NLP Methods and the Ancient Egyptian Language in its Different Stages: A Survey from a Diachronic Perspective

Lisanne Rüh

Structure

- Introduction and Motivation
- Evolution of the Ancient Egyptian Language
- Research works on different tasks:
    - Transliteration
    - Alignment
    - Translation
    - Character Recognition
    - Combined Tasks
- Discussion and Conclusion

Introduction and Motivation

- Long history of ancient Egypt ➜ language change
- Support research, make ancient texts accesible with NLP methods
- Possible tasks:
  - Classify/date texts
  - Character detection in images
  - Transliteration
  - Translation to modern language
  - Digitization
  - Alignment

Evolution of the Ancient Egyptian Language

Old Egyptian (OE): c. 3000-2000 BC

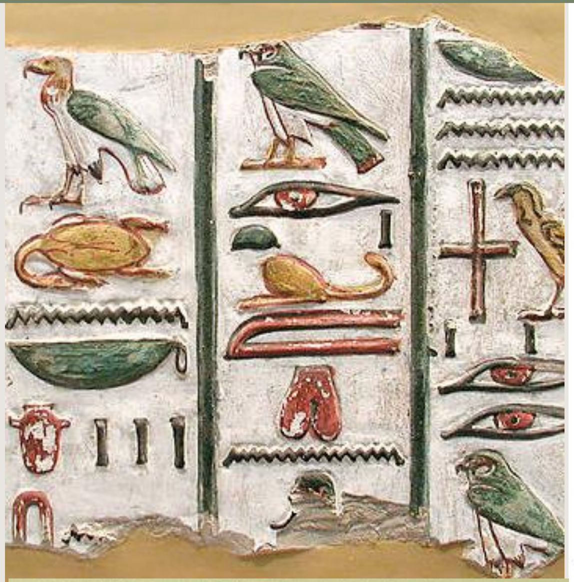
Middle Egyptian (ME): c. 2000-1300 BC

Late Egyptian (LE): c. 1300-700 BC

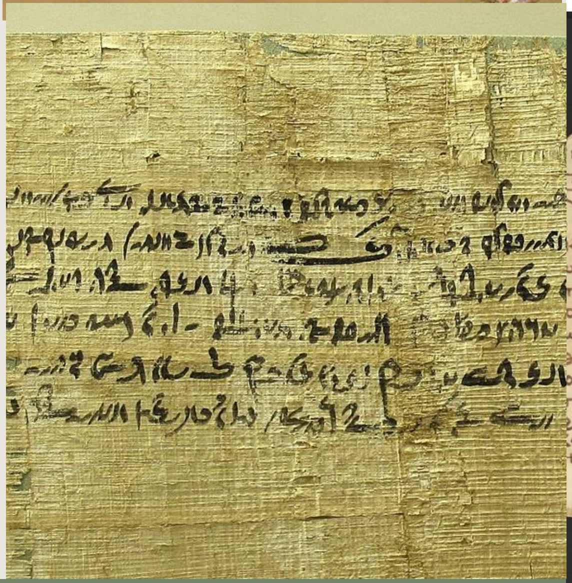
Demotic: c. 700 BC – 500 AD

Coptic : from c. 300 AD

Hieroglyphs



Hieratic



Demotic

Coptic

Hieroglyphic example word

- Sign functions: phonograms, determinatives
- Character representation with code from Gardiner’s sign list
- No word separation markers
- Variable reading order

