

Projet Loracraftt – Une architecture client-serveur pour traduire des textes hiéroglyphiques mot par mot avec un smartphone

Didier DTL Morandi
21 janvier 2025

Résumé

Les solutions combinant CNN et mécanismes d'attention sont maintenant arrivées à maturité pour la reconnaissance et la classification des hiéroglyphes égyptiens, idem pour les algorithmes de traduction assistée par ordinateur. Mais la traduction mot à mot nécessite un corpus dont les smartphones actuels ne pourront disposer au vu de sa taille présumée et souhaitée. C'est pourquoi nous proposons, dans le projet Loracraftt, une architecture client-serveur qui permettra le traitement de la lecture, la reconnaissance et translittération des textes sur le smartphone et la traduction et la gestion des textes de référence sur un serveur distant, communiquant de façon synchrone avec le client par un protocole WebSocket-TCP.

1. Introduction

L'année 2022 fut l'année d'une double célébration. Celle de la « compréhension » de l'écriture hiéroglyphique par Jean-François Champollion en 1822 et celle de la fabuleuse découverte de la tombe de Toutankhamon par Howard Carter exactement cent ans après. Il suffit de parcourir la *Grammaire égyptienne* conçue par Champollion (Champollion 1836) et éditée par son frère Jacques-Joseph Champollion-Figeac à titre posthume pour comprendre pourquoi la traduction des textes hiéroglyphiques du Moyen Empire est si compliquée et certainement pas à la portée des touristes profanes qui viennent visiter les sites remarquables de l'Égypte.

L'arrivée des ordinateurs personnels dans les années 80 et celle des langages orientés objets, tel *Python* en 1991, combinés à de nouvelles techniques de programmation ayant montré leur utilité pour la reconnaissance de formes et la traduction automatisée - tels notamment les réseaux de neurones convolutifs et les mécanismes d'attention - a permis le développement de projets informatiques destinés à la traduction des hiéroglyphes. Citons pour mémoire le précurseur *Tomb Reader* de Morris Franken et Jan van Gemert (Franken 2015), *Fabricius* par le tandem Google/Ubisoft (Google 2017), *Hieroglyphs AI* de Evgeniy & Alexander Sulimov (Sulimov 2020), *Pixoglyphe* de Fleur Brun *et al.* (Brun 2020), *DeepScribe* de Krishnan *et al.* (Krishnan 2020), *GlyphNet* de Andrea Barucci *et al.* (Barucci 2022) et *Loracraftt*¹ de l'auteur (Morandi 2022).

Contrairement aux précédents, le projet *Loracraftt* entend traduire les textes mot à mot et non pas signe après signe hiéroglyphique. Mais la nécessité de rendre disponible au logiciel embarqué un corpus qui sera essentiellement composé de dictionnaires, tels ceux de Erman (1926), Faulkner (1962), Hannig (2003, 2006) et *VégA*² (2015), plus un grand nombre de textes de référence contenant translittérations et traductions (dans un premier temps en anglais, puis en français et en allemand) tels ceux des sites *Thesaurus Linguae Aegyptiae*³ et *Projet Ramses*⁴, impose le choix de conserver toutes ces données sur un serveur distant doté de grandes capacités de stockage, d'où la nécessité de mettre en place une architecture client-serveur, architecture que nous présentons ci-après.

