Dimensions of definiteness in Ch'ol: A dialectal comparison

Morelia Vázquez Martínez[*] & Carol-Rose Little[†]
[*]*ITSM* & [†]*Cornell University*
Society for the Study of the Indigenous Languages of the Americas

January 3, 2020

Before we begin…

- I am Morelia Vázquez Martínez
- I am from Campanario, Chiapas, Mexico
- My first language was Ch'ol
  - I speak the Tila dialect of Ch'ol (mutually intelligible with other Ch'ol dialects)
- I learned Spanish when I was 12
- Today, I am going give this presentation to you in my language

Introduction

- According to Coon (2010) there are two ways that Ch'ol (Mayan) marks definiteness

(1) Tyi i-k'ux-u-yoñ **jiñi** ts'i'.
     PFV A3-bite-VT-B1 DET dog
     'The dog bit me.'   With a determiner

(2) Tyi k-pejk-ä   **juñ**.
     PFV A1-read-VT **book**
     'I read a/the book.'   Bare nouns may also be definite

Introduction

- We are interested in whether there is a difference in marking definiteness in Ch'ol across different dimensions (e.g., anaphoric vs unique definites)
- Also whether there are differences across dialects
  - For instance, Vázquez Álvarez (2011) notes that the Tila dialect of Ch'ol has an extra determiner, *li*

Our plan today

- ▶ Background
    - ▶ Ch'ol and dialects
    - ▶ Definiteness: uniqueness and anaphoricity
    - ▶ Previous work on definiteness in Ch'ol
- ▶ Methodology
- ▶ Results
- ▶ Conclusions and broader impacts

Background: Ch'ol

- ▶ 222,000 speakers of Ch'ol (also known as Chol or *lakty'añ*)

- ▶ 3 mutually intelligible dialects: Tila, Tumbalá, Sabanilla

    - ▶ Our data comes from the Tila and Tumbalá dialects
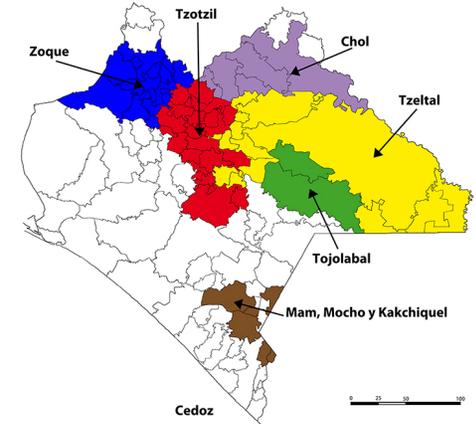
Figure 1: Map of Chiapas and languages spoken in Chiapas

Background: Dimensions of definiteness

- ▶ Languages use different strategies to express definiteness
- ▶ Spanish uses a definite article in (3a) whereas in Shan, bare nouns can be interpreted as definite or indefinite (3b)

(3) a. **El** perro me mordió.   b. háw hǎn **lik**.
     the dog   me bit   1   see **book**
     'The dog bit me.'   'I see a/the book.'
     Spanish   Shan: Moroney (2018)

Background: Dimensions of definiteness

- ▶ Recent work has shown that definiteness can be split up into different dimensions and some languages mark these dimensions morphologically (Arkoh & Matthewson, 2013; Jenks, 2018; Schwarz, 2009, 2013)
    - ▶ We briefly review UNIQUE definiteness and ANAPHORIC definiteness

Background: Uniqueness

- Unique definites are entities which are unique to a situation (the Pope, the bed (in a room with one bed))

- Ebert (1971) shows that Fering (a Germanic language) uses a weak from of the definite article (*a*) to mark unique definite entities like 'sun' or 'king' in (4)

  (4) **A köning** kamm to bischük.
      the$_{weak}$ king came to visit
      'The king came for a visit.' Fering (Ebert (1971) from Schwarz (2013: 541))

Background: Anaphoricity

- In other cases, Fering uses the weak form of the definite article (*di*) to refer to anaphorically to a referent (as in (5))

  (5) Oki hee an hingst keeft. *A / **Di hingst**
      Oki has a horse bought. *DET$_{weak}$ / DET$_{strong}$ **horse**
      haaltet.
      limp

      'Oki bought a horse. The horse limps.' (Ebert (1971) from Schwarz (2013: 538))

