical contexts but explain meaning in the abstract by giving rules, are also a COBUILD strategy.

To see the big picture, the content of explanations used in the experiment is quantified in Table 3. The digits stand for the number of test items for which the specific information categories (e.g., contextual definitions) were present, and not for the number of the categories themselves, which were sometimes multiplied for an item (e.g., a tool gave a few contextual definitions of one word).

TABLE 3. Content of explanations (in numbers)

Information category	ChatGPT-3.5	Copilot	Google	LDOCE	COBUILD
rule-based definitions	24	17	0	0	0
contextual definitions	15	4	0	2	23
analytical definitions	0	4	25	23	1
synonyms	0	12	16	6	1
antonyms	0	4	0	0	0
examples	8	23	24	24	23
usage notes	9	3	0	0	0
etymology	2	11	0	0	0
semantic conclusion/summary	10	0	0	0	0

AI-generated definitions represented mostly the COBUILD strategies (contextual and rule-based). Only Copilot offered analytical definitions in four cases. COBUILD consistently supplied contextual definitions except for *fungible* and *ostensibly*, where an analytical definition and synonyms were given, respectively.<sup>2</sup> Google's definitions were always analytical. Those in LDOCE were analytical, too, except for *ostensibly* and *remiss*, defined in context. The conversational COBUILD defining format might be a reason why both chatbots (and COBUILD itself) contributed to comprehension and immediate retention much more than the Google dictionary and LDOCE. If so, the results diverge from those by Cumming et al. (1994), which point to no difference in comprehension based on contextual and analytical definitions.

The similarities in defining strategies between the bots and COBUILD notwithstanding, AI-generated definitions evidently diverged from those in COBUILD; they were markedly longer and more lexically complex. They often spanned several sentences and included infrequent words from beyond the defining vocabulary of either COBUILD or LDOCE.<sup>3</sup> The language of ChatGPT-3.5 explanations tended to be even more difficult than that of Copilot, which sometimes aggregated definitions from learners' dictionary websites: *dictionary.cambridge.org*, *oxfordlearnersdictionaries.com* or *merriam-webster.com*.<sup>4</sup> Quite surprisingly, the higher degree of lexical complexity in AI-generated definitions did not entail greater comprehension problems. Interestingly, Google definitions, based on native speakers' dictionaries, were also lexically more difficult than those in learners' dictionaries. Yet, the Google dictionary proved much less beneficial for comprehension than the bots.

Possibly, the length and intrinsic variety of AI-generated explanations might have compensated for their lexical difficulty.<sup>5</sup> On average, AI explanations were over three times longer

<sup>2</sup> They come from the *Collins English dictionary*, not the *Collins COBUILD Advanced Learner's Dictionary*.

<sup>3</sup> LDOCE uses 2000 common words in definitions. COBUILD makes no restrictions on its defining vocabulary, but achieves simplicity and naturalness in definitions by using simpler and more common words than those defined (Kamiński, 2021, p. 199).

<sup>4</sup> The first two dictionaries have defining vocabularies, while the last one follows the COBUILD strategy for vocabulary control (Kamiński, 2021).

<sup>5</sup> Cf. Kamiński (2021, p. 238) found that definition length is positively correlated with comprehension, which means that longer definitions facilitate comprehension more than short ones.

than those in LDOCE or Google (see Table 4). When the last two sources were looked up, users typically saw a single analytical definition followed by examples of usage and (sometimes) synonyms (Table 3). ChatGPT-3.5, by contrast, supplied contextual definitions *together with* rule-based ones for as many as 14 words. Copilot gave one definition at a time, but exemplified almost three times more words than ChatGPT-3.5 (23 vs 8). In contrast to ChatGPT-3.5, which neglected synonymy, it also offered synonyms for 12 words, four of which were also given antonyms. On top of that, both chatbots included some unique features, like etymology or usage notes, absent from dictionary entries. ChatGPT-3.5 provided semantic conclusions/summaries, and Copilot clearly indicated the sources drawn on, made farewells to its user and expressed readiness to continue chatting if the need arises (see Figures 3 and 4).