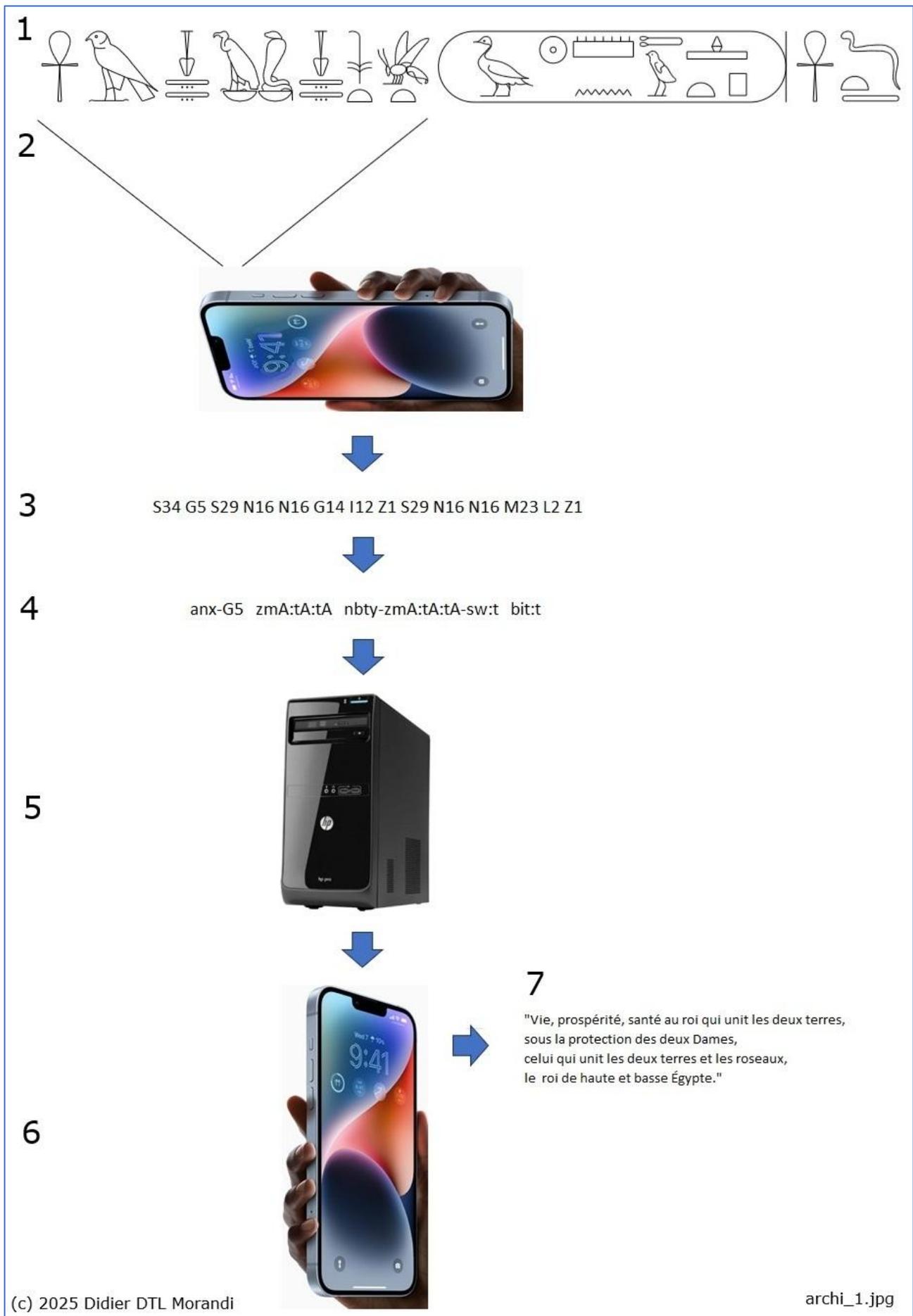
¹ <https://www.shpylgoreih.fr/loracraftt.htm>

² <https://vega-vocabulaire-egyptien-ancien.fr/>

³ <https://thesaurus-linguae-aegyptiae.de/home?lang=en>

⁴ <http://www.egypto.ulg.ac.be/Ramses.htm>

2. Architecture



L'image page précédente montre :

1. l'inscription à traduire
2. l'utilisation d'un smartphone pour capturer un groupe de signes (appel de l'application, prise d'une photo et déclenchement manuel du traitement local⁵)
3. identification des signes et classification en codes Gardiner
4. translittération
5. envoi automatique de la chaîne de caractères translittérés au serveur via une connexion synchrone de type WebSocket⁶
6. réception par le smartphone du texte traduit mot à mot
7. lecture sonore de la traduction affichée à l'écran dans la langue choisie au paramétrage de l'application

a) Description détaillée de l'architecture

Le projet nécessite un smartphone (ou une tablette) avec l'application *Loracraftt* installée dessus, il est donc prévu de développer deux versions de l'application : une pour iOS et une pour Android.

De façon idéale, aucune taille minimum pour la capacité des équipements ne devra être imposée afin d'éviter d'obliger l'utilisateur à un achat coûteux⁷.

L'application *Loracraftt* se chargera en local de :

- prendre les photos
- reconnaître les signes
- les classer en codes Gardiner
- les translittérer
- envoyer le résultat au serveur pour traduction
- recevoir la traduction
- diffuser la traduction à l'utilisateur (texte et son)

Le serveur, qui contiendra la base de données et le logiciel de traduction, sera mis gracieusement à la disposition des utilisateurs depuis le site Web de *Loracraftt*⁸.