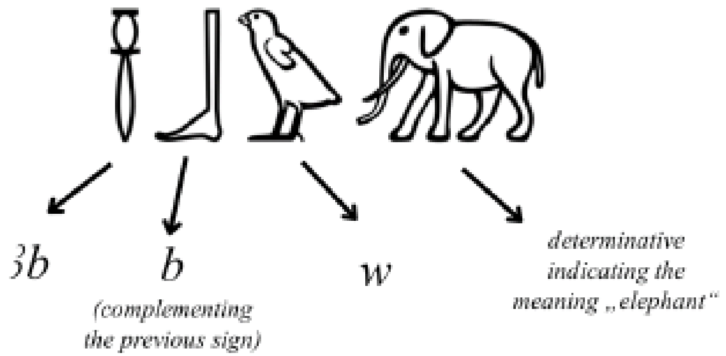
Hieroglyphs	
Gardiner codes	U23 D58 G43 E26
Transliteration	*ꜣbw*
Translation	elephant

Table 1: Hieroglyphic example word with its corresponding Gardiner codes, transliteration and English translation.

Research works on different tasks

Paper	Character Recognition	Word Segmentation	Transliteration	Translation	Other
Rosmorduc (2008)		x	x		
Nederhof (2008)					Alignment
Franken and van Gemert (2013)	x	(x)			
Gohy et al. (2013)					Genre Classification
Elnabawy et al. (2018)	x			x	
Wiesenbach and Riezler (2019)		(x)	(x)	x	
Barucci et al. (2021)	x				
Moustafa et al. (2022)	x			x	
Mohsen et al. (2023)	x				
Sobhy et al. (2023)	x	x	x	x	

Table 2: Overview of the presented papers and their included tasks, (x): implicitly included, ?: unclear.

Transliteration: Rosmorduc (2008)

- Finite-state-transducer with rewriting rules for ME
- Replacing input with sign values, combine into simple groups, then words ➜ use combination with least cost of rules
- Results:
  - Metric: ratio of incorrect words
  - Tale of the Shipwrecked Sailor (ME): 9% error rate
  - Papyrus Westcar (ME): 18% error rate ➜ unknown sign groups
  - Instruction to Amenemope (LE): worse performance

Alignment: Nederhof (2008)

- Orthographic model for aligning hieroglyphs with their transliteration
- Mapping of input sign sequence to possible reading(s) with annotated list ➜ match against words in transliteration
- Penalties for violations ➜ final configuration
- Limitation: sign list ➜ non-standard readings cause errors
- Results:
  - Metric: ratio of wrongly aligned words
  - Tale of the Shipwrecked Sailor (ME): 1% error rate
  - Papyrus Westcar (ME): 3% error rate
  - No error propagation, small differences irrelevant ➜ robustness

Genre Classification: Gohy et al. (2013)

- Classify LE texts by genre ➜ 7 classes (e.g. letter, monumental or administrative texts)
- Data: 332 texts from Ramses corpus ➜ unbalanced
- Results:
  - Naive Bayes with lemmata: 84.3% accuracy
  - SVM: 80.6% accuracy with lemmata, 64% with verbal inflexions
  - Segment & Combine (learn classes of multiple sequences for each input): 67% accuracy with lemmata, 53.4% with POS
  - Study register variation between genres ➜ similarity of vocabularies

Translation: Wiesenbach and Riezler (2019)

- Aim: direct translation from hieroglyphs
- Multi-task setups:
  - Main task: hieroglyphs to German
  - Assistance tasks: transliteration to German, hieroglyphs to transliteration, hieroglyphs to POS tags
- Encoder/decoder sequence-to-sequence systems with attention
- Data: OE and ME sentences from Thesaurus Linguae Aegyptiae, about 29k with and 62k without hieroglyphic encoding
- Improved word segmentation with joint learning

System	BLEU
Baseline: hieroglyphs ➜ German	19.77
Upper bound: transliteration ➜ German	27.67
Baseline + transliteration (input)	+2
Baseline + transliteration, POS tags (output)	+3

Character Recognition: Franken and van Gemert (2013)

- Detect hieroglyphs from images
- Sign segmentation: saliency- based text detection algorithm
- Sign classification: visual matching with different image descriptors
- Additional LMs: lexicon lookup, n-gram based probabilities ➜ context about neighbouring signs

