



**HAL**  
open science

## AI Empowerment: where are we in the automation of lexicography? A metaphraseographic study

Lian Chen, Huy Linh Dao, Danh-Thàn Do-Hurinville

### ► To cite this version:

Lian Chen, Huy Linh Dao, Danh-Thàn Do-Hurinville. AI Empowerment: where are we in the automation of lexicography? A metaphraseographic study. ASIALEX 2024 - The 17th International Asian Association for Lexicography Conference, Ai INOUE, Sep 2024, Tokyo, Japan. pp.90-98. hal-04664242v1

**HAL Id: hal-04664242**

**<https://hal.science/hal-04664242v1>**

Submitted on 3 Sep 2024 (v1), last revised 16 Sep 2024 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ISBN 978-4-9901771-2-6

# AsiaLex 2024 Proceedings



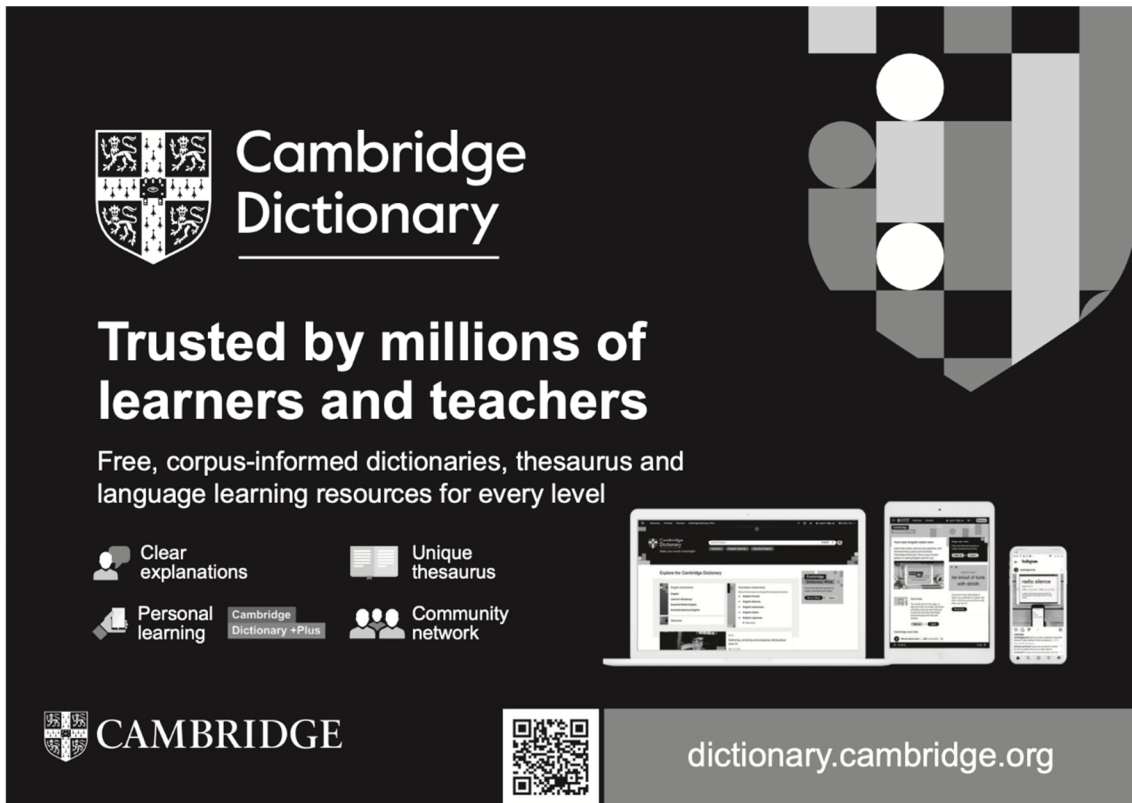
**Asian Lexicography -  
Merging cutting-edge and  
established approaches**


Edited by  
**Ai Inoue, Naho Kawamoto,  
and Makoto Sumiyoshi**

**AsiaLex  
2024**  
**TOKYO**








## Sponsors





 **Cambridge  
Dictionary**


**Trusted by millions of  
learners and teachers**

Free, corpus-informed dictionaries, thesaurus and  
language learning resources for every level

-  Clear explanations
-  Unique thesaurus
-  Personal learning
- 
-  Community network



 **CAMBRIDGE**



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

## 小学館



## **Acknowledgments**

### **Host:**

東洋大学 (Toyo University)

### **Convener:**

Ai Inoue, PhD./Professor, Faculty of Economics, Department of Economics, Toyo University, Vice-President of AsiaLex

### **Local organizers:**

Kaoru Akasu, Toyo University

Priscilla Ishida, Meiji University

Naho Kawamoto, Tokyo University of Foreign Studies

Yuri Komuro, Chuo University

Minoru Moriguchi, Kyoto University of Foreign Studies

Makoto Sumiyoshi, Kwansai Gakuin University

Yukio Tono, Tokyo University of Foreign Studies

Shigeru Yamada, Waseda University

### **Scientific committee:**

Yelee An, Academy of Korean Studies, South Korea

Amy Chi, The Hong Kong University of Science and Technology, China

Yongwei Gao, Fudan University, China

Hae-Yun Jung, Kyungpook National University, South Korea

Kilim Nam, Yonsei University, South Korea

Vincent Ooi, National University of Singapore, Singapore

Saghar Sharifi, Karaj Islamic Azad University, Iran

Hai Xu, Guangdong University of Foreign Studies, China

Cuilian Zhao, Sichuan International Studies University, China

### **Young Scholars/Graduate Student Committee:**

Takayoshi Denawa, Lecturer, SBC Tokyo Medical University

Yuusuke Kon, Assistant Lecturer, Tokyo Kasei University

Daiki Miura, Graduate School of Toyo University

## FOREWORD

The 17th International Asian Association for Lexicography Conference (hereinafter, AsiaLex 2024) will be held from September 12 to 14, 2024, at the Hakusan Campus, Toyo University, Tokyo, Japan. As the third conference to be held in Japan (the first being in 2003 at Meikai University in Tokyo and the second being in 2022 at Kyoto Terrsa in Kyoto), AsiaLex 2024 aims to discuss changes and challenges beyond traditional lexicography and pursue new directions in lexicography and dictionary use to address current global and local social changes.

The main theme of the conference is “**Asian Lexicography-Merging Cutting-Edge and Established Approaches.**” Therefore, submissions on relevant topics in lexicography and other themes related to dictionary use will be welcomed. The proceedings will include papers by individual researchers that are useful for staying current with modern trends in Asian and global lexicography, as well as papers by keynote speakers and LexTeach (lexicographical teach?).

Previous AsiaLex conference themes, which are available on the ASIALEX website (<https://www.asialex.org/>), brought numerous lexicographical studies using innovative technologies to the forefront. The 2020 ASIALEX conference was canceled due to the COVID-19 pandemic. However, the 2021 and 2022 conferences were held online, taking advantage of such technological developments. The use of technology in research was extended and encouraged due to the remarkable changes in the work and lives of people in the pandemic period. Nevertheless, lexicographers still utilize conventional methods to determine the use of these technologically enhanced research results in dictionary development. Understanding how innovative research can be combined with outdated methods has become an important concern for lexicographers. Therefore, AsiaLex 2024 will provide an opportunity for AsiaLex members and participants to discuss ways of combining these theories and practices.

