Formal Machine Interpretation for the Semasiographic Mixtec Codices of Precolonial and Early Colonial Mesoamerica

Christopher Driggers-Ellis*, Gabriel Ayoubi, Girish Salunke, Christan Grant

University of Florida Data Studio | ufdatastudio.com | driggersellis.cw@ufl.edu *

Abstract

The precolonial and early colonial Mixtec codices describe the history and stories of the region in a semasiographic medium that is full of symbolic representations and meant to be narrated. Recently, the community has investigated hierarchical representation of related media, including Aztec codices and Mayan hieroglyphic script, in a step towards symbolic machine interpretation of these historic Mesoamerican artifacts. In this work, we propose formal symbolic machine interpretation of XML encodings representing facsimile images from the Mixtec Codex Zouche-Nuttall. We demonstrate the efficacy of symbolic machine interpretation from XML step-by-step, showing how our parser and interpreter process text capturing a scene from the Mixtec Codex Zouche-Nuttall. We hope our contribution and the example we provide motivate collaboration among the archaeological, historical, linguistic, and natural language processing research communities to apply machine interpretation to Mixtec codices and similar manuscripts.

History

In this work, we investigate symbolic interpretation of XML representations of scenes from Mixtec Codex Zouche-Nuttall (CZN). The CZN was first published in facsimile by Zelia Nuttall in 1902 [1], [2] and has been the subject of human interpretation for more than a century since. The CZN manuscript has been dated to the 14th or 15th centuries [2], but there are numerous examples of similar manuscripts from both precolonial and colonial Mesoamerica and various cultures in the region, including the Aztec and Maya [3], [4], [5].

Glyph Types

Figure 1: The four-part figure above shows the four glyph types we concern ourselves with in the interpretation of Mixtec codices. *Top-Left:* Male (a) and Female (b) Human Glyphs. *Top-Right:* Name-Date (a) and Year (b) Glyphs. *Bottom-Left:* A Toponym. *Bottom-Right:* The whole picture. Multiple glyph types appear in concert to convey a marriage in Page 26 of the CZN (obverse) (Examples provided courtesy of the British Museum).

Symbolic Interpretation

Related works have initiated encodings of manuscripts from Mesoamerican cultures, including Aztec and Mayan codices [4], [5], [6]. Perri et al. explore encoding of agglutinative pictoral elements in the Aztec Codex Mendoza [4], [6], and de la Iglesia et al. develop an XML-based TEI representation schema for Mayan hieroglyphics [5].

Glyph Classification

Classification of glyphs into the correct types is essential for creating the XML representations our symbolic contribution handles. Related works have already taken steps in this direction for the Human and Name-Date Glyphs [7], [8]. Webber et al. perform binary classification tasks on human glyphs according to their gender (Male/Female) and their pose (Sitting/Standing) [7]. Salunke et al. perform both binary and multiclass classification tasks necessary to read name-date glyphs [8]. Glyphs are classified according to whether they are name-dates or years, and additional classifiers manage classifying the sign and number of the glyph.

References

[1] Z. Nuttall, *Codex Nuttall; facsimile of an ancient Mexican codex belonging to Lord Zouche of Harynworth, England*. Cambridge, Mass., Peabody Museum of American Archaeology and Ethnology, Harvard University, 1902.

[2] R. L. Williams, *The complete codex Zouche-Nuttall : Mixtec lineage histories and political biographies / by Robert Lloyd Williams ; foreword by Rex Koontz.* (Linda Schele series in Maya and pre-Columbian studies), eng, 1st ed. Austin: University of Texas Press, 2013, ISBN: 9780292744387.

[3] E. H. Boone, *Stories in red and black : pictorial histories of the Aztecs and Mixtecs / Elizabeth Hill Boone.* eng, 1st ed. Austin: University of Texas Press, 2000, ISBN: 0292708769.

[4] A. Perri, L. Perondi, and D. Capo, "Rethinking unicode: How to digitally encode non-linear written artefacts? a tentative encoding of codex mendoza, folio 2r," in *Proceedings of Grapholinguistics in the 21st Century, 2024*, 2024.

[5] M. de la Iglesia et al., "The code of maya kings and queens: Encoding and markup of maya hieroglyphic writing," *Journal of the Text Encoding Initiative*, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:239751206.

[6] L. Fedorova, A. Perri, et al., "Emblematic techniques as textual strategies in non-linear and linear scripts," *GRAPHOLINGUISTICS AND ITS APPLICATIONS*, vol. 9, pp. 75–99, 2024.

[7] A. Webber et al., "Analyzing finetuned vision models for Mixtec codex interpretation," in *Proceedings of the 4th Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP 2024)*, M. Mager et al., Eds., Mexico City, Mexico: Association for Computational Linguistics, Jun. 2024, pp. 42–49. DOI: 10.18653/v1/2024.americasnlp-1.6. [Online]. Available: https://aclanthology.org/2024.americasnlp-1.6.

