

Modellierung eines digitalen Zeichenkatalogs für die Hieroglyphen des Klassischen Maya

Ein neues Konzept zur Klassifikation von Schriftzeichen sowie der qualitativen Bewertung und Einstufung von Entzifferungshypothesen

Franziska Diehr¹, Maximilian Brodhun², Sven Gronemeyer³, Katja Diederichs³, Christian Prager³, Elisabeth Wagner³ und Nikolai Grube³

Abstract: Im Projekt “Textdatenbank und Wörterbuch des Klassischen Maya” stellt der Zeichenkatalog eine der wichtigsten Komponenten zur Erstellung des digitalen Textkorpus und dessen linguistischer Analyse dar. Für die noch nicht vollständig entzifferte Schrift und Sprache des Klassischen Maya ist der Zeichenkatalog das essentielle Hilfsmittel zur Identifikation und Klassifikation der im Schriftsystem verwendeten Zeichen. Aufgrund besonderer Eigenschaften der Hieroglyphenschrift und der noch unvollständigen und damit auch auf Hypothesen basierenden Entzifferung ist eine Neukonzeption zur Inventarisierung und Klassifikation von Mayaschriftzeichen notwendig. Herausforderungen bei der Modellierung stellen sich vor allem bei der Erfassung von Zeichen und deren Zuordnung zu Graphemen sowie der qualitativen Bewertung und Einstufung von Entzifferungshypothesen. Dieser Beitrag konzentriert sich auf die Vorstellung des Zeichenkatalogs und seiner Bedeutung für die Erforschung des Schriftsystems des Klassischen Maya. Dabei wird der gewählte Modellierungsansatz erläutert und die Methode der Anforderungsanalyse unter besonderer Betrachtung des Aspekts der interdisziplinären Arbeit in einem Digital Humanities Projekt beschrieben.

Keywords: Digitale Geisteswissenschaften; Ontologiebasierte Modellierung; RDF; CIDOC CRM; Virtuelle Forschungsumgebung; TextGrid; Linguistik; Klassisches Maya; Hieroglyphenschrift

1 Einleitung

Forschungsgegenstand des Projekts “Textdatenbank und Wörterbuch des Klassischen Maya” (TWKM) ist die Hieroglyphenschrift der vorspanischen Mayakultur, die in der Zeit zwischen 500 v. Chr. und 1500 n. Chr. auf dem Gebiet der heutigen Staaten Mexiko, Guatemala, Belize und Honduras verwendet wurde. Ziel des Vorhabens ist die Erschließung der bislang rund 10.000 bekannten Texte und ihrer Schriftträger in einem maschinenlesbaren Korpus. Darauf basierend soll ein Wörterbuch des Klassischen Maya

¹ Staats- und Universitätsbibliothek Göttingen, Metadaten und Datenkonversion, Papendiek 14, 37073 Göttingen, Deutschland, diehr@sub.uni-goettingen.de

² Staats- und Universitätsbibliothek Göttingen, Forschung und Entwicklung, Papendiek 14, 37073 Göttingen, Deutschland, brodhun@sub.uni-goettingen.de

³ Rheinische Friedrich-Wilhelms-Universität, Abteilung für Altamerikanistik, Oxfordstrasse 15, 53111 Bonn, Deutschland, (sgronemeyer|katja.diederichs|cprager|uzsb00|ngrube)@uni-bonn.de

erstellt werden, das sowohl den gesamten Sprachschatz und dessen Verwendung in der Schrift abbildet [No13]. Gefördert wird das bis auf voraussichtlich 2028 angelegte Langzeitvorhaben von der Nordrhein-Westfälischen Akademie der Wissenschaften und Künste. Die an der Universität Bonn angesiedelte Arbeitsstelle kooperiert dabei eng mit der Niedersächsischen Staats- und Universitätsbibliothek Göttingen. In interdisziplinärer Zusammenarbeit werden die Hieroglyphentexte mit geisteswissenschaftlichen Methoden und informationstechnologischen Werkzeugen aufbereitet, ausgewertet und interpretiert [Pr14a]. Das Projekt nutzt dazu die virtuelle Forschungsumgebung TextGrid und passt dafür Komponenten an projektspezifische Bedürfnisse an.

Für die Erstellung des Textkorpus und des Wörterbuchs einer nicht entzifferten Sprache ist ein Zeichenkatalog, ein Inventar aller verwendeten Schriftzeichen, ein unverzichtbares Werkzeug. Dieser Beitrag widmet sich der Beschreibung des im TWKM-Projekt entwickelten Zeichenkatalogs sowie dessen Modellierung und erläutert dabei die Rolle der interdisziplinären Zusammenarbeit in einem in den Digitalen Geisteswissenschaften verorteten Projekts.