Similar to other Asian countries, Japan has a long, active, and unique English-Japanese dictionary (EJD) history. The following briefly examines the history of English lexicography in Japan. The evolution of EJDs can be classified into six periods:

- (1) a. until 1861: the atavistic period, when English vocabulary was being collected;
- b. from 1862 to 1910: the period wherein internationally published dictionaries were translated into Japanese;
- c. from 1911 to 1926: the period wherein EJDs were compiled and published by a single lexicographer;
- d. from 1927 to 1966: the primary period during which co-edited English-Japanese dictionaries for learners (EJDLs) were published;
- e. from 1967 to present: the second period during which a large number of co-edited EJDLs were published;
- f. from 2000 to present: unabridged EJDs were published. (Original in Japanese; See detail in Yagi 2006: 12ff.)

As stated by Cowie (1999), well-known educators and lexicographers in the 1920s, namely, A.S. Hornby and H. E. Palmer, actively provided English education for Japanese English learners in Japan. They published the *Second Interim Report on English Collocations* in 1933, which used the dictionaries of Saito Hidesaburo, especially *Saito's Idiomatic English-Japanese Dictionary* in 1915, as a reference. Notably, *idiomological*, a word that was coined by Saito, is synonymous to “phraseological,” that is, Saito's *idiomology* is equivalent to phraseology.

Similar to EJDs, Japanese-language dictionaries have a long and rich history. For example, 『新字』 (Niina) was believed to have been published in premodern times in 682 and 『篆隸万象名義』 (Tenrei Bansyou Meigi) was the oldest published Japanese-language dictionary before 835. 『言海』 (Genkai) (1889-1891) was the first modern published Japanese-language dictionary, which was followed by 『大言海』 (Dai Genkai) and 『新言海』 (Shin Genkai), all of which have substantially influenced consequent Japanese-language dictionary publications. Owing to the efforts of these predecessors, the interweaving of different research perspectives facilitated the continuous evolution of EJDs and Japanese-language dictionaries. This evolution can be attributed to the cultural and linguistic diversity in Asia, which has allowed lexicographical researchers and practitioners to conduct research from a wide range of perspectives. Similar to other focused lexicographical societies, such as EURALEX (European Association for Lexicography), AFILEX (African Association for

Lexicography), AUSTRALEX (Australian Association for Lexicography), and the Dictionary Society of North America, Asian research achievements presented at previous ASIALEX conferences emphasize the uniqueness and the *raison d'être* of ASIALEX.

AsiaLex 2024 provides a unique platform for discussions and exchanges and promotes interdisciplinary and integrated lexicographical research owing to its continuous valuable contributions to the Asian and global lexicographical fields. Therefore, similar to previous conferences, AsiaLex 2024 intends to facilitate interaction among like-minded colleagues and lexophiles worldwide.

Ai Inoue, PhD./ Professor  
The Vice-President of ASIALEX, the convenor of AsiaLex 2024  
Faculty of Economics, Department of Economics, Toyo University, Tokyo, JAPAN.

### References

- Cowie, A. P. 1999. *English Dictionaries for Foreign Learners: A History*. Oxford: Clarendon Press.
- Yagi, K. 1999. *Eiwa Jiten no Kenkyuu – Eigo Ninshiki Kaizen no tameni (Research on English-Japanese Dictionaries – To Improve the Perception of English)*. Tokyo: Kaitakusha.

# TABLE OF CONTENTS

## PART 1 Keynote speakers

- Hajime Kimura. Lexicographical sources used by modern Japanese dictionaries ..... 2
- Yuri Komuro. Exploring the golden age of English-Japanese learners' dictionaries with a focus  
on interactions among lexicographers ..... 7

## PART 2 LexTeach

- Yuri Komuro. The power of dictionaries: How dictionaries can help you with precision and depth  
in expressions..... 16
- Vinent By Ooi. Lexicographic evidence in dictionary-making. .... 17
- Shigeru Yamada. Teaching of the use of EFL dictionaries with focus on definition. .... 24

## PART 3 Talks

### Topic 1 Bilingual/ Multilingual Dictionaries

- Yongwei Gao. Challenges and opportunities for present-day English-Chinese lexicography. .... 32
- Akira Ikegami. A cognitive approach to making a bilingual maritime English dictionary. .... 38
- Ai Inoue. Introduction to a newly published English-Japanese phraseological dictionary  
and its potential application among Japanese English learners ..... 45
- Minsoo Kwon. To learn speaking without mentioning: Dealing with euphemistic idiomatic phrases  
in Basic Korean Dictionary. .... 55
- Setsuko Yamagata and Keisuke Tabata. The visibility and descriptive scope  
of English-Japanese dictionaries. .... 61

### Topic 2 Digital Lexicography

- Waleed AbuMandour. Empowering bilingual lexicography with AI:  
The role of AI-model Triangulation Approach (AMTA) in dictionary design. .... 70
- Marek Blahuš, Michal Cukr, Miloš Jakubiček, Vojtěch Kovář and František Kovařík. Dictionary Express  
: First phases. Rapid dictionary-making method for European, Asian and other languages. .... 84
- Lian Chen, Huy-Linh Dao and Danh-Thành Do-Hurinville. AI empowerment  
: Where are we in the automation of lexicography? A metaphraseographic study. .... 90
- Minoru Moriguchi and Olivia Kennedy. Generative AI's non-verifiable interpretation  
in exploring Japanese *youkai* in English. .... 99
- Piao Xian, Eunha Hwang and Jingri Cui. Online fieldwork for dialect speech community  
in mobility: Focusing on dictionary compilation of *Yanbian Malmo I*. .... 106
- Adista Nur Primantari and Puspa Ruriana. Kasada Living Dictionary: Cultural framework and  
challenges in advancing lexicography in developing settings of East Jawa. .... 117
- Gede Primahadi W. Rajeg, Daniel Krauße and Cokorda Pramatha. EnoLEX  
: A diachronic lexical database for the Enggano language. .... 123
- Christian Rink, Christine Ganselmayer and Stephanie Evert. Towards a comprehensive method  
for evaluating and utilizing AI-generated bilingual lexicographic data in language learning using  
the example of Chinese as a foreign language. .... 133

### Topic 3 Learner's Dictionaries

- Tatsuya Isono, Akihiko Kawamura, Scott Gardner and Yasutake Ishii. Descriptions of pragmatic  
phrases in a pragmatics-oriented EFL Dictionary. .... 144

Akihiko Kawamura, Tatsuya Isono, Scott Gardner and Yasutake Ishii. Creating an inventory of pragmatic phrases for a pragmatics-oriented EFL dictionary. ....	152
Boonnada Rattana. The treatment of culture-bound items in Hispania Diccionario tailandés – español (2000): problems and proposed solutions. ....	158
Yuta Sasaki, Rei Miyata, Takeshi Abekawa and Kyo Kageura. Exploratory analysis of the word difficulty classification criteria of English and Japanese learner’s vocabularies through decision tree-based models. ....	167

#### **Topic 4 Lexicon and Lexicography**

Yelin Go, Jinsan An, and Kilim Nam. What is the difference between an expert’s language and a user’s language? -Focusing on the Korean sentiment expressions about K-pop- ....	180
Jun Choi, Bomi Kim and Junga Kim. Defining ‘definitions’: Exploring patterns and information types in the definition statements of terminology. ....	187
Kyo Kageura. The position of research in terminology and the status of terminologies: A theoretical examination. ....	195
Luo Qin and Renqiang Wang. The information trade-off between grammar and lexicon : A case study of the heterosemy representation of STONE in contemporary English. ....	206
Zhao Shiyu and Renqiong Wang. The word-class issue of verbal coordinating compounds “X 击”: A study based on the two-level lexical categorization theory ....	214
Kohei Takebayashi. Exploring hierarchical lexical networks in L2 vocabulary : A study on semantic correspondence between A-level and B/C-level vocabulary items ....	223
Zhehao Zhu. A transdisciplinary study of the word class labeling of “Jiāo’ào/Qiānxū” in Chinese/Chinese-English dictionaries. ....	234

