Title: TlajtoliTranslate: a free, efficient, and customizable transcoder
from classical to modern Nahuatl orthography

Author: Espinoza-Campuzano Cesar Augusto¹

Institutional affiliation:

1: Posgrado en Ciencias Biológicas, Universidad Nacional Autónoma de México, Circuito Exterior S/N,
Ciudad Universitaria, Del. Coyoacán, 04510 CDMX, Mexico

Corresponding author:

Espinoza-Campuzano Cesar Augusto: ce13210@my.bristol.ac.uk

Abstract

Nahuatl is a language currently spoken by at least 1.6 million people in México. Traditionally, and in academic circles, the classical Nahuatl orthography has been prevalent. The classical orthography is based on the Spanish interpretation of Nahuatl phonetics and most Nahuatl texts (*corpus*) are written in the classical orthography. Conversely, Mexico's Public Education Ministry (SEP) teaches Nahuatl using the practical orthography which is simpler and more intuitive than the classical one. This disassociation between both systems prevents many Nahuatl-speakers from easily accessing the Nahuatl *corpus*. Here I present TlajtoliTranslate, a new free and intuitive transcoder that allows the swift and easy conversion of texts from the classical to the practical Nahuatl orthography. The transcoder rules can be customized by the user to obtain the desired results. This transcoder was programmed in R and uses the Shiny package to produce a webpage interface. TlajtoliTranslate is freely available online at <https://cesar-espinoza-campuzano.shinyapps.io/TlajtoliTranslate/>.

Keywords

Nahuatl, orthography, transcoding, Practical Nahuatl, Classical Nahuatl, Nawatl

Introduction

In Mexico, 6.1 % of the population speaks an indigenous language. Among these languages Nahuatl is the most prevalent with more than 1.6 million speakers in 2020 (INMUJERES 2021; INEGI 2020). Despite the rich archive of texts and documents written since the conquest (XVI century), many Nahuatl users are not able to access this *corpus* of documents for two main reasons. First, the illiteracy rate of the original language speakers in Mexico was still higher than 20% in 2020 (INEGI 2020; INMUJERES 2021). Second, a large proportion of the Nahuatl *corpus* (collection of texts) is written using the classical orthography which is different to the practical one that is used in basic education programs (Trager 2016; De la Cruz Cruz 2014).

Orthographic proposals

Different options for writing Nahuatl are currently used or have been proposed. Linguists use the International Phonetic Alphabet (IPA; De la Cruz Cruz, 2014) however, the symbols of the IPA are rarely used or taught to native Nahuatl speakers. Another option is a recently developed abugida system (Trager 2016) that follows closely the phonetics and syllabic clustering of the Nahuatl language. The proposed system uses new symbols which greatly depart from the Latin characters used in the classical and practical orthographies. Despite the advantages of this abugida, there is no tradition of the teaching or use of this system. Besides these proposals, the two main Nahuatl orthographies (Olko and Sullivan 2015) are:

Classical Nahuatl

Pre-conquest Nahuatl peoples used glyphic writing which consisted on a mixture of logographs and phonetic signs to write their language (Olko and Sullivan 2015; De la Cruz Cruz 2014). After the conquest, Latin letters used by Spaniards were adapted to represent the Nahuatl phonetics (Table 1). Because of its Spanish origin, the classical orthography has rules that accommodate Spanish speakers but lack any intuitive basis for native speakers. For example, the /k/ sound in the classical orthography follows the Spanish rules: depending on the vocal it precedes, it is sometimes represented with “c” and sometimes with a “qu”. There are similar problems to represent the sound /w/ which is sometimes represented as “hu” and sometimes as “uh”. Furthermore, the sound /s/ can be represented either by “c” or “z”.

Table 1. Comparison between orthographies. The practical or modern orthography depicted in the table is the implementation of modern orthography by Mexico’s Public Education Ministry (SEP)

Phoneme	Classical orthography	Practical orthography (SEP)
/a, a:/	A a	A a
/e, e:/	E e	E e
/i, i:/	I i	I i
/o, o:/	O o	O o
/p/	P p	P p

/t/	T t	T t
/k/	C c; Qu qu	K k
/kʷ/	Cu cu; -uc; Qu qu	Ku ku
/ʔ/	H h	J j
/t͡s/	Tz tz	Ts ts
/t͡ʃ/	Ch ch	Ch ch
/t͡l/	Tl tl	Tl tl
/s/	Z z; C c	S s
/ʃ/	X x	X x
/l/	L l	L l
/j/	Y y	Y y
/w/	Hu hu; uh	U u
/m/	M m	M m
/n/	N n	N n

