

---

# Out-of-the-Box GPTs for Lexicography

Gilles-Maurice de Schryver, *BantUGent & LT<sup>3</sup>, Ghent University,  
Ghent, Belgium; and Department of African Languages,  
University of Pretoria, Pretoria, South Africa*  
([gillesmaurice.deschryver@UGent.be](mailto:gillesmaurice.deschryver@UGent.be))  
(<https://orcid.org/0000-0001-7272-9878>)

---

**Abstract:** A detailed analysis is offered of how ChatGPT-3.5 has treated the world's languages for lexicographic purposes, starting with hyper-visible English, moving to semi-visible Portuguese, and ending with low-visible Bantu languages. Case studies for the latter group are region-visible Swahili, nation-visible Xhosa and local-visible Lusoga. This impact-based scale is reflected in the uptake of GenAI technology for lexicography. The results of using out-of-the-box ChatGPT-3.5 are fantastic for English, fake for Portuguese, and simply fail for the Bantu languages. But there is a way out, and the seeds for that are sown in this article.

**Keywords:** GENERATIVE ARTIFICIAL INTELLIGENCE (GENAI), LARGE LANGUAGE MODELS (LLMs), GENERATIVE PRE-TRAINED TRANSFORMER (GPT), OUT-OF-THE-BOX GPTs, LEXICOGRAPHY, DICTIONARIES, LEXICOGRAPHERS, ENGLISH, PORTUGUESE, BANTU, SWAHILI, XHOSA, LUSOGA

**Obufunze: GPT ey'omutindo eya GenAI mw'isomo ly'okuwandiika amawanika.** Okutangaazibwa ku ngeli ChatGPT-3.5 bw'ekozeseibwa mu nnimi dh'ensi yoonayoona mu kugezaaku okutuukiliza ebigendelelwa by'okuwandiika amawanika, kuweebwa mu lupapula luno. Okutangaaza kuno kutandiikila ku nkozesa eyolekebwa ku lulimi oluganzi mu nsi yoonayoona olw'Olungeleza, kw'amala kwaizibwa ku nnimi engazimuuku akatono ng'Olupootigo, kw'asembyayo enkozesa eyeeyolekela mu nnimi edhikozesebwa abantu abatono einho ng'edho edhigwa mu luse lwa Bantu. Ennimi dha Bantu edhiliku enkozesa entono einho edhigelaagela-nhizibwa ku y'edho emba oba enjanziku n'omuli, olulimi lw'Oluswayili, Oluxhosa n'Olusoga. Ennimi dhino dhonsatule n'edhisinziilwaku okulaga tekinologiya wa GenAI bwe yeeyolekelamu ng'akozesebwa mu milimu gy'okuwandiika amawanika. Kizuuse nti enkozesa ya tekinologiya eno nnambulukufu inho mu milimu egikola ku mawanika g'Olungeleza, aye, ya kibogwe ku nnimi enjanziku akatono ng'Olupootigo. Bwe kituuka ku nnimi dh'omu luse lwa Bantu kabili enkola eno eililila inho y'aba oti kumpi tekelela ilala. Aye, wakaaliwooku eisuubi eliva mu byandhulwa mu kitundu ky'olupapula luno eky'okubili. Mu lupapula olw'okubili n'omuli ebyava mu kwongela okunoonheleza ku ngeli eyinza okukozesebwa okuziba olutanda oluli wagati w'ennimi dh'ensi yoonayoona enjanzu, n'edhitali, aye nga dhoonadhoona dhiluubilila okukoseza GenAI okuwandiika amawanika.

**Ebigambo ebikulu:** KALIMAGEZI-KISIBUKILA, OKUTETENKELEZA KW'ENNIMI ENNHINGI, KISIBUKILA AKAALI KUKENGULWA, GPT EY'OMUTINDO, EISOMO LY'OKUWA-

NDIIKA AMAWANIKA, AMAWANIKA, ABAWANDIISI B'AMAWANIKA, OLUNGELEZA, OLUPOOTIGO, BANTU, OLUSWAYILI, OLUXHOSA, OLUSOGA

## 1. Prologue

In De Schryver (forthcoming-a, forthcoming-b) a sample of 100 contributions on Generative Artificial Intelligence (GenAI) in lexicography was studied, covering the period from the release of ChatGPT on 30 November 2022 up to 31 May 2025. Over those two years and a half, all but one of the contributions simply used *out-of-the-box* Generative Pre-trained Transformer (GPT) models. Those types of Large Language Models (LLMs) are known for their ability to generate sophisticated human-like texts. Just a single one of the 100 contributions moved to the next level for lexicography, namely to a *customised* GPT model (De Schryver 2024a). However, and despite the expressed wishes about it at gatherings on digital lexicography, starting with eLex 2023, not a single attempt has been made to date to move to the next-next level, namely to a *fine-tuned* GPT model. In the present article the first step is revisited for ChatGPT, and in a next article (De Schryver 2025) the second step is revisited with attention to both OpenAI's ChatGPT and Google's Gemini, expanding the four-page abstract on it prepared (but not delivered) for a workshop on LLMs and lexicography at Euralex 2024 (De Schryver 2024c).

## 2. Out-of-the-box GPTs

### 2.1 Hyper-visible English

By and large, the most successful experiments with LLMs for lexicography have been reported on for applications using English (1.45 billion speakers; spoken in 59 sovereign countries and 27 non-sovereign entities). See especially De Schryver (2023b), Lew (2023), Rees and Lew (2024), Lew (2024), Ptasznik et al. (2024) and Lew et al. (2024). At the same time, some colleagues correctly pointed out a few caveats, see especially McKean and Fitzgerald (2023), Jakubíček and Rundell (2023) and Nichols (2023). That said, virtually all scholars agree that it is not a good idea not to bring in today's GenAI for English lexicography.

### 2.2 Semi-visible Portuguese

The results of the first in-depth and detailed attempt to study the lexicographic potential of using ChatGPT for a language other than English, were presented at the inaugural Americalex-S conference in São Paulo, Brazil (De Schryver 2023a). The assumption was that ChatGPT would do an excellent job for the rather large language Portuguese (260 million speakers; spoken in 10 coun-

tries and territories). Following the success of using ChatGPT-3.5 for compiling a monolingual English COBUILD-style learners' dictionary — demonstrating its incredible performance on lemmas from a variety of word classes: verbs, nouns, conjunctions, adjectives and interjections, as well as metaphorical meanings and idiomatic expressions (see De Schryver 2023b: supplementary material) — the first goal was to test ChatGPT-3.5 on a monolingual Portuguese COBUILD-style learners' dictionary. There was an ongoing student strike on campus during the conference, so the polysemous lemma *strike* was chosen, and put into a Portuguese translation of the same prompt which had been used for English in De Schryver (2023b: supplementary material), as shown in Figure 1. Here and elsewhere, English translations of the Portuguese are to be found in Addendum A. Note that in De Schryver (2023b) the goal was to arrive at a single prompt with which each type of lemma, no matter its part of speech or any other aspects, may be generated. The 'prompt engineering' thus preceded the prompt eventually proposed, with which an entire COBUILD dictionary could be compiled. Copies of De Schryver's (2023b) prompt have subsequently been used by colleagues, such as Zhong (2025) using DeepSeek, and also in translated form, such as in Spanish by Fuertes-Olivera (2024), or in Lithuanian by Pavydytė (2024).

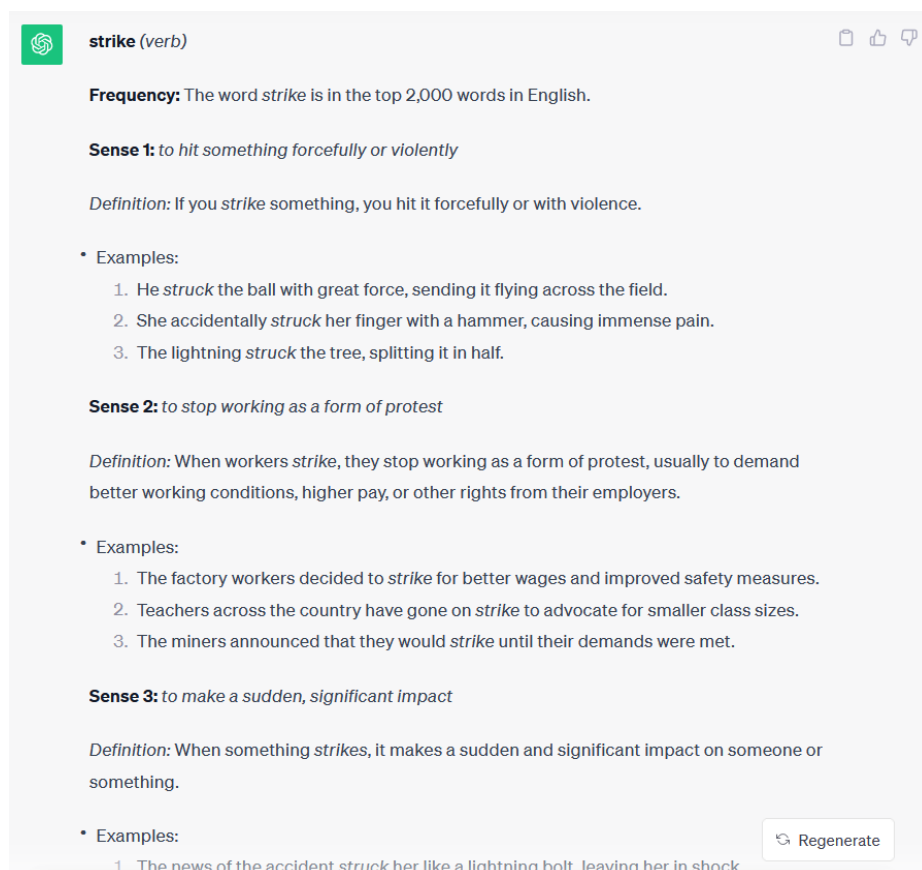
Default (GPT-3.5)



Por favor, forneça um verbete de dicionário no estilo COBUILD para "strike". Comece com as informações de frequência na parte superior. Cada sentido deve estar em um bloco numerado. Cada bloco começa com a classe gramatical e as formas morfológicas do respectivo sentido. Em seguida, há uma definição de sentido e exemplos de sentido que ilustram o uso e o significado de cada sentido específico. Para as frases de exemplo, certifique-se de usar diferentes estruturas de frases, referindo-se a diferentes pessoas; refira-se a situações passadas, presentes e futuras; varie frases de exemplo longas e curtas; e inclua outras elaborações para que você, caro ChatGPT, produza material de dicionário de alta qualidade para mim.

**Figure 1:** ChatGPT-3.5 prompt to obtain a COBUILD-style entry for a monolingual Portuguese learners' dictionary, here for the lemma *strike*

Even though the prompt was written in Portuguese, the reply from ChatGPT was entirely in English, as seen in Figure 2 (only showing the first part, the start of the verbal use).



**Figure 2:** Start of the ChatGPT-3.5 output for the prompt seen in Figure 1

This was followed up with: 'Thanks, but actually I wanted to see a monolingual dictionary entry for the Portuguese entry '*greve*' (which is 'strike' in English)', after which ChatGPT replied with what is seen in Figure 3.

This time, ChatGPT presented a noun (correct), with an indication that it is feminine (correct), in a wider frequency band compared to English (ambiguous), a good definition (but not a COBUILD-style one), acceptable examples and an interesting encyclopaedic note. Observe that examples 1 and 2 in Figure 3, are very similar to examples 1 and 2 under the verbal sense 2 seen in English in Figure 2, except that the second one there is not even a verb, but part of a verb phrase where *strike* is used as a noun (so an error). From the very first reply by ChatGPT onwards, and now reinforced by the similar examples in the second reply, one starts wondering if one is not dealing with translation issues (in the sense that the model generates everything in and for English first, and only then maps the English onto another language as if it were English).