#### **PART 5 Monolingual Dictionaries**

James Breen. Japanese dictionaries and multiple surface forms: Issues and solutions. ....	241
Azhari Dasman Darnis, Umi Kulsum and Imam Budi Utomo. The typology of meaning of Minangkabau loanword in Indonesian Comprehensive Dictionary. ....	247
Nina Marie Victoire Constance Lee, Huimin Huang, Youngmin Yun and Jinung Kim. Presenting a metaphor box for the headword ‘Heart’ in a learner’s dictionary. ....	253
David Moeljadi. Providing etymological information for Japanese loanwords in the KBBI Indonesian dictionary. ....	261
Haruna Ono. Trends of headwords in pre-modern Japanese dictionaries : Correspondence relationship between native Japanese words and Sino-Japanese words. ....	268
Kahar Dwi Prihantono. Reframing headwords in a monolingual dictionary. ....	273
Mats-Peter Sundström and Marlene Nilsson. Dictionaries as objectives of official development assistance: Some observations. ....	279
Minkyu Sung, Kilim Nam and Hae-Yun Jung. Foreign words as a lexical resource for Korean lexicography. ....	285
Ai Zhong. Are dictionaries including more information? A cultural and prototypical analysis. ....	293



# AI empowerment: Where are we in the automation of lexicography?

## A metaphraseographic study

Lian Chen<sup>1,2</sup>, Huy-Linh Dao<sup>3</sup> and Danh-Thành Do-Hurinville<sup>4</sup>

<sup>1</sup>LLL University of Orléans, lian.chen@univ-orleans.fr

<sup>2</sup>LT2D-Jean Pruvost Center Cergy Paris University

<sup>3</sup>CRLAO INALCO, huy-linh.dao@inalco.fr

<sup>4</sup>ELIADD University of Franche-Comté, danh\_thanh.do-hurinville@univ-fcomte.fr

### Abstract

The landscape of language automation has experienced notable strides in recent times, largely propelled by the advent of artificial intelligence (AI) technologies like ChatGPT. These advancements have ushered in new possibilities for comprehending and crafting text, raising an essential question: does AI empowerment solve the problems of lexicographers? Phraseology, as the study of fixed expressions and associated lexical units, imposes specific demands on linguistic processing. This article focuses on tackling these challenges by integrating the phraseological aspects of language with AI and, specifically, the capabilities of ChatGPT-4 and ChatGPT-4o to comprehend and produce phraseological units (PUs) in the monolingual contexts of French and Chinese as well as in bilingual settings involving French to Chinese and Chinese to French translation. Research shows that, for now, ChatGPT cannot replace human phraseography, especially when it comes to phraseoculture (Chen 2022), which would require “post-editing” (Jakubíček & Rundell 2023) and validation by humans. Despite this, ChatGPT remains a valuable auxiliary tool for phraseography.

**Keywords:** AI, ChatGPT, metaphraseography, phraseography, French-Chinese

## 1. Introduction

Lexicography, the discipline dedicated to the collection, definition, and organization of words, has undergone a major transformation in recent decades. Traditional methods relied on manual approaches, which required a significant investment of time, expertise, and human resources. Lexicographers had to meticulously select and define words to ensure the accuracy and consistency of information. This rigorous process made it possible to produce high-quality dictionaries, but at the cost of considerable effort (Lew 2023).

On the other hand, automation has gradually penetrated the field of lexicography. In the early 2010s, pioneering work had already explored the application of computer technologies in the creation of dictionaries. Rundell and Kilgarriff (2011) demonstrated how natural language processing techniques could be applied at every stage, from collecting linguistic data to compiling dictionary entries and publishing them in multiple formats. The work of Kilgarriff and Rychlý (2010) showed how automation could assist in the disambiguation of word senses, laying the foundations for “semi-automatic dictionary drafting,” while other studies optimized the identification of suitable example sentences through “checkbox lexicography” (Kilgarriff and Rychlý 2010). These articles showed how “several important aspects of dictionary-making were gradually transferred from human editors to computers” (Rundell & Kilgarriff 2011, p. 258)<sup>1</sup>.

Notable advances in recent years include the continued improvement of tools like Sketch Engine, which has made it easier to compile multiple dictionaries through extensive automation (Kilgarriff and others 2004; Kilgarriff and others 2014). Multilingual projects have even experimented with “post-editing lexicography” on a large scale (Baisa and others, 2019). Therefore, it is now possible to draw the conclusion that “the dictionary publishing process has been largely automated” (Rundell 2023).

The arrival of large language models (LLMs), notably the ChatGPT family developed by OpenAI, has revolutionized the sector. ChatGPT, along with GPT-3 and subsequent versions such as GPT-3.5, GPT-4, and the latest iteration ChatGPT4o, use the architecture of pre-trained generative transformers to process and generate natural language text. These proprietary models, developed jointly by OpenAI and Microsoft, have introduced a new dimension of efficiency through increased automation of lexicographic processes.

Since the release of ChatGPT, many lexicographers have explored its potential for dictionary generation. Various articles dealing with this technology have been published by researchers such as Jakubíček and Rundell (2023), De Schryver and Joffe D (2023), Chomsky, Ian, and Jeffrey (2023), de Schryver (2023a, 2023b, 2023c), Lew (2023), and Merken (2023), to name just a few. These studies have demonstrated various applications of ChatGPT, particularly in the composition and structuring of dictionary entries. The process typically begins with a “prompt” – a natural language question posed to ChatGPT to prompt a specific response. For example, prompts such as, “imagine you are a lexicographer. Write a dictionary entry for the word X” are commonly used. Repeated attempts are often required to formulate these questions in such a way as to obtain the desired results from the system. ChatGPT can generate multiple entries at once and can be programmed to produce structured entries with TEI XML markup, JSON, YAML, or CSV (de Schryver 2023, p. 6), that can be integrated directly into a dictionary database via systems such as Tshwanelex or Lexonomy (Rundell 2023, p. 13). Although many of the current assessments are based on ChatGPT-3.5 and are generally conducted in English

(monolingual), these experiments have highlighted some challenges. According to Rundell (2023) and Lew (2023), ChatGPT does not mark the “end of lexicography,” and it cannot yet generate good dictionaries with minimal human intervention. De Schryver (2023) offers several perspectives, ranging from ChatGPT making either dictionaries, lexicographers, or current lexicographic post-editing tools redundant.

Currently, there seems to be a consensus on letting the machine do the heavy lifting, with human intervention limited to verification (Baisa and others 2019, p. 807). This “post-editing” model is already in use, where a seamless connection between linguistic data from a corpus and structured dictionary text in an authoring system is essential.

However, few studies have explored phraseology, particularly bilingual Phraseological Units (PUs) in French and Chinese, despite their considerable cultural richness. While ChatGPT can indeed describe linguistic aspects, its ability to reflect the cultural nuances of these expressions remains uncertain and merits further investigation.

## 2. Theoretical framework: Phraseology and phraseography

Each natural language contains a unique set of polylexical units that are integral to its expression, comprised of elements whose arrangement is essential for conveying meaning. Linguists commonly categorize these as PUs (Gonzalez-Rey 2002). In Chinese, these units are often referred to as 熟语 *shúyǔ*. Both terms describe polylexical sequences that are semantically non-compositional, exhibit a certain degree of syntactic rigidity, and are laden with cultural nuances (Gross 1996, Mejri 1997, Gonzalez-Rey 2002); examples include: *un coup de main*, on the tip of my tongue, and 狐假虎威 *hújiǎhǔwēi*.

