Contrasting a new AI-powered dictionary designed for on-screen reading with electronic dictionaries that have evolved from print editions

Ana Frankenberg-Garcia[1]

[1] University of Surrey, UK
E-mail: a.frankenberg-garcia@surrey.ac.uk

Abstract

Despite the growing number of papers investigating how existing dictionaries compare with lexicographic material generated through LLM-powered chatbots, there do not seem to be any studies contrasting actual dictionaries conceived with the assistance of AI with well-established dictionaries compiled and edited by lexicographers. This paper compares Reverso, a new English dictionary powered by AI, with the Oxford Dictionary of English (ODE) and the Merriam-Webster dictionary (MW) from the perspective of vocabulary assistance while reading a general-interest text in an online publication. The three dictionaries were evaluated on four criteria: coverage of the target sense, findability of the correct meaning, readability of the definition, and overall look-up experience. The results show that Reverso does well in all four parameters and outperforms ODE and MW in readability and look-up experience. Implications for better integration of vocabulary assistance for reading from computer screens and how LLMs can contribute to that end are discussed.

Keywords: LLMs; Vocabulary Assistance; On-screen Reading; Embedded Dictionaries

1. Introduction

Large language models (LLMs) are fast disrupting the field of lexicography. Although LLMs based on transformer architecture have been around for some time, it was only when OpenAI publicly launched ChatGPT optimized for conversational use and released it as a free web-based chatbot on 30 November 2022 that its affordances began to be investigated more widely in lexicography studies.

In the past couple of years, several papers have emerged comparing dictionary entries crafted by professional lexicographers with entries resulting from chatbot prompts designed by experts (e.g., Lew, 2023; Jakubíček & Rundell, 2023). Although LLMs tend to perform better in some lexicographic tasks than in others, as recognized in the very conference topic of eLex 2025[1], it is becoming increasingly evident that LLMs present unprecedented opportunities to enhance traditional workflows in lexicography.

[1] https://elex.link/elex2025/event-info/

Indeed, Reverso Define, a new English dictionary created using LLMs, was launched in record time in September 2024. What is less known is how a dictionary that draws on LLMs for its production compares with dictionaries created by lexicographers from scratch. In this study, Reverso was compared with the Oxford Dictionary of English (ODE) and the Merriam-Webster Dictionary (MW) from the perspective of vocabulary assistance while reading from a computer screen.

2. Background

In the past couple of years, several papers have emerged comparing dictionaries with dictionary-like information obtained through dialogue with chatbots (De Schryver, 2023). Lew (2023) used ChatGPT 3.5 in prompts that required advanced knowledge of lexicography and contained examples of the target dictionary entries sought. The model generated monolingual English dictionary entries for fifteen verbs of communication, emulating the style of the corpus-driven COBUILD dictionary. Four experts were then asked to blindly evaluate the results in comparison with the published dictionary. Their ratings indicated that the definitions produced by ChatGPT looked promising, but other aspects of the dictionary entries were problematic.

In another study involving prompts crafted by experts, McKean and Fitzgerald (2024) used ChatGPT 3.5 to undertake typical lexicography routines. This included tasks such as prompting the tool to (a) provide a list of headwords that appeared alphabetically between *growl* and *guardian;* (b) return phrases featuring selected words (like *cut corners* for *cut*); (c) generate a list of derivational and grammatical inflections arising from a base form; (d) write definitions; and (e) list examples of usage. The researchers noted various problems in the output and concluded that ChatGPT did not meet the required level of quality of human editorial work.

Nichols (2023) reflected upon the impact of LLMs on lexicography by conducting a SWOT[2] analysis. She concluded that among the strengths of ChatGPT are the facts that it is free, easy to use, provides the meaning of a word according to its contextual use in a sentence, and synthesizes vast amounts of data on the fly. On the other hand, its weaknesses include the problems that chatbots hallucinate, produce responses that tend to be far too long for simple meaning look-ups, do not handle synonym differentiation very well, and are only as good as the training data underlying them.

Despite its current shortcomings, dictionary users may nevertheless benefit from lexicographic output generated by LLMs. For example, Ptaznik and Lew (2025) surveyed 225 students in Poland and found that although monolingual and bilingual dictionaries remain popular and are generally considered more reliable, the direct use of ChatGPT for language assistance has become commonplace and is praised for its speed. In a study on L2 English learners' reactions to AI-generated definitions, Rees and Lew

[2] Strengths, Weaknesses, Opportunities and Threats.

(2024) designed a multiple-choice reading task to test lexical knowledge in order to compare the effectiveness of definitions generated by ChatGPT 3 with definitions copied from the Macmillan English Dictionary (MED). The experiment also included testing comprehension without definitions as a control. The chatbot prompt used was a straightforward *explain word X.* The results were not conclusive because there was no significant difference between the effect of ChatGPT and MED on comprehension, and neither was there a significant difference between the time the participants took to process ChatGPT and MED definitions. However, in another word-comprehension experiment comparing the direct use of the free version of ChatGPT 3.5 and the Longman Dictionary of Learner's English (LDOCE) via the participants' phones, Ptasznik et al. (2024) found that ChatGPT significantly outperformed the dictionary, but was not faster.

In a descriptive study that bears some resemblance to the present paper, Jakubíček and Rundell (2023) created a test dictionary[3] consisting of 99 full entries generated by ChatGPT 3.5 and compared them with equivalent entries in two first-class dictionaries edited by lexicographers: ODE and MED. To create the test dictionary, the researchers engaged in a guided dialogue with the chatbot, refining the prompt to "Generate a dictionary entry for H [headword] including possible word forms, word senses, pronunciation, collocations, synonyms, antonyms and examples of usage" (Jakubíček & Rundell, 2023: 524). The sample compared included words of varying complexity and several parts of speech, as well as some multi-word expressions. The results of the analysis were not dissimilar to those in Lew (2023). The researchers found that despite occasional glitches - like old-fashioned definitions starting with *the act of* or *the state of* and the extended length of some definitions - ChatGPT presented promising results, not only because a number of the problems identified can be tackled in a refined prompt, but also because writing definitions is hard and requires special skills (Rundell, 2007). ChatGPT performed less well in other parameters of the study, particularly with regard to word-sense disambiguation (where the researchers noted a propensity to superfluous sense distinctions), incomplete grammatical information (such as missing verb transitivity and intransitivity distinctions), and example sentences, which, as also noted in Lew (2023), were often unimaginative and inauthentic.

