

The challenge of AI-generated neology

Cécile Poix¹, Natalya Shevchenko²

¹ Université Lumière Lyon 2, CeRLA

² Université Lumière Lyon 2, CeRLA

E-mail: C.Poix@univ-lyon2.fr, natalya.shevchenko@univ-lyon2.fr

Abstract

A neologist spends a great length of time analysing the variety of coinages, not only to observe which of the word-formation processes are at work and their productivity, but also to explore the usage of natural language. It is indeed frequency that determines the difference between hapaxes and occasionalisms (Poix 2021), the scale of neologicity of coinages before a possible social acceptance into a speaker’s lexicon (Sablayrolles 2019), and the dynamic relations of entrenchment (a cognitive process) and conventionalization (a social process) under the influence of usage events in speech communities (Schmid 2020).

In a recent corpus-based neological study reviewing the evolution of the French suffix *-ance*, we encountered AI-generated quotations in the entry of an occasionalism in an online dictionary (La langue française). This paper addresses issues in current and future lexicological studies: the inclusion of neologisms in dictionaries, the authenticity of corpora as evidence and the relevance of quotations in dictionaries as well as the evolution of lexicographic practices.

Keywords: Gen-AI quotations; occasionalism; lexicographic practices

1. introduction

In the European Commission’s *Living guidelines on the Responsible Use of Generative AI in Research* (2025), we read:

Artificial intelligence (AI) has become an integral part of our daily lives, transforming how we live and work. Over the past few years, there has been a rapid and disruptive acceleration of AI capabilities, driven by significant advances in widespread data availability, computing power and machine learning techniques. Remarkable strides were made in particular in the development of foundation models – AI models trained on extensive amounts of unlabelled data. These advancements have given rise to what is known as ‘General purpose AI’, capable of performing a wide array of tasks. This includes “generative AI” which can generate various forms of new content (text, code, data, images, music, voice, videos, etc.), usually based on instructions (also known as prompts) provided by the user. The quality of the output produced by these models is such that it may be difficult to distinguish it from human-generated content.

Thus, texts we read are increasingly partly of fully machine-generated. Indetectable as they may be, machine-texts are made of words and words reflect society. The evolution of society and language generally goes hand-in-hand as coinages define society’s

innovations. To the extent of our knowledge, there has not yet been AI-generated neologisms disseminated in general language, though no-one can assert it is impossible. The study of neology is however challenged by machine-text in the analysis of coinages.

Many parameters have to be considered in the study of neology. Coinages can be analysed as hapaxes, occasionalisms, nonce-words or neologisms. The distinction is complex (Poix, 2021). Frequency is one factor. The hapax legomenon (in Ancient Greek (something) only said once) is in that respect different from an occasionalism¹, though they both result from the nonce-formation process. Occasionalisms are coined for a specific purpose (Poix, 2018), often with stylistic functions, at the junction between creativity and poetic licence. They are comparable to nonce-formations. According to Crystal (2002: 132), “a nonce word (from the sixteenth century phrase for the nonce, meaning ‘for the once’) is a lexeme created for temporary use, to solve an immediate problem of communication”. Social validation and diffusion are also essential in the study of neology as Bauer states (2004: 78) that “for some authorities a nonce word is by definition ephemeral, and then contrasts with a neologism”.

Metadiscursive comments that accompany neologisms are other clues to measure the scale of neologicity of coinages before a possible social acceptance into a speaker’s lexicon (Sablayrolles, 2019). Observing usage also reveals when a lexeme is no longer considered a neologism and is as such an integral part of language. Also, according to Schmid (2020: 2), “repeated usage activities in usage events” describe the linguistic system through the interaction of entrenchment (a cognitive process – linguistic knowledge in the minds of speakers) and conventionalization (a social process – regularities of communicative behaviour among a speech community).

While neologists will likely have a ponder about hapaxes and nonce-formation processes, lexicographers are concerned with the establishment of neologisms.

The current article does not analyse the ability or probability of finding lexical innovations in AI-generated text, but raises concerns in the evolution of neology and lexicographic studies: the authenticity of corpora with undetectable AI-generated lexical data, the relevance of using AI-tools for lexical studies and the evolution of lexicographic practice. How do we assess if internet-based evidence is naturally occurring? If online dictionaries do not all comply to the same ethical practice, can AI-generated entries describe natural occurrences of human language?

