Morphological Glossing of Mayan Languages under XML: Preliminary Results

Frauke Sachse\(^1\) & Michael Dürr\(^2\)

\(^1\) Rheinische Friedrich-Wilhelms-Universität, Bonn
\(^2\) Freie Universität, Berlin

Introduction

This paper summarises the results of a workshop that was held at the Department for the Anthropology of the Americas of the University of Bonn between 4-6 September 2014. The workshop was a joint initiative of the research project Textdatenbank und Wörterbuch des Klassischen Maya (TWKM = Text Database and Dictionary of Classic Mayan) and the research group developing the software application Tool for Systematic Annotation of Colonial K’iche’ (TSACK) and aimed at discussing and defining standardised conventions for the linguistic description and glossing of Mayan language forms under XML.\(^1\)

Grammatical descriptions of Mayan languages exhibit a plethora of descriptive standards. Produced by different linguists of different backgrounds with different research objectives, they reflect the diverse theoretical orientations of the linguistic discipline, ranging from formal descriptions of the structural or generative type to prescriptive grammars for the use in language teaching. Functionally identical forms are found to be analysed and glossed rather differently, depending on the purpose of description or the theoretical model applied. Even edited volumes usually maintain the personal preferences of authors, which may result in the ‘third person singular ergative’ being variously glossed in one and the same volume as “3erg”, “3sE”, “3SE”, “3sgE”, or --following a common standard of distinguishing pronominal sets a (ergative) and b (absolutive) as “3a”, “3sA”, “3sg.a”, “3SG.A”, “a3s”, “A3s”, “a3” and “A.3” (see Avelino 2011 among others). Although there are justifications for maintaining different conventions, these constitute a source of potential confusion; in the case of the just mentioned example the abbreviation, “A” might be mistaken for the equally common gloss of the absolutive pronoun. Few attempts have been made to compare and integrate this material and provide

\(^1\) The participants of the workshop who contributed to the discussion and examples that are used in the present paper include in alphabetical order: Katja Diederichs, Sven Gronemeyer, Christian Prager, Elisabeth Wagner (for TWKM) as well as Michael Dürr, Christian W.R. Klingler and Frauke Sachse (for TSACK).
a standardised and generally applicable descriptive terminology that can help to analyse grammatical
development in the Mayan language family.

Any attempt to make the data of different Mayan languages comparable requires the definition of set
conventions for glossing and typological description. As a prerequisite to the analysis of Classic Mayan
by systematic comparison with modern and historic languages of the Mayan family, the TWKM-project
will need to decide on such conventions. By choosing conventions that other corpus projects on Mayan
languages operating within the same XML-based environment can share, the data would become
comparable and permit comprehensive analysis of semantic and grammatical structure across corpora
in the TextGrid repositories. Thus, standardising the rules for glossing would create the necessary
infrastructure for a network of Mayan language database projects within the TextGrid environment.

The aim of the workshop was to identify and discuss difficulties and problems in interlinear glossing of
Mayan languages and use them as a basis for defining the conventions and rules of linguistic
description under XML. The languages that were primarily focused on during the workshop, thus,
included K’iche’ (colonial and modern), Ch’ol, Modern Yukatek and Classic Mayan. Accordingly, the
following summary presents results that are only preliminary and are not yet meant as a defined
standard, but as a basis for further discussion.

Basic premises of the XML environment

Linguistic glossing is dependent on its purpose. The conventions proposed and discussed in this paper
take the respective objectives of the TWKM and TSACK projects into account and conform with the
restrictions imposed by the XML environment of an annotated corpus.

The main objective of the TWKM project is to build a corpus-based dictionary of Classic Mayan. Using
the virtual research environment TextGrid, all Classic Maya texts will be compiled in a digital corpus
and annotated to create a comprehensive database of lexical entries and morphosyntactic forms and
structures. The annotation process starts with the graphemic classification of hieroglyphic signs and
needs to include the phonemic transcription of sign values and their morphemic transliteration into
words. The transliterated texts are then morphologically analysed and glossed, which constitutes the
basis for the translation of sentence structures and the individual lexemes, from which the dictionary
is built. The annotation process is complex and requires the inclusion of multiple options on all levels.
An exact XML-schema and the technological infrastructure are at this stage still under construction.

The Tool for Systematic Annotation of Colonial K’iche’ (TSACK) is being developed as a software that
supports the semi-automated analysis and XML-annotation of language forms in colonial documents
(see Sachse et al. 2015). The primary objective of the research project is to define XML-based
standards for corpus-oriented documentation of colonial dictionaries of the Highland Mayan language
K’iche’. Colonial dictionaries do not follow common orthographic standards and exhibit inconsistencies
in semantic correspondences of K’iche’ and Spanish entries. TSACK assists in the analysis of the
orthography and speeds up the XML-annotation process, which allows for the processing of larger
quantities of lexicographic data. There are plans to implement this tool into the TextGrid environment
and further develop and adapt it for the annotation of colonial data from other Mayan languages.

---

2 TSACK was developed in a pilot study for a project on the lexicography of colonial K’iche’ that will be undertaken
by the authors of this paper. The research was funded at the University of Bonn between October 2013 and
September 2014 (Maria von Linden-Programm). The programming was carried out by Christian Klingler, who was
imminently involved in the theoretical development of the software.
which would help in processing large amounts of language data and make them available for comparative analysis.

Both projects share the objective of building databases that will serve the lexico-semantic and grammatical analysis of Mayan language data. Accordingly, linguistic glossing conventions need to be adapted to this particular purpose.

Dictionaries consist of lexical entries, or lemmata, the basic forms of lexical words. Dictionary-building thus always requires lemmatisation, i.e. the definition of the basic lexical form. The process of lemmatisation is dependent on the typology of the language. Mayan languages are primarily agglutinating. To build a dictionary from a text corpus, the words in each text need to be broken down into their morphological parts to make the lexical stems and roots retrievable within the corpus. Each of the elements that can make up a word (root, lexical stem, derivational morphemes, grammatical morphemes) need to be glossed individually. While for most cases of glossing it would suffice to break complex forms down to the lemma (1a), the compilation of lexical databases for which TSACK is being developed requires the morphological analysis of each form down to the root (1b).

(1) Glossing of stems and roots

K’iche’

a. k-in-b’aqir-ik
INC-1s.ABS-become.thin-MOD.V.INTR
‘I become thin’

b. k-in-b’aq-ir-ik
INC-1s.ABS-N:bone-INTRVZ.INCH-MOD.V.INTR
‘I become thin’

A lemma consists of a minimum of a root and can combine a root and one or more derivational morphemes. Each derivational morpheme derives a new lemma which is annotated accordingly. The distinction of grammatical and derivational morphology and the classification of lexical categories needs to be part of the annotation scheme, as shown in the following example of a K’iche’ form. Accordingly, lexical and derivational categories need to be glossed unambiguously.

(2) XML-annotation of the entry *quinbakiric* from the Anonymous Franciscan K’iche’ Dictionary:

```xml
<entry>
  <kichee_entry>
    <word>
      <original_form xml:id="w1">quinbakiric</original_form>
      <ref target="w1" type="transliteration" status="certain">
        <gram_affix function="INC" affix_is="prefix">k</gram_affix>
        <gram_affix function="1s.ABS" affix_is="prefix">in</gram_affix>
        <lemma xml:id="l1" class="V.INTR">
          <lemma xml:id="l2" class="N">
            <root xml:id="r1" class="N">b’aq</root>
          </lemma>
          <der_affix function="INTRVZ.INCH" affix_is="suffix">ir</der_affix>
        </lemma>
        <gram_affix function="MOD.V.INTR" affix_is="suffix">ik</gram_affix>
      </ref>
      <ref target="l1" type="translation" status="certain">become.thin</ref>
      <ref target="l2" type="translation" status="certain">bone</ref>
      <ref target="r1" type="translation" status="certain">bone</ref>
    </word>
  </kichee_entry>
</entry>
```
In the example, grammatical morphemes are glossed in green, derivational categories in blue, and lexical classes in red. The detailed annotation allows the rebuilding of both, root-based and stem-based glossing.