Background: Definiteness in Ch'ol

- According to Coon (2010) there are two ways that Ch'ol (Mayan) marks definiteness

  (6) Tyi i-k'ux-u-yoñ **jiñi** ts'i'.
      PFV A3-bite-VT-B1 DET dog
      'The dog bit me.' With a determiner

  (7) Tyi k-pejk-ä **juñ**.
      PFV A1-read-VT **book**
      'I read a/the book.' Bare nouns may also be definite

Background: Definiteness in Ch'ol

- Here, we investigate whether there are differences in the way Ch'ol marks:
  - Anaphoric and unique definites
  - Indefinites

- We also investigate whether there are differences along these dimensions across dialects

Methodology: Tila data

- Speakers: 4 women (24–65), 4 men (27–80)
- Our data comes from El Campanario (2019)
- 6 recordings
  - Baj (145 lines), Kajpe' (63 lines), Lukum (226 lines), Radio (133), Lembal (195), Imojtyolty'añob lakña'ob (178 lines)

Methodology: Tumbalá data

- Speakers: 2 women and one man (40–70) (2016–2019)
- Our data comes from San Miguel (Salto de Agua) and La Ilusión (Tumbalá)
- 8 recordings
  - Xiba (95 lines), Bats' (57 lines), Kumale (54 lines), Ñoxi'aläl (80 lines), Bajlum (20 lines), San Miguel (24 lines), Lakchuchu' (83 lines)
  - ' Bajche' mi ik'atyiñob iyijñam wajali' Gutiérrez (n.d.) (61 lines) de Tumbalá

Methodology: Identifying definiteness

- Our methodology is similar to that of Šimík and Burianová (To Appear) who conducted a corpus study with Czech on bare nouns interpreted as definite
- We recorded each noun as definite or indefinite, indicating context in a note as well
- We only included nouns that were arguments of verbs or in subject position of a predicate

Methodology: Unique referents

- Nouns referring to a unique entity in the global context were coded as unique
- Similarly nouns referring to a unique entity in the immediate situation were also coded as unique as in (8)

(8) There is a salient bird nearby the speakers and this is the first mention of the bird

Aj-kotorro ya'   tyi wejl-i ju'be.
NC-**parrot** there PFV fly-IV DIR:down

'The parrot flew down.'   Imojtyolty'añob lakña'ob (Tila)

Methodology: Anaphoric referents

► We coded nouns as anaphoric if they were previous introduced in the context, as in (9)

(9)   Context: Speaker just said 'there was a snake.'

ya'   me=ku   k'uk'ux jolo   li   lukum=i.
there MIR=AFF EDM   coiled.up DET snake=ENCL

'The snake was just there, all coiled up!'   Lukum (Tila)

Methodology: Data not included

► Possessed nouns

► Nonverbal arguments (like objects of prepositions and possessors)

► Proper names

► Generic referents

► Referents to kinds or classes

► Nouns in object position of a light verb

► Cases where it was too difficult to determine

Methodology: What we coded for

Story	Sentence	Noun	D/I	Det	S/O?	Order	A/U	Pred	
kajpe'	¿ichoch mi ak'äñ li rok ta' che'iñi?	li rok	d	li	o	vo	a	t	previous sentence: mi kch'äme'maj kchiki' yik'oty krok (virginia)
kajpe'	¿y jiñku jiñi kajpe'i säk'bil o ch'ajach mi a wa' tyikisañla?	jiñi kajpe'i	d	jiñi	s	sp	a	t	conversation about coffee
kajpe'	wa'li che mukbä la' juch' jiñ kajpe' mach sumuk bajche jiñi	jiñ kajpe'	d	jiñi	o	vo	a	t	conversation about coffee
lukum	chi'iñ kaña kwuty tyi ñumi tyi maja' li lukum	li lukum	d	li	s	vs	a	i	talking about the snake that is known to both speakers
lukum	mismo jiñäch ta' yubi ajkoralillo	aj coralillo	d	no	s	vs	a	i	established the snake is a coralillo
lukum	ya' meku kuk'ux jolo li lukumi, pero tyoj letse kx'ijal,	li lukumi	d	li	s	vs	a	a	said after añ lukum