Obrigado, mas na verdade eu queria ver um verbete de dicionário monolíngue para o verbete português "greve" (que é "strike" em inglês).



Peço desculpas pela confusão anterior. Aqui está um verbete de dicionário monolíngue para a palavra "greve" em português:

**greve** (*substantivo feminino*)

**Frequência:** A palavra *greve* está entre as 10.000 palavras mais frequentes em português.

**Definição:** A *greve* é uma paralisação organizada por trabalhadores, na qual eles deixam de trabalhar como forma de protesto, geralmente para reivindicar melhores condições de trabalho, salários mais altos, ou outros direitos dos empregadores.

**Exemplos:**

1. Os funcionários da fábrica entraram em *greve* para exigir um aumento salarial.
2. Os professores decidiram iniciar uma *greve* devido à falta de investimentos na educação.
3. A *greve* dos caminhoneiros causou escassez de suprimentos em todo o país.

**Nota:** A palavra "greve" é frequentemente usada no contexto de protestos trabalhistas e é um direito garantido por leis trabalhistas em muitos países. Ela pode ser realizada por diversos setores profissionais, incluindo professores, médicos, motoristas, e outros, como um meio de pressionar empregadores ou governos a atenderem às demandas dos trabalhadores.

Regenerate

**Figure 3:** ChatGPT-3.5 output for the noun *greve* 'strike', for a monolingual Portuguese learners' dictionary

Next, the verb *recomendar* 'recommend', being one of the 15 verbs of communication used by Lew (2023), was submitted. The prompt (with an extra line) is shown in Figure 4. The reply by ChatGPT looked rather satisfactory, as may be seen from Figures 5a, b and c.



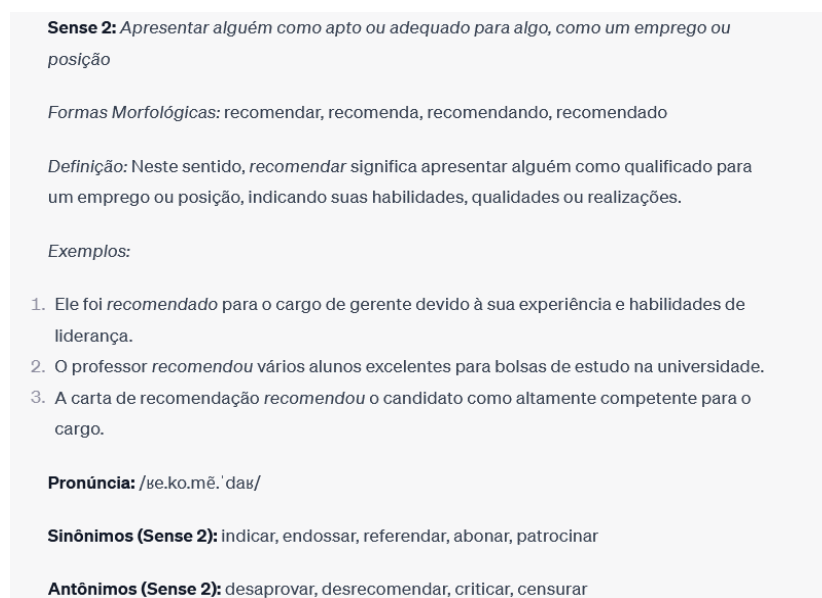
Por favor, forneça um verbete de dicionário monolíngue no estilo COBUILD para "recomendar". Comece com as informações de frequência na parte superior. Cada sentido deve estar em um bloco numerado. Cada bloco começa com a classe gramatical e as formas morfológicas do respectivo sentido. Em seguida, há uma definição de sentido e exemplos de sentido que ilustram o uso e o significado de cada sentido específico. Para as frases de exemplo, certifique-se de usar diferentes estruturas de frases, referindo-se a diferentes pessoas; refira-se a situações passadas, presentes e futuras; varie frases de exemplo longas e curtas; e inclua outras elaborações para que você, caro ChatGPT, produza material de dicionário de alta qualidade para mim. Inclua também uma pronúncia do verbete usando o IPA e inclua uma lista de sinônimos e antônimos nos sentidos em que essas listas forem relevantes.



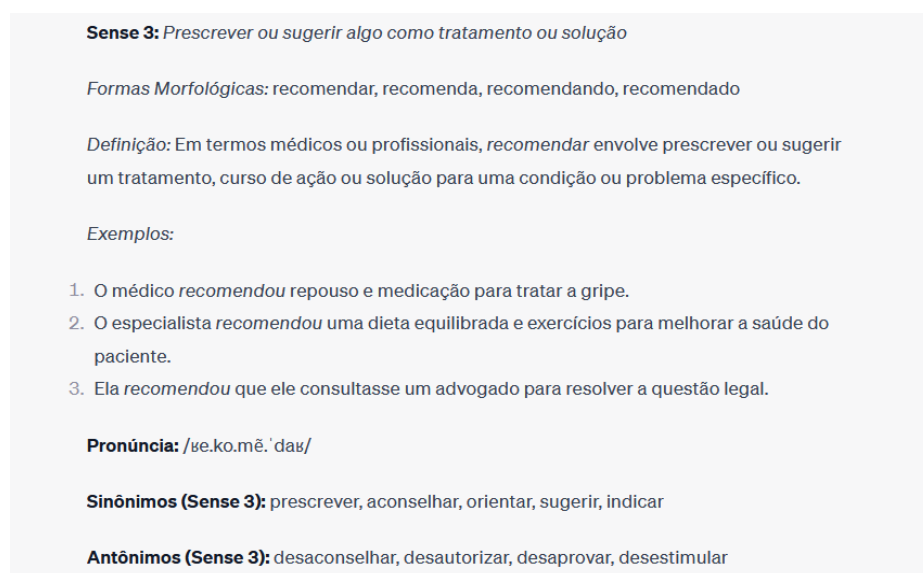
**Figure 4:** ChatGPT-3.5 prompt to obtain a COBUILD-style entry for a monolingual Portuguese learners' dictionary, here for the verb of communication *recomendar* 'recommend'



**Figure 5a:** ChatGPT-3.5 output for the Port. verb *recomendar* 'recommend', # 1, for the prompt seen in Figure 4



**Figure 5b:** ChatGPT-3.5 output for the Port. verb *recomendar* 'recommend', # 2, for the prompt seen in Figure 4



**Figure 5c:** ChatGPT-3.5 output for the Port. verb *recomendar* 'recommend', # 3, for the prompt seen in Figure 4

However, and again — so clearly a serious issue — two of the three definitions are not following the COBUILD-style, despite the explicit request to do so in the prompt. An attempt was then made to even add as second sentence to the prompt: 'The most important aspect of the COBUILD-style is that for the definition, you need to use the full-sentence defining style, which consists of two clauses: the first clause illustrates the usage pattern, the second clause is a paraphrase of meaning'. But that had no effect whatsoever, despite the fact that merely mentioning COBUILD-style the very first time, resulted in a definition very close to it, see Figure 2 for the verb *strike*: 'If you strike something, you hit it forcefully or with violence'. Not as perfect as what is seen in an actual COBUILD, but close: 'If you strike someone or something, you deliberately hit them'.<sup>1</sup> Conclusion: ChatGPT-3.5 is unable to compile proper COBUILD-style entries for a monolingual Portuguese learners' dictionary.

What about a bilingual dictionary? The first focus is on an English to Portuguese entry (in that direction), for the lemma *bank*, with the prompt in English as seen in Figure 6.



Give me a bilingual English-Portuguese dictionary entry for the word 'bank'. Start with the frequency information at the top. Each sense should be in a numbered block. Each block starts with the part of speech and the morphological forms of the respective sense. This is followed by a sense definition in English, and a translation of that definition into Portuguese. Note that for the definition, you have to use full-sentences, not codes, and these sentences consist of two clauses: the first clause illustrates the usage pattern, the second clause is a paraphrase of meaning. Also list one or more translation equivalents in Portuguese at each sense. Add English examples with their Portuguese translations that illustrate both the use and the meaning of each particular English sense. For the example sentences, make sure to use different sentence structures, referring to different people; refer to past, present, and future situations; vary long and short example sentences; and include other elaborations so that you, dear ChatGPT, produce high-quality dictionary material for me. Also include a pronunciation of the entry using IPA, and include a list of synonyms and antonyms at the senses where such lists are relevant.

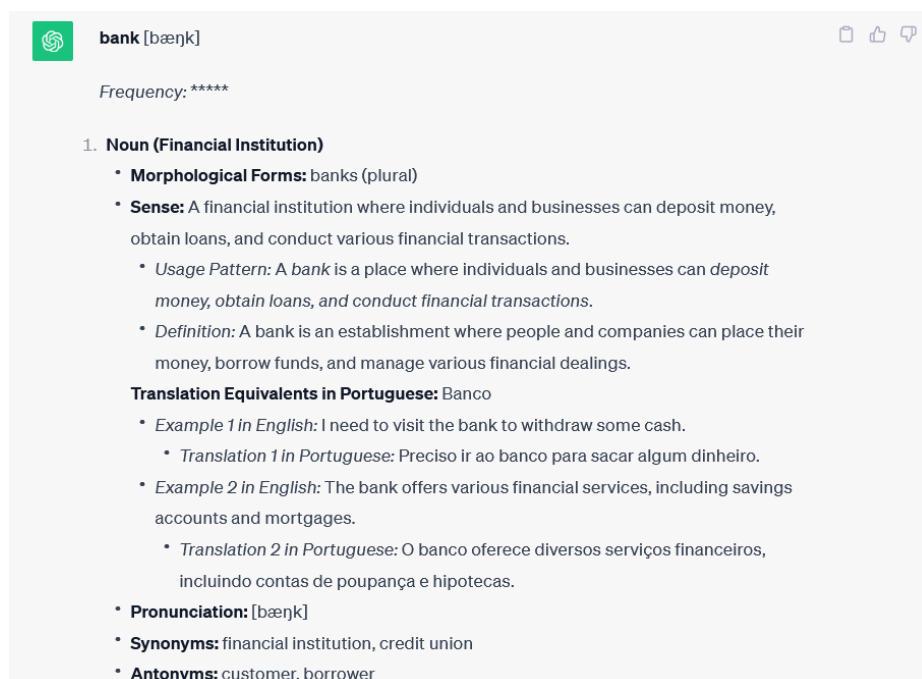
**Figure 6:** Prompt to obtain a COBUILD-style entry for a bilingual English-Portuguese learners' dictionary, here for the lemma *bank*. With also more information on what is meant with a 'COBUILD definition'

The lemma *bank*, a classic since first discussed seven decades ago by Noam Chomsky (1957: 95), and often used by Patrick Hanks ever after (Hanks 1987, 2000, 2013), has multiple nominal and verbal uses. ChatGPT offers three senses, two nominal ones, shown in Figures 7a and b, and one verbal one, shown in Figure 7c.