(3) Root-based and stem-based glossing of annotated example

<table>
<thead>
<tr>
<th>original dictionary entry</th>
<th>quinbakiric</th>
</tr>
</thead>
<tbody>
<tr>
<td>transcription</td>
<td>kinb’aqirik</td>
</tr>
<tr>
<td>morphological analysis (1)</td>
<td>k-in-b’aq-ir-ik</td>
</tr>
<tr>
<td>gloss (1)</td>
<td>INC-1s.ABS-N:bone-INTRVZ.INCH-MOD.V.INTR</td>
</tr>
<tr>
<td>morphological analysis (2)</td>
<td>k-in-b’aqir-ik</td>
</tr>
<tr>
<td>gloss (2)</td>
<td>INC-1s.ABS-V.INTR:become.thin-MOD.V.INTR</td>
</tr>
<tr>
<td>translation</td>
<td>‘I become thin’</td>
</tr>
</tbody>
</table>

The annotation of Classic Mayan texts has special requirements. Morphological analysis and glossing are dependent on the phonemic transcription, the transliteration of syllabic sign values and ultimately the graphemic classification. As all of these processes imply a certain level of uncertainty, annotation needs to allow for multiple interpretations. Furthermore it needs to be borne in mind that lexical and morphological analysis, and thus glossing, of Classic Mayan is still a reconstructive process that draws on evidence from modern and colonial Mayan languages in order to identify the lexical roots, grammatical markers and functions of the language depicted by the hieroglyphic script. As illustrated in the following example (4), the exact morphological analysis is not always clear and alternative glossings need to be included and retained until the grammatical patterns are better understood. It is the aim of the TWKM project to corroborate or dismiss current reconstructions and hypotheses about Classic Maya grammar based on a large annotated corpus of inscriptions. The glossing of lexical and morphological forms in the Classic Maya corpus is therefore as much an analytical result as it is an analytical tool to test and verify formal as well as functional categories.

(4) Interdependence of reconstructive sign analysis and morphological glossing in Classic Mayan sign

| classification (Thompson 1962) | 644°19:130.116:126 |
| translation                    | CHUM-mu-wa-ni-ya   |
| transcription                  | chumwaniy          |
| morphological analysis (1)     | chum-wan-ø=iy      |
| gloss (1)                      | POS:sitting-INTRVZ-3s.ABS-ANT |
| morphological analysis (2)     | chum-wan-iy-ø      |
| gloss (2)                      | POS:sitting-INTRVZ-COM-3s.ABS |
| translation                    | ‘he sat down’      |
A note on orthographic standards

Linguistic glossing is independent from the orthographic standard used to represent the object language that is being glossed. However, for the purpose of defining standard conventions for TWKM and TSACK a common orthography needs to be used. Since the early colonial times, various orthographies have been in use, generating a significant number of potentially ambiguous characters. While in most modern orthographies the grapheme <k> represents the non-glottalised velar stop /k/, earlier (including colonial) orthographies used it either to represent the glottalised velar stop /k’/ (colonial Yukatek) or for the non-glottalised uvular stop /q/ (colonial K’iche’).

The current paper employs the phoneme-based standard alphabet defined by the Academia de las Lenguas Mayas (1988) to represent the Mayan languages of Guatemala. With the exception of grapheme <x>, the characters, or letters, of the ALMG alphabet are unambiguous and also apply to most Mayan languages in Mexico. The common Mexican conventions of using <b> instead of <b’> and <ts>/<ts’> instead of <tz>/<tz’> are not followed in here.

The orthographic conventions are shown below in integrated inventories.

<table>
<thead>
<tr>
<th>Stops</th>
<th>Bilabial</th>
<th>Alveolar</th>
<th>Alveopalatal</th>
<th>Retroflex</th>
<th>Palatal</th>
<th>Velar</th>
<th>Uvular</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>[- glottalised]</td>
<td>p</td>
<td>t</td>
<td>ty*</td>
<td>ky</td>
<td>k</td>
<td>q</td>
<td></td>
<td>’</td>
</tr>
<tr>
<td>[+ glottalised]</td>
<td>p’</td>
<td>t’</td>
<td>ty**</td>
<td>ky’</td>
<td>k’</td>
<td>q’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[+ voiced]</td>
<td>b’</td>
<td>d’</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affricate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[+ glottalised]</td>
<td>tz</td>
<td>ch</td>
<td>tx**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>tz’</td>
<td>ch’</td>
<td>tx’**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nasal</td>
<td>m</td>
<td>n</td>
<td>ñ*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fricative</td>
<td>s</td>
<td>x, xh**</td>
<td>x**</td>
<td></td>
<td>j</td>
<td>h</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lateral</td>
<td>l</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vibrant</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glide</td>
<td>w</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes:
* Alveopalatal ty, ty’ and ñ have not been defined in the ALMG alphabet, but have been added for Ch’ol (Mexico)
** Mamean and Q’anjol’alan only.
*** Apico-alveopalatal affricates (tch and tch’) and fricative (sh) have been excluded from this table, as they are restricted to a single variety of Mam (Todos Santos).

Table 1: Integrated consonant inventory of Mayan languages

Vowel length is a distinctive feature in several Mayan languages, although the short vs. long distinction is quite often realised as a lax vs. tense articulation. According to the recommendations of the ALMG, vowel length will not be indicated for the K’iche’an languages.

In Modern Yukatek, tones are indicated by acute (´; = high) or gravis (`; = low) accent over the vocalic nucleus of a syllable.

<table>
<thead>
<tr>
<th>High</th>
<th>Short</th>
<th>Front</th>
<th>Long</th>
<th>Lax</th>
<th>Short</th>
<th>Central</th>
<th>Long</th>
<th>Lax</th>
<th>Short</th>
<th>Back</th>
<th>Long</th>
<th>Lax</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>i</td>
<td>ii</td>
<td>i</td>
<td>*i</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mid</td>
<td>e</td>
<td>ee</td>
<td>e</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>a</td>
<td>aa</td>
<td>å</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes:
* For Ch’ol and Chontal (both Mexico), a high central short vowel <i> has been added to the ALMG alphabet.

Table 22: Integrated vowel inventory of Mayan languages
Adaptation of the Leipzig Glossing Rules

The standard for linguistic glossing and description of Mayan languages to be developed by the current initiative follows the rules and conventions laid out in the Leipzig Glossing Rules (LGR), which are here expanded and modified to meet the specific properties of Mayan languages and the constraints imposed by the given research objectives.

The definition of the LGR was a joint effort by Linguistic departments of the Max Planck Institute for Evolutionary Anthropology in Leipzig and the University of Leipzig (see LGR: 1). The rules were defined in response to the lack of a common standard for linguistic glossing and the need for such typological conventions to facilitate cross-linguistic comparison. The descriptive and comparative research disseminated by the Department of Linguistics of the MPI in Leipzig, including the World Atlas of Language Structures (wals.info/), apply the LGR as a standard. The LGR were intended as a set of rules and standard conventions for the glossing of morphological categories in linguistic publications. The glossing of syntactic features has been deliberately excluded. The LGR cover the core of grammatical and functional categories and do not claim to be exhaustive; the optional need for defining and modifying the standard set of conventions is explicitly acknowledged (p. 1). A number of different initiatives have expanded the LGR. The main feature not included in the LGR are derivational categories. Since derivation is a basic principle of word formation in Mayan languages and thus essential to the analysis of lexical categories as it is required by the TWKM and TSACK projects, glosses for derivational categories need to be included.

One essential prerequisite of interlinear glossing set out in the LGR is that glosses encode functional meaning and grammatical properties of morphemes. Existing grammatical descriptions of Mayan languages do not generally observe this rule, instead morphemes are frequently glossed by their structural category or “grammatical function” is defined based on the form of a morpheme and not its context. This is in particular the case, when only the structural properties of a morpheme are known, but the functional category is not understood.

