

What is Definite and what is not in South Eastern Wastek

Jonah Bates
University of Kansas
Lawrence, KS, USA

This paper is a critical look at the glossing of the morpheme *an* as a definite article in a recently published corpus of the South Eastern dialect of Wastek Mayan, where it sometimes appears in contexts without corresponding nominals, without the semantics of definiteness, or in typologically marked word orders. Three structures involving *an* are considered: 1. *an* PREP¹ NOUN word order, 2. post-topic *an*, and 3. subordinating *an*. This paper concludes that *an* is best separated into two different morphemes: its expected use as a definite article (with *an ti* NOUN order explained through D-to-P raising) and a clausal head in the left periphery. Further, evidence shows that both article and complementizer *an* have a clitic hosting allomorph *n-*. Finally, rather than positing a synchronic relationship between DP- and CP-dwelling *an*, they are claimed to have both descended from a Proto-Mayan demonstrative.

Keywords: definite article, left periphery, head movement, Mayan

1. Introduction - Source & Background

The volume under consideration is *South Eastern Huastec Narratives* by Anna Kondic. This almost 200-page book, published in 2016, is the first of its kind for SE Wastek; comprising a compilation of short stories and conversations that provide a rich corpus of natural speech. The South Eastern dialect is one of three dialects of the Mayan language Wastek², also known by the autonyms Teenek or Kaaw Teenek. All three dialects are spoken in the Mexican states of San Luis Potosí and Veracruz. These three varieties are Western or San Luis Potosí Huastec (HVA), Central or Veracruz (HUS), and the South Eastern variety examined here, which is spoken in Chontla, Chinampa, Amatlan, and Tamiahua (HSF).

2. Case One: The Definite Article and PPs

This section investigates uses of *an* that appear to be DP internal, being clearly associated with noun phrases. To begin we will assume that the morpheme in question does actually serve as the definite article for SE Wastek, at least in stereotypical cases like example (1) below.

- (1) *an ataaʃ teenek*
DEF house Wastek
'the Wastek house' (Kondic 2016:6)

¹ The abbreviations I use throughout are mostly following Kondic (2016) with a few additions: A 'absolutive', COM 'completive', DEF 'definite article', DEM 'demonstrative', E 'ergative', FUT 'future', HAB 'habitual', INC 'incompletive', INSTR 'instrumental', IRR 'irrealis', NM 'nominal modifier', PERF 'perfect', PL 'plural', POSS 'possessive', P/PREP 'preposition', R 'realis', REL 'relative', SG 'singular', Sp 'Spanish loan', 1,2,3 'person'

² I will use this term throughout. Also written as Huastec or less commonly Waxtec, Huaxtec, or Huastek.

In favor of this assumption, we find three lines of evidence (from the weakest to the strongest): 1. *an* is the same morpheme used for a definite article in the majority variety of Wastek in San Luis Potosí (Edmonson 1988:489), 2. its distribution corresponds to definite translations in the Spanish and English and contrasts with the indefinite translations accompanying the use of partitive *i* and indefinite (numeral ‘one’) *jun*, and 3. throughout the corpus, the morpheme *an* ‘DEF’ occurs in contexts exhibiting stereotypically ‘definite’ semantics and pragmatics, viz. inclusiveness, or uniqueness, and identifiability, including anaphoricity or previously mentioned-ness (see Lyons 1999).

One way to approach the make-up of the SE Wastek DP is to compare it to like structures in the other Wastekan dialects. As noted in Edmonson (1988:489) for the San Luis Potosí dialect, a noun *must* be marked by either the definite article *an*, a demonstrative, or the indefinite article *juun* (also the cardinal number ‘one’) and/or partitive *i*, except when appearing after the prepositions *ti* ‘GENERAL/LOC’, *ma* ‘GOAL/SOURCE’, and *k’aal* ‘INSTR/COMITATIVE/OBLIQUE’ or if the noun is marked for possession. In these cases, the values of definiteness are either given by pragmatic context (in the PPs) or entailed. These interactions between the definite article and prepositions are the first areas we will examine where the dialects diverge. When SE speakers used prepositions like *k’aal* in the corpus, it did not replace the definite article but surfaced in the expected word order: PREP DEF NOUN.