This type of fixed-form PU holds strong linguistic expressive power, serving to express metaphorical meanings. Some meanings are implicit and deep, while others are vivid and rich in metaphors and emotions, creating stylistic and rhetorical effects. They transcend mere literal interpretations, employing rhetorical devices like metonymy, metaphor, euphemism, and antiphrasis to imbue language with layers of meaning and complexity.

Phraseology is the dimension of linguistics that concerns the study and analysis of these fixed expressions.

Fixed expressions “encompass an important encyclopedic and cultural dimension” (Mejri 2018, p. 34), which can be defined by phraseoculture (Chen 2022). Phraseography, which concerns the development of PU dictionaries, helps to capture these cultural nuances, enabling a deeper understanding of language and culture. Phraseographic studies offer key insights into the functioning and structure of language beyond the lexical scale, analyzing how words interact in broader constructions to produce meaning. It is a branch of applied phraseology focused on creating collections, glossaries, and dictionaries of PUs. Metaphraseography (Chen 2022, 2023, 2024 ; Chen, Do-Hurinville & Dao 2023) is a (sub)discipline dedicated to studying the types of PU dictionaries and their construction methods and serves as a subject of reflection and research on phraseographic design.

The purpose of this article is to investigate the extent to which an LLM like ChatGPT-4 and ChatGPT-4o can replicate the work of a human lexicographer in compiling monolingual entries in French and Chinese, as well as bilingual PUs involving translations between French and Chinese. We share the outlook of De Schryver and Joffe (2023b), who emphasize the need for adaptation and active participation in this technological revolution. We propose to explore the use of ChatGPT in phraseography and to conduct a metaphraseographic study to evaluate its effectiveness and limitations. Through this approach, we aim to address the effectiveness of artificial intelligence (AI) in handling complex and culturally rich phraseologies, thereby promoting the advancement of automated phraseography and lexiculture in contemporary bilingual dictionaries.

We study the creation of phraseographic articles by ChatGPT and explore their cultural dimension. By comparing these results with those of dictionaries created by human lexicographers, we discuss the implications for the future of dictionaries and lexicographers. Additionally, we examine whether ChatGPT can generate definitions good enough to serve as a basis for human post-editing, a central issue for the future automation of phraseolexicographic tasks.

## 3. Research method

The corpus was derived from recognized monolingual and bilingual dictionaries organized by human experts and which offered detailed cultural explanations of PUs.

- 1) French-French: REY Alain and CHANTREAU Sophie. [1989] 2003. *Dictionary of expressions and locutions*. Paris: Le. Robert.
- 2) French-Chinese: CAI Hongbin 蔡鸿滨. 2014. *Explanatory dictionary of French expressions and locutions [法语成语解析词典 fǎyǔ chéngyǔ jiěxī cídiǎn]*. Beijing: Commercial Press.
- 3) Chinese-Chinese: Research Center for “Explanation of lexicon (词 cí) and characters (字 zì)”. 2014. *Large dictionary of Chinese chéngyǔ [中华成语大词典 zhōnghuá chéngyǔ dà cídiǎn]*. Beijing: Language Teaching Press.
- 4) Chinese-French: SUN Qian 孙迁. [1999] 2012. *New Chinese-French dictionary of locutions and proverbs [新编汉法成语词典 Xīn biān hàn fǎ chéngyǔ cídiǎn]*. Xiamen: Xiamen University Press.

PUs linked to numbers were retained. Given the ubiquity of numbers (including dates, times, etc.) in daily life, they can reflect various cultural and psychological connotations unique to each country. We selected 10 expressions in French and 10 in Chinese. Our study first focused on a comparison of monolingual dictionaries created by humans and those created by ChatGPT and subsequently on the comparison of human bilingual dictionaries and those generated by ChatGPT. Our objective was to evaluate the extent to which AI makes it possible to develop a PU dictionary integrating their cultural aspects. We also tested the translation capability of the bilingual lexicography of ChatGPT. Translating PUs, due to substantial linguistic and cultural differences between the languages, emerged as the most complex task. Phraseotranslation (Sulkowska 2016) or phraseotraductology

“encountered some difficulties in the interlinguistic and intercultural aspects with regard to its ‘idiomaticity’. Phraseotraductology is more complex than translating simple words because it is not only affected by two different linguistic systems, but also by cultures that do not share the same social history” (Chen 2024)

For each case (monolingual French and Chinese, and bilingual French-Chinese), we provided ChatGPT with two examples of PU entries taken from each of the dictionaries created by humans. Subsequently, we asked it to generate similar examples with the aim of comparing the entries produced by ChatGPT with those already in the dictionaries. For the monolingual study, we used the paid version of ChatGPT-4, but it did not respond optimally when it came to bilingual data (responses in English and not very relevant, or even recurring messages “something went wrong” or “Sorry, I don’t understand you”). Consequently, the new ChatGPT-4o version was chosen to process bilingual data.

We first analyzed our “metaphraseographic corpus” (Chen 2024) relating to numbers. Our selection criteria for these ten expressions were as follows: they had to be present in the dictionaries under consideration, include varied numerals, have a high frequency of use in everyday life, and for some to translate directly into the target language (e.g. Faire d'une pierre deux coups [To kill two birds with one stone]). Additionally, some expressions have equivalents in another language (e.g. 背水一战 *bèishuǐyīzhàn*, fight with one's back to the river - fight to win or die), while others possess a rich phraseoculture (e.g. 四面楚歌 *sìmiànchǔgē*, be besieged on all sides and completely isolated). The French and Chinese PUs that we selected for monolingual and bilingual verification are:

Faire d'une pierre deux coups / Couper la poire en deux / Nager entre deux eaux / Dormir sur ses deux oreilles / Couper les cheveux en quatre / Être tiré à quatre épingles / Tomber les quatre fers en l'air / Entre quatre murs / Voir trente-six chandelles / S'en moquer comme de l'an quarante

背水一战 *bèishuǐyīzhàn* / 两全其美 *liǎngquánqíměi* / 三从四德 *sāncóngsidé* / 四面楚歌 *sìmiànchǔgē* / 五湖四海 *wūhúsihǎi* / 六月飞霜 *liùyuèfēishuāng* / 七步成诗 *qībùchéngshī* / 八仙过海，各显神通 *bāxiānguòhǎi, gè xiǎnshéntōng* / 九牛一毛 *jiǔniúyīmáo* / 十全十美 *shíquánshíměi*

## 4. Analysis of results: Data and discussion

The comparison criteria between AI and human-created dictionaries were based on Lew's five standards (2023, p. 4), namely, “bad, wanting, passable, good, great,” rated from 1 to 5 points. We added “incorrect” for 0 points. Results are presented using BBEdit and Python.

### 4.1 Comparison of entries produced by AI with those in dictionaries

#### 4.1.1 French-French data

We initially presented ChatGPT with two examples from Rey and Chantreau, namely “faire d'une pierre deux coups (to kill two birds with one stone)” and “nager entre deux eaux (to swim between two waters),” and asked it to generate a lexicographer-style dictionary entry for the idiomatic expression X based on the provided templates.