2. Case Study

In a recent lexicological study (Poix & Rebours, 2025) reviewing the evolution of the French suffix *-ance*, 450 nouns were extracted from five corpora (frTenTen2023,

¹ Chanpira coined the term in 1966 to refer to a lexical innovation created for a poetic function at a specific place in a literary text.

Subtitles 2018, *Wiktionnaire*, *Listedemots.net*, *Le Grand Robert*). Occurrences not present in all corpora were further scrutinized as candidates of emerging change and around fifteen coinages were identified (*dirigeance* [manageance], *gouvernance* [governance], *malaisance* [embarrassance]², etc.). A KWIC-concordance enabled further analysis and classification as occasionalisms (*plussoyance* [agreeance]), paleologisms (aidance), loanwords (guidance), etc., and more generally to place the candidates in the field of terminology (compliance) or general language (*gênance* [cringeance]). The study aimed to examine neologisms as markers of usage evolution in speech communities. General language occurrences were thus kept and low frequency lexical items (hapaxes and occasionalisms) were discarded. With several occurrences found online, the coinage *apeurance* seemed to be a suitable candidate.

Deemed unproductive in the fifties (François, 1950; Goosse, 1952) *-ance* nominalisations in competition with other nominalisations (suffixed in *-ment* and *-ion*) have gained productivity in French this century, causing a rise of interest by linguists (Dal & Namer, 2010; Knittel, 2016, Knittel & Marín, 2022, Gréa et al., 2023, Steuckardt, 2024). Often coined against the blocking principle³, criticised by the Académie Française⁴, though sometimes dictionarised⁵, *-ance* nominalisations are not only productive among the youth, they are also prolific in terminology (e.g., cosmetics) and are gaining momentum in general language.

According to Gréa et al. (2023: 1) “in the standard lexicon, Object Experiencer Psychological Verbs (OEPV) are not used as basis for *-ance* nominalizations. [...] The only example found in this base, *attirance* ‘attraction’ from *attirer* ‘to attract’, has been coined by Baudelaire (Rey et al., 1998)”. Their analysis reveals that neological *-ance* nominalisations now inherit both the Stimulus (subject) and the Experiencer (object) arguments. For instance, *l’apaisance du reggae* can be understood as ‘≈ Reaggae is appeasing’ (Stimulus) as well as ‘≈ Someone is appeased by reggae’ (Experiencer).

Following these findings, and since the standard nominalisation of the psychological verb *apeurer* is *apeurement*, *apeurance* potentially reflects the evolution of *-ance* usage (Stimulus and Experiencer arguments). The base psychological verb *apeurer* is transitive, used mostly in its past participle form in literary contexts. It is derived from the noun *peur* (fear) as a parasynthetic compound (a- + verbal conversion suffix) where the prefix means ‘make’ or ‘make more’. The English equivalent ‘afear’, though less

² Possible English translations.

³ According to Bauer (2004: 22) “Blocking is the principle invoked to account for the non-existence of one word because of the prior existence of a synonymous (or, occasionally, homophonous) word. The terms ‘pre-emption by synonymy’ and ‘pre-emption by homonymy’ are sometimes used instead of ‘blocking’, which was introduced by Aronoff (1976).” Aronoff recently invalidated his own theory (2023).

⁴ In the Académie Française prescriptive section “Dire, Ne pas dire” (To say, Not to say), in 2023, *gênance*, under the header “*extensions de sens abusives*” (reprehensible shifts in meaning) was qualified as a “*doublet inutile*” (unnecessary doublet). (Accessed at <https://www.academie-francaise.fr/dire-ne-pas-dire/g%C3%AAAnance> July 2025).

⁵ *Gênance* was added to the Petit Robert in 2023.

frequent and mainly regional (OED) shares morphological and semantic similarities. Thus, ‘afearance’ is the translation of *apearance* we will use henceforth in our description.

In order to verify the arguments hypothesis in the dataset (Poix & Rebours, 2025), contexts were verified. Corpora, which are now broadly accessible as open resources, and the wide array of online platforms dedicated to lexicographic data should facilitate the study of neologisms, from detection to usage. Sablayrolles pointed out in 2019 that internet data should be exploited with great care. Exploring the internet for the analysis of *apearance* revealed more than expected. It showed that after the second lexicographic “revolution” of collaborative online dictionaries (Murano, 2014: 148), the age of GenAI might subvert lexicographic practices.