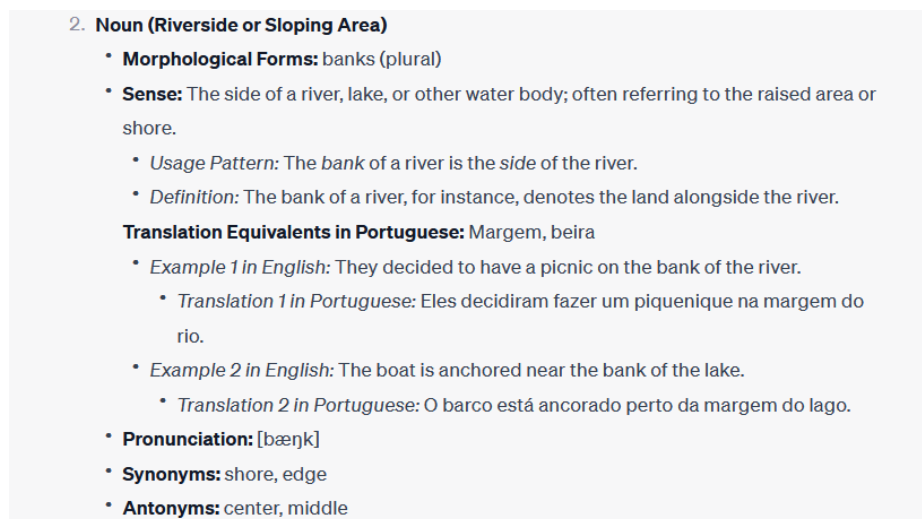
Clearly, ChatGPT did not understand what is meant with the COBUILD-style full-sentence definition (FSD).<sup>2</sup> The two clauses became two blocks, one for the usage pattern, the other for the paraphrase of meaning. Despite the fact that three attempts at defining are now made, under the heading 'Sense', the heading 'Usage Pattern' and the heading 'Definition', it is only at the third sense that a vague attempt is made at 'Definition': '*When someone banks something, they are either depositing it in a financial institution or relying on it for future use*'. Much more convoluted than what COBUILD does: '*If you bank money, you pay it into a bank*'.<sup>3</sup>

Plus, the first example for this verbal use of *bank* follows the pattern [VERB + in] and the second [VERB + on], so they don't really belong there, while COBUILD itself actually lists [VERB + with] as a verbal sense in itself: '*If you bank with a particular bank, you have an account with that bank*'.<sup>4</sup>





**Figure 7a:** ChatGPT-3.5 output for the English lemma *bank*, # 1, for the prompt seen in Figure 6



**Figure 7b:** ChatGPT-3.5 output for the English lemma *bank*, # 2, for the prompt seen in Figure 6

### 3. Verb

- **Morphological Forms:** banks (3rd person singular present), banked (past tense), banking (present participle), banks (3rd person singular present)
- **Sense:** To deposit or store something in a bank, or to rely on or store something for future use.
  - *Usage Pattern:* Someone banks something in a bank, or they bank on something for a specific purpose.
  - *Definition:* When someone banks something, they are either depositing it in a financial institution or relying on it for future use.
- Translation Equivalents in Portuguese:** Depositar, guardar, contar com
  - *Example 1 in English:* She banks her savings in a local bank.
    - *Translation 1 in Portuguese:* Ela deposita suas economias em um banco local.
  - *Example 2 in English:* He's banking on winning the competition with his skills.
    - *Translation 2 in Portuguese:* Ele está contando com sua vitória na competição com suas habilidades.
- **Pronunciation:** [bæŋk]
- **Synonyms:** deposit, rely on
- **Antonyms:** withdraw, disbelieve

**Figure 7c:** ChatGPT-3.5 output for the English lemma *bank*, # 3, for the prompt seen in Figure 6

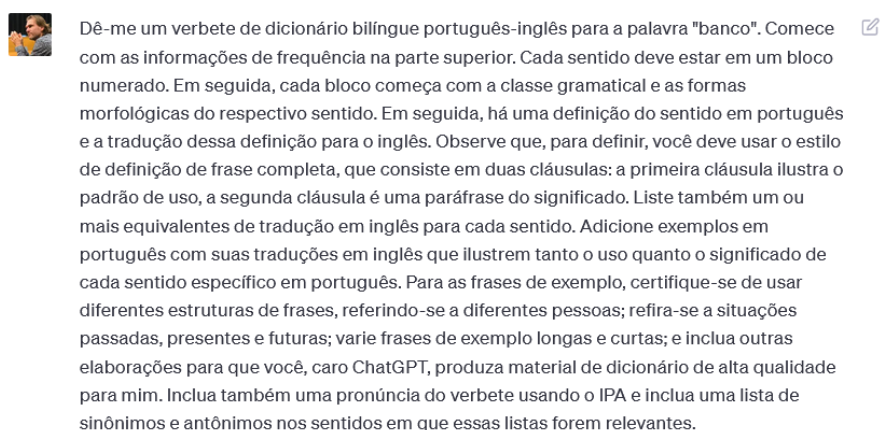
Furthermore, given that a request was made for an English to Portuguese dictionary, there is also still too much in English. The next prompt became: '*Right, but what happened to the Portuguese? It is supposed to be a bilingual dictionary!*'. After that ChatGPT continued with presenting two versions, asking '*Which response do you prefer?*'. Part of the two versions is shown in Figure 8.

In the version on the left of Figure 8, the 'Definition' is now in Portuguese only, while in the version on the right of Figure 8, the synonyms and antonyms are in Portuguese only. Both additions are fine, but it had been better if both options were merged. Conclusion: As was the case for the monolingual test, it is clear that also for this direction of a bilingual dictionary, ChatGPT-3.5 is not up to the task, as it is unable to compile proper COBUILD-style entries for an English–Portuguese learners' dictionary.


<p>1. <b>Noun (Institution)</b></p> <ul style="list-style-type: none"> <li>• <b>Morphological Forms:</b> banks (plural)</li> <li>• <b>Sense:</b> A financial institution where individuals and businesses can deposit money, obtain loans, and conduct various financial transactions.</li> <li>• <b>Usage Pattern:</b> A bank is a place where individuals and businesses can deposit money, obtain loans, and conduct financial transactions.</li> <li>• <b>Definition:</b> Um banco é uma instituição financeira onde pessoas e empresas podem depositar dinheiro, obter empréstimos e realizar várias transações financeiras.</li> </ul> <p><b>Translation Equivalents in Portuguese:</b> Banco</p> <ul style="list-style-type: none"> <li>• <b>Example 1 in English:</b> I need to visit the bank to withdraw some cash.</li> <li>• <b>Translation 1 in Portuguese:</b> Preciso ir ao banco para sacar algum dinheiro.</li> <li>• <b>Example 2 in English:</b> The bank offers various financial services, including savings accounts and mortgages.</li> <li>• <b>Translation 2 in Portuguese:</b> O banco oferece diversos serviços financeiros, incluindo contas de poupança e hipotecas.</li> </ul> <ul style="list-style-type: none"> <li>• <b>Pronunciation:</b> [bæŋk]</li> <li>• <b>Synonyms:</b> financial institution, credit union</li> <li>• <b>Antonyms:</b> customer, borrower</li> </ul>	<p>1. <b>Noun (Financial Institution)</b></p> <ul style="list-style-type: none"> <li>• <b>Morphological Forms:</b> banks (plural)</li> <li>• <b>Sense:</b> A financial institution where individuals and businesses can deposit money, obtain loans, and conduct various financial transactions.</li> <li>• <b>Usage Pattern:</b> A bank is a place where individuals and businesses can deposit money, obtain loans, and conduct financial transactions.</li> <li>• <b>Definition:</b> A bank is an establishment where people and companies can place their money, borrow funds, and manage various financial dealings.</li> </ul> <p><b>Translation Equivalents in Portuguese:</b> Banco</p> <ul style="list-style-type: none"> <li>• <b>Example 1 in English:</b> I need to visit the bank to withdraw some cash.</li> <li>• <b>Translation 1 in Portuguese:</b> Preciso ir ao banco para sacar algum dinheiro.</li> <li>• <b>Example 2 in English:</b> The bank offers various financial services, including savings accounts and mortgages.</li> <li>• <b>Translation 2 in Portuguese:</b> O banco oferece diversos serviços financeiros, incluindo contas de poupança e hipotecas.</li> </ul> <ul style="list-style-type: none"> <li>• <b>Pronunciation:</b> [bæŋk]</li> <li>• <b>Synonyms:</b> instituição financeira, caixa econômica</li> <li>• <b>Antonyms:</b> cliente, mutuário</li> </ul> <p>2. <b>Noun (Riverside or Sloping Area)</b></p>
--	---

**Figure 8:** Part of the ChatGPT-3.5 output for the prompt seen in Figure 6, plus a request for more Portuguese, in two versions (left vs. right)

As a final test, the reverse was tried, producing a COBUILD-style Portuguese–English learners' dictionary, with the prompt shown in Figure 9, and the reply as seen in Figures 10a, b and c.



**Figure 9:** Prompt to obtain a COBUILD-style entry for a bilingual Portuguese–English learners' dictionary, here for the lemma *banco* 'bank'



The screenshot shows the ChatGPT-3.5 interface. At the top, the word 'banco' is followed by its phonetic transcription ['bẽku]. To the right are icons for a document, a thumbs up, and a thumbs down. Below this, the frequency is indicated as '\*\*\*\*\*'. The main content is organized into sections. The first section is titled '1. Substantivo (Instituição Financeira)'. It contains several bullet points: 'Formas Morfológicas: bancos (plural)', 'Sentido: Uma instituição financeira onde pessoas e empresas podem depositar dinheiro, obter empréstimos e realizar várias transações financeiras.', 'Padrão de Uso: Um banco é um local onde pessoas e empresas podem depositar dinheiro, obter empréstimos e realizar transações financeiras.', and 'Definição: Um banco é um estabelecimento onde pessoas e empresas podem colocar seu dinheiro, obter empréstimos e gerenciar várias transações financeiras.' This is followed by 'Equivalentes de Tradução em Inglês: Bank'. Then, two examples are provided: 'Exemplo 1 em Português: Preciso ir ao banco para sacar algum dinheiro.' with its English translation 'Tradução 1 em Inglês: I need to go to the bank to withdraw some money.', and 'Exemplo 2 em Português: O banco oferece diversos serviços financeiros, incluindo contas de poupança e hipotecas.' with its English translation 'Tradução 2 em Inglês: The bank offers various financial services, including savings accounts and mortgages.' The section ends with 'Pronúncia: ['bẽku]', 'Sinônimos: instituição financeira, agência bancária', and 'Antônimos: cliente, mutuário'.