What the above studies have in common is that they focus on the *potential* advantages and disadvantages of the use of LLMs in lexicography. The word *potential* is emphasized because there does not seem to be any investigation that compares an existing dictionary that was conceived with the support of LLMs with well-established dictionaries compiled and edited by lexicographers. The present study addresses this gap by comparing Reverso with ODE and MW.

While previous studies examined the quality of lexicographic content generated by LLMs or its effects in controlled experiments with learners, the present study proposes to undertake the comparison from another angle. Namely, from the perspective of the

[3] https://www.lexonomy.eu/chatgpt35

capacity of dictionaries to assist in vocabulary comprehension when reading texts on a computer screen. More specifically, the study looks at a general-interest online article as a source of language-reception queries, where readers are naturally exposed to unknown words and phrases. When an unfamiliar lexical item is met, Scholfield (1999: 13) explains that the "information targeted is the meaning of that item – specifically the *meaning* relevant to the context where it has been met" [4]. It is a timely moment to revisit Schofield on dictionary use in reception because, with a growing number of people using their screens rather than paper for reading, the nature of dictionary consultation for looking up meanings is changing. This has implications in terms of dictionary content, dictionary format, and the interrelation of the two, especially because, as discussed in Nichols (2023) and demonstrated in Ptasznik et al. (2024), readers using an electronic device to read can just as easily check word meanings directly via chatbots.

Thus, instead of comparing dictionaries formally in terms of their adherence to accepted lexicographic norms, or experimentally from the viewpoint of how learners and experts react to them, the present comparison is functional in the sense that it examines how the three dictionaries under analysis serve language reception needs that may arise from authentic general-interest reading material published online. Details about how this was approached are presented in the next section.

3. Method

This section describes the materials and tools used in the study and explains how the analysis was undertaken.

3.1 Materials and tools

The reading comprehension text used in the study - Rethinking screen time: A better understanding of what people do on their devices is key to digital well-being by Rinanda Shaleha - is an online article published in *The Conversation* on 19 November 2024[5]. It was chosen at random from the countless general-interest reading materials available online at the time.

To focus on vocabulary from the above text that readers of English as an additional language might genuinely want to look up, a copy was analysed automatically using The Compleat Lexical Tutor[6] Classic Vocabulary Profiler by Cobb (n.d.). The tool is a practical, web-based alternative to Range (Heatley et al., 2002), which is in turn a

[4] Although dictionaries can also help with writing (e.g. Frankenberg-Garcia et al. 2019), language production was not the focus of the present study.

[5] https://theconversation.com/rethinking-screen-time-a-better-understanding-of-what-people-do-on-their-devices-is-key-to-digital-well-being-243644

[6] https://www.lextutor.ca

computational deployment of West's (1953) General Service List and Coxhead's (2000) Academic Word List. What the Vocabulary Profiler does is it colour-codes words in a text according to West's 1K English lemmas (in blue), the 2K lemmas (in green), the lemmas in the Academic Word List (in yellow), and all other words are considered off-list (in red). The vocabulary selected for this study comprised the off-list words in red, i.e., words that are neither in the General Service List nor the Academic Word List.

Although the Vocabulary Profiler website states that the original 2002 tool has been updated, words like *smartphone*, *tablet* and *online*, which have become commonplace in the last decade, are classed as off-list. Since the idea was to select only words that could cause genuine reading comprehension difficulties for present-day learners of English, all the vocabulary items initially marked as off-list were checked against the more recent Oxford 3000. This is a list made up of the 3,000 core words of English (chosen because of their corpus frequency and relevance to learners from A1 to B2 CEFR levels).

At this juncture, it is worth noting that the classification of words according to CEFR levels may vary. For example, *online* is considered A1 in Oxford 3000, but A2 in the Cambridge online dictionary. Another issue is that there are different types of vocabulary knowledge (González-Fernández & Schmitt, 2020), but the CEFR classification does not distinguish between them. As Rees and Lew (2024: 57) explain, "Receptive tasks require a less precise vocabulary knowledge or a knowledge of fewer aspects of a word's meaning than productive ones." Thus, words that are recognized and understood by readers at a given CEFR level may not be used in language production until they are at a more advanced level. Additionally, although the difficulty of a word may depend on the sense in which it is used, vocabulary lists do not normally distinguish between the different meanings of polysemous words (Frankenberg-Garcia et al., 2021). A final note is that the vocabulary profiling tools consulted do not classify multiword units.

Aware of the above caveats, it was not this study's objective to critique word lists and vocabulary profiling tools, but simply to use them as practical, off-the-shelf instruments to identify reading comprehension vocabulary that could be problematic for learners. The vocabulary selected for the study consisted of 39 lexical items in the reading comprehension text that were off-list in the Vocabulary Profiler and Oxford 3000. These words were inspected contextually to disambiguate part of speech. For example, *prolonged* appeared in the text as an adjective (not a verb), and *distracting* was a verb (not an adjective). The form *glance* was neither an isolated verb or noun, but part of the adverbial expression *at first glance*. Lemmas featured more than once in the text were counted only the first time they appeared, assuming readers would not need to consult them again. The lexical items selected were as follows:

13 adjectives: *chronic, cognitive, depressive, distracted, flawed, glycemic, graphic, mindful, nuanced, offline, prolonged, quantitative, standardized*

17 nouns (3 plural): *alerts, buzz, cognition, connectedness, critique, cyberbullying, engagement, era, fatigue, gaming, meditation, mindfulness, multitasking, nuance, policymakers, scrolling, teens*

8 verbs (4 inflected): *blends, boost, categorize, conceptualized, distracting, enriching, generalize, procrastinate*

1 adverbial expression: *at first glance*

Note that, when seen in context, some of the above items could be interpreted as part of a compound rather than isolated words. For example, *buzz* was featured in the context of *buzz phrase*, and *graphic* in the context of *graphic design*. However, the vocabulary profiling tools consulted do not classify multiword units, and it was assumed that, except for the adverb *at first glance*, readers would be able to construe the (non-idiomatic) meanings of compounds by looking up each word separately.