2 Kurze Einführung in das Schriftsystem des Klassischen Maya

Die Maya sind eine der fünf Kulturen weltweit, die unabhängig voneinander ein Schriftsystem entwickelten. Im Vergleich zu den übrigen Schrift- und Aufzeichnungssystemen in Mesoamerika ist die Hieroglyphenschrift der Mayakultur das einzige lesbare Schriftsystem, das über einen Zeitraum von rund 2000 Jahren verwendet wurde. Die frühesten Textzeugen entstanden im 3. Jahrhundert v. Chr. Diese Schrifttradition, die in der Klassischen Periode (100 – 810 n. Chr.) ihren Höhepunkt erreichte, dauerte bis zur Ankunft der spanischen Eroberer im 16. Jahrhundert an und wurde bis in das späte 17. Jahrhundert im Untergrund fortgeführt.

Aufgrund des ikonischen Charakters der etwa 650 Zeichen wird das Schriftsystem der Maya als Hieroglyphenschrift bezeichnet. Typologisch handelt es sich um ein logo-syllabisches Schriftsystem, in dem es zwei Zeichenklassen gibt, Silbenzeichen und Logogramme. Logogramme stehen für konkrete sprachliche Begriffe und verweisen mit wenigen Ausnahmen immer nur auf ein Denotat. Die zweite Zeichenklasse sind Vokal- und Silbenzeichen, die offene Silben repräsentieren. Sie dienten zur silbischen Schreibung lexikalischer und grammatischer Morpheme. Darüber hinaus fanden sie als vorangestellte oder angehängte phonetische Komplemente von Logogrammen Verwendung. Dadurch war es möglich, Wörter ausschließlich mit Silbenzeichen oder nur mit Logogrammen zu schreiben. In der Regel wurden Logogramme und Silbenzeichen miteinander kombiniert (Abb. 1). Eine ausgeprägte kalligraphische Komplexität der Texte wurde vor allem durch allographe Notation und mittels allomorpher Repräsentation von Zeichen erreicht. Dies erlaubte den Schreibern, Texte ästhetisch anspruchsvoll und ohne die Wiederholung von Zeichen zu verfassen. Für häufig verwendete Silben gab es mindestens zwei oder mehr

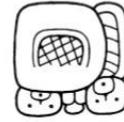
Zeichen. Die Schriftzeichen wurden so miteinander kombiniert, dass sie nahezu rechteckige Blöcke bildeten, wobei ein solcher Hieroglyphenblock wahrscheinlich der emischen Vorstellung eines Wortes entspricht. Aus der Kombination der auf unterschiedliche Satzteile verweisenden Hieroglyphenblöcke miteinander entstanden Sätze. Mehrere Sätze wurden zu komplexen Texten zusammengefügt, deren Syntax und Textstruktur jener modernen Maya-Sprachen ähneln [Pr14b].



PAKAL
Logogramm



PAKAL-la
Logogramm und Silbenzeichen



pa-ka-la
Silbenzeichen

Abb.1: Verwendung von Logogramm und Silbenzeichen [Mo02, 128]

3 Zur Begriffsbestimmung von Graphem, Graph und Zeichen⁴

Die Analyse eines jeden Schriftsystems beginnt mit dem Erstellen einer Liste all seiner Grapheme, der verschriftlichen Form des Zeichens. Grapheme sind auf abstrakter Ebene auf Regeln basierende, kontrastive Einheiten. Allographe stellen die konkret vorkommenden Varianten eines Graphems dar, die diesem zugewiesen werden können. Ein Graph ist hierbei ein konkret realisiertes, jedoch noch keinem Graphem zugewiesenes Schriftzeichen, basierend auf existierenden Definitionen von Graphem und Graph [Co06, 173-175][Cr10, 204]. Wenn ein Schriftzeichen als Graphem klassifiziert wird, dann ist die zugehörige Graphemklasse in diesem Sinne ein nicht vorzufindender, aber anhand diagnostischer Merkmale definierter abstrakter Prototyp, dessen Identität in allen möglichen Ausformungen seiner konkret abgebildeten Allographe erhalten ist. Zur Erstellung eines Grapheminventars werden alle Vorkommen von Graphen in Texten gesammelt und festgehalten [Co06, 174]. Hierbei wird zunächst nach Regeln für die Distinktion von Graphen gesucht, um daraus eine Menge verschiedener Grapheme zu eruieren. Aus dieser begrenzten Menge ergibt sich schließlich das Grapheminventar eines Schriftsystems, welches nach Rogers auch z.B. Numerale, Interpunktion und Diakritika einbezieht [Ro05, 10]. Die Verwendung des Begriffs Zeichen für den TWKM-Katalog begründet sich erstens in der Nutzung der Begriffe Schriftzeichen und Zeichen in ihrer allgemein akzeptierten Verwendung beim Verweis auf Schriftelemente. Weiterhin gibt es auch eine gewachsene Begriffstradition in der Epigraphik, wie etwa die Zeichenkataloge ägyptischer Hieroglyphen. Zweitens sollte eine Begrifflichkeit gewählt werden, die einerseits im Bereich der Linguistik im weitesten Sinne etabliert ist, andererseits außerhalb der Graphematik verwendet wird, und

⁴ Die Definitionen sind im Rahmen des TWKM Working Paper mit dem Arbeitstitel "Standards for Maya Epigraphic Analysis I: Principles of Maya Graphemics" entstanden, das im Laufe des Jahres publiziert wird.

somit nicht im Kontext unserer Nutzung falsch verstanden werden kann. Das Zeichen verstehen wir im Sinne der Semiotik insofern, als es eine Bindung zwischen der Abbildungsebene zur Inhaltsebene ermöglicht, wobei zunächst lediglich seine Inhaltsseite beschrieben und ihm explizit zugewiesen wird. Das geschieht, in dem wir dem Zeichen sogenannte Zeichenfunktionen zuweisen. (Eine ausführliche Beschreibung der Zeichenfunktion erfolgt in Kapitel 6.)