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

The classical orthography was implemented in the colonial period by Motolina, Olmos, Sahagún and Carochi (Lockhart 1999; De la Cruz Cruz 2014). This orthography has been advanced by several authors (Andrews 1975; Karttunen 1992; Campbell and Karttunen 1989; Lockhart 1999; Olko and Sullivan 2015) and it is used by several universities (De la Cruz Cruz 2014) like the Institute of Ethnological Teaching and Research of Zacatecas (IDIEZ) of the Autonomous University of Zacatecas (UAZ), the University of Warsaw (UW), and the Autonomous University of Tlaxcala (UAT). Despite being prominent in research and institutions, the classical orthography is not completely standardized and currently there are several proposals which propose minor changes to a common framework (De la Cruz Cruz 2014). The classical orthography has four salient advantages: 1) most of the Nahuatl *corpus* dating from the colonial period onwards is written in this orthography which allows classical orthography users to readily access it; 2) most research about the Nahuatl language is written using some variation of this orthography; 3) most Nahuatl dictionaries are based on this classical orthography (Karttunen 1992; Universidad Nacional Autónoma de México 2012), and 4) Using the classical orthography promotes a continuity in the Nahuatl *corpus*.

16

17

The classical orthography presents two main drawbacks. First, pronunciation is not intuitive because it imposes rules from a different language onto Nahuatl. Second, it is not taught in Mexico's

1 elementary schools and, because of this, the prevalence of the classical orthography can isolate native
2 speakers from the Nahuatl *corpus* and the research about their language (Olko and Sullivan 2015).

3 *Practical or modern Nahuatl*

4 Accounting for the large proportion of Nahuatl speakers in the country's population, the Mexican
5 Public Education Ministry (SEP; *Secretaría de Educación Pública*) has implemented elementary
6 education syllabi in Nahuatl (De la Cruz Cruz 2014). Because of this, Nahuatl is one of the most studied
7 indigenous languages in Mexico. The SEP uses the practical (or modern) in its syllabi. Although this
8 practical orthography is based on the Latin alphabet, it follows some phonetical conventions to
9 simplify the reading and teaching of the language. The rules for transcoding from the classical to the
10 practical orthography are usually straightforward (Table 1 and Method) however, some exceptions
11 can be identified (see Exceptions in Method and Discussion).

12 The practical orthography has a more consistent pronunciation that diverges from Spanish
13 and the classical orthography. For example, /s/ sounds are consistently represented by "s" (Table 1).
14 This consistency has shown some empiric evidence for speeding the learning of reading skills in
15 elementary school settings (Trager 2016). Although an in-depth discussion of the advantages and
16 disadvantages of these systems is outside of the scope of this paper, the transcoder presented here
17 will encourage the swift and easy transition from the classical to the practical orthography.

18 Finally, the need for a transcoder between the classical and practical orthography has been
19 previously noted (Trager 2016). Trager (2016), identified an existing transcoder
20 (<https://lexiquetos.org/modernizador-nahuatl/>) however, this transcoder does not seem to be
21 operational. Because of this, a functional, efficient, and customizable transcoder is proposed here.

22 *Method*

23 TlajtoliTranslate is available online at [https://cesar-espinoza-](https://cesar-espinoza-campuzano.shinyapps.io/TlajtoliTranslate/)
24 [campuzano.shinyapps.io/TlajtoliTranslate/](https://cesar-espinoza-campuzano.shinyapps.io/TlajtoliTranslate/). The transcoder was written using the R programming
25 language (R Core Team 2020) in RStudio (RStudio Team 2019). The transcoder uses the

str_replace_all() function of the tidyverse package (Wickham et al. 2019) to transcode the input text from the classical to the practical Nahuatl orthography. To allow the transcoding code to be executed behind a graphic user interface, the Shiny package (Chang et al. 2017) was used to produce a reactive webpage (Figure 1A and D). Users can input text in two ways to TlajtoliTranslate, either by writing directly into the text input box or by uploading a text (.txt) file. The input text is passed to a server which runs the R script and returns the transcoded text to the user interface along with the transcoding time. If a text file was used, the user can download a transcoded text file. Finally, the user can specify some optional transcoding rules (see below; Figure 1F).