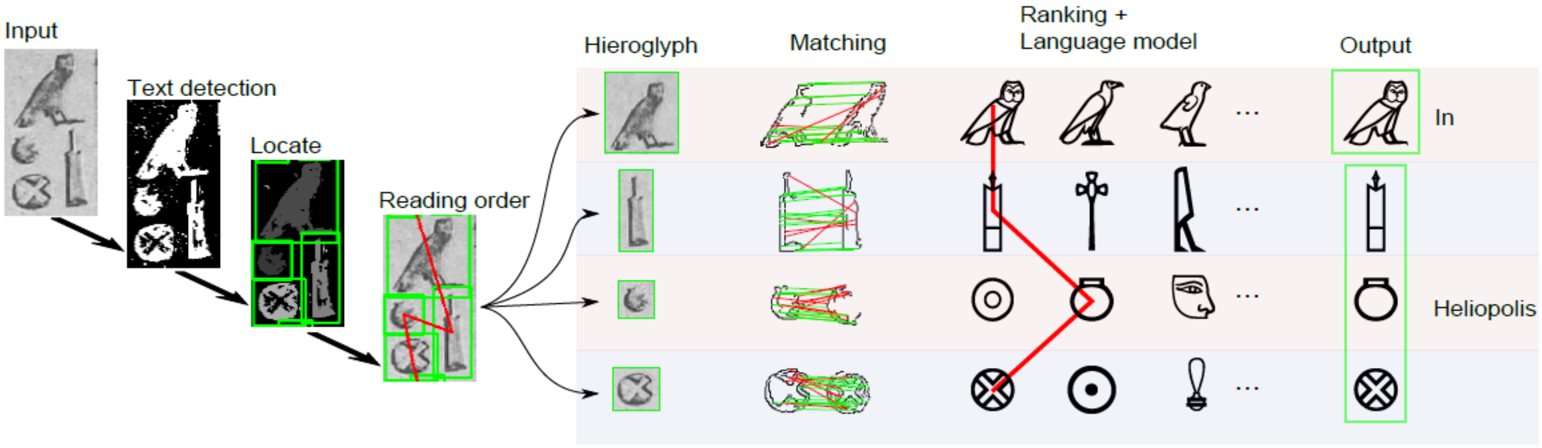


Figure 2: Pipeline for hieroglyph recognition. The 3rd output hieroglyph from the top is corrected by the language model in order to find the word 'Heliopolis' (birth-city of Unas).

Character Recognition: Franken and van Gemert (2013)

- Data:
  - image dataset: inscriptions with OE hieroglyphs from pyramid of Unas, 171 different characters, 4k total annotated hieroglyph images
  - textual corpus: 158 OE pyramid texts

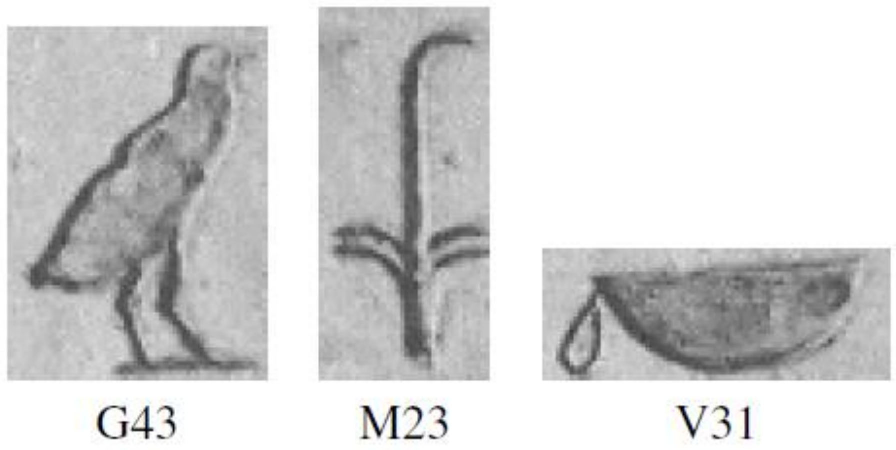


Figure 1: Example images with their labels from the Franken and van Gemert (2013) hieroglyph dataset.