⁵ Une traduction au fil de l'eau par balayage n'est pas prévue dans la version 1 du logiciel.

⁶ Nous pensons qu'une liaison de type HTTPS, par définition asynchrone, ralentira considérablement la communication entre le serveur et le client, outre qu'elle ne permettrait pas au serveur de transmettre des données au client tant que celui-ci ne les demande pas.

⁷ Pour information, à la date de publication du présent document, un iPhone 16 Pro a une capacité maximum de 1 To (1024 Giga octets) et coûte autour de 1 200 euros.

⁸ Il n'est pas prévu de produire une application qui serait commercialisée. Ce travail est un travail de recherche et les sources et ressources seront postées sur GitHub.

La connexion synchrone WebSocket-TCP sera initialisée au lancement de l'application et maintenue en permanence, avec un dispositif de restauration automatique en cas de coupure tant que l'application est active (perte du réseau GSM par exemple).

b) Évolution

Une évolution en cours d'étude consisterait pour le serveur à rechercher dans son corpus les occurrences par exemple des quatre premiers mots traduits, réglage paramétrable à quatre mots minimum et plus.

En effet, un exercice réalisé avec le moteur de recherche de Google a donné les résultats suivants :

chaîne recherchée	nombre de mots	nombre d'occurrences
« recherche » ⁹	1	1 330 millions
« recherche de »	2	139 millions
« recherche de trois »	3	12,1 millions
« recherche de quatre mots »	4	9

Ensuite, le serveur irait interroger sa base de données afin de rechercher les textes qu'il connaît et qui comportent ces mots traduits ensemble, ce qui pourrait donner le résultat suivant :

Translittération reçue

Traduction

Wsjr jt n=k msddw Wnjs nb.w Osiris, saisis-toi de quiconque déteste Ounas

Résultat de la recherche dans le corpus (extrait)

*Wsjr jt n=k msddw Wnjs nb.w Osiris, saisis-toi de quiconque déteste Ounas
mdw m rn=f dw et de celui qui déshonore son nom !
Dḥwtj j.sb jt sw n Wsjr Thot, pars, saisis celui qui menace Osiris,
jn mdw m rn n(j) Wnjs dw emporte celui qui déshonore le nom d'Ounas,
d n=k sw m dr.t=k et mets-le dans ta main !*

Envoi du serveur au smartphone :

« J'ai identifié que les six premiers mots que j'ai traduits font notamment partie d'un texte gravé sur le mur nord du passage entre la chambre funéraire et l'antichambre¹⁰ de la pyramide d'Ounas à Saqqarah, mais également au premier registre de la paroi nord¹¹ de la chambre funéraire. Veux-tu que je t'en lise un extrait ? » (etc.)

⁹ Les guillemets sont obligatoires pour permettre une recherche des mots de façon contiguë.

¹⁰ Spruch 23, § 16a [W/F-A/N 7] et 16b [W/F-A/N 9].

¹¹ [W/F/Ne I 1-4].

3. Une avancée notable depuis 2022

Le 16 mars 2023, Asmaa Sobhy, Mahmoud Helmy, Michael Khalil, Sarah Elmasry, Youtham Boules et Nermin Negied, de la *Digital Egypt Builders Initiative*¹², ont publié un article intitulé « *Un traducteur automatique basé sur l'IA pour la langue hiéroglyphique ancienne - Des images numérisées au texte anglais* » (Sohby 2023), document dans lequel ils écrivent dans l'introduction : « Dans ce travail de recherche, notre objectif est de déchiffrer cette langue remarquablement intéressante pour permettre aux touristes de comprendre plus facilement les écritures des anciens Égyptiens grâce à la détection et à la reconnaissance automatiques des hiéroglyphes puis à leur traduction en anglais. »

Mais cette étude n'a – semble-t-il – donné lieu à aucune mise en application.

Le 12 janvier 2025, l'auteur a publié un premier article (Morandi 2025) sur le projet *Loracraftt*, nous espérons que la présente publication continuera à motiver les chercheurs sur le sujet.

4. Conclusion

On voit que la traduction automatisée des textes égyptiens hiéroglyphiques du Moyen Empire est arrivée à maturité, reste à rassembler une équipe de développeurs pour la mettre en pratique.