Two occurrences were found in books of limited distribution:

- (1) *Là, je vois, j’observe, je réfléchis, alors qu’ailleurs c’est l’éternelle apearance.*
Peumery, Jules. (1930). *Voyage au Maroc. De Calais vers le soleil.* Calais : Imprimerie Moderne.
- (2) *Et j’attends le soupir, et souffre que je tombe !
Et bien que la souffrance use les musicales
Je me retiens de dire, préfère jouer du rhombe,
Il n’est pas d’apearance, préfère être amical*
Vasseur, Stéphane (2016 : 8). *Vers de Vie.* Les Éditions du Net.

These examples do not provide sufficient context (It is the eternal afearance, there is no afearance) to verify any Stimulus / Experiencer argument. To expand on AI and neology, these quotations were submitted to a neural machine translation tool (DeepL) which provided two translations of *apearance*. In (1) it is reduced to ‘fear’ and the loanword is kept in (2).

- (1) There, I see, I observe, I reflect, while elsewhere it's eternal fear. (DeepL translation).
- (2) And I wait for the sigh, and suffer that I fall! / And though suffering wears out the musical / I refrain from saying, prefer to play the rhombe⁶, / It is no apearance, prefer to be friendly. (DeepL translation).

Apearance also has an entry in *Wiktionnaire*, the French *Wiktionary* with another quotation:

- (3) La grande apearance de ce pitbull a fait reculer plus d'un intrus.
This pit bull's great stamina has scared off many an intruder. (DeepL translation).

This anonymous quotation clearly indicates *apearance* has both the Stimulus and the Experiencer arguments (the subject pitbull provokes the fear of the frightened intruder

⁶ The traditional musical instrument ‘rhombus’ is also left as a loanword.

object). Interestingly though, DeepL is now giving ‘stamina’ as an English equivalence.

For further evidence, there is also an entry in *La langue française* (Llf) which is a platform dedicated to the promotion of the French language (spelling, grammar, conjugation, literature and much more). It has a dictionary (though no lexicographers on the team⁷). *Apeurance* displayed three quotations as follows (accessed at <https://fr.wiktionary.org/wiki/apeurance> 25 July 2024):

(4) “L’obscurité n’est point effrayante par elle-même, mais en raison de l’apeurance qu’elle confère à nos craintes enfouies. – Henri de la Fontel – Citation fictive générée à l’aide d’intelligence artificielle”

[Darkness is not frightening in itself, but because of the fear it confers on our buried fears. – Henri de la Fontel – Fictional quote generated with the help of artificial intelligence] (DeepL translation).

(5) “Dans le théâtre de la vie, les masques tombent rarement sans révéler une certaine apeurance, celle de la vérité brute et non embellie. – Claire Valmont – Citation fictive générée à l’aide d’intelligence artificielle”

[In the theatre of life, masks rarely fall without revealing a certain apeurance, that of the raw, unembellished truth. – Claire Valmont – Fictional quote generated with the help of artificial intelligence] (DeepL translation).

(6) “Les contes anciens puisent leur force dans l’apeurance des ombres qu’ils projettent sur les murs de notre imagination, nous rappelant que toute lumière a son envers. – Étienne Lebrun – Citation fictive générée à l’aide d’intelligence artificielle”

[Ancient tales draw their strength from the fearsome shadows they cast on the walls of our imagination, reminding us that every light has its reverse. – Étienne Lebrun – Fictional quote generated with the help of artificial intelligence] (DeepL translation).

The above quotations were not extracted from authentic corpora, their authors – Henri de la Fontel, Claire Valmont or Étienne Lebrun – do not exist but are simply made-up by AI.

Contacted in November 2024, Nicolas Le Roux, founder of *La langue française*, explained that they relied for less than a year on OpenAI technology to enhance the glosses of their entries⁸, using OpenAI models to remove abbreviations in glosses extracted from Wiktionnaire or CNRTL (TLFi) and generate fictitious quotations where real ones were not available.

⁷ According to their website, the editorial team of Llf comes from various backgrounds (political sciences graduate, journalist, writer, speech therapist, literary critique, phraseologist). Lexicography is not mentioned and no information was found relating to the person(s) responsible for the dictionary content. (<https://www.lalanguefrancaise.com/a-propos>) (30 September 2025).

⁸ Asked in November 2024 about their criteria for selecting new entries, Le Roux declared using the index of entries from Wiktionnaire accessible via a dump, then simply removing proper nouns and keeping verbs as lemma.

What started as simple contextual research, initiated a series of questions concerning the evolution of the corpus-based study of neology.