Figure 2: Screenshot of data used in study for Tila dialect

Results: Total data

Table 1: Tumbalá total data   Table 2: Tila total data

Indefinite	13	9% of total	Indefinite	26	46% of total
Definite	128	91% of total	Definite	30	54% of total
Total	141		Total	56	

Indefinite: Tumbalá and Tila

- **Tumbalá**
  - Of the total 13 indefinite examples, 100% occurred without a determiner
  - None were ergative subjects – only objects of transitive verbs, subjects of intransitive verbs or in the theme position of the existential predicate *añ*
- **Tila**
  - We found 26 examples of indefinite referents
  - All except one were bare nouns
  - They were all objects of transitive verbs or in the theme position of an existential predicate

Summary: Indefinites

- In both the Tila and Tumbalá dialects, we did not find indefinites occurring with determiners (one exception with the Tila dialect)
- Furthermore, no ergative subjects were recorded as being indefinite
- Both dialects exhibit a similar pattern with respect to marking indefiniteness

Definites in Tumbalá and Tila

Table 3: Definite nouns Tumbalá

Anaphoric	114
Unique	14
Total	128

Table 4: Definite nouns Tila

Anaphoric	11
Unique	19
Total	30

Unique definites

Table 5: Tumbala unique definites

Bare	14
Total	14

Table 6: Tila unique definites

Bare nouns		5
With a determiner	*li*	12
	jiñi	2
		14
Total		19

Anaphoric definites

Table 7: Tumbalá anaphoric definites

Bare nouns			41
With a determiner	aj(iñi)/jiñi/je'	70	
	li/ili	3	
			73
	Total	114	

Table 8: Tila anaphoric definites

Bare nouns			2
With a determiner	li	7	
	jiñi	2	
			9
	Total	11	

Summary: Definites

▶ **Tumbalá** speakers are more likely to:
  ▶ Use bare nouns to refer to unique definites (100% of examples)
  ▶ Use determiners to mark anaphoric definites though bare nouns may also refer anaphorically
    ▶ 64% of anaphoric definites had a determiner
▶ **Tila** speakers are more likely to:
  ▶ Use determiners with definite nouns
    ▶ 74% of unique definites
    ▶ 81% of anaphoric definites were marked with a determiner
  ▶ Use the determiner *li* where this determiner was practically nonexistent with Tumbalá speakers

Implications

▶ While both dialects allow bare nouns to be definite, there is greater tendency in the Tila dialect to use a determiner when referring to definite entities, though bare nouns may also be interpreted as definite
  ▶ This suggests a grammaticalization process of the determiner *li* in Tila into a definite article[1]
▶ In Tumbalá speakers use bare nouns to refer to unique referents as well as anaphoric referents but primarily use determiners to refer back to an anaphoric referent
  ▶ The Tumbalá pattern is similar to languages that mark anaphoric definites but not unique definites (Jenks, 2018)

---

[1]See also appendix for another theoretical implication of the grammaticalization of *li* with respect to subextraction.

Broader impact: Dialectal inclusion

▶ All data from this study will be archived at AILLA and publicly available for speakers and learners of Ch'ol
▶ This is especially important for the Tila dialect as many speakers view this dialect as being inferior to the Tumbalá dialect
▶ Therefore, we hope that with more dialectal studies like this one, speakers will appreciate dialectal differences, rather than judge them

Broader impact: Indigenous languages

- Many indigenous languages are in danger of becoming extinct, Ch'ol included
- For instance, in Campanario (where I am from), many believe that with Spanish they will have more work opportunities and be more accepted in the wider Mexican society

Broader impact: Indigenous languages

- But now there are more opportunities for speakers of indigenous languages than there were when I was growing up
- I hope that other speakers will take advantage of these opportunities so that they can continue preserving their languages

Wokolix la'wälä!

Acknowledgements

Abbreviations

Glosses: 1: first person; 2: second person; 3: third person; A: ergative/possessive; B: absolutive; CL: classifier; DET: determiner; EDM: extreme degree modifier; IV: intransitive verb; NC: noun classifier; PREP: preposition; PFV: perfective; TV: transitive verb

References I

Arkoh, R., & Matthewson, L. (2013). A familiar definite article in Akan. *Lingua, 123*, 1–30.

Bošković, Ž. (2005). On the locality of left branch extraction and the structure of NP. *Studia Linguistica, 59*, 1–45.