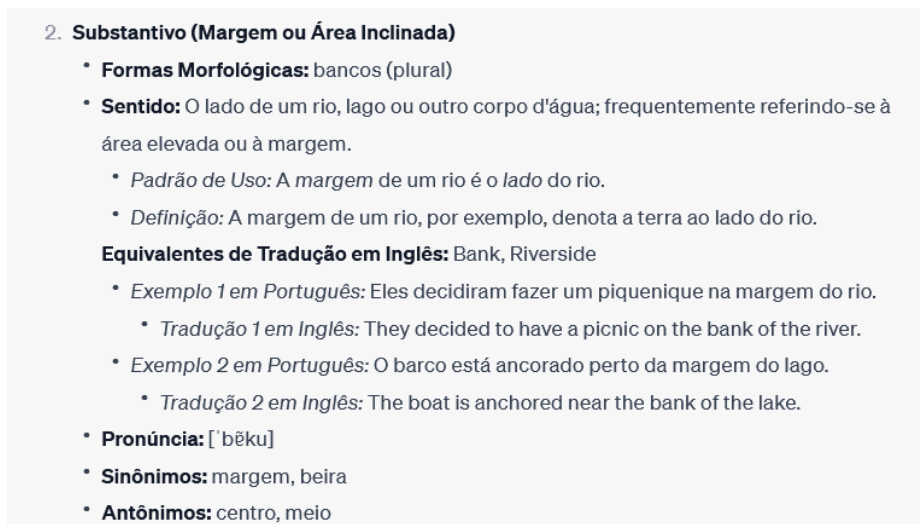
**banco** ['bẽku]

Frequência: \*\*\*\*\*

1. **Substantivo (Instituição Financeira)**

- **Formas Morfológicas:** bancos (plural)
- **Sentido:** Uma instituição financeira onde pessoas e empresas podem depositar dinheiro, obter empréstimos e realizar várias transações financeiras.
  - **Padrão de Uso:** Um banco é um local onde pessoas e empresas podem depositar dinheiro, obter empréstimos e realizar transações financeiras.
  - **Definição:** Um banco é um estabelecimento onde pessoas e empresas podem colocar seu dinheiro, obter empréstimos e gerenciar várias transações financeiras.
- **Equivalentes de Tradução em Inglês:** Bank
  - **Exemplo 1 em Português:** Preciso ir ao banco para sacar algum dinheiro.
    - **Tradução 1 em Inglês:** I need to go to the bank to withdraw some money.
  - **Exemplo 2 em Português:** O banco oferece diversos serviços financeiros, incluindo contas de poupança e hipotecas.
    - **Tradução 2 em Inglês:** The bank offers various financial services, including savings accounts and mortgages.
- **Pronúncia:** ['bẽku]
- **Sinônimos:** instituição financeira, agência bancária
- **Antônimos:** cliente, mutuário

**Figure 10a:** ChatGPT-3.5 output for the Portuguese lemma *banco*, # 1, for the prompt seen in Figure 9



This block continues the output from Figure 10a. It starts with the section title '2. Substantivo (Margem ou Área Inclínada)'. It contains several bullet points: 'Formas Morfológicas: bancos (plural)', 'Sentido: O lado de um rio, lago ou outro corpo d'água; frequentemente referindo-se à área elevada ou à margem.', 'Padrão de Uso: A margem de um rio é o lado do rio.', and 'Definição: A margem de um rio, por exemplo, denota a terra ao lado do rio.' This is followed by 'Equivalentes de Tradução em Inglês: Bank, Riverside'. Then, two examples are provided: 'Exemplo 1 em Português: Eles decidiram fazer um piquenique na margem do rio.' with its English translation 'Tradução 1 em Inglês: They decided to have a picnic on the bank of the river.', and 'Exemplo 2 em Português: O barco está ancorado perto da margem do lago.' with its English translation 'Tradução 2 em Inglês: The boat is anchored near the bank of the lake.' The section ends with 'Pronúncia: ['bẽku]', 'Sinônimos: margem, beira', and 'Antônimos: centro, meio'.

2. **Substantivo (Margem ou Área Inclínada)**

- **Formas Morfológicas:** bancos (plural)
- **Sentido:** O lado de um rio, lago ou outro corpo d'água; frequentemente referindo-se à área elevada ou à margem.
  - **Padrão de Uso:** A margem de um rio é o lado do rio.
  - **Definição:** A margem de um rio, por exemplo, denota a terra ao lado do rio.
- **Equivalentes de Tradução em Inglês:** Bank, Riverside
  - **Exemplo 1 em Português:** Eles decidiram fazer um piquenique na margem do rio.
    - **Tradução 1 em Inglês:** They decided to have a picnic on the bank of the river.
  - **Exemplo 2 em Português:** O barco está ancorado perto da margem do lago.
    - **Tradução 2 em Inglês:** The boat is anchored near the bank of the lake.
- **Pronúncia:** ['bẽku]
- **Sinônimos:** margem, beira
- **Antônimos:** centro, meio

**Figure 10b:** ChatGPT-3.5 output for the Portuguese lemma *banco*, # 2, for the prompt seen in Figure 9

### 3. Verbo

- **Formas Morfológicas:** banco (3ª pessoa do singular do presente), bancou (pretérito), bancando (particípio presente), banco (3ª pessoa do singular do presente)
- **Sentido:** Depositar ou armazenar algo em um banco, ou confiar ou armazenar algo para uso futuro.
  - *Padrão de Uso:* Alguém *bancou* algo em um *banco*, ou eles *bancam* algo para um propósito específico.
  - *Definição:* Quando alguém deposita algo em um banco, está armazenando em uma instituição financeira ou confiando em algo para uso futuro.
- Equivalentes de Tradução em Inglês:** Deposit, Rely on
  - *Exemplo 1 em Português:* Ela deposita suas economias em um banco local.
    - *Tradução 1 em Inglês:* She banks her savings in a local bank.
  - *Exemplo 2 em Português:* Ele está contando com sua vitória na competição com suas habilidades.
    - *Tradução 2 em Inglês:* He's banking on winning the competition with his skills.
- **Pronúncia:** ['bẽku]
- **Sinônimos:** depositar, confiar
- **Antônimos:** sacar, desacreditar

**Figure 10c:** ChatGPT-3.5 output for the Portuguese lemma *banco*, # 3, for the prompt seen in Figure 9

When the output seen in Figures 10a, b and c was presented to two professors in Brazil, one a lexicographer for Portuguese (Regiani Aparecida Santos Zacarias), the other a linguist for Portuguese (who prefers to remain anonymous), they thought it all looked rather good at first. When they were asked (too) many questions, their opinion changed. While the first sense (Figure 10a), for the nominal use as a financial institution, is acceptable, the second sense (Figure 10b), for the nominal use as a sloping area, and third sense (Figure 10c), as a verb, are riddled with errors. The second block deals with the meaning 'bank of a river'. Even though this is supposed to be a sense for the lemma *banco*, the word used under 'Sense', 'Usage Pattern' and 'Definition' as well as in the two examples is a different word entirely: *margem*. At the same time, under the synonym list for *banco* one finds *margem* 'margin; riverbank; shore' and *beira* 'edge; riverbank; precipice verge'. Ironically, and as can already be deduced, only the synonyms presented refer to the bank of a river, using *banco* for that is simply not possible.

The third sense of '*banco*' supposedly presents the verbal use of this lemma. However, the sense 'to bank money' simply does not exist with the lemma *banco* in Portuguese. Now, ChatGPT not only 'created' this sense, it went as far as

inventing conjugations for *banco* to then use these conjugations as well (\**bancou*, \**bancam*), then continues with a convincing COBUILD-style full-sentence definition but forgets to use the lemma and instead uses the (correct) verbs *depositar* 'deposit' and *confiar* 'rely on', and finally does the same when providing the first example. Likewise for the phrasal verb 'count on', so [VERB + on], which also does not use *banco* in Portuguese, but rather *contar com* as seen in the second example of Figure 10c. When ChatGPT offers its list of synonyms, it merely repeats what it had used all along in the microstructure already. Note that there is a verb *bancar* 'fund, finance' in Portuguese, for which the morphological forms are *banca*, *bancou*, *bancando*, *banca*.<sup>5</sup> Conclusion: As was the case for the other dictionary direction, ChatGPT-3.5 is once more not up to the task, as it is fully unable to compile proper COBUILD-style entries for a Portuguese–English learners' dictionary.

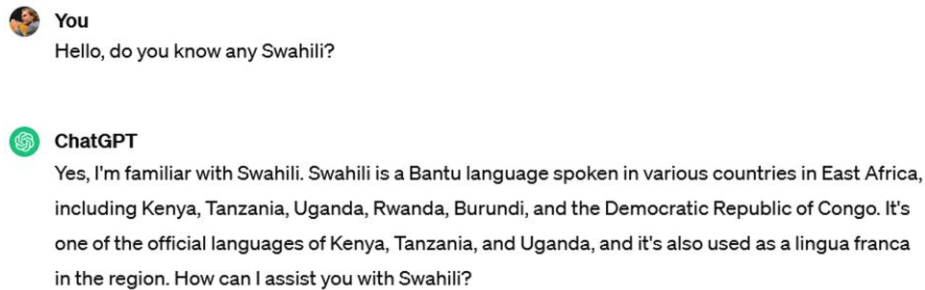
So, what is really going on, and what is ChatGPT-3.5 doing for bilingual lexicography? Well, the assumption is that one is actually looking at English lexical analysis through the eyes of a non-English language, here Portuguese. Reformulated, the hypothesis is that everything one sees there was created internally based on English (given the strong bias in the training data) and was merely mapped onto Portuguese since the prompts asked to go bilingual. If a further analogy is allowed: Given the overwhelming amount of English in an LLM, a model like ChatGPT-3.5 is simply incapable to 'think' in any other language, let alone analyse it linguistically in any proper way.

## 2.3 Low-visible Bantu

Back in 2023–2024, with the certainty that an LLM like ChatGPT is 'fantastic' for English, and the knowledge that it is basically 'fake' for bilingual lexicography, does that mean that it simply 'fails' for exotic languages? During a talk at the University of Missouri (De Schryver 2024b) various attempts were made at having ChatGPT compile proper dictionary articles for a variety of Bantu languages, amongst others Swahili (100 million speakers; spoken in more than 14 countries), Xhosa (19 million speakers; spoken in 2 countries) and Lusoga (4.5 million speakers; spoken in part of one country).

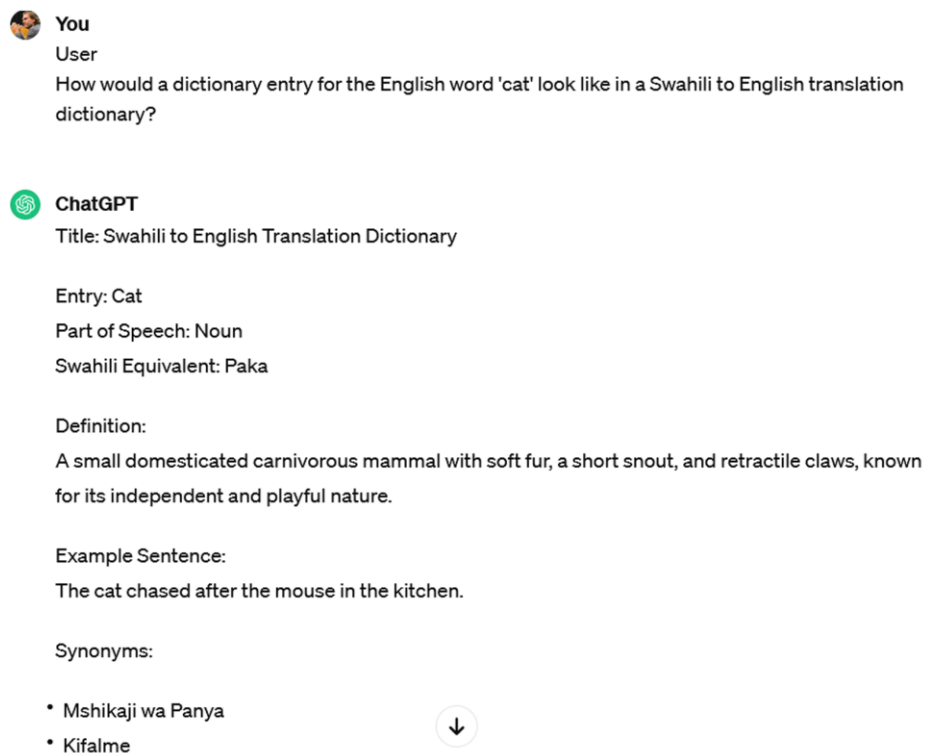
### 2.3.1 Region-visible Swahili

First, and again using ChatGPT-3.5, the chatbot was questioned about some of the basics of Swahili, to see whether it gets at least the essentials right for the language. The opening prompt and answer are displayed in Figure 11.



