

手稿の翻刻における AI の利用について

－ フランス近代の事例 －

On the Use of AI in Paleographic Transcription:
A Case Study from Modern France

小田淳一

東京外国語大学アジア・アフリカ言語文化研究所

This paper examines the extent to which AI proved useful at various stages of the paleographic transcription of the private manuscript notebooks (*Carnets*, four volumes) left by Joseph-Charles Mardrus (1868–1949), the translator of the French version of *The Thousand and One Nights*, which was published in Paris during the Belle Époque and achieved remarkable popularity. The transcription work was conducted under an exclusive authorization granted by Mardrus's legal heirs. The present report focuses on *Carnet A*, whose transcription is scheduled for publication in March 2026.

1. はじめに

「翻刻」とは、一般に古文書や碑文などの手書き資料を活字化する作業を指し、近年では AI を利用した翻刻コンテンツやチャレンジも開催されている。翻刻における AI 利用の最大の目的は、従来の方法では多大な時間を要してきた判読不能箇所を支援する点にある。本稿では、フランス近代の手稿を対象とした翻刻作業において試みた生成 AI（以下 AI）による支援の具体例を取り上げ、当該分野における AI の有用性とその限界について考察する。

2. 研究対象と史料の特性

2.1 翻刻テキスト

本稿が研究の対象とする史料は《ベル・エポック》のパリで刊行され一世を風靡した『千一夜物語 *Les Mille et Une Nuits*』フランス語版の訳者ジョゼフ＝シャルル・マルドリユス¹ Joseph-Charles Mardrus (1868-1949) が遺した未公開

の手稿ノート『カルネ *Carnets*』であり²、本稿は 2026 年 3 月に翻刻を刊行予定の『カルネ A』 [Oda et Daaif 2026] を対象とする。

2.2 翻刻テキストの特性

(1) 形式に関する特性

『カルネ A』は他のカルネと同様、私的な覚書として記された手稿ノートであり、各ページは、整っていない細かい筆記体によって埋め尽くされている。こうした書字は、自己使用を目的としたメモにしばしば見られるものであり、判読が困難、時にはほとんど判読不能な箇所も少なくない。この点自体が本史料の顕著な特徴の一つであるが、それに加えて、多数の専門用語や分野固有のジャーゴンが含まれているほか、フランス語以外の言語が頻出する。取り分け、アラビア語は多くの頁に散在しており、さらにヘブライ語、ギリシア語、ラテン語なども随所に用いられている³。

¹マルドリユスの家系はもともとグルジア（カフカス地方）出身の裕福な一族であり、アルメニア・キリスト教の有力者であったマルドリユスの曾祖父は 18 世紀のロシア軍侵攻の際にポーランド王 Stanislas I 世のもとに身を寄せ、次いで王に随行してフランスのルイ 15 世の宮廷に赴いた。その息子（マルドリユスの祖父）はナポレオンのエジプト遠征軍に会計担当将校として従軍し、さらにその息子（マルドリユスの父）もナポレオン 3 世軍の会計担当将校としてイタリアやクリミア半島等への遠征に参加し、第 2 帝政崩壊後にフランス国籍を選んでカイロに居を定めた。マルドリユスは 1868 年にカイロで生まれ、幼年時代をカイロで過ごした後、毎年避暑に出かけていたレバノンで中等教育及び医学の基礎教育をイエズス会の教育機関（現聖ジョゼフ大学）で受けてからマルセイユとパリで医学を修め、1894 年に医学博士の学位を取得する。マルドリユスがパリの文壇と付き合い始めたのはこの頃からであったと思われる。学位取得後の 5 年間は郵便船の船医として勤務し、この間にマラルメに勤められたとの説がある『千一夜物語』の訳を行ったとされている。

²『カルネ』全 4 巻について筆者は、マルドリユスの法定相続人であった故マリオン・シェネ Marion Chesnais 氏から独占的使用許諾を受けている。『カルネ』は他のマルドリユス・コレクションと共にシェネ氏の死去に伴って 2017 年にフランス国立図書館に移管されたが全 4 巻のうち本報告が扱う『カルネ A』と『カルネ B』はいずれも依然非公開となっており、全頁がほぼアラビア語の『カルネ E』と、頁数が少ない『カルネ F.G.H.』 [BnF サイトでは F.G.M. と誤記] のみが予約閲覧に限定して公開されている。³ これらの言語要素は、必ずしも当時の学術的転写規則に従って記されているわけではなく、マルドリユス自身の自由な表記に委ねられている場合が多い。そのため、いくつかの箇所では解読および翻刻がきわめて困難となっている。筆者らはこうした難解な箇所にとくに注意を払い、その大部分を翻刻することに成功し、同時に、単なる翻刻にとどまらず、読者の理解を助けるため、ほぼ体系的に脚注を付した。そこで明示したのは、語や表現の出現位置、文字の書かれ方、欄外注記であるか否か、抹消された語句や文章であるかどうかといった点である。