Character Recognition: Franken and van Gemert (2013)

- Results:
  - Segmentation: 83% recall, 85.5% precision
  - Classification: 90% accuracy (manual seg.), 70% accuracy (automatic seg.)
  - LM: -5% with lexicon, small improvements with n-grams

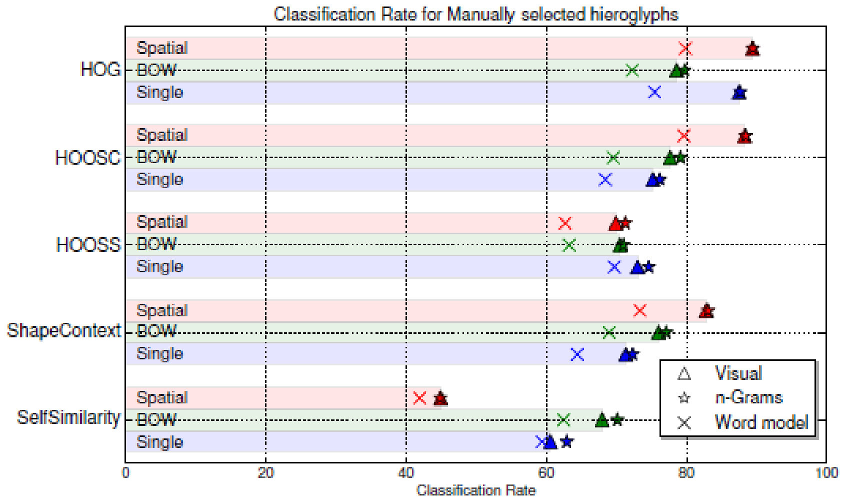


Figure 5: Results for manually cut-out hieroglyphs. The average score is $74 \pm 1\%$.

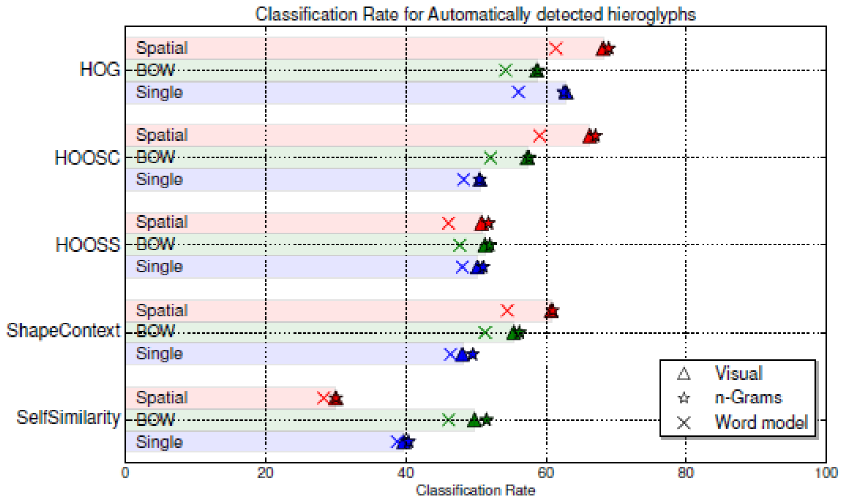


Figure 6: Results for automatically detected hieroglyphs. The average score is 53 ± 5.