3.2 Procedure

After selecting the above vocabulary items, they were looked up in Reverso, ODE and MW while reading the text from a computer screen. Reverso was accessed directly from the URL of the reading material through a Chrome extension. ODE was accessed directly from the reading material on Chrome when using an Apple computer. MW was accessed on a separate Chrome browser tab. The look-ups were analysed according to the following criteria:

1. Coverage
2. Findability
3. Readability
4. Look-up experience

For coverage, the target sense was classified as either present or absent. Regardless of whether definitions were more general or more precise, in this study, both were considered valid, since as discussed earlier, comprehension only does not require all dimensions of word knowledge. However, in the few cases where a more precise definition was presented as a sub-sense of a broader definition, it was the sub-sense that was taken into account.

In terms of findability, the position of the target sense in the entry was ranked. For headwords like *buzz*, which is classed as a noun and a verb, all definitions listed were taken into account, rather than just the definitions listed under the target grammatical

category. For example, the target noun sense of *buzz* comes 5th in Reverso, with the verbal senses of the word being presented first. For headwords with sub-senses, each sub-sense was counted as a separate position in the ranking, given that the reader would normally have to read past the main sense to reach the relevant sub-sense. Thus, if the target sense sought was presented under definition 1.1 (following definition 1), then the target sense was ranked second. In cases where to reach the target sense the user had to click on a hyperlink taking the reader to a different headword, that extra click was also computed in the ranking. A final point about findability is that Reverso exhibited a few odd cases of false polysemy, which is typical of LLM-generated dictionary entries (Jakubíček & Rundell, 2023; Michta & Frankenberg-Garcia, 2025). For example, under *cognitive*, Reverso displays the target sense twice, using different wordings that are not sufficiently distinct: (1) *relating to mental processes like thinking and understanding*; and (2) *involving conscious intellectual activity*. In such cases, it was the first definition shown that counted, assuming readers would stop browsing the entry once they encountered the sense sought.

To operationalize readability, a deliberate decision was made not to use the outdated but still popular Flesch-Kincaid readability test (Kincaid et al., 1975). This crude instrument, developed over 70 years ago, simply calculates readability in terms of sentence and word-syllable length, disregarding vocabulary, which is essential to this study. This means that sentences with an equal number of words and syllables, like *The cat is on the mat* (seen in early language learning materials) and *Prions pass from cell to cell* (from a biomedical research abstract), score the same American 5th Grade level of readability (very easy to read), without considering the how familiar the words in these sentences might be.

In this study, readability was assessed based on the following criteria:

a. Length
b. Vocabulary
c. Circularity

Although definition length (measured in number of word tokens) does not in itself mean much in terms of readability, it is nevertheless important in the context of reading assistance, since the longer it takes to scan a definition, the more time readers are kept away from their main reading material. Importantly, as seen above, the vocabulary of definitions is also key. Definitions that include words that dictionary users are not familiar with are likely to be less intelligible, and may lead them to chase the meaning of extra words, further diverting them from their reading. The study's definitions were therefore checked against the Vocabulary Profiler followed by Oxford 3000 to identify off-list items, following the same procedure used to identify vocabulary likely to be looked up in the main text. With regard to circularity, all definitions containing words of the same root as the word being defined were initially classed as circular. For example, Reverso's definition of *standardized*, "brought into conformity with a

standard", was at first deemed circular because the definition contains *standard*, which has the same root as *standardized*. However, since *standard* is not off-list in the vocabulary profiling tools consulted, this case of circularity was counted as void, as it was assumed that readers looking up *standardized* would understand *standard*. In contrast, ODE defines *meditation* as "the action or practice of meditating", where *meditating* belongs to the same root as *meditation.* The difference, however, is that *meditating* is off-list. Therefore, the definition was classed as circular, since it was assumed that the reader who does not understand *meditation* will not understand *meditating* either.

A final factor that could affect readability is syntactic complexity. However, this was not computed because definitions are often presented as sentence excerpts (Rundell, 2008a) rather than full sentences that are readily analysable in terms of syntactic complexity. Moreover, it is unclear whether definitions based on sentence excerpts are more intelligible than full-sentence definitions.

The look-up experience was finally analysed from a descriptive, qualitative perspective, focusing on how straightforward it was to access the dictionary while reading the main text on a computer.

4. Results

This section describes the results of the analysis in terms of coverage, findability, readability and look-up experience.

4.1 Coverage

Table 1 summarizes target-sense coverage in the three dictionaries. Reverso offered the most complete coverage of the meanings sought, providing 37/39 relevant definitions, closely followed by ODE, with 36/39. MW offered 33/39 definitions.

Note that the missing senses were not necessarily missing as headwords. Reverso, for example, provided other senses of *alerts*, but not the one required (*signal on an electronic device reminding the user to do something or attracting their attention*). Likewise, ODE listed other senses of *engagement*, none of which pertained to the *being engrossed* sense in the reading material. MW featured the headwords but not the required senses of *alerts*, *mindful* and *teens*.

In a few cases of incomplete coverage, the three dictionaries redirected the user to a related headword. For example, when clicking on *scrolling* in Reverso, the reader is taken to the adjectival sense of the word, but the noun sense sought is missing. When looking up *glycemic*, ODE displays the entry for *glycaemia*, where there is a mention of *glycemic* as a "derivative", but no accompanying adjectival definition. Likewise, when seeking the adjective *standardized*, the user is taken to the verb *standardize*, but with no

definition of the adjectival sense sought. In MW, the user seeking *connectedness* is taken to *connected*, *glycemic* to *glycemia*, and *scrolling* to *scroll*.

Dictionary	Coverage (n=39)	Missing senses
Reverso	37 (94.9%)	alerts_N scrolling_N
ODE	36 (92.3%)	engagement_N glycemic_Adj standardized_Adj
MW	33 (84.6%)	alerts_N connectedness_N glycemic_N) mindful_Adj scrolling_N teens_N

Table 1: Target-sense coverage

4.2 Findability

Table 2 shows how many definitions the user had to read past before reaching the target sense. It can be seen that in both Reverso and ODE, the target sense was normally displayed first (median and mode = 1st). In MW, there was greater variability. Although the target sense was shown first in most entries (mode), the range was quite broad (1st to 20th), skewing the median to 2nd. The target sense that was further away from the top of the entry in Reverso was *buzz* (5th). The expression *at first glance* was 11th under the headword *glance* in ODE, and 20th in MW.