The definition of glossing rules cannot be independent of linguistic description and functional categorisation. The analyses of morphological functions can however differ quite substantially. For instance, the Yukatek aspectual prefix k- has been variously identified as an incompletive (e.g. Smailus 1989), imperfective (e.g. Verhoeven 2007: 117) or habitual (Bricker 1998). Or the K’iche’ suffix -ik that marks intransitive verbs in final position of the clause has been categorised as a modal marker (e.g. Dürr 1987), a status suffix (Kaufman 1990:71), category suffix (= sufijo de categoría) (López Domingo 1997: 84), or simply a phrase final marker (Romero 2006). The definition of a common understanding of grammatical forms is therefore a prerequisite to systematic glossing. Comparative analysis of grammatical development in Mayan languages shows that functionally identical categories can be marked by structurally rather distinct elements. The historical development of elements, however, must not be entirely disregarded, when identifying functional categories. The present summary takes basic reflections on the typology of Mayan languages into account and discusses the analyses of linguistic features, where necessary. As indicated in the LGR, glossing rules cannot solve the problem of multiple analyses. Forms that can be analysed, and thus glossed, in multiple ways are a common feature in Mayan languages, e.g. Yukatek b’ak’il waaj which can be analysed as ‘meat-bread’ or ‘meaty bread’.
Basic glossing rules

The following basic rules for glossing of functional and semantic properties in Mayan languages are restricted to linguistic glossing on the morphological level, aspects of syntactic glossing are not taken into account at this stage. The rules are taken and expanded upon from the LGR.

Word Alignment and Separation of Morphemes

The LGRs define interlinear glosses to be left-aligned vertically and word by word (see LGR, Rule 1). Morphemes are separated by hyphen (see LGR, Rule 2). No distinction is being made between grammatical and derivational morphology, both use a ‘dash’ for hyphenation.

(5) Vertical word alignment and hyphenation

KʼICHEʼ

a. k-e-war-ik    ri  ixoq-ibʼ
INC-3p.ABS-sleep-MOD.V.INTR ART woman-PL
‘the women sleep’

If morphologically bound elements constitute distinct prosodic or phonological words, a hyphen and a single space may be used together in the language example, while the gloss treats the form as a single word (see LGR; Rule 2A).

(6) Prosodically separate units constituting a complex form

YUKATEK

a. k-u- y-il-ik-ø
INC-3s.ERG-3s.ERG-see-INCl.V.TR-3s.ABS
‘s/he sees it’

CHʼOL

b. tziʼ- kʼel-e-ø
COM.3s.ERG-see-COM.V.TR-3s.ABS
‘he saw it’

While affixes are separated by hyphens, clitic boundaries are generally marked by an equals sign = (see LGR, Rule 2). The definition of clitics and their differentiation from affixes are not necessarily straightforward in Mayan languages. Within the XML-annotation scheme, clitics will be treated like affixes, in that they are marked for grammatical function and structurally specified as “enclitics”.

(7) Prosodic units consisting of more than one element (including clitics)

CHʼOL

a. wiñik-oñ=ku
man-1s.PRED=ASS
‘I am a man’
While in the LGR reduplication is marked separately by a tilde ~, we treat it here like affixation. In Mayan languages, reduplicated elements generally have derivational or grammatical function and can therefore be treated as morphemes. Both, partial and full reduplication are common in Mayan.

(8) Reduplication

YUKATEK

a. le-letz’-kil
   [C1V1-V.INTR-ADVJZ]\(^3\)
   INTENS-*sparkle-ADJVZ
   ‘sparkling’
Ch’OL

b. woj-woj-ña
   [C1V1C2-ROOT-ADVLZ]
   INTENS-bark-ADVLZ
   ‘yapping’

Allomorphs and epenthetic segments

Many Mayan languages also have developed allomorphs for affixes that are sensitive to the vocalic or consonantal character of the adjacent syllable margin of the morpheme boundary. In the following example (9a) from K’iche’ the allomorphs of the second person singular possessive prefixes a- and aw- are both glossed as 2s.POSS. In example (b), k- and ka- are both glossed as INC.

(9) Allomorphs

K’ICHE’

a. a-b’i’    vs.    aw-ochoch
   2s.POSS-name
   ‘your name’
   2s.POSS-home
   ‘your home’

b. k-at-xaj-aw-ik    vs.    ka-ø-xajawik
   INC-2s.ABS-dance-AP-MOD.V.INTR
   ‘you dance’
   INC-3s.ABS-dance-AP-MOD.V.INTR
   ‘s/he dances’

In a number of Mayan languages clusters of vowels or consonants in specific morphological contexts are avoided by insertion of an epenthetic vowel or consonantal glide, e.g. y in Ch’ol. Epenthetic vowels or consonants do not carry a meaning of their own and are therefore not glossed as separate elements. Epenthetic segments occurring at a morpheme boundary are therefore assigned to the preceding or following morpheme and thus treated as allomorphs in the glossing. In the following example from Ch’ol, the second person singular absolute suffix -ety is realised as -ety, when following a vowel. In both cases the morpheme is glossed as 2s.ABS.

\(^3\) This line is added for explanation and not to be reproduced in the glossing.
(10) Epenthetic segments

**CH’OL**

a. tzi’- y-ik’-e-yet
   COM.3s.ERG-3s.ERG-give-APPL-2s.ABS
   ‘s/he gave it to you’

b. mi’- y-ik’-e-ñ-ety
   INC.3s.ERG-3s.ERG-give-APPL-INC.V.TR.D-2s.ABS
   ‘s/he gives it to you’

Category labels

The LGR define the use of only upper case letters for the glossing of grammatical category labels. This convention is followed with only one exception, which is the glossing of singular and plural in person categories as s and p. The LGR employ “SG” and “PL” to mark number in person categories. However, to avoid confusion with the nominal plural, which in some Western Mayan languages can be structurally and formally identical with the third person plural, a different gloss is chosen here.

If a morpheme corresponds to more than one “metalanguage element”, the individual glosses for these elements are separated by periods (see LGR, Rule 4). The LGR suggest further conventions to mark such “one-to-many correspondences”, which are however not adopted here.

Bound personal pronouns are labeled with the elements ‘grammatical person’ (e.g. 1s, 3p) and pronominal category (i.e. absolutive, ergative, possessive), separated by the period. Following an option under Rule 4 of the LGR, person and number are not separated by a period, i.e. 1s instead of 1.s.

(11) Elements in person categories

**CH’OL**

a. tzi’- tziñsa-yob’
   COM.3s.ERG-die.CAUS-3p.ABS
   ‘s/he killed them’

   **K’ICHE’**

b. nu-wuj
   1s.POSS-book
   ‘my book’

In some Western Mayan languages, aspectual markers and bound ergative pronouns have fused, creating portmanteau forms with multiple grammatical references that are separated by periods in the gloss, see e.g. CH’OL tzi’ COM.3s.ERG (11a).

Most one-to-many correspondences in Mayan languages regard functional classes that are subdivided into more specific functional categories. For example, in K’iche’ modal suffixes that mark the verb category fall into different modal categories, which are specified after a period. The modal marker -ik occurs with intransitive roots and stems as is accordingly labelled as MOD.V.INTR (8a). The transitive stem tz’ib’a that is derived from the noun tz’ib ‘writing, script, letter’ is marked with the modal suffix -j for derived transitive verbs and accordingly glossed as MOD.V.TR.D (8b). Imperative verbs and verbs with incorporated directional verb take the same set of modal markers (i.e. -oq on intransitive and -a’
on transitive verbs), which are glossed for their respective grammatical function as MOD.IMP or MOD.DIR (8c-d).