- (2) *k’aal an k’wachip*
 PREP DEF *niscón*
^{JB}‘with/as the *niscón*’³ (Kondic 2016:16)

As is the case with *k’aal*, the definite article can also co-occur with the preposition *ti* but not in the same word order. Instead, these constructions appear with the word order: DEF PREP NOUN and convey semantics of specificity or definiteness on the noun while bare *ti* NOUN constructions consistently give an indefinite, generic, or ‘kind’ reading. For example, compare (3) with (4) below.

- (3) *an ti che’ey*
 DEF PREP bed
 ‘on this bed’ (Kondic 2016:20)
- (4) *ti toom*
 PREP herb
 ‘some grass’ (Kondic 2016:16)

Pairs like this are consistent with *an* contributing definiteness to the noun in the prepositional phrase. The word order difference between *ti* constructions and those with *k’aal* will be addressed below in **Section 6** after we consider a few other interesting cases of DEF glossing. For now, suffice it to say that there is no reason to doubt that *an* in these cases is serving as the definite article. The apparently non-canonical word order when used with *ti* seems to have more to do with something special about that preposition than with *an* itself.

³ Since the corpus translation given in the book is meant to convey the narratives in coherent ways for the Spanish or English, I have added my own more literal translation in some places. This is always marked with JB. If it differs significantly, I include all three: Kondic’s English, speaker’s Spanish, and my translation.

3. Possessor Marking, Demonstratives and the Definite Article

Another significant difference between the SE dialect and that described by Edmonson pertains to possessor marking. Similar to other Mayan languages (Grinevald & Peake 2012), Wastek references the person and number features of the possessor on the possessed noun with the same person markers used to mark ergative transitive subjects in the verbal domain⁴. SE Wastek shares this in common with the Potosino dialect. In both, these possessor markers surface as proclitics to the possessed noun. In the Potosino dialect, these possessive marker clitics always replace definite marking. However, SE Wastek speakers in Kondic's corpus produce POSS marking in a variety of contexts, including to the right of the demonstrative *axee* 'this' and cliticized to the preposition *t(i)*- mentioned above. Most interestingly, in many possessed contexts in the corpus, the possessor marker is cliticized to the right edge of a morpheme *n-* and this morpheme is glossed DEM by Kondic, as in the examples below. This is by far the most frequent format for possessive marking.

Possessive marking with *n-*

(5) <i>n-u</i> <i>maa</i>	(6) <i>n-i</i> <i>ajaatik</i>	(7) <i>n-in</i> <i>ale'</i>
DEM-E1 mother	DEM-EPL God	DEM-E3 cornfield
'my mother'	'our God'	^{JB} 'his cornfield'
(Kondic 2016:15)	(Kondic 2016:54)	(Kondic 2016:73)

It is probable that *n-* has been glossed by Kondic as DEM because of its resemblance to a purported morpheme *ni*, which Kondic also glosses as DEM. This is perhaps due to an apparent cognateship with Potosino Wastek *nixe'* 'that'. However, evidence for the existence of this demonstrative is lacking. The majority of the cases I have investigated where *ni* is glossed as DEM can be better parsed as *n-* plus another morpheme *-i*; usually EPL. Furthermore, evidence from the distribution and semantics of *n-* appear more consistent with it being a reduced form of *an* 'DEF' rather than a reduced form of a far demonstrative. Specifically, this *n-* morpheme can appear in any instance where simple possession is marked and the semantics do not specify any deixis, let alone far deixis. Most tellingly, this morpheme appears in the corpus in phrases with the near demonstrative *axee'* 'this' as in (8), just like the definite article. In (9), the story had already been mentioned.

(8) <i>axee'</i> <i>n-u</i> <i>tomk-iil</i>	(9) <i>axee'</i> <i>an</i> <i>t'ilap</i>
DEM ^{JB} n-E1 spouse-POSS	DEM DEF story
'my wife' (Kondic 2016:121)	'this story' (Kondic 2016:121)

In many Mayan languages, and more broadly cross-linguistically, possessor marking can naturally co-occur with the definite article, given that possession often entails the same semantic qualities as definiteness. This is the first case where I recommend a change to the glossing seen in Kondic (2016). I suggest that the morpheme *n-* seen before the possessive E-clitics should be glossed DEF instead of DEM. This would unite the free form *an* 'DEF' with the bound form *n-*. This ability for monosyllabic VC morphemes in SE Wastek to re-syllabify the coda consonant as an onset when bound to a vowel-initial morpheme will become important later.