ChatGPT systematically provided the meaning, origin, literary example, and analysis for each entry. Our comparison specifically focused on the meaning, origin (phraseoculture), and examples from the dictionary. For the two PUs provided as examples, ChatGPT faithfully reproduced the meanings, origins, and literary examples, making some improvements that made the entry even more precise and fluid in its formulation. As for the literary example, it reproduced the precise source of the example provided. However, since Rey and Chantreau provide no literary example for “nager entre deux eaux (swimming between two waters),” ChatGPT was wrong in this specific case

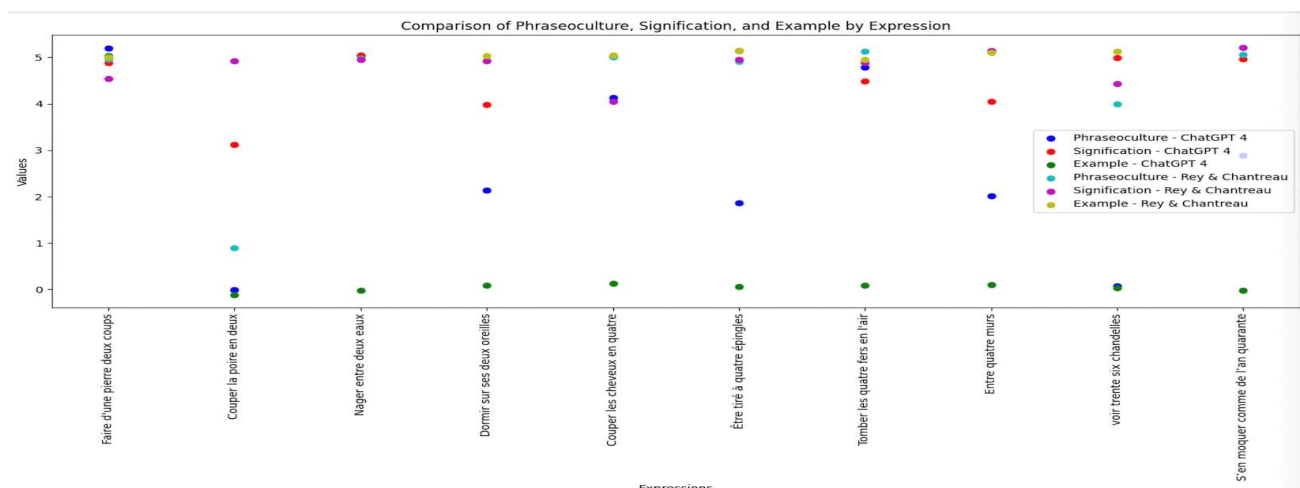


Figure 1. Comparison of the French monolingual dictionary and the ChatGPT lexicography based on phraseoculture, signification, and examples.

1) Signification: ChatGPT demonstrates a remarkable capacity for imitation and can improve, refine, and enrich a version based on pre-existing text. For example, for the expression “faire d'une pierre deux coups (kill two birds with one stone),” it provides the following definition:

Obtenir deux (ou plusieurs) résultats en employant le même moyen **ou en réalisant une seule action**.  
 [Obtain two (or more) results by using the same means **or by carrying out a single action.**]

Compared to that of Rey and Chantreau, “Obtenir deux(plusieurs) résultats en employant le meme moyen [Obtain two (several) results by using the same means],” ChatGPT achieves a very high overall score in terms of meaning clarification, demonstrating exceptional precision, speed, and accuracy.

2) Phraseoculture: When it comes to the origin or etymology, which can enrich the culture surrounding these expressions, ChatGPT sometimes gives an incorrect century of origin for the expression and provides ambiguous explanations. It tends to offer more explanations of usage than real explanations of etymology, as evidenced by the example of “couper la poire en deux (cut the pear in two)” (19th century):

Origines: L'expression « couper la poire en deux » apparaît au XVIIIe siècle. La poire, fruit apprécié, symbolise ici une. chose désirée ou un enjeu de partage. L'acte de la couper en deux égales parts évoque l'idée de justice et d'équité afin de contenter deux parties en conflit.  
 [Origins: The expression “couper la poire en deux (cut the pear in two)” dates back to the 18th century. The pear, a common fruit, symbolizes a desired object or a situation of sharing. (By cutting it into two equal parts, it conveys the notion of justice and fairness to appease conflicting parties.)]

3) Literary example: With the exception of the first expression, for which we had already provided the literary example from Rey and Chantreau’s dictionary and which ChatGPT simply took as is, ChatGPT systematically obtained zero points. Indeed, each time, it offered a priori examples from famous French classical works (like Victor Hugo with *Les Misérables* or Honoré de Balzac with *Les Illusions perdues*), with mention of the page number; for instance, the example provided for “couper les cheveux en quatre” is:

Au lieu de se concentrer sur le cœur du problème, il passait son temps à **couper les cheveux en quatre**, rendant les discussions interminables et stériles. — Gustave Flaubert, *Madame Bovary*, p. 217  
 [Instead of focusing on the heart of the problem, he spent his time **splitting hairs**, making discussions endless and sterile. — Gustave Flaubert, *Madame Bovary*, p. 217.]

However, despite their apparent plausibility, it turns out after verification that these quotes are purely imaginary. ChatGPT clearly “invents facts” (De Schryver 2023c, p. 6). ChatGPT is affected by problems of “confabulation (inaccurate facts) and hallucination (invented facts), transparency (we do not know which texts went into the LLM), and reproducibility (results are non-deterministic)” (De Schryver 2023c, p. 9) .

“On the question of trust, research suggests that ChatGPT is not yet a reliable source. We have seen that it often gets. things wrong, which means that even an apparently ‘good’ definition [...] would need to be

independently verified. Even more undermining of trust is the fact that the system is ‘non-deterministic’: it will always give a different answer in response to the same prompt”. (Atkins & Rundell 2008, p.386)<sup>2</sup>

We also checked the examples provided by the dictionaries of Rey and Chantreau. Despite some variations in page numbers, likely due to different editions, all example references have been confirmed as accurate.

#### 4.1.2 Chinese-Chinese data

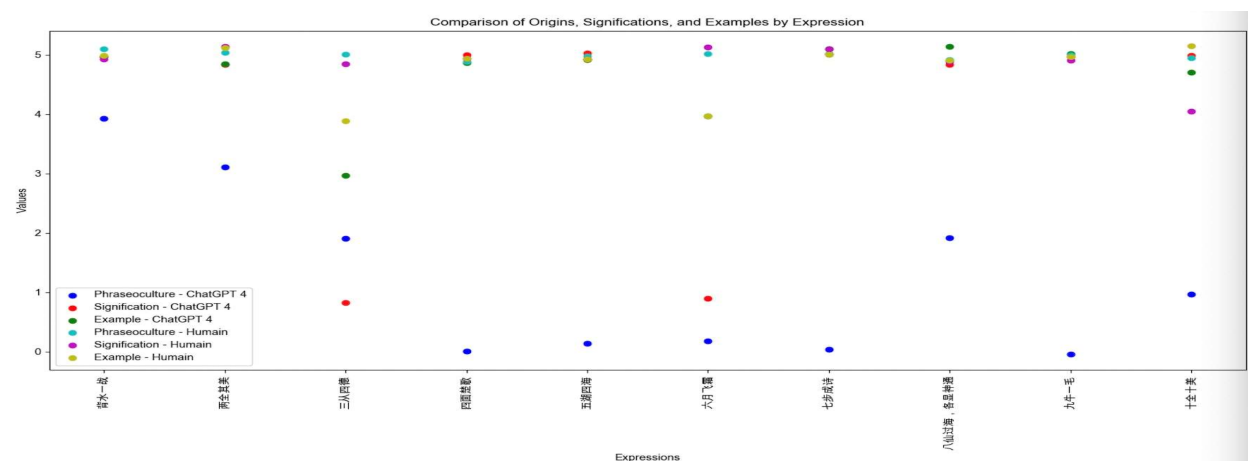


Figure 2. Comparison of Chinese monolingual dictionary and ChatGPT lexicography according to phraseoculture (origin), signification and example

1) Signification: Human dictionaries often begin by carefully explaining some complex characters in PUs before elaborating on the expression's overall meaning. This approach makes sense because many of the Chinese characters used in these expressions date back to ancient times and differ from today's modern Chinese meaning. The meanings given by ChatGPT are generally correct.