**Figure 11:** ChatGPT-3.5 opening prompt and answer on the Swahili language

While the answer is only approximately correct (apart from the army, no one uses Swahili in Uganda), the next question and answer immediately delve into lexicography for Swahili, as seen in Figure 12.



**Figure 12:** ChatGPT-3.5 prompt and output for an entry in a Swahili–English translation dictionary, here for *cat*

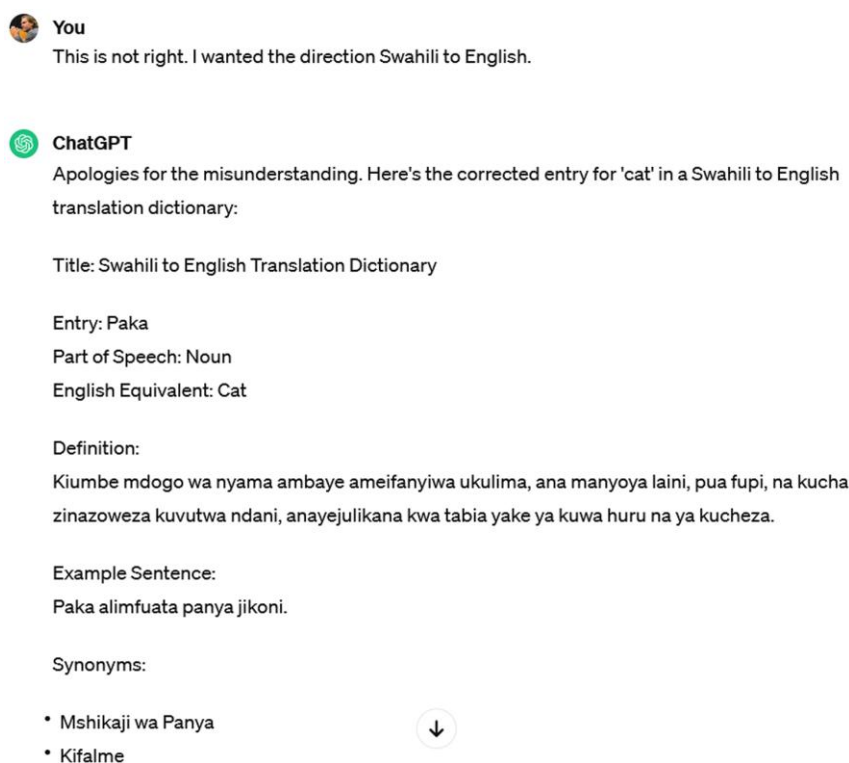
As may be seen from Figure 12, right from the moment one tries out lexicography in Swahili, the chatbot already gets the dictionary direction wrong. The

question was to go from Swahili to English, but the reply came in the reverse dictionary direction, with, strangely, only synonyms (of the English word *cat*) in Swahili. Now, in passing, it would be welcome if a chatbot could start taking over from an online Swahili dictionary, such as the unidirectional Swahili to English dictionary at [africanlanguages.com](http://africanlanguages.com)<sup>6</sup>, having been ranked by Google<sup>7</sup> as the top Swahili dictionary for two decades now, even though the year before its launch, there were already 20 online dictionaries for Swahili (De Schryver 2003: 12). See Addendum B.

In any case, the direction error was pointed out, and ChatGPT then offered the revised material displayed in Figure 13.

As may be seen from Figure 13, ChatGPT basically swapped the lemma and its translation equivalent around, and (correctly) translated the English to Swahili for the definition and example sentence. All of this, together with the POS, is correct, while the first synonym is original: *mshikaji wa panya* 'mouse catcher', while the second is not clear: *kifalme* 'princess'.

Given that this reply (mostly) made sense, the next prompt asked to see more *uses* and thus more *meanings*, see Figure 14. For more on the core concept of 'mapping *meaning* onto *use*' see Hanks (2002).



**Figure 13:** Additional ChatGPT-3.5 prompt and output for an entry in a Swahili–English translation dictionary, here for *paka* 'cat'





You

OK, makes sense. Does it have other uses and thus also other meanings?



ChatGPT

Yes, the word "paka" in Swahili primarily refers to a domestic cat, but it can also be used in other contexts or idiomatic expressions, giving rise to additional meanings or connotations. Here are a few examples:

1. **Verb Conjugation:** In Swahili, "paka" can also be used as a verb, meaning "to carry" or "to wear" when describing actions related to carrying something or wearing clothes. For example:
  - Nina paka begi langu. (I am carrying my bag.)
  - Anapaka kofia nyekundu. (He/she is wearing a red hat.)
2. **Idiomatic Expressions:** The word "paka" might be used in idiomatic expressions with meanings that extend beyond the literal translation. For instance:
  - Kuwa na paka mbili ni shida. (Literal translation: Having two cats is a problem.)
    - This expression might mean that having too many of something can lead to complications or difficulties.
3. **Metaphorical Use:** Sometimes "paka" can be used metaphorically to describe someone's behavior or characteristics, often implying slyness or cunning similar to the characteristics attributed to cats in various cultures.

Overall, while "paka" primarily means "cat" in Swahili, its usage can extend to various contexts and idiomatic expressions, allowing for different interpretations and meanings.

**Figure 14:** Second additional ChatGPT-3.5 prompt and output for an entry in a Swahili–English translation dictionary, here for *paka* 'cat'

As may be seen from Figure 14, this reply went horribly wrong. Instead of seeing an entry with multiple (numbered) senses, all the earlier information has been dispensed with. As for the three new numbered paragraphs, the first makes no sense at all, as it mixes up the noun *paka* 'cat' with the verb *-paka* 'smear, spread, apply' — which is anything but another sense of the noun *cat*; the second paragraph is an idiomatic expression with the noun *cat* — which is good; while the third paragraph talks about metaphorical uses in English, but there is not a single usage example in either language — so not useful. The wrap-up at the end is correct, but does not add anything new.

Courtesy obliges, one needs to end with the verb of communication *-pendekeza* 'recommend'. Figure 15 shows the usual prompt (which was also used for Portuguese in Figures 4 and 9), but now translated into Swahili, to obtain a COBUILD-style entry for a monolingual Swahili learners' dictionary. For the translation, here and further on, see Addendum C.



You

Nipe ingizo la kamusi ya Kiswahili-Kiingereza kwa lugha mbili kwa neno '-pendekeza'. Anza na maelezo ya marudio hapo juu. Kila maana inapaswa kuwa katika block iliyohesabiwa. Kisha kila kizuizi huanza na sehemu ya hotuba na maumbo ya kimofolojia ya maana husika. Hii inafuatwa na ufafanuzi wa maana katika Kiswahili, na tafsiri ya ufafanuzi huo kwa Kiingereza. Kumbuka kwamba ili kufafanua, inabidi utumie mtindo wa kufafanua sentensi kamili, ambao una vishazi viwili: kifungu cha kwanza kinaonyesha muundo wa matumizi, kifungu cha pili ni kifungu cha maana. Pia orodhesha tafsiri moja au zaidi zinazolingana katika Kiingereza katika kila maana. Ongeza mifano ya Kiswahili pamoja na tafsiri zao za Kiingereza zinazoonyesha matumizi na maana ya kila maana ya Kiswahili. Kwa sentensi za mfano, hakikisha unatumia miundo tofauti ya sentensi, ukirejelea watu tofauti; rejea hali zilizopita, za sasa, na zijazo; tofautisha sentensi ndefu na fupi za mifano; na ujumuishe maelezo mengine ili wewe, mpendwa ChatGPT, unitayarishie nyenzo za ubora wa juu za kamusi. Pia jumuishe matamshi ya ingizo kwa kutumia IPA, na ujumuishe orodha ya visawe na vinyume kwenye hisi ambapo orodha kama hizo zinafaa.

**Figure 15:** ChatGPT-3.5 prompt to obtain a COBUILD-style entry for a monolingual Swahili learners' dictionary, here for the verb *-pendekeza* 'recommend'

The reply by ChatGPT looked decidedly weird, as may be seen from Figure 16 and the explanation that follows.



ChatGPT

-pendekeza

*Marudio:* -pendekeza (Kitenzi)

#### 1. Kitenzi - Tendeka

*Maana:* Kutoa maoni au mapendekezo ya kufanya kitu fulani kwa sababu unaamini kuwa ni bora au inafaa kufanywa.

*Tafsiri:* recommend

*Visawe:* kupendekeza, kushauri

*Vinyume:* kukataa, kupinga

*Mifano:*

- *Daktari alimpendekeza mgonjwa apumzike zaidi.*  
(The doctor recommended that the patient rest more.)
- *Mwalimu alitupendekeza tufanye utafiti wa ziada kwa ajili ya mtihani ujao.*  
(The teacher recommended that we do extra research for the upcoming exam.)

**Matamshi:** /pende'keza/

**Visawe na Vinyume:**

- Visawe: kupendekeza, kushauri
- Vinyume: kukataa, kupinga



**Figure 16:** ChatGPT-3.5 output for the Swahili lemma *-pendekeza* 'recommend', for the prompt seen in Figure 15

From Figure 16, it is quickly clear that what ChatGPT-3.5 produces is of little value: The lemma is repeated twice, the part of speech (*kitenzi* 'verb') is repeated twice, the synonym and antonym slots are repeated twice, and the first so-called synonym simply takes the lemma and puts it into class 15, which is nothing special, as all verbs in any Bantu language belong to class 15, and thus start with *ku-*. The definition does not follow the COBUILD-style, and even so, turns the verb into the noun *mapendekezo* 'recommendation' to explain the verb *-pendekeza* 'recommend', which is bad lexicography. And in terms of layout, there is also no reason for the number 1, at the first and only *kitenzi* 'verb' block. Conclusion: ChatGPT-3.5 can simply not handle Swahili lexicography.

### 2.3.2 Nation-visible Xhosa

Given that Xhosa is a much smaller Bantu language compared to Swahili, basically only spoken in South Africa and recognized as a minority language in Zimbabwe, the approach this time was to try to make it super-simple, by merely asking for a small number of random short dictionary entries for Xhosa, as seen in Figure 17.



**You**

Please give me a short list of Xhosa dictionary entries.



**ChatGPT**

Certainly! Here are some common Xhosa dictionary entries:

1. "Molo" - Hello
2. "Enkosi" - Thank you
3. "Mhlekezzi" - Sir / Mr.
4. "Mhlekezzi" - Madam / Mrs.
5. "Unjani?" - How are you?
6. "Ndiphilile" - I am fine
7. "Ndiyavuya ukukwazi" - Nice to meet you

8. "Hamba kahle" - Goodbye
9. "Indlu" - House
10. "Inkosana" - Child
11. "Ibhokisi" - Box
12. "Iziko" - Museum
13. "Indawo" - Place
14. "Umntu" - Person
15. "Umntwana" - Child
16. "Umlilo" - Fire
17. "Inja" - Dog
18. "Inyoni" - Bird
19. "Igqirha" - Doctor
20. "Isango" - Door



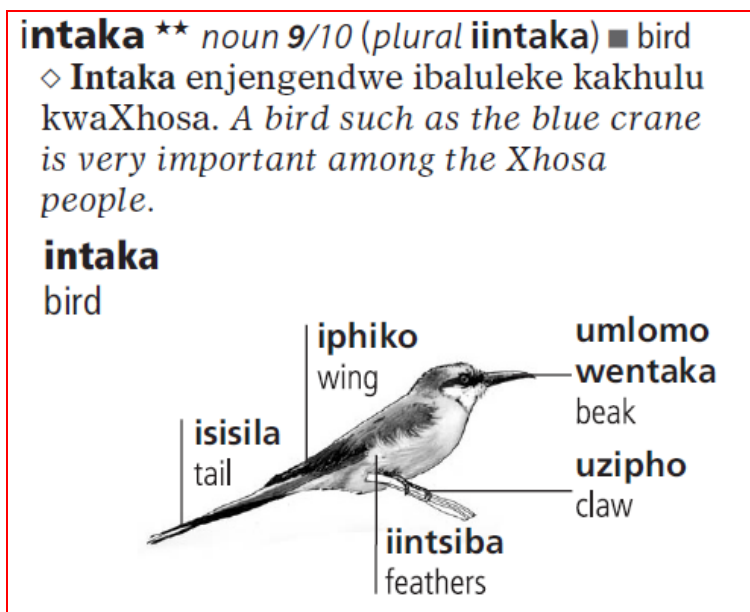
**Figure 17:** ChatGPT-3.5 prompt and reply to obtain a small number of random short dictionary entries for Xhosa

Figure 17 hardly displays a list of dictionary entries, only just a random vocabulary list. Unfortunately, it is abundantly clear that ChatGPT cannot even pick 20 frequent words from Xhosa, as item number 18, *inyoni* 'bird', is from the neighbouring and bigger language Zulu (27 million speakers; spoken in at least 6 countries), see Figure 18.