⁴ See Appendix A for an overview of the ergative and absolutive person markers in SE Wastek. Although the ergative/possessor markers are commonly called Set A markers in wider Mayan literature, for easy reference to the corpus at hand, I will follow the corpus convention and the possessive markers will be marked E for ergative.

4. Case Two: Post-topic *an*

The next interesting environment is best described as *an* occurring to the left of the realis/irrealis markers *ti* and *ka*, as in (10).

(10) Temporal Adverb

Taam an t-i um-an-tx-iy an lukuk
 then ^{JB}an R-E.PL add-EP-APPL-TS(COM) DEF mud
 ‘Then we added the mud’ (Kondic 2016:15)

The options for the material to the left of *an* seen in the samples I reviewed were as follows:

- a. Topicalized/focused deictic, manner or temporal adverbs
- b. Topicalized/focused arguments (pronouns/phrases)
- c. *wh*- question words

The adverbs, pronouns, and phrases mentioned above can also at times appear simply to the left of *ti* and *ka* without an intervening *an* or in other clause positions. The semantic difference between those constructions and ones like (10) appears to be that *an* is used when changing topics or for focus between alternatives. In addition, the use of *an* creates what appears to be a cleft structure when focusing, for example, instrumental prepositional phrases headed by *k’aal*. Comparing the example in (11), with *an* topicalization, to that in (12), without *an* topicalization, highlights the construction. In the first example, the instrumental topic NP *an apach* ‘the palm’ is separated from *k’aal* by the emphatic pronoun *jaach* to identify that it is the palm, as opposed to other alternative building materials, that is being used to build a traditional Wastek house.⁵

(11) Topic cleft Instrumental PP

[an apach’] jaach *k’aal* an t-a juj-n-al an ataaaj.
 [DEF palm] PRO *with* ^{JB}an R-A3 put.palm-INSTR-INC DEF house
 ‘The palm is used to thatch the house.’
 ‘La palma es con que se empaja la casa.’
^{JB}‘The palm is that with which the house is thatched.’ (Kondic 2016:7)

(12) Non-topic Instrumental PP

axee’ i ataaaj t’aj-atx [*k’aal* i akamlaap].
 DEM NM house make-PPL [*with* NM pole]
 ‘This house is made with poles.’
 ‘Esta casa está hecha con horcones.’ (Kondic 2016:4)

Since the post-topic/focus morpheme *an* does not provide definiteness to anything in these utterances, I recommend that it be considered simply a homophone of *an* DEF. For the rest of this paper, it will be referred to as *post-topic an* and will gloss it simply as *an*.

⁵ This cleft structure is reflected in the Spanish translation ‘la palma es con que’ made by native Wastek speakers for whom Spanish is a second language. Presumably this Spanish structure was chosen because it preserves more closely the structure present in the original Wastek. I provide a corresponding English translation to emphasize the same.

Before closing this section, it must be noted that there are plenty of clauses that lack overt marking of (ir)realis *ti* or *ka* and in these clauses, *an* never appears. Yet, as may be expected, many of these still sport a similar ‘topic-comment’ structure as those examples seen above. For instance, compare the question-answer minimal pair in (13) and (14) between two speakers in the narrative *Inside a Huastec House of Encinal*.