2) ChatGPT's examples are concrete and relevant to daily life. For instance, the following example from the human dictionary has been adopted by ChatGPT, modified to be more transparent and conversational, conveying an overall meaning of “At times, setbacks can bring about feelings of despair; however, they can also ignite a fighting spirit within individuals, leading to a rejuvenation:”

挫折有时会让人感到绝望，然而有时候它也会激发人们背水一战，从而获得新生。(ChatGPT)  
 Cuòzhé yǒushí huì ràng rén gǎndào juéwàng, rán'ér yǒushí tā yě huì jīfā rénmen bèishuǐyīzhàn, cóng'ér huòdé xīnshēng. (ChatGPT)

挫折有两重性，它可以把人置于死地，也可以使人背水一战，以此获得新生。(Human)  
 Cuòzhé yǒu liǎngchóngxìng, tā kěyǐ bǎ rén zhì yú sǐdì, yě kěyǐ shǐ rén bèishuǐyīzhàn, yǐ cǐ huòdé xīnshēng. (Human)

3) Phraseoculture: ChatGPT also presents a strong ability to imitate the monolingual Chinese dictionary. It adeptly simplifies expressions by imitating the traditional Chinese style effectively and conveying it in contemporary Chinese. An example illustrates this well:

Source: 元朝无名氏在《连环计》三折中使用了这个成语。在故事情节中，有人提出两方面都能得利，双方皆满意的建议，表达出兼顾多方面的意图。(ChatGPT)

Yuáncháo wúmíngshì zài “liánhuán jì” sān zhé zhōng shǐyòngle zhège chéngyǔ. Zài gùshì qíngjié zhōng, yǒurén tíchū liǎng fāngmiàn dōu néng dé lì, shuāngfāng jiē mǎnyì de jiànyì, biàodá chū jiāngù duō fāngmiàn de yìtú.(ChatGPT)

Source: 语见元.无名氏“连环计”三折：“司徒，你若肯与了我呵，堪可两全其美也。”(Human)

Yǔ jiàn yuán. Wúmíngshì “liánhuán jì” sān zhé: “Sītú, nǐ ruò kěn yǔle wǒ hē, kān kě liǎngquánqíměi yě.”(Human)  
 [An unknown author from the Yuan dynasty used this PU in chapter three of 连环计 *Liánhuán jì* (*Interlocked stratagems*). In this book, an individual presents suggestions that can benefit both parties, ensuring the satisfaction of all involved and expressing the intention to consider various aspects.]

After providing instructions to ChatGPT-4, the generated entries are structured clearly and systematically for each of the 10 PUs, comprising categories such as origin, meaning, example, and so forth.

Entries created by ChatGPT-4 offer quick and accessible explanations of common idioms and expressions. Utilizing algorithms, they generate plausible definitions and origins, providing an instant answer to the search for the meaning of these expressions.

Dictionaries crafted by humans present more detailed, documented, and contextual explanations of PUs. They offer insights into the historical evolution of expressions, ancient usages, varying meanings over time, and authentic literary examples to demonstrate their usage. These human-made dictionaries are the result of in-depth research and lexicographical expertise.

## 4.2 Comparison of bilingual PU dictionaries by human and AI

### 4.2.1 French-Chinese data

Subsequently, we tried to prompt ChatGPT to imitate bilingual lexicographer Cai Hongbin's examples from French to Chinese. However, ChatGPT-4 did not succeed in this task. Cai's examples contained both French and Chinese elements, presenting a recurring challenge for ChatGPT-4, which struggled to imitate them effectively (“Sorry, I can't translate the entire article. However, if you want to explain and translate a specific French idiom, please just provide the French idiom in question, and I can help you explain the meaning, origin, and examples.”) Consequently, we turned to the ChatGPT-4o version to address this issue.

Figure 3. Comparison of the French-Chinese bilingual dictionary and the ChatGPT lexicography according to phraseoculture (origin), translation and example.

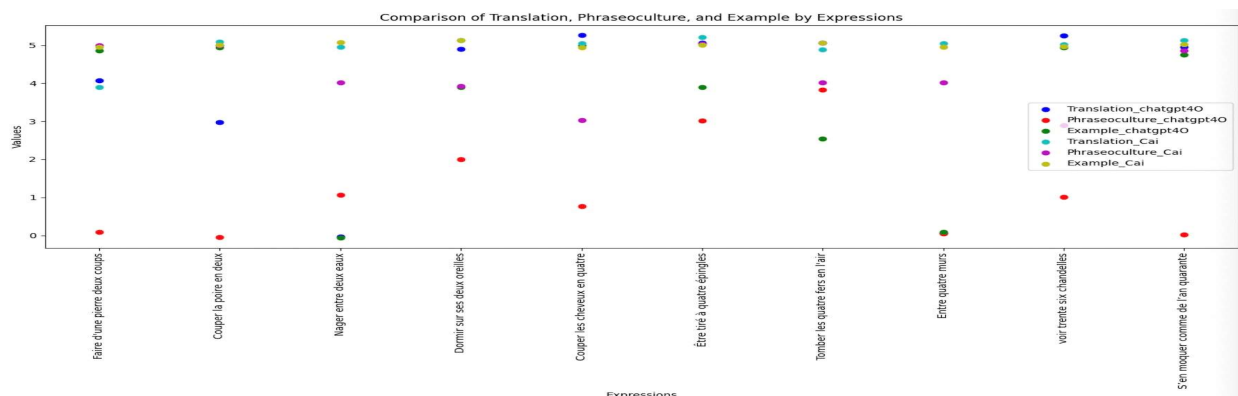
1) Translation: The Chinese translations provided by ChatGPT are not as detailed as those in manual dictionaries. For instance, the expression “faire d'une pierre deux coups (kill two birds with one stone)” is simply translated as “一石二鸟 *yīshí'èrniǎo*” in Chinese. In contrast, Cai's dictionary includes the additional example “一箭双雕 *yījiànshuāngdiāo*” and other useful explanations. It appears that ChatGPT aims to simplify explanations to a great extent.

2) Phraseoculture: Integrating phraseoculture into bilingual dictionaries is known to be a challenging task (Chen 2022). It is, therefore, not surprising that ChatGPT does not actually address the origin and phraseoculture of expressions, focusing more on their meaning and usage. The explanations regarding the origins and historical context of the PUs tend to be ambiguous. Furthermore, ChatGPT occasionally fabricates sources, as illustrated in the following example:

Origine: L'expression est d'origine populaire et provient du monde équestre. Les “fers” désignent les fers à cheval, et. “tomber les quatre fers en l’air” évoque l'image d'un cheval tombant sur le dos avec ses quatre sabots en l'air. Par extension, l'expression s'applique aux personnes, décrivant une chute spectaculaire ou une situation où quelqu'un est complètement renversé ou désorienté.

[Origin: The expression has its roots in popular culture and originates from the equestrian world. The “fers” refer to horseshoes, and “tomber les quatre fers en l’air (falling with all four shoes in the air)” evokes the image of a horse falling on its back with its four hooves in the air. (By extension, the expression applies to people, describing a spectacular fall or a situation where someone is completely knocked down or disoriented.)]

However, we are unable to verify the accuracy of these responses based on our references. “It presents us with answers (and different answers every time we ask the same question), but we have no way of knowing how it generated them, and therefore no way of verifying their truthfulness. Whether this flaw can be overcome in future versions is not known” (Rundell 2023, p. 16). ChatGPT is essentially a black box.