Coon, J. (2010). *Complementation in Chol (Mayan): A theory of split ergativity* (Unpublished doctoral dissertation). MIT, Cambridge, MA.

Corver, N. (1992). Left branch extraction. In *Proceedings of NELS 22* (pp. 67–84).

Ebert, K. (1971). Zwei formen des bestimmten Artikels. In D. Wunderlich (Ed.), *Probleme und Fortschritte der Transformationsgrammarik* (pp. 159–174). Hueber.

Gutiérrez, R. d. C. E. (n.d.). *Cuentos cultura chol.* Retrieved from `https://www.unicach.mx/_/ambiental/descargar/Gaceta6/Art5.pdf`

Jenks, P. (2018). Articulated definiteness without articles. *Linguistic Inquiry, 49*(3), 501-536. Retrieved from `https://doi.org/10.1162/ling_a_00280` doi: 10.1162/ling_a_00280

References II

Moroney, M. (2018). *Interpretation of internally-headed relative clauses in shan.* Presented at BLS 44.

Schwarz, F. (2009). *Two types of definites in natural language* (Unpublished doctoral dissertation). University of Massachusetts Amherst.

Schwarz, F. (2013). Two kinds of definites cross-linguistically. *Language and Linguistics Compass, 7*(10), 534–559.

Šimík, R., & Burianová, M. (To Appear). Definiteness of bare NPs as a function of clausal position: A corpus study of Czech. In T. Ionin & J. MacDonald (Eds.), *Proceedings of formal approaches to Slavic linguistics (FASL) 26.*

Uriagereka, J. (1988). *On government* (PhD thesis). University of Connecticut.

Vázquez Álvarez, J. J. (2011). *A grammar of Chol, a Mayan language* (Unpublished doctoral dissertation). University of Texas Austin, Austin, TX.

Theoretical implication: Left branch extraction

▶ The presence or absence of a definite determiner can have consequences for other structures in the language

▶ One implication is that languages that allow so-called "left branch extraction", a phenomenon where a modifier, like an adjective, appears far away from the noun it modifies, do not have definite articles (10)[2]

(10)   *Left Branch Extraction Implication:*
        If a language permits left branch extraction, it lacks articles. (Bošković, 2005; Corver, 1992; Uriagereka, 1988)

---

[2]This is formulated as a one-way implication: it does not mean that all languages without articles allow left branch extraction.

Theoretical implication: Left branch extraction

► The data in (11) where *cha'kojty* 'two' has been extracted from the absolute subject is judged grammatical by Tumbalá speakers, but Tila speakers exhibit varying judgements for sentences like (11)

(11)   Cha'-kojty$_i$ ta'   yajl-i [t$_i$ wakax.]
         two-CL   PFV   cow

         '*Two* cows fell.'   Tumbalá ✓; Tila %

► We suggest that this is because the determiner *li* is becoming a grammaticalized definite article in Tila
   ► Evidence that the grammaticalization process of *li* affects other aspects of the grammar like the possibility for left branch extraction in (11)

Examples: Tumbalá

(12)   The first mention of the man in the story 'Bats"

         A=che'   jiñi ya-tyal   la=k-tyaty=i
         PART=PART DET there-come PL.PART=A1-father=top

         'And then along came a man…'   *Bats'* Tumbala, indefinite

(13)   Tyikäw=ix=ta   **k'iñ** che'.
         hot=already=REA **sun** PART
         'The sun was very hot (that day).'   Tumbalá unique

(14)   After the first mention of 'men' in the story '
         Päkpäk   che'=tyak-ob   **aj**   wiñik-ob
         lying.down PART=PL.INDEF-PL DET man-PL
         'The men were lying down.'   *Xiba* Tumbalá anaphoric

Examples: Tila

(15)   Context: Speaker just said 'there was a snake.'
         ya'   me=ku k'uk'ux jolo   li   lukum=i.
         there MIR=AFF EDM   coiled.up DET snake=ENCL

         'The snake was just there, all coiled up!' Lukum (Tila) anaphoric

(16)   Ta'=bi   och-i   tyi   komisariado li   x-ixik=i.
         PFV=REP enter-TV PREP komisariado DET NC-woman

         'The woman became the community leader.' Radio (Tila) unique

(17)   Aw-om=ki arus.
         A2-want=if rice

         'If you want some rice...'   Lembal (Tila) indefinite