TABLE 4. Mean length of explanations (in words)

ChatGPT-3.5	Copilot	Google	LDOCE	COBUILD
160.3	152.7	45.6	47.8	83.1

The image shows a screenshot of a chatbot interface (Copilot) explaining the word 'mitigate'. The text is as follows:

Mitigate is a verb that means **to make something less harmful, unpleasant, or bad**. For example, you could say "Getting a lot of sleep and drinking plenty of fluids can mitigate the effects of the flu."<sup>1</sup>

Some synonyms of mitigate are **relieve, alleviate, ease, help, and soothe**<sup>2</sup>. Some antonyms are **aggravate, worsen, and intensify**<sup>2</sup>.

The word mitigate comes from the Latin word **mitigare**, which means **to soften**. It originally had the same meaning in English, but later changed to mean **overcareful about detail**<sup>3</sup>.

Some examples of using mitigate in a sentence are:

- Drainage schemes have helped to mitigate this problem.
- He seems like he's not paying attention, but don't be fooled - he's very thorough and meticulous.
- She mitigated the offense by apologizing sincerely.

I hope this helps you understand the meaning and usage of mitigate. If you have any other questions, feel free to ask me. 😊

At the bottom, there are three numbered sources:

- 1 [bing.com](https://www.bing.com)
- 2 [merriam-webster.com](https://www.merriam-webster.com)
- 3 [dictionary.cambridge.org](https://dictionary.cambridge.org)

Annotations with blue boxes and arrows point to specific parts of the text:

- rule-based def. + example** points to the first sentence.
- synonyms + antonyms** points to the second sentence.
- etymology** points to the third sentence.
- more examples** points to the list of examples.
- farewell note** points to the closing sentence.
- sources** points to the numbered list of sources.

FIGURE 3. The explanation of *mitigation* by Copilot (with annotation)

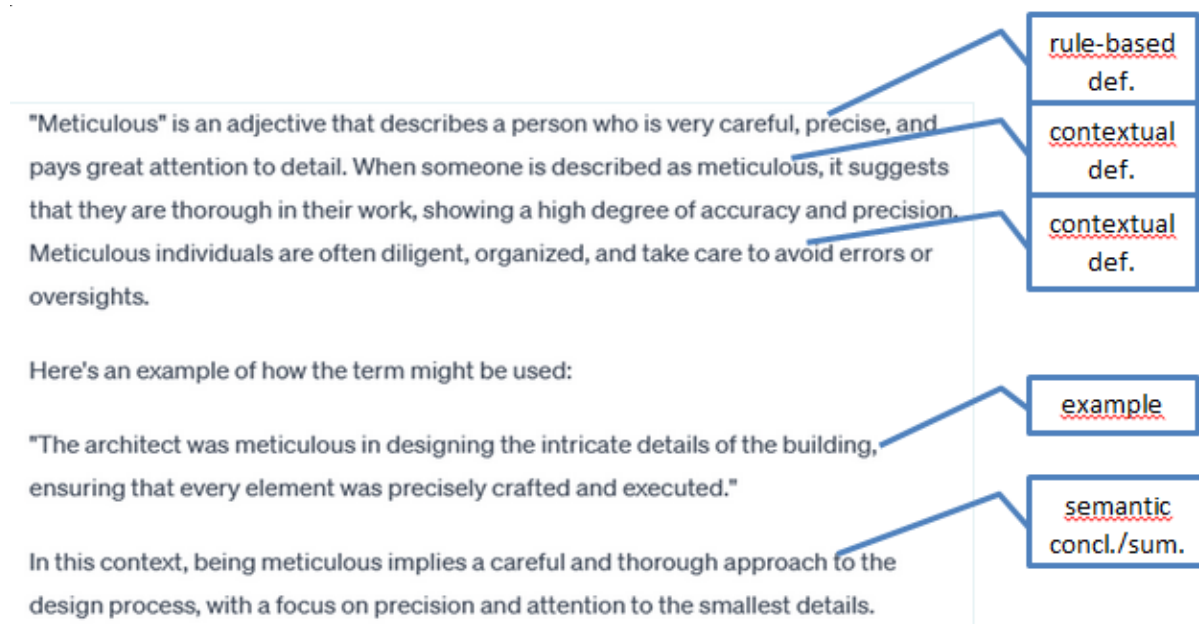


FIGURE 4. The explanation of *meticulous* by ChatGPT-3.5 (with annotation)