4 Quellenlage und bestehende Zeicheninventare

Es ist sehr wahrscheinlich, dass niemals ein vollständiges Grapheminventar angelegt werden kann, da wir auf Grundlage bisher entdeckter und noch zu entdeckender Textzeugen arbeiten. Dabei sind insbesondere handschriftliche Quellen weitestgehend zerstört, bisher sind hier nur vier Kodizes bekannt. Die Wahrscheinlichkeit ist also hoch, dass Zeichen und deren Grapheme des Klassischen Maya bisher nur unvollständig in ihrer konkreten Form abgebildet und klassifiziert werden konnten und sich bisherige Einordnungen durch neue Funde möglicherweise ändern. In diesem Zusammenhang blicken wir auf eine noch relativ kurze Geschichte der Erforschung der Mayaschrift zurück, deren Durchbruch erst in den 1950er Jahren stattfand. Dennoch brachte sie bisher elf Kataloge für die Inventarisierung und Systematisierung der Hieroglyphen hervor. Dabei haben sich vor allem die Transkriptionskonventionen und Zeichennomenklaturen für Handschriften von Günter Zimmermann sowie der etablierte Katalog von Eric Thompson durchsetzen können. Letzterer stellt trotz fehlerhafter Klassifikationen, Doppelverzeichnungen und Unvollständigkeit des Textbestands bis heute das Standardreferenzsystem für Mayahieroglyphen dar [Pr14b]. Als besonders problematisch erweisen sich vor allem Mehrfachinventarisierungen, bei denen mehrere Allographe eines Graphems als verschiedene Zeichen inventarisiert wurden [Gr90] [Ke62][Ku89][Ri06][RSS96].

Ein Nachteil traditioneller Zeichenkataloge ist, dass sie einerseits fehlerhafte Klassifikationen enthalten und andererseits durch ihre gedruckte Fassung unveränderbar und somit nicht dynamisch erweiterbar sind. Dies verhindert, dass z.B. gegebene Fehlklassifikationen korrigiert werden oder neue Verknüpfungen zu Zeichen und zu verschiedenen Zeichenfunktionen erstellt werden können. Hier kann ein digitaler Zeichenkatalog Abhilfe schaffen, indem er flexibel auf Änderungen reagieren kann und dabei gleichzeitig persistente Identifikationsmöglichkeiten bietet.

5 Anforderungsanalyse für die Modellierung des Zeichenkatalogs

Um einen an die konkreten Bedürfnisse der Mayaschriftforschung angepassten Zeichenkatalog zu realisieren, müssen konkrete Bedarfe an das Metadatenschema, dessen Umsetzung in einem geeigneten Erfassungssystem sowie an die zugrundeliegende technische Umgebung ermittelt werden. Unser Vorgehen ist an die Methode des Experteninterviews

angelehnt, dass laut Reinhold “insbesondere im Kontext der Modellierung von Forschungsprozessen und Forschungsdaten in den Digital Humanities [...] eine adäquate Methode der Informationsbedarfsanalyse dar[stellt], da aufseiten der Forschenden ein hohes Maß an impliziten Wissen zu erwarten ist” [Re15, 330]. Dabei sei aber zu beachten, dass von Experten nur Wissen mitgeteilt wird, wenn der Fragende bereits eine hohe Expertise über das Thema vorweist [Fl07, 218,219]. In unserem Falle bedeutet dies, dass von informationswissenschaftlicher und -technologischer Seite Verständnis über die grundlegenden Konzepte der Linguistik und Sprachforschung zu erlangen ist. Als interdisziplinäres Forschungsvorhaben zielen wir darauf ab, dass alle Beteiligten einen gemeinsamen Wissensstand erreichen. Um zu verstehen wie das domänenspezifische Wissen in der Datenstruktur repräsentiert und es mittels Erfassungssystems abbildet wird, erlangen die Schriftforscher Kenntnisse über Modellierung, Funktionsweise und praktischen Anwendung der Daten und ihrer Strukturierung. Damit wird letztendlich ein professionalisierter Umgang mit den erzeugten Forschungsdaten erreicht.