3. The dictionarisation of neologisms

Which neologisms should gain entry in dictionaries is the everlasting question. This challenging issue mostly depends on the purpose of the dictionary itself. For Rey (2013), two criteria are paramount: their dissemination in society and/or their conceptual importance in a science or domain. For Pruvost (2003: 14), dictionaries, which has been complementing lexicography for years, is gradually acquiring an autonomous status. “It’s no longer the research, it’s the adaptation of the data, with so many ‘mixtures’ with the data as its potential publics.” Potentially, La langue française presents a “commercial product” (Boulanger, 2025: 271) that satisfies the curiosity of its contemporary users and thus resorts to GenAI to complete missing data in order to offer an appealing product.

One can question the necessity of creating an entry for *apeurance*. With only two literary occurrences (1930 and 2016), it remains an occasionalism which is unlikely to be used by the language community as a neologism. Furthermore, the quotations are stylistically rich, but semantically poor, as they potentially only describe an intensification of peur (fear). The *Wiktionnaire* entry and quotation (3) date back from 2013 from an anonymous contributor. As such, what is the purpose of adding data that is non-representative of the contemporary usage of French, by any French-speaking community. Online platforms recycle online data in a potential overrepresentation of useless words. Neural probabilistic models and stochastic models are the new natural language generators. How does it fit in with lexical innovation and the creative patterns of human-made neologisms?

In order to test AI software abilities to create an entry for ‘afearance’, we explored possibilities with open-access Yeschat – AI-powered dictionary tools. In fact, many dictionary-related tools are accessible on their front page (Dictionary for dummies, Dictionary creator, Simple dictionary Gpt, Borgges dictionary, etc.).

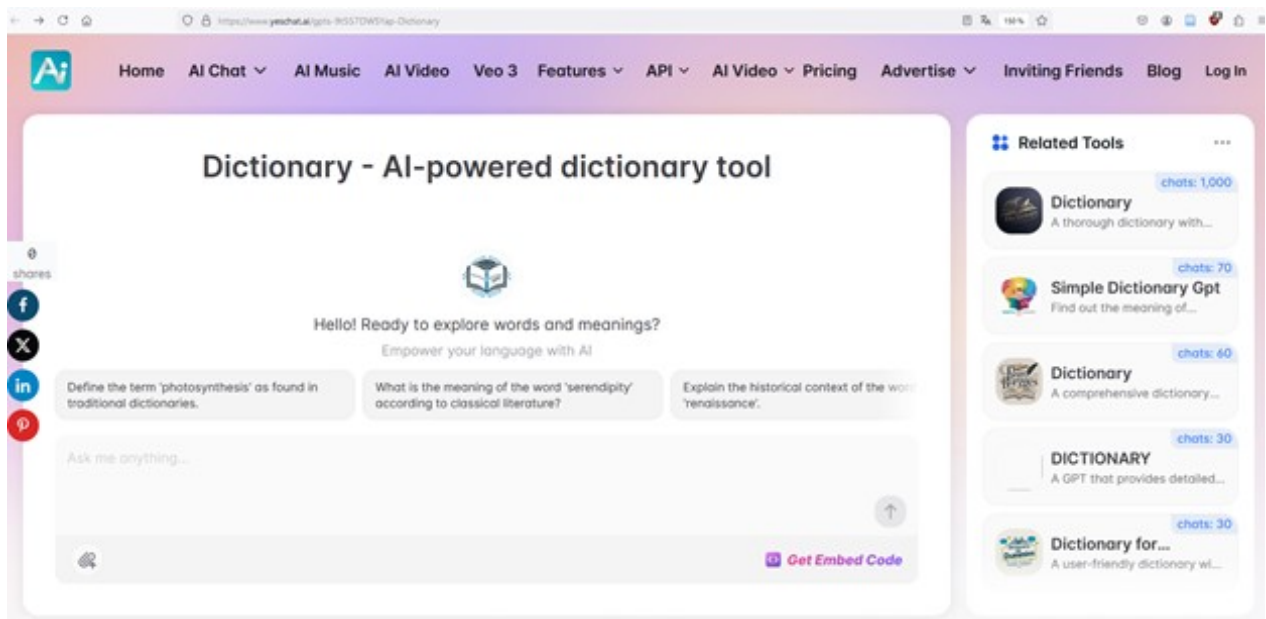


Table 1: screenshot – Yeschat – AI-powered dictionary tools

It is thus possible for anyone to play lexicographer! Here some results:

With the prompt “Provide a lexicographic entry for the word *afearance* as would appear in [https://www.oxfordlearnersdictionaries.com/]” (18 June 2025)

afearance

/ə'fɪərəns/

noun (archaic)

1. The act or process of causing fear; the state of making someone afraid.

Example: The sudden appearance of the beast brought an overwhelming *afearance* to the village.