Figure 4 shows that ChatGPT-3.5 excelled at multiplying definitions; a rule-based definition of *meticulous* is followed by two contextual ones. Semantic summaries/conclusions, in turn, explained some specific examples of usage (like in Figure 4), or closed comprehensive explanations, e.g., *in summary, 'rogue' typically conveys a sense of deviance, unpredictability, or unscrupulous behaviour, but its specific meaning depends on the context in which it is used*. Either way, they might relieve the user of the interpretative burden inherent in dictionary lookup. A usage note (not illustrated in the Figures) highlighted register or pragmatic contexts of usage, e.g., *The phrase is commonly used in informal contexts* (Copilot: *pan out*) or *The term is commonly used in discussion about politics, international relations and complex social issues* (ChatGPT-3.5: *quagmire*).

Possibly, the variety inherent in the AI-generated explanations offset their lexical complexity. When a definition contained rare words, other information categories, like synonyms, antonyms, multiple examples of usage, or etymology, might help grasp meaning and retain it in working memory more successfully than brief explanations in LDOCE or the Google dictionary. Also, several definitions of different types in ChatGPT-3.5, often coupled with a semantic conclusion, might be more beneficial for comprehension and immediate retention than one definition at a time in Copilot.<sup>6</sup>

Extended AI-generated explanations, in contrast to more concise dictionary entries, were likely to create redundancy. As noted in the Introduction, unnecessarily duplicated information can overtax the limited cognitive capacity of working memory and hinder learning (e.g., Paas & Sweller, 2012). Yet, it has recently been shown that this effect applies not so much to content

<sup>6</sup> Lexicographic studies suggest that compensatory and synergy effects among dictionary entry elements influence language comprehension and learning. Dziemianko and Lew (2013) found that part-of-speech identification is affected by whether definitions are shown alone or in more naturalistic dictionary entries featuring all basic entry components (including POS and register labels, pronunciation and examples of usage). Laufer (1993) concluded that definitions and examples given in entries support comprehension significantly more than either definitions or examples alone. Dziemianko (2014) observed that collocations are learned better when given twice in the entry (before and within examples) than when they appear only once (in examples). It stands to reason that similar compensatory mechanisms and synergy effects might also work for AI-generated output.



redundancy, which actually reduces cognitive load and increases learning outcomes, as modal redundancy, concerned with an excessive number of modalities in which information is displayed (e.g., animation, narration, and written text; Albers et al., 2023). The latter does not apply to the current study, where information was presented in one modality (displayed as on-screen text).<sup>7</sup> Duplicative content, in turn, did occur in AI-generated explanations (see Figure 4). Such contentual redundancy might be a reason for better comprehension and immediate retention. It might appeal to the participants because natural speech directed to learners is typically more redundant than speech directed to proficient speakers (Tal et al., 2023). Redundancy should also increase when comprehension is difficult (Tal et al., 2024), and it is comprehension problems that the current experiment investigated.

It is also noteworthy that the AI-generated content resembled iteration rather than exact repetition. It was heterogeneous and variable, and gave learners different options for making meaning. It might have represented language as a dynamic (rather than fixed) system more successfully than frozen dictionary content. Such observations seem to support the role of iteration in language comprehension and learning, and in more general terms, the complexity theory (Larsen-Freeman, 2012).

It seems surprising that the best delayed retention was achieved when COBUILD, LDOCE and the Google dictionary had been consulted. Their explanations were the shortest and the least redundant. The odds are that the participants (university students at the upper-intermediate level) were more used to lexicographic conventions than the AI style of explanations. Maybe for this reason, they could better recall the meaning explained by the dedicated reference works in the long run.