Dictionary	Median	Mode	Range
Reverso	1st	1st	1-5
ODE	1st	1st	1-11
MW	2nd	1st	1-20

Table 2: Position of the target sense in an entry

4.3 Readability

Tables 3 and 4 summarize the results of the readability analysis. It can be seen in Table 3 that the Reverso definitions were, on average, much shorter than the ones in ODE and MW, which were in turn similar in length. Additionally, there was considerably more variation in the length of definitions of ODE and MW than in those of Reverso.

Dictionary	Mean	SD	Range
Reverso	6.59	1.36	10-4
ODE	12.00	5.59	28-3
MW	12.39	6.81	31-3

Table 3: Definition length in number of words

A one-way ANOVA was computed, indicating that the differences in length observed were significant (F-ratio=14.72693) at p <0.01. A post-hoc Tukey HSD test indicated the differences between Reverso and ODE were significant (Q=6.33), as were the differences between Reverso and MW (Q=6.79), at p <0.01. In both cases, the effect size was large (Cohen's d = 1.25 and 1.12, respectively). However, the differences between ODE and MW were not (Q=0.46).

With regard to vocabulary, Table 4 shows that the definitions in Reverso contained 13 words that were classed as off-list in the Vocabulary Profiler and Oxford 3000, while ODE contained 19 such words and MW 24. Of these, two definitions in ODE and one in MW were considered circular, as the target sense of the word being defined used a word of the same root that was also off-list in the definition.

Dictionary	Number of off-list words in definitions	Off-list vocabulary in definitions	Circularity in definitions
Reverso	13	attentive_Adj, defect_N, diverted_V, diverting_V, execution_V, exertion_V, non-judgmental_Adj,recurring_V, rumor_N, simultaneously_Adv, subtle_Adj, surroundings_N, tiredness_N	N/A
ODE	19	affinity_N, bully_N, clinical_Adj, diverting_V, engraving_V, exertion_N, graphics_N, imperfection_N, intimidating_Adj, lettering_N, meditating_V, meditative_Adj, preoccupied_Adj, recurring_V, sensations_N, subtle_Adj, tending_V, therapeutic_Adj, tiredness_N	meditation: meditating cyberbully: bully
MW	24	anonymously_Adv, appealingly_Adv, defect_Adj, demarcation_N, engaging_N, exertion_N, exhaustion_N, expressible_N, generalized_V, generic_Adj, mantra_N, heightened_Adj, imperfection_N, intellectual_Adj, judicious_Adj, non-judgmental_N, nuances_N, subtle_Adj, prominent_Adj, repetition_N, speculative_Adj, spirited_Adj, telecommunications_N, weariness_N	nuanced: nuance

Table 4: Off-list vocabulary in definitions

4.4 Look-up experience

When reading the general-interest article selected for this experiment on a computer screen, Reverso was directly accessible from the reading screen, ODE could be accessed through a right-click of the mouse when using an Apple device, and MW was accessed by alternating between the reading material open in one browser tab and the dictionary in another. A descriptive account of the look-up experience when using each separate dictionary is provided below. At the end of this section, Table 5 summarizes the main points of the overall experiences.

4.4.1 Reverso

Accessing Reverso through a single click on any word in the text without having to select the word is very efficient. A small window with one or more definitions appears on the reader's screen without covering the part of the text with the word sought. This helps users to read on without losing their place (Figure 1). The positioning of the definitions window is dynamic: if the word is at the top of the screen, the window opens immediately below it. Conversely, if the word is further down the screen, the window opens immediately above it.

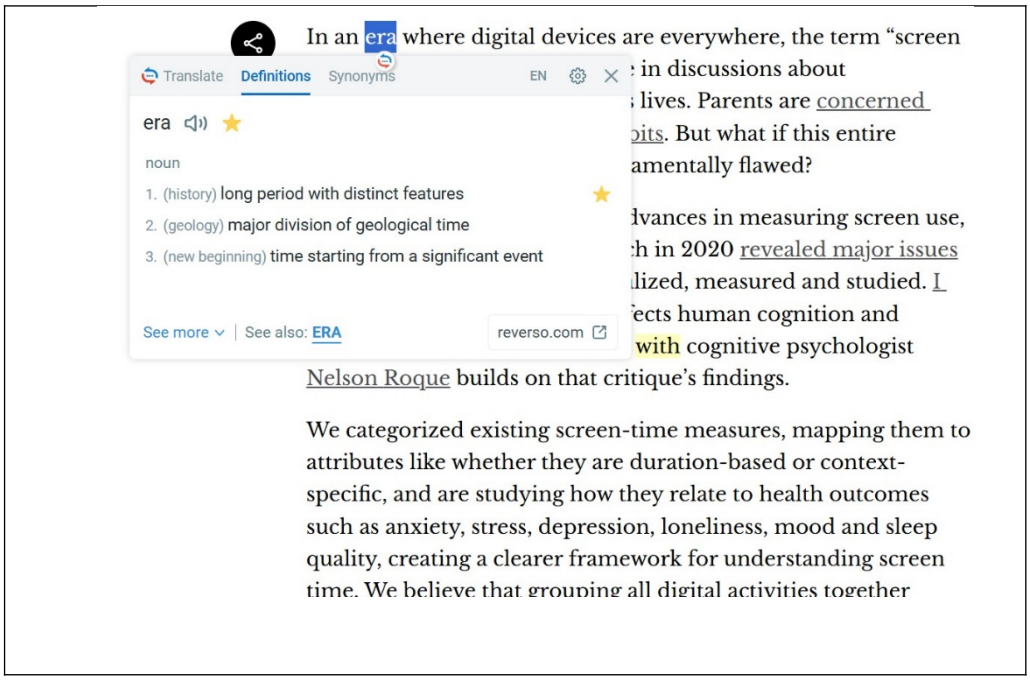

Figure 1: *Era* in Reverso's Chrome extension for on-screen reading

Reverso's Chrome extension window condenses the contents of the dictionary's full entries, retaining only sense indicators and definitions, but linking to further information should the reader wish to dig deeper. In particular:

a. All words in the definitions are clickable, taking the user to the definition of the word clicked.

b. A megaphone icon links the word to a sound file on how to pronounce it
c. A star icon enables users to add the word (and the excerpt in which it occurs in the reading material) to "Favorites" (readers can later go to Favorites to review the words looked up and their contexts).
d. A link to "See more" expands the contents of the window so as to include one example sentence under each sense.
e. A link to "Translate" changes the definitions window to a machine-translation window.
f. A link to "Synonyms" changes the definitions window to a computer-generated thesaurus.
g. A link to "reverso.com" takes the reader to a separate browser tab containing the full dictionary entry.