(2) 内容に関する特性

『カルネ』に収められているのは、主としてフランス語で書かれた個人的な覚書であるが、前述のような他言語による記述も含まれている。その内容は多岐に亘り、百科全書的な記述、政治的意見、同時代の作家・芸術家・詩人・彫刻家に関する所感などが記されている一方で、医師でもあったマルドリユスが専門書や新聞、学術雑誌に掲載された記事を書き写した部分も少なくない。この点において『カルネ』は多様な観点からの研究を可能にする貴重な史料の価値を有している。これらの覚書はマルドリユス個人の教養や思考のみならず、彼が生きた時代に支配的であった思想、イデオロギー、人間関係の規範を知る手がかりを提供するものである。さらに、学際的な視点からこれらの断片的な記述を分析することによって、マルドリユスの時代のパリ社会に新たな光を当て、ひいてはフランスの文化的歴史の新たな側面を明らかにする可能性を秘めている。

3. AI 利用の方法論

手稿テキストの翻刻において AI を利用する際、その最大の目的は判読不能箇所へのアプローチであるが、本稿では以下のような方法を用いた。

3.1 OCR 機能

AI を OCR として用い、判読不能箇所を含む文字列の画像を直接解読させる試みを行ったが、内容語（名詞・動詞等）を含む文字列はほぼ解読できなかった。これは、『カルネ』の大半のページが、前述したように「整っていない細かい筆記体」による走り書きであることに起因するものと考えられる。その一方で、アラビア語⁴、ギリシア語、ヒエログリフについては、マルドリユスが比較的丁寧に記した箇所に限って、ある程度まで解読可能な例が多く認められた。

3.2 文脈推定補完

人手によって一部のみ翻刻された文字列を対象とし、判読不能箇所の前後の可読部分を提示したうえで、その内容を文脈に基づいて推定させた。この方法による解読の成否は、文字列が示す内容の性質によってばらつきが認められたものの、概して提示する文脈の範囲が広いほど、推定精度が向上する傾向が確認された。

3.3 OCR 機能と文脈推定補完の併用

上記の二つの方法を併用して、判読不能箇所の前後の可読文字列部分に加え、当該箇所の画像を提示して推定させた。この方法は、3.1 で述べた OCR 機能を単独で用いた場合よりも、当然ながら解読率の向上が認められたが、やはり当該箇所の内容の性質によって結果には大きな差が生じた。すなわち、固有名詞、あるいは定型的な表現が想定される箇所では比較的高い精度で推定が可能であった一方で、マルドリユス独特の省略表現や言い回しにおいては、推定結果の信頼性は限定的であった。また、画像情報を併用した場合であっても、書字が著しく崩れている箇所については解読への寄与は同じく限定的であった。

4. 翻刻支援における顕著な成功例

本章では AI の利用によって判読不能箇所の解読が進んだ幾つかの成功例の中で顕著なものを挙げる。

4.1 ヘブライ文字のゲマトリア（数秘術）

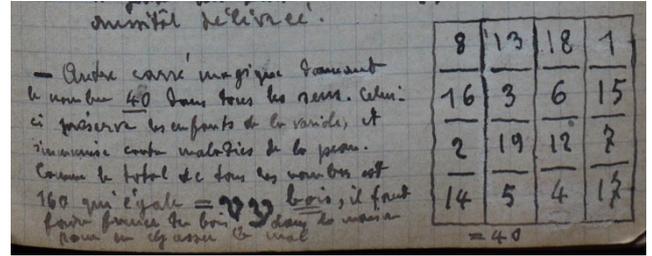


図1: 魔方和 40 の 4 次魔方陣を含む当該部分

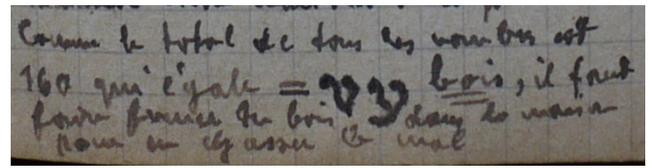


図 2: 魔方陣の説明(拡大)

(1) 魔方陣の説明部分の翻刻と日本語訳

— Autre carré magique donnant le nombre 40 dans tous les sens. Celui-ci préserve les enfants de la variole, et immunise contre maladies de la peau. Comme le total de tous les nombres est 160 qui égale = 73 bois, il faut faire fumer du bois dans la maison pour en chasser le mal.