2. The state of being afraid or frightened.

Example: His face showed the *afearance* of a man who had seen death.

Origin:

Late Middle English, from the verb *afear* (to frighten) + the suffix *-ance*, denoting a state or quality.

This version mirrors the clean and structured approach of modern dictionaries, offering pronunciation, part of speech, definitions, example usage, and etymology.

Here is the output using a prompt without the URL “Provide a lexicographic entry for the word *afearance* as would appear in the Oxford Learner’s Dictionary”.

I see you’re asking for an entry in the style of the Oxford Learner’s Dictionary, but the term *afearance* isn’t commonly used or recognized in modern English. It appears to be an archaic or rare form, derived from *afear*, which was more commonly used in Middle English. The term itself isn’t typically found in the Oxford Learner’s Dictionary since it has fallen out of use. However, here’s an

entry that reflects its historical form and context:

afearance

noun

/ə'fɪərəns/ (*archaic, rare*)

Definition:

1. The state or condition of being afraid or fearful; dread or apprehension.
2. The manifestation or appearance of fear.

Examples:

- "The afearance of the villagers grew stronger as the storm approached."
- "He spoke with an afearance that betrayed his true feelings."

Origin:

From Middle English *afearen*, meaning "to frighten," with the suffix *-ance*, denoting a state or condition.

Note:

This term is considered obsolete in contemporary usage but can be found in older literary texts, particularly from the Middle Ages and early Modern English periods.

The occasionalism 'afearance' is declared archaic in the first output and archaic or rare with the second prompt. AI hallucinations (inaccurate or non-existent outputs) are often the result of vague prompts. The last output ended with a question: "Would you like further details on its historical usage or context?", which certainly raised our curiosity. And all secrets about afearance were instantly revealed:

Here's a bit more detail on the historical usage and context of **afearance**:

Historical Context and Usage of "Afearence"

The word **afearance** is primarily found in **Middle English** texts and was used to describe the state or condition of fear, dread, or apprehension. It is derived from the verb *afear*, which meant "to frighten" or "to cause fear" in Old and Middle English.

Key Historical Insights:

1. Etymology:

- **Afearence** comes from the combination of *afear*, meaning "to frighten," and the **-ance** suffix, which forms nouns indicating a state or condition. The verb *afear* itself comes from Old English *āfæran* (meaning "to frighten" or "to make afraid"). This verb is related to the root of the modern English word **fear**.

2. Literary Usage:

- **Afearence** can be found in various **Middle English** and **Early Modern English** texts, often appearing in works related to fear, dread, or anxiety in a narrative or descriptive context. For example, in texts of the **14th to 16th centuries**, one could find the term used in describing an emotional or psychological state in characters, often indicating foreboding or a sense of impending doom.

3. Examples from Literature:

- In **Chaucer's Canterbury Tales** (late 14th century), though the exact term might not appear, the thematic focus on fear and apprehension is frequent, and we find

expressions that echo the same emotional state captured by the term *afearance*. Words like *afear* were used to describe characters who were startled or overcome with fear.

- **William Shakespeare** occasionally played with the idea of fear in his works, and while he didn't use *afearance* itself, you can see how the term may have aligned with the language of his time.

4. **Obsolescence:**

- By the 17th century, the term **afearance** had largely fallen out of regular use, and terms like *fear* or *dread* took precedence. The suffix **-ance** itself became less common in forming new words related to emotions, as modern English moved toward more concise forms and terms.

5. **Synonyms:**

- Words that would have been used in place of *afearance* in later periods include:

- o **Fear**

- o **Dread**

- o **Apprehension**

- o **Terror**

6. **Regional and Dialectal Usage:**

- Though *afearance* was largely obsolete in standard usage by the 1600s, it may have persisted in some regional dialects or been revived in specific literary contexts. Writers interested in evoking older forms of speech sometimes chose words like *afearance* to create a particular atmosphere or flavor of antiquity in their writing.

Example in Context:

- **From Middle English Texts:** "The *afearance* of the night before the battle was heavy in their hearts, none daring to speak."

This would suggest a kind of pervasive unease or fear among the characters in the face of an uncertain or dangerous event.