The definitions window can be easily closed by clicking on the customary cross in the top-right, or anywhere on the main reading material. If the place clicked is another word, the window with the definition of the previous word will close and one with the new word will open.

The only difficulty in accessing Reverso emerged when the word sought was within a hyperlink in the text (e.g., *connectedness*) or a multiword unit (e.g., *at first glance*). Clicking on a word with a hyperlink takes the reader to an external URL. However, if that same word is selected with the user's cursor, then the Reverso icon appears underneath it, enabling the definitions window to open when clicked. Multiword units also need to be selected first before they can be looked up.

4.4.2 ODE

Accessing ODE from the reading screen is less straightforward. To begin with, the user must be on an Apple device. At the time of writing, the dictionary is not available for reading on-screen in other environments. Second, to arrive at ODE, the reader must first right-click on the word sought. This will open a menu, from which the user must select "Look Up", and only then will a window with a selection of dictionaries open. ODE can be seen at the top of the list of dictionaries that can be selected (if available in the computer's dictionary settings). This same dictionary selection window offers a truncated preview of the relevant entry. If the meaning sought is shown in the preview, then no further clicking is necessary (Figure 2). This will happen if the definition needed appears at the top of the entry and is not very lengthy. If the target definition is not visible from the preview because it appears further down in the entry or because it is a longer definition and cannot be seen in full in the preview, then the user must click to open the full ODE window (Figure 3).

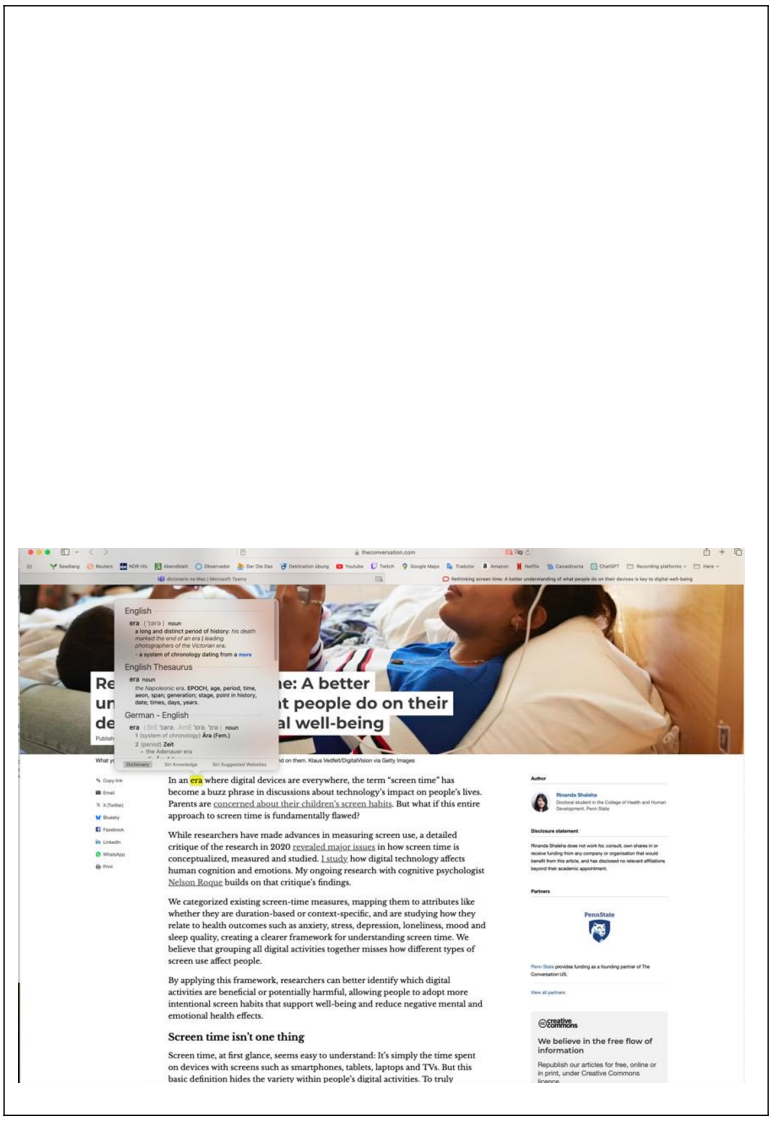

Figure 2: *Era* in an Apple computer's dictionary selection window (ODE preview at the top)

The positioning of the ODE window is dynamic and does not conceal the word sought. However, words with hyperlinks and multiword units need to be selected before clicking to look up their definitions. As seen above, this also applies to Reverso.

Unlike Reverso, which selects which contents readers see in the definitions window (e.g., example sentences are not shown by default), the ODE window in Apple devices is very similar to full entries in the ODE browser version. However, because the window does not occupy the full screen, reading its contents normally involves more scrolling. Another difference is that the contents of the ODE window are static: the words in the definitions are not clickable, there is no link to pronunciation sound files, and so on. It is as if the entry is taken from a print dictionary, with no further interactivity.

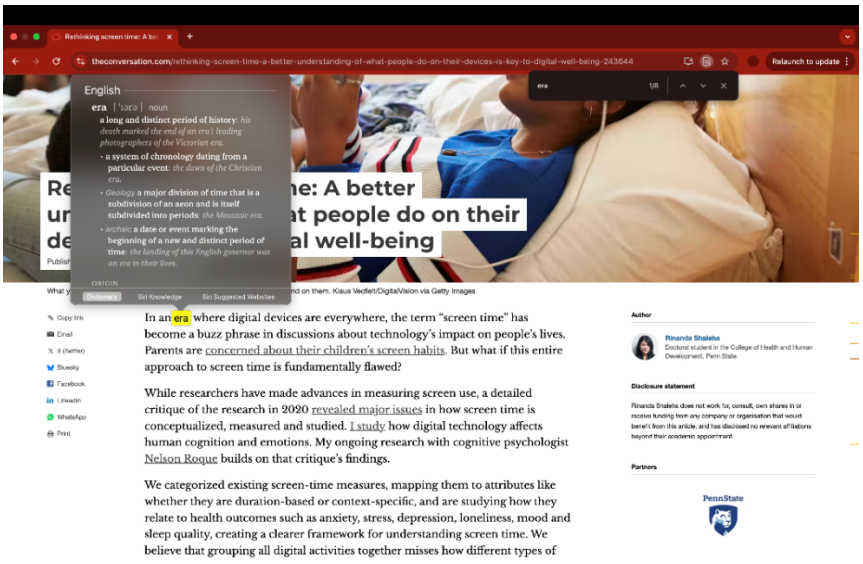

Figure 3. *Era* in ODE window on an Apple computer screen

Closing the window is easy. The reader simply clicks on the conventional cross icon in the top right, or clicks anywhere else outside the window.