Figure 1. TlajtoliTranslate graphic user interface (GUI). The GUI is written in practical Nahuatl, Spanish, and English. Users may input text directly to the app (A) to obtain an immediate output (B) and a transcoding time (C). The users can also input plain .txt files (D), a transcoded text file can be downloaded by the user (E). Finally, TlajtoliTranslate allows the user to optionally specify different transcoding rules (F). (G) TlajtoliTranslate is available at: <https://cesar-espinoza-campuzano.shinyapps.io/TlajtoliTranslate/>. H) A link is provided to allow users to provide feedback on detected errors.

1 Default transcoding rules

2 The default output of the transcoder follows the practical system as implemented by Mexico's Public
3 Education Ministry (SEP):

- 4 • "U" is placed instead of the digraphs "hu" and "uh" (but see Optional Rules and Exceptions).
- 5 • "L" replaces all instances of "ll" (but see Optional Rules)
- 6 • "J" replaces all "h", except if h forms part of the digraph "ch"
- 7 • "K" is placed instead of "c" when "c" is followed by "a", "o", or "u".
- 8 "K" is placed instead of "c" when "c" is not part of the digraph "ch".
- 9 "K" replaces the digraph "qu" in "qui" and "que".
- 10 • "S" replaces "z".
- 11 "S" replaces "c" when it is followed by "e" or "i"

12 Exceptions

13 Some exceptions are used to accommodate words in which a phonetic change was detected (see
14 Discussion). On these words "uh" character represents the /ʔ/ sound instead of /w/:

- 15 • Words derived from *ihqui* ("in this way") are transcoded as *ijki* instead of *iuki*.
- 16 • Words with *ihca* like *ihcatzan* ("even though") are transcoded as *ijka* instead of *iuka*.
- 17 • Verbs of class two in the past or imperfect tenses which end in *uhqui*, *uhqueh*, *uhtoc*, *uhtoqueh*
18 (classical orthography) are transcoded as *jki*, *jkej*, *jtoc*, *jtokej*

19 Optional rules

20 The user may specify two extra rules that do not follow the practical system used by the SEP:

- 21 • "W" is used instead of "u" to represent the /w/ sound. This change is not used by the SEP but
22 is commonly used by Nahuatl speakers.

- “l” is not replaced at all. While the SEP replaces the “ll” for “l”, the consonant /l/ is often germinated in Nahuatl, thus it is sometimes convenient to preserve this “ll” to preserve the original syllabic clustering of the language (Trager 2016):

Results and Discussion

TlajtoliTranslate can easily and quickly transcode texts from the classical to the practical orthography. To assess the response time of the transcoder, a simple benchmark test was performed by recording the system times in R. The transcoding time was tested on two documents of different sizes: the first document was a small story of 37 words (Table 2) which was transcribed in 935 milliseconds (less than a second).

Table 2. Small 37-word text. This text was used to benchmark the human transcoding times. The text can be translated as: Great Nezahualpilli was in the field and said: “As soon as I started to sow, and a black bird ate all my seeds. How will I scare of the blackbirds? Even though I have a scarecrow in the field, the blackbirds know that it is not a real human. I know! I will live in the field and become my scarecrow!”

Classical nahuatl text	Text transcribed by TlajtoliTranslate
Huehuetzin Nezahualpilli	Ueuetsin Nesaualpili
itztoc pan milli	itstok pan mili
huan quiihto:	uan kiihto:
"Zan nipeuhqui niquitoca,	"San nipejki nikitoka,
huan ce zanatl quicua nochin noxinachhuan.	uan se sanatl kikua nochin noxinachuan.
Queniuhqui niquinnahcomanaz zanameh?	Kenijki nikinnajkomanas sanamej?
luhcatzan nicpiya ce tlapihpixquetl pan milli,	ljcatsan nikpiya se tlapijpiketl pan mili,
zanameh quimatih tlen axneliya macehualli.	sanamej kimatij tlen axneliya maseuali.
¡Nimatiya!	¡Nimatiya!
¡Ninemiz pan milli huan nimoeliz	¡Ninemis pan mili uan nimoelis
notlapihpiqueuh!"	notlapijpiqueu!"

To put this transcoding time, I compared to the time it took to several advanced-level classical Nahuatl students from the UAZ (n=7). These students had worked with the practical Nahuatl orthography for at least one academic term. The human transcoding process took on average 2.33 minutes (range=2.41 to 4.31 minutes) which meant a rate of 0.17 words per second. Finally, the human transcoding of these 37 words had 3.86 on average.