(12) Modal categories

K’iche’

a. x-oj-war-ik
   COM-1p.ABS-sleep-**MOD.V.INTR**
   ‘we slept’

b. x-ø-in-tz’ib’-a-j
   COM-3s.ABS-1s.ERG-writing-TRVZ-**MOD.V.TR.D**
   ‘I wrote it’

c. ch-at-b’ix-o-n-oq
   IMP-2s.ABS-song-TRVZ-AP-**MOD.IMP.V.INTR**
   ‘sing!’

d. x-at-ul-inw-il-a’
   COM-2s.ABS-DIR:come-1s.ERG-see-**MOD.DIR.V.TR**
   ‘I came to see you’

Another set of grammatical categories which require the marking of more than one metalanguage elements are derivational operators that derive new lexical classes. The gloss specifies the class of derivation and the semantic function. Nominalisers (NMLZ), for instance, fall into different functional categories, such as agentives (AGT), abstractives (ABSTR), instrumentals (INSTR), verbal nouns (VN), etc. The functional specification of the derivation is added after a period.

(13) Derivational operators deriving new lexical classes

K’iche’

a. kun-a-n-el
   N:medicine-TRVZ-AP-**NMLZ.AGT**
   ‘healer’

b. u-kem-ik
   3s.POSS-V.TR:weave-**NMLZ.VN**
   ‘weaving’

c. saq-ar-ik
   ADJ:white-**INTRVZ.INCH-MOD.V.INTR**
   ‘turn white/bright’

Derivational operators not deriving a new lexical class are not specified as derivations and just labeled by function.
(14) Derivational operators not deriving new class

K’ICHE’

a. aj-chak
AGT-N:work
‘worker’

b. aq’ab’-al
N:night-ABSTR
‘darkness’

c. saq-soj
ADJ:white-MODER
‘moderately white’

Derivations with zero-marking.

(15) Derivations with zero-marking

K’ICHE’

a. saq-ø
ADJ:white-NMLZ
‘light’

YUKATEK

b. k-in-tz’ú’utz’-ø-ik-ø
INC-1s.ERG-N:kiss-TRVZ-INC.V.TR-3s.ABS
‘I kiss him/her’

Linguistic descriptions of Mayan languages often specify the derivational basis of a derivational operator in a gloss. For example, “INTRVZ.POS” for intransitivisers from positional roots. However, since the root/stem that functions as the derivational basis is glossed in the XML-annotation scheme for its lexical category, the overspecification is not necessary and therefore generally omitted.

(16) Overspecification of lexical basis in derivational glosses

CLASSIC MAYAN

a. chum-wan-ø=iy
POS:sitting-INTRVZ-POS-3s.ABS=ANT
‘s/he sat down’

Semantic labeling of lexical categories

The meaning of lexical categories is glossed in English. According to the LGR, the lexical category label is not reproduced in the gloss. The XML-annotation contains that information. Multiple meanings of a root or stem are likewise annotated in the XML-scheme, however, the gloss only contains the core meaning most applicable in the context.
(17) **Semantic labeling of lexical categories**

**K’ICHE’**

a. aj-q’ij or: aj-q’ij  
AGT-day AGT-N:day  
‘diviner = day-er’  

```xml
<lemma xml:id="l1" class="N">q’ij</lemma>
<ref target="l1" type="translation">sun</ref>
<ref target="l1" type="translation">day</ref>
<ref target="l1" type="translation">heat</ref>
```

If the translation of the lemma or root contains more than one lexical element, these are separated by a period.

(18) **Semantic glosses consisting of more than one element**

**CH’OL**

a. tza’ jul-i-ø  
COM  **arrive.here**-COM.V.INTR-3s.ABS  
‘she arrived here’

b. tza’ k’ot-i-ø  
COM  **arrive else where**-COM.V.INTR-3s.ABS  
‘she arrived there’

The meanings of some verbs are formed with directionals accompanying the verb. The lexical meaning is not glossed, but expressed through the translation.

(19) **Complex semantics of verbs accompanied by directionals**

**K’ICHE’**

a. k-ø-u-k’am uloq  
INC-3s.ABS-3s.ERG-receive DIR:**towards.speaker**  
‘he brings it’

b. k-ø-u-k’am ub’ik  
INC-3s.ABS-3s.ERG-receive DIR:**away.from.speaker**  
‘he takes it’

**CH’OL**

c. wol-ix a-ch’im-ø maji-el  
PROG-already 2s.ERG-take-3s.ABS DIR:**place.of.addressee**-DIR.V.INTR  
‘you are already taking it away’

d. wol-ix a-ch’im-ø sujt-el  
PROG-already 2s.ERG-take-3s.ABS DIR:**place.away.from.addressee**-DIR.V.INTR  
‘you are already taking it home’

In lexicalised noun phrases or lexicalised predicative expressions that consist of a verb and a specific noun in the function of direct object or subject the lexical annotation is solved under XML, but not considered in the gloss.
(20) Lexicalised phrases

Ch’OL

a. tyoj-Ø  i-pusik’al  wiňik
   POS:be.straight-3s.ABS  3s.ERG-heart  man
   ‘straight is the heart of the man’
   “the man is honest”

When the meaning of lexical roots is not known, they are glossed with “?”.

(21) Lexical roots with unknown meaning

K’iche’

a. u-mop-il
   3s.POSS-?-ABSTR/RELZ
   ‘budding (of flowers)’

When compounds are only in part semantically transparent, the intransparent part is glossed with “?”.
The meaning of the compound as a lemma is annotated in the XML-scheme and can be retrieved.

(22) Compounds with semantically intransparent parts

Ch’OL

a. i-b’oj-tye’-lel    i-b’ojtye’-lel
   3s.POSS-?-wood-RELZ    3s.POSS-pole.wall-RELZ
   ‘his wall’     ‘his wall’

Derived stems that have lexicalised by undergoing phonological change and are not morphologically
transparent to the speaker are glossed with the semantic gloss of the root and the gloss of the
derivational category separated by a period. Segmentable morphology is always glossed, even if
derivations are non-productive.

(23) Glossing of non-segmentable morphology

Ch’OL

a. tziňsa-ñ  but:  chim-sa-ñ
   die.CAUS-MOD.V.TR.D  die-TRVZ.CAUS-MOD.V.TR.D
   ‘kill’      ‘kill’

b. otzi-b’e-ñ  but:  och-sa-b’e-ñ
   enter.CAUS-APPL-MOD.V.TR.D  enter-TRVZ.CAUS-APPL-MOD.V.TR.D
   ‘put’      ‘put’

When grammatical morphemes have grammaticalised as part of the verb stem and are non-
segmentable, the semantic gloss of the lexical stem and the grammatical category are separated by a
period.
(24) Non-segmentable categories

CH’OL

`che’eñ` but: `che’-ob’
say.3s.ABS say-3p.ABS
‘he says’ ‘they say’

Non-overt elements

Non-overt elements are generally marked with ø, if they form part of a paradigm. All Mayan languages mark the third person singular absolutive as zero.

(25) Non-overt elements

K’ICHE’

a. x-ø-u-b’i-j
COM-3s.ABS-3s.ERG-say-MOD.V.TR.D
’s/he said it’

b. ø=winaq
3s.ABS=human
’s/he is human’

Bipartite elements

No examples of bipartite lexemes have been analysed in Mayan languages. Bipartite grammatical morphemes are however an attested feature and marked by repetition of the gloss.

(26) Bipartite elements

YUKATEK

x-tzaj-ab’
INSTR-fry-INSTR
‘instrument for frying = pan’

Infexion and stem changes

Infaxis are not marked following the LGR conventions as <infix>, since this would not only interfere with XML-annotation using <xxx>-tags, but also complicate searching for the lexical root. In these cases, the stem is glossed by meaning and grammatical function and the root meaning is inserted as a separate reference into the annotation scheme. Infexion is for instance attested in a nominalisation process in Tzeltal, where h is inserted after the root vowel of transitive stems. The following example gives both the gloss and the XML-annotation with the separate reference to the root.