Character Recognition: Barucci et al. (2021)

- Hieroglyph classification from images
- Testing different existing CNNs: ResNet-50, Inception-v3, Xception ➜ pre-trained + fine-tuning on hieroglyphs vs. training from scratch
- New CNN for hieroglyphs: Glyphnet
- Data: Franken dataset + own data (total 40 characters) ➜ augmentation, final set with 7k images
- Results:
  - Training from scratch better: 94.5% vs. 90.6% accuracy for ResNet-50
  - Glyphnet better and faster: 97.6% accuracy, 96.8% F1-score
  - Augmentation improves classification: about 3% for Glyphnet

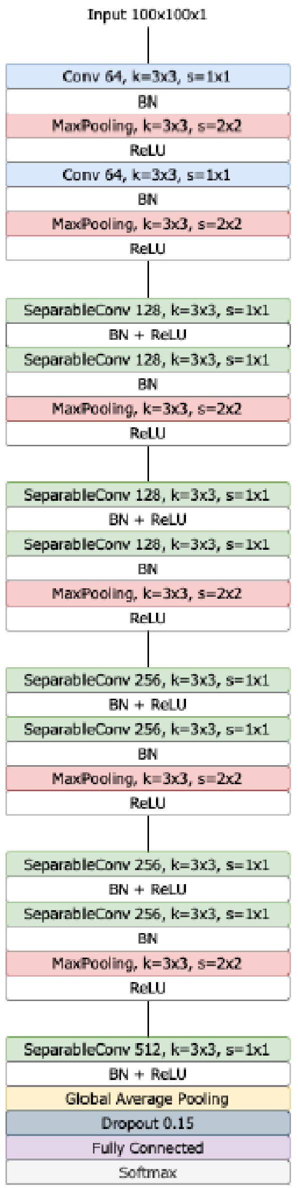


FIGURE 8. The proposed Glyphnet architecture.

Combined Tasks: Elnabawy et al. (2018)

- Character recognition from images: segmentation, Histogram of Oriented Gradients (HOG) feature matching for classification
- Translation to English: translations for individual characters from table (unknown origin) ➜ highly questionable approach
- Data: not specified
- Evaluation:
  - Only small set of characters
  - Comparison with system for Chinese character recognition
  - 66,64% overall accuracy

➜ approach and results not convincing

TABLE III
STATISTICS ON DIFFERENT APPROACHES FOR HIEROGLYPHS CHARACTER RECOGNITION

Hieroglyph Gardiner's code	Proposed work	Liu et. al [32]
F12	80%	65%
F13	73%	59%
F16	33%	10%
H06	60%	67%
L01	65%	49%
M01	52%	63%
M03	76%	69%
N37	87%	61%
N41	54%	59%
P13	68%	43%
P98	85%	63%

Combined Tasks: Mohsen et al. (2023)

- Separating hieroglyphs from image background
- Classification: lightweight CNN SqueezeNet (pre-trained on ImageNet), EfficientNet
- Translation to English/Arabic with Google Translate API ➜ unclear
- Create mobile application „Aegyptos“, additional pronunciation feature ➜ questionable, no phonetic reconstruction of earlier language stages
- Data: Elnabawy et al. dataset (994 images, 42 characters), additional own dataset (60k images, 1k characters), Luxor and Aswan dataset ➜ origin and collection methods unclear
- Results:
  - Classification: 95% accuracy for SqueezeNet, 82.1% for EfficientNet
  - Translation: 89% accuracy ➜ evaluation process unclear

Combined Tasks: Moustafa et al. (2022)

- Classification on single hieroglyph images with lightweight CNNs (MobileNet, ShuffleNet, EfficientNet)
- Want to include English/Arabic translation to mobile application „Scriba“ ➜ no approaches for implementation
- Data: subset of Franken dataset (11k images after augmentation), general images (5k) and cartouches (1k) from pyramid of Unas, temples in Luxor and Aswan, Google Search results

Fig. 13: Royals Dataset Sample

- Results:
  - MobileNet and EfficientNet better than ShuffleNet
  - Augmentation improves classification a lot ➜ 95% accuracy for MobileNet
  - Near perfect accuracy for own data ➜ cartouches only 10 classes

Combined Tasks: Sobhy et al. (2023)