4.4.3 MW

Unsurprisingly, the MW look-up experience is the least smooth of the three dictionaries compared, as the user has to switch browser tabs to access MW. By the time the look-up is completed, when switching back to the reading material, it is harder to remember not only the place in the text that prompted the look-up, but also the gist of the sentence or clause in which it occurred. One way to overcome losing track of the place in the text of the word sought is to select it before switching tabs to the dictionary. Upon returning to the reading material, the word remains selected, making it easier to resume one's place in the text. As in ODE, readers are also taken to the full MW entry, which usually contains a surplus of information that can be distracting for readers engaging

with another text. A final point is that MW does not capitalize fully on the advantages of the electronic medium, as only pre-selected words in the definitions are clickable, which limits the way users can navigate to other entries should the need arise.

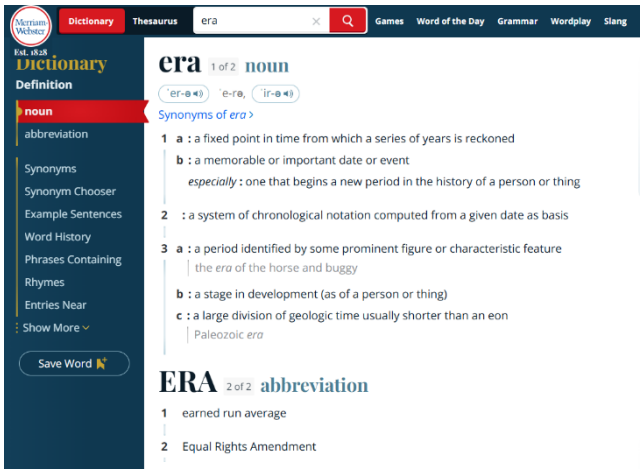

Figure 4: *Era* in MW via a separate browser tab

4.4.4 Look-up experience summary

Look-up feature	Reverso	ODE	MW
Integration with reading material	Via browser extension	Via Apple devices	Via a separate browser tab (no integration)
Access to definitions	1. Click on any word to open a small window with definitions (words within hyperlinks and multiword units need to be selected first).	1. Right-click on a word to open a menu with a "Look up" option (words within hyperlinks and multiword units need to be selected first). 2. Click on "Look up" to open a window with a truncated definitions preview (the target definition can often be found there). 3. Another click is needed to open the full entry.	1. Open a separate tab to navigate between the reading material and the dictionary. 2. Type or paste a copy of the word sought into the search box 3. Click on the search icon.
Contents of the embedded dictionary window	Condensed version of the full dictionary entry, showing only sense indicators and definitions. Links to further contents.	Truncated preview of definition at first, then full entry when clicked.	N/A
Format of the	The window is fully interactive. All extras, such	Static, print-like rendering of the browser version of the	N/A

embedded dictionary window	as examples, pronunciation, saving words looked up, synonyms, translation, and words in the definitions are clickable.	entry (no further interactivity possible).	
Return to reading	The word looked up remains visible on screen. The dictionary window can be closed with one click.	The word looked up remains visible on screen. The dictionary window can be closed with one click.	Readers must switch tabs. They may lose their place in the text.

Table 5: Look-up experiences compared

5. Discussion

This study emulates an authentic reading scenario, where readers look up unfamiliar words in a random general-interest text published online. As Gouws and Tarp (2024:194) argued, "most readers will prefer a quick answer so that they can continue to enjoy or concentrate on the text they are reading. In this respect, the integrated dictionary is undoubtedly a huge step forward, as the reader only has to click on the word in question to initiate a lexicographical consultation, instead of having to look it up elsewhere". The authors explain that embedded dictionaries (like Reverso and ODE in this study) are more like glossaries than dictionaries. Indeed, users can access definitions as they read, and are not required to convert the word form clicked into a lemma to look it up. The latter has implications for coverage, however, as clicking on a word may not necessarily take the user to the required definition. In this study, for example, clicking on *scrolling* with the Reverso extension took the user to the adjectival sense of the word, but not to the required noun sense. Similar issues occurred in the other two dictionaries. A challenge for future embedded lexicography is to disambiguate homographs in context so that readers can be more efficiently taken to the relevant part of speech of a given word form. English gerund and noun homographs ending in *ing*, verb and adjective homographs ending in *ed*, and third-person verb or plural noun homographs ending in *s* are examples of forms that could be problematic in this respect.

Based on the 39 lexical items analysed in this study, the three dictionaries offered good coverage of the vocabulary that English readers below the C1 CEFR level may wish to look up. Reverso achieved nearly 95% coverage, closely followed by ODE. MW was slightly less effective, with just under 85% coverage. Note, however, that these results are based on the analysis of a single general-interest text chosen at random. More authentic texts need to be submitted to similar analyses to establish whether there is a trend when comparing dictionaries from the perspective of how well they cover vocabulary likely to be unfamiliar to readers below a certain level of English. On the other hand, this study's method to analyse coverage empowers dictionary users to test coverage regarding texts that are relevant to them personally. This is arguably more meaningful to users than the number of headwords and senses advertised by dictionary

publishers[7].