Ptasznik et al. (2024) hypothesise that the retention of lexical information might be enhanced by chatting with bots, as it frees learners “from the error-prone and time-consuming task of locating relevant information in a ‘generic’ dictionary entry” and provides “immediately relevant feedback without the accompanying noise.” In the controlled conditions of the current study, all screenshots saved the participants the search and ensured instantaneous, pertinent and noise-free feedback, but retention (both immediate and delayed) still differed across the tools. Thus, it seems that it is not so much the search path as the content of the feedback that may hold the key.

## CONCLUSIONS

de Schryver (2023) expressed concern that ChatGPT output had so far been evaluated and commented on by professional lexicographers rather than actual users, whose satisfaction with the tool had not been properly investigated. The current study does not inquire into learners’ contentment with generative systems, but evaluates their actual usefulness for decoding and learning meaning. It thus complements the contribution of the few user studies in the field (Ptasznik et al., 2024; Rees & Lew, 2024), which do not examine AI-based learning outcomes. Comparing the effectiveness of two chatbots, the Google dictionary and two reputable learners’ dictionaries, the experiment proves empirically the bots’ usefulness for meaning explanation and immediate retention. For these two purposes, ChatGPT-3.5 is as good as COBUILD and much

---

<sup>7</sup> Cf. the redundancy principle, which refers to an inefficient combination of information sources which needlessly burden a limited working memory modality (visual or verbal). In effect, learning from graphics and narration is more effective than from graphics, narration, and redundant printed text (Mayer, 2009, p. 118).

better than LDOCE and Google, while Copilot is outdone only by COBUILD. However, explanations given by the bots do not stick in memory for good. Most of what is remembered immediately after consulting them (around 60%) is lost in the long run. Meaning explained by the dedicated reference is learned much better; as much as 90% of it is retained for long in the case of LDOCE and Google, and 60% in the case of COBUILD. Evidently, AI-generated explanations successfully explain meaning and are conducive to remembering it immediately after exposure. Yet, it is dictionaries (including the Google dictionary) that help learners retain it for a longer time. It is also noteworthy that COBUILD proved to be no worse than AI in any task. Overall, in the presence of such a serious competitor as AI, learners' dictionaries and the Google dictionary do not seem to be completely overshadowed (yet?).

Like any experiment, the study has a number of limitations. Even though the tools were anonymised, original formatting might have given them away. Unfortunately, it is not known if the participants were familiar enough with the tools to identify them by format, or to what extent their prior experience affected the results. Also, providing screenshots with disabled cross-references made the interface non-interactive. It is hard to predict if any cross-references would be followed in a more natural setting, and what effect a resultant chain of lookups might have. Still, the controlled, non-interactive interface was found to be an unavoidable trade-off for the possibility of exploring the effect of the AI-generated explanatory content. Yet, without eye-tracking, it is impossible to determine how information was retrieved from the screenshots. The participants might have quickly scanned the content to home in on its relevant part, or read all of it thoroughly. Either way, aware of being studied, they might have modified their routine habits to enhance performance. The Hawthorne effect (Mackey & Gass, 2016) cannot be ruled out.

To imitate regular users' behaviour, no human input was provided to arrive at the optimal prompt; prompts were not tried out to see which works best. The AI-generated output was not cleaned up, either. Some parts of the chat's responses happened to be quite irrelevant to the query as a result. For instance, Figure 3 shows that the second bulleted example does not illustrate the looked-up word – *mitigate*, but *meticulous*. Besides, the content that cross-referred to other sources was sometimes hallucinated. To illustrate, Copilot claims that an example of *circumspect* (*He is chary of signing papers without being circumspect first*) comes from *dictionary.cambridge.org*, where it cannot be found. Some definitions proved hallucinations, too. The Copilot definition of *ostensibly* was allegedly drawn from *merriam-webster.com*, but it is not there. Strangely enough, the bots' explanations, apparently not as reliable and fine-tuned to the task as dictionary entries, proved more useful for meaning comprehension and immediate retention than LDOCE or the Google dictionary.