- Glyph detection on images: R-CNN (pre-trained on ImageNet)
- Classification: pre-trained ResNet-50, hierarchical classification, Siamese network + character-level LM (considering previous two characters)
- Word segmentation + transliteration: forward/reverse maximum matching on dictionary, Sentencepiece subword tokenizer trained on Gardiner code sequences
- Translation to English: transformer model with Wiesenbach and Riezler's hyperparameters

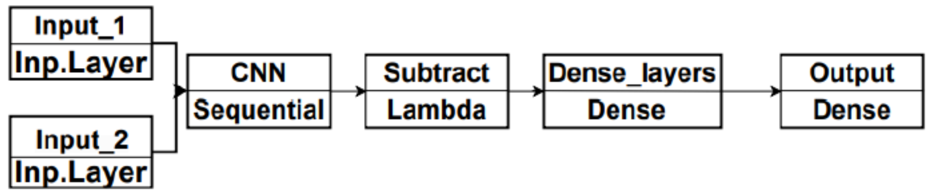


FIGURE 9. The Siamese network's architecture.

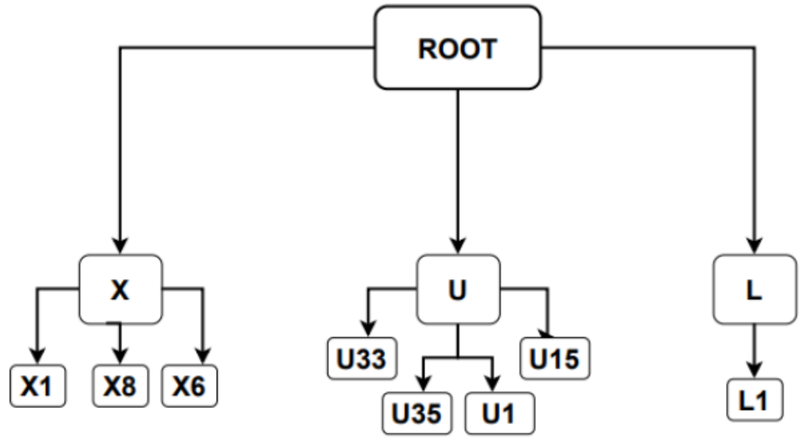


FIGURE 8. Part of the hierarchical classification.

Combined Tasks: Sobhy et al. (2023)

- Image data:
  - Glyph detection: 2k randomly cropped images from Franken dataset
  - Classification: 8k augmented images from Franken dataset
- Text data:
  - Transliteration: dictionary of 10k words with Gardiner codes and transliteration
  - Translation: GitHub corpus (ME sentences with transliteration and English) and HuggingFace dataset (sentences with transliteration and German, English obtained with MT) ➜ different variants/augmentations
- Evaluation for end-to-end process (image ➜ translation) missing

Combined Tasks: Sobhy et al. (2023)

- Results:
  - Glyph detection: 95.9% precision, 74.4% recall
  - Best translation setting: 59.1 corpus-level BLEU on non-augmented single data source
  - Text augmentation: synonym replacements close to best setting, (back)translations with MT misleading

Classification	**Accuracy**
ResNet-50	72%
Hierarchical classification	68%
Siamese network: • One reference per class • Three references per class • + Character-level LM	 85% 87.5% 88.5%
Segmentation + Transliteration	
Forward maximum matching	60%
Sentencepiece	20%

Discussion

- Most works focus on images and character recognition, only few efforts on translation
- Challenge: sparcity of digital, annotated data ➜ augmentation useful
- Approaches mainly for early periods (OE and ME) and scripts (hieroglyphs), language change not considered ➜ potential for diachronic studies
- ML requires less Egyptological knowledge ➜ can cause confusion

Conclusion

- Research on NLP methods for ancient Egyptian languages currently still in early exploration phase
- Certain tasks, language stages and scripts under-represented, diachronic aspects not considered yet ➜ potential for future work
- Promising results for future academical and practical use

References

Andrea Barucci, Costanza Cucci, Massimiliano Franci, Marco Loschiavo, and Fabrizio Argenti. 2021. A deep learning approach to ancient egyptian hieroglyphs classification. *IEEE Access*, 9:123438–123447.

