Introduction

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

  (1) Tyi i-k’ux-u-yoñ jĩňi tš’i’.  
     PfV A3-bite-vt-b1 det dog  
     ‘The dog bit me.’

  (2) Tyi k-pejk-á juň.  
     PfV A1-read-vt book  
     ‘I read a/the book.’

  Bare nouns may also be definite

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
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

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. hāw hān lik.
  the dog  me bit
  ‘The dog bit me.’
  Shan: Moroney (2018)

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 form of the definite article (a) to mark unique definite entities like ‘sun’ or ‘king’ in (4)

(4) A könig kam 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. *DETweak / DETstrong 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-yōñ jiñi ts’i’.
PFV A3-bite-vt-b1 DET dog
‘The dog bit me.’
With a determiner

(7) Tyi k-pej-kā 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), Imojtyoly’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), Noxi’aläl (80 lines), Bajlum (20 lines), San Miguel (24 lines), Lakchuchu’ (83 lines)
  - ‘Bajche’ mi ik’atyïñob 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 wej-l ju’be.
  nc-parrot there pfv fly-iv dir:down
  ‘The parrot flew down.’ Imojtyoly’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

<table>
<thead>
<tr>
<th>Story</th>
<th>Sentence</th>
<th>Noun</th>
<th>Def</th>
<th>Subj</th>
<th>Vs</th>
<th>A</th>
<th>Pred</th>
</tr>
</thead>
<tbody>
<tr>
<td>kajpe'</td>
<td>jchich mi=al i k’ok’s xochi?</td>
<td>li nek</td>
<td>d</td>
<td>-</td>
<td>-</td>
<td>a</td>
<td>t</td>
</tr>
<tr>
<td>kajpe'</td>
<td>y jchich li kajpe’ xochi? mi a w li k’ok’s xochi?</td>
<td>jk’al kajpe’</td>
<td>d</td>
<td>jk’al</td>
<td>-</td>
<td>a</td>
<td>t</td>
</tr>
<tr>
<td>kajpe'</td>
<td>wu’ chich mi=al xochi?</td>
<td>jk’al kajpe’</td>
<td>d</td>
<td>jk’al</td>
<td>-</td>
<td>a</td>
<td>t</td>
</tr>
<tr>
<td>lukum</td>
<td>ch’k’a laka tmixti</td>
<td>li lukum</td>
<td>d</td>
<td>li</td>
<td>s vs</td>
<td>a</td>
<td>i</td>
</tr>
<tr>
<td>lukum</td>
<td>maima jchich xochi xuk s ch’k’al</td>
<td>li lukum</td>
<td>d</td>
<td>-</td>
<td>vs</td>
<td>a</td>
<td>i</td>
</tr>
<tr>
<td>lukum</td>
<td>ma’ melu xochi jolo li lukum, pero tyoj lek=ta’aj,</td>
<td>li lukum</td>
<td>d</td>
<td>li</td>
<td>s vs</td>
<td>a</td>
<td>n</td>
</tr>
</tbody>
</table>

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

Results: Total data

<table>
<thead>
<tr>
<th></th>
<th>Tumbalá total data</th>
<th>Tila total data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Indefinite</strong></td>
<td><strong>13</strong></td>
<td><strong>26</strong></td>
</tr>
<tr>
<td><strong>Definite</strong></td>
<td><strong>128</strong></td>
<td><strong>30</strong></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>141</strong></td>
<td><strong>56</strong></td>
</tr>
</tbody>
</table>

Table 1: Tumbalá total data

Table 2: Tila total data

```markdown
<table>
<thead>
<tr>
<th>Indefinite</th>
<th>9% of total</th>
<th>Indefinite</th>
<th>46% of total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definite</td>
<td>91% of total</td>
<td>Definite</td>
<td>54% of total</td>
</tr>
<tr>
<td>Total</td>
<td>100%</td>
<td>Total</td>
<td>100%</td>
</tr>
</tbody>
</table>
```
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>
<thead>
<tr>
<th>Table 3: Definite nouns Tumbalá</th>
<th>Table 4: Definite nouns Tila</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anaphoric</td>
<td>Anaphoric</td>
</tr>
<tr>
<td>Unique</td>
<td>Unique</td>
</tr>
<tr>
<td>Total</td>
<td>Total</td>
</tr>
<tr>
<td>114</td>
<td>11</td>
</tr>
<tr>
<td>14</td>
<td>19</td>
</tr>
<tr>
<td>128</td>
<td>30</td>
</tr>
</tbody>
</table>

Unique definites

<table>
<thead>
<tr>
<th>Table 5: Tumbala unique definites</th>
<th>Table 6: Tila unique definites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bare</td>
<td>Bare</td>
</tr>
<tr>
<td>With a determiner</td>
<td>With a determiner</td>
</tr>
<tr>
<td>li</td>
<td>li</td>
</tr>
<tr>
<td>jini</td>
<td>jini</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>14</td>
<td>19</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bare nouns</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>With a determiner li</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>With a determiner jini</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
</tr>
</tbody>
</table>
Anaphoric definites

Table 7: Tumbalá anaphoric definites

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Bare nouns</td>
<td>41</td>
</tr>
<tr>
<td>With a determiner</td>
<td></td>
</tr>
<tr>
<td>aj(iñi)/jiñi/je'</td>
<td>70</td>
</tr>
<tr>
<td>li/ili</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>73</td>
</tr>
</tbody>
</table>

Total 114

Table 8: Tila anaphoric definites

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Bare nouns</td>
<td>2</td>
</tr>
<tr>
<td>With a determiner</td>
<td></td>
</tr>
<tr>
<td>li</td>
<td>7</td>
</tr>
<tr>
<td>jiñi</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>9</td>
</tr>
</tbody>
</table>

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
- 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)

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

---

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

Acknowledgements

We are very grateful to the speakers who we recorded. In El Campanario, we would like to thank Agusto López, Emiliana López, Virginia Martínez, Elmar Vázquez, Silvia Jiménez, and Abraham Martínez. In San Miguel we would like to thank Evelina Arcos, Rosa López, and Celia Álvaro. In La Ilusión, we would like to thank Lucio Meneses. We would also like to thank audiences at Form and Analysis in Mayan Linguistics 5 for comments and discussion of this work.

This work is supported in part by the National Science Foundation under grant no. BCS-1852744 and an Engaged Cornell graduate student research grant.
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


References II


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)

---
²This 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 absolutive subject is judged grammatical by Tumbalá speakers, but Tila speakers exhibit varying judgements for sentences like (11)

(11) *Cha'-kojty i ta' yajl-i [ ti 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' ji'nı ya-tyal la=k-tyaty=i
    PART=PART DET there-come PL.PART=1-Father=top
    'And then along came a man...'  Bats' Tumbala, indefinite

(13) Hot=already=rea *k'ın che'.
    Tyikäw=ix=ta 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
    laying.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'u'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.
    PPV=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