All three dictionaries performed well in target-sense findability, displaying the meaning sought at the top of the entry in most look-ups. This is important, as it will quickly enable readers to go back to reading without having to go through other senses of the word consulted. The order in which a sense is presented is of course only relevant in cases of homographs and polysemy. However, the majority of words in natural languages are polysemous (Ravin and Leacock, 2023), as indeed they were in this study. This has consequences, because "locating the right sense can be a major source of difficulty, and we know that many users 'solve' this problem simply by selecting the first sense they come to" (Rundell, 2008b: 232). Indeed, Lew, Grzelak and Leszkowicz (2013) found in an eye-tracking study that users normally only glance at the first one or two senses shown in an entry. So anything beyond that can be particularly disruptive when dictionary users' focus of attention is not the dictionary itself, but rather another text they are reading.

The target sense placed furthest away from the top of the entry in Reverso was the noun *buzz* (5^{th}). However, if readers selected the compound *buzz phrase* in the reading material instead of just clicking on *buzz*, they would have been taken directly to the definition required. In ODE and MW, it was the expression *at first glance* that was the furthest down from the top of the entry. Users were taken to *glance*, and the expression could only be located in positions 11 (ODE) and 20 (MW). Interestingly, the browser version of ODE takes the user directly to the definition of *at first glance* if it is typed into the search box. This is an important difference between the ODE browser version and its Apple version, which does not seem to denest multiword expressions from headwords. Making multiword expressions easier for users to access directly from whatever they are reading on screen is thus another challenge for embedded lexicography. Using LLMs to predict multiword expressions from single words clicked might be a solution. In Reverso, this already seems to be happening, for at the bottom of the window with the definition of *buzz*, the definition of *buzz phrase* is also displayed.

While all three dictionaries performed well in terms of findability, this study only compared the position of the target sense once the relevant entry was found.However, it is important to note that, in a natural reading scenario. findability is also impacted by the number of clicks necessary to arrive at the entry sought.

Although the study did not examine the syntax and exact wordings of definitions, the LLM-generated definitions in Reverso were, on average, practically half the length and indeed significantly shorter than the definitions in ODE and MW written by human lexicographers. Not only that, but the vocabulary used in the definitions was also better adapted to readers below the C1 level, as only 13 words used in Reverso definitions, compared to 19 in ODE and 24 in MW were classed as off-list. Additionally, unlike

[7] At the time of writing, Reverso claims on its website that it contains over 500K meanings, while ODE and MW do not specify the number of definitions they contain.

ODE and MW, none of the definitions in Reverso used off-list words of the same root as the word being defined, which reduces the chances of users having to carry out follow-up look-ups. Although the definitions were not checked for syntactic complexity, the syntax of the much shorter definitions in Reverso was likely comparatively less complex than in the other two dictionaries. Another indication of the good readability of the definitions in Reverso comes from a separate study by Rees & Frankenberg-Garcia (2025), where Spanish university students used Reverso and Oxford Learner's Dictionary (OALD) to look up unfamiliar words in a reading comprehension test. Both dictionaries were, on average, over 80% effective in elucidating meanings. However, Reverso was more efficient (users completed the task significantly faster).

These findings tie in with Jakubíček and Rundell's (2023: 526) evaluation of definitions created by LLMs when they conclude that "definitions are in general one of ChatGPT's stronger points". The Reverso LLM prompt for generating definitions must have certainly established an upper limit to their number of words, and may have also included instructions to restrict the use of rare words, thus capitalizing on automating conciseness and intelligibility. This is important, because writing definitions is arguably one of the hardest tasks undertaken by human lexicographers, and "of all the components in a dictionary entry, definitions have so far proved the least tractable in terms of automation" (Jakubíček & Rundell 2023:527).

It is also relevant that definitions for decoding needs can often be satisfied with less information than definitions that aim to cater for encoding needs as well (Scholfield, 1999). As Atkins and Rundell (2008:408) explain, the reader "doesn't need to find out everything there is to know [about a word], but simply to understand what a writer is saying in a particular passage". One implication for embedded dictionaries for reading comprehension assistance is that their contents needn't (and arguably shouldn't) be the same as those of general-purpose dictionaries like ODE and MW. In fact, cutting down on information that is not relevant to decoding would be an advantage, as consulting a shorter definition devoid of extra information about a word is faster and less disruptive, enabling readers to get back to the texts they are reading more efficiently. This is precisely what happens in Reverso. Not only are definitions significantly shorter, but also the contents of the definition window that opens up when clicking on an unfamiliar word have been purposively selected from the full browser version. In contrast, the ODE window available from reading screens in Apple devices is first just a preview of the full ODE entry, which when clicked leads the user to the full entry itself. It wouldn't be unreasonable to suggest that the ODE embedded on Apple reading screens could be enhanced through prompts that lead to the automatic shortening and selective pruning of its current entries.

It is not only the contents of embedded dictionaries that need to be better adapted for on-screen reading, however. As shown in this study, the overall look-up experience matters too. Reverso outperformed the two other dictionaries in terms of its integration to on-screen reading: it can be used in any system, definitions are normally accessible

via a single click, the definitions window is dynamic (does not conceal the word looked up), the window focuses on definitions is but fully interactive should users wish to explore pronunciations, examples, translations, and so on, and readers can close the window and go back to reading in just one click.

6. Conclusion

The aim of this investigation was to compare a new AI-powered dictionary designed for on-screen reading (Reverso) with two well-known electronic dictionaries that have evolved from print editions (ODE and MW) from the perspective of vocabulary assistance for language reception. The analysis undertaken examines how the three dictionaries assist with vocabulary comprehension questions arising from 39 lexical items selected from an authentic general-interest text published online while accessing the material via the reader's computer screen.

The most obvious finding to emerge from this study is that of the four parameters compared – (1) coverage (was the target sense provided?), (2) findability (was the target sense easy to spot?), (3) readability (how long were the definitions and what vocabulary did they use?) and (4) look-up experience (how straightforward was it to access the dictionary while reading?) – the biggest differences observed were with regard to the last two. The AI-generated definitions in Reverso were significantly shorter than those of the other two dictionaries and they contained fewer off-list words and no circular definitions. The visualization of definitions was in turn far better integrated with on-screen reading experience in terms of number of clicks and interactivity.