4 "As expected, there are also infixes that occur before the final element of their hosts. In the Mayan language Tzeltal, a group of numeral classifiers is derived from verbs by infixation of h before the final consonant (when the latter is a stop or an affricate; in all other cases, h is deleted; see Kaufman 1971). Examples of this phenomenon include the following: huht ‘holes’, from hut ‘be perforated’; lihk ‘ropes, cords’, from lik ‘carry’, and peht ‘handfuls of wood’, from pet ‘embrace (below the arms)’."
(27) Infixation

TZELTAL

a. huht
perforate.NMLZ
‘hole’

The same rule applies to grammatical changes of stems in the formation of passive and antipassive, which is a common feature in some Mayan languages. For example:

(28) Passive and antipassive stem changes

YUKATEK

a. k-u-ko’on-ol
HAB-3s.ERG-sell.PASS-INC.V.INTR
‘it is sold’

CH’OL

b. tza’ mijk-i-ø
COM cover.PASS-COM.V.INTR-3s.ABS
‘s/he was covered (wrapped, hidden)’

Incorporation of verbs and adverbs

In several Mayan languages, adverbial particles can be incorporated into the verb structure. In Western Mayan languages, such adverbials occur between the aspect- and ergative-markers. In the glossing, these adverbials are treated as affixes and separated by hyphens. In the Eastern Mayan language K’iche’, the occurrence of such adverbs is only attested with incorporated directionals and indicates separate prosodic forms (30b).

(29) Incorporation of adverbs

CH’OL

tza´-ix-ab’i i-k’uñ-chuk-u-ø-yob’
COM-already-REPRT 3p.ERG-ADV:finally-capture.COM.V.TR-3s.ABS-3.PL
‘they finally captured him’
6 Incorporation of directionals and adverbs

K’ICHE’

a. x-in-ul-r-il-a’
   COM-1s.ABS-DIR:come-3s.ERG-see-MOD.DIR.V.TR
   ‘he came to see you’

b. x-ø-b’e-k’u-ya’-oq
   COM-3s.ABS-DIR:go-ADV:then-give.PASS-MOD.DIR.V.INTR
   ‘s/he then went to be given’

Comments on the glossing of selected functional categories

The following section summarises the suggestions for some glossing conventions that were discussed during the workshop. The selection includes cases that require particular comment. The argument does not claim to be comprehensive in neither of the cases.

Grammatical relations

Although the present paper does not treat the glossing of syntactic features, the following abbreviations have been reserved to mark grammatical relations. The nomenclature follows Dixon (1994) and part of the general LGR.

S  =  subject of intransitive predicate
A  =  agent; subject of transitive predicate
O  =  object; patient of transitive predicate

Accordingly, the abbreviations SBJ and OBJ have been reserved to gloss syntactic constituents and are not used for morphological glossing.

Lexical classes

The lexical classes comprise root categories and closed word classes with grammatical functions. Root categories in Mayan languages include:

N  =  noun
V.INTR  =  intransitive verb
V.TR  =  transitive verb
ADJ  =  adjective
ADV  =  adverb
POS  =  positional
PART  =  particle
PRO  =  pronoun

Closed word classes include:

ART  =  article
CLF  =  classifier
CONJ  =  conjunction
DEM  =  demonstrative
EXIS = existential  
INT = interrogative  
NUM = numeral  
PREP = preposition  
RN = relational noun

Person categories

As it is the premise to gloss grammatical function, the practice of glossing pronouns by pronominal sets “A” and “B” that is common practice in Mayan linguistics is not followed here. Instead pronominal markers are glossed by person category and grammatical function.

Person-marking on verbs distinguishes absolutive pronouns (ABS) that mark S and O and ergative pronouns (ERG) that mark A. In Mayan languages with a split ergative system, ERG also marks S in a subset of intransitive verbal constructions.

Possessor-marking on nouns is glossed separately as POSS, as not all Mayan languages employ the same sets of pronouns for this function. In most Mayan languages nominal predication (PRED) is marked with absolutive pronouns.

Person categories are glossed with numbers 1-3 and an abbreviation indicating singular or plural. The LGR use \textit{SG} for singular and \textit{PL} for plural. It is suggested here to gloss singular and plural person categories in Mayan languages as \textit{s} and \textit{p} instead.

The labeling of singular and plural categories with lower case letters is inconsistent with the LGRs. However, lower case letters are chosen here to avoid confusion, as the LGR do not allow for clear distinction between nominal plural marking and plural suffixes in bipartite plural person marking as it occurs in most Western Mayan languages. In these languages, nominal plural, the third person absolutive pronoun and the plural complement of third person plural possessive/ergative marking are all marked by the same suffix. To allow for differentiation of all three functions, we suggest to gloss the plural complement as 3.PL. This solution is however not ideal and can lead to potential confusion, as the LGRs employ the same gloss to refer to the third person plural (3p). An alternative solution might still be preferable in this case.

Differentiating nominal and verbal plural marking

\begin{enumerate}
\item K’iche’
\begin{enumerate}
\item k-in-war-ik
\textit{INC-1s.ABS-sleep-MOD.V.INTR}  
‘I sleep’  
\item x-ø-q-eta’ma-j
\textit{COM-3s.ABS-1p.ERG-learn-MOD.V.TR.D}  
‘we learned it = we know’
\end{enumerate}
\end{enumerate}

The labeling of singular and plural categories with lower case letters is inconsistent with the LGRs. However, lower case letters are chosen here to avoid confusion, as the LGR do not allow for clear distinction between nominal plural marking and plural suffixes in bipartite plural person marking as it occurs in most Western Mayan languages. In these languages, nominal plural, the third person absolutive pronoun and the plural complement of third person plural possessive/ergative marking are all marked by the same suffix. To allow for differentiation of all three functions, we suggest to gloss the plural complement as 3.PL. This solution is however not ideal and can lead to potential confusion, as the LGRs employ the same gloss to refer to the third person plural (3p). An alternative solution might still be preferable in this case.

\begin{enumerate}
\item Ch’ol
\begin{enumerate}
\item iy-alob’il-ob’
\textit{3s.POSS-child-PL}  
‘his/her children’
\item iy-alob’il-ob’
\textit{3p.POSS-child-3.PL}  
‘their children’
\end{enumerate}
\end{enumerate}
b. tzi’- tziña-yob’ cf. tzi’- tziña-ø-yob’
COM.3s.ERG-die.CAUS-3p.ABS COM.3p.ERG-die.CAUS-3s.ABS-3.PL
‘s/he killed them’ ‘they killed him/her/it’
YUKATEK

c. k-u-kiims-ik-o’ob’ cf. k-u-kiims-ik-ø-o’ob’
INC-3s.ERG-die.CAUS-INC-3p.ABS INC-3p.ERG-die.CAUS-INC-3s.ABS-3.PL
‘s/he kills them’ ‘they kill him/her/it’

In Ch’ol, aspect markers or prepositions can fuse with the ergative prefix, which is analysed as a non-segmentable category. The phenomenon is also attested for other Mayan languages.

(33) Non-segmentable aspect-markers and prepositions

Ch’OL

a. mi’- y-il-ø
INC.3s.ERG-3s.ERG-say-3s.ABS

b. tzi’- mel-e-ø
COM.3s.ERG-make-COM.V.TR-3s.ABS

c. tyi’- y-ity
PREP.3s.POSS-3s.POSS-buttocks

Some Western Mayan languages have an inclusive/exclusive contrast in the first person plural. The inclusive/exclusive gloss is inserted behind the person category 1p, separated by a period.

(34) Inclusive/exclusive contrast

Ch’OL

a. lak-ña’
1p.INCL.POSS-mother
‘our mother (inclusive)’

b. k-ña’ lojoñ
1p.EXCL.POSS-mother 1p.EXCL.POSS
‘our mother (exclusive)’

c. tza’ letz-i-yoñla
COM ascend-COM.V.INTR-1p.INCL.ABS
‘we (inclusive) ascended’

d. mi-j- k’el-e-yety-lojoñ
INC-1p.EXCL.ERG-see-COM.V.TR-2s.ABS-1p.EXCL.ERG
‘we (exclusive) saw you’

Inclusive/exclusive marking is also attested in Tzotzil. In the following example, the inclusive is marked on the plural marker.
Inclusive/exclusive contrast in Tzotzil (Vinogradov 2014:43)

Tzotzil

ch-i-tzak-at-otik
INC-1p.ABS-catch-PASS-1.INCL.PL
‘we would be caught’

K’iche’ is the only Mayan language that has a formal person, which is not marked on the reference verb/noun, but by a free pronominal particle in postposition. As a gloss for this formal person the abbreviation FORM is selected.