Der Prozess der Informationsbedarfsanalyse stellt somit den am stärksten interdisziplinär geprägten Bereich in der gemeinsamen Projektarbeit dar. Konkret sind wir dabei in folgender Weise vorgegangen: Zunächst erörterten wir Grundlegendes über den Wissensgegenstand, also den Zeichenkatalog und seiner Funktion im Projektkontext. Anhand des entstandenen Protokolls wurde ein Arbeitspapier erstellt, anhand dessen weitere Bedarfe ermittelt und vorhandene konkretisiert wurden. In einer Phase des intensiven und regelmäßigen Wissensaustauschs wurden einzelne Konzepte besprochen und Fragen zum Gegenstand und dessen Modellierung erörtert. Neben Telefonkonferenzen nutzten wir dafür auch kollaborative, digitale Arbeitsumgebungen wie Etherpads und ein Wiki. Hier wurden Vorschläge zu Definitionen und Strukturierung sowie zur technischen Umsetzung unterbreitet, die dann von anderer Seite auf ihre Plausibilität und Anwendbarkeit geprüft und auf ihre Funktionalität getestet wurden. Nach Rückmeldung wurden die Konzepte ggf. angepasst bzw. weiter optimiert. Abschließend fand die gemeinsame Redaktion der Dokumentation des Metadatenschemas⁵ statt, in dem Definitionen aus dem Bereich der Linguistik und Schriftforschung in Konzepte der Wissensmodellierung integriert wurden.

6 Ein neuer Zeichenkatalog für das Klassische Maya als Ergebnis ontologisch-basierter Modellierung

Mit dem TWKM-Zeichenkatalog führen wir ein grundlegend neues Konzept für die Inventarisierung von Mayaglyphen ein, die eine flexible Zuordnung der Zeichen zu ihren Graphemen ermöglicht. Dies erreichen wir, indem wir Zeichen (Sign) und Graphe (Graph) getrennt voneinander erfassen (Abb. 2). Erst wenn ein Graph mit einem Zeichen verbunden wird (isGraphOf), erfolgt dessen Identifizierung als Allograph des Zeichens. Diese Verbindung ist optional, so dass auch Graphe inventarisiert werden können, die bisher

⁵ Vollständige Dokumentation des Metadatenschemas des Zeichenkatalogs ist einsehbar unter: <http://idiom-projekt.de/catalogue>

noch keinem Zeichen zugeordnet werden konnten. Die Gesamtheit aller identifizierten Allographe bildet das Graphem des Zeichens. Je nach Forschungsstand kann dies flexibel angepasst werden, so dass Allographe durch neue Verknüpfungen hinzugefügt und Falschzuweisungen korrigiert werden können. Das Konzept der getrennten Erfassung von Zeichen und Graph und deren flexibler Zuordnung zueinander schafft dabei Möglichkeiten, denen die gedruckten Kataloge bisher nicht gerecht werden konnten. Das erarbeitete Konzept benötigt eine Datenstruktur, die es erlaubt semantische Relationen zwischen eindeutig referenzierbaren Entitäten herzustellen. Daher stellte sich ein Metadatenschema in RDF-Realisierung als optimale Form der Wissensrepräsentation für den TWKM-Zeichenkatalog heraus.

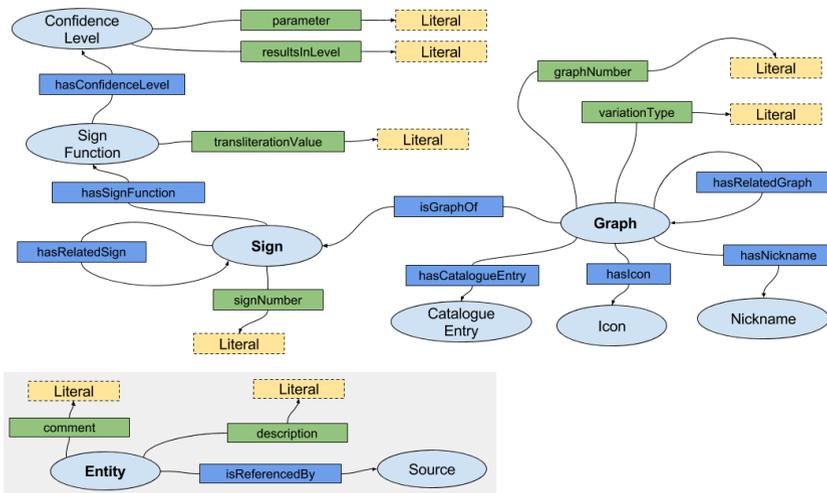


Abb.2: Domänenmodell des TWKM-Zeichenkatalogs⁶

Um mit anderen Datenbeständen und Systemen interoperabel zu sein, wurde die Suche nach einer geeigneten Basis-Ontologie angestrebt, an der sich die Modellierung des Zeichenkatalogs anbinden ließe. Es entzog sich jedoch unserer Kenntnis, ob ein Katalog für Schriftzeichen bereits Gegenstand einer ontologischen Modellierung war.