- (13) tenchee’ an t-a jay-un-th-a-al?
 here ^{JB}an R-E2 eat-EP-CAUS-TS-INC
 ‘And they eat here?’ ‘
 ¿Aquí les da de comer?’
^{JB}‘Is it here that you feed (them)?’ (Kondic 2016:20)

- (14) Ajaa, tenchee’ n-i jay-un-th-a-al?
 yes here REL-E1 eat-EP-CAUS-TS-INC
 yes here ^{JB}n-EPL eat-EP-CAUS-TS-INC
 ‘Yes, here we offer people food.’
 ‘Ajaa, aquí les damos de comer.’
^{JB}‘Yes, it is here that we feed (them).’ (Kondic 2016:20)

The only difference between the two structures above is the presence/absence of the realis marker *ti*. In both, the adverb ‘here’ appears clause initially with special discourse relevance and the lack of realis marking in the answer may be due to economy since it was mentioned already in the question. However, in Kondic’s glossing, the first utterance has *an* as the definite article between the topic and the clause while the second utterance’s *n-* is glossed as a relative in the same position. In a similar approach that leads us to unify the pre-possessive marker *n-* with the definite article *an* in Section 3, the structural evidence lends itself to an analysis where post-topic *an* is also recognized as one morpheme with post-topic *n-* into post-topic *(a)n(-)*.

Additionally, as exemplified in (15) below, the semantics work much better as marking a discourse role of the elements to the left than glossing *n-* as the marker of a relative clause. Namely, the clause that follows *n-* in these instances is not a relative clause describing the NP to its left. Instead, this utterance comes as the answer to the question, “With what did you make it?”, and the purpose of the structure is to point out that the speaker did not, in fact, make it, but it was her mother that did. The predicate structure of the clefted material to the left of *n-* can be seen in the use of the independent 3SG pronoun *ja* in *ja n-u maa*, which more literally translates to ‘she is my mother’.

- (15) ja n-u maa n-in t’aj-a-amal
 s/he ^{JB}DEF-E1 mother ^{JB}an-E3 do-TS-PERF
 ‘My mother made it.’
 ‘Mi mamá lo ha hecho.’
^{JB}‘It is my mother that has made it.’ (Kondic 2016:15)

To close this section, the most important observation to make about these environments is that they all can be interpreted as some sort of cleft structure supporting a topic or focus discourse role for the material to the left. An analysis of this cleft structure and the nature and syntactic placement of post-topic *(a)n* will be returned to and discussed in more detail in Section 6.

5. Case Three: Subordinating *an*

The final environment in the scope of this paper is a morpheme *an* that appears at the left edge of subordinate clauses, which serve as the complements of matrix verbs like ‘to see’ and ‘to feel’. Two such examples can be seen in (16) and (17) below.

- (16) baal baa’ k-a chuu’t-at an t-u atx-im
 so.that NEG IRR-A3 see-PASS.COM an R-A1PL have.bath-AP(INC)
 ‘... so as not to be seen when we bathe.’
 ‘...para que no se vea cuando nos bañamos.’
^{JB} ‘...so that it is not seen that we bathe.’ (Kondic 2016:22)

- (17) ne’ech k-in ach’-a’ an t-in ne’ech ti polk’-an-al
 FUT IRR-E3 feel-TS(COM) DEF R-E3 FUT R get.better-MID-INC⁶
 ‘...he will feel that the painful part of his body is loosening.’
 ‘... va a sentir 6om ova aflojando la parte dolorida.’
^{JB} ‘He is going to feel that he is going to ‘get better’

Again, this *an* surfaces to the left of the (ir)realis marking, except that this time what comes before is not a topicalized constituent but a matrix clause to which the following clause is a complement.

6. Possible Analyses

So far, three interesting environments have been presented for the morpheme *an* glossed as DEF.

1. The marked word order DEF PREP NOUN when *an* is used with the preposition *ti*.
2. ‘Post-topic’ *an*
3. Subordinating *an*

Analysis of the first environment will be saved until the last part of this paper. Before that, I will propose an account that unites ‘post-topic’ *an* and subordinating *an* as one morpheme by locating them syntactically. They both surface to the left of the (ir)realis markers and the head carrying aspect marking. If they are analyzed as part of the clause with these heads, then it follows that *an* must be quite high on the clausal spine. When we consider this distribution in combination with the role *an* plays in connecting topics with a matrix clause and connecting matrix and subordinate clauses, as well as its origin as a demonstrative, SE Wastek clausal *an* is most parsimoniously analyzed as a CP element. This is not a surprising result since many Mayan languages exhibit homophony between complementizers and definite articles with both morphemes descending from a demonstrative in a parent language (see discussion in Kaufman 2015:1008). Intriguingly, there is evidence for just such a demonstrative among the wider Mayan languages. The Eastern Mayan language Q’eqchi, for example, deictically contrasts *ha’an* ‘that’ from *ha’in* ‘this’ (Tzoc 2003:82).