Conclusion:

While *afearance* is no longer in common usage, it serves as a fascinating example of how the English language has evolved, with older forms like this being replaced by simpler, more direct terms. It holds a niche place in the linguistic history of English, particularly within literary and historical studies. If you're interested in diving into its occurrences in older texts, resources like Middle English Dictionary and Shakespeare's complete works could offer more precise examples.

Let me know if you'd like more specific examples from any works or a deeper dive into this period of language history!

Needless to say, the output is *afearingly* convincing and its analysis could probably provide an article in itself. In a nutshell, the perceived quality of the output makes it look scientifically sound. Some assertions would need to be verified (e.g., "the suffix *-ance* became less common as modern English moved towards more concise forms and terms"). One could easily miss the hallucinations (e.g., section 6 on regional and dialectal usage). It also has useless data (e.g., referring to Chaucer's works where the "exact term might not appear" and Shakespeare who "didn't use *afearance* itself"). Its register is not always appropriate (e.g., "fascinating examples", "niche place").

Generally ending with the suggestion of further outputs, it entices the user to enter the AI chatting game, thus forgetting its economic and environmental impact. The AI Action Summit, held in Paris on February 10-11, 2025 emphasized the urgent need for “environmentally sustainable AI”, since, according to a McKinsey report through the rise of AI usage, “the energy demand of data centers in the EU and the UK combined could reach up to 150 TWh by 2030 – equivalent to 1/3 of France’s total electricity consumption in 2024” (De Roucy-Rochegonde & Buffard, 2025: 30). Is using AI for enhancing entries of hapaxes or occasionalisms in dictionaries worth the trouble?

4. Authenticity of corpora

From the main theoretical, methodological and application issues related to corpus work, Tognini-Bonelli’s analysis (2001: 2) explains the importance of sampling criteria, representativeness and authenticity of the texts.

A corpus can be defined as a collection of texts assumed to be representative of a given language put together so that it can be used for linguistic analysis. Usually the assumption is that the language stored in a corpus is naturally occurring, that it is gathered according to explicit design criteria, with a specific purpose in mind, and with a claim to represent larger chunks of language selected according to a specific typology. Not everybody, of course, goes along with these assumptions but in general there is a consensus that a corpus deals with natural, authentic language.

Does the concept of “natural, authentic language” expand to AI-generated text? OpenAI, a large language model (LLM), is a type of deep learning algorithm trained on massive quantities of text, able to learn the characteristics of the general language and “understand” and generate text. However, LLMs deal with tokens which can be a word, part of a word (subword) or even a character. Thus, their production is not based on lexicogenic material (semantic or morphologic). They only mathematically predict the next word given the input prompt. Unlike natural human language, LLMs lack common sense and creativity, do not understand the text they generate, can be biased, and cannot verify what they generate. Open resources will increasingly reflect the usage of a global language, artificially enhanced, which will affect the study of neology.

The European Commission’s Living guidelines on the Responsible Use of Generative AI mention various principles that apply to the responsible use of AI: transparency (on the use of generative AI), verification (of the correctness of the generated output) and responsibility (for the correct use of generative AI and the published output). Llf qualifies for ‘transparency’, as the fictive quotations were clearly identified as AI-generated. Can the correctness of a quotation illustrating the usage of neologism be verified? Is there a risk to assess for the responsibility principle? The only risk to be considered is the unnecessary LLM activity, and its potential environmental impact in

order to produce an output that has neither needs to fulfil nor identifiable purpose.

5. Quotations in dictionaries

Examining the OED (EOD) was a research project (2005-2022) that has attempted “to assess the use of the quotations made by the editors in writing definitions and providing editorial commentary”. Their online sources, while making “extensive use of electronic resources” since 1984 are all listed and traceable. From emails received from the readers to recording quotations from database, looking for evidence is at the heart of lexicographic practices.

The CeRLA⁹ research centre invited Danica Salazar from the OED in 2025. She shared her expertise on “Using the OED for academic research” and “the OED and World Englishes”. It was interesting to find out that it takes 6 months to 2 years to create an entry. No data is ever retrieved from the OED, showing oldest usage always first, as a testimony of the evolution of language. There is no comparison between sorrow lexicographic practices and AI-powered dictionary chatbots.

Occasionalisms are sometimes listed in the OED. They are often found under the ‘arbitrary’ type of formation (e.g., mithril, Oompa Loompa, tove, etc.).