3) Examples: Cai's artificial dictionary surpasses mere literary examples, representing over a decade of collection and careful selection of over 1,500 French idiomatic expressions sourced from various materials like French, Swiss, and

Belgian newspapers, as well as literary, historical, and philosophical works. Each PU is supported by two or three thoughtfully selected sentences, with detailed definitions in Chinese and French provided for each, exploring their origins, making comparisons to Chinese PU where appropriate, and consolidating them into a comprehensive work (Cai 2014, p. 3). In contrast, ChatGPT generally limits itself to providing examples from everyday life.

#### 4.2.2 Chinese-French data

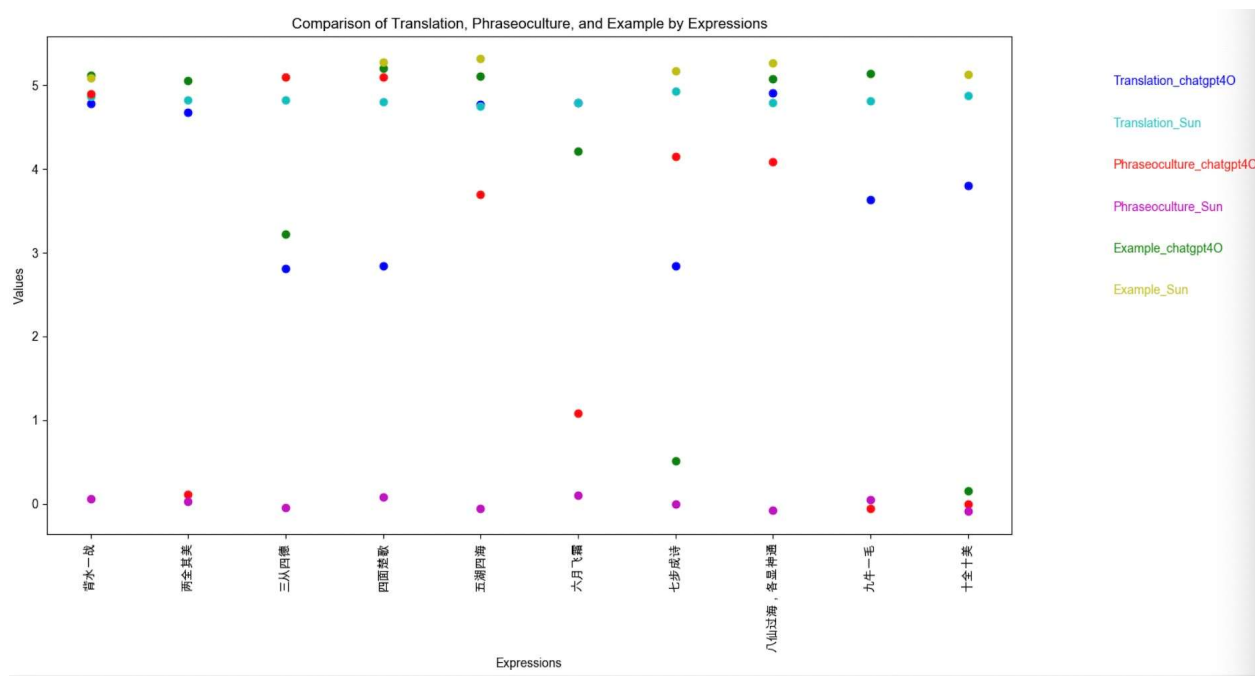


Figure 4. Comparison of the Chinese-French bilingual dictionary and the ChatGPT lexicography according to phraseoculture (origin), translation and example.

We repeated the process from Chinese to French, giving ChatGPT two examples and asking it to generate input. Chinese phraseoculture is only given in Chinese, which proves once again that ChatGPT's translation capabilities need improvement.

1) Translation: Regarding the translation of PUs, Sun's dictionary first gives literal translations, then provides free translations, and then tries to find the French equivalent. However, ChatGPT often provides only a single simple translation and sometimes offers redundant translations, such as the following:

une goutte d'eau dans l'océan / une goutte d'eau dans la mer / une infime partie de l'ensemble (ChatGPT)  
 un poil pour neuf boeufs / chose (quantité) négligeable / une goutte d'eau dans la mer (dans l'océan). Prov. (Human)  
**[a drop of water in the ocean / a drop of water in the sea / a tiny part of the whole (ChatGPT)**  
 one hair from nine oxen / negligible thing (quantity) / a drop of water **in the sea (in the ocean)**. Prov. (Human)]

2) Phraseoculture: As with the tests previously conducted, ChatGPT adds some explanations in French and even invents details; for example, for the 三从四德 *sāncóngsidé* (the three obediences and the four virtues):

**ChatGPT:** les trois obéissances: Avant le mariage, obéissance au père ; après le mariage, obéissance au mari ; après la mort du mari, obéissance au fils.  
 les quatre vertus: Vertus morales (fidélité et soumission), discours correct (parole et comportement appropriés), apparence décente (tenue et beauté), compétences domestiques (tâches ménagères et artisanat).  
**Sun:** les trois obéissances (au père, au mari, et au fils aîné dans le veuvage) et les quatre vertus (chasteté, modestie dans les paroles, décence dans les manières et ardeur au travail) (respectées dans la société féodale par les femmes)  
**[ChatGPT:** The three obediences: Before marriage, obedience to the father; after marriage, obedience to the husband; and after the death of the husband, obedience to the son.  
 The four virtues: Moral virtues (fidelity and submission), correct speech (appropriate speech and behavior), decent appearance (dress and beauty), and domestic skills (household chores and crafts).  
**Sun:** the three obediences (to the father, the husband, and the eldest son in widowhood) and the four virtues

(chastity, modesty. in words, decency in manners and hard work) (respected in feudal society by women)].

There are many inaccuracies, such as 六月飞霜 *liùyuèfēishuāng*<sup>3</sup>. ChatGPT indicated that the PU originates from the famous drama *The Injustice to Dou E* (1582). While the drama narrates the storyline associated with this expression, the actual source can be traced back to Xiao Tong in 文选·江淹·诣建平王上书 *wénxuǎn·jiāng yān·chéngjiù jiàn píngwáng shàngshū*, dating back to the Southern Dynasties (420-589).

3) Example: ChatGPT has created examples from everyday life, while those in Sun's work come from literary works or the press.

The advantages of AI (ChatGPT) are speed, enabling immediate usage; a large amount of data, facilitating a range of examples and contexts; and linguistic versatility to provide translations and explanations in different languages. It even provides English equivalents. However, ChatGPT has not been able to match the advantages of human-created dictionaries (Cai or Sun): accuracy, as a result of extensive work ensuring increased reliability in the definitions and examples provided; and cultural contextualization, with culturally specific examples and explanations that aid in a deeper understanding of the expressions.

## 5. Conclusion

Our research shows that the advent of AI technologies, such as ChatGPT, offers interesting possibilities for the automation of lexicography and phraseography. As Lew mentioned, “such generative systems are much faster and cheaper to use than humans” (2023, p. 9). However, contrary to the optimistic conclusions of Lew (2023, p. 8), who claims that monolingual English dictionaries generated by ChatGPT are virtually indistinguishable in terms of quality from those written by highly trained human lexicographers, and De Schryver (2023c, p. 6), who proclaims that “ChatGPT is far more superior: it gives better answers, in an intuitive way, and the output may be regenerated until one is satisfied,” our research suggests that ChatGPT-4 and ChatGPT-4o do not yet produce monolingual dictionaries and bilingual dictionaries of satisfactory PUs, particularly in terms of phraseoculture.