Trotz des Fokus‘ auf die Beschreibung von Prozessen zur Dokumentation von Objekten des kulturellen Erbes, bietet die Ontologie CIDOC CRM⁷ viele Metakonzepte, die sich für die Modellierung des Zeichenkatalogs als nachnutzbar erwiesen. Die Klassenhierarchie (Abb. 3) zeigt, dass fast alle Klassen des TWKM-Zeichenkatalogs als Subklassen von CIDOC CRM definiert werden konnten. Als “identifiable expressions in natural language” wurden “Sign” und “Graph” als spezifische Subkonzepte von “E33_Linguistic_Object” [IC11] gefasst. Die Klasse “ConfidenceLevel” ermöglicht es, eine Aussage über die Plausibilität einer Zeichenfunktion zu treffen, damit kann sie als “set[s] of propositions about

⁶ Grafik erstellt von Franziska Diehr

⁷ CIDOC Conceptual Reference Model

real or mental things and that are documented as single units or serve as topic of discourse” (E89_Propositional_Object) [IC11] verstanden werden.

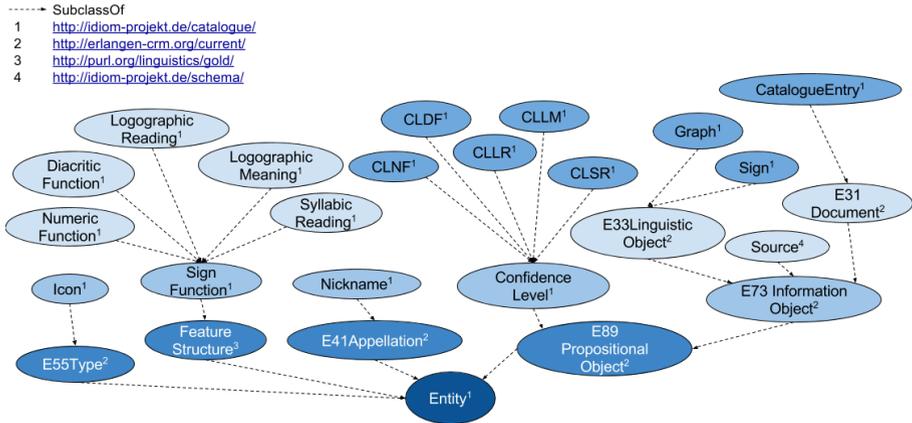


Abb. 3: Klassenhierarchie des TWKM-Zeichenkatalogs⁸

Als einzige Ausnahme wurde “SignFunction” nicht als Subkonzept eines CIDOC CRM Terms, sondern als Subklasse von “FeatureStructure” aus GOLD⁹ modelliert. Mit “SignFunction” wird einem Zeichen eine Funktion zugeordnet (Abb. 2, hasSignFunction). Damit kann es als eine spezifische Form von “FeatureStructure [...] a kind of information structure [...] used to group together qualities or features of some object“ [GO10] aufgefasst werden. Im Klassischen Maya lassen sich fünf Zeichenfunktionen beschreiben: (1) Zahlzeichen “NumericFunction”, (2) diakritische Zeichen “DiacriticFunction”, (3) Logogramme mit identifizierter Lesung “LogographicReading”, (4) Logogramme mit unidentifizierter Lesung “LogographicMeaning”, (hier wird den Zeichen wird eine Bedeutung zugeordnet) und (5) syllabische Zeichen mit identifizierter Lesung “SyllabicReading”.

6.1 Qualitative Bewertung und Einstufung von Entzifferungshypothesen

Bei der Arbeit mit einer noch nicht vollständig entschlüsselten Schrift müssen Forscher zwangsläufig verschiedene Hypothesen zur Entzifferung von Zeichen vorlegen. Wir sehen es als eine notwendige Aufgabe an, alle nachvollziehbaren Lesungsvorschläge zu dokumentieren. Dabei ist es uns ein besonderes Anliegen, die Entzifferungshypothesen nicht nur zu sammeln, sondern deren Qualität nach formalen Kriterien bewertbar zu machen.

⁸ Grafik erstellt von Franziska Diehr

⁹ General Ontology for Linguistic Description (GOLD) zur Beschreibung von Linguistik mit dem Ziel grundlegende Kategorien und Relationen für die wissenschaftliche Beschreibung von menschlicher Sprache zu definieren [GO10].

Dazu haben wir ein neutrales, transparentes System geschaffen, das anhand formaler Kriterien ermöglicht, eine Einstufung der Plausibilität und Qualität von Entzifferungshypothesen vorzunehmen.

Dem Zeichen werden, wie vorangehend erläutert, Zeichenfunktionen zugewiesen (Abb. 2, `hasSignFunction`), die einen graphemischen Transliterationswert (`transliterationValue`) haben. Ein Zeichen kann mehrere gleiche Arten von Zeichenfunktionen haben, so kann es bspw. zwei logographische Lesungen mit jeweils einem distinkten Transliterationswert haben. Um diese mit einem “`ConfidenceLevel`” zu versehen, werden aus einer Liste jene Kriterien (`parameter`) ausgewählt, auf die sich der angegebene Transliterationswert stützt. Für jede Art der Zeichenfunktion wurde ein eigenes Set¹⁰ entwickelt, das sich u.a. am sprachlichen Nutzungskontext (z.B. korrekte Wortart, plausibler Text-Bild-Bezug etc.) oder am Nachweis in modernen Maya-Sprachen orientiert [Ke76, 165-185][Ho01]. Die Kriterien sind mittels Aussagelogiken so miteinander in Bezug gesetzt, dass je nach Kombination automatisch eine qualitative Einstufung (`resultsInLevel`) vorgenommen wird. Dabei steht “`Level 1`” für die höchstmögliche Konfidenzstufe und damit für eine evidente Entzifferungshypothese.