⁶ I am not certain about the translation of the last verb root *polk’* ‘get.better’. It may decompose to *pol* ‘to curl’ + *k’* ‘INTENSIVE’ (Cf. Edmonson 1988:660&671). Whatever the case, it has no impact on the thesis of this section.

Given that clausal *an* is a complementizer, it may be tempting to analyze the DEF PREP NOUN word order as COMP PREP NOUN instead since it is not a typologically uncommon method for PPs and other nominal modifications to be introduced in relative clauses. However, two pieces of evidence speak against this. First, such a proposal would not account for the difference discussed in Section 2 between the *an ti* construction and construction with just *ti*. Specifically, the *an ti* versions have definite semantics while those with just *ti* do not. Secondly, it is not just the definite article that patterns before *ti*. The narratives give us minimal phrasal pairs where the definite article *an* is replaced by the near demonstrative *axee* ‘this’ as in (18) and (19) below, and *naa* ‘that’ (20).

- (18) tenchee’ [pp axee’ ti komunidad], [pp axee’ ti rejoyoon]
 here [pp DEM PREP community(Sp)], [pp DEM PREP area(Sp)]
 ‘Here in this community, in this area...’
 ‘Aquí en esta comunidad, en esta región...’ (Kondic 2016:73)
- (19) juun i k’ij ul-ich juun i alte’ pik’o’ [pp axee’ ti luk’aar].
 one NM time arrive-INC one NM forest dog [pp DEM PREP place(Sp)]
 ‘Once a coyote came to this community.’
 ‘Un día llegó u coyote a esta comunidad.’ (Kondic 2016:154)
- (20) Porke ja’ n-in pith-a-amal [pp naa’ ti podeer].
 because(Sp) s/he an-E3 give-TS-PERF [pp DEM PREP power(Sp)]
 ‘Because he has given this [*sic*] power.’
 ‘Porque el ha dado ese poder.’ (Kondic 2016:55)

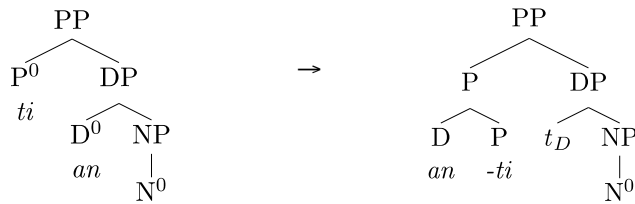
In these examples, the semantic difference becomes clear. As was the case with the definite article, the demonstratives follow the preposition *k’aal* in the expected way as in (21). It is only with *ti* that the order is DEM PREP NOUN.

- (21) u exl-a-amal we’ [pp k’aal axee’ an ya’ulaach-txik]
 E1 learn-TS-PERF little [pp with DEM DEF disease-PL]
 ‘I have learned a little about these illnesses.’ (Kondic 2016:49)

Since the marked word order occurs with multiple nouns and D elements but only the one preposition *ti*, this suggests that something about *ti* is causing the shift. One analysis of the D-before-*ti* word order is that the preposition *ti* itself has phonologically weakened, necessitating head raising of D⁰ to P⁰.⁷

⁷ There is one possible case in the corpus including a demonstrative, a definite article, as well as the preposition *ti*. The phrase *axee’ ta k’itxaaj* is glossed as ‘DEM PREP day’ in Kondic (2016:87). However, the form of the preposition is odd since it should be *ti*. On review of the video of that story, *Celebration of Thipaak, the Maize God*, it can be heard that the speaker originally said *axee’ ti* but goes back and self-corrects. It is possible to parse what was produced on the second time as *axee’ t’ a⁽ⁿ⁾ k’itxaaj*. If this is the case, it strengthens the idea of head raising since the presence (and raising) of the demonstrative leaves the definite article in the canonical word order with respect to *ti*.