**inyoni** ★★ *noun* 9/10 (*pl. izinyoni*) ■ bird  
 ♦ Silalel' izinyoni zitshiloza. *We listen to the birds chirping.*  
 ► **kwelenyoni** ■ topmost part of a tree  
 ♦ Wathi lapho esephezulu kwelenyoni, wase elula isandla ukuze abambe igatsha elithele kakhulu izithelo. *When he was at the top of the tree, he stretched out his hand to grab a branch that was full of fruit.*

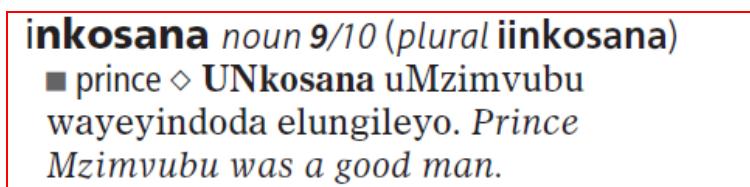
**Figure 18:** Dictionary entry for *inyoni* 'bird' in the *Oxford Zulu School Dictionary* (De Schryver 2010: 96)

Rather, the word for bird in Xhosa is *intaka*, see Figure 19.



**Figure 19:** Dictionary entry for *intaka* 'bird' in the *Oxford Xhosa School Dictionary* (De Schryver and Reynolds 2014: 89)

Even several translations are debatable: item 10, *inkosana*, is translated as 'child', but see Figure 20, where it is prince.



**Figure 20:** Dictionary entry for *inkosana* 'prince' in the *Oxford Xhosa School Dictionary* (De Schryver and Reynolds 2014: 87)

Item 12, *iziko*, is translated as 'museum', but see Figure 21, where it is '1. fireplace; 2. centre; institution'.

**iziko** ★★ *noun* 5/6 (plural **amaziko**)  
**1** ■ fireplace ◇ Lumka, **iziko** liseshushu. *Be careful, the fireplace is still hot.* **2** ■ centre; institution ◇ **IZiko** leeLwimi likhuthaza zonke iilwimi. *The Centre for Languages encourages the use of all languages.*

**Figure 21:** Dictionary entry for *iziko* '1. fireplace; 2. centre; institution' in the *Oxford Xhosa School Dictionary* (De Schryver and Reynolds 2014: 119)

While the translation for museum is actually a loanword, namely *imyuziyam*, see Figure 22.

**museum** ★ (say myoo-zee-uhm) *noun* (plural **museums**) ■ imyuziyam ◇ In this **museum** you can see many interesting things from Xhosa history. *Kule myuziyam uya kubona izinto ezininzi ezinika umdla ezivela kwimbali yamaXhosa.*

**Figure 22:** Dictionary entry for *museum* in the *Oxford Xhosa School Dictionary* (De Schryver and Reynolds 2014: 418)

Therefore, and as an overall conclusion so far, with ChatGPT-3.5 English not only does much better than other big languages (like Portuguese), but small languages (like Zulu) take over even smaller ones (like Xhosa).

### 2.3.3 Local-visible Lusoga

The next level, and still within Bantu, concerns what happens to even smaller languages of the primarily oral type, i.e. Bantu languages for which there is hardly any written documentation, no matter the actual number of sometimes impressively high number of speakers. Lusoga is such a case, spoken by up to 4.5 million speakers in the southeast of Uganda. The prompt was kept even more basic than any of the previous ones, just asking for the top frequent words in the language, as displayed in Figure 23.



**You**

User

Can you give me a list of top frequent words used in Lusoga?



**ChatGPT**

Lusoga is a Bantu language spoken in Uganda, primarily by the Basoga people. Here are some common Lusoga words:

1. "Wasuze otya" - How are you? (informal)
2. "Wasibye otya" - How are you? (formal)
3. "Wee" - Yes
4. "Nedda" - No
5. "Webale" - Thank you
6. "Nkikola" - I am doing
7. "Ndiga" - I am eating
8. "Lala bulungi" - Good night
9. "Oli wa?" - Who are you?
10. "Gyebale ko" - I am sorry
11. "Nsonyiwa" - Welcome
12. "Mugende gwe?" - Where are you going?
13. "Gwe nna?" - Who, me?
14. "Omukyala" - Woman
15. "Omuntu" - Person
16. "Obwavu" - Sickness
17. "Embeera" - Situation
18. "Obubonero" - Example
19. "Amagezi" - Knowledge
20. "Obusungu" - Wisdom

These are some common words in Lusoga that you may encounter in everyday conversations.

**Figure 23:** ChatGPT-3.5 prompt and reply to obtain a list of the frequent words in Lusoga

Firstly, this list has moved away from a list of words, to mostly a list of phrases. Secondly, an analysis of the ChatGPT-3.5 output by Minah Nabirye, a native speaker of Lusoga, reveals that only 7 out of the 20 are correct, with two of those correct ones actually only in the neighbouring and much bigger language Luganda (20 million speakers; spoken in Uganda immediately to the west of Lusoga, as well as in Rwanda). See Table 1 for the details.

**Table 1:** Analysis of the ChatGPT-3.5 output of so-called 'frequent words in Lusoga'

ChatGPT-3.5	ChatGPT-3.5 translation	Language	Status	Lug(anda)	Lus(oga)	English
1 <i>Wasuze otya</i>	How are you? (informal)	Lug/Lus	Wrong translation	-	-	Good morning
2 <i>Wasibye otya</i>	How are you? (formal)	Lug/Lus	Wrong form	<i>Osiibye otya?</i>	<i>Osiibye otya?</i>	-
3 <i>Wee</i>	Yes	??	Unknown form	-	-	-
4 <i>Nedda</i>	No	Lug	Correct	-	-	-
5 <i>Webale</i>	Thank you	Lug/Lus	Wrong form	<i>Weebale</i>	<i>Weebale</i>	-
6 <i>Nkikola</i>	I am doing	Lug	Correct	-	-	-
7 <i>Ndiga</i>	I am eating	Lug	Wrong translation	-	-	Sheep
8 <i>Lala bulungi</i>	Good night	Lug/Lus	Wrong form	<i>Sula bulungi</i>	<i>Sula bulungi</i>	-
9 <i>Oli wa?</i>	Who are you?	Lug/Lus	Wrong translation	-	-	Where are you?
10 <i>Gyebale ko</i>	I am sorry	Lug	Wrong translation	<i>Gyebaleko</i>	<i>Gyebaleku</i>	Thank you for your work
11 <i>Nsonyiwa</i>	Welcome	Lug	Wrong translation	-	<i>Nsonhiwa</i>	Forgive me
12 <i>Mugende gwe?</i>	Where are you going?	??	Unknown form	<i>Mugende (*gwe)</i>	<i>Mugye (*iwe)</i>	You go(pl) (gwe/iwe = you)
13 <i>Gwe nna?</i>	Who, me?	??	Unknown form	-	-	-
14 <i>Omukyala</i>	Woman	Lug/Lus	Correct	-	-	-
15 <i>Omuntu</i>	Person	Lug/Lus	Correct	-	-	-
16 <i>Obuwavu</i>	Sickness	Lug/Lus	Wrong translation	-	-	Poverty
17 <i>Embeera</i>	Situation	Lug/Lus	Correct	-	-	-
18 <i>Obubonero</i>	Example	Lug/Lus	Correct	-	-	-
19 <i>Amagezi</i>	Knowledge	Lug/Lus	Correct	-	-	-
20 <i>Obusungu</i>	Wisdom	Lug/Lus	Wrong translation	-	-	Anger

That means that all in all, there are merely 5 correct Lusoga forms out of 20, or thus one quarter (25%). Conclusion: If three quarters (75%) of the returned material is erroneous on the most basic of the basic lexicographic tasks, there is simply no point in trying anything else with ChatGPT-3.5 for Lusoga, even though it is spoken by a respective number of as many as 4.5 million speakers.

### 3. What to do next?

The main take-away from this article is that the lexicographic prompt introduced by De Schryver (2023b), being a single prompt with which to compile an entire English-style COBUILD learners' dictionary, may indeed work well for English, but is not good enough for other languages. At least, not using ChatGPT-3.5. The field of GenAI chatbots, however, has moved on since then, so a serious attempt must first be made to see how later out-of-the-box models like GPT-4o or Gemini-2.5 perform the job on other languages than English. Second, an attempt should be made to move from an out-of-the-box model, to one where one starts customising the model. These two aspects are precisely the topics of a follow-up article (see De Schryver 2025).

### PS.

In this day and age, and as the secretary of the exam committee for African Studies at UGent where the author of this publication is tasked with checking



all BAs, MAs and PhDs on their correct use of GenAI, the reader needs to be informed that GenAI was *not* used for any of the writings in this article, except for some of the sub-headings. Addendum D, available online (click [here](#)), displays the full chat between the author and OpenAI's GPT-4o in this respect. In itself, it provides an interesting perspective on using GenAI as a writing assistant. Research on using AI as a writing assistant in our field started with the work of Henrik Køhler Simonsen, first published as Leroyer and Simonsen (2019), and from then on, year after year, as Simonsen (2020, 2021, 2022, 2023). Subsequent to the release of ChatGPT, other colleagues followed (Huete-García and Tarp 2024, Tarp and Nomdedeu-Rull 2024).

Reading through the entire chat of Addendum D, available online (click [here](#)), reveals how good GPT-4o has become as a lexicographer; it is like consulting a wide variety of six different dictionaries, going in all sort of directions: (1) magnitudes, scales, sizes, (2) demographic speaker populations, (3) socio-technological reach, visibility, language power, (4) digital inclusion, corpus resources, language equity, (5) localness without grandeur, geographic nuance, judgmental rings, and (6) progression from hyper to local, capturing visibility and rank, balancing clarity, tone and structure. Simply stunning, and it is unlikely that any human lexicographer can do this in the fractions of seconds that GPT-4o spends on each of those levels. For the interested reader, even some screen recordings exist, from which the varied speed of the GenAI replies can be monitored. That said, one of the adjudicators observed all of the following, with which I fully agree:

That's not a fair point. It is, in my opinion overly simplistic and reductive to use generation speed as the sole criterion. What about training time or about the time it takes to critically check the responses? If one argues at this level, one could also point to the questionable use of resources in training and running current LLMs: not only does this speed come at the price of enormous resource consumption (electricity and water), but the language training material was also most likely acquired in violation of copyright laws. When praising the speed of LLM's responses, I believe it is important not to remain silent about these issues. (Anonymous adjudicator for this article, to whom sincere thanks are due.)