Several broader recommendations can be drawn from this study. Among them is the need for better disambiguation of homographs in context when a reader clicks on an unfamiliar word in an electronic text (e.g. is *distracting* a gerund or an adjective?) to ensure the reader is taken to the relevant sense in fewer clicks. Another challenge is to improve access to multiword expressions. While single words can be easily clicked and looked up in the embedded versions of Reverso and ODE, multiword items need to be selected first. Predicting from a word like *glance* that the user wants to look up *at first glance* rather than just *glance* on its own would make consultation more efficient. A further key takeaway from this study is that embedded dictionaries for reading comprehension assistance would benefit from short definitions, leaving out other contents of an entry, so that users can resume reading without too much disruption. Using AI to create such entries from scratch or to selectively prune the contents of well-established dictionaries so they are better adapted for on-screen reading vocabulary assistance could be an efficient way of reaching this goal.

To conclude, it is important to remember that this study did not contemplate full dictionary entries that include less promising aspects of LLM outputs. It examined word meaning for language reception only. Yet, because more and more texts are being

read on screens instead of on paper, it is time to rethink what is required of dictionary contents and format to ease the strain of consultation for vocabulary comprehension during reading. The present study has shown that Reverso, an AI-powered dictionary designed for on-screen reading, outperforms dictionaries that have evolved from print editions, particularly in terms of the readability of definitions and the overall look-up experience, highlighting some of the benefits of using LLMs to enhance embedded dictionaries for language reception.

7. Acknowledgements

The author was a consultant for Reverso at the time the study was undertaken.

8. References

Atkins, S.B.T. & Rundell, M. (2008). *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.

Beatriz González-Fernández & Schmitt, N. (2020). Word knowledge: Exploring the relationships and order of acquisition of vocabulary knowledge components. *Applied Linguistics*, 41(4), 481–505.

Cobb, T. (n.d.). *Web Vocabprofile*. An adaptation of Heatley, Nation & Coxhead's (2002) Range. Accessed at http://www.lextutor.ca/vp/ (15 July 2025)

Collins COBUILD Advanced Online. Accessed at: https://www.collinsdictionary.com/dictionary/English (15 July 2025)

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213–238.

De Schryver, G.-M. (2023). Generative AI and lexicography: The current state of the art using ChatGPT. *International Journal of Lexicography*, 36(4), 355-387.

Frankenberg-Garcia, A., Lew, R., Roberts, J., Rees, G. & Sharma, N. (2019). Developing a writing assistant to help EAP writers with collocations in real time, *ReCALL*, 32(2) 23-39.

Frankenberg-Garcia, A., Rees, G. and Lew, R. (2021). Slipping through the cracks in e-Lexicography. *International Journal of Lexicography*, 34(2), 206-234.

Gouws, R.H. & Tarp, S. (2024). Despite current challenges: Lexicography has a bright future. *Lexicographica*, 40(1), 187–212.

Heatley, A., Nation, I.S.P., & Coxhead, A. (2002). *RANGE and FREQUENCY programs*. Accessed at: http://www.victoria.ac.nz/lals/staff/paul-nation.aspx (15 July 2025)

Jakubíček, M. & Rundell, M. (2023). The end of lexicography? Can ChatGPT outperform current tools for post-editing lexicography? In *Proceedings of eLex 2023*, 518–532.

Kincaid, J.P., Fishburne, R.P., Rogers, R.L., & Chissom, B.S. (1975). *Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy enlisted personnel* (Research Branch Report 8-75). Naval Air Station Memphis: Naval Technical Training Command.

Lew, R. (2023). ChatGPT as a COBUILD lexicographer. *Humanities & Social Sciences Communication*, 10, 704.

Lew, R., Grzelak, M., & Leszowicz, M. (2013). How dictionary users choose senses in bilingual dictionary entries: An eye-tracking study. *Lexikos*, 23(1), 228–254.

Longman Dictionary of Contemporary English. Accessed at: https://www.ldoceonline.com/ (15 July 2025)

McKean, E. & Fitzgerald, W. (2024). The ROI of AI in lexicography. *Lexi*, 11(1), 7–27.

Merriam-Webster Dictionary Online. Accessed at: https://www.merriam-webster.com/ (15 July 2025)

Michta, T. & Frankenberg-Garcia, A. (2025). Learners‘ reactions to false polysemy. Paper presented at *eLex 2025.*

Nichols, W. (2023). Invisible lexicographers, AI, & the future of the dictionary. Keynote lecture at *eLex 2023*. Accessed at: https://youtu.be/xYpwftj_QQI (15 July 2025)

Oxford 3000. https://www.oxfordlearnersdictionaries.com/wordlists/oxford3000-5000 (15 July 2025)

Oxford Advanced Learner's Dictionary. Accessed at: https://www.oxfordlearnersdictionaries.com/ (15 July 2025)

Oxford Dictionary of English. Accessed at: https://apps.apple.com/us/app/oxford-dictionary/id978674211 and https://premium.oxforddictionaries.com/english/ (15 July 2025)

Ptasznik, B. & Lew, R. (2025). Dictionaries versus AI tools through the eyes of English majors. *International Journal of Lexicography*, 38(2), 140–158.

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.

Ravin, Y. & Leacock, C. (Eds.). (2000/2023). *Polysemy: Theoretical and Computational Approaches*. Oxford: Oxford University Press. Online edition: Oxford Academic.

Rees, G.P. & Frankenberg-Garcia, A. (2025). Can dictionary consultation during reading be improved? In *Advancing CALL: New research agendas,* Full paper presented at EUROCALL 2025, Milan, 27-30 August 2025.

Rees, G.P. & 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.

Reverso Define Chrome Extension. Accessed at https://www.reverso.net/download-browser-extension (15 November 2024)

Reverso Define. Accessed at https://dictionary.reverso.net/english-definition/ (15 November 2024)

Rundell, M. (2007). Interview with Michael Rundell, by Melanie Butler, MED magazine, first published in *EL Gazette*, April 2007. Acessed at: https://macmillandictionaries.com/MED-Magazine/June2007/45-MED-Profile-print.htm (15 July 2025)

Rundell, M. (2008a). More than one way to skin a cat: why full-sentence definitions have not been universally adopted. In T. Fontanelle (Ed.), *Practical Lexicography* (pp.

197–209). Oxford: Oxford University Press.
Rundell, M. (2008b). Recent trends in pedagogical lexicography. In T. Fontanelle (Ed.), *Practical Lexicography* (pp. 221–243). Oxford: Oxford University Press.
Scholfield, P. (1999). Dictionary use in reception. *International Journal of Lexicography*, 12(1), 13–34.
West, M. (1953). *A General Service List of English Words*. London: Longman, Green & Co.