Ein Beispiel: Für die Zeichenfunktion des Logogramms **CH’AM**, das in einigen Kontexten **ch’a-CH’AM**, in anderen **CH’AM-ma** geschrieben wird, und als transitives Verb häufig mit dem präfigierten Pronomen *u-* erscheint, welches anzeigt, dass das Wort mit einem Konsonanten beginnen muss, ergibt die Auswahl der Kriterien “`preposed complementation [-prefix(es)]`”, “`postponed complementation [-suffix(es)]`” und “`preconsonantal ergative pronoun`” die Einstufung in “`Level 2`”. Damit liegt eine relative hohe, aber noch keine gesicherte Entzifferung vor.

6.2 Integration bestehender Zeicheninventare

Um auf die Unzulänglichkeiten der bisherigen Kataloge zu reagieren, nimmt der TWKM-Zeichenkatalog eine vollständige Neuinventarisierung der bekannten Mayaschriftzeichen vor. Dabei wird jedes Zeichen mit einer Katalognummer versehen (Abb. 2, `signNumber`). Als in der Forschung etablierte Arbeitsinstrumente, werden die elf vorhandenen Kataloge als Konkordanz in den TWKM-Katalog integriert. Dies geschieht indem die neu inventarisierten Graphen mit Einträgen aus den anderen Katalogen versehen werden (`Catalogue-Entry`). Dabei werden u.a. Katalognummer, Abbildung, Kommentar und Quellennachweis dokumentiert. Durch die Zuordnung des Katalogeintrags zum Graph statt zum Zeichen, können wir die Mehrfachinventarisierungen aufbrechen und Falschzuordnungen aufheben. Durch die Verbindung zu den neuinventarisierten Graphen erhalten wir außerdem die forschungsgeschichtliche Perspektive auf die Zeichenklassifikation und ermöglichen unseren Nutzern auch die Arbeit mit und Suche nach bereits etablierten Katalognummern.

¹⁰ siehe dazu Dokumentation des Metdatenschemas <http://idiom-projekt.de/catalogue>

6.3 Klassifikation von Graphvarianten

Ein Grund für Mehrfachinventarisierungen und Falschzuordnungen von Graphemen zu ihren Zeichen ist sicherlich der Variantenreichtum, mit dem Mayaschreiber Glyphen darstellten. Bisher existierte keine systematische Untersuchung, wie Varianten gebildet wurden, auch wenn einige Vorarbeiten existieren [Be34a][Be34b][Be36][Be37][La95, 204ff.], die z.B. Substitutionen, Rotationsprinzipien oder Symmetrien untersuchen. Einige Kataloge listen für bestimmte Zeichen auch geläufige Graphvarianten, aber ohne sie komparativ zu inventarisieren. Das TWKM-Projekt konnte erstmalig allgemeingültige Regeln und Prinzipien zur Graphbildung durch graphotaktische Analysen ableiten und definieren. So wurden insgesamt 45 einzelne Variationsmöglichkeiten identifiziert, die sich in neun Klassen systematisieren lassen (Mono-, Bi-, Tri-, & Variopartite, Division, Animation Head, Animation Figure, Multiplication, Extraction) und für alle Graphvarianten universell vergleichbar sind. Die Identifizierung von Varianten stellt ein weiteres Argument für die getrennte Erfassung von Zeichen und Graph dar und begründet unser Konzept. Bei der Erfassung eines Graphs wird dessen “variationType” aus einer vordefinierten Liste ausgewählt. Der Typ wird durch eine Abkürzung, z.B. “st” (Monopartite), dargestellt. Wenn möglich, wird dieses Graph dann mit einem Zeichen verbunden, das bereits eine Katalognummer besitzt. Sobald die Verknüpfung erfolgt ist, wird automatisch die “graphNumber”, bestehend aus der “signNumber” des Zeichens und dem “variationsType” gebildet, bspw. 556st.

6.4 Bestimmung des ikonischen Charakters der Graph

Insbesondere zur Durchführung paläo- und ikonographischer Studien zum bildhaften Charakter der Mayaglyphen, haben wir ein kontrolliertes Vokabular entwickelt. In SKOS modelliert, umfasst es 13 Facetten¹¹, die das Graphikon formal beschreibbar machen. Im Metadatenschema wird das Graph mittels “hasIcon” (Abb. 2) mit den entsprechenden Termen des Vokabulars (Icon) verknüpft.