## Endnotes

1. See <https://www.collinsdictionary.com/dictionary/english/strike>, sense 3.
2. For more on the FSD, see Hanks (1987), Barnbrook (2002), Lew (2023), and Rundell (2024).
3. See <https://www.collinsdictionary.com/dictionary/english/bank>, finance and storage, sense 3.
4. See <https://www.collinsdictionary.com/dictionary/english/bank>, finance and storage, sense 4.
5. See <https://www.definify.com/word/bancar>.
6. See <https://africanlanguages.com/swahili/>.
7. See <https://www.google.com/>.

## References

- Barnbrook, G. 2002. *Defining Language: A Local Grammar of Definition Sentences* (Studies in Corpus Linguistics 11). Amsterdam: John Benjamins.
- Chomsky, N. 1957. *Syntactic Structures* (Janua Linguarum 4). The Hague: Mouton.
- De Schryver, G.-M. 2003. Online Dictionaries on the Internet: An Overview for the African Languages. *Lexikos* 13: 1-20.
- De Schryver, G.-M. 2010. *Oxford Bilingual School Dictionary: Zulu and English / Isichazamazwi Sesikole Esinezilimi Ezimbili: IsiZulu NesiNgesi, Esishicilelwe abakwa-Oxford*. Cape Town: Oxford University Press Southern Africa.
- De Schryver, G.-M. 2023a. *Contemporary Lexicography: A Case Study in AI [Keynote Lecture]*. Paper presented at the Inaugural Conference of the Association of Lexicography for the Americas: South, Central, Caribbean, and Mexico (Americalex-S), São Paulo, Brazil, 20–25 October 2023. <https://youtu.be/watch?v=yAxzTx7A2LU>
- De Schryver, G.-M. 2023b. Generative AI and Lexicography: The Current state of the Art Using ChatGPT. *International Journal of Lexicography* 36(4): 355-387 + ten pages of supplementary material online.
- De Schryver, G.-M. 2024a. *Customising LLMs for Lexicography [Keynote Lecture]*. Paper presented at the 8th International Workshop on Cognitive Aspects of the Lexicon (CogALex-VIII), co-located with the Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), Turin, Italy, 20 May 2024.
- De Schryver, G.-M. 2024b. *Generative Artificial Intelligence & Three Lexicographic Fs: FANTASTIC for Monolingual English Dictionaries, FAKE for Translation Dictionaries, FAIL for Exotic Dictionaries [Invited Talk]*. Paper presented at the Department of English, University of Missouri, Columbia, MO, USA, 5 April 2024. [https://youtu.be/zUI0GrPrH\\_0](https://youtu.be/zUI0GrPrH_0)
- De Schryver, G.-M. 2024c. The Road towards Fine-tuned LLMs for Lexicography. Krek, S. (Ed.). 2024. *Book of Abstracts of the Workshop 'Large Language Models and Lexicography' @ EURALEX 2024, 8 October 2024, Cavtat, Croatia*: 6-11. Ljubljana: ELEXIS Association.
- De Schryver, G.-M. 2025. Customised GPTs for Lexicography. *Lexikos* 35(2): 397-422.
- De Schryver, G.-M. forthcoming-a. 100 Scholarly Echoes of Generative AI in Lexicography.
- De Schryver, G.-M. forthcoming-b. Citing the Future: The Legacy of Patrick Hanks.
- De Schryver, G.-M. and M. Reynolds. 2014. *Oxford Bilingual School Dictionary: IsiXhosa and English / Oxford isiXhosa-isiNgesi English-isiXhosa Isichazi-magama Sesikolo*. Cape Town: Oxford University Press Southern Africa.
- Fuertes-Olivera, P.A. 2024. Making Lexicography Sustainable: Using ChatGPT and Reusing Data for Lexicographic Purposes. *Lexikos* 34(1): 123-140.
- Hanks, P. 1987. Definitions and Explanations. Sinclair, J.M. (Ed.). 1987. *Looking Up: An Account of the COBUILD Project in Lexical Computing and the Development of the Collins COBUILD English Language Dictionary*: 116-136. London: Collins ELT.
- Hanks, P. 2000. Do Word Meanings Exist? *Computers and the Humanities* 34(1–2): 205-215.
- Hanks, P. 2002. Mapping Meaning onto Use. Corréard, M.-H. (Ed.). 2002. *Lexicography and Natural Language Processing. A Festschrift in Honour of B.T.S. Atkins*: 156-198. s.l.: Euralex.
- Hanks, P. 2013. *Lexical Analysis: Norms and Exploitations*. Cambridge, MA: The MIT Press.

- 
- Huete-García, Á. and S. Tarp. 2024. Training an AI-based Writing Assistant for Spanish Learners: The Usefulness of Chatbots and the Indispensability of Human-assisted Intelligence. *Lexikos* 34: 21-40.
- Jakubíček, M. and M. Rundell. 2023. The End of Lexicography? Can ChatGPT Outperform Current Tools for Post-editing Lexicography? Medved', M., M. Měchura, C. Tiberius, I. Kosem, J. Kallas, M. Jakubíček and S. Krek (Eds.). 2023. *Proceedings of the eLex 2023 Conference: Electronic Lexicography in the 21st Century*: 518-533. Brno: Lexical Computing.
- Leroyer, P. and H.K. Simonsen. 2019. Google Translate: trussel eller redning for oversættelsesordbøger? [Google Translate: Threat or Salvation for Translation Dictionaries?]. *LexicoNordica* 26: 95-115.
- Lew, R. 2023. ChatGPT as a COBUILD Lexicographer. *Humanities and Social Sciences Communications* 10(704): 10 pp.
- Lew, R. 2024. Dictionaries and Lexicography in the AI Era. *Humanities and Social Sciences Communications* 11(426): 8 pp.
- Lew, R., B. Ptasznik and S. Wolfer. 2024. The Effectiveness of ChatGPT as a Lexical Tool for English, Compared with a Bilingual Dictionary and a Monolingual Learner's Dictionary. *Humanities and Social Sciences Communications* 11(1324): 10 pp.
- McKean, E. and W. Fitzgerald. 2023. The ROI of AI in Lexicography [Keynote Lecture]. 2023. *Proceedings of the 16th International Conference of the Asian Association for Lexicography: "Lexicography, Artificial Intelligence, and Dictionary Users"*: 18-27. Seoul: Yonsei University.
- Nichols, W. 2023. Invisible Lexicographers, AI, and the Future of the Dictionary [Keynote Lecture]. Medved', M., M. Měchura, C. Tiberius, I. Kosem, J. Kallas, M. Jakubíček and S. Krek (Eds.). 2023. *Book of Abstracts of the eLex 2023 Conference: Electronic Lexicography in the 21st Century*: 3-3. Brno: Lexical Computing.
- Pavydytė, K. 2024. *ChatGPT generuojami iliustraciniai laukai: turinio ir kokybės analizė bei palyginimas su BLKŽ [Illustrative Sentences Generated by ChatGPT: The Analysis of Its Content, Quality, and Comparison with the Dictionary of the Standard Lithuanian Language]*. MA Dissertation. Vilnius: Vilnius University.
- Ptasznik, B., S. Wolfer and R. Lew. 2024. A Learners' Dictionary Versus ChatGPT in Receptive and Productive Lexical Tasks. *International Journal of Lexicography* 37(3): 322-336.
- Rees, G.P. and R. Lew. 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.
- Rundell, M. 2024. Obituary: Patrick Wyndham Hanks. *International Journal of Lexicography* 37(2): 135-142.
- Simonsen, H.K. 2020. Augmented Writing: Nye muligheder og nye teorier [Augmented Writing: New Possibilities and New Theories]. Sandström, C., U.-M. Forsberg, C. af Hällström-Reijonen, M. Lehtonen and K. Ruppel (Eds.). 2020. *Nordiska studier i lexikografi 15, Rapport från 15 konferensen om lexikografi i Norden, Helsingfors, 4-7 juni 2019 [Nordic Studies in Lexicography 15, Report from the 15th Nordic Lexicography Conference, Helsinki, 4-7 June 2019]*: 307-315. Helsinki: Nordic Association for Lexicography.
- Simonsen, H.K. 2021. AI Writers in Language Learning. Chang, M., N.-S. Chen, D.G. Sampson and A. Tlili (Eds.). 2021. *Proceedings of the 21st International Conference on Advanced Learning Technologies (ICALT 2021), Online, 12-15 July 2021*: 238-240. Los Alamitos, CA: IEEE.

- Simonsen, H.K.** 2022. AI Text Generators and Text Producers. Chang, M., N.-S. Chen, M. Dascalu, D.G. Sampson, A. Tlili and S. Trausan-Matu (Eds.). 2022. *Proceedings of the 22nd International Conference on Advanced Learning Technologies (ICALT 2022), Bucharest, Romania, 1–4 July 2022*: 218-220. Los Alamitos, CA: IEEE.
- Simonsen, H.K.** 2023. Bilingvæle ordbøger vs. AI-skriveassistenter? [Bilingual Dictionaries vs. AI Writing Assistants?]. *LexicoNordica* 30: 117-136.
- Tarp, S. and A. Nomdedeu-Rull.** 2024. Who Has the Last Word? Lessons from Using ChatGPT to Develop an AI-based Spanish Writing Assistant. *Círculo de Lingüística Aplicada a la Comunicación [Circle of Linguistics Applied to Communication]* 97: 309-321.
- Zhong, A.** 2025. Prompts for Language Learners: A Practical Guide to Using DeepSeek as a Dictionary. *Lexikos* 35(1): 274-285.

**Addendum A:** Using ChatGPT-3.5 — For Portuguese Lexicography [translations into English]

- Translation of Figure 1:

Please give me a COBUILD-style monolingual dictionary entry for 'strike'. Start with the frequency information at the top. Each sense should be in a numbered block. Each block then starts with the part of speech and the morphological forms of the respective sense. This is followed by a sense definition and sense examples that illustrate both the use and the meaning of each particular sense. For the example sentences, make sure to use different sentence structures, referring to different people; refer to past, present, and future situations; vary long and short example sentences; and include other elaborations so that you, dear ChatGPT, produce high-quality dictionary material for me.

- Translation of Figure 3:

Thanks, but actually I wanted to see a monolingual dictionary entry for the Portuguese entry '*greve*' (which is 'strike' in English).

I apologize for the confusion earlier. Here is a monolingual dictionary entry for the word '*greve*' in Portuguese:

*greve* (feminine noun)

**Frequency:** The word *strike* is among the 10,000 most frequent words in Portuguese.

**Definition:** A *strike* is a stoppage organized by workers in which they stop working as a form of protest, usually to demand better working conditions, higher wages, or other rights from employers.

**Examples:**

1. Factory workers went on *strike* to demand a pay rise.
2. The teachers decided to go on *strike* because of the lack of investment in education.
3. The truckers' *strike* caused supply shortages across the country.

**Note:** The word 'strike' is often used in the context of labour protests and is a right guaranteed by labour laws in many countries. It can be carried out by various professional sectors, including teachers, doctors, drivers, and others, as a means of putting pressure on employers or governments to meet workers' demands.

- Translation of Figure 4:

Please give me a COBUILD-style monolingual dictionary entry for '*recomendar*'. Start with the frequency information at the top. Each sense should be in a numbered block. Each block then starts with the part of speech

and the morphological forms of the respective sense. This is followed by a sense definition and sense examples that illustrate both the use and the meaning of each particular sense. For the example sentences, make sure to use different sentence structures, referring to different people; refer to past, present, and future situations; vary long and short example sentences; and include other elaborations so that you, dear ChatGPT, produce high-quality dictionary material for me. Also include a pronunciation of the entry using IPA, and include a list of synonyms and antonyms at the senses where such lists are relevant.