Formal person

K’iche’

a. k-inw-il la
INC-1s.ERG-see 2s.ABS.FORM
‘I see you (formal)’

b. x-oj-il alaq
COM-1p.ABS-see 2p.ERG.FORM
‘you (pl. formal) saw us’

Person categories are combined in the gloss with the grammatical function of the marker, i.e. ABS, ERG and POSS.

Person categories and their grammatical functions

K’iche’

a. k-in-war-ik
INC-1s.ABS-sleep-MOD.V.INTR
‘I sleep’

b. k-at-in-ch’ay-o
INC-2s.ABS-1s.ERG-hit-MOD.V.TR
‘I hit you’

c. nu-tat
1s.POSS-father
‘my father’

Preconsonantal and prevocalic forms and other forms of phonological assimilation in bound pronouns are not distinguished by different glosses. In Ch’ol the first person singular ergative marker k- becomes j- before consonants k and k’, i.e. k- → j- / _[k].

Phonological change/alternation in bound pronouns

K’iche’

a. x-Ø-a-b’an-o ~ x-Ø-aw-il-o
COM-3s.ABS-2s.ERG-make-MOD.V.TR COM-3s.ABS-2s.ERG-see-MOD.V.TR
‘you made it’ ‘you saw it’
Although most linguists gloss the person category on nominal predicates as an absolutive pronoun, this practice is inconsistent with the premise that only grammatical function glossed. We therefore suggest to use the abbreviation PRED to gloss person in these constructions (see also Vinogradov 2014).

(39) Person categories in nominal predicates

K’iche’

a. in achi
   1s.PRED man
   ‘I am a man’

Ch’ol

b. k-pi’il-ety
   1s.POSS-friend-2s.PRED
   ‘you [are] my friend’

c. b’uch-ul-ety
   POS:sitting-ADJVZ-2s.PRED
   ‘you are (in the position of) sitting’

d. kol-em-ø jiñi otyoty
   grow-PTCP-3s.PRED ART house
   ‘this house [is] big’

Independent pronouns in Mayan languages are combinations of one set of dependent pronouns and determiners in form of articles or demonstratives. In many Mayan languages these forms have fused, in some they are still separated. In K’iche’ the independent pronoun is identical with the absolutive in the first and second person, in the third there is a separate free form. The free forms can combine with articles ri or le to form or occur individually. In these cases, articles and pronouns are glossed individually. In languages where the independent pronoun is a lexicalised complex form, the entire form is glossed (e.g. in Ch’ol).

(40) Glossing of independent pronouns

K’iche’

a. (ri) in in kos-inaq
   ART 1s.PRO 1s.PRED tired-PTCP
   ‘I am tired’

b. ri are’ ø kos-inaq
   ART 3s.PRO 3s.PRED tired-PTCP
   ‘s/he is tired’
Possessive constructions

Mayan languages distinguish alienably and inalienably possessed nouns, which fall into different classes depending on their respective marking patterns. A certain set of inalienably possessed nouns are marked with an absoluble suffix, when occurring in unpossessed contexts.

(41) Absoluble suffixes on unpossessed inalienably possessed nouns

K’ICHE’

<table>
<thead>
<tr>
<th>Possessor</th>
<th>Possessum</th>
<th>Suffix</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. r-aqan</td>
<td>aqan-aj</td>
<td>foot/leg-ABSL</td>
</tr>
<tr>
<td>3s.POSS-foot/leg</td>
<td>‘his/her foot/leg’</td>
<td></td>
</tr>
<tr>
<td>b. u-k’ajol</td>
<td>k’ajol-axel</td>
<td>son.of.father-ABSL</td>
</tr>
<tr>
<td>3s.POSS-son.of.father</td>
<td>‘his son’</td>
<td></td>
</tr>
<tr>
<td>c. i-chol</td>
<td>chol-el</td>
<td>maizefield-ABSL</td>
</tr>
<tr>
<td>3s.POSS-maizefield</td>
<td>‘his/her maizefield’</td>
<td>‘maizefield (unpossessed)’</td>
</tr>
<tr>
<td>d. j-k’ib’</td>
<td>k’ib’-il</td>
<td>arm-ABSL</td>
</tr>
<tr>
<td>1s.POSS-arm</td>
<td>‘my arm’</td>
<td>‘arm, branch (unpossessed)’</td>
</tr>
<tr>
<td>e. a-chich</td>
<td>chich-il</td>
<td>older sister-ABSL</td>
</tr>
<tr>
<td>2s.POSS-older sister</td>
<td>‘your older sister’</td>
<td>‘older sister (unpossessed)’</td>
</tr>
</tbody>
</table>

CH’OL

<table>
<thead>
<tr>
<th>Possessor</th>
<th>Possessum</th>
<th>Suffix</th>
</tr>
</thead>
<tbody>
<tr>
<td>c. joñoñ</td>
<td>k-ujil</td>
<td>e’tyel</td>
</tr>
<tr>
<td>1s.PRO</td>
<td>1s.ERG-be.able.to</td>
<td>work</td>
</tr>
<tr>
<td>‘I am able to work’</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Inalienably possessed nouns which describe a relation to the human body or entity generally take a suffix (mostly –Vl) that marks the partitive relationship and is glossed as a relationaliser.

(42) Relationaliser suffixes on inalienably possessed nouns

K’ICHE’

<table>
<thead>
<tr>
<th>Possessor</th>
<th>Possessum</th>
<th>Suffix</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. u-b’aq</td>
<td>u-b’aq-il</td>
<td>3s.POSS-bone-RELZ</td>
</tr>
<tr>
<td>3s.POSS-bone</td>
<td>‘his/her bone’</td>
<td>‘his/her bone’</td>
</tr>
<tr>
<td>Ch’OL</td>
<td>alienable/non-partitive</td>
<td>inalienable/partitive</td>
</tr>
<tr>
<td>b. i-k’ajk</td>
<td>i-k’ajk-al</td>
<td>3s.POSS-fire-RELZ</td>
</tr>
<tr>
<td>3s.POSS-fire</td>
<td>‘fire’</td>
<td>‘his/her fire = his/her fever’</td>
</tr>
<tr>
<td>Ch’OL</td>
<td>alienable/non-partitive</td>
<td>inalienable/partitive</td>
</tr>
</tbody>
</table>
**Relational nouns and complex prepositions**

Relational nouns are a common feature in Mayan as well as most Mesoamerican languages, which constitute a structural as well as a functional category. The term refers to a closed class of functionally restricted, inalienably possessed nouns which reference a syntactic relation and thus have prepositional function. These nouns can be body part terms referencing clear spatial relations as well as other roots referencing a non-spatial relation ('with', 'by/through/because of', 'for the benefit of', 'alone' etc.). Under XML, the lexical roots of relational nouns are annotated for their word class (RN) and for their functional meaning (e.g. BEN, COMIT, CAUS).

(43)  Relational nouns with possessive person-marking

**K'iche’**

a. are’  ajq’ij  r-ech   tinamit
   3s.PRO  diviner  3s.POSS-RN.BEN  town
   ‘he is the diviner for/of the town’

b. x-ø-b’e   k-uk’
   COM-3s.ABS-go  3p.POSS-RN.COMIT
   ‘s/he went with them’

c. k-e-kun-a-x    r-umal
   INC-3s.ABS-N:healing-TRVZ-PASS  3s.POSS-RN.CAUS
   ‘they were healed by him’

Complex prepositions are structurally distinct from relational nouns, inasmuch as they combine a basic preposition with a body part-noun (N) that is marked with a possessor.