6.5 Nachweissystem für Literaturquellen

Um dem grundlegenden Anspruch an wissenschaftliche Arbeit, dokumentierte Aussagen mit entsprechenden Quellen zu belegen, gerecht zu werden, legen wir besonderen Wert darauf, sämtliche Informationen stets nachvollziehbar und vor allem überprüfbar zu halten. Insbesondere beim Nachweis von Lesungshypothesen und den Einträgen aus den konkordanten Katalogen sind Quellenangaben unverzichtbar. Wie das Domänenmodell (Abb. 2) zeigt, können alle Entitäten (Entity) unseres Schemas mittels der Relation “isReferencedBy” einer Quelle (Source) verbunden werden. Damit wird ein lückenloser Nachweis der verwendeten Quellen ermöglicht. An dieser Stelle kommt eine weitere Komponente unserer virtuellen Arbeitsumgebung zum Einsatz: Für unsere

¹¹ human and animal, body part, age and sex, pose and gesture, plant and plant part, landscape, artefact physical state, property, orientation

Projektbibliographie verwenden wir das open-source Literaturverwaltungsprogramm Zotero, das innerhalb unseres Erfassungssystems mittels API abgefragt wird. Sowohl URI der betreffenden Ressource als auch eine formatierte bibliographische Angabe werden bei der entsprechenden Instanz, z.B. einer Lesungshypothese, gespeichert.

7 Der Zeichenkatalog als Grundbaustein für Textkorpus und linguistische Analyse

Da es für die Hieroglyphenschrift des Klassischen Maya keinen Unicode-Schriftsatz gibt und die Lesung vieler Zeichen noch nicht bekannt oder umstritten ist, kann das in TEI kodierte Textkorpus weder aus einem festen Schriftzeichensatz, noch aus phonemisch-transliterierten Werten bestehen. Um ein flexibles Korpus zu erhalten, das auf neue Entzifferungen und verschiedene Lesungshypothesen reagieren kann, nutzen wir den TWKM-Zeichenkatalog als eine Art Grundbaustein bei der Korpuserstellung. Innerhalb des kodierten Texts wird jede Glyphe mittels Katalognummer und einer Referenz zu der entsprechenden URI im Zeichenkatalog erfasst. Der Text wird damit erst durch die Abfrage der Daten aus dem Zeichenkatalog ‘lesbar’. Dies geschieht in einem weiteren Prozessierungsschritt mittels des linguistischen Analysetools, einer weiteren Komponente unserer virtuellen TWKM-Arbeitsumgebung. Hier werden die numerischen Transliterationen (Katalognummern) zunächst in die entsprechende graphemische und weiter in phonemische Transliterationen überführt. Erstes geschieht semi-automatisch, da es mehrere mögliche Transliterationenwerte (z.B. syllabische und logographische Lesungen) für ein Zeichen gibt. Hier spielt auch die im Zeichenkatalog vorgenommene qualitative Einstufung der Transliterationenwerte eine entscheidende Rolle: Die Analyse kann anhand hoher oder niedriger Konfidenzstufen durchgeführt werden. Die Lesungshypothesen können somit in ihrem Verwendungskontext überprüft werden, was idealerweise zu neuen Erkenntnissen für die Entzifferung führen kann. Das linguistische Analysetool wird es darüber hinaus ermöglichen parallele, als gleichwertig anzusehende Textanalysen anzulegen. Somit kann den verschiedenen Lesungshypothesen Rechnung getragen werden. Durch die Verbindung von Zeichenkatalog, Textkorpus und linguistischer Analyse entsteht letztendlich ein dynamischer Text, der je nach Forschungsfrage individuell generiert werden kann.

Bei der Erstellung des Textkorpus zeigt sich noch ein weiterer Vorteil, der sich aus der getrennten Erfassung von “Sign” und “Graph” ergibt: Soweit es die Beschaffenheit des Textträgers erkennen lässt, wird auf das Graph im Zeichenkatalog referenziert. Damit wird die konkrete Verwendung bestimmter Graphvarianten explizit dokumentiert und Fragen zum Verwendungskontext von Glyphen können untersucht werden: Gibt es bestimmte Perioden oder geographische Räume in denen sie verwendet wurden? Zeigt sich eine Abhängigkeit von der Art des Textträgers, seines Aufstellungsorts oder vom Textinhalt?

8 Ausblick auf kommende Arbeiten

In den kommenden Monaten ist die Neuinventarisierung von Mayaschriftzeichen mittels des Erfassungssystems des digitalen Zeichenkatalogs eine der Hauptaufgaben im Projekt. Dabei werden auch Zeichnungen für alle erfassten Grapheme von den Schriftforschern im TWKM-Projekt angefertigt und in den Zeichenkatalog integriert.

Derzeit wird intensiv an der Spezifikation des TEI-Schemas für die Textauszeichnung des Korpus gearbeitet. Ziel ist es, in der verbleibenden Projektlaufzeit bis voraussichtlich 2028, alle bisher bekannten Textträger (ca. 10.000) zu dokumentieren und deren Texte mittels des TEI-Schemas auszuzeichnen. Um diese Arbeit intuitiver zu gestalten und die Texterstellung zu beschleunigen, wird zurzeit ein Plug-In für TextGrid gearbeitet, das mittels eines Parsers aus manuell eingegebenen numerischen Transliterationen der Texte XML-Dokumente generiert, die nach projektspezifischem TEI-Schema strukturiert sind.