[8] G. Salunke, C. Driggers-Ellis, and C. Grant, "Classifying name-date and year figures in mixtec codices," in *Computational Humanities Research 2025*, T. Arnold, M. Fantoli, and R. Ros, Eds., Anthology of Computers and the Humanities, 2025, pp. 1210–1218. DOI: 10.63744/eNge3Gj1LSHb.

[9] M. Smith, *Picture Writing from Ancient Southern Mexico; Mixtec Place Signs and Maps* (Civilization of the American Indian series). University of Oklahoma Press, 1973. [Online]. Available: https://books.google.com/books?id=w01vzgEACAAJ.

[10] J. Pohl, *The Politics of Symbolism in the Mixtec Codices* (Vanderbilt University publications in anthropology). Vanderbilt University, 1994, ISBN: 9789354623769. [Online]. Available: https://books.google.com/books?id=DTMNAAAAYAAJ.

DATA STUDIO

UF Herbert Wertheim College of Engineering
Department of Computer & Information Science & Engineering
UNIVERSITY of FLORIDA

Method and Results

```xml
<?xml version="1.0" encoding="UTF-8"?>
<document>
  <scene>
    <year x="0" y="0" symbol="Flint" number="6" />
    <name_date x="0" y="0" symbol="Eagle" number="7" />
    <human x="0" y="0" gender="Lady" pose="sitting" orientation="right" />
    <name_date x="0" y="0" symbol="Eagle" number="9" />
    <object x="0" y="0" identity="house">
    <human x="0" y="0" gender="Lord" pose="sitting" orientation="left" />
    <name_date x="0" y="0" symbol="Alligator" number="6" />
  <scene/>
<document/>
```

```
grammar Scene;

document
    : XML_DECL DOCUMENT_OPEN (SCENE_OPEN sent SCENE_CLOSE)+ DOCUMENT_CLOSE
    ;

sent    : clause
        | OBJ (date clause | clause)
        | date (OBJ clause | clause)
        ;

clause  : clause_f+ (date_tail | obj_tail)? ;
date_tail : date (OBJ clause_f+ | clause_f+)? ;
obj_tail  : OBJ (date clause_f+ | clause_f+)? ;
date    : Y ND? ;
clause_f: H (ND | near_date)? ;
near_date : NEAR_OBJ ND? ;
```

Figure 2: We begin with XML document (top) and Context Free Grammar (bottom) shown above. The ANTLR parser combines the lexing and parsing required to convert the XML in to an AST according the grammar. This AST is illustrated using our example from Page 26 of the CZN (obverse) in the Fig. 3 below.

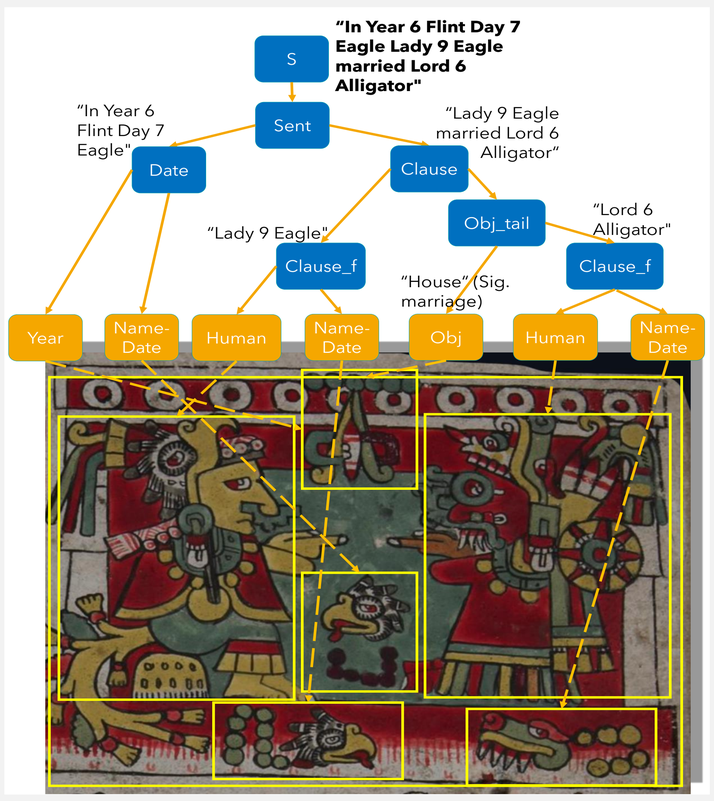

Figure 3: The abstract syntax tree (AST) for our example from Page 26 of the CZN (obverse) (Provided courtesy of the British Museum).

Limitations and Future Work

While we demonstrate a proof of concept in this work, we fully acknowledge numerous considerations that limit its scope. We review them here.

We do not currently have a method of automatically converting codex excerpts into the symbolic XML representations our parser expects. The current parser therefore requires extensive human supervision in the form of curating XML data for parsing.

Also, we do not address the toponyms in Mixtec codices. Mixtec toponym interpretation is quite involved, sometimes relying on clues from archaeology or wordplay in ancient Mixtec language [3], [9], [10]. The uncertainty and complexity of toponym interpretation necessitates specialized investigation and solution.

Use the QR Code to read more about our project.