(22) D-to-P head raising



Both the definite article and the preposition *ti* can phonologically host clitics in the form of the possessive markers. In cases where the noun is possessed, and we have D-to-P movement, the possessor clitic then cliticizes to the right edge of the entire D+P head, as in (23).

- (23) *an*_i-*t*-in *t*_i k'ima³
 D_i-PREP-E3 *t*_i house
 'from his home' (Kondic 2016:36)

For this analysis to work when there is no overt D⁰, we may posit that covert non-definite D heads still raise to *ti* since *ti* never shows NOUN PREP order with a bare noun but remains pronounced on the left as in (24), reglossed from (4).

- (24) Ø_i-*t*_i *t*_i toom
 D_i-PREP *t*_i herb
 'some grass' (Kondic 2016:16)

The final outline of the PP-internal syntax of SE Wastek has prepositions higher than demonstratives, which in turn c-command the definite article and the noun phrase. The P head *ti* requires head movement of the highest DP-internal head to P⁰.

7. Conclusion

Kondic's pioneering work on SE Wastek in the book *South Eastern Huastec narratives: a trilingual edition* is a valuable corpus for future research analyzing Wastek phrase structure. In this paper analyzing the distribution and syntax of *an* within that corpus, I have presented evidence that there is a clausal *an* in the CP domain and the *an* contributing definiteness in the DP, and that both are descended from one Proto-Mayan demonstrative **an*. Furthermore, I proposed an analysis that the *n*- of *n*-POSS is DEF and the *n*- glossed as REL is 'post-topic' *an*, identifying them as the clitic-hosting allomorphs of those morphemes. Finally, I proposed a possible analysis for the marked word order of D-elements with the preposition *ti* through head movement.

References

- Edmonson, B. (1987). *Huastec Pronouns*. *Anthropological Linguistics*, Vol. 29, No. 4, *Mayan Languages and Their Speakers*. pp. 342-361
- Edmonson, B. (1988). *A Descriptive Grammar of Huastec (Potosino Dialect)*. PhD Thesis. University of Tulane.
- Grineveld, C. & Peake, M. (2012). Ergativity and voice in Mayan languages: a functional-typological approach. In Authier, G. & Haude, K. (eds.) *Ergativity, Valency and Voice*. *Empirical Approaches to Language Typology* 48. pp. 15-49.
- Kaufman, T., & Justeson, J. (2003). *A preliminary Mayan etymological dictionary*. Foundation for the Advancement of Mesoamerican Studies. Retrieved from <http://www.famsi.org/reports/01051/pmed.pdf>
- Kaufman, T. (2015). *Mayan Comparative Studies*. Foundation for the Advancement of Mesoamerican Studies. Retrieved from <http://www.famsi.org/reports/01051/pmed.pdf>
- Kondic, A. (2012). *Narraciones en huasteco de San Francisco: San Francisco Huastec Narratives*. Tlalocan (Universidad Nacional Autónoma de México), 18, pp. 48-61 and 62-76. Retrieved from <http://revistas.unam.mx/index.php/tlalocan/article/viewFile/36642/33199>
- Kondic, A. (2016). *South Eastern Huastec narratives: a trilingual edition*. Norman, OK: University of Oklahoma Press.
- Lyons, C. (1999). *Definiteness*. Cambridge, UK: University of Cambridge Press.
- McQuown, N. (1984). *A sketch of San Luis Potosí Huastec*. In HMAI, Supplement 2: 83-142.
- Tzoc, J. (2003). *Gramática Descriptiva del idioma Maya Q'eqchi*. Cobán Alta Verapaz, GT: Academia de las Lenguas Mayas de Guatemala (ALMG).

Appendix A. Person Marking in SE Wastek

SE Wastek is an ergative/absolute language. Subjects of intransitive predicates are marked with the same Absolute marking as the objects of transitive ones while subjects of transitives are marked Ergative. As is common in Mayan languages, SE Wastek also marks the identity of the possessor through E marking on the possessed noun. But see table **Error! Reference source not found.** and the discussion in Section 3 for an alternative analysis.

	Erg/Poss	Abs
1	u	in
2	a	it
3	in	∅/a
1pl	i	u
2pl	i	ix
3pl	i	ip