Sobald die grundlegende Inventarisierung aller Zeichen und Grapheme erfolgt ist, wird der Zeichenkatalog auf dem Projektportal veröffentlicht. In der Zwischenzeit arbeiten wir an einem umfangreichen Aufsatz mit dem Arbeitstitel “Standards for Maya Epigraphic Analysis I: Principles of Maya Graphemics”, der im Laufe des Jahres auf der Projektwebseite¹² veröffentlicht werden soll. Dieser wird, neben dem hier besprochenen Konzept zur Modellierung des Zeichenkatalogs, auch eine ausführliche Besprechung der Graphvariantenbildung, der Konfidenzstufen und -kriterien sowie weiterer, hier nur ansatzweise betrachteten, Forschungsergebnisse enthalten. Damit wird dieser Aufsatz auch insbesondere für die Fachcommunity der Mayaschriftforschung von Interesse sein.

Literaturverzeichnis

- [Be34a] Beyer, Hermann: The Position of the Affixes in Maya Writing I. *Maya Research* 1, 20-29, 1934.
- [Be34b] Beyer, Hermann: The Position of the Affixes in Maya Writing II. *Maya Research* 1, 101-108, 1934.
- [Be36] Beyer, Hermann: The Position of the Affixes in Maya Writing III. *Maya Research* 3(1), 102-104, 1936.
- [Be37] Beyer, Hermann: Studies on the inscriptions of Chichen Itza. *Contributions to American Anthropology and History* 4(21), 29-175, 1937.
- [Bu90] Bußmann, Hadumod: *Lexikon der Sprachwissenschaft*, 2. Auflage, 1990.
- [Co06] Coulmas, Florian: *The Blackwell Encyclopedia of Writing Systems*, 4. Auflage, Blackwell, Oxford, 2006.

¹² <http://mayawoerterbuch.de/>

- [Cr10] Crystal, David: *The Cambridge Encyclopedia of Language*, 3. Auflage, Cambridge University Press, Cambridge, 2010.
- [No13] Nordrhein-Westfälische Akademie der Wissenschaften und der Künste: 10,7 Millionen Euro für zwei neue Langzeitforschungsprojekte, 2013, <http://www.awk.nrw.de>, Stand: 11.05.2017.
- [FL03] Farrar, Scott; Langendoen, D. Terrence: A linguistic ontology for the Semantic Web, *GLOT International* 7(3), 97-100, 2003.
- [Fl07] Flick, Uwe: *Qualitative Sozialforschung, Eine Einführung*, 7. Auflage, Rowohlt, 2007.
- [GO10] GOLD: General Ontology for Linguistic Description, 2010, <http://linguistics-ontology.org/>, Stand: 11.05.2017.
- [Gr90] Grube, Nikolai: *Die Entwicklung der Mayaschrift: Grundlagen zur Erforschung des Wandels der Mayaschrift von der Protoklassik bis zur spanischen Eroberung. Acta Mesoamericana* 3, Von Flemming, Berlin, 1990.
- [IC11] ICOM/CIDOC CRM Special Interest Group: CIDOC Conceptual Reference Model, Version 5.0.4, 2011, <http://www.cidoc-crm.org/cidoc-crm/>, Stand: 11.05.2017.
- [Ho01] Houston, Stephen: Introduction. In: (Houston, Stephen, Oswaldo Chinchilla & David Stuart): *The Decipherment of Ancient Maya Writing*, University of Oklahoma Press, Norman, S. 3-19, 2001.
- [Ke62] Kelley, David: Review of *A Catalog of Maya Hieroglyphs*, by J. Eric S. Thompson. *American Journal of Archaeology* 66, 436-438, 1962.
- [Ke76] Kelley, David: *Deciphering the Maya Script*, University of Texas Press, Austin, 1976.
- [Ku89] Kurbjuhn, Kornelia: *Maya: The Complete Catalogue of Glyph Readings*, Schneider & Weber, Kassel, 1989.
- [La95] Lacadena, Alfonso: *Evolución formal de las grafías escriturarias mayas, Implicaciones históricas y culturales*, Madrid, 1995.
- [Mo02] Montgomery, John: *How to Read Maya Hieroglyphs*. Hippocrene, New York, NY. 2002.
- [Pr14a] Prager, Christian: *Zielsetzung*, 2014, <http://mayawoerterbuch.de>, 11.05.2017.
- [Pr14b] Prager, Christian: *Zeichentypen und Kataloge*, 2014, <http://mayawoerterbuch.de>, 11.05.2017.
- [Re15] Reinhold, Anke: Das Experteninterview als zentrale Methode der Wissensmodellierung in den Digital Humanities. *Information Wissenschaft Praxis*, 66 (5-6), 327-333, 2015.
- [Ri06] Riese, Berthold: Drei neue Maya-Hieroglyphen Kataloge. *Anthropos* 101, 238-246, 2006.
- [Ro05] Rogers, Henry: *Writing Systems, A linguistic Approach*, Blackwell Textbooks in Linguistics, 2005.
- [RSS96] Ringle, William M.; Smith-Stark, Thomas C.: *A Concordance to the Inscriptions of Palenque, Chiapas, Mexico*, Middle American Research Institute, Tulane University, New Orleans, 1996.