「— もう一つの魔方陣は、あらゆる方向で合計が 40 になる。これは子供達を天然痘から守り、皮膚病に対する免疫を与えるものだ。すべての数の合計は 160 となり、これは 73 木に等しいので、悪を追い払うには家の中で木を燻さなければならない。」

(2) AI による解読

図 2 を幾つかの AI に解読させたところ、ヘブライ文字であることを認識したのは Gemini のみだったが、73 をそれぞれ発音記号の V (Vav) と Y (Yod) と間違え、それらの数値 (V=6, Y=10) を示して、魔方陣の各行列の総和 160 と辻褃を合わせようとした。

ChatGPT にこれらの情報と上の翻刻テキストを提示して説明させたところ、73 がヘブライ語で「木」を意味することを解読し、さらにそれぞれの数値の和が魔方陣の各行列の総和と等しいことも示した。

(3) 筆者がこの部分に付けた注

「ヘブライ語の 73 (Etz=木) を構成する文字 v (Ayin) と s (Tsadi) のゲマトリア (数秘術) におけるそれぞれの数値は v (Ayin) = 70, s (Tsadi) = 90 で、合計は 160 となる。つまり、木を燻すという行為が、数 (160)、文字 (73)、物質 (bois) を一致させる護符的・魔術的操作となるのである」。

⁴ マルドリユスのアラビア語運用能力については、彼の『千一夜物語』フランス語訳がどのアラビア語原典テキストに依拠したのが不明確だったことなどが原因で、当時から一部の中東研究者に疑問視されていたが、本研究において彼の私的覚書である『カ

ルネ』で用いられたアラビア語を分析した結果、問題となるような重大な欠陥はないという結論に至り、長年取り沙汰されてきた彼のアラビア語能力についての一定の客観的な証拠を得ることができた。

4.2 人名の判読

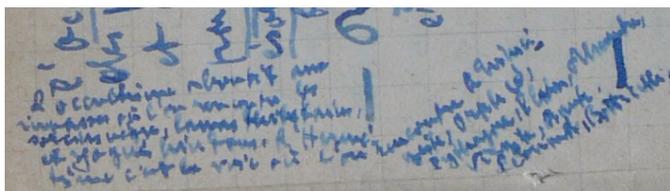


図 3: 当該部分(一文全体)の画像
(ページ左余白に上下, 斜めに書かれている)

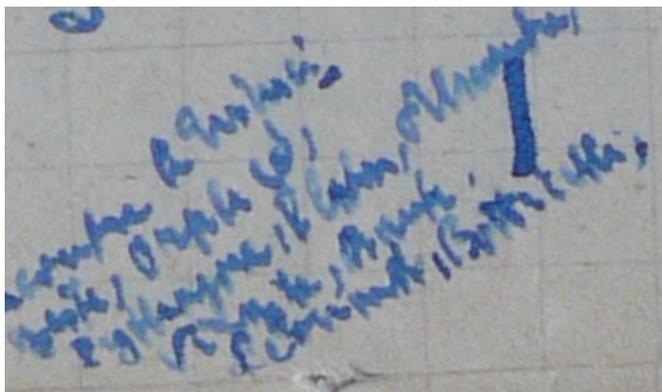


図 4: 人名を含む部分を拡大した画像

(1) 当該部分の翻刻と日本語訳

L'hermétisme c'est la voie où l'on rencontre le Trismégiste, Orphée, Pythagore, Plotin, Alexandre, Virgile, Dante, Eckart, Botticelli.

「ヘルメス主義とは、トリスメギストス、オルフェウス、ピタゴラス、プロティノス、アレクサンドロス、ウェルギリウス、ダンテ、エックハルト、ボッティチェリに出会う道である。⁵⁾」

(2) AI による解読

エックハルト以外の人物名は、既に人手による翻刻で解読済みであったため、この一文の文脈上エックハルトに相当する箇所を「語頭が Ec で始まる人物名」として推測させた。Claude はこれらの人名が神秘思想の歴史における重要人物であることを解読し、Ec で始まりこの思想的系譜に該当する人物として、「神との合一」を説いて中世神秘主義を深化させたマイスター・エックハルトという回答に到達した。

⁵⁾ トリスメギストス (ヘルメス・トリスメギストス) : ギリシャのヘルメス神とエジプトのトート神が融合したとされる伝説上の人物。ヘルメス思想の源流で、「三重に偉大な者」を意味する；オルフェウス : ギリシャ神話の詩人。竖琴の名手で、冥界に赴き亡妻を取り戻そうとした神話で知られる。オルフェウス教の祖；ピタゴラス (c.570-c.495 B.C.) : 古代ギリシャの哲学者・数学者。数を万物の根源とする思想を唱え、神秘主義的教団を形成した；プロティノス (c. B.C.205-B.C.270) : 新プラトン主義の創始者。「一者」から万物が流出するという宇宙論を説き、神秘的合一を重視した；アレクサンドロス大王 (B.C.356-B.C.323) :