Second, the transcoder was used on a text file containing a Nahuatl textbook used to teach intermediate-level Nahuatl to non-native speakers UAZ students. The main text of the book (excluding Introduction and Dictionary sections) was written in the classical orthography and is 7851 words long. The transcoding of the long text took 735 milliseconds (slightly longer than half a second). Considering the recorded human transcoding times, transcoding these 7852 words would take an average student more than 22.7 hours of uninterrupted work. Furthermore, this transcoding would present more than 800 errors. In the light of the difference between human and machine transcoding times, I suggest that TlajtoliTranslate can potentially make the Nahuatl *corpus* available for native speakers by saving a significant amount of time and manual labour.

Potential pitfalls

Despite the good performance of the transcoder, caution should always be observed, and the manual revision of the produced texts is encouraged to avoid possible pitfalls. Furthermore, a link for feedback was included in TlajtoliTranslate (Fig 1H). Using this link, users can report errors or new exceptions identified in the output texts.

Exceptions and errors: The transcoder uses rules based on the SEP's implementation of the practical orthography. It replaces "uh" for "u" which is correct in most contexts: *e.g.*, like in the word for equity (*neneuhcayot* is transcoded as *neneukayot*) and the word for eagle (*cuauhtli* is transcoded as *kuautli*).

The first exception was found in some class-two verbs. Specifically, with verbs ending with "hua" like *chihua*. In the classical orthography, when conjugating one of these verbs in the past and imperfect tenses they drop the final "a". Afterwards, the verbs suffer an inversion from "hu" to "uh". Finally, a "qui" or other declination is added to the verb. The ending of these conjugated verbs in the classical orthography can be *uhqui*, *uhqueh*, *uhtoc* and *uhtoqueh* however, they must be transcoded as *jki*, *jkej*, *jtoc* and *jtokej* respectively.

The second exception deals with a phonetical phenomenon where "uh" actually represents the sound /ʔ/. Some of the identified words are related to *iuhqui* ("like this"): *iuhqui* should be transcoded as *ijki*,

not *iuki*. The word for “how” (*queniujqui*) should be transcoded as *kenijki*, not *keniuki*. Furthermore, words with “*ihca*” like *ihcatzan* (“even though”) should be transcoded as *ijkatsan* and not *iukatsan*. This phonetical phenomenon was identified in the Huastecan Nahuatl. Because of this, the phenomenon might not be shared amongst all Nahuatl variants or even amongst all words which include “*ihqui*” or “*ihca*”. These potential issues of different exceptions amongst variants will be assessed by using the users’ feedback.

Foreign words and loans: Caution is recommended when transcoding texts which are rich with foreign names or loans from different languages. Proper names like “Hector” are common in Mexico and they should preserve their correct spelling however, they will be automatically transcoded with an incorrect spelling (e.g., “Jektor”). Because of this, the manual revision of names is advised. Furthermore, Nahuatl has many loaned words: some of them are already adapted to Nahuatl like *maneharoa* (“to drive”) or *cahuayoh* (“horse”) and are correctly transcoded as “*manejaroa*” and “*kauayoj*”. However, some users may prefer to preserve the spelling of non-Nahuatl words and should correct the texts manually.

Nahuatl variants: There are between 4 and 30 Nahuatl variants (De la Cruz Cruz 2014). TlajtoliTranslate has been optimized to work with the Huastecan Nahuatl variant. It is possible that other variants may present different exceptions. Thanks to the feedback link, it will be possible to cater for the different exceptions from different Nahuatl variants.

Finally, although an in-depth discussion of the advantages and disadvantages of the different Nahuatl orthographies is beyond the scope of this paper, it must be stated that I disagree with the view of Olko & Sullivan (2015). These authors state that the practical orthography, and others not rooted in the classical one, “constitute an obstacle to language revitalization and native literacy” (Olko & Sullivan, 2015, p. 208). Part of their rationale against the modern orthography was that “it discourages indigenous people from reading and studying the great *corpus* of older works that constitute the written cultural legacy of the Nahua civilization” (Olko & Sullivan, 2015, p. 208). At least for this part

of their argument, I dare to say that transcoders like TlajtoliTranslate will eliminate this obstacle and allow even more indigenous people to access the Nahuatl *corpus*.

Future directions

It is possible that TlajtoliTranslate users may identify different optimal transcoding rules amongst different Nahuatl variants. Such differences can be catered by coding them as new optional rules so the users can pick a specific Nahuatl variant when using TlajtoliTranslate.