In the future, learners' choices of online tools in situations of lexical deficit should be investigated. The interpretative capacity of AI-generated explanations and their conversational style are likely to encourage English learners to turn to bots (rather than dictionaries or search engines) for linguistic assistance. Moreover, as shown above, chatbots can well meet learners' decoding needs, including those not fulfilled by existing dictionaries (like content interpretation). Another possible advantage of chatbots is the sequential user interface. However, only further research can reveal whether learners prefer sequential chats with AI-driven tools to 'flat' dictionary consultation (de Schyver, 2023, p. 380). It may also be interesting to see whether training language learners in using online dictionaries, querying Google and interacting with chatbots (e.g., by prompt engineering) might affect their preferences.

However intuitive chatting with bots seems, their uncured, sometimes hallucinated, and redundant output may be off-puttingly long and demand extra effort to extract proper meaning from. While contentual overlap may be advantageous for meaning comprehension and immediate retention, there must be “a point of ideal overlap of information in which as much information as possible is repeated without exceeding the maximum capacities of the learner” (Albers et al., 2023, p. 349). Going beyond that point may impede decoding and learning. In further research, the degree of redundancy in AI-generated explanations, ways of manipulating it (e.g., by prompts) and its effect on comprehension and learning should be investigated.

Finally, further research might compare the usefulness of AI-generated dictionary *entries* with entries from online dictionaries. While the former are considered quite successful by expert lexicographers (de Schryver, 2023; Lew, 2023), it might be interesting to see if they have any advantage over the latter in comprehension and learning when put to the test by language learners. It might also be revealing to involve LLM in comparative research into immediate and delayed retention of *usage*, not only meaning. It would then be possible to gain a more profound insight into how far AI can constitute an alternative (or challenge?) to learners’ dictionaries.

#### REFERENCES

- Albers, F., Trypke, M., Stebner, F., Wirth, J., & Plass, J. (2023). Different types of redundancy and their effect on learning and cognitive load. *The British Journal of Educational Psychology*, 93 (Suppl.2), 339-352. <https://doi.org/10.1111/bjep.12592>
- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877-1901.
- Cobuild. (n.d.). *COBUILD advanced English dictionary*. <https://www.collinsdictionary.com/>
- Cumming, G., Cropp, S., & Sussex, R. (1994). On-line lexical resources for language learners: Assessment of some approaches to word definition. *System*, 22(3), 369-377.
- Davis, D. (2024). Meet the 7 most popular search engines in the world. *Search Engine Journal*. Retrieved August 2, 2024, from <https://www.searchenginejournal.com/seo/meet-search-engines/>
- de Schryver, G. (2023). Generative AI and lexicography: The current state of the art using ChatGPT. *International Journal of Lexicography*, 36(4), 355-387. <https://doi.org/10.1093/ijl/ecad021>
- de Schryver, G., & Joffe, D. (2023, February 27). *The end of lexicography, welcome to the machine: On how ChatGPT can already take over all of the dictionary maker’s tasks*. [Paper presentation]. 20th CODH Seminar. Center for Open Data in the Humanities, Research Organization of Information and Systems, National Institute of Informatics, Tokyo, Japan. <https://youtu.be/watch?v=mEorw0yefAs>
- Diaz, M. (2023). How to use Copilot (formerly called BingChat). Retrieved August 2, 2024, from <https://www.zdnet.com/article/how-to-use-the-new-bing-and-how-its-different-from-chatgpt/>
- Dziemianko, A. (2014). On the presentation and placement of collocations in monolingual English learners’ dictionaries: Insights into encoding and retention. *International Journal of Lexicography*, 27(3), 259-279. <https://doi.org/10.1093/ijl/ecu012>
- Dziemianko, A., & Lew, R. (2013). *When-definitions revisited*. *International Journal of Lexicography*, 26(2), 154-175.
- Hanks, P. (1987). Definitions and explanations. In J. Sinclair (Ed.), *Looking up: An account of the COBUILD project in lexical computing* (pp. 116-136). Collins ELT.
- Jakubiček, M., & Rundell, M. (2023). The end of lexicography? Can ChatGPT outperform current tools for post-editing lexicography? In M. Medveď, M. Měchura, I. Kosem, J. Kallas, C. Tiberius, & M. Jakubiček (Eds.), *Proceedings of the eLex 2023 conference* (pp. 518-533). Lexical Computing.
- Kalyuga, S., & Sweller, J. (2021). The redundancy principle in multimedia learning. In R. Mayer & L. Fiorella (Eds.), *The Cambridge handbook of multimedia learning* (pp. 212-220). Cambridge University Press. <https://doi.org/10.1017/9781108894333.021>
- Kamiński, M. (2021). *Defining with simple vocabulary in English dictionaries*. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/tlrp.22>