A neologist specialised in literary lexical innovations is particularly interested to read, in the OED, entries such as “merry-go-rounder”. However obsolete and rare, this occasionalism, coined by Dickens in *Oliver Twist* to express “a cause of astonishment” has a purpose and a meaning – in the book. It shows the ability to play with language and extend lexicogenic processes beyond productive word-formations. The quotation (“Oh, my eye! here's a merry-go-rounder! – Tommy Chitling's in love!”) reflects Dicken’s creativity. It might not be any more useful than the entry for *apeurance* though.

Prompting Yeschat to “Provide an example quotation for the word "merry-go-rounder”, the output is assertive, convincing, though not what Dickens – who definitely does not use American slang in *Oliver Twist* – expressed in 1838:

The word "merry-go-rounder" is a rare and somewhat informal term. It generally refers to someone who is frequently involved in shifting positions, often in politics or business, or someone who takes part in repetitive or cyclic activity, like a "merry-go-round."

An example quotation comes from the Dictionary of American Slang:

"One merry-go-rounder after another took over the job, each bringing in his own team and throwing the old one out."

This usage reflects the metaphor of constant rotation or change, much like the movement of a merry-go-round.

Let me know if you'd like literary citations or usage from another era.

⁹ CeRLA : Centre de recherche en linguistique appliquée, Université Lumière Lyon 2, France.

While we did not verify if ‘merry-go-rounder’ is indeed listed in the Dictionary of American Slang, the chatbot did not find Dickens quotation, though available through a Google search when associating ‘merry-go-rounder’ and ‘Oliver Twist’. A better prompt would likely give better results. Knowledge is paramount when using AI tools.

Fictive quotations have now been fully removed from the dictionary entries of La langue française. Le Roux explained that “they did not bring much value to the dictionary entries since AI is not always able to find the right usage context for certain words, which made it confusing.” There are challenges and opportunities in determining the relevance of machine intervention in lexicographic practices and the relevance of citing quotation evidence, or absence, thereof. Practices are changing. Time will tell.

6. Conclusion

Looking back at the origins of lexicographic practices, in the preface of his English Dictionary (1841: xxiv), Samuel Johnson speaks wisdom:

no dictionary of a living tongue ever can be perfect, since, while it is hastening to publication, some words are budding, and some falling away; that a whole life cannot be spent upon syntax and etymology, and that even a whole life would not be sufficient that he, whose design includes whatever language can express, must often speak of what he does not understand;

AI does not ‘understand’ language. LLMs will surely learn about neology and be able to reproduce productive and predictive word-formation processes in any language. For now, they surely know how to mimic lexicographic entries and provide more data than any human will ever be able to process in a lifetime. About two decades ago, Sue Atkins (2002: 25) pointed out that “wealth of data alone does not make a good dictionary. It simply swamps us with a dazzling array of facts and no systemic way of evaluating them.” AI-generated text in dictionaries may now provide a wealth of potentially unreliable data that swamp internet users with a dazzling array of systemically presented facts and no way of evaluating their authenticity.

7. References

- Aronoff, M. (2023). Three ways of looking at morphological rivalry. *Word Structure*, 16(1), 49-62.
- Atkins, B. S. (2002). Then and now: Competence and performance in 35 years of lexicography. *10th EURALEX. Proceedings, Copenhagen*, pp 1-28.
- Bauer, L. (2004). *A Glossary of Morphology*. Washington, D.C.: Georgetown University Press.
- Boulanger, J.-C., Courbon, B. (dir.) & Auger, B. (2024). *De la conception du lexique à sa dictionnarisation*. Limoges: Lambert-Lucas.