(44)  Complex prepositions with possessive person-marking

**Ch’ol**

a. tyi’-pam   mesa
   PREP.3s.POSS-N:face  N:table
   ‘on the face of the table = on the table’

   **K’iche’**

b. chi u-pam   ri  r-ochoch
   PREP  3s.POSS-N:stomach ART  3s.POSS-N:house
   ‘inside his house’

**Reflexives and indirect Objects**

Reflexives are treated in some grammars as part of the set of relational nouns. Syntactically, however, they are possessed transitive complements. Their function is not to establish a relationship with a
following NP, as it is the case with relational nouns/prepositions. Essentially, Mayan reflexives work the same way as in English and combine a possessor and a noun with the meaning ‘self’; they also include reciprocal readings. Reflexives are nevertheless glossed as a grammatical category.

(45) Reflexive constructions

K’ICHE’

a. k-ø-inw-il w-ib’ → k-ø-inw-il w-ib’
INC-3s.ABS-1s.ERG-see 1s.POSS:N:REFL
‘I see (it) my self = I see myself’
‘I see myself’

YUKATEK

b. k-in-jatz’-ik-ø in-b’a
HAB-1s.ERG-beat-INC.V.TR-3s.ABS 1s.POSS:N:REFL
‘I beat (it) my self = I beat myself’

CH’OL

c. tz’i’- jatz’-i-ø-yob’ i-b’i
COM.3p.ERG-hit-COM.V.TR-3s.ABS.3.PL 3p.ERG-N:REFL
‘they hit their selves = they hit each other’

In most Mayan languages indirect objects are realised by oblique phrases introduced by prepositions. As grammaticalised forms they are often referred to as “dative pronouns”, which however does not adequately describe the form that is used.

(46) Indirect objects

K’ICHE’

a. k-ø-in-ya’ chi r-ech
INC-3s.ABS-1s.ERG-give PREP 3s.POSS-RN.BEN
‘I give it to his benefit/possession = I give it to him’

YUKATEK

b. k-in-tz’a’-ik-ø t-eech
HAB-1s.ERG-give-INC.V.TR-3s.ABS PREP-2s.ABS
‘I give it to you’

Agentives

There are different types of agentive nominalisation in Mayan languages. All Mayan languages share the feature of agentive prefixes or proclitics, which precede nominal and adjectival stems, or even nominal phrases, to derive agentive nouns.

(47) Agentive prefixes/proclitics

K’ICHE’

a. aj-chak
AGT-work
‘worker’
b. **aj-r-el-ib’al**  
   AGT-3s.POSS-emerge-NMLZ-INSTR sun  
   ‘eastener’  
   **YUKATEK**

c. **h-tz’óon**  
   AGT-hunt.AP  
   ‘hunter’

Yukatek seems to be the only Mayan language that distinguishes masculine and feminine agents morphologically. Masculine agents are marked with **h-** while feminine agents are marked with **š-**. The gender distinction is marked in the gloss.

(48) **Gender distinction in agentive prefixes/proclitics in Yukatek**

**YUKATEK**  

\[
\begin{align*}
   h-kòon-ol & \quad x-kòon-ol \\
   AGT.M-sell.AP-ABSTR & \quad cf. \quad AGT.F-sell.AP-ABSTR \\
   ‘salesman (= the one of selling)’ & \quad ‘saleswoman (= the one of selling)’
\end{align*}
\]

Etymologically, **h-** derives from the gender-non-specific agentive **aj** found across the language family, while **x-** is clearly related to the likewise common female nominal classifier (*i)x. Only in Yukatek both markers developed into a gender-based paradigm. In Classic Mayan classifier and agentive can co-occur in the same word, e.g. **Ix Aj k’uhun** [IX-AJ-K’UH-HU’N-(na)] ‘female venerator/keeper’ (Jackson & Stuart 2001).

**Positionals**

Positional roots are a distinctive feature in the Mayan language family. Yet, in some cases there is no clear consensus about what constitutes a positional root. In many Mayan languages, positional roots do not occur on their own and require a derivational operator. The meaning of the positional root is glossed with an English verbal noun.

(49) **Glossing of positional roots**

**K’ICHE’**

| a. kø-u-kotz’-ob’a’ | ri ab’aj |
| INC-3s.ABS-3s.ERG-POS:lie.down-TRVZ | ART stone |
| ‘he laid down the stone’ |
| **Ch’OL** |

| b. mi’- b’uch-tyi-l | tyi lum |
| INC.3s.ERG-POS:sitting-INTRVZ-INC.V.INTR | PREP earth |
| ‘he sits on the ground’ |

| c. b’uch-ul-oñ | |
| POS:sitting-ADJVZ-1s.PRED | |
| ‘I am (in a) sitting (position)’ |
### Preliminary list of glossing conventions