- Kosem, I., Lew, R., Wolfer, S., Müller-Spitzer, C., & Silveira, M. (2019). The image of the monolingual dictionary across Europe: Results of the European survey of dictionary use and culture. *International Journal of Lexicography*, 32(1), 92-114. <https://doi.org/10.1093/ijl/ecz002>
- Larsen-Freeman, D. (2012). On the roles of repetition in language teaching and learning. *Applied Linguistics Review*, 3(2), 195-210. <https://doi.org/10.1515/applirev-2012-0009>
- Laufer, B. (1993). The effects of dictionary definitions and examples on the comprehension of new L2 words. *Cahiers de Lexicologie*, 63, 131-142.
- Lew, R. (2023). ChatGPT as a COBUILD lexicographer. *Humanities and Social Sciences Communications*, 10(1), 704. <https://doi.org/10.1057/s41599-023-02119-6>
- Lew, R., & Tomczak, E. (2022). Where should English dictionaries place multi-word expressions? A study of user consultation behaviour 3L: Language, Linguistics, Literature®, 28(3), 69-83. <http://doi.org/10.17576/3L-2022-2803-05>
- Longman. (n.d.). *Longman dictionary of contemporary English*. <https://www.ldoceonline.com/>
- Mackey, A., & Gass, S. (2016). *Second language research: Methodology and design* (3<sup>rd</sup> edition). New York: Routledge.
- Mayer, R. (2009). *Multimedia learning* (2<sup>nd</sup> edition). Cambridge University Press.
- Nichols, W. (2023, June 27-29). *Invisible lexicographers, AI, and the future of the dictionary*. [Paper presentation] eLex 2023 Conference: Electronic Lexicography in the 21st Century, Brno, The Czech Republic. [https://www.youtube.com/watch?v=xYpwftj\\_QQI](https://www.youtube.com/watch?v=xYpwftj_QQI)
- Paas, F., & Sweller, J. (2012). An evolutionary upgrade of cognitive load theory: Using the human motor system and collaboration to support the learning of complex cognitive tasks. *Educational Psychology Review*, 24(1), 27-45. <https://doi.org/10.1007/s10648-011-9179-2>
- Ptasznik, B. (2022). After the digital revolution: Dictionary preferences of English majors at a European university. *Lexikos*, 32, 220-249. <https://lexikos.journals.ac.za/pub/article/view/1721>
- Ptasznik, B., Wolfer, S., & Lew, R. (2024). A learners' dictionary versus ChatGPT in receptive and productive lexical tasks. *International Journal of Lexicography*, 37(3), 322-336. <https://doi.org/10.1093/ijl/ecae011>
- Rees, G., & Lew, R. (2024). The effectiveness of OpenAI GPT-generated definitions versus definitions from an English learners' dictionary in a lexically orientated reading task. *International Journal of Lexicography*, 37(1), 50-74. <https://doi.org/10.1093/ijl/ecad030>
- Rundell, M. (2023). Automating the creation of dictionaries: Are we nearly there? In Y. An (Ed.), *Proceedings of the 16th International Conference of the Asian Association for Lexicography* (pp. 1-9). Yonsei University.
- Statcounter. (2024). Search engine market share worldwide. Retrieved August 5, 2024, from <https://gs.statcounter.com/search-engine-market-share>
- Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 19-30). Cambridge University Press. <https://doi.org/10.1017/CBO9780511816819.003>
- Tal, S., Grossman, E., & Arnon, I. (2024). Infant-directed speech becomes less redundant as infants grow: Implications for language learning. *Cognition*, 249, 105817. <https://doi.org/10.1016/j.cognition.2024.105817>
- Tal, S., Grossman, E., Rohde, H., & Arnon, I. (2023). Speakers use more redundant references with language learners: Evidence for communicatively-efficient referential choice. *Journal of Memory and Language*, 128, 1-11. <https://doi.org/10.1016/j.jml.2022.104378>
- Tavast, A. (2023, June 27-29). Round Table on Large Language Models and AI in lexicography. [Panel session]. 8th Electronic Lexicography in the 21st Century Conference, Brno, Czech Republic. <https://www.youtube.com/watch?v=dNkksTDYas>