- Dal, G., & Namer, F. (2010). Les noms en-ance/-ence du français : quel(s) patron(s) constructionnel(s)?. *Proceedings of the 2^o Congrès Mondial de Linguistique Française*, pp. 893-907.
- De Roucy-Rochegonde, L. & Buffard, A. (2025). *AI, Data Centers and Energy Demand: Reassessing and Exploring the Trends*. Ifri Papers, Paris: Ifri.
- European Commission, Directorate-General for Research and Innovation. (2025). *Living guidelines on the responsible use of generative AI in research*. Brussels, ERA Forum Stakeholders' document.
- François, A. (1950). *La désinence « ance » dans le vocabulaire français, une « pédale » de la langue et du style. Essai historique suivi du répertoire des mots contemporains finissant par ance avec un appendice des mots finissant par ence et escence*. Genève : Droz.
- Goosse A. (1952). François (Alexis). La désinence « ance » dans le vocabulaire français, une « pédale » de la langue et du style. Essai historique suivi du répertoire des mots contemporains finissant par ance avec un appendice des mots finissant par ence et escence. *Revue belge de philologie et d'histoire*, 30(1-2), pp. 224-226.
- Gréa, P., Knittel, M.L., Marín R. & Villoing, F. (2023). Innovative uses of French neological -ance nominalizations, ISMo 2023: International Symposium of Morphology.
- Johnson, S., Walker, J., Todd, H. J. & Chalmers, A. (1841). *Johnson's English dictionary, as improved by Todd and abridged by Chalmers*. Philadelphia, Kimber & Sharpless.
- Knittel, M.L. (2016). Les noms en-ance: un panorama. *Proceedings of the SHS Web of Conferences, Vol. 27*, pp. 12011. EDP Sciences.
- Knittel, M.L. & Marín, R. (2022). L'héritage transcatégoriel des propriétés sémantiques : le cas des Noms en-ance et des verbes et adjectifs apparentés. *Proceedings of the SHS Web of Conferences, Vol. 138*, pp. 07001. EDP Sciences.
- Murano, M. (2014). La lexicographie 2.0 : nous sommes tous lexicographes. *Cahiers de Recherche de l'École Doctorale en Linguistique Française*, 8, pp. 147-162. Trieste, Edizioni Università di Trieste.
- Poix, C. (2021). Études francophones de néologie-Complexité terminologique. *Neologica*, 15, pp. 97-115.
- Poix, C. and M.A. Rebours. (2025). Reviviscence du suffixe -ance : déviance ou malaisance ? *CINEO VI Congresso internacional de neologica em línguas românicas*. Coimbra.
- Pruvost, J. (2003). Some Lexicographic Concepts Stemming from French Training in Lexicology (1): The helpful distinction between theoretical lexicography and practical dictionary-making. *Kernerman Dictionary News*, 11, pp. 10-15. https://lexicala.com/wp-content/uploads/kdn11_2003_Somelexicographic-concepts_JP.pdf
- Rey, A. (2013). Interview at Sauramps bookstore, Montpellier for the presentation of *Dictionnaire amoureux des Dictionnaires*. Plon. https://www.youtube.com/watch?v=H_hBewH98ZQ.

- Sablayrolles, J.F. (2019). *Comprendre la néologie. Conceptions, analyses, emplois*. Limoges: Lambert Lucas.
- Schmid, H.J. (2020). *The dynamics of the linguistic system: Usage, conventionalization, and entrenchment*. Oxford University Press.
- Steuckardt, A. (2024). Le suffixe en –ance dans l’histoire de la langue française : déclin et résurgences. *Regards croisés sur le changement linguistique*. Sheerbrooke.
- Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam: John Benjamins.

Corpora and online resources

- DeepL Accessed at <https://www.deepl.com/en/translator> (18 July 2025)
- EOED (Examining the Oxford English Dictionary) (2005-2022). Accessed at <https://oed.hertford.ox.ac.uk/about/aims-and-findings/> (18 July 2025)
- frTenTen. Accessed at <https://www.sketchengine.eu/frtnten-french-corpus/> (18 July 2025)
- La langue française. Accessed at <https://www.lalanguefrancaise.com/> (18 July 2025)
- Le Grand Robert*. Accessed at <https://www.lerobert.com/collection-le-grand-robert-de-la-langue-francaise.html> (18 July 2025)
- Le Trésor de la Langue Française informatisé (TLFi). Centre National de Ressources Textuelles et Lexicales. Accessed at <https://www.cnrtl.fr/definition/> (18 July 2025)
- Listedemots.net. Accessed at <https://www.listesdemots.net/accueil.htm> (18 July 2025)
- Oxford English dictionary. Accessed at <https://www.oed.com/> (18 July 2025)
- OpenSubtitles parallel corpora 2018. <https://www.opensubtitles.org/fr>
- Wiktionnaire*. Accessed at https://fr.wiktionary.org/wiki/Wiktionnaire:Page_d%E2%80%99accueil (18 July 2025)
- Yeschat – AI-powered dictionary tool Accessed at <https://www.yeschat.ai/gpts-9t557DWSYap-Dictionary> (18 July 2025)

This work is licensed under the Creative Commons Attribution ShareAlike 4.0 International License.

<http://creativecommons.org/licenses/by-sa/4.0/>