Note that a final line was added to the prompt, as was done at the end of the initial study as well (De Schryver 2023b); highlighted in grey above.

- Translation of Figure 5a:

*recomendar* (verb)

**Frequency:** The word *recomendar* is among the 2,000 most frequent words in Portuguese.

**Sense 1:** *To express a favourable opinion about something or someone; to advise or suggest something to someone*

**Morphological forms:** recommend, recommends, recommending, recommended

**Definition:** When you recommend something, you express a positive opinion about it, often by advising or suggesting that someone else do it.

**Examples:**

1. I *recommend* this restaurant; the food is delicious.
2. She *recommended* a good book for me to read on vacation.
3. They are *recommending* that everyone wears masks to protect against the virus.

**Pronunciation:** /ʁe.ko.mẽ.'daʁ/

**Synonyms (Sense 1):** advise, suggest, indicate, endorse, approve

**Antonyms (Sense 1):** advise against, disapprove of, discourage

- Translation of Figure 5b:

**Sense 2:** *Presenting someone as fit or suitable for something, such as a job or position*

**Morphological forms:** recommend, recommends, recommending, recommended

**Definition:** In this sense, to recommend means to present someone as qualified for a job or position, indicating their skills, qualities or achievements.

**Examples:**

1. He was *recommended* for the position of manager because of his experience and leadership skills.
2. The professor *recommended* several excellent students for scholarships at the university.
3. The letter of recommendation *recommended* the candidate as highly competent for the job.

**Pronunciation:** /ʁe.ko.mẽ.'daʁ/

**Synonyms (Sense 2):** indicate, endorse, refer, endorse, sponsor

**Antonyms (Sense 2):** disapprove, disrecommend, criticize, censure

- Translation of Figure 5c:

**Sense 3:** *To prescribe or suggest something as a treatment or solution*

*Morphological forms:* recommend, recommends, recommending, recommended

*Definition:* In medical or professional terms, recommending involves prescribing or suggesting a treatment, course of action or solution for a specific condition or problem.

*Examples:*

1. The doctor *recommended* rest and medication to treat the flu.
2. The specialist *recommended* a balanced diet and exercise to improve the patient's health.
3. She *recommended* that he consult a lawyer to resolve the legal issue.

**Pronunciation:** /ʁe.ko.mẽ.'daʁ/

**Synonyms (Sense 3):** prescribe, advise, guide, suggest, indicate

**Antonyms (Sense 3):** advise against, disallow, disapprove of, discourage

- Translation of Figure 9:

Give me a bilingual Portuguese–English dictionary entry for the word 'banco'. Start with the frequency information at the top. Each sense should be in a numbered block. Each block then starts with the part of speech and the morphological forms of the respective sense. This is followed by a sense definition in Portuguese, and translation of that definition into English. Note that to define, you have to use the full-sentence defining style, which consists of two clauses: the first clause illustrates the usage pattern, the second clause is a paraphrase of meaning. Also list one or more translation equivalents in English at each sense. Add Portuguese examples with their English translations that illustrate both the use and the meaning of each particular Portuguese sense. For the example sentences, make sure to use different sentence structures, referring to different people; refer to past, present, and future situations; vary long and short example sentences; and include other elaborations so that you, dear ChatGPT, produce high-quality dictionary material for me. Also include a pronunciation of the entry using IPA, and include a list of synonyms and antonyms at the senses where such lists are relevant.

The grey highlight added on top of an extra sentence in the prompt here, marks the attempt to make sure a COBUILD-style definition, so a full-sentence definition (FSD), is obtained by ChatGPT-3.5.

- Translation of Figure 10a:

**bank** ['bẽku]

Frequency: \*\*\*\*\*

**1. Noun (Financial Institution)**

- **Morphological forms:** banks (plural)
- **Sense:** A financial institution where people and companies can deposit money, obtain loans and carry out various financial transactions.
  - *Usage Pattern:* A bank is a place where people and companies can deposit money, obtain loans and carry out financial transactions.
  - *Definition:* A bank is an establishment where people and businesses can put their money, get loans and manage various financial transactions.
- **English Translation Equivalents:** Bank
  - *Example 1 in Portuguese:* I need to go to the bank to withdraw some money.
    - *Translation 1 in English:* I need to go to the bank to withdraw some money.
  - *Example 2 in Portuguese:* The bank offers various financial services, including savings accounts and mortgages.
    - *Translation 2 in English:* The bank offers various financial services, including savings accounts and mortgages.
- **Pronunciation:** ['bẽku]
- **Synonyms:** financial institution, bank branch
- **Antonyms:** customer, borrower

- Translation of Figure 10b:

**2. Noun (Bank or Sloping Area)**

- **Morphological forms:** banks (plural)
- **Sense:** The side of a river, lake or other body of water; often referring to the raised area or the margin.
  - *Usage Pattern:* The margin of a river is the side of the river.
  - *Definition:* The margin of a river, for example, denotes the land next to the river.
- **Translation Equivalents in English:** Bank, Riverside
  - *Example 1 in Portuguese:* They decided to have a picnic on the margin of the river.
    - *Translation 1 in English:* They decided to have a picnic on the bank of the river.
  - *Example 2 in Portuguese:* The boat is anchored near the margin of the lake.
    - *Translation 2 in English:* The boat is anchored near the bank of the lake.
- **Pronunciation:** ['bẽku]
- **Synonyms:** margin, edge
- **Antonyms:** centre, middle



- Translation of Figure 10c:

### 3. Verb

- **Morphological forms:** \*banco (3rd person singular present tense), \*bancou (past tense), \*bancando (present participle), \*banco (3rd person singular present tense)
- **Sense:** To deposit or store something in a bank, or to entrust or store something for future use.
  - *Usage Pattern:* Someone \*banks something in a bank, or they \*bank something for a specific purpose.
  - *Definition:* When someone deposits something in a bank, they are storing it in a financial institution or entrusting it for future use.
- **English Translation Equivalents:** Deposit, Rely on
  - *Example 1 in Portuguese:* She deposits her savings in a local bank.
    - *Translation 1 in English:* She banks her savings in a local bank.
  - *Example 2 in Portuguese:* He's counting on you winning the competition with your skills.
    - *Translation 2 in English:* He's banking on winning the competition with his skills.
- **Pronunciation:** ['bẽku]
- **Synonyms:** deposit, entrust
- **Antonyms:** withdraw, discredit

Note that the stars (\*) were added to this translation, to mark (some of) the various errors in Portuguese. See the main text for the explanation.

## Addendum B: Online Swahili–English Dictionary / Kamusi ya Kiswahili–Kiingereza

- Without support, the following online Swahili–English dictionary — having been free and top-ranked by Google<sup>7</sup> for over two decades — will have to be taken down (<https://africanlanguages.com/swahili/>).

The screenshot shows a web browser window with the URL <https://africanlanguages.com/swahili/>. A pink notification box at the top states: "Dear Swahili dictionary user, Unfortunately, we are out of funds for hosting costs for this site and urgently need donations to keep it running. Please consider donating at [PayPal.me/davidjoffe](https://PayPal.me/davidjoffe). Thank you! If you would like to sponsor this site (for example, with a retail advertising banner here), [Contact Us](#)". Below this, a grey banner mentions a new article on "Generative AI and Lexicography" by Song Trein na Pretoria, with a link to "Creative Gardening with Indigenous Plants". A yellow box contains a PayPal donation form with the text: "Please donate now with Payfast to help keep this Swahili dictionary online and going - thank you. - David Joffe". The form has input fields for "Amount (Rands)" set to 80 (equivalent to R80 or ~\$4.30) and "Quantity" set to 1, with a "Pay Now" button. Below the form, text says "This dictionary needs support to help pay our costs, please help support this site:" followed by "PayPal donations" and "Bitcoin donations" buttons. A blue header reads "Online Swahili - English Dictionary" with a language switcher for "Kiswahili | English". A search bar contains "search / tafuta" and a "Search" button. A welcome message says "Welcome to [africanlanguages.com/swahili/](https://africanlanguages.com/swahili/) (now more mobile-friendly! please give the mobile version a try)". A link to "contact us" is provided for advertising. A grey box offers a "downloadable Windows version" of the dictionary. A news item from 13 Oct 2014 states: "This dictionary now also returns results from the Microsoft Swahili terminology list, below the regular results!"

← → ↺ <https://africanlanguages.com/swahili/> 90% ☆ 🔍 Search

Dear Swahili dictionary user,  
Unfortunately, we are out of funds for hosting costs for this site and urgently need donations to keep it running.  
Please consider donating at [PayPal.me/davidjoffe](https://PayPal.me/davidjoffe). Thank you!  
If you would like to sponsor this site (for example, with a retail advertising banner here), [Contact Us](#)

New 3 Oct 2023: New article on [Generative AI and Lexicography](#)  
[Song Trein na Pretoria](#)  
[Creative Gardening with Indigenous Plants](#)

Please donate now with Payfast to help keep this Swahili dictionary online and going - thank you. - David Joffe  
Amount (Rands)  (R80 = ~\$4.30)  
Quantity:   
[Pay Now](#)

This dictionary needs support to help pay our costs, please help support this site: [PayPal donations](#) [Bitcoin donations](#)

**Online Swahili - English Dictionary**  
Change page language: [Kiswahili](#) | [English](#)

[Search](#)

Welcome to [africanlanguages.com/swahili/](https://africanlanguages.com/swahili/) (now more mobile-friendly! please give the mobile version a try)  
If you would like to advertise on this page, please [contact us](#).

[Click here to try or buy a downloadable Windows version of this dictionary »](#)  
(Sales help support this dictionary)

**News 13 Oct 2014:** This dictionary now also returns results from the Microsoft Swahili terminology list, below the regular results!

**Addendum C:** Using ChatGPT-3.5 — For Swahili Lexicography [here the translations into English]

- Translation of Figure 15:

Give me a bilingual Swahili–English dictionary entry for the word '-pendekeza'. Start with the frequency information at the top. Each sense should be in a numbered block. Each block then starts with the part of speech and the morphological forms of the respective sense. This is followed by a sense definition in Swahili, and translation of that definition into English. Note that to define, you have to use the full-sentence defining style, which consists of two clauses: the first clause illustrates the usage pattern, the second clause is a paraphrase of meaning. Also list one or more translation equivalents in English at each sense. Add Swahili examples with their English translations that illustrate both the use and the meaning of each particular Swahili sense. For the example sentences, make sure to use different sentence structures, referring to different people; refer to past, present, and future situations; vary long and short example sentences; and include other elaborations so that you, dear ChatGPT, produce high-quality dictionary material for me. Also include a pronunciation of the entry using IPA, and include a list of synonyms and antonyms at the senses where such lists are relevant.

- Translation of Figure 16:

**-pendekeza**

*Repeat:* -pendekeza (Verb)

**1. Verb - To do**

*Meaning:* To give an opinion or recommendation to do something because you believe it is good or appropriate to do it.

*Translation:* recommend

*Synonyms:* to recommend, advise

*Antonyms:* to reject, to oppose

*Examples:*

- *The doctor recommended that the patient rest more.*  
(The doctor recommended that the patient rest more.)
- *The teacher recommended that we do extra research for the upcoming exam.*  
(The teacher recommended that we do extra research for the upcoming exam.)

**Pronunciation:** /pende'keza/

**Synonyms and Antonyms:**

- Synonyms: to recommend, advise
- Antonyms: to reject, oppose