Égyptologues et informaticiens, le *Projet Loracraftt* vous parle ? Parlons-en¹³.

Remerciements spéciaux à Alexander Sulimov, Amandine Marshall, Andrea Barucci, Arnold McDonald, Bernard Mathieu, Jan van Gemert, Jean Winand, Jean-Yves Empereur, Maryline Sellier, Morris Franken, Nagai Masakatsu, Raymond Monfort, Renaud de Spens, Sébastien Biston-Moulin et Serge Rosmorduc pour leur assistance, conseils et encouragements.

5. Références

Bonnamy 2013. *Dictionnaire des hiéroglyphes*, Actes Sud, 992 pages.

<https://www.fnac.com/a6076114/Yvonne-Bonnamy-Dictionnaire-des-hieroglyphes>

Brun 2020. *Pixoglyphe* – Détection et reconnaissance de hiéroglyphes. [en ligne]

https://www.shpylgoreih.fr/documents/Brun_Rapport_Pixoglyphe_projet_VEGA_2020.pdf

Champollion 1836. *Grammaire égyptienne*, Firmin Didot, Paris. [en ligne]

https://www.shpylgoreih.fr/documents/Champollion_Grammaire_egyptienne_1836.pdf

Erman 1926. *Wörterbuch der Aegyptischen Sprache*, Akademie-Verlag, Berlin. [en

ligne] https://www.shpylgoreih.fr/documents/Erman_Worterbuch_der_Aegyptischen_Sprache_1926.pdf

¹² <https://debi.gov.eg/>

¹³ didier.morandi@gmail.com - <https://www.shpylgoreih.fr/>

- Faulkner 1962. *Concise Dictionary of Middle Egyptian*, Griffith Inst. Oxford. [en ligne] https://www.shpylgoreih.fr/documents/Faulkner_Concise_Dictionary_of_Middle_Egyptian_1962.pdf
- Franken 2015. *Tomb Reader* – Creating an App that lets you read hieroglyphs. [en ligne] <https://tombreaderapp.com/>
- Google 2017. *Fabricius* – Décoder les hiéroglyphes égyptiens avec l'apprentissage automatique. [en ligne] <https://artsandculture.google.com/experiment/fabricius/gwHX41Sm0N7-Dw>
- Hannig 2003. *Ägyptisches Wörterbuch*, tome 1, Verlag, Mainz-am-Rhein. [en ligne] https://www.shpylgoreih.fr/documents/Hannig_Aegyptisches_Worterbuch_I_2003.pdf
- Hannig 2006. *Ägyptisches Wörterbuch*, tome 2, Verlag, Mainz-am-Rhein. [en ligne] https://www.shpylgoreih.fr/documents/Hannig_Aegyptisches_Worterbuch_II_2006.pdf
- Krishnan 2020. *DeepScribe* - Localization and Classification of Elamite Cuneiform Signs Via Deep Learning. [en ligne] <https://voices.uchicago.edu/ochre/project/deepscribe/>
- Molen 2000. *A Hieroglyphic Dictionary of Egyptian Coffin Texts*, Brill. [en ligne] https://www.shpylgoreih.fr/documents/Molen_Hieroglyphic_Dictionary_of_Egyptian_Coffin_Texts_2000.pdf
- Morandi 2022. *Loracrafft* – Lecture, classification, translittération et traduction de textes hiéroglyphiques assistées par ordinateur. [en ligne] <https://www.shpylgoreih.fr/loracrafft.htm>
- Morandi 2025. *Traduction des hiéroglyphes assistée par ordinateur : Les CNN méritent votre plus grande attention*. [en ligne] https://www.shpylgoreih.fr/documents/Morandi_CAHT_Evolution_2025.pdf
- Nederhof 2024. *Hieroglyphs OCR - Transcribe image of hieroglyphic text to Unicode*. [online] <https://github.com/nederhof/hocr>
- Sohby 2023. *An AI-Based Automatic Translator for Ancient Hieroglyphic Language - From Scanned Images to English Text*. [en ligne] https://www.shpylgoreih.fr/documents/Sobhy_An_AI_Based_Automatic_Translator_for_Ancient_Hieroglyphic_Language_2023_Fr.pdf
- Sulimov 2020. *Hieroglyphs AI*. Automatic identification of Gardiner codes. [en ligne] <https://www.hieroglyphs.ai/>