5. 補助的ツールとしての活用

本稿の主題である翻刻作業そのものとは直接関係しないものの、AI は付随的な作業においても一定の有用性を示した。具体的には、索引作成に際しての候補語の抽出や整理、ならびに人手による入力作業に伴って生じた些末なミスタイプの検出などにおいて、作業効率の向上に寄与した。一方で、AI による候補語の選定基準は必ずしも一貫しておらず、作業の安定性という点では課題も認められた。これに対し、Excel 上で語彙一覧を管理し、VBA (VBE) による関数やマクロを併用することで、候補語の整理や統合を効率的に行うことが可能となり、AI の出力を人手で制御・補完する手段として有効であった。以上の点から、AI は単独で完結するツールというよりも、既存の編集支援環境と組み合わせることで真価を発揮する補助的ツールとして位置づけられる。なお、索引作成時に用いた Excel 上の VBA (VBE) による関数やマクロについても、その設計および実装の一部において AI の支援を受けた。もともと、AI が生成したコードはそのまま使用するのではなく、実際のデータ構造や作業手順に即して適宜修正を施しており、最終的な動作の検証と責任は翻刻者が担っている。この点においても、AI は作業を代替するものではなく、人手による編集作業を補助するツールとして機能したと言える。

6. おわりに

6.1 仮説生成装置としての AI

AI は、翻刻作業において判読不能箇所の解読を自動的に行う装置というよりも、複数の解釈候補を提示する仮説生成装置として機能する。その出力は、翻刻者の専門的判断によって評価・取捨選択されることを前提としており、AI の利用は翻刻作業の代替ではなく補助として位置づけられる。

6.2 翻刻者の専門的知識と AI 利用の有効性の相関

翻刻者が有する専門的知識の量および質は、AI を翻刻作業に援用した際の有用性に大きな影響を及ぼす。これまで示した各事例から明らかなように、AI は判読不能箇所の解読や補完に一定の効果を発揮するものの、その成果は翻刻者自身の言語的・歴史的・文献学的知識に大きく依存している。すなわち、翻刻者の専門性が高いほど、AI の出力を適切に評価・取捨選択し、翻刻作業に有効に統合することが可能となり、結果として AI 利用の有用性は格段に向上する。

マケドニア王。大帝國を築き、ヘレニズム文化を広めることで神秘思想の発展にも影響を与えた；ウェルギリウス (B.C.70-B.C.19) : 古代ローマの詩人。叙事詩『アエネーイス』の作者。中世には予言者・魔術師的存在としても尊敬された；ダンテ (1265-1321) : イタリアの詩人。『神曲』で地獄・煉獄・天国を巡る霊的遍歴を描いた；マイスター・エックハルト (c.1260-c.1328) : ドイツの神学者・神秘主義者。神との直接的合一を説いた思想家；ボッティチェリ (c.1445-1510) : ルネサンス期イタリアの画家。『ヴィーナスの誕生』『春』など、新プラトン主義の神秘思想を反映した作品で知られる。

謝辞

本報告は JSPS 科研費 20K00492 の助成を受けた研究成果のひとつである。

参考文献

- [小田 2003] 小田淳一：「マルドリユス・コレクションに関する予備調査報告」, 『筑波大学フランス語・フランス文学論集』, 第 18 号, pp. 163-175, 2003.
- [Nishio *et al.* 2022] Tetsuo Nishio, Naoko Okamoto, Jun'ichi Oda, Margaret Sironval, Marion Chesnais and Yo Kaji: *Catalogue du fonds Joseph-Charles Mardrus, traducteur des Mille Nuits et Une Nuit*, Abencerage A. Ghozzi, Paris, 2022.
- [小田他 2022] 小田淳一, セリア・シュルファ, ラハセン・ダーイフ：「ジョゼフ＝シャルル・マルドリユスの手稿『カルネ』の翻刻プロジェクト」, 『筑波大学フランス語・フランス文学論集』, 第 37 号, pp. 84-98, 2022.
- [Oda et Daaïf 2026] Jun'ichi Oda et Lahcen Daaïf, *Carnets manuscrits de Joseph-Charles Mardrus (traducteur des Mille et Une Nuits): Carnet-A*, Institut de recherches sur les Langues et les Cultures d'Asie et d'Afrique, Université des Études Étrangères de Tokyo (ILCAA) ; Centre Interuniversitaire d'Histoire, Archéologie, Littératures des mondes chrétiens et musulmans médiévaux (CIHAM), Université Lyon 2, 2026, in press.