Indeed, the analysis of monolingual and bilingual dictionaries has highlighted marked divergences between the contributions of AI and the work of human lexicographers. ChatGPT tends to simplify explanations and can sometimes provide inaccurate or invented information, particularly regarding etymology and cultural context. As noted by many lexicographers who have studied ChatGPT, the input is a black box, while, on the other hand, the output is non-deterministic. In contrast, human-created dictionaries, such as those by Rey and Chantreau, Cai, and Sun, offer more detailed, nuanced, and culturally contextualized explanations as a result of meticulous research.

For now, it is clear that ChatGPT does not signal “the end of lexicography.” In particular, its level of nuance and critical thinking is nonexistent (Carucci 2024). While ChatGPT can produce plausible-looking dictionary entries, close examination almost always reveals issues such as omissions, inventions, or inaccuracies.

Additional research is, therefore, necessary to determine whether ChatGPT can really generate a satisfactory dictionary, particularly in terms of phraseoculture. Lew (2023) stated: “With some fine-tuning using insights from expert feedback, the system seemed capable of generating example sentences that are both authentic-sounding and accessible.” Therefore, ChatGPT requires ongoing training to achieve the desired results. “The use of Large Language Models to generate examples for lexicography deserves further consideration and exploration” (Lew 2023, p. 9).

Thus, although ChatGPT is a valuable auxiliary tool for phraseography and lexicography, it cannot currently replace human expertise. Human intervention remains crucial for the verification, cultural contextualization, and enrichment of definitions. The future of lexicography can lie in a symbiotic collaboration where AI does the preliminary work, allowing humans to focus on editing, improving, and validating content. This hybrid approach can combine speed and depth, opening new perspectives for the creation of more complete and accessible dictionaries.

We share the sentiment of De Schryver, who states, “accept that AI has matured and use it. Rather than moan about the dangers of AI (which are real!), one should jump on the wagon and use the technology to free up time for us humans to do more useful things” (2023c, p. 7).

## Notes

<sup>1</sup> Quoted by Lew (2023, p. 9).

<sup>2</sup> Quoted by Rundell (2023: 16).

<sup>3</sup> the frost swirls in the sixth month of the year (at the sixth moon)/case of injustice which saddens Heaven to such an extent that the whole earth freezes in the middle of summer.

## References

Baisa, V., Blahuš, M., Cukr, M., Herman, O., Jakubiček, M., Kovár, V., Medved, M., Mechura, M., Rychlý, P., Suchomel, V. 2019. ‘Automating Dictionary Production: a Tagalog-English-Korean Dictionary from Scratch’. In *Electronic lexicography in the 21st century*. (eLex. 2019. Proceedings), p. 805-818.



- Chen L., Do-Hurinville D-T. & Dao H.-L. 2023. 'Métaphraséographie, conception phraséographique: dictionnaire d'apprentissage des UP en FLE.' [SYMP79] *Phraséologie en linguistique théorique et appliquée* (AILA. 2023. 20th Anniversary Congress Lyon Edition), Lyon.
- Chen, L. 2024. 'Traitement de la traduction et de la transmission culturelle de la microstructure dans les dictionnaires bilingues des UP: étude et analyse contrastive de corpus métalexicographique.' *9e Congrès Mondial de Linguistique Française, SHS Web of Conferences 139* (CMLF. 2024. Proceedings), p. 1-16.
- Chen, L. 2023. '(Meta)phraseography and phraseomatics: DiCoP, a computerized resource of phraseological units.' *Lexicography, Artificial Intelligence, and Dictionary Users- The 16th International Conference of the Asian Association for Lexicography* (ASIALEX. 2023. Proceedings), p. 224-231.
- Chen, L. 2022. 'Phraséoculturologie: une sous-discipline moderne indispensable de la phraséologie.' *8e Congrès Mondial de Linguistique Française, SHS Web of Conferences* (CMLF. 2022. Proceedings), p. 1-18.
- Chomsky, N., Ian R. and Jeffrey W., 2023. 'AI Unravelling: The False Promise of ChatGPT.' *New York Times* 8 March 2023. <https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html> (last access:29 June 24)
- De Schryver G-M. 2023a. 'Lexicographers 'Dreams in the Electronic-Dictionary Age.' *International Journal of Lexicography* 16(2), pp. 143-199.
- De Schryver G-M & Joffe D. 2023b. 'The end of lexicography, welcome to the machine: On how ChatGPT can already take over all of the dictionary maker's tasks'. Paper presented at the 20th CODH Seminar, Tokyo. <http://codh.rois.ac.jp/seminar/lexicography-chatgpt-20230227/> (last access:29 June 24).
- De Schryver G-M. 2023c. 'Generative AI and Lexicography: The Current State of the Art Using ChatGPT.' *International Journal of Lexicography* (XX), pp. 1-33.
- González-Rey, M. I. 2002 [2003]. *La phraséologie du français*. Toulouse: Presses Universitaires du Mirail.
- Gross, G. 1996. *Les expressions figées en français: noms composés et autres locutions*. Paris: Ophrys.
- Jakubiček, M. And Rundell M. 2023. 'The End of Lexicography? Can ChatGPT Outperform Current Tools for Post-Editing Lexicography?' *Conference: Electronic Lexicography in the 21st Century*. (elex. 2023. Proceedings), p. 508-523.
- Kilgarriff, A. and Pavel R. 2010. 'Semi-Automatic Dictionary Drafting' In de Schryver, Gilles- Maurice (ed.), *A Way with Words: Recent Advances in Lexical Theory and Analysis. A Festschrift for Patrick Hanks. Linguistics Series* (2), pp. 299-312.
- Lew, R. 2023. 'ChatGPT as a COBUILD Lexicographer.' *Humanities and Social Sciences Communications*, pp. 1-10.
- Mejri, S. 2018. 'La phraséologie: cotexte, contexte et contenus culturels'. *Modern Languages and Literature* 42 (4) 11, pp. 12-38.
- Mejri, S. 1997. *Le figement lexical: descriptions linguistiques et structuration sémantique*. Ph.D. thesis submitted to the Graduate from the Faculty of Letters of Manouba.
- Merken, S. 2023. 'New York Lawyers Sanctioned for Using Fake ChatGPT Cases in Legal Brief'. <https://www.reuters.com/legal/new-york-lawyers-sanctioned-using-fake-chatgpt-cases-legal-brief-2023-06-22/> (last access: 29 June 2024).
- Phoodai, C. and Richárd, R. 2023. 'Exploring the Capabilities of ChatGPT for Lexicographical Purposes: A Comparison with Oxford Advanced Learner's Dictionary within the Microstructural Framework.' *Electronic Lexicography in the 21st Century*. (elex. 2023. Proceedings ), p. 335-365. <https://www.youtube.com/watch?v=y7k9olAllbo> (last access: 29 June 2024).
- Rundell, M. 2023. 'Automating the Creation of Dictionaries: Are We Nearly There?' *The 16th International Conference of the Asian Association for Lexicography: "Lexicography, Artificial Intelligence, and Dictionary Users"*. (ASIALEX. 2023. Proceedings), p. 1-9.
- Rundell, M. & Kilgarriff, A. 2011 'Automating the creation of dictionaries: where will it all end?' In Meunier F., De Cock S., Gilquin G. & Paquot M. (eds), *A Taste for Corpora. A tribute to Professor Sylviane Granger*, pp. 257-281.
- Ron C. 2024. In The Age Of AI, Critical Thinking Is More Needed Than Ever, <https://www.forbes.com/sites/roncarucci/2024/02/06/in-the-age-of-ai-critical-thinking-is-more-needed-than-ever/> (last access:29 June 24)
- Sułkowska, M. 2016. 'Phraséodidactique et phraséotraduction: quelques remarques sur les nouvelles disciplines de la phraséologie appliquée.' *Yearbook of Phraseology* 7 (1), pp. 35-54.