Jaime Duque-Domingo, Pedro Javier Herrera, Enrique Valero, and Carlos Cerrada. 2017. Deciphering egyptian hieroglyphs: Towards a new strategy for navigation in museums. *Sensors*, 17(3).

Reham Elnabawy, Rimon Elias, and Mohammed Salem. 2018. Image based hieroglyphic character recognition. In *2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS),* pages 32–39.

Morris Franken and Jan C. van Gemert. 2013. Automatic egyptian hieroglyph recognition by retrieving images as texts. In *Proceedings of the 21st ACM International Conference on Multimedia, MM '13*, page 765–768, New York, NY, USA. Association for Computing Machinery.

Stéphanie Gohy, Benjamin Martin Leon, and St´ephane Polis. 2013. Automated text categorization in a dead language. the detection of genres in late egyptian. In *Texts, Languages & Information Technology in Egyptology. Selected papers from the meeting of the Computer Working Group of the International Association of Egyptologists (Informatique & E´gyptologie), Li`ege, 6-8 July 2010*. F.R.S.-FNRS - Fonds de la Recherche Scientifique [BE], Presses Universitaires de Li`ege.

Hanna Jenni. 2010. *Lehrbuch der klassisch-ägyptischen Sprache*. Schwabe, Basel.

Shehab ElDin Mohsen, Reem Mansour, Ahmed Bassem, Basma Dessouky, Samira Refaat, and Taraggy M Ghanim. 2023. Aegyptos: Mobile application for hieroglyphs detection, translation and pronunciation. In *2023 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC),* pages 1–8.

Ragaa Moustafa, Farida Hesham, Samiha Hussein, Badr Amr, Samira Refaat, Nada Shorim, and Taraggy M. Ghanim. 2022. Hieroglyphs language translator using deep learning techniques (scriba). In *2022 2nd International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*, pages 125–132.

References

Mark-Jan Nederhof. 2008. Automatic alignment of hieroglyphs and transliteration. In *Information Technology and Egyptology in 2008, Proceedings of the meeting of the Computer Working Group of the International Association of Egyptologists (Informatique et Egyptologie),* pages 71–92, Vienna. Gorgias Press.

Mark-Jan Nederhof. 2015. OCR of handwritten transcriptions of ancient egyptian hieroglyphic text. In *Altertumswissenschaften in a Digital Age: Egyptology, Papyrology and beyond*.

Serge Rosmorduc. 2008. Automated transliteration of egyptian hieroglyphs. In *Information Technology and Egyptology in 2008, Proceedings of the meeting of the Computer Working Group of the International Association of Egyptologists (Informatique et Egyptologie),* pages 167–183, Vienna. Gorgias Press.

Serge Rosmorduc. 2015. Computational linguistics in egyptology. *UCLA Encyclopedia of Egyptology,* 1(1).

Asmaa Sobhy, Mahmoud Helmy, Michael Khalil, Sarah Elmasry, Youtham Boules, and Nermin Negied. 2023. An ai based automatic translator for ancient hieroglyphic language—from scanned images to english text. *IEEE Access*, 11:38796–38804.

Thea Sommerschield, Yannis Assael, John Pavlopoulos, Vanessa Stefanak, Andrew Senior, Chris Dyer, John Bodel, Jonathan Prag, Ion Androutsopoulos, and Nando de Freitas. 2023. Machine Learning for Ancient Languages: A Survey. *Computational Linguistics,* 49(3):703–747.

Philipp Wiesenbach and Stefan Riezler. 2019. Multitask modeling of phonographic languages: Translating middle Egyptian hieroglyphs. In *Proceedings of the 16th International Conference on Spoken Language Translation*, Hong Kong. Association for Computational Linguistics.