It may be of interest to implement a backwards transcoder to pass from the practical to the classical orthography. The backwards transcoding might pose some pitfalls. For example, the many-to-one correspondence of some letters between orthographies (e.g., “s” to “c” or “z”) and the effects of exceptions. Although the backwards transcoding pitfalls can probably be overcome, this kind of backwards-transcoding might not be as useful as the current implementation.

Finally, I agree with the view of Olko & Sullivan (2015) that the practical Nahuatl needs more standardization in the form of dictionaries. Because of this I encourage the practical orthography users to take on this endeavour within a multi-variant framework.

Conclusions

Here, I report TlajtoliTranslate, a transcoder from the classical to the practical Nahuatl orthography. The transcoder is freely available online and is capable of transcoding either the user’s direct input or .txt files.

The automated transcoding process is swift and reliable. The machine process of transcoding 7851 of words takes place in around a second while the human transcoding is prone to errors and would probably take more than 22 hours of uninterrupted work.

TlajtoliTranslate is customizable, users can select optional rules to modify its output and accommodate for the users’ writing preferences.

The transcoder implemented here is adaptable and will be able to accommodate for the errors and exceptions identified by the users or differences among Nahuatl variants.

Finally, since the Mexican public education uses the practical Nahuatl, many native speakers do not have access to the Nahuatl *corpus*. The transcoder implemented here can potentially allow native speakers to access the *corpus*.

List of abbreviations:

SEP: Mexico's Public Education Ministry (Secretaría de Educación Pública)

Declarations

- Ethics approval and consent to participate

- Consent for publication

Not applicable

- Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request. The small text transcoded is available in Table 2

- Competing interests

The author declares that he has no competing interests

- Funding

Not applicable

- Authors' contributions

CAEC conceived the project, coded the transcoder (TlajtoliTranslate) and wrote the manuscript.

- Acknowledgements

I would like to acknowledge Rebeca Vargas Tovar, Nahuatl professor at the Autonomous University of Zacatecas (UAZ) and Expert Translator of Nahuatl for the Penal Justice Centre of the State of Zacatecas (Centro de Justicia Penal Federal; Estado de Zacatecas) for her help in organizing the human transcoding time test.

- Authors' information (optional)

Not applicable

References

- Andrews, Richard J. 1975. *Introducción to Classical Náhuatl*. Austin: University of Texas Press.
- Campbell, Joe, and Frances Karttunen. 1989. *Foundation Course in Nahuatl Grammar*. Austin: Institute of Latin American Studies, The University of Texas.
- Chang, Winston, Joe Cheng, J Allaire, Yihui Xie, Jonathan McPherson, and others. 2017. "Shiny: Web Application Framework for R." *R Package Version 1* (5): 2017.
- INEGI. 2020. "Censo de Población y Vivienda."
- INMUJERES. 2021. "Población Indígena." 2021. http://estadistica.inmujeres.gob.mx/formas/tarjetas/Poblacion_indigena.pdf.
- Karttunen, Frances. 1992. *An Analytical Dictionary of Nahuatl*. Norman: University of Oklahoma Press.
- la Cruz Cruz, Victoriano De. 2014. "La Escritura Náhuatl y Los Procesos de Su Revitalización." *Contributions in New World Archaeology* 7 (2010): 199–210.
- Lockhart, James. 1999. *Los Nahuas Después de La Conquista: Historia Social y Cultural de La Población Indígena Del México, Central, Siglos XVI-XVIII*. México, D.F: Fondo de Cultura Económica.
- Olko, Justyna, and John Sullivan. 2015. "Empire, Colony, and Globalization. A Brief History of the Nahuatl Language." *Colloquia Humanistica*, no. 2: 181–216. <https://doi.org/10.11649/ch.2013.009>.
- R Core Team. 2020. "R: A Language and Environment for Statistical Computing." Vienna, Austria. <https://www.r-project.org/>.
- RStudio Team. 2019. "RStudio: Integrated Development Environment for R." Boston, MA. <http://www.rstudio.com/>.
- Trager, Edward H. 2016. "A New System for Writing Nahuatl." 2016. <http://unifont.org/nahuatl>.
- Universidad Nacional Autónoma de México. 2012. "Gran Diccionario Náhuatl [En Línea]." 2012. <http://www.gdn.unam.mx>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemond, et al. 2019. "Welcome to the {tidyverse}." *Journal of Open Source*

1 *Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.

2