<table>
<thead>
<tr>
<th></th>
<th>First person</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><strong>first person</strong></td>
</tr>
<tr>
<td>2</td>
<td><strong>second person</strong></td>
</tr>
<tr>
<td>3</td>
<td><strong>third person</strong></td>
</tr>
<tr>
<td>A</td>
<td><strong>agent-like argument in canonical transitive verb</strong></td>
</tr>
<tr>
<td>ABS</td>
<td><strong>absolutive</strong></td>
</tr>
<tr>
<td>ABSL</td>
<td><strong>absoluble</strong></td>
</tr>
<tr>
<td>ABSTR</td>
<td><strong>abstractive</strong></td>
</tr>
<tr>
<td>ADJ</td>
<td><strong>adjective</strong></td>
</tr>
<tr>
<td>ADJVZ</td>
<td><strong>adjectivizer</strong></td>
</tr>
<tr>
<td>ADV</td>
<td><strong>adverb(ial)</strong></td>
</tr>
<tr>
<td>ADVLZ</td>
<td><strong>adverbializer</strong></td>
</tr>
<tr>
<td>AFF</td>
<td><strong>affirmative</strong></td>
</tr>
<tr>
<td>AGT</td>
<td><strong>agentive</strong></td>
</tr>
<tr>
<td>ANT</td>
<td><strong>anterior</strong></td>
</tr>
<tr>
<td>AP</td>
<td><strong>antipassive</strong></td>
</tr>
<tr>
<td>APPL</td>
<td><strong>applicative</strong></td>
</tr>
<tr>
<td>ART</td>
<td><strong>article</strong></td>
</tr>
<tr>
<td>ASS</td>
<td><strong>assertive</strong></td>
</tr>
<tr>
<td>AUX</td>
<td><strong>auxiliary</strong></td>
</tr>
<tr>
<td>BEN</td>
<td><strong>benefactive</strong></td>
</tr>
<tr>
<td>CAUS</td>
<td><strong>causative</strong></td>
</tr>
<tr>
<td>CLF</td>
<td><strong>classifier</strong></td>
</tr>
<tr>
<td>COMIT</td>
<td><strong>comitative</strong></td>
</tr>
<tr>
<td>COM</td>
<td><strong>completive</strong></td>
</tr>
<tr>
<td>COND</td>
<td><strong>conditional</strong></td>
</tr>
<tr>
<td>CONJ</td>
<td><strong>conjunction</strong></td>
</tr>
<tr>
<td>COP</td>
<td><strong>copula</strong></td>
</tr>
<tr>
<td>CVB</td>
<td><strong>converb</strong></td>
</tr>
<tr>
<td>DEF</td>
<td><strong>definite</strong></td>
</tr>
<tr>
<td>DEM</td>
<td><strong>demonstrative</strong></td>
</tr>
<tr>
<td>DET</td>
<td><strong>determiner</strong></td>
</tr>
<tr>
<td>DIM</td>
<td><strong>diminutive</strong></td>
</tr>
<tr>
<td>DIR</td>
<td><strong>directional</strong></td>
</tr>
<tr>
<td>DIST</td>
<td><strong>distal</strong></td>
</tr>
<tr>
<td>DISTR</td>
<td><strong>distributive</strong></td>
</tr>
<tr>
<td>DU</td>
<td><strong>dual</strong></td>
</tr>
<tr>
<td>DUB</td>
<td><strong>dubitative</strong></td>
</tr>
<tr>
<td>DUR</td>
<td><strong>durative</strong></td>
</tr>
<tr>
<td>EMPH</td>
<td><strong>emphasis</strong></td>
</tr>
<tr>
<td>ERG</td>
<td><strong>ergative</strong></td>
</tr>
<tr>
<td>EXCL</td>
<td><strong>exclusive</strong></td>
</tr>
<tr>
<td>EXIS</td>
<td><strong>existential</strong></td>
</tr>
<tr>
<td>F</td>
<td><strong>feminine</strong></td>
</tr>
<tr>
<td>FOC</td>
<td><strong>focus</strong></td>
</tr>
<tr>
<td>FORM</td>
<td><strong>formal</strong></td>
</tr>
<tr>
<td>FREQ</td>
<td><strong>frequentative</strong></td>
</tr>
<tr>
<td>FUT</td>
<td><strong>future</strong></td>
</tr>
<tr>
<td>IMP</td>
<td><strong>imperative</strong></td>
</tr>
<tr>
<td>INC</td>
<td><strong>aspect, incompletive</strong></td>
</tr>
<tr>
<td>INCH</td>
<td><strong>inchoative</strong></td>
</tr>
<tr>
<td>INCL</td>
<td><strong>inclusive</strong></td>
</tr>
<tr>
<td>INDF</td>
<td><strong>indefinite</strong></td>
</tr>
<tr>
<td>INSTR</td>
<td><strong>instrumental</strong></td>
</tr>
<tr>
<td>INT</td>
<td><strong>interrogative, question markers</strong></td>
</tr>
<tr>
<td>INTENS</td>
<td><strong>intensifier</strong></td>
</tr>
<tr>
<td>INTR</td>
<td><strong>intransitive</strong></td>
</tr>
<tr>
<td>INTRVZ</td>
<td><strong>intransitivizer</strong></td>
</tr>
<tr>
<td>IPFV</td>
<td><strong>imperfective</strong></td>
</tr>
<tr>
<td>IRR</td>
<td><strong>irrealis</strong></td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>LD</td>
<td>left dislocation</td>
</tr>
<tr>
<td>LEN</td>
<td>lentitive</td>
</tr>
<tr>
<td>LOC</td>
<td>locative</td>
</tr>
<tr>
<td>M</td>
<td>masculine</td>
</tr>
<tr>
<td>MOD</td>
<td>modal marker</td>
</tr>
<tr>
<td>MODER</td>
<td>moderative</td>
</tr>
<tr>
<td>N</td>
<td>noun (lexical root category)</td>
</tr>
<tr>
<td>NEG</td>
<td>negation, negative</td>
</tr>
<tr>
<td>NMLZ</td>
<td>nominalizer/nominalization</td>
</tr>
<tr>
<td>NN</td>
<td>unknown</td>
</tr>
<tr>
<td>NUM</td>
<td>numeral</td>
</tr>
<tr>
<td>O/P</td>
<td>patient-like argument in canonical transitive verbs</td>
</tr>
<tr>
<td>OBJ</td>
<td>object</td>
</tr>
<tr>
<td>OBL</td>
<td>oblique (syntactic gloss)</td>
</tr>
<tr>
<td>OPT</td>
<td>optative</td>
</tr>
<tr>
<td>p</td>
<td>plural in person categories</td>
</tr>
<tr>
<td>PART</td>
<td>particle</td>
</tr>
<tr>
<td>PASS</td>
<td>passive</td>
</tr>
<tr>
<td>PFV</td>
<td>perfective</td>
</tr>
<tr>
<td>PL</td>
<td>plural (on nominal categories)</td>
</tr>
<tr>
<td>POS</td>
<td>positional (ROOT)</td>
</tr>
<tr>
<td>POSS</td>
<td>possessive</td>
</tr>
<tr>
<td>POT</td>
<td>potential (aspect)</td>
</tr>
<tr>
<td>PRED</td>
<td>predicative (syntactic)</td>
</tr>
<tr>
<td>PREP</td>
<td>preposition</td>
</tr>
<tr>
<td>PRF</td>
<td>perfect</td>
</tr>
<tr>
<td>PRO</td>
<td>pronoun</td>
</tr>
<tr>
<td>PROG</td>
<td>progressive</td>
</tr>
<tr>
<td>PROH</td>
<td>prohibitive</td>
</tr>
<tr>
<td>PROX</td>
<td>proximal/proximate</td>
</tr>
<tr>
<td>PST</td>
<td>past</td>
</tr>
<tr>
<td>PTCP</td>
<td>participle</td>
</tr>
<tr>
<td>PURP</td>
<td>purposive</td>
</tr>
<tr>
<td>QUOT</td>
<td>quotative</td>
</tr>
<tr>
<td>RECP</td>
<td>reciprocal</td>
</tr>
<tr>
<td>REFL</td>
<td>reflexive</td>
</tr>
<tr>
<td>REL</td>
<td>relative</td>
</tr>
<tr>
<td>RELZ</td>
<td>relationalizer</td>
</tr>
<tr>
<td>REP</td>
<td>repetitive</td>
</tr>
<tr>
<td>REPRT</td>
<td>reportative</td>
</tr>
<tr>
<td>RES</td>
<td>resultative</td>
</tr>
<tr>
<td>RN</td>
<td>relational noun</td>
</tr>
<tr>
<td>SBJ</td>
<td>subject</td>
</tr>
<tr>
<td>s</td>
<td>singular in person categories</td>
</tr>
<tr>
<td>S</td>
<td>single argument of canonical intransitive verb</td>
</tr>
<tr>
<td>SG</td>
<td>singular (on nominal categories)</td>
</tr>
<tr>
<td>STAT</td>
<td>stative</td>
</tr>
<tr>
<td>SUPER</td>
<td>superlative</td>
</tr>
<tr>
<td>TEMP</td>
<td>temporal</td>
</tr>
<tr>
<td>TOP</td>
<td>topic</td>
</tr>
<tr>
<td>TR</td>
<td>transitive</td>
</tr>
<tr>
<td>TRVZ</td>
<td>transitivization</td>
</tr>
<tr>
<td>V</td>
<td>verb (root)</td>
</tr>
<tr>
<td>VN</td>
<td>verbal noun</td>
</tr>
</tbody>
</table>
References

Academia de las Lenguas Mayas de Guatemala (ALMG)  

Avelino Becerra, Heriberto (ed.)  

Bricker, Victoria R., Po’ot Yah, E. & Dzul de Po’ot, O.  
1998 A Dictionary of the Maya Language: As Spoken in Hocabá, Yucatán. Salt Lake City: University of Utah Press.

Dürr, Michael  

Jackson, Sarah & David Stuart  

Kaufman, Terrence  

Kaufman, Terrence  

Lehmann, Christian  

Leipzig Glossing Rules (LGR)  
2008 The Leipzig Glossing Rules: Conventions for interlinear morpheme-by-morpheme glosses. (URL: https://www.eva.mpg.de/linguistik/pdf/LGR08.02.05.pdf)

López Ixcoy, Candelaria Dominga (Saqijix)  
1997 Ri Ukemiik ri K’ichee’ Ch’ab’al: Gramática K’ichee’. Guatemala: Cholsamaj

Montgomery, John  
2002 How to Read Maya Hieroglyphs. New York: Hippocrene Books

Romero, Sergio F.  
2006 Sociallinguistic Variation and Linguistic History in Mayan: The Case of K’ichee’. PhD thesis, Department of Linguistics, University of Pennsylvania

Sachse, Frauke, Michael Dürr & Christian W. R. Klingler  
accepted Digitale Erschließung und systematische Annotation kolonialer Lexikographien am Beispiel der Mayasprache K’iche’. DARIAH Working Papers. [expected 2016]

Smailus, Ortwin  

Verhoeven, Elisabeth  

Vinogradov, Igor  