## APPENDIX

### A SAMPLE OF THE MAIN TEST FOR *CIRCUMSPECT* BY TEST VERSION

TEST VERSION A. *Circumspect* explained by Copilot


What does **circumspect** mean? Read the information below and supply an equivalent in your native language in the answer window.


***We would welcome a more circumspect suggestion.***


Circumspect is an adjective that means **careful not to take risks**. It can be used to describe people, actions, or situations that are careful and considerate. For example:

- Officials were circumspect about saying what the talks had achieved <sup>1</sup>.
- She was circumspect in her business dealings <sup>2</sup>.
- He is chary of signing papers without being circumspect first <sup>3</sup>.

The word circumspect comes from Latin *circumspicere*, meaning **to look around** <sup>2</sup> <sup>4</sup>. Some synonyms of circumspect are cautious, wary, careful, and prudent. Some antonyms are unguarded, incautious, reckless, and rash.

Learn more 

1  [bing.com](https://www.bing.com)

2  [merriam-webster.com](https://www.merriam-webster.com)

3  [dictionary.cambridge.org](https://dictionary.cambridge.org)

Answer:



What does **circumspect** mean? Read the information below and supply an equivalent in your native language in the answer window.

***We would welcome a more circumspect suggestion.***

"Circumspect" is an adjective used to describe a person or their actions when they are cautious, careful, and considerate in order to avoid potential risks or problems. Someone who is circumspect approaches situations with thoughtfulness, foresight, and a careful weighing of potential consequences. This term is often associated with a prudent and deliberate approach to decision-making, indicating a person who is attentive to details and exercises discretion in their actions. Being circumspect involves being aware of one's surroundings, potential challenges, and the impact of one's choices.

Answer:

TEST VERSION C. *Circumspect* explained by the Google dictionary

What does **circumspect** mean? Read the information below and supply an equivalent in your native language in the answer window.

***We would welcome a more circumspect suggestion.***

adjective

adjective: **circumspect**

wary and unwilling to take risks.

"the officials were very circumspect in their statements"

Similar:

cautious

wary

careful

chary

guarded

on one's guard


discreet

Answer:

TEST VERSION D. *Circumspect* explained by LDOCE

What does **circumspect** mean? Read the information below and supply an equivalent in your native language in the answer window.

***We would welcome a more circumspect suggestion.***

**cir-cum-spect** /'s3:kəmspekt \$ 's3:r-/ adjective formal  

thinking carefully about something before doing it, in order to avoid risk **SYN** cautious

 The governor was usually circumspect when dealing with the media.

Answer:

TEST VERSION E. *Circumspect* explained by COBUILD

What does **circumspect** mean? Read the information below and supply an equivalent in your native language in the answer window.

***We would welcome a more circumspect suggestion.***

# circumspect

Word Frequency ●●●●●

(sɜːˈkʌmspekt) 🔊 ⓘ

**ADJECTIVE**

If you are **circumspect**, you are cautious in what you do and say and do not take risks.

[formal]

*The banks should have been more circumspect in their dealings.* [ɪn] 🔊

*Are you saying, in your usual circumspect manner, that it isn't be a good idea?* 🔊

**Synonyms:** cautious, politic, guarded, careful More Synonyms of **circumspect**

**circumspectly** 🔊 **ADVERB** [ADVERB after verb]

*You should behave as circumspectly as possible in political matters.* 🔊

Answer: