Split Ergativity in Ch'orti' Maya:

A Contribution to a Diachronic Typology of Alignment Change

Inaugural-Dissertation

zur Erlangung des Doktorgrades der Philosophischen Fakultät der Universität zu Köln im Fach Linguistik

vorgelegt von

Natalie Korobzow

am 19.02.2024

angenommen von der Philosophischen Fakultät der Universität zu Köln

am 30.04.2024

Leipzig, 14.03.2025

Contents

List of Figures

List of Tables

Abbreviations and orthography

Glossing[1]

1	first person
2	second person
3	third person
A	(index-)set A
ABS	absolutive
ABST	abstractive
ACC	accusative
ACT	active
ADV	adverb
AGT	agentive
AP	antipassive
AF	agent focus
AFF	affirmative
AFV	affective
APPL	applicative
ART	article
B	(index-)set B
C	(index-)set C
CAUS	causative
CL	classifier
CMT	comitative
CNTF	counterfactual conditional
COM	completive
COP	copula
D	(index-)set D (only in Xinkan, not Mayan languages)
DAT	dative
DEF	definite article
DEM	demonstrative

[1] I mostly follow the Leipzig Glossing Rules except for cases where conventions in Mayan linguistics (or in Hull (2016) for Ch'orti' specifically) deviate from the rule.

DET determiner

DIR direct (case)

DIRL directional

DISTR distributive

DS dependent status

ERG ergative

EXCL exclusive (plural)

F feminine

FOC focus

FUT future

GEN genitive

GER gerund

IMP imperative

INC incompletive

INCH inchoative

INCL inclusive (plural)

IND indicative

INF infinitive

INS instrumental

INV inverse

IPF imperfect

IPFV imperfective

IRR irrealis

IS intransitive status

ITER iterative

IV intransitive verb

LOC locative

M masculine

MPAS mediopassive

N neuter

NEG negation

NMLZ nominalizer/nominalization

NOM nominative

NPT non-past

NSG non-singular

NUM numeral

OBJ object

OBL oblique

OPT optative

P patient

PASS passive (PAS in Hull's verb classes)

PFV perfective

PL plural

POS positional verb

POSS possessive

POT potential

PREP preposition

PRF perfect

PROG progressive

PROX proximate, proximative

PRS present

PST past

PTCP participle

Q question

REFL reflexive

REL relative

RPT reportative

RES resultative

RN relational noun

SAP speech act participant ($1^{st}/2^{nd}$ person)

SBJ subject

SBJV subjunctive

SG singular

SS status suffix (unspecified)

TERM	terminative
TOP	topic marker
TS	transitive status
TV	transitive verb
U	undergoer (object)
VN	verbal noun
VOL	volition

Languages

ACN	Acalan Chontal
AKA	Akatek
CHJ	Chuj
CHL	Chol
CHN	Chontal
CHR	Ch’orti’
CHT	Choltí
HUA	Huastec
HGM	Hieroglyphic Maya
ITZ	Itzaj
IXL	Ixil
KAQ	Kaqchikel
KCH	K’iche’
LAC	Lacandón
LL	Lowland Mayan languages: Yucatecan + Cholan (Kaufman 2017: 69)
MAM	Mam
MOP	Mopan
PCH	Proto-Cholan
PGK	Proto-Greater-K’iche’an
PM	Proto-Mayan
PQC	Poqomchi’
PTS	Proto-Tseltalan
QAN	Q’anjob’al

QEQ Q’eqchi’

TSE Tseltal

TZU Tz’utujil

WM Western Maya: Greater Q’anjob’alan + Greater Tseltalan (Kaufman 2003: 42)

WM+ not explained in Kaufman (2003; 2015; 2017), seems to be used in the sense of “Western Maya and some additional language(s)”

YUC Yucatec Maya

General

* reconstructed form

** theoretically expected but unattested form

ε “trecillo” – occasionally used in colonial sources by Spanish clerics for sounds that are not covered by the Spanish orthography and phonological system, e.g, ejectives

AFC Agent Focus Construction

AI Arte I (from the Morán Manuscript)

AII Arte II (from the Morán Manuscript)

AILLA The Archive of the Indigenous Languages of Latin America[2]

ALMG Academia de Lenguas Mayas de Guatemala

INALI Instituto Nacional de Lenguas Indígenas (Mexico)

MM Morán Manuscript

MPP Maldonado-Paxbolon Papers

NP noun phrase

PLFM Proyecto Lingüístico Francisco Marroquín

SSTN Standard Sino-Tibetan Nominalization

TAM tense/aspect/mood

A note on glossing and orthography

Most of the glossing is my own and not taken from the sources except for languages in which I have no expertise (i.e., anything that is not Mayan or Indo-European). I do not use the zero morpheme (ø) for unmarked forms and instead gloss the information that is expressed by the

[2] https://ailla.lib.utexas.edu/ (last accessed: 2025-03-03).

absence of morphological marking on the appropriate stem or suffix, e.g., Ch'orti' *ixin* (go.B3) 'he went' instead of *ixin*-ø (go-B3).

Mayan languages possess both the glottal stop as a phoneme and "glottalized" or ejective consonants, e.g., *p', t', k'*. In the past, there have been different ways to represent these two phenomena. Following the conventions set by ALMG and INALI, most authors today to not differentiate between the two and use <'> to mark both[3] (Aissen, England & Zavala Maldonado 2017a: 9). I will do the same in all cases and adapt other authors' orthography to this standard.

[3] Sometimes authors choose to reserve ' for ejectives and use a straight apostrophe <ʼ> or <ʔ> to represent the glottal stop. The late Terrence Kaufman used <7> to indicate glottal stop in all his publications.

Acknowledgements

This document is the slightly revised version of my dissertation accepted by the Philosophical Faculty of the University of Cologne in April 2024.

The amount of gratitude I feel towards the people that believed in me for this project is difficult to put into words. My Mayan journey began with a class in K'iche' which I had to take because the introduction to Nahuatl that I was actually interested in took place at the same time as another class that was obligatory. Thanks to my teacher Dr. Frauke Sachse, who is amazing at communicating her passion for Mayan in general and the finer things in linguistics like ergativity specifically, I fell in love with this language family in the very first session and have never looked back since (sorry Nahuatl).

My *Doktorvater* Prof. Eugen Hill taught me everything I know about the historical-comparative method and did not doubt for a second that I would be able to apply it to Mayan languages. My *Doktormutter* Prof. Uta Reinöhl has really done too much for me to only deserve the factual title "second supervisor". I enjoyed being part of her institute in Freiburg – at least in spirit. No matter how much I struggled, I could always be sure that both of them would be on my side and support me. I am very grateful for every piece of advice and every kind word of encouragement. The day of my defense is one of the happiest memories that I have and I am grateful to all the people who helped me create it: my defense committee consisting of Eugen, Uta, Prof. Anna Bonifazi, Prof. Marco García García, Prof. Birgit Hellwig and Prof. Nikolaus P. Himmelmann, my friends from Cologne who came to support me and my friends from Würzburg who celebrated with me.

At the beginning of my PhD in 2019/2020, I struggled to narrow down the focus of my dissertation. I am especially grateful to Prof. Nikolai Grube for taking the time to talk to me and steer my interest in a promising direction at that time. It is thanks to him that I went on a wild goose chase in search of the origins of split ergativity. Dr. Sven Gronemeyer and Dr. Christian Prager also took the time to discuss the topic with me at length. Alejandro Garay provided me with contacts, publications that I would otherwise not have had access to and a seemingly endless supply of good mood.

Prof. Daniel Kölligan always made room for Mayan languages in his classes and gave valuable feedback to my ideas. My wonderful friends from Cologne, Bonn and Würzburg took care of me and cheered me on every step of the way. For the surprisingly stressful publication phase, the final part of the journey which challenged me in creative new ways I

had not anticipated, I was able to rely on new-found friends from Leipzig who made me feel at home within days of my arrival. Finally, I am grateful to all the amazing, talented researchers who write inspiring papers.

My family, both close and extended, watched my journey with growing bewilderment but everlasting support. My parents taught me essential life skills to survive the strange world of academia. I owe my stubbornness to my father who never ceased to tell me that I can do anything if I really want to, and my ability to bounce back from failure to my mother for teaching me that there is no crisis a cup of tea cannot solve. My brother has been a constant source of happiness and inspiration to me since the day he was born. It is so funny to discover the many ways in which we are different to then come to the conclusion that we are still kind of the same. My longest and dearest friend listened to me vent and provided me with distraction and baked goods. Astrid, who evades a label because she has been so many different things to me, helped me find my way more than once. I am also grateful to Chichi, Sammy and their lovely family with which I spent several very happy years. My non-academic friends have showered me with excitement and support while also providing reality checks when they were much needed.

Thanks to all of you, I always had someone to turn to for every flavor of existential crisis. However, the most bitter flavors required a very special person, Jakob, who with his mere existence helped me climb out of any hole I managed to dig for myself. You gave me hope, understanding, great food and a healthy dose of escapism. You listened to doom-and-gloom-talk, insanity and conspiracy theories about morphosyntax. How you are still with me is beyond me. I must have a great “personality” or something.

The last beings to thank are the latest additions to the family, Vasilisa and Masha. When questioned on the topic of how she thinks split ergativity arose in the Cholan branch, Vasilisa was unfortunately unavailable for a comment as she was working on her own version of string theory. Masha, on the other hand, who has always been the more polite of the two, pointed out: “mrrr peep!” I couldn’t agree more.

To everyone mentioned above,

but, most importantly, to my parents

1 Introduction

This dissertation provides a new historical explanation for the alignment split and the associated third set of indexes ("set C") in Ch'orti' based on the historical-comparative method and typological considerations. It further aims to enrich our knowledge about alignment typology and pathways of language change, especially at the intersection of grammar and pragmatics.

Ch'orti' (Glottocode chor1273) is an endangered and understudied Mayan language that is of great importance for historical studies of the language family because it is considered the only living direct descendant of the Mayan language recorded in the hieroglyphic texts (Houston, Robertson & Stuart 2000a: 321). However, this relationship has been established based on arguments concerning the verbal morphology of Ch'orti' specifically and the generally accepted theory meant to explain the origin of the third index-set of Ch'orti' (Robertson 1998) is methodologically flawed. Though Ch'orti' alignment is usually described as aspect-based split-ergative and compared to the same type of split in other Mayan languages, I will argue here that the parallels are less pervasive even synchronically, but especially when it comes to the diachrony of the respective splits.

Chapter 2 provides a basic introduction to alignment and Mayan as well as the methods I employ in my study. In 2.1, I discuss alignment terminology (2.1.1) and types of alignment splits (2.1.2) that have been described in the literature. I also address some trends in the study of alignment change (2.1.3). In 2.2, I introduce the Mayan language family providing an overview of the languages (2.2.1), the general typological features of the family (2.2.2) and important language contact zones (2.2.3). Afterwards, I address the state of historical research on Mayan and especially the reconstruction of Proto-Mayan (2.2.4). Since Ch'orti' is the focus of this dissertation, I devote section 2.2.5 to its history from probable beginnings to its successive decline especially during the Spanish conquest (2.2.5.1). I also discuss the name(s) of the language (2.2.5.2) and give an overview of the most important historical and contemporary sources (2.2.5.3). As it will be relevant in my discussion, in 2.2.6, I briefly introduce Colonial Cholti[4] as well, which is considered to be Ch'orti's closest relative and intermediate link between Ch'orti' and Hieroglyphic Maya. Finally, I sketch the methodology employed in this dissertation in 2.3.

[4] For an explanation of my choice of language names and their spelling see chapter 2.2.1.

Chapter 3 gives an overview of the different types of alignments and alignment splits that have been described for Mayan languages. I start with a discussion of ergative-absolutive alignment (3.1) because it is shared by all languages and reconstructed as the basic alignment of Proto-Mayan. Some languages show various split patterns of which I give an overview in section 3.2. Since aspect-based split ergativity is the focus of this thesis, I address it in more depth in section 3.3 with separate discussions of its manifestations in Yucatecan (3.3.1), Western Cholan (3.3.2) and Poqom (3.3.3). A different kind of split that partly interacts with split ergativity in Mayan is the active-stative split. In section 3.4, I discuss which of the two (split ergativity or active-stative/"split intransitivity") is the appropriate term for the alignment split in Mayan languages that is conditioned by aspect (3.4.1) because there has been disagreement about this in the past. In 3.4.2, I present Mayan languages that attest both types of splits.

In chapter 4, I introduce the phonological system (4.1) as well as the basics of the verbal system of Ch'orti' providing an overview of the index-sets, set A, B and C, and their use (4.2) as well as the strategies of marking aspect (and tense) that exist in the language (4.3).

Chapter 5 is a discussion of the existing literature concerning the historical origin of set C. I start with the proposal in Robertson (1998) in 5.1 followed by the one put forward in Wichmann (1999) in 5.2, both of which are very similar, and then discuss a newer theory presented in Quizar (2023) in 5.3. All three proposals presume that set C originated in set A. As I argue, this is precisely the reason why they ultimately suffer from similar problems and leave the same questions unanswered (5.4). As a transition into the following chapter, I recount an illuminating exchange of opinions in 5.5 that concerns the morphophonology of Ch'orti'.

In chapter 6, I investigate how the index-sets of Ch'orti' interact with verbal (and nominal in the case of set A) stems that begin with vowels (6.1) or with *j-* (6.2). A special case is the interaction of the first-person singular of set A and set C with these stems, which I address in a separate section (6.3). In 6.4 and 6.5, I discuss the results with a focus on why it is unlikely that set C developed from set A.

Chapter 7 finally presents a new theory on the origin of set C.[5] This theory differs radically from all proposals so far because it does not derive set C from set A but from set B. I first

[5] After finishing my own work, I came upon a very similar, though not identical, explanation of the origin of set C in the late Robert Wald's (2007: 900–916) dissertation on the verbal complex in Hieroglyphic Maya. The fact that I overlooked this initially is entirely my responsibility. To retain transparency, I have

describe how set B moved from the position after the verb to the preverbal position by way of the Mayan preverbal focus position (7.1) and then address the functional-semantic change of focus to aspect marking in 7.2. Section 7.3 provides origin scenarios for those forms of set C that cannot easily be derived from set B. In 7.4, I discuss whether some further peculiar features of Ch'orti' may be better understood through the lens of my proposed theory and therefore may provide additional support for it. I also re-address some open questions concerning aspect-based split ergativity in Yucatec Maya and discuss the relevance of focus for it, too (7.4.4). All the evidence is brought together in 7.5. In 7.6, I contrast my approach with Wald (2007).

The remaining chapters apply the results of chapter 7 to other Mayan languages that share some of the peculiar features of Ch'orti'. In chapter 8, I discuss whether the proposed explanation of set C can likewise elucidate the similar phenomenon in Choltí and address the implications for the relationship between Choltí and Ch'orti'. Chapter 9 is an attempt to apply the same explanation to Chontal, a Western Cholan language. In chapter 10, I describe a similar formation in Poqomchi', a Greater K'iche'an language that is not closely related with Cholan but is spoken in the vicinity of Cholan languages. Chapter 11 deals with alignment in Hieroglyphic Maya addressing the ongoing debate of whether a split alignment is likely and why it was assumed to be there in the first place. Chapter 12 discusses the results of this thesis from a typological perspective connecting them to the general observations on alignment change discussed in 2.1.3. The conclusion in chapter 13 sums up and evaluates the contribution of this thesis both to Mayan historical linguistics and general linguistic typology and identifies areas where further research is either needed or promising.

In addition to the main text, there are five appendices. Most of these I originally made for myself but decided to include them in the thesis as they consist of additional information that may be especially useful for readers unfamiliar with Mayan languages (or Ch'orti' specifically). Appendix A gives an overview of the various Mayan language names that are in use today or have been employed in the past. It also includes a figure visualizing the classification of Mayan languages with colors. In Appendix B, I provide the word list from the very first source of Ch'orti' (Galindo 1834) in the original orthography because the source is otherwise not easily accessible. I also give the modern counterparts of the lexemes that Galindo recorded. Appendix C offers definitions of key terminology employed in Mayan

not engaged with Wald's approach when revising this dissertation for publication. Instead, I included an additional chapter 7.6 where I compare our approaches.

linguistics that may not be common knowledge among researchers working on other language families. Appendix D is meant to facilitate quick look-up of the function of specific Ch'orti' morphemes that occur in the examples. Appendix E is a grammatical sketch of Ch'orti' that I prepared for myself which I enclose here in case others find it useful. Finally, the thesis ends with the bibliography.

2 Basics

2.1 Morphosyntactic alignment

Alignment has been studied extensively and systematically at least since the second half of the 20th century, though earlier publications also addressed this topic. Right from the beginning, it became apparent that researchers struggled to find appropriate terms in which to describe this phenomenon which may seem so exotic or puzzling from the point of view of a speaker of, e.g., English or German. For instance, Schuchardt (1896) and Uhlenbeck (1916) both described alignment in Caucasian and Algonquian languages respectively as "passive" despite the fact that the former feature ergative-accusative alignment (discussed in chapter 2.1.1) while the latter show active-stative alignment (discussed in 2.1.2), not to mention the fact that "passive" is not an ideal description for either (on which see chapter 2.1.3). Sapir (1917) in his review of Uhlenbeck used terms like "casus energeticus" and "casus inertiæ" to approach the matter. As we will see in the following chapters, the search for appropriate terms is far from complete.

2.1.1 "Ergativity" and other basic types of alignment

The phenomenon of "ergativity" specifically seems to have a special place in alignment research despite being only one of a handful of different alignment types alongside nominative-accusative alignment, albeit a very common one. Central general discussions of alignment in the past have very often focused on ergativity: among others, Comrie (1973; 1978; 2013), Dixon (1979; 1994) as well as more recently McGregor (2009). Gildea (2004) discusses (and rejects) the idea that there are universal cognitive motivations for ergativity. Coon, Massam and Travis (2017) offer a variety of topics and perspectives around ergativity. More general treatments of alignment and the terminology employed to discuss it are Witzlack-Makarevich (2010), Bickel (2011), Haspelmath (2011) and Zúñiga (2018). I found the latter two especially useful for my purposes; Zúñiga (2018) because it focuses on the diachrony of alignment and Haspelmath (2011) because it gives a concise overview of the different uses of common alignment terms as well as their history. Though, of course, a lot more has been written on alignment, this literature is a good starting point. Here, I will not provide an exhaustive discussion of the subject – this section only serves to introduce key concepts that are necessary to understand my discussion of Mayan alignment. Before

approaching the mysterious phenomenon of "ergativity", it is necessary to define what is meant by "alignment".

The terms "alignment" or "grammatical relations" describe which of the core arguments of the verb (S, A and O[6]) are treated in the same way and which are treated differently. Following Dixon's (1994: 6) definition, S refers to the subject or only argument of an intransitive[7] verb while A and O refer to the subject and object of a transitive verb respectively. Of course, the notion of "subject" has been established from the viewpoint of European languages with nominative-accusative alignment and is therefore not necessarily applicable to other kinds of alignment. Leaving the problematic notions of "subject" and "object" aside for a second, let us have a look at an actual example first.

The following examples from German, a language in which case marking is used on noun phrases to distinguish verbal arguments, illustrate the marking of the various arguments in a nominative-accusative system. In (1), nominative case is used to mark the S of *gehen* 'to go', an intransitive verb. The transitive sentence in (2) shows that A, the person that sees the dog, is encoded with nominative case just like the S in (1). O on the other hand, the dog, takes accusative case and is thus marked differently from S or A. This is what constitutes nominative-accusative alignment.

(1) German intransitive verb (my data)

Ich	*gehe.*[8]
1SG.NOM	go.1SG. PRS
'I go.'	

[6] Some scholars prefer P to O. Gilbert Lazard used X, Y and Z for A, P and S respectively (Haspelmath 2011: 541).

[7] Some researchers, e.g., Zúñiga or Bickel (see above), reserve the use of the terms "intransitive" and "transitive" for "complex form-meaning correspondences" (Zúñiga 2018: 2) as described by Hopper and Thompson (1980) in their influential paper on transitivity. For purely syntactic relationships one may then instead use "one-argument clause" and "two-argument clause" as well as "monovalent/bivalent/trivalent predicates" for predicates with one, two or three arguments respectively. However, unless stated otherwise, "intransitive" and "transitive" are used in purely syntactic, not semantic, terms in this dissertation.

[8] Throughout this dissertation, I use color coding to make it easier to see the difference between the different arguments and alignment types.
For nominative-accusative alignment, I use red for S and A and green for O.
For ergative-absolutive alignment, I use red for A and green for S and O.
For tripartite alignment, which will become important later, I use red for A, green for O and blue for S.
The arguments of neutral alignment are all marked in red.
In horizontal alignment, red is used for A and O while green is used for S.

(2) German transitive verb (my data)

Ich	*sehe*	*den*	*Hund.*
1SG.NOM	see.1SG. PRS	DEF.ACC	dog.ACC

'I see the dog.'

For a ditransitive predicate as in (3), A, T and R[9] can be employed to refer to the three arguments (Haspelmath 2011: 558). A is still marked with nominative case. The dog (accusative case) can now be considered the "theme" of the sentence, at least in German (therefore T). The third argument, the recipient of the dog (therefore R), is marked with dative case.

(3) German ditransitive verb (my data)

Ich	*gebe*	*dem*	*Mädchen*	*den*	*Hund.*
1SG.NOM	give.1SG. PRS	DEF.DAT	girl.DAT	DEF.ACC	dog.ACC

'I give the dog to the girl.'

The question of how the arguments of ditransitive clauses align with those of monotransitive and intransitive ones is not the subject of this dissertation and will therefore not be discussed further.[10]

Aside from nominative-accusative, there are four other types of basic alignment. The following examples from K'iche' Maya illustrate "ergativity" or ergative-absolutive alignment. In K'iche', as in all Mayan languages, indexation is used via two bound person form "sets" (or "index-sets" as proposed in Haspelmath 2013: 215), the so-called "set A" and "set B", to cross-reference S, A and O on the verb.[11] When comparing the nominative-accusative alignment in (1) and (2) to the ergative-absolutive alignment in (4) and (5), the color coding makes it immediately obvious that the arguments are grouped differently. In nominative-accusative alignment, S and A are marked alike (in red) and O (green) differently,

---

It is very important to keep in mind that the color coding is only there to visualize alignment. Therefore, the colors are applied according to the syntactic function of the arguments, not according to what elements are cognate or identical. That means that it is possible that the same case be marked in one color in one example, but in another in the following example.

[9] Alternatively, A, T and G (for "goal"), e.g., in Zúñiga (2018).

[10] In this thesis, I do not distinguish between mono- and ditransitive verbs for Mayan and contrast both with intransitive verbs equally to determine alignment because in Mayan, at most two arguments can be cross-referenced on the verb. The distinction is not relevant for my research question, though this remains a worthwhile subject for the future.

[11] In Mayan linguistics, it is customary to call these "pronominal affixes" or "clitics" or even "pronouns". Following the argumentation in Haspelmath (2013: 217–218) I avoid this and instead speak of indexes or bound person forms.

while in ergative-absolutive alignment S and O are marked in the same way (in green), while A receives its own marking (red).

(4) K'iche' intransitive verb (Can Pixabaj 2017: 466)

x-oj-b'iin-ik

COM-B1PL-walk-IS

'We walked.'

(5) K'iche' transitive verb (Can Pixabaj 2017: 466)

x-oj-k-il-o

COM-B1PL-A3PL[12]-see-TS

'They saw us.'

Now let us return to "subjects" and "objects" and contrast German with K'iche'. In languages with nominative-accusative alignment, transitive and intransitive subjects share properties that make the notion "subject" useful for them but this is not the case for other types of alignment (Haspelmath 2011: 537–538). In languages with ergative-absolutive alignment, it is difficult to determine which of the arguments of the transitive verb is the subject or whether the category makes sense at all (on this see, e.g., Sasse 1978; Hopper & Thompson 1980; Haspelmath 2011: 537–539). This is precisely why S, A and O are needed as cover terms in the first place – they make it possible to compare alignment (and other phenomena) across languages. Though they are reminiscent of "subject", "agent" and "object" (or "patient" in the case of P), they are not to be seen simply as abbreviations of these concepts – instead, the intention is to have a neutral way to refer to core arguments in all languages universally without having to deal with the syntactic difficulties such as the biased notion of "subject" (Haspelmath 2011: 538–539).[13]

Dixon's definition of S, A and O given above is problematic for a different reason, as well. He defines A and O via the concept of "transitive clause" for which, in turn, he does not provide a definition (Haspelmath 2011: 543). As Haspelmath points out, it is impossible to define transitivity in a way that the definition will fit all languages – while transitive clauses

[12] As a workaround for the ambiguity that results from abbreviating both the agent and set A with A, the latter will exclusively be used with the word "set" except for glosses. On the other hand, A denoting "agent" will never be used with the word "set" and never in glosses, so the terminology should remain clear. In my view, this is preferable to using an unusual abbreviation for either agent or set A instead of the ones that are firmly established in the literature.

[13] Note, however, that S, A and O have also been viewed critically and argued to present further disadvantages (e.g., Harris 1997; Mithun & Chafe 1999).

with “typical” agents and patients are easily comparable cross-linguistically, this is much more difficult for less clear cases such as experiential clauses – something that is not acknowledged or at least never made explicit in Dixon’s work or that of scholars employing his terminology. Therefore, Haspelmath (2011: 545) suggests to follow the “Comrian approach” in making explicit “what has perhaps been implicit in Dixon’s work: that A and P can be identified readily only for typical transitive clauses (i.e., clauses with physical effect verbs like ‘kill’ or ‘break’), while other two-argument clauses are not considered in this context”.

It seems to me that Mayan languages have relatively few cases of non-prototypical transitive clauses. For Mayan, internally, I define intransitive clauses as those that have one argument marked by one set of bound person forms and transitive clauses as those that have two arguments, both of which are marked via the two sets of bound person forms (not differentiating, as already stated above, between transitive and ditransitive verbs).[14] The absence of marking (the „zero-morpheme“) counts as well if it is part of the paradigm. There is no semantic component in my definition.[15] Whenever I discuss examples which do not necessarily correspond to “prototypical” transitive clauses in other languages, one should keep in mind that cross-linguistic comparability may be limited. However, it is still, I believe, valid to use these non-prototypical examples to argue in favor of a language-internal claim.

The three remaining basic alignment types are tripartite, neutral and horizontal alignment (Zúñiga 2018: 2). In tripartite alignment, S, A and O each receive their own marking. Ch’orti’, the Mayan language that is the focus of this dissertation, has three, not two sets of bound person forms. In the examples below, set C is used to mark S on the intransitive verb in (6), whereas in (7), set A is used to mark A while set B is used to mark O. In other words, the marking of S is neither the same as that of A nor that of O.

(6) Ch’orti’ intransitive incompletive (Hull 2016: 379)

i-wayan

C2SG-sleep

‘you sleep’

[14] As Haspelmath (2011: 549) notes, the direction of the definition is important: While Dixon defines S, A and O via intransitive and transitive clauses, it is more rigorous to do it the other way around. To avoid circularity, one should then not define A and O in terms of participants in a “prototypical transitive situation” but a “prototypical action with two participants”.

[15] Under Bickel’s approach, S, A and O are used as generalized semantic roles, not as syntactic functions, which Haspelmath (2011) cautions against.

(7) Ch'orti' transitive incompletive (Hull 2016: 289)

uwire't

uw-ira-et

A3-see-B2SG

'he sees you'

Neutral alignment is the opposite of tripartite alignment because here, the same marking is employed for S, A and O. This type is widespread, especially when the "marking" is zero, i.e., when a language does not mark S, A and O in any morphological way at all (Zúñiga 2018: 3).[16] English is an example of such a language because it lost almost all of the ancient Germanic case system (except for, e.g., the paradigm of pronouns, which still distinguishes subject *I* from direct object *me*). In examples (8) and (9), *cat* is always marked in the same way, no matter whether it is S, A or O.

(8) English intransitive (my data)

The cat sleeps.

(9) English transitive (my data)

The cat meets another cat.

Finally, horizontal or "double-oblique" alignment is when A and O are marked in the same way, while S is marked differently. This type of alignment is relatively rare (Zúñiga 2018: 3) and even sometimes described as especially "useless" (cf., e.g., Payne 1997: 140).[17] Rošani

[16] In the case of neutral marking, it is still possible to distinguish the arguments, e.g., through specific word order. However, the alignment is then neutral in terms of morphological marking, which is what we are concerned with here.

[17] "The most important distinction to make among A, S, and P is between A and P. This is because A and P are the only arguments (among these three) that are instantiated in the same clause, and it is very important from the point of view of communication to identify which argument is acting upon which other argument. On the other hand, the two other distinctions, S vs. A and S vs. P, are communicationally irrelevant. […] All of the other systems, however, are dysfunctional in one way or another. In particular, system IV [NK: i.e., horizontal alignment] ignores the important distinction between A and P and makes a useless one between S on the one hand and A and P on the other. System V [neutral alignment] ignores all distinctions, while III [tripartite] overdistinguishes them." (Payne 1997: 140).

The view expressed here reminds one of the idea that languages are efficiency-driven constructs that strive to optimize themselves, which gives language change a teleological dimension and, if thought through to the end, leaves little room for variation even though a great deal of variation clearly exists in the languages of the world. This view is decidedly rejected in this thesis: I like the metaphor of languages as "bundles of historical accidents" (Harris 2008: 54) – a quote that goes back to paleontologist Stephen Jay Gould (1983: 101) who described organisms as "bundles of historical accidents, not perfect and predictable machines". Though language change is likely not entirely random, its study shows that it is quite chaotic. You never know what might happen when a certain construction becomes grammaticalized. Chapter 7 of this thesis is

(Iranian) is a language that features this kind of alignment in the past tense. For S in (10), absolutive case is used, whereas both arguments of the transitive sentence in (11), A as well as O, are marked with oblique case (hence the name "double-oblique"[18]).

(10) Rošani past tense intransitive (Payne 1980: 155)[19]

dāδ	*xawrič-ēn-an*	*tar*	*Xaraɣ*	*sat*
these.ABS	boy-PL-3	to	Xorog	go.PST

'These boys went to Xorog.'

(11) Rošani past tense transitive (Payne 1980: 155)

duf	*xawrič-ēn*	*um*	*kitōb*	*x̌ēyt*
these.OBL	boy-PL	that.OBL	book	read.PST

'These boys read that book.'

A final point that sounds obvious but nevertheless needs to be explicitly stated is the fact that alignment can only be determined by comparing intransitive and transitive clauses. It is never enough to provide only one sentence as evidence for some type of alignment. Alignment **only** exists as a relationship between the arguments of the two verb types. In theory, it can be established based on many features of a language. Most commonly, alignment is analyzed based on a language's morphology, the most straightforward case perhaps being case marking though the analysis is very commonly applied to indexation as well.[20] Outside of the realm of morphology, alignment labels can likewise be applied to the syntactic treatment of verbal arguments (Dixon 1994: 16–17). This encompasses, e.g., the ability of S, A and O to be focused, topicalized, questioned or relativized. For instance, a language that allows to focus S and A arguments, but not O arguments, can be described as showing nominative-accusative syntactic behavior. Meanwhile, a language that allows to focus S and O, but not A, shows

---

a great example for that. In speaking about language change, I try to avoid expressions that imply that speakers have much insight into their language or form it according to their needs (which is, I believe, a rare phenomenon that has only become a big factor with school education and standardized variants). At the very least, the "efficiency" of or "motivation" behind a certain linguistic feature has to be judged based on its source or starting point, i.e., in terms of diachrony, as also advocated for in Sonia Cristofaro's work (e.g., 2013).

[18] Horizontal alignment is a more neutral term, however, since it is also applicable to languages that do not use the "oblique" case for argument marking.

[19] I feel it is necessary to point out that this is not the same Payne that considers horizontal alignment "useless".

[20] In fact, case marking and indexation are not as different from each other functionally as often thought (Haspelmath 2013: 216). Therefore, it is unproblematic to apply terms like "nominative-accusative" or "ergative-absolutive" to index marking, as well, even though the terminology was traditionally developed based on case morphology and its labels.

ergative-absolutive alignment in this subdomain of its syntax. There may be a connection between the different types of alignment in a language but they may also be entirely independent of each other. It is often said that determining the alignment of agreement is especially difficult (Siewierska 2003; Bickel et al. 2013: 15) as there are different dimensions of agreement[21] that can be taken into account to determine alignment and these do not necessarily show coherent patterns. For example, agreement could be ergative-absolutive morphologically but nominative-accusative in terms of whether it is prefixed or suffixed. I do not see a problem in this because following what I wrote above, we do not *expect* the patterns to be uniform if they are independent of each other. Of course, it needs to be stated explicitly what part of the language one focuses on when investigating alignment – but this is also true for the study of case itself because there, you could theoretically likewise have some case forms that are prefixes and others that are suffixes and the pattern could likewise diverge from the pure morphological form of the affixes.

For the same reason I try to avoid expressions like "ergativity" or a language "being ergative" because these, I feel, invite the interpretation of ergative-absolutive alignment being a holistic feature of a language with implications for other parts of its grammar – hence the (in my view) futile search for "correlates of ergativity" (e.g., Larsen & Norman 1979; Aissen 2017a). Of course, studying co-occurring features can be interesting. But we should not expect them to necessarily always be there. Along with this come certain expectations such as that ergative-absolutive alignment should be uniform in all languages that show it. But the linguistic features surrounding alignment can be very diverse and ultimately do not matter for the diagnosis of ergative-absolutive alignment. Defined in the simplest and traditional way, ergative-absolutive alignment is simply the same treatment of S and O with a different treatment of A. Nothing more. As a sidenote, the fact that there is no similar obsession with "correlates of accusativity" exposes the debate as the ethnocentric double standard that it is. Instead of a language displaying "ergativity" or "being ergative", I suggest saying, e.g., that a language shows ergative-absolutive alignment (with an optional addition as to what parts of the language this alignment is found in). This may only be a minor difference in phrasing it, but from my observations of people's struggles in conceptualizing ergative-absolutive alignment, I believe it may help to draw the attention to the points that matter. This will become especially relevant in the following chapter where I discuss types of so-called conditioned or "split" alignment with more than one type of alignment in the same language.

[21] See Haspelmath (2013) on the reasons why Mayan-type indexation should not be analyzed as agreement.

2.1.2 Conditioned alignment

One of the most important developments in the study of alignment is the realization that "ergativity" – or any other alignment type – is not a holistic property of a language; instead, by the late 20th century, holistic typology was replaced by the more accurate view that languages can show different alignment types in different subdomains of their grammar (Zúñiga 2018: 2). In fact, many languages attest two or more[22] types of alignment conditioned by different criteria (Zúñiga 2018: 3). This is then considered a "split" alignment. It is important to keep in mind that only different alignments found in the same subdomain can constitute a split alignment. For example, a language where both case marking on noun phrases and agreement clitics with the verb are used need not have the same alignment in these two mechanisms of argument marking (see preceding chapter). The case marking could, for instance, follow ergative-absolutive alignment, while the clitics could use nominative-accusative marking. This is the case in Mizo/Lushei (Tibeto-Burman; see DeLancey (2005: 7) for details). This would commonly not be considered a "split". Otherwise, all languages would likely need to be characterized as split-alignment languages because one could likely demonstrate various kinds of treatment and grouping of arguments in all of them. On the other hand, the use of different alignment types in different tenses of a language would indeed constitute a split since it affects the same domain of grammar.

In general, morphological splits can be conditioned by either the lexical or semantic properties of the verb, lexical or semantic properties of the verbal arguments or by grammatical factors. In an alignment split conditioned by the lexical properties of the verb like actionality[23], S is either marked like A (S_A) or like O (S_O). This subtype is traditionally known as "active-stative"[24] in functional typology, but other labels are in use as well, such as "active-inactive", "split-S", "fluid-S", "semantic alignment" (Donohue & Wichmann 2008), "split-intransitive" (Merlan 1985) and "agentive", though they are not always necessarily

[22] Harris (2008) discusses the origin of three different alignment patterns found in Georgian.

[23] Actionality is sometimes also called "lexical aspect" to distinguish it from grammatical aspect. The difference between the two is that the former refers to inherent properties of the event that is described while the latter characterizes the perspective assumed by the speaker in describing the event.

[24] There is no agreement on whether the active-stative type is a kind of alignment, a kind of split or neither (Zúñiga 2018: 4). Sapir (1917) regarded split-S patterns as an additional basic alignment type, Dixon (1994) understood them as special cases of ergative-absolutive alignment and Bickel and Nichols (2009) as well as Mithun and Chafe (1999) view them as something else entirely (Zúñiga 2018: 4). A brief sketch of the problem as well as a proposed solution is found in Haspelmath (2011: 560–562). For now, I stick to the traditional view and terminology.

defined in the same ways.[25] I follow Dixon (1994) and use "active-stative" to refer to the general phenomenon with "split-S" and "fluid-S" as subgroups whenever differentiation is necessary. A classic example of active-stative alignment is found in Dakota (Siouan). There is one class of intransitive verbs that marks S (13) just as A is marked in (12) while O is unmarked (third person). The alignment is therefore nominative-accusative:

(12) Dakota transitive verb (Merlan 1985: 325)

wa-kté

1SG.SBJ-kill

'I kill him'

(13) Dakota intransitive verb (S_A) (Merlan 1985: 324)

wa-ní

1SG.SBJ-live

'I live, I am alive'

On the other hand, a different class of intransitive verbs rather marks S the same way that O is marked in transitive clauses resulting in ergative-absolutive alignment:

(14) Dakota intransitive verb (S_O) (Merlan 1985: 324)

ma-šíča

1SG.SBJ-bad

'I am bad'

(15) Dakota transitive verb (Merlan 1985: 325)

ma-yá-kte

1SG.OBJ-2SG.SBJ-kill

'you kill me'

The precise semantic difference between the two classes may differ: sometimes the S_A class refers to an "activity" while the S_O class refers to a "state or condition" (hence the name "active-stative") while in other cases S_A is used for activities that the subject has control over (or volition in) whereas S_O is used for activities done without control or volition (Dixon 1994: 71).

[25] I do not have a strong opinion on terminology in this case, though I think "split-intransitivity" is misleading because it suggests that all kinds of splits where something happens with intransitive verbs have something in common, which, in my view, is wrong (see also section 3.4.1).

When it comes to the further distinction of split-S versus fluid-S, in a split-S language, each intransitive verb

> "has fixed class membership – either S_a or S_o – generally on the basis of its prototypical meaning. If one wanted to use a verb which deals with a prototypically non-controlled activity to describe that activity done purposely, then it would still take the S_o marking (and something like an adverb 'purposely' could be added)." (Dixon 1994: 72)

The size of these classes can vary considerably, even among languages that are related (see Merlan 1985 for various configurations). Sometimes the labels "unergative" and "unaccusative" are used for the two classes, which come from the Chomskyan tradition (Zúñiga 2018: 4). I consider them unintuitive and will not use them in this thesis.

However, it is rarely the case that the verbs of an active-stative language fall into two neat groups:

> In practice, some verbs refer to activities that are always likely to be controlled and these are always likely to be marked as S_a; other verbs refer to activities or states that are likely never to be controlled and these are always likely to be shown as S_o. But there will be many verbs in a middle region, referring to activities where there can be control or lack of control, and these may accordingly be marked either as S_a or S_o. (Dixon 1994: 79)

This "middle region" described by Dixon is what the label "fluid-S" may refer to. Contrary to split-S languages, fluid-S languages have a class of verbs that allows the speakers to choose one or the other kind of marking depending on what they want to express. Since the same verbs can be employed with either marking, fluid-S is not strictly speaking a case of a lexically conditioned split but rather a semantically conditioned split. However, there is obviously an intimate relationship between split-S and fluid-S.

A different kind of lexical conditioning is that following an NP hierarchy originally established by Silverstein (1976) and visualized here in Table 1 following Dixon:

Table 1. The Nominal Hierarchy in Dixon (1994: 85).

1st person pronouns	2nd person pronouns	demonstratives 3rd person pronouns	proper nouns	common nouns		
				human	animate	inanimate
← more likely to be in A than in O function						

In languages with this kind of split, one usually observes that NPs farther to the left on the hierarchy co-occur with nominative-accusative alignment whereas elements farther to the right occur with ergative-absolutive alignment (Harris & Campbell 1995: 243). This

phenomenon, often called "NP split", has been described, e.g., for Anatolian (Indo-European) by Garrett (1990). I exemplify it here by Mocho' Maya which has a split alignment described in the following way: when S and A are the highest-ranking participants (first/second person), they follow nominative-accusative alignment, while ergative-absolutive alignment is used for the lowest-ranking participants (Zavala Maldonado 2017: 236–237).[26] S in (16) is marked by so-called "set A" markers and therefore in the same way as A in (17) while different marking[27] is used for O (so-called "set B").

(16) Mocho' intransitive with high-ranking participants (Martin 1998: 203)

eewi	*ii-wa'-i*	*bwéno*
yesterday	A1SG-eat-IS	well

'Yesterday, I ate well.'

(17) Mocho' transitive with high-ranking participants (Martin 1998: 204)

k-ii-patzbe'	*eeqan*
POT-A1SG-deceive.B3	tomorrow

'I will lie (to him) tomorrow.'

On the other hand, if the arguments of the verb are third person, the alignment is ergative-absolutive. In (18), S is marked by set B just like O in (19), while A is marked differently (set A):

(18) Mocho' intransitive with low-ranking participants (Martin 1998: 202)

ook-i	*ch-antiil-oq*	*oso*
enter-IS.B3	A3-wife-IRR	bear

'She became the bear's (so-called) wife.'

(19) Mocho' transitive with low-ranking participants (Martin 1998: 208)

ch-ik'-a	*noonh*	*ch-antiil-e'*
A3-carry.B3-TS	DIRL.going	A3-wife-PL

'They took their wives.'

The difference in construction that leads to these two different alignments affects intransitive verbs: both in (17) and (19), A is marked by set A and O by set B. On the other hand, the intransitive verb receives set A marking when S is an SAP (Speech act participant) as in (16) while it receives set B marking for a third-person S as in (18).

[26] Interestingly, O is always marked by set B, whether it refers to an SAP or not.

[27] Set B does not have an overt morpheme for the third person, i.e., B3 is unmarked in Mayan.

A similar, semantically conditioned alignment split[28] that manifests itself on transitive verbs is the so-called "inverse" or "direct-inverse" alignment defined in the following way in Klaiman (1992: 228–229):

> "Usually, the telltale indication that one is dealing with an inverse system is the fact that verbs of transitive, non-reflexive predications are marked with a special morpheme, or assigned to a special paradigm, in case an argument denoting a speech act participant, an SAP (first or second person) referent, corresponds to a nonsubject core argument or logical role."

This type of alignment is attested in the Americas as well as the Himalayas (Zúñiga 2018: 4). There is great variation in the manifestation of this alignment split cross-linguistically (Klaiman 1992: 229). In the following Plains Cree examples, the constructions in (20) 'I see him.' and (21) 'He sees me.' are almost identical: in both cases, the speech act participant is marked as *ni-* although it is S in (20) and O ("non-subject" in Klaiman's terms) in (21). The fact that in (21), the arguments need to be interpreted the other way around, i.e., inversely, is conveyed through the suffix *-ik(w)* (Klaiman 1992: 229).

(20) Plains Cree SAP subject (Klaiman 1992: 228)

Ni-wāpam-āw.

I-see-SAP.SG

'I see him'

(21) Plains Cree SAP non-subject (Klaiman 1992: 228)

Ni-wāpam-ik.

I-see-INV

'He sees me.'

The conditioning in this split reminds one of the NP hierarchy and in fact, the hierarchy is at the base of the distribution of the argument marking (Klaiman 1992: 236). The difference between the NP split discussed above and exemplified by Mocho' and the inverse marking is that in the case of the latter, the relative position of both verbal arguments decides the marking. In Plains Cree, the marking automatically "flips" when O is above A on the hierarchy, which is indicated by an inverse suffix. On the other hand, in the case of Mocho', S is marked by set A when it refers to an SAP, no matter what. Additionally, O does not receive set A marking even when it refers to an SAP (Zavala Maldonado 2017: 237). It is always

[28] As with active-stative, there is disagreement as to whether this is a basic alignment type or a type of split (Zúñiga 2018: 4). I would agree with Zúñiga that it rather constitutes an alignment split because the alignment is not consistent throughout the language.

marked by set B. Therefore, there is no change that affects the transitive verb and no "switching" of the arguments.

Splits that are not lexically conditioned are governed by mostly grammatical factors. One of the most common cases is a split conditioned by tense or aspect (Harris & Campbell 1995: 242). Many Indo-Iranian languages attest such a split pattern. Katë (Nuristani, spoken in the Hindu Kush region) usually shows nominative-accusative alignment where S and A are marked through direct case and thus in the same way, while the oblique case is used to mark O as illustrated in the optative forms of (22) and (23) that are relics of the old present.

(22) Katë optative intransitive (old present) (Grjunberg 1980: 238)

(vúze)	*nargé-m*
1SG.DIR	run-OPT.1SG

'I shall run'

(23) Katë optative transitive (old present) (Grjunberg 1980: 129)

(vúze)	*suv*	*paṣyé (< paṣí-e)*	*yu-m.*
1SG.DIR	all	mountain.slope-OBL.M	eat-OPT.1SG

'I shall eat the whole mountain'

In the perfective past of Katë (Grjunberg 1980: 218–219; Halfmann 2024: 533–534), on the other hand, the alignment is ergative-absolutive because S in (24) is marked through direct case just like the O of (25), while A is now marked by the oblique case.

(24) Katë perfective past intransitive (Mohammad 1991: 12)

Asád	*á-y-o.*
Asad.DIR	come-PFV-PST.3SG.M

'Asad came'

(25) Katë perfective past transitive (Mohammad 1991: 12)

Yéme	*Asád*	*vř̌-y-o.*
1SG.OBL	Asad.DIR	see-PFV-PST.3SG.M

'I saw Asad'

Dixon (1994: 99), among others, points out that in the case of alignment split by tense or aspect, it is always the past tense or perfective aspect that features ergative-absolutive alignment. Although "always" is too strong a phrasing as there are clear counter-examples (see discussion of Cariban in 2.1.3), this still seems to hold true as a tendency. Non-past tense

or imperfective aspect then features nominative-accusative alignment. Katë neatly follows this distribution.

A split can likewise be conditioned by formal clause features like subordination (Harris & Campbell 1995: 243). For instance, ergative-absolutive alignment may be used in independent clauses but neutral alignment in complement clauses as is the case in the Ixtahuacán dialect of Mam Maya (Zavala Maldonado 2017: 239). The following examples demonstrate that in independent clauses, S in (26) is marked by set B just like O in (27), while A is marked by set A. This is a case of ergative-absolutive alignment.

(26) Ixtahuacán Mam intransitive independent clause (England 1983a: 2)

ma *chin* *b'eet=a*

PROX B1SG walk=1SG

'I walked.'

(27) Ixtahuacán Mam transitive independent clause (England (2013: 119) in Zavala Maldonado (2017: 239))

ma *chin=etz* *t-tzyu-'n=a*

PROX B1SG=DIRL A2PL-grab-DS=2SG/1SG

'You grabbed me.'

On the other hand, in the following complement clauses in (28) and (29), all arguments are marked by set A. This is a case of neutral alignment.

(28) Ixtahuacán Mam intransitive complement clause (England 1983a: 10)

[ela *t=b'aj* *meq't]* *n-xi'* *t-waa-'n* *xjaal*

when A3SG=DIRL be.heated INC-DIRL.B3 A3SG-eat-DS person

'When it was heated, the person ate it.'

(29) Ixtahuacán Mam transitive complement clause (England 1983b: 259)

ok *t-ku'-x* *ky-awa-'n* *xjaal* *kjo'n*

when.POT A3SG-DIRL-DIRL A3PL-plant-DS person cornfield

'When the people plant the cornfield.'

How we should label these splits that are conditioned by tense/aspect, the position on the NP hierarchy or formal clause structure is a matter of debate. Often, e.g., in Mayan linguistics, the label "split ergativity" is used for them. Some researchers, e.g., Zúñiga (2018: 4), on the other hand, give more weight to whether the split phenomena appear due to factors in the transitive

or the intransitive verb. Anything affecting intransitive verbs is subsumed under "split intransitivity" with split-S and fluid-S as subclasses. A tense-/aspect-based split that manifests on intransitive verbs would therefore be considered a case of fluid-S, not split ergativity, which could only apply to splits that manifest in transitive verbs as we have seen in Katë above. Many publications do not adhere to this constraint, however, and I am not convinced that it is useful (see discussion in chapter 3.4). As we will see in chapter 3, the label "split ergativity" is commonly and productively applied to Mayan languages with split phenomena as well, although the change that leads to the split manifests in intransitive, not transitive verbs.

In this section I have discussed conditions under which a language could show more than one basic alignment type. It is important to keep in mind that basic labels like "nominative-accusative", "ergative-absolutive" or "tripartite" are not on the same level as "split ergativity" or "active-stative".[29] The relationship between the two is such that the latter describe the condition under which one or the other type of the former arises.

Before we conclude this chapter, a few words on terminology and typical expressions are in order. It is often stated that something "triggers" or "causes" an alignment split, which implies a causal relationship when what we actually see is mere correlation. It is true that there are certain general tendencies. For instance, as stated above, if there is tense-/aspect-based split ergativity in a language with ergative-absolutive and nominative-accusative alignment, it is usually the perfective aspect (or past tense) that features the ergative-absolutive alignment[30], while imperfective aspect (or present/future tense) features nominative-accusative alignment (Trask 1979: 388). However, it is important to understand that there is nothing in the semantics of the perfective aspect that attracts ergative-absolutive alignment and nothing about the imperfective that dictates it be used with nominative-accusative alignment. This is proven by the discussion of Cariban in 2.1.3. Instead, aspectual meaning is typically expressed with certain constructions which in turn mechanically cause one or the other kind of alignment through their structural properties, often aided by existing syncretisms in the language (see, e.g., Korobzow 2022). For example, in Mayan languages, the syncretism of A marking and possession plays a role in most of the attested alignment splits. In this dissertation, I will employ the words "based" or "conditioned" to speak about the synchronic

[29] As already noted in footnote 24, different views have been expressed on whether this is an alignment or a split in the past.

[30] "The primary correlation here is undoubtedly with the perfective aspect, since the development of a perfective aspect into a past tense is a natural one which is widely attested" (Trask 1979: 395).

distribution of split alignment. However, I never imply a causal relationship between the synchronic conditioning and the respective alignment type. As to the diachronic conditioning, this will be addressed in the following chapter.

2.1.3 Alignment change

Much is known about alignment change and the historical mechanisms behind it, but a lot more is still to be discovered. From the wealth of publications, I can only highlight a selected fraction of contributions. In a paper that is influential to this day, Trask (1979) explored different origins of split ergativity depending on the conditioning factors and other properties of the alignment. Garrett (1990), as already mentioned in the previous section, addressed the origin of the NP split in Anatolian (Indo-European). More recent publications addressing alignment change in general are McGregor (2009; 2017), Coon, Massam and Travis (2017), and Dahl (2021). For some language families or subgroups, alignment has been studied especially extensively, Indo-Iranian perhaps being one of the most prominent examples, e.g., Dahl and Stroński (2016) for Indo-Aryan, though see also Reinöhl's (2018) review of this publication, in Haig (2008) and Jügel (2015) for Iranian in general and in Korn (2009) for Balochi. Payne (1980) studied how ergative features are lost again in Pamir languages. Though a historical study was not his intention, Liljegren (2014) included diachronic remarks in his description of alignment features of the Greater Hindukush. Outside of Indo-European languages, notable recent publications that I found inspiring, among many others, include Casaretto et al. (2020) who argue that the frequently quoted "ergative-to-passive" hypothesis (see below) is untenable for Tima (Niger-Congo, Sudan) and Coghill (2016) who explores how ergativity emerged and was lost again in Aramaic (Semitic). Gildea and Queixalós (2010) host a variety of topics concerning alignment diachrony in languages of Amazonia. Aldridge and Yanagida (2021) studied how nominalizations can lead to alignment change in Austronesian and Japanese. Alice Harris (e.g., 2008) has worked extensively on sources of alignment of languages in the Caucasus.

Dixon (1994) is not only a typological introduction to ergativity and other kinds of alignments and alignment splits, it also includes historical reflections and generalizations. For instance, Dixon discusses the hypothesis that the development of ergative-accusative alignment in a language with otherwise nominative-accusative alignment can happen through reanalysis of a passive and that, conversely, ergative-absolutive alignment can change to nominative-accusative by way of reanalysis of an antipassive (Dixon 1994: 182–206). He discusses other

changes, too, but these receive the most attention. In their summary of the state of the art, Campbell and Harris (1995: 240–281) likewise claim "[t]he most dramatic examples of alignment change take place through reanalysis" (Harris & Campbell 1995: 243). Just like Dixon, they discuss the reanalysis of passive and antipassive. Since these are the most commonly proposed mechanisms, it is worth discussing them here.

For Indo-Iranian tense-/aspect-based split ergativity, a common origin can be proposed even though, synchronically, many other changes have obscured the original construction in the respective languages. Though the reanalysis of a passive form as a source of ergative-absolutive alignment has especially been discussed for Indo-Iranian, we will see that this is not an appropriate description for what happened in these languages. Constructions such as those in examples (24) and (25) from Katë discussed in section 2.1.2 feature an Indo-Iranian deverbal adjective *-ta (< Indo-European *-to) that was integrated into the verbal paradigm (Trask 1979: 397). The original construction is still present in older Indo-Iranian languages like Old Persian, which generally shows nominative-accusative alignment without any kind of split as examples (30) and (31) demonstrate:

(30) Old Persian intransitive (Bīsutūn in Schmitt 2009: 38)

adam-šam	*xšāyaθiya*	*āham*
1SG.NOM-3.PL.GEN/DAT	king.SG.NOM	be.1SG.IPF

'I was their king.'

(31) Old Persian transitive (Bīsutūn in Schmitt 2009: 45)

avaθā	*adam*	*[…]*	*Gaụmātam*	*tayam*	*magum*	*avājanam*
then	1SG.NOM		Gaumata.SG.ACC	REL.M.SG.ACC	mage.SG.ACC	kill.1SG.IPF

'then I killed that Gaumata the mage'

The construction in (32) is generally assumed to have been the origin of ergative-absolutive alignment in Indo-Iranian (Harris & Campbell 1995: 244). It involves the above-mentioned formation ending in *-ta*, which is a resultative participle, and the expression of the semantic subject of the action obliquely in genitive-dative case.

(32) Old Persian (Bīsutūn in Schmitt 2009: 66)

ima,	*taya*	*manā*	*kr̥tam*
DEM.N.SG.NOM	REL.N.SG.NOM	1SG.GEN/DAT	do.PTCP.RES.N.SG.NOM

Bāxtriyā

Bactria.LOC.SG

'This is what was done by me in Bactria.'

This originally "passive" construction is generally claimed to have been reanalyzed as active eventually (Harris & Campbell 1995: 244–245):

"Passive": 'It was done by me.'

↓

Active: 'I did (it).'

In this reanalysis, the obliquely expressed semantic agent was understood as fulfilling the syntactic agent role while the original syntactic agent (in nominative case) was interpreted as expressing the direct object.

However, it is important to understand that we are not in fact dealing with a "passive" participle in the case of Indo-Iranian and therefore, these languages do not actually represent evidence for the "passive-to-ergative" pathway.[31] The participle involved is a **resultative** one best described as having a stative value so that this type of split ergativity would arise from "the incorporation into the inflectional paradigm of a nominalized deverbal form with stative force" (Trask 1979: 397). The fact that this participle is oriented towards marking S and O[32] but not A is the reason why ergative-absolutive alignment arose in the past tense or perfective aspect (depending on the individual language) when this construction was grammaticalized. In the case of Katë, the original construction of examples (24) and (25) can then be literally understood as 'Asad is come' and 'To me, Asad is seen'.[33]

A key insight here is the fact that the construction should be understood as **nominal** in origin because the *ta-form is a verbal adjective and therefore a nominal, not a verbal form (Reinöhl n.d.). Not only does it not derive from a passive – it does not derive from a verbal

[31] Trask (1979: 391) instead argues that split ergativity that is conditioned by the *NP hierarchy* arises from a "passive made obligatory".

[32] The *ta-participle is a so-called "absolutive-oriented" participle and can be used to reference S and O but not A (for details on participle orientation see Shagal (2019) or Korobzow (2022) as well as chapter 12 of this dissertation).

[33] Katë *-y-* is the regular result of Proto-Indo-Iranian *-i-ta- (Jakob Halfmann, p.c.).

construction at all. This reanalysis of the *ta-construction is therefore a striking parallel to cases where alignment change is argued to have happened because of the reanalysis of nominalizations. One language family that attests this development are the Cariban languages (Gildea 1992). Proto-Cariban alignment is reconstructed as nominative-accusative in terms of word order and general agreement; however, in nominalizations, S and O are treated alike instead, as illustrated in the following examples (Harris & Campbell 1995: 246). Both S in (33) and O in (34) are marked with the agreement prefix *y-* when nominalized while A receives dative marking(Harris & Campbell 1995: 247).[34]

(33) Tiriyo intransitive nominalization (Gildea 1992: 129)

y-itö-ø *se-pa* *wai*

1-go-NMLZ want-NEG 1.be

'I don't want to go.' (lit. 'I am not wanting my going')

(34) Tiriyo transitive nominalization (Gildea 1992: 130)

[...] mahak-uya *y-eri-ø-ke*

mosquito-DAT 1-bite-NMLZ-INS

'… (because of) my being bitten by mosquitos.' (lit. '… with the biting of me by mosquitos.')

In some Cariban languages, "a biclausal structure consisting of a nominalized complement of a finite copula in the matrix clause was reanalyzed as a monoclausal structure" (Harris & Campbell 1995: 247). This resulted in forms like the following, which now show grammaticalized ergative-absolutive alignment in past tense. S in (35) is marked in the same way as O in (36), namely by the prefix *i-*, whereas A receives an additional ergative suffix cognate to the dative one above.

(35) Pémong intransitive past (Gildea 1992: 189)

i-tö-'pä

3-go-PST

'He went.'

(36) Pémong transitive past (Gildea 1992: 189)

i-kä'pa-'pä-i-ya

3-smear-PST-3-ERG

'He smeared him.'

[34] Note that this is not the case when S and O are independent nouns (Harris & Campbell 1995: 247).

Another interesting result of Gildea (1992) is the fact that the neat distribution mentioned on page 18 is, like many alleged "universals", a tendency at best. Though this is not seen in examples (35) and (36), in his survey, Gildea (1992: 255–256) found that it is not at all the case that ergative-absolutive alignment is always found in past tense or perfective aspect if a split is present. Likewise, he provides further evidence that ergative-absolutive alignment does not arise from passive structures in all cases.

Since we discussed the role of reanalysis of passives in the genesis of ergative-absolutive alignment, it is instructive to look at the opposite case as well, namely the reanalysis of antipassives via object demotion as a source for nominative-accusative alignment. This is discussed in Harris and Campbell (1995: 245–246) for Kartvelian languages. Modern Georgian ergative-absolutive alignment is demonstrated in (37) and (38). S and O are marked by absolutive case while A is marked by ergative case.

(37) Modern Georgian intransitive (Harris & Campbell 1995: 245)[35]

tamar	*didi*	*mepe*	*iq'o*
Tamar.ABS	great	monarch.ABS	was

'Tamar was a great monarch.'

(38) Modern Georgian transitive (Harris & Campbell 1995: 245)

deda-m	*p'erang-i*	*garecxa*
mother-ERG	shirt-ABS	washed

'Mother washed the shirt.'

Like many languages with ergative-absolutive alignment, Georgian also has an antipassive, demonstrated in (39). In this form, the O of an originally transitive verb is demoted to oblique position marked by dative. The verb becomes intransitive and receives S marking (absolutive case) (Harris & Campbell 1995: 245).

(39) Modern Georgian object demotion (Harris & Campbell 1995: 245)

deda	*p'erang-s*	*recxavs*
mother.ABS	shirt-DAT	washes

'Mother is washing the shirt.'

The authors claim that this antipassive construction was associated with the imperfective aspect. Taken together with the intransitive imperfective in (40), which is simply formed by

[35] For examples (37) and (38), they assign case labels based on the history of the language, not based on modern grammatical description.

changing the verb form, not the case, the whole construction was reanalyzed as a regular imperfective with nominative-accusative alignment where S and A came to be marked by absolutive case while O was marked by dative (Harris & Campbell 1995: 245–246):

(40) Modern Georgian intransitive imperfective (Harris & Campbell 1995: 246)

tamar	*didi*	*mepe*	*aris*
Tamar.ABS	great	monarch.ABS	is

'Tamar is a great monarch.'

Further mechanisms identified in Harris and Campbell (1995: 248–251) as sources for alignment change such as the reanalysis of instrumentals[36] or the reanalysis of verbs with incorporated nouns (e.g., in Kartvelian) will not be discussed here. As a sidenote, I would like to caution against viewing "borrowing" per se as a mechanism of alignment change, as the authors do (Harris & Campbell 1995: 251), because, strictly speaking, constructions may be borrowed which then may lead to alignment change, but it is not a new alignment per se that is borrowed.

Finally, it is worth looking beyond grammar when searching for the origins of alignment change. DuBois (1987) suggested that ergative-absolutive alignment arises through certain discourse strategies because he found A to be dispreferred for the expression of new agents leaving new information to be conveyed mostly via S or O ("preferred argument structure"). Tied to this is the observation that A is rarely expressed lexically as a full NP – most clauses have only one lexical NP and it is almost exclusively S or O. This suggestion was established based on, among others, Mayan languages, especially Sakapultek (K'iche'an). Harris and Campbell (1995: 251–255) are rather critical of this idea. Using the proposed pathway of "ergative-to-passive", they argue that the reanalysis of the obliquely expressed argument (syntactic O, semantic A) would not have happened because this would include the genesis of a lexically expressed A which should be dispreferred following DuBois's account. They conclude that the causality may go the over way around, i.e., that ergative-absolutive alignment may promote the discourse strategies described by DuBois. In fact, recent work suggests that DuBois's results cannot be replicated for all languages and therefore cannot be taken to apply to all languages (see, e.g., Everett 2009; Haig & Schnell 2016; McGregor 2009: 486).

[36] See, e.g., Garrett (1990) for Anatolian and Gorokan (Eastern Highlands of Papua New Guinea).

DuBois (2017: 46) in turn reaffirmed the validity of his theory claiming that the critique in Harris and Campbell (1995) "while bringing no empirical research to bear on the question at hand, translate originally statistical observations into the language of categorical statements" and by doing this paint "a portrait of the discourse basis of ergativity […] that is almost unrecognizable to someone familiar with the theory". Further studies of the interaction of discourse and grammar may bring clarity to this question in the future. Concerning the data that contradict DuBois as mentioned above, I suggest to view the discourse correlation described by him as only one possible mechanism of how alignment can result from discourse: just as is the case in grammaticalization where, e.g., futures may develop from a variety of sources, we would not assume that one explanation fits all cases and that DuBois's observations are the only force governing alignment change. However, it is clear that, if we want to add DuBois's observations to our growing catalog of "information structure to grammar" pathways, we need to go beyond synchronic correlation and make explicit what happened diachronically.

After having presented the basic concepts around alignment and alignment change, I will introduce the Mayan language family in the following chapter.

2.2 Mayan languages

2.2.1 Languages

The Mayan language family consists of about thirty members[37] presented here and in Figure 8 on page 302. The overview is based on Campbell (2017: 44) with some minor changes (e.g., I included Hieroglyphic Maya as a separate language).

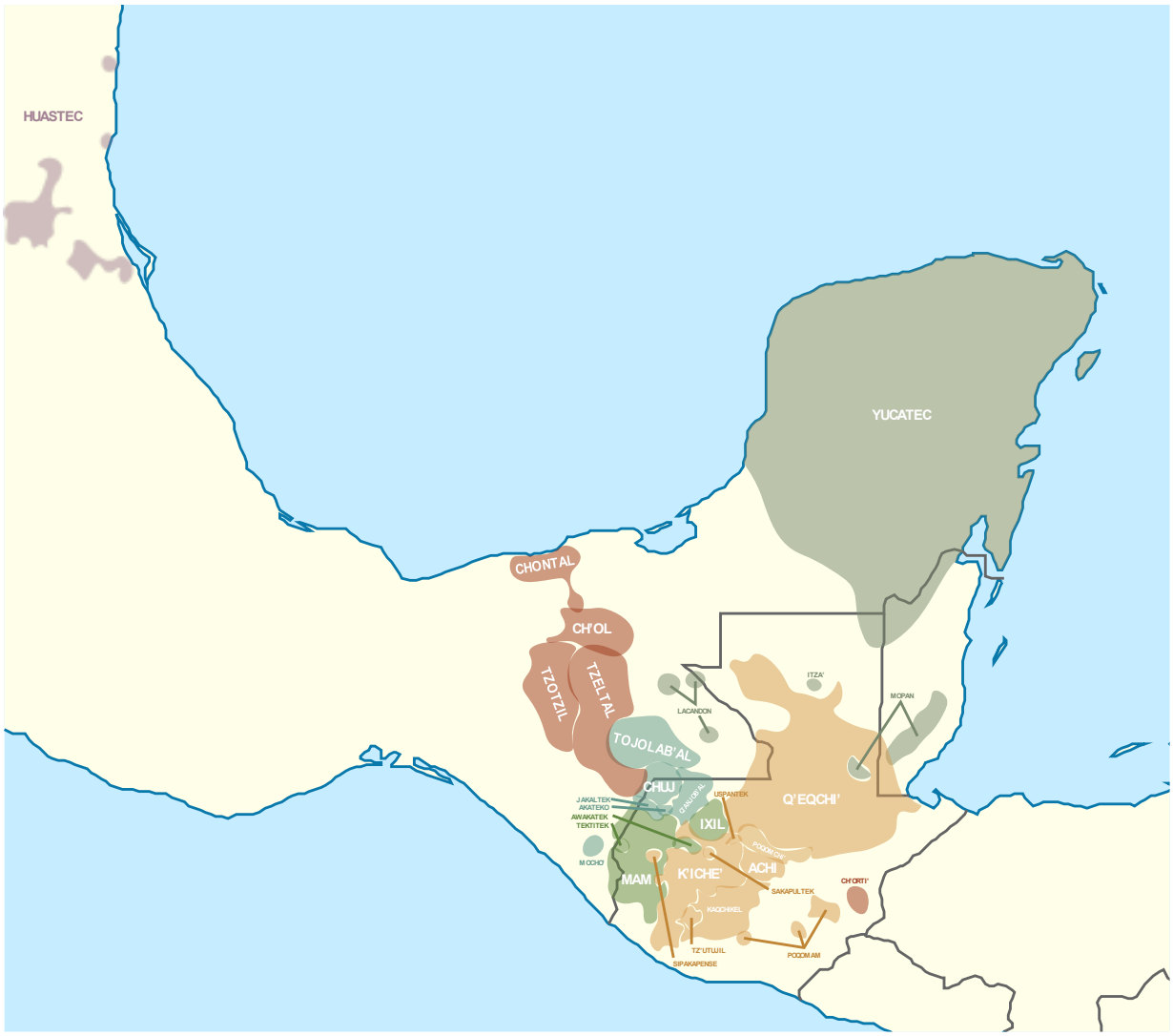
[37] The exact count depends on one's definition of "language" and "dialect".

- **Proto-Mayan**
  - **Huastecan**
    - **Huastec**
    - **†Chicomuseltec**
  - **Core/Central Mayan**
    - **Yucatecan**
      - **Yucatec Maya**
      - **Lacandón**
      - **Itzaj**
      - **Mopan**
    - **Western Mayan**
      - **Cholan-Tseltalan**
        - **Cholan**
          - **Chontal**
          - **Chol**
          - **(†)Hieroglyphic Maya**
          - **†Choltí**
          - **Ch'orti'**
        - **Tseltalan**
          - **Tseltal**
          - **Tsotsil**
      - **Greater Q'anjob'alan**
        - **Q'anjob'alan**
          - **Q'anjob'al**
          - **Akatek**
          - **Popti'**
          - **Mocho'**
        - **Chuj-Tojolab'al**
          - **Chuj**
          - **Tojolab'al**
    - **Eastern Mayan**
      - **Greater K'iche'an**
        - **Q'eqchi'**
        - **Uspantek**
        - **Poqom**
          - **Poqomchi'**
          - **Poqomam**
        - **K'iche'an Proper**
          - **K'iche'**
          - **Kaqchikel**
          - **Tz'utujil**
          - **Sakapultek**
          - **Sipakapense**
      - **Mamean**
        - **Mam**
        - **Teko**
        - **Awakateko**
        - **Ixil**

This subgrouping is more or less uncontroversial, although for Tojolab'al, it remains unclear whether it belongs to Greater Q'anjob'alan or to Cholan-Tseltalan (Campbell 2017: 45). Law (2017a) somewhat unsatisfactorily proposes to treat it as a "mixed language".

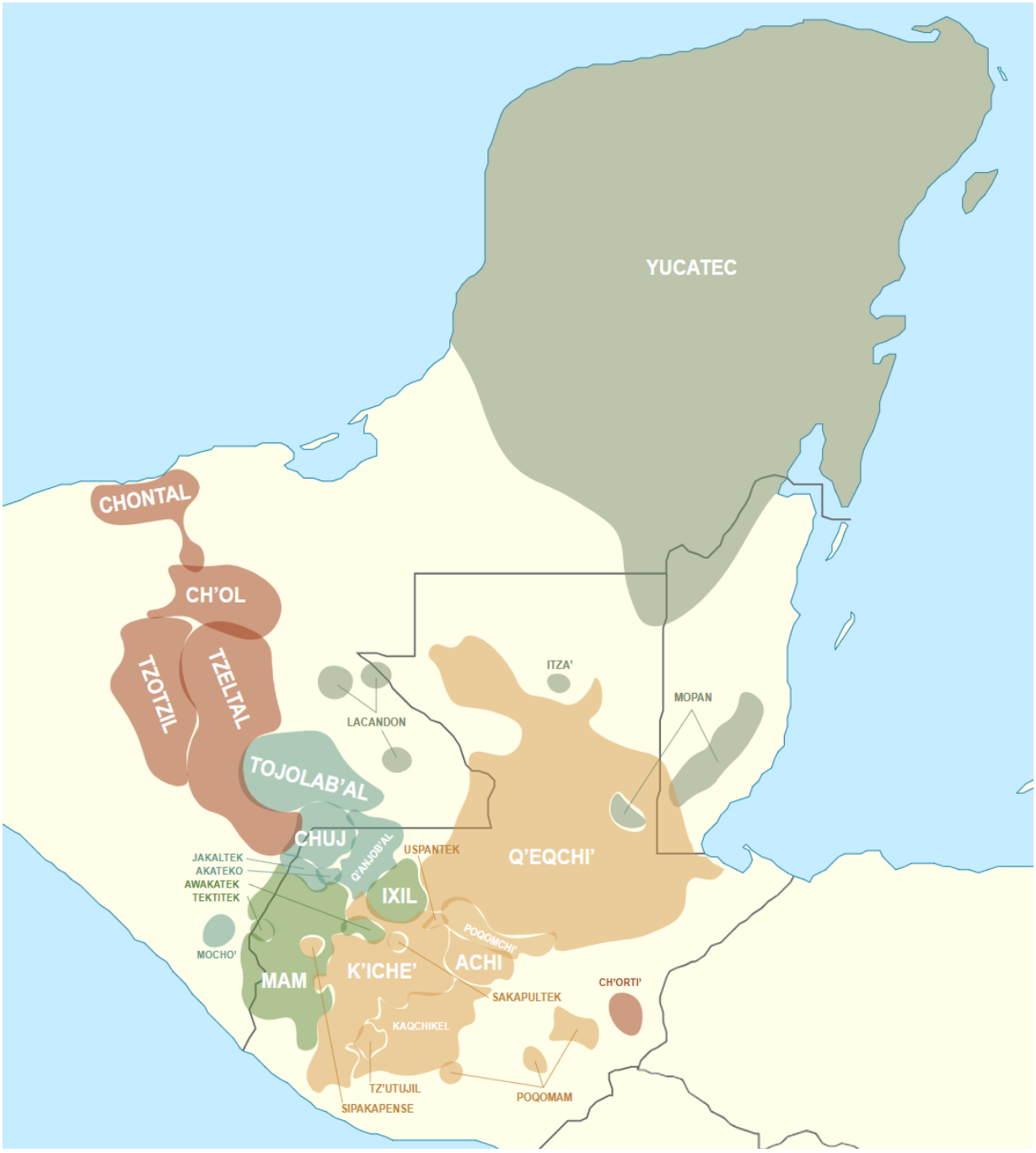
Mayan languages are spoken in Guatemala, Mexico and Belize as displayed in Figure 1 and Figure 2. Huastec is an outlier spoken further north close to the coast of the Gulf of Mexico. Historically, the Maya territory extended further at least into what is today El Salvador and Honduras (see chapter 2.2.5).

Figure 1. Present geographic distribution of Mayan languages in Mexico and Central America. Colors indicate internal divisions in the family.[38]



[38] https://en.m.wikipedia.org/wiki/File:Mayan_languages_map.svg (last accessed: 2024-02-07). © user Noahedits, 2020, based on Law (2014: 23), Creative Commons Attribution-Share Alike 4.0 International License (https://creativecommons.org/licenses/by-sa/4.0/deed.en). No changes made.

Figure 2. Present geographic distribution of Mayan languages in Mexico and Central America. Colors indicate internal divisions in the family.[39]



Speaker numbers vary between under thirty for heavily endangered languages like Itzaj and about a million for K'iche' (Aissen, England & Zavala Maldonado 2017a: 1). A synthesis of the history of linguistic research is found in Aissen, England and Zavala Maldonado (2017a).

[39] https://en.m.wikipedia.org/wiki/File:Mayan_languages_map.svg (last accessed: 2024-02-07). © user Noahedits, 2020, based on Law (2014: 23), Creative Commons Attribution-Share Alike 4.0 International License (https://creativecommons.org/licenses/by-sa/4.0/deed.en). Changes: image cropped to zoom in on Core Mayan (without Huastec).

Table 34 in Appendix A provides an overview of the various language names that are or have been in use over time. Some of the languages have only become known to linguists recently (after 1965), namely Akatek, Sakapultek, Sipakapense and Tektitek (see Kaufman 1969; 1975). Achi and Chalchiteko are viewed as separate languages by some, but commonly, Achi is considered to be a variety of K'iche', while Chalchiteko is taken to be a variety of Awakatek. Though the difference of language and dialect is a fluid one, at least the respective speakers view themselves as separate ethnic groups from K'iche' or Awakatek.[40] Tuzantec and Mocho' are also either separate languages or dialects. Both are critically endangered (on all this cf. Campbell 2017: 45).

At least two extinct languages are known. Chicomuseltec was documented by Sapper (1897; 1912) and went extinct in the 1970s/80s (Campbell & Canger 1978). Choltí is only documented in a manuscript from the 17th century and went extinct shortly after, though some scholars have argued that it survives in the form of modern Ch'orti' (e.g., Houston, Robertson & Stuart 2000a; Robertson 1998; Robertson, Law & Haertel 2010).[41] In that case, it would not have gone extinct. The same can be said of Hieroglyphic Maya – if Ch'orti' is a surviving descendant of the language recorded in those texts, as is widely assumed, then Hieroglyphic Maya has not, technically speaking, gone extinct. Some further languages are known from colonial sources, e.g., Coxoh, which could either be another extinct Mayan language or a dialect of Tseltal (Campbell 2017: 45). Thanks to the hieroglyphic corpus, the Mayan languages are in a unique position among the languages of the Americas as they feature an immense historical corpus that reaches at least as far back as 300 BCE (Law & Stuart 2017: 128). Obviously, this makes Mayan especially suitable for diachronic studies.

Since this dissertation is concerned with Cholan languages, a note on language names and my use of these is in order. I employ the name "Choltí" instead of "Ch'olti'" and "Hieroglyphic Maya" instead of "Classic Ch'olti'" because the latter terms respectively are based on the proposed mother-daughter-granddaughter relationship of Hieroglyphic Maya, Choltí and Ch'orti'. For the latter, the version "Ch'orti'" instead of "Chorti'" is apparently a recent phenomenon and does perhaps not correspond to the phonetic reality (see chapter 2.2.5.2, where this is addressed in more detail). Therefore, even if the relationship between Hieroglyphic Maya, Choltí and Ch'orti' were uncontroversial it would not necessarily make sense to apply the modern spelling to older stages of the language. But the relationship is *not*

[40] Of course, ethnic and linguistic identities need not be identical.

[41] The same authors argue that the language recorded in the Mayan hieroglyphic texts is the exclusive parent of both Ch'orti' and Choltí.

uncontroversial for two reasons: First, though it seems to be generally accepted that Hieroglyphic Maya indeed records a language that specifically belongs to the Cholan branch of Mayan (see, e.g., Schele 1982: 8; Campbell 1984; Houston, Robertson & Stuart 2000a; Grube 2000; Wichmann 2002), it is still a matter of debate what stage of Cholan the texts represent.[42] Second, in this dissertation, I will address some methodological shortcomings surrounding the theories behind the proposed relationship of Ch'orti', Cholti and Hieroglyphic Maya.[43]

For the time being, I will use the orthography of the colonial source for Cholti because it is the only information that we have on the language. The alternative name Cholti' (see Appendix A) would also be a valid option. The name "Hieroglyphic Maya" may lead to some confusion because it has been suggested that the script also records languages from other branches (Lacadena 2013: 5). I am aware of this but will employ the term with the basic assumptions that most of them are written in the same form of Cholan – as many other publications do, e.g., Lacadena and Davletshin (2013). The appropriate name of the language can be reevaluated once its exact identity stands on more solid ground.

For further information on Mayan languages, the first stop is certainly the monumental overview in Aissen, England and Zavala Maldonado (2017b) and the bibliography therein. Additional literature will be mentioned in the respective sections.

2.2.2 Typological profile

In my view there is only limited use in broad typological profiles as they gloss over all differences and are often essentialist in nature. Since some people do find them helpful, I nevertheless reiterate some of the points that are commonly mentioned in descriptions of Mayan languages. Still, it should be kept in mind that not all Mayan languages feature these; also, the individual languages likely have many interesting traits that have not been

[42] For completely opposite opinions see Mora-Marín, Hopkins and Josserand (2009) and Law, Robertson and Houston (2009) with Wichmann (2006) perhaps presenting a balanced view between the two. Mora-Marín, Hopkins and Josserand (2009) claim that the texts record the ancestor of all Cholan languages, while Law, Robertson and Houston (2009) argue it is only the ancestor of Eastern Cholan (Cholti and Ch'orti), not Western Cholan (Chol and Chontal). Wichmann (2006) assumes that the separation of the Cholan branch is captured in the hieroglyphic texts.

[43] Recently, it has also been pointed out by Quizar (2020: 238) that the exact affiliation of Ch'orti' and Cholti is still up for debate, since, e.g., the conservative antipassive system of Ch'orti' could not have arisen from the one we see in Cholti.

extensively studied or even mentioned so far. Key phonological traits include (Aissen, England & Zavala Maldonado 2017a: 5):[44]

- a contrast between a series of voiceless stops and a series of glottalized stops, where the bilabial stop is an implosive while the others are ejectives[45]
- a series of voiceless affricates and ejective affricates
- voiceless fricatives
- voiced nasals, liquids and glides
- five canonical vowels and phonemic vowel length in most languages.

Mayan languages are head-initial, i.e., they use prepositions instead of postpositions, possessed nouns precede the possessors and complementizers precede the clauses they introduce. They use ergative-absolutive alignment and mark aspect rather than tense morphologically on verbs. Of course, this does not mean that it is impossible to communicate specific temporal information in Mayan: it can be expressed through temporal adverbs or understood from the context. Aspect markers usually precede the verb and the prefixed person markers.

Case is not a salient category in Mayan. The core arguments are cross-referenced on the verb via two (in the case of Ch'orti' three) sets of bound person markers. Set A always precedes the verb stem, while set B markers may precede or follow the stem depending on the language in question and the syntactic context. Since they are indexed on the head, independent pronouns generally do not occur unless topicalized or focused, which is common for head-initial languages. Clauses may thus have only one lexical argument or none at all.

Mayan languages distinguish nouns and verbs, although there is some overlap between the two categories. There are some morphemes that only verbs can combine with, e.g., TAM markers, indicators of transitivity and syntactic dependency, voice-changing morphemes and grammaticalized motion auxiliaries (Polian 2017: 209). Person marking is not exclusive to verbs, since set A (A of transitive verbs) is also used to mark possession on nouns and set B (S and O) also forms non-verbal predicates (Polian 2017: 209) in the so-called "stative

[44] If not stated otherwise, this section provides information based on the short introduction by Aissen, England and Zavala Maldonado (2017a).

[45] For more in-depth phonetic analyses of Mayan ejectives and implosives see, e.g., Shosted (2011) or Bennett, Henderson & Harvey (2023).

construction". The distinction between transitive and intransitive verbs is a particularly salient feature in Mayan languages: They are distinguished morphologically not only through person marking, but also sometimes by the use of different TAM (and other) morphemes (Polian 2017: 209). Changes in word class always require derivational affixes: "There is essentially no conversion" (Aissen, England & Zavala Maldonado 2017a: 6). There are, however, varying amounts of ambivalent roots, i.e., roots that belong to more than one class in the languages.

Voice is indicated morphologically as a suffix on the verb. Mayan languages have rich voice systems: there is usually more than one passive (with different semantic nuances) and also several antipassive constructions.[46] Both passives and antipassives decrease the valency of the verb. A valency-increasing mechanism is the causative construction, which is found in many Mayan languages.

Mayan languages are verb-initial, but that does not mean that every sentence starts with a verb. Two (optional) positions before the verb are reserved for topicalization and focus respectively with topic always coming first and focus second if both are present.

There are various types of complex predicates, especially secondary predicates, which have been studied by, e.g., Mateo Toledo (2008; 2022). Serial verb constructions exist at least in Ch'orti' (Quizar 1994a). Complement clauses can be both finite and non-finite. Many auxiliary constructions are described where matrix verbs grammaticalize as auxiliary, while the lexical meaning is expressed by an embedded verb.

However, these generalized traits are contrasted by a great diversity in the individual languages. One such area with a lot of diversity is alignment. As will be discussed in chapter 3, Mayan languages show a variety of split alignment systems which use other types of alignments alongside ergative-absolutive. Chapter 7 will add to this diversity by characterizing Ch'orti' as having a unique kind of alignment not only synchronically but also in terms of its origin.

2.2.3 Language contact

The most extensive recent study on language contact within Mayan as well as with other language families is Law (2014). Though contact will not be central to my argumentation in

[46] The absolutive antipassive either has no patient or an oblique patient. Some languages also have an object incorporation antipassive or an "agent focus antipassive" on which see chapter 7.1.3.

this thesis, I will mention "Lowland" languages here and there and therefore have to introduce the concept. For the sake of completion, I briefly name the other two contact areas as well. So, the following three areas are essential for understanding contact phenomena in Mayan languages (Campbell 2017: 53):

1. The **Greater Lowland Mayan Linguistic Area** mostly involves Cholan and Yucatecan, but also Tseltalan, Q'anjob'alan, Poqom, Q'eqchi' and Ixil[47] (Law 2017b: 116). Law (2014) describes all kinds of diffused innovations ranging from phonological changes to syntactic or semantic features such as a new distinction of inclusive and exclusive in person marking and aspect-based split-ergativity. He even suggests there is evidence for borrowing of bound morphemes like person markers, aspect affixes or numerals classifiers, though this should be considered with caution. It is possible that some of the evidence can be understood without borrowing once we gain a better understanding of the history of Mayan languages, especially the phonological reconstruction of the individual branches and subgroups (see section 2.2.4).

2. The Huehuetenango diffusion area called "the **Huehuetenango Sphere**" includes all Mamean and Greater Q'anjob'alan languages except Tojolab'al and Mocho'. A comprehensive overview is found in Barrett (2002).

3. Outside of Mayan languages, significant convergence in multiple language areas has been described for the broader **Mesoamerican Linguistic Area** initially described in detail in Campbell, Kaufman and Smith-Stark (1986), though many points of this study need to be reevaluated or expanded (see, e.g., Munro 2017). Similarities among Mesoamerican languages discussed in Campbell, Kaufman and Smith-Stark (1986) include a vigesimal counting system, non-verb-final word order and a rich use of "semantic calques" based on areally unique shared metaphors. Often, Mayan is assumed to be the donor for borrowings.

As to foreign influence on Mayan, there is evidence for a "moderate amount of Mije-Sokean influence on certain subsets of Mayan languages" as well as some Totonakan and Zapotecan influence (Kaufman 2017: 63). In some cases, it is even argued that Mije-Sokean influence

[47] Possibly also Huastecan (Law 2017b: 116, 119–120). The idea that Huastecan not only was in contact with the other languages but in fact should be grouped together with Cholan-Tseltalan and Yucatecan based on some shared similarities is maintained in Law (2014) and Robertson (e.g., 1992; Robertson & Houston 2015) but mostly rejected in the scientific community (Campbell 2017: 45).

extended to almost the whole of Mayan (Kaufman 2017: 63). Contact with Xinkan and Lencan languages probably only went the other way around, with Mayan languages, especially Cholan or Greater Tseltalan, influencing Xinkan and Lencan (Campbell 1984: 7–9; 1997: 189). Nahuan languages arrived relatively late to the area according to Campbell (1997) and mostly after the conquest by the Spaniards; at least in Guatemala, the presence of Pipil can be posited securely as early as 800 CE but beyond that, there is no conclusive evidence for early presence or influence of Nahua. Macri and Looper (2003) suggest earlier influence recorded in the hieroglyphic texts; however, their evidence is not convincing and the contribution needs to be reevaluated. Obviously, Spanish has played a major role in the change and decline of Mayan and other indigenous languages. A recent contribution by Dakin, Parodi and Operstein (2017) gives an updated overview of language contact and resulting change in Mesoamerica "and beyond" taking into account the effects of exposure to Spanish as a high-prestige language.

For this dissertation, only the Lowland Mayan contact zone is of importance, but a better understanding of contact effects will help in sorting out open questions in subgrouping in the future (Campbell 2017: 53).

2.2.4 History and Proto-Mayan

Based on the overview of the historical study of Mayan languages given in Campbell (2017), the earliest observations of Mayan languages being related are found in colonial sources already, e.g., in Ximénez (1702), who compared them to the Romance languages, which are "daughters" of Latin. Though some sound correspondences and regular sound changes were described in earlier sources, e.g., Stoll (1884; 1885), Halpern (1942) is considered to present the first real reconstruction of some Proto-Mayan sounds (Campbell 2017: 43). Comparative work adhering to the general standard in historical-comparative linguistics started with McQuown (1955; 1956) whose reconstruction included, e.g., a tonal contrast that was later eliminated in refinements made by Kaufman (1964; 1969; 1976; 1990), who also changed the reconstructed *p' to an implosive *b' (Campbell 2017: 46). Further significant changes were made by Campbell (1977: 89–90), who, based on his reconstruction of the Greater K'iche'an subbranch, showed that, e.g., PM had a distinct phoneme *r in addition to the already reconstructed *y, even though other languages like Yucatecan have lost the distinction (Campbell 2017: 46).

Reconstruction of the linguistic prehistory has yielded some solid conclusions, though much remains to be investigated. Most scholars agree that Huastecan was probably the first Mayan subbranch to separate from the other languages, followed by Yucatecan, which left Cholan-Tseltalan with Q'anjob'alan in a Western Mayan branch and K'ichean with Mamean in an Eastern Mayan branch (Campbell 2017: 54). As a source for the reconstruction of the prehistory, Campbell highlights Kaufman (1976; 2017). However, Kaufman (1976) in particular relies heavily on glottochronological calculations, a common tool back then but today considered to be unreliable (Campbell 2013: 453–458). According to Campbell (2017: 54), "many dispute the accuracy of glottochronology, but at least the dates can be taken as reflecting a general relative chronology". I would be skeptical that it could even be used in this way. Since the diversification scenario mostly still relies on said calculations, the matter needs to be reinvestigated in the future.

Though a lot of progress has been made in the historical study of Mayan languages (summarized in Campbell & Kaufman 1985; Campbell 2017), the field suffers from "preliminarity" on all fronts. Law (2013) is an excellent overview of the kind of evidence that is actually available for all claims made concerning Mayan history. The somewhat unfortunate state of things can be neatly summed up in the following way:

> "For the most part, Kaufman's arguments and evidence for his model of the Mayan language family diversification are unpublished and unavailable to the author and others attempting to evaluate his proposals, though it is often possible to infer what evidence he is likely to have used to develop his hypotheses." (Law 2013: 145)

To demonstrate how little researchers joining the field actually have to go on, I would like to go into more detail on this issue. Campbell (2017: 51) points out that, compared to many other language families, we are lucky to have an etymological dictionary of Mayan that encompasses 1500 entries (Kaufman 2003). However, this resource is still preliminary (as even stated in the title) and can be very difficult to use because it mainly consists of wordlists without discussion. It is therefore a comparative dictionary or a collection of cognate sets, but not an etymological dictionary in the proper sense. It is especially unreliable for languages that are studied less or were only documented extensively later like Ch'orti'. Cognates are often not identified if they are not immediately obvious via the most basic of sound correspondences. Additionally, it contains confusing reconstructed language stages like "LL" (Lowland Mayan languages). It does not make sense to reconstruct a stage like this because the Lowland Mayan languages are an areal, not a genetic grouping. Though contact and borrowing must, of course, be taken into account for reconstruction, it is, by definition of the word, neither customary not useful to reconstruct a common "ancestor" for an areal

phenomenon. Moving on from lexical to phonological and grammatical data, a little bit of everything can be found in Kaufman's (2015) unpublished manuscript on "Mayan Comparative Studies", which is more than 1000 pages long and offers insights into Kaufman's work in progress which, unfortunately, also means that it is a challenging read and that one often runs into large gaps, especially data-wise.

Likewise, while most languages and subbranches have been the subject of reconstructive attempts, sometimes in more and sometimes in less elaborate ways[48], we are mostly lacking truly detailed reconstructions that use a bottom-up approach, go beyond an initial survey and work out all the details as well as solve problematic cases and take into account as well as present all of the available language material. Especially the latter point, also mentioned in the quote above by Law, is a significant problem: it should go without saying that publishing reconstructions is not enough, it is essential to also provide detailed cognate sets and discussion of these sets to explain what the reconstructions are based on. In their reconstruction of Proto-Cholan, Kaufman and Norman (1984) list the languages that the reconstructions are based on but mostly do not provide the actual forms. On the other hand, Campbell (1977: 46) in his reconstruction of Proto-K'iche'an does provide actual lexical material but states that he has standardized dialect data thus eliminating variation, which already constitutes a reconstruction in itself that would have required transparency and justification. The latest publication by Kaufman (2017), where he revises some of his reconstructions, suffers from the same problem as most other publications on Proto-Mayan reconstruction: the readers are confronted with long lists of reconstructed lexemes but no evidence on what forms from individual languages his reconstructions are based on. Instead, they are referred to his etymological dictionary (Kaufman 2003), though Kaufman (2017: 64, 106) admits that it is significantly incomplete and includes some errors.

England (1992) treats the relationship between Mayan languages and the reconstruction of Proto-Mayan and does discuss sound correspondences with cognate sets but the book by no means constitutes a full reconstruction as only a selection of language features are included. The Swadesh list at the end of the book with Proto-Mayan reconstructed forms is not based on all languages. As it stands, this unfortunate situation ultimately leads to the necessity to redo basic groundwork as actual and sufficient evidence for claims put forward in publications is often difficult to come by; on the other hand, trust is not something one can base a discipline

[48] For instance, concerning the phonological reconstruction, Schweitzer (2006: 54) points out that work on the consonants is a lot more extensive than on the vowels.

on. As a defense, the respective authors likely did not expect that so few people would follow in their footsteps to continue said groundwork.[49]

Aside from Cholan and K'iche'an, noteworthy reconstructions of Mayan branches are Norcliffe (2003) for Proto-Huastecan, Kaufman (1972) for Proto-Tseltalan and Fisher (1973) for Yucatecan. Hofling (e.g., 2006; 2017; 2018) has likewise worked extensively on the diachrony and comparative grammar of Yucatecan. I am not aware of a designated publication reconstructing Proto-Mamean systematically; the absence of any such publication is confirmed in Adell (2016a: 1) who provided a discussion of coronal contrasts in Ixil and Mamean in general. The Q'anjob'alan branch is sometimes named as the one where a "serious reconstruction" (Campbell 2017: 45) is needed the most[50], e.g., to clarify the status of Tojolab'al and the exact relationship of the Q'anjob'alan languages with Cholan-Tseltalan (in the broader group of Western Maya). However, a new reconstruction of Cholan is equally needed – and not only because we now have more and better data on Ch'orti' and Chontal. For instance, the change of PM *ee to PCH *i as well as PM *oo > PCH *u is of significant importance as it is essentially the main defining feature of Cholan (Campbell 2013: 181). Nevertheless, in Kaufman and Norman's paper on Proto-Cholan reconstruction it is noted that "this change is neither regular nor pervasive, i.e., there are more cases where it does not take place than where it occurs." (Kaufman & Norman 1984: 87) This clearly indicates that the phonological reconstruction is far from complete. Unfortunately, little further reconstructive work is being pursued in Cholan. While some scholars like Mora-Marín work on questions concerning details such as the reconstruction of the Mayan applicative and antidative (2003), the Proto-Cholan positional status marker *-täl (2005a) or the Proto-Cholan pronouns, deictics and definite articles (2009a), basic questions about the historical phonology of the

[49] Studies like Storniolo (2008) on Eastern Cholan or Becquey (2014) on Comparative Cholan cannot be regarded as new or more in-depth reconstructions even though the titles of their works might suggest so. Neither constitutes a systematic "bottom-up" approach that compares the attested languages and reconstructs a common ancestor. Becquey proceeds precisely the other way around in mostly discussing the changes that Proto-Cholan underwent on its way from Proto-Mayan to the individual languages based on existing hypotheses and stating which he finds more likely, often without giving arguments. His comparison of Cholan does not include Hieroglyphic Maya. Storniolo's approach to Eastern Cholan, on the other hand, is not systematic but instead focuses on specific features of the grammar of Ch'orti' and Choltí. Unfortunately, she does not present convincing evidence for the claims that she makes, e.g., the presence of set C pronouns in the hieroglyphic texts at Copan (on set C see chapters 4 and 5).

[50] Again, Adell (2016b) is worth mentioning where PM *t and *t^{y} are discussed in the context of Q'anjob'alan evidence.

Cholan languages (e.g., the conditions of the change of PCH *l > CHR *r*) remain unanswered[51]. As long as this is not remedied, all treatments that deal with details will rest on a shaky foundation.[52] I should state that the critical view expressed in this chapter is not only my own; it is shared by other scholars, e.g., Schweitzer (2006: 54)[53] or Adell (2016b)[54].

Since this thesis is concerned with Mayan alignment, a few words on the study and reconstruction of Mayan morphosyntax are in order. A general treatment of Proto-Mayan syntax is Norman and Campbell (1978). Robertson (1980) offers a detailed reconstruction of Mayan pronouns and their incorporation into the "verbal complex", while Robertson (1992) reconstructs Mayan tense, aspect, mood and voice systems. Robertson's work, despite claiming to be based on the historical-comparative method, suffers from similar shortcomings as his approach to set C in Ch'orti', which is discussed in detail in chapter 5.1. It is therefore to be used with caution.[55] So far, it can be said with some degree of certainty that Proto-Mayan was "an ergative language, with associated antipassive constructions" (Campbell 2017: 51) and that, just like in modern Mayan languages (see section 3.1), A was marked with

[51] Note also Robertson (2010) who in attempting to follow Mora-Marín's (2009b: 151) demand for "a more thorough reconstruction of the history of the Cholan-Tzeltalan languages" proceeds to discuss solely morphological evidence.

[52] Questions concerning morphology cannot be addressed adequately as long as the historical phonology has not been worked out in detail. For example, Campbell (1984: 6) in his overview of Mayan sound correspondences for Maya epigraphers simply states that Ch'orti' has *r* where other languages have *l*. Though this is not true for all phonetic environments, nobody has so far formulated and published clear rules as to when this sound change happens and when it does not. The same applies to the more recent overview of sound correspondences in Davletshin (2013a: 74–75). Korobzow (2023) presented some first results concerning conditioned sound changes in Ch'orti': the commonly cited sound change of PM *l > CHR *r* only happens in intervocalic and word-final position while word-initially, it remains *l*. The Ch'orti' reflexes of PCH *j, h are discussed in chapter 6.2.

[53] Schweitzer (2006: 54) laments the "sound correspondences that still need be considered as basic [German: "die immer noch als basal zu betrachtenden Lautentsprechungen"; NK: "basal" here seems to mean "basic/preliminary", not "foundational", judging from the context]" and addresses the same problems in Kaufman's work that I also pointed out.

[54] "Systematic arguments underlying the commonly accepted reconstructions of the proto-Mayan (pM) coronals */t/, */tj /, and */ʧ/ have not been published, nor have the cognate sets demonstrating the correspondences upon which they are based" (Adell 2016b: 1).

[55] Schweitzer (2006: 54) says about Robertson's work that he tends to use the questionable method of "reconstructing forward", i.e., using a preliminary reconstruction of the Proto-language to argue for proposed developments in the daughter languages, which carries the danger of circular reasoning. More importantly: "What is particularly problematic with many of Robertson's arguments is that he places too much emphasis on the theoretical postulation of systems and filling in gaps, while using the instrument of analogy in a very free manner. [German: "Problematisch bei vielen Argumentationen von Robertson ist insbesondere, daß er zu sehr das theoretische Postulieren von Systemen und das Ausfüllen von Lücken in den Vordergrund stellt und dabei das Instrument der Analogie in sehr freier Weise einsetzt."]" (Schweitzer 2006: 54). On this see also chapter 5.1 of this dissertation.

a special index-set (set A) which was also used to mark possession on nouns while S and O were marked by a second index-set (set B).

Despite being preliminary, the proposed reconstructions of sound systems, lexicon, subgrouping etc. are used as an argumentative base by many, "for convenience", as Adell (2016b: 1) states. I will cautiously do the same in my dissertation whenever a historical perspective is needed, since I am left with no alternative, but a new reconstruction of the Cholan branch is desirable for the future.

2.2.5 Ch'orti'

Ch'orti' is a Mayan language of the Cholan branch. This branch is said to be characterized by the shared innovation of raising of long mid vowels PM *ee and *oo to PCH *i and *u respectively (Campbell 2013: 181). Campbell (1984: 14) exemplifies this with the following dataset (Table 2):

Table 2. Correspondence sets for Ch'orti' compared to other Cholan languages. Data taken from Hull (2016) for Ch'orti' and from Kaufman (2003) for the rest, supplemented by Hopkins, Josserand and Cruz Guzmán (2011) for Chol and Kettunen and Helmke (2020) for Hieroglyphic Maya.

Proto-Maya	Ch'orti'	Chol	Tseltal	K'iche'	Yucatecan	Meaning
*keej	*chij*	*chijmay*	*chij*	*keej*	YUC *kéej*	'deer'
*b'eeh	HGM *b'ih*	*b'ij*	*b'e*	*b'e(e)*	LAC *b'eh*	'road'
*tooŋ	*tun*	*tun*	*ton*	–	YUC *tùun*	'stone'
*sootz'	*sutz'*	*sutz'*	*sotz'*	*sootz'*	YUC *sootz'*	'bat'
*oox-	*ux-*	*ux-*	*ox-*	*ox-*	YUC *oox-*	'3'

As the data in Table 2 show, other languages sporadically show similar reflexes, e.g., YUC *tùun* or TSE *chij*. This is something that still needs to be investigated in detail for each branch (see preceding chapter) and even within Cholan the sound change is "neither regular nor pervasive" (Kaufman & Norman 1984: 87). Without further clarification of when this change takes place, one might even question its use for the definition of subgrouping.[56]

[56] It has been proposed that irregular, "lexically-specific sound change" (François 2014: 178) can likewise be applied for subgrouping if it affects the same lexemes in all relevant languages. However, even if true this does not mean that we do not need to investigate why it did not happen in other lexemes that seemingly show the same phonological environment.

Cholan is commonly divided into a Western Cholan branch that consists of Chol and Chontal and an Eastern Cholan branch consisting of Colonial Choltí and Ch'orti'. On the status of Hieroglyphic Maya as being ancestral to either Ch'orti', Eastern Cholan or all Cholan languages see the brief discussion in chapter 2.2.1. In the following sections, I want to provide a little more context for where Ch'orti' is spoken now, where it used to be spoken and how we know.

2.2.5.1 The Ch'orti' area

> For one, it is controversial even to refer to a "Ch'orti' area." The very word Ch'orti' is problematic because, according to some ethnographers and linguists (for example, Wisdom, Fought), the /ch/ phoneme is not glottalized, as the diacritic would suggest. Whence did this Ch'orti' spelling come, and why do some use it whereas others do not? Regarding the "area," if we are referring to the region where Ch'orti' culture is practiced and language is used, then to what exact time frame are we referring? Ch'orti' culture and language existed in what is today eastern Guatemala, western Honduras, and northwestern El Salvador at the time of the Spanish-Mexican invasion, but this is not necessarily the case before and after that event. Is it not misleading to insinuate by the use of "Ch'orti' area" that a unitary Ch'orti' culture has existed over time, as if it were primordial and enduring? (Metz 2009a: 1)

It is easy to assume continuity for a region based on what is known about it today. However, as the quote makes clear: we need to be cautious. Even with the hieroglyphic texts at our disposal, which record a language not entirely identical to Ch'orti', interpretative work is needed to understand where the history of the Ch'orti' language actually begins. The area where Ch'orti' – among other languages – is spoken and is assumed to have been spoken earlier can roughly be described as the area around the borders of today's Guatemala, Honduras and El Salvador. It is therefore not in the center of the Mayan language area but on the periphery (Metz 2009a: 1).[57]

In the following chapters, I will sketch what we know or can plausibly assume about the prehistory of the Ch'orti' language starting with the earliest sources. Information can be scarce up until documentation of the language starts and it is necessary to rely heavily upon inferences from archaeology or colonial sources, which are external descriptions. Of course, it is notoriously difficult to correlate archaeological data with linguistic ones. Unfortunately, we lack historical sources with an internal perspective on the Ch'orti'. The introduction in Hull (2016) as well as Metz (2006; 2022) and many of the papers in Metz, McNeil and Hull (2009), provide further information on the Ch'orti' area including further literature and specific names of places where Ch'orti' was reported to have been spoken by contemporary

[57] For a map, see chapter 2.2.1.

sources. Herranz (2000) deals with the history and present of Ch'orti' and other indigenous communities in Honduras (in Spanish).

2.2.5.1.1 Before colonization

Little is known about the population of this area in the Preclassic[58], but it is commonly accepted that in the Classic period, the people of this region – or at least the ruling class – were Maya (Metz 2009a: 3). Among the important historical sites of the Classic Maya culture in the area settled by the Ch'orti' today are Copan in Honduras and Quiriguá (north of Copan in Guatemala, see Figure 3). At these sites hieroglyphic texts are attested which reveal the (linguistic) identity of the site's occupants. However, as mentioned in chapter 2.2.1, the exact language of the hieroglyphic texts is still a matter of dispute aside from the fact that it belongs to the Cholan branch of Maya. Therefore, the answer to the question of when Hieroglyphic Maya stops and when Ch'orti' begins ultimately lies in the exact identification of the variety of Cholan that Hieroglyphic Maya records.

Figure 3. Important Preclassic archaeological sites in Southern Mesoamerica.[59]

Based on archaeological records, it is generally assumed that around 100 CE, a group of "Proto-Ch'orti'" Maya[60] came to the region from the west of Guatemala and took over control

[58] Mesoamerican civilization is divided into the following eras archaeologically: Preclassic (2000 BCE – 250 CE), Classic (250–909) and Postclassic (909–1697) (Martin & Grube 2008: 8–9).

[59] https://en.wikipedia.org/wiki/Kaminaljuyu#/media/File:Formative_Period_southern_Mesoamerica_2.svg (last accessed: 2024-02-11). © user Madman2001, 2008, CC BY-SA 3.0 license (https://creativecommons.org/licenses/by-sa/3.0/). No changes made.

of the Copan valley and its resources from a non-Maya population that probably spoke a language ancestral to modern Lenca (Sharer 2009: 131).[61] Sharer, unfortunately, does not explain why this language is assumed to have been Lencan. The exact reason for the eastward movement of Mayan speakers is also unknown, but it correlates with dramatic changes among the earliest Mayan polities of the Preclassic like Kaminaljuyú (Sharer 2009: 130), see Figure 3. For instance, carved stone monuments and texts disappeared completely and the Lake Miraflores – the source for Kaminaljuyú's irrigation network – dried up at around the same time, i.e., in 100 CE (Sharer 2009: 130). This is believed to have prompted population movement.

More than 300 years later, as the hieroglyphic texts tell us, in 426–427 CE, ruler K'inich Yax K'uk' Mo' founded a new dynasty in Copan, which presumably led to additional migration of Maya people to the Copan Valley (Sharer 2009: 131). These new Maya presumably spoke a Cholan language from the Lowlands that was closely related to the one already spoken in the Copan valley by the Maya that had arrived from the highlands following the displacement around 100 CE (Sharer 2009: 131–132).

However, the hieroglyphic texts only speak for a small elite group of the population, the ruling class (Sharer 2009: 125). It need not necessarily be the case that the commoners of Copan and Quiriguá also spoke a Cholan language. Indeed, although inferring culture or even language from archaeological records alone is difficult, Sharer (2009: 125), among many others, points out that in the last few years, a growing body of research has shown that, while it can be securely assumed that the elites of Copan and Quiriguá were indeed Cholan Maya, their subjects were most likely ethnically diverse. It is, however, unclear what languages were in contact with the predecessor of Ch'orti' at that time.

Terga (1980: 29–36) and Girard (1977) suggest that by the early Postclassic, the Pipil (Nahua, Uto-Aztecan) invaded the region from the Pacific coast of contemporary Guatemala and from El Salvador (Metz 2009a: 5). According to Lardé y Larín (1955: 19), this influx of Pipil led to the "Ch'orti'" abandoning existing towns and founding Camotán and Jocotán further north (in today's Chiquimula department; maps displaying some of the towns and areas are found in

[60] With "Proto-Ch'orti'", Sharer likely means "the assumed predecessor of Ch'orti'", whatever stage of Cholan their language might have represented.

[61] Lencan is not related to Mayan (or any other language family), as far as is known today.

section 2.2.5.1.3), after they had been pushed out of the area around Lake Metapán (Metz 2009a: 5). These are also the remaining areas where Ch'orti' is still spoken today.[62]

Early colonial documents suggest that right before the Spanish conquest, what is known as Ch'orti' today may have been called "Apay" or "Apayak". Metz (2009a: 5) claims that the term has no meaning in modern Ch'orti'. On the other hand, Robertson, Law and Haertel (2010: 30) point out that the term *apay* appears in the Morán Manuscript of Colonial Cholti with the gloss 'friend'. Thompson (1990: 94) proposes that *-ak* may be Ch'orti' *ak'* 'tongue', which I find a plausible interpretation for the second part of the word.

Diego García de Palacio reported in 1576 that Apay is spoken in Copan and the valleys of Acacebastla and Chiquimula ("Chimula") and also understood in Yucatan and other provinces which are Uyajal, Lacandón, Verapaz, Chiquimula, Copan in modern orthography (Hull 2016: 1; Scholes & Roys 1968: 26; Fernández Guardia 1883: 7).[63] Fernández Guardia (1883: 7) hypothesized that "Apay is probably the same Ch'orti'"[64], although he did not provide any arguments and in fact described Ch'orti' as a language "of which you can say that we know nothing about it".[65] He only refers to Brasseur de Bourbourg who thought that Ch'orti' was identical to Cholti and that he himself, who visited the region two times, tends to agree that they are related (Fernández Guardia 1883: 7). However, he also identifies Cholti with Chol (Fernández Guardia 1883: 7). Should that have been the case at least in de Palacio's times, the question remains as to when the single language started diversifying – after all, Ch'orti' and Chol today are not identical and Cholti is likewise considered to be its own variety.

Hull considers the debate as to the identity of Apay settled: "There is little doubt that Apay and Ch'orti' are one and the same language" (Hull 2016: 1). To me, this seems a lot less clear. The mutual intelligibility that García de Palacio describes does not necessarily mean that it

[62] According to Robertson, Law and Haertel (2010: 31), the area where Ch'orti' (or its predecessor) was spoken before the arrival of the Spanish must have stretched a lot farther north. This is based on the assumption that Colonial Cholti is a very close relative of Ch'orti' and the two must have in fact been dialects at the time Cholti was recorded, i.e., in the 17th century. Under this view, the conquest divided the area into a northern and a southern part and eventually lead to the complete disappearance of the northern population, while the southern survives as Ch'orti' today. As I do not believe that Ch'orti' is a direct descendant of Cholti, I advise to regard this with caution.

[63] "[…] and so the language Apay that they speak here runs and is understood in Yucatán and the mentioned provinces [Span. <[…] y ansi la lengua apay que aqui hablan corre y se entiende en yucatan y las provincias dichas>]" (García de Palacio 1576: 26). The "mentioned" provinces are: <ayajal lacandon verapaz y la tierra de chiquimula y esta de copan> (García de Palacio 1576: 26); in modern orthography: "Uyajal"? (Scholes & Roys 1968: 18), Lacandón, Verapaz, Chiquimula and Copán.

[64] Span. "La Apay es probablemente la misma Chorti."

[65] Span. "[…] de la que se puede decir que no sabemos nada".

was Ch'orti' that was spoken over an area that stretched as far as Yucatán etc. Sapper (1893: 3) reports that there is a high degree of intelligibility based on "frequent mutual contact, i.e., habituation"[66] at least among closely related Mayan languages like Poqomchi' and Q'eqchi'. It is possible that García de Palacio's statement referred not to Ch'orti' specifically but to a subgroup of Cholan or even a larger group.[67]

2.2.5.1.2 Spanish conquest (1524–1821)

1524 marks the year when Spanish troops under the leadership of Pedro de Alvarado came to Guatemala from Central Mexico (Brewer 2009: 138). Unfortunately, we lack any kind of documents from the point of view of the Ch'orti' on their experience of Guatemala's colonial times so that it is necessary to rely on Spanish sources for eastern Guatemala (Brewer 2009: 137). This means that one should keep in mind that the sources are not unbiased and likely paint the Ch'orti' in a worse light than they would have themselves.

After some initial smaller revolts by the Ch'orti'[68], in 1530, a local indigenous leader named Q'alel (<Copan Calel>) put together an army of 30.000 to fight the invaders at Copan[69]; unfortunately, this fight ended after several days with the victory of the Spanish and their local allies (Brewer 2009: 139). This did not mean the end of the general resistance, however. In fact, it remained a difficult – and not very lucrative – task to control the Ch'orti' (Brewer 2009: 139). Apparently, there were still sporadic Ch'orti' uprisings in the 18th century (Brewer 2009: 145).

With growing numbers of Spanish settlers, the Ch'orti' population declined constantly due to "disease, violence, out-migration, and forced labor" (Brewer 2009: 141). The Ch'orti' population in Chiquimula, e.g., dropped by 50% between 1549 and 1589 only, while other regions were even more drastically affected with decreases of 65% in Esquipulas and 88% in Jalapa (Brewer 2009: 145). The population of Jilotepeque increased by 23%, but this was due

[66] German: "durch häufigen wechselseitigen Verkehr, also Gewöhnung".

[67] This is especially true if one takes into account that more than 300 years of potential language change lie between García de Palacio and Sapper so that one can assume that the languages of the region were to some degree more similar to each other in García de Palacio's time.

[68] At this moment, it remains unclear to me if we are able to identify the indigenous population that this account refers to as Ch'orti' or their predecessors beyond doubt. Therefore, the statements in this chapter are subject to reservations up to the first mention of the name "Ch'orti'" by Cortés y Larraz.

[69] This is not the same Copan as the big polity in the Classic – and the historical site still known today – but a fortress "near modern-day Rincón del Jicaque in Honduras near the Lempa River" (Brewer 2009: 139). Feldman (2009: 149–150) describes various settlements that we know of with the name "Copan"/"Copán".

to flight from other regions (Brewer 2009: 145). Eventually, the rest of the Ch'orti' became isolated in Jocotán and Camotán, while African slaves replaced them as laborers (Brewer 2009: 141).

For many Mayan languages, we can today profit from the work of missionaries who wanted to preach in the local languages and thus learned them and left resources attesting to the colonial varieties for us to study. However, the evangelization of the Ch'orti' was done by clergymen less inclined to preach in the indigenous language: they used Spanish instead, which is why they were less successful than missionaries that used local languages (Brewer 2009: 142). Unfortunately, this also led to more Ch'orti' Maya ultimately abandoning their language altogether when evangelization and learning of the Spanish language became inevitable (Brewer 2009: 142).

Compared to western Guatemala, the colonial experience of the Ch'orti' was "especially harsh" (Metz 2006: 9). Many lost their livelihood since the Spanish confiscated the lands; some were bilingual in Spanish and Ch'orti' as early as by the end of the 1500s and in El Salvador, the Ch'orti' were no longer counted as a distinct population by the mid-1800s (Metz 2009a: 5).

By the end of the colonial period, the Ch'orti' had "lost much of their indigenous cultural heritage and were slowly absorbed into Spanish society" (Brewer 2009: 146). As Feldman (2009: 148) points out, since the Spanish rule – and the pressures placed on the indigenous people by it – caused population instability and displacement, it is difficult to assume a continuity between today's Ch'orti' Maya and ancient societies that lived in the same area, as already pointed out in the introductory quote of chapter 2.2.5.1.

The earliest attested occurrence of the name <Chorti> is from this period; more specifically, from 1768 when Pedro Cortés y Larraz (1712–1787), Archbishop of Guatemala, visited his diocese and documented what languages were spoken in what towns (Cortés y Larraz 1771: e.g., 118; Metz 2009b: 169). This is a little earlier than 1795 as claimed in Robertson, Law and Haertel (2010: 30). Cortés y Larraz documents speakers of Ch'orti' for the following settlements: Los Esclavos, Tejutla, Jilotepeque (Pueblo de San Marcos: Poqomam, Pueblo de <Ycpala>: Ch'orti'), Quesaltepeque, Jocotan, Chiquimula, Zacapa, San Christobal

Acasaguastlan.[70] Almost all of these are confirmed in Galindo's report from 1834 (see following chapter and Figure 4).

2.2.5.1.3 1821–1996

In 1821, Guatemala, Honduras and El Salvador gained independence from Spain but this did not mean that the Ch'orti' area and population stopped to diminish. In the first half of the 19th century, it was reported that Ch'orti' was the common vernacular in the department of Chiquimula (Hull 2016: 1). Galindo's (1834) report provides more information, including settlements that are not mentioned as Ch'orti'-speaking elsewhere, e.g., Ocotepeque. Thompson (1990: 92) argued that, since Ocotepeque is situated in between "two Chorti concentrations", the information is credible. Galindo is the earliest source where we can securely assume that the language he refers to is in fact Ch'orti' since he also recorded a wordlist (see Appendix B). Figure 4 shows all (so far) identifiable places from Galindo's report.

[70] This list includes all settlements mentioned in volume I which is available digitally here: https://es.wikisource.org/wiki/Descripci%C3%B3n_Geogr%C3%A1fico-Moral_de_la_Di%C3%B3cesis_de_Goathemala (last accessed: 2024-02-11). Volume II is partly, volume III almost entirely undigitized.

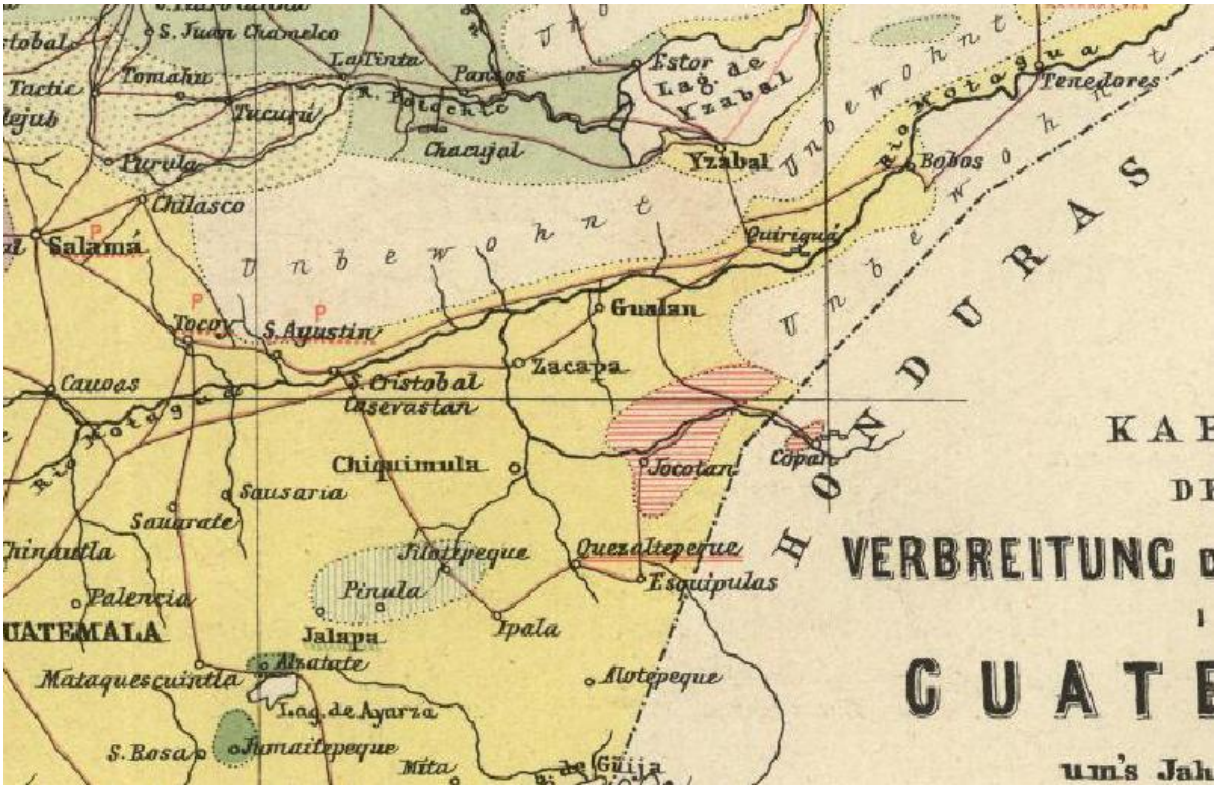
Figure 4. Galindo's "Chortí empire" with settlements in Guatemala (red), Honduras (blue) and El Salvador (yellow).[71] Map data ©2025 Google.

According to Hull, Ruano Suárez reported in 1892 that at the end of the 19th century, Ch'orti' was "virtually unknown to those under fifty years old in Chiquimula" (Hull 2016: 1). This assessment either refers to Chiquimula the town (i.e. not the department) or it is too pessimistic as the area where Ch'orti' is still spoken today (Jocotán and Camotán) is in the Chiquimula department. Unfortunately, I did not have access to this part of Suárez' publication to check the original source.

[71] Unfortunately, not all settlement names can be displayed properly on the map. The dataset that the map is based on is available in my PhD repository on Github: https://github.com/Korobzow/PhD
Full quote: "Throughout this extent of land the Chortí language was spoken and is still spoken, and from these and other data it is inferred that the peoples of Cuaginiquilapa, Los Esclavos, Quesaltepeque, San Jacinto, Santa Elena, San Estéban, San Juan Ermita or del Río, Jocotán, Camotán, San José, Chimalapa, Sacapa, and San Pablo, in the State of **Guatemala**, formed a part of this empire. Chiquimula and Esquipulas were governed by subordinate princes of the King of Copan. In **Honduras**, Omoa, the mineral district of San Andrés, Sensenti, Ocotepeque, Tipalpa, La Brea, and other places were comprised in their dominions. In **Salvador** the same empire comprised Texis, Dulce Nombre, Metapas, Tejutla, and Sitalá." (Galindo in Morley 1920: 595, emphasis added).
Unlocalized: <Cuaginiquilapa>, Santa Elena, Chimalapa, San Pablo, Omoa, San Andrés, Tipalpa, La Brea. Some of them can be found in the countries today but are suspiciously far off from the general area where Ch'orti' places cluster.

A map prepared by German geographer Karl Sapper (1893) shows where Ch'orti' was still spoken at the end of the 19th century: around Jocotán and Copán at the border between Guatemala and Honduras (cf. Figure 5). The Ch'orti' area is marked in red-and-white horizontal stripes. Note that the town of Chiquimula (to the west of Jocotán) indeed is not part of the area. Compared to Figure 4 above, the area where Ch'orti' was spoken had shrunk dramatically by the end of the 19th century.

Figure 5. Languages of Guatemala around 1892. Part of a map by Sapper (1893).[72]



In the 1930s and 40s, especially hard policies against "Indians" were adopted by the national governments (Metz 2009a: 5–6). During the Guatemalan Civil War (1960–1996), a genocide was committed against the Maya which in total took more than 200,000 lives (Commission for Historical Clarification (CEH) 1999: 17).

2.2.5.1.4 Ch'orti' today

> No one speaks Ch'orti' in El Salvador, nor in Honduras, for all intents and purposes. Chiquimula, Esquipulas, and Zacapa have no Ch'orti' speakers. The town of Jocotán is now more than 90 percent Ladino, and the neighboring town of Camotán has only a few speakers of Ch'orti'. Many of the hamlets surrounding Jocotán have 50 percent or more of the population who speak

[72] Full map in high resolution digitalized here: https://zs.thulb.uni-jena.de/receive/jportal_jparticle_00668886 (last accessed: 2024-02-07).

> Ch'orti', and in places like Pelillo Negro, Pacrén, parts of Olopa, and a number of others [sic] hamlets, Ch'orti' is the dominant language of daily interactions. (Hull 2016: 2)

Today, the Ch'orti'[73] live near the Guatemalan-Honduran border, more specifically in the Guatemalan department of Chiquimula and extending slightly into northern Honduras (Hull 2016: 1). However, not all of those who identify as Ch'orti' ethnically also still speak Ch'orti'. Somewhat dated population figures from 1994 give a Ch'orti' population of 68,154 for Chiquimula, but by Hull's (2016: 1) estimates, the number of ethnic Ch'orti' who actually speak the language is closer to 12,000. They split into about thirty communities where knowledge of the language can vary drastically with at least ten communities having completely abandoned Ch'orti' (Hull 2016: 1). Most of the remaining speakers in Guatemala are bilingual in Ch'orti' and Spanish (Hull 2016: 1). However, Metz confirms that, since the 1990s, "a strong Ch'orti' identity has resurged and expanded greatly from the fifteen or so rural communities in Jocotán and Olopa Guatemala, where nearly everyone still speaks Ch'orti'" (Metz 2009a: 6).

As to Honduras, with "more acculturation, and extreme poverty, identity and motivation revolve foremost around land, with language and culture recovery being secondary" (Metz, McNeil & Hull 2009: 157). Though we lack data to establish when exactly the language died out in Honduras, sources between 1834 and 1940 demonstrate that there were at least some speakers in the villages around the village of Copán Ruinas (Herranz 2000: 313). The most extensive information comes from Rafael Girard, who traveled the Honduran Ch'orti' area during many years, and wrote in 1940 that it is "a practically dead language"[74] except for Copán and El Paraíso[75] where only traces of it survive (Herranz 2000: 313–314). Herranz (2000: 312) provides population figures from 1988 for the department of Copán (Honduras) which show that by then, there were only three people who spoke Ch'orti' and all of them were actually Guatemalan farmers that had arrived by 1960, making Honduran Ch'orti' a moribund language at that time, before it died out entirely.[76] Now, although many ethnic

[73] Although etymologically, Ch'orti' only refers to the language with *-ti'* meaning 'mouth, speech', it is also used as the ethnonym for the people.

[74] Span.: "casi una lengua muerta".

[75] Knowledge of the language was already limited then, however: Girard's collected vocabulary is much more complete than that of Membreño, but nevertheless, he had to use Membreño for some very basic vocabulary like the verbs <guehej> 'to eat', <jujuy> 'to chew' and <nehen> 'to be', because these could neither be given nor even recognized by his Honduran informants (Herranz 2000: 314).

[76] In fact, there is no systematic study that would confirm that Ch'orti' indeed has died out completely everywhere in Honduras but it is rather likely (Herranz 2000: 315).

Ch'orti' remain in northern Honduras in the vicinity of the city of Copán Ruinas, the Ch'orti' language is no longer in use (Hull 2016: 1).

The complete "lack of identity in El Salvador" (Metz, McNeil & Hull 2009: 157) is not surprising considering the early extinction of the language as compared to its relatively long survival in Honduras and survival to this day in Guatemala.

2.2.5.2 Language name

As already mentioned, the earliest attestation of <Chorti> is from 1768 by Cortés y Larraz. Whether earlier reports that speak of "Apay" or "Apayak" refer to Ch'orti' or not is unclear at this point (see chapter 2.2.5.1.1). Concerning the modern language name ("Ch'orti'" as opposed to earlier "Chorti" or "Chortí"), a note on orthography is in order. To this day, experts from different fields are divided on the question of how the name of the language in question should be written accurately. According to Wisdom (1940: 6), the /ch/ in Ch'orti' is not glottalized. This leads to a translation of *Chor-ti'* as 'field language' consisting of *chor* 'field, clearing' and *ti'* 'mouth; language', which corresponds to the name of the alleged closest relative Choltí used by Friar Francisco Morán in 1695, who gave the translation "language of the corn farmers"[77] (Metz 2009a: 2).

Guatemalan Maya linguists reject the spelling Chorti' because they perceive the connotations of 'field language' as derogatory and claim that a glottalized /ch'/ is also phonetically more accurate (Metz 2009a: 2). *Ch'or-* would then be related to the reduplicated form *ch'orch'or* 'larynx' (Metz 2009a: 2). Most academics now follow this orthography, though one of the leading linguists on Ch'orti', John S. Fought, rejects this spelling specifically on phonetic grounds (Metz 2009a: 3). Interestingly, native speaker Marcos García (2015) also uses the spelling Chorti' in his thesis.

It seems that the choice of <ch> or <ch'> is a matter of ideology, not scientific accuracy. As Metz (2009a: 2) states, "*Ch'orti'* is […] a complex product of phonetics, ethnicity, and politics". Since Ch'orti' is the established language name by now and as language activists explicitly demand this spelling, it will be used throughout this dissertation, although it seems that Chorti' might factually be more correct. On the other hand, as mentioned in section 2.2.1, I continue to employ the colonial orthography <Choltí> because Morán explicitly stated that it

[77] Span.: "lengua de los milperos". *Milpa* is a word of Nahuatl origin that refers to a field for growing crops, especially corn. A more accurate translation would be "language of the fields" (see footnote 95).

is the “lengua de los milperos” and this is the only information that we have on that language. As to the other Cholan languages, according to Coon (2017: 648) the question of “Ch’ol” (Tumbalá) versus “Chol” (Tila) is a matter of dialect. Where I do not differentiate between dialects, I will use “Chol”. Chontal generally is not glottalized, though I have seen instances of “Ch’ontal” in more than one publication but from the fact that the correct spelling occurs alongside the unglottalized one in the same text, one can deduce that those instances are simple errors, perhaps prompted by the writing of Ch’orti’ and Ch’ol with ’.

2.2.5.3 Linguistic documentation

These are the sources on Ch’orti’ available to us in chronological order of attestation starting with the oldest one:

Cortés y Larraz (1768): Record of the name <Chorti>.[78]

Juan Galindo (**1834**): Political activist and explorer Galindo visited many Mayan ruins and wrote letters in several languages to report his findings. He recorded 51 Ch’orti’ words during his visit to Copan (Hull 2016: 2). His report is printed in Morley (1920: 593–604).[79] The digitized (scanned typewriter document) version of the English letter contains a mistake (lexeme <Nojtá> ‘big’ is left out). I have worked with Morley’s and the French version of Galindo’s original handwritten letter (Galindo 1834). The data can be found in Appendix B including modern lexical counterparts.

[John Lloyd Stephens (1839) recorded a small list of words which unfortunately turned out not to be Ch’orti’ but Poqomam (Hull 2016: 2).[80]]

Ruano Suárez (**1892**): Suárez collected Poqomam vocabulary in San Luis Jilotepeque (Guatemala) and Ch’orti’ from the area of Chiquimula (Guatemala) and possibly Copán Ruinas (Honduras) and made a similar mistake as Stephens, only the other way around, in that

[78] Record of the name itself is significant for, e.g., dating the sound change of PCH *l > CHR *r*.

[79] Digitized here: https://cdm15999.contentdm.oclc.org/digital/collection/SCMisc/id/50383 (last accessed: 2025-03-13).

[80] The recorded words definitely do not correspond to Ch’orti’ both in phonology and in the actual lexemes used: <nal> ‘maize’ would be *nar* and <cahchi> ‘mouth’ has a second part *-chi’*, which would have been *ti’* in Ch’orti’. This sound change proves that we must be dealing with a Greater K’iche’an language since this is their exclusive innovation (Campbell 1984: 6). The transcribed <tzeken> ‘bird’ also does not fit since ‘bird’ is *mut* in Ch’orti’ and even when considering semantic change, there is no cognate of the Poqomam term with a different meaning. The mix-up is not surprising considering that Poqomam and Ch’orti’ apparently were spoken in the vicinity of each other (see the list of settlements from Cortés y Larraz at the end of chapter 2.2.5.1.2).

he assumed that Ch'orti' is just a dialect of Poqomam (Hull 2016: 2). Nevertheless, he did collect valid Ch'orti' data. He describes the indigenous population of Chiquimula as former residents of Copan who had to leave the city and move to the Chiquimula valley because of the Spaniards. As mentioned in section 2.2.5.1.3, at the time of his visit there were practically no more people under fifty years old who still spoke Ch'orti' in Chiquimula and "the language is known by only an occasional 'anciano' (elderly person) there" (Hull 2016: 2). Aside from that, he reports speakers in Jocotán, Camotán, Olopa, Esquipulas and Quetzaltepeque. The data collected by Suárez were never officially published by himself but **Alberto Membreño** in **1897**, "a well-respected lawyer and judge in Honduras who served as president of the country for six months in 1916" (Hull 2016: 2), published them in the index of his *Hondureñismos: Vocabulario de los provincialismos de Honduras* (Membreño 1897). Comparison between the words that he gives and Suárez, however, reveals multiple differences. The original manuscript by Suárez held at the Princeton University Library[81] was not accessible to me. All my data on his manuscript come from Hull (2016), Membreño (1897) and the transcript of Suárez's wordlist by Korovina (2021). This, of course, makes Suárez a rather unreliable source that needs to be used with caution.

German geographer **Karl Theodor Sapper (especially 1890s)**, who traveled Mesoamerica extensively and recorded vocabulary from many languages as well as providing detailed language maps of the regions, also studied Ch'orti' among many other Mayan languages (Hull 2016: 2–3). His informants came from Jocotán, Camotán, Quetzaltepeque, El Obraje and Hacienda Grande.

Systematic fieldwork on Ch'orti' started in the **1930s** with **Charles Wisdom** who produced "the most important studies on the Ch'orti', both linguistically and ethnographically" (Hull 2016: 3). He began working in the conservative hamlet of Tunuco in 1931, then in and around Jocotán (1932) and Olopa (1933) and produced "a prolific amount of material, both published and unpublished" (Hull 2016: 3), which is all the more impressive considering he only spent about a year in the field. The sheer number of entries, many of which record words that have by now disappeared from the language[82], makes his (unfortunately unpublished) dictionary one of the most important linguistic sources on Ch'orti' (Hull 2016: 3). Since Wisdom was not a trained linguist, his carefully defined transcription system is, unfortunately, inconsistent

[81] https://catalog.princeton.edu/catalog/9939210893506421 (last accessed: 2024-02-11).

[82] "I have spent several months working separately with more than ten different elderly Ch'orti' consultants going through Wisdom's dictionary item by item and can confidently state that dozens of terms are utterly unknown today, increasing the value of Wisdom's work even more" (Hull 2016: 3).

in many ways. The transcription of the dictionary by Brian Stross (1992) "offered logical and useful suggested corrections to many of these errors, but there remain plenty of others" (Hull 2016: 3).[83] I would discourage the use of Stross's transcript.[84] The fact that preliminary drafts replace critical editions without clear statements as to the undertaken changes is, unfortunately, not unheard of in Mayan studies. The problem with this is that very obvious mistakes[85] can in especially unfortunate cases lead researchers to build whole theories on them (see footnote 185). Microfilm scans of Wisdom's dictionary are, thankfully, available (Wisdom 1950) and can be checked in unclear cases. For quick searches, I recommend the typescript by Korovina (2019) that she made based on Stross's with some added corrections, although it, too, still contains mistakes.[86] Marhenke (2015) offers a transcript of Wisdom's (1950) texts. However, for reliable use of the source, a critical edition of it is needed, which is why I only consulted Wisdom on particular questions regarding historical phonology but could not use it as a corpus.

Another important scholar is **Rafael Girard** (1949a) who wrote a five-volume series on the Ch'orti', of which the first volume contains a long list of words (Girard 1949b: 91–138). Unfortunately, this vocabulary is not simply the result of fieldwork done by Girard but stems

[83] For example, Hull reports that Wisdom "clearly invents noun roots where no such terms existed in the language" (Hull 2016: 3) extrapolating nominal forms from CVC roots that have been intransitivized through the passive infix *-j-*, e.g., <biht> 'load carried on the head, anything carried or supported'. Hull doubts the existence of this lexeme because "no such term is found in any other Ch'orti' source and is never used today" (Hull 2016: 3) – arguably a weak argument and somewhat curious considering his statement from footnote 82. He further doubts the data concerning "very specific nominal forms" (Hull 2016: 3), e.g., <bet> 'owe, an owing, debt', which, again, is unknown from other sources and also unknown today. A third group of "fake" nouns is supposedly derived by Wisdom by deleting the final thematic vowel of the verb "to create a false nominal root" (Hull 2016: 3), e.g. <bihn> 'thought, pondering, worry, preoccupation' from *b'ijnu* 'to think', "a term that nowhere else exists in any source and is not used by any speakers today" (Hull 2016: 3). Wichmann (1999: 9, 11) shares the view that Wisdom lists "fake" nouns because he "takes all verb roots to be nouns and invents nominal meanings for them, meanings that are sometimes so specific that it is hard not to be lured into believing that they actually do exist" (Wichmann 1999: 11). "This fact alone calls into question a portion of Wisdom's data and must be taken into account when referencing it." (Hull 2016: 3). However, one must be careful not to discard recorded data just because one cannot find it elsewhere, especially when documentation is as scarce as is the case with Ch'orti'. To sort this out, we need a critical edition of Wisdom's manuscript.

[84] Quizar (2020: 272) likewise notes that she would discourage this.

[85] E.g., Stross mistook a clear *i* for an *a*. This has also been pointed out in Marhenke (2020). See footnote 185.

[86] For example, Korovina (2019: 48) transcribes <war a ulok'oih tuuni'> for 'his nose is running' when Widom's (1950: 516) manuscript says <war a " tuuni'> with <"> indicating that the same as in the above examples is to be inserted here. The examples above that one unanimously show <alok'oih>. Therefore, no *u* is involved in the construction – a critical difference, as will be discussed in chapter 4.

from various sources, e.g., Alberto Membreño (1897)[87], among others, who, as stated above, in turn relied on the material from Ruano Suárez's (1892) manuscript. Therefore, the data in Girard are in part more than fifty years older than his original publication as well as not second- but third-hand data. Still, Girard's works also contain "a good number of original terms" (Hull 2016: 3) and Hull finds it surprising that he relied on other sources at all, considering that he had "significantly more experience living with the Ch'orti' than any previous compiler of Ch'orti' data". In terms of ethnography, Hull deems Girard's data "first-rate"[88] since he was able to confirm many of the things Girard reported during his own fieldwork. Still, even Hull agrees that some of Girard's claims, although not false per se, went too far, e.g., concerning his "relating the Ch'orti' to the Classic-period Maya, certain calendrical associations with the ancient Maya, and an insistence of the Popol Vuj being known among the Ch'orti'" (Hull 2016: 3).[89]

Extensive Ch'orti' fieldwork was conducted by **John Fought** in the **1960s** (1964, 1965, 1966, 1967) both in Guatemala[90] and with his chief consultant Isidro González in New Haven, Connecticut (Hull 2016: 3). Fought provided the first detailed description of Ch'orti' phonology, morphophonemics and morphology (1967) and the first compilation of Ch'orti' texts (1972), "a trove of linguistic and cultural data" (Hull 2016: 3). In 2006, he told Hull that he considered publishing a dictionary based on his work and might do so in the future but for the time being, his data are available on AILLA (Hull 2016: 3).[91]

In the 1990s, with the start of the Pan-Maya Movement, organizations arose that strove to better the status of Maya groups and/or document Mayan languages such as Majawil Q'ij, the Academia de Lenguas Mayas Guatemala (ALMG, the Academy of Maya Languages in Guatemala) and the Proyecto Lingüístico Francisco Marroquín (PLFM) (Hull 2016: 4). Projects on language documentation and revitalization started that yielded a number of publications:

[87] Hull states that Girard "borrowed heavily from Membreño (1897), not across the board, but rather in the majority of cases" (Hull 2016: 3).

[88] "I personally consider his data relating to the Ch'orti' to be of considerable breadth, incredibly detailed at times, and most certainly trustworthy in the vast majority of cases." (Hull 2016: 3).

[89] I found the following amusing anecdote in Dugan (2013: 167) who during his fieldwork among the Ch'orti' attended a religious ceremony that was held at a location reported to be "traditional" for such ceremonies. When Dugan inquired about how the Ch'orti' knew that the site had been used in this way earlier, the answer was that they had read about it "in Girard".

[90] In two villages of Jocotán (Pelillo Negro and Guareruche) in 1964/65 (Fought 1967: 1).

[91] https://ailla.lib.utexas.edu/collections/822/ (last accessed: 2025-03-13; requires registration).

- Vitalino Pérez Martínez (1994) produced the first ever grammar of Ch'orti' for the PLFM that was written by a native speaker. This was followed by a dictionary (Pérez Martínez et al. 1996) and a small collection of legends in Ch'orti' and Spanish (Pérez Martínez 1996). The dictionary specifically was "a milestone in Ch'orti' studies" (Hull 2016: 4). Though it suffers from inconsistent transcription, odd organization and incorrect grammatical analyses, it contains many items that do not appear in previous sources (Hull 2016: 4).

- The ALMG published a short Ch'orti' vocabulary in (2001) republished in 2007 which suffers from the same problems as Pérez Martínez et al. (1996) but to a greater extent (Hull 2016: 4).[92] Additionally, typographical errors appear on almost every page, neologisms are listed without being marked as such and the classification of terms regarding grammatical categories makes it inaccessible to readers who are not already proficient in Ch'orti' (Hull 2016: 4).

- Further educational materials are being produced by groups like ALMG, PLFM, CONALFA (Comité Nacional de Alfabetización). These are sometimes published locally and therefore under the radar of international scholars but they are valuable for documentation and revitalization of Ch'orti' in the community nonetheless (Hull 2016: 4).

Further sources by non-native speakers include:

- Helen Oakley's (1965) very short grammatical sketch of Ch'orti' with two texts.

- The first ever pedagogical grammar produced by evangelical missionaries John Lubeck and Diane Cowie (1989). This, unfortunately suffers from inconsistent orthography and notation though Lubeck's "command of Ch'orti' is impressive" (Hull 2016: 4) and thus his translation of the New Testament into Ch'orti' is also a valuable source on the language (Wycliffe Bible Translators 2012).

- A dictionary with over 1100 terms, sample sentences and grammatical classification has been compiled by Brent Metz, Teo Ramírez, Fred García and Mundo García in (1992). It was never published, "but one can only hope that it will be" (Hull 2016: 4).

- Søren Wichmann produced an unpublished morphological sketch of Ch'orti' (1999) which was meant to be revised at some point and all data "checked in the field"

[92] The orthographical inconsistency is due to various speakers writing different parts of the book.

(Wichmann 1999: 2) but this apparently never happened. Nevertheless, Hull states that he "would strongly recommend" (Hull 2016: 7) it and that many of his grammatical categories "correspond in full or in part" (Hull 2016: 10) to Wichmann's analyses. I found that Wichmann's sketch is only easy to use if you already have knowledge of Ch'orti'.

- Otto Schumann Gálvez's (2007) work states that it is a treatment of Ch'orti' verbs but the publication goes beyond that. Advantages include the fact that speakers from different ages provided data and that aside from Jocotán there is data from Camotán, where Ch'orti' has almost died out by now. Unfortunately, here, too, typographical errors abound and the grammatical analyses are not always correct (Hull 2016: 4). Hull reports that Schumann Gálvez also made a draft of a word list that so far has not been published.

- James Dugan (2013) wrote a grammar of Ch'orti' as his PhD thesis. He conducted fieldwork in Jocotán and its vicinity in 2002, 2004 and 2005 (Dugan 2013: 4–5). The grammatical analysis is sometimes questionable both on synchronic and diachronic grounds, as will be pointed out in the course of this dissertation. Still, it is one of the most accessible grammars for beginners and provides lots of data.

- Kerry Hull apparently plans to write a descriptive grammar of Ch'orti' in the future (Hull 2016: 7, 10).

Finally, we are fortunate enough to possess a modern trilingual dictionary in Ch'orti', Spanish and English, compiled by Kerry Hull during more than thirty months of fieldwork beginning in 1999 (Hull 2016: 5).[93] In 2005, he had published a shorter, preliminary version on FAMSI

[93] During the fieldwork he stayed mainly in Jocotán, but also in the hamlet of Oquén. He reports that he brought consultants to Jocotán more than half of the time and worked throughout the Ch'orti' area otherwise. He lists the localities of Suchiquer Abajo, Suchiquer Arriba, Pelillo Negro, Guareruche, San Juan Hermita, Tunucó, Quetzaltepeque, Pacrén, Amatillo, Agua Blanca, Canapará, La Mina, Las Flores, Oquén, Suchiquer, Tierra Blanca, Tunucó Abajo, Titicopote Abajo, and in the caseríos of Canapará Arriba, Quebrada Seca, Escobillal, La Quebrada, Pinalito, and Lajas" (Hull 2016: 5). His principal consultants were from Oquén, Las Flores and Guareruche, but overall, the dictionary is based on elicited data from over fifty people. The elicitation, often recorded, started with basic vocabulary and verbal paradigms and moved forward from that. Further recordings include oral traditions, of which he has collected about eighty to date. With improving fluency, he was able to conduct fieldwork in Ch'orti' and gathered additional data indirectly while his informants explained various topics in Ch'orti'. A special research interest of his is archaic language used in rituals and ceremonies, so he studied Ch'orti' ritual healers, *ajnirom*, and their healing chants. All the data he provides have been independently checked by himself and Ch'orti'

(Foundation for the Advancement of Mesoamerican Studies)[94], but the analyses in the more recent publication of 2016 supersede the ones from 2005 (Hull 2016: 10). The dictionary contains about 9.000 main entries and ample additional ethnographic information and scientific data concerning flora and fauna (Hull 2016: 11). More importantly, it includes sample sentences for most lexical entries since "the meaning of all words resides in their contextual usage, not in the terms themselves" (Hull 2016: 11). The sentences are all based on recordings of more than fifty different native speakers', either of natural speech, elicitations or oral histories. Very importantly, dialectal variation was not harmonized, though unfortunately, no dialectal information is given in the individual entries. He admits that some of the grammatical analyses made in the dictionary are provisional:

> There are, however, a sizeable number of verbal forms that have neither been discussed nor described in previous studies on Ch'orti'. This dictionary contains those data, sometimes with provisional classification pending further investigation. There is certainly much more to do. (Hull 2016: 10)

This last sentence sums it up quite well: despite all these valuable sources, "much more remains to be done" (Hull 2016: 4). An especially worthwhile endeavor as pointed out by Dugan would be fieldwork with female informants because he himself only worked with men and recorded many stories "recounted by men to other men, when women are not present" (Dugan 2013: 169) while the opposite is practically unrecorded. In general, data provided by women is underrepresented in all corpora.

Compared to other Mayan languages like Yucatec Maya or K'iche', little research has been conducted on Ch'orti' aside from the essentials like documentation and description. Notable theoretical linguistic literature includes works by Quizar on word order (1979; 1994b), split ergativity (Quizar & Knowles-Berry 1988), serial verbs (1994a), the antipassive (2020) and language contact to Xinkan (2023) as well as studies by Hull, e.g., on verbal art and poetic structures (2003) and numeral classifiers (2018). Marcos García (2015) studied the interaction of the index-sets with the stems they attach to. Of course, Ch'orti' is also sometimes mentioned in treatments of other languages, especially Hieroglyphic Maya.

---

informants several times. Unlike other scholars like Rafael Girard, he has not relied on other sources but his own data (see Hull (2016: 5) on all this).

[94] http://www.famsi.org/reports/03031/03031Hull01.pdf (last accessed: 2025-03-13).

2.2.6 Choltí

Since Choltí will become important in this dissertation, a more in-depth description of its documentation is necessary, too, in order to understand what kind of source we are dealing with. Choltí is an extinct Mayan language that is known to us from a colonial source often called the "Morán Manuscript" (MM) (Morán 1695). For a detailed analysis of its contents, genesis and probable authorship see Robertson, Law and Haertel (2010), the most recent edition of the manuscript, though it unfortunately does not contain a complete facsimile.

The MM contains two grammatical descriptions (usually called "Arte I" and "Arte II" respectively) of a language explicitly called "Choltí"[95] (literally "Chol language") as well as several Catholic texts and a word list in this language (Robertson, Law & Haertel 2010: xi). Though the people living in the area were called Choltí, Lacandón and Acalá and perceived each other as politically and culturally distinct, it is possible that they all spoke the same language (Robertson, Law & Haertel 2010: 8). Although we do not know exactly in which regions various parts of the MM (or its sources) were written, the most important site is the Manché province (today southeastern Petén) because the language descriptions were made during missionary work among the Manché Chol (Robertson, Law & Haertel 2010: 14, 18). Figure 6 provides an overview of the mentioned areas and groups.

[95] "Grammar in the Ch'olti' tongue, which means 'the language of cornfield workers' [Span. <Arte en lengua ɛholti que quiere decir lengua de milperos>]" (Robertson, Law & Haertel 2010: 9). The authors mention that the title was written as "language of cornfields [Span. <lengua de milpas>]" originally, not "de milperos". This is closer to the actual translation: "Chol-ti'" can literally be understood as "(Corn)field language". However, the original author specifically corrected this apparent mistake (Robertson, Law & Haertel 2010: 9).
The so-called "tresillo" ɛ is a sign used by the Spanish to transcribe unusual sounds, e.g. the uvular ejective *q'* in Mayan. However, in the Morán Manuscript the letter is sometimes even used for simple *k* or for *ch'* as in this case.

Figure 6. Chol provinces in the 16th century.[96]

Though the MM is usually simply dated to 1695, its history is a little more complex than that. Several people seem to have been involved in the genesis of the manuscript in some way or another since different hands can be identified in the manuscript, especially when it comes to later additions, which are often written in a different hand than the main text (Robertson, Law & Haertel 2010: 13–22). Among them, two are especially important. Robertson, Law and Haertel (2010: 15–19) argue convincingly that Joseph Ángel de Zenoyo was the main compiler of the surviving manuscript who copied the original "Arte" and provided a second version of it (Arte II) with corrections and comments – the original manuscript from which he copied does not survive. Though exact dates are unknown for most parts of the manuscript, one is dated to 1695 while other parts must have been written around 1685 when Ángel joined an expedition to the Manché region where "Choles" were (once again) supposed to be pacified (Robertson, Law & Haertel 2010: 18).

Although the original source did not survive, we know that Ángel did not write the entire manuscript himself but copied it either from the original written by Friar Francisco Morán or another copy of it. Frequent references to Morán make it clear that he is the original author of Arte I and large parts of the religious sections (Robertson, Law & Haertel 2010: 11).

[96] https://commons.wikimedia.org/wiki/File:Manche_Chol_16th_map.png (last accessed: 2025-03-03). © user Asdfjrjjj after Becquey (2012) and Jones (1998), 2023, Creative Commons Attribution-Share Alike 4.0 International License (https://creativecommons.org/licenses/by-sa/4.0/deed.en). No changes made.

However, Morán had been dead "for some twenty years" (Robertson, Law & Haertel 2010: 18) when Ángel joined the Manché expedition. Therefore, it is well justified that the actual linguistic information found in the manuscript receive an earlier date that coincides with Morán's life and work among the Choltí. Morán had been born in 1591 in northern Spain and came to Guatemala in 1618; he wrote extensively in Choltí and must have written the surviving sections between 1633 and 1656/57 (Robertson, Law & Haertel 2010: 14). The earliest layers of our source on Choltí therefore come from the middle, not the end, of the seventeenth century.

Other scholars have voiced concerns in the past as to how reliable the MM is as a linguistic source (Fought 1984; Fought 2000; Kaufman & Norman 1984), e.g.:

> Although the manuscript contains a wealth of linguistic data, the nature of the source often makes the interpretation of this data rather problematic: we have here a work compiled by non-native speakers of Cholti whose command of the language was imperfect in an imperfect copy (there are many spelling mistakes even in the Spanish text) made at different times and places and probably representing different varieties of Cholti. (Kaufman & Norman 1984: 97)

Although the "spelling mistakes" Kaufman and Norman refer to may perhaps be attributed to unstandardized Spanish orthography, some caution and interpretation is needed when using the MM as a source. As much as we would like to believe that the author(s) of this manuscript knew Choltí like native speakers, we simply cannot be sure. On the one hand the author(s) seem to have had enough command of the language to use poetic forms that are very typical of Mayan (Robertson, Law & Haertel 2010: xii), on the other hand glottalized consonants are practically not distinguished from unglottalized ones except for *k'* and *ch'* (Boot 2004: 2). However, inconsistent orthography need not be proof of insufficient language competence. Likewise, phonological competence can be viewed separately from, e.g., grammatical competence. From what is known about both Morán and Ángel, they seem to have had good enough command of Choltí to be taken seriously as speakers. Morán "is rumored to have written extensively in Ch'olti'" and "would have lived among Ch'olti' speakers for at least twelve years before writing most, if not all, of the material" (Robertson, Law & Haertel 2010: 14). Ángel, on the other hand, "felt that his knowledge of Ch'olti' gave him license to revise and improvise" (Robertson, Law & Haertel 2010: 19) what the revered Father Morán had written. Both also knew other Mayan languages. Ángel specifically was chair of languages in Kaqchikel and K'iche' at the University of Guatemala (Robertson, Law & Haertel 2010: 15).

Perhaps the best judgment of the MM lies somewhere in the middle: it is certainly a waste of good material to discredit the MM completely, but it should also be used with some caution. Even presuming that the authors' command of Choltí was excellent, we are still dealing

mostly with Christian texts, even in the examples in the grammar. That means that the manuscript does not necessarily contain natural Choltí – as a (both linguistic and cultural) translation it belongs to a special genre.

2.3 Methodology

This dissertation strives to contribute to Cholan language history as well as further our knowledge about how one kind of alignment can change into another. As Zúñiga (2018: 2) notes and as discussed in chapter 2.1.3, several mechanisms like reanalysis, extension and borrowing have been studied as the origin of alignment change. Unfortunately, the "three kinds of argumentative bases – attestation, reconstruction, and speculation – are unevenly distributed in the literature" (Zúñiga 2018: 2). Many studies do not show that a specific change occurred or must have occurred, but only argue "that it might have occurred, often without discussing strengths and weaknesses of their proposal, either in the light of empirical evidence or vis-à-vis alternative explanations" (Zúñiga 2018: 2), which makes it more difficult to assess their claims.

The goal of this dissertation is to provide a historical explanation for a peculiar grammatical phenomenon of Ch'orti' (set C and aspect-based split ergativity) using attestation, reconstruction and speculation in a balanced way. This includes 1) deconstructing the theories proposed by others (Robertson 1998; Wichmann 1999; Quizar 2023) and pointing out where they went wrong and 2) developing a new theory for the origin of set C using data from Hieroglyphic Maya as well as some observations on synchronic (as well as diachronic) morphophonological alternations in Ch'orti'. While doing that, I rely on the following theoretical building blocks:

1. **Historical-comparative linguistics**.
    a. I follow the Neogrammarians in assuming that sound change is fundamentally regular and deviations from certain rules can always be explained, provided one has access to sufficient data.
    b. I do not accept analogy based on superficial similarity as a magical tool to solve every problem that arises, i.e., I restrict my use of analogies to those cases where it is both necessary and likely that one may have happened and I always provide information on the specific context in which they may have happened.

c. For the historical development of Ch'orti', I consulted language data from the earliest available sources (Galindo 1834; Suárez 1892; Membreño 1897; Wisdom 1940) as well as publications on Hieroglyphic Maya under the commonly accepted premise that Ch'orti' is a descendant of Hieroglyphic Maya. As we will see, Cholti proves less useful in understanding Ch'orti' grammar than was hitherto argued.

d. Though the reconstructions of Proto-Cholan and Proto-Mayan available at the moment are preliminary, I will include them in my reasoning whenever necessary, e.g., because others have used the data for their argumentation. I try to provide lexical data when using reconstructions by others who have failed to do so.

2. **Typology**. I discuss parallel attestations of the changes I propose from other languages. Though this is not necessarily proof that the same must have happened in Ch'orti', when working with languages where historical data are limited[97], it is helpful to check whether a proposed change is even typologically plausible. Conversely, my study of Ch'orti' intends to contribute to our knowledge of attested changes in the realm of alignment change to possibly help explain peculiar formations in other languages, as well.

3. **Corpus study**. Though it was not possible for me to conduct my own fieldwork, a lot of data is available already that can be used to investigate historical questions. My analyses are based on the legends collected by Pérez Martínez (1996), which I annotated morphologically in FLEx.[98] I have also used the sample sentences in Hull (2016) extensively, supplemented by Fought (1967; 1972). The latter contain a wealth of data, which I could not analyze and annotate systematically but this should be done in the future. Ch'orti' data always come from Hull (2016) if not stated otherwise.

This way, I hope that my dissertation will not only be a useful contribution to the history of a specific Mayan language (Ch'orti'), but also to the theory of alignment change in general.

[97] Though a much earlier stage of Cholan is attested in the form of Hieroglyphic Maya, this does not necessarily help us in understanding Ch'orti' because the specific attestation of Ch'orti' (aside from the name) begins in 1834 and thus relatively late. Changes that happened between the decline of the use of the hieroglyphic script and the beginning of documentation of Ch'orti' can only be inferred but not observed.

[98] The original publication only included the Ch'orti' texts with Spanish versions of the text, which sometimes seem to be rather loosely translated.

3 Alignment in Mayan

In this chapter, I will introduce the various alignment and alignment split types that have been described for Mayan languages. I especially focus on a kind of split ergativity that is conditioned by aspect and is considered an areal feature of Lowland Mayan languages and on the difference between Mayan-type split ergativity and split-S/fluid-S marking.

3.1 Ergative-absolutive alignment

As briefly mentioned in section 2.2.2, in Mayan languages, verbal arguments are cross-referenced on the predicate via indexes; noun phrases or independent pronouns are optional (Zavala Maldonado 2017: 226). Almost all Mayan languages have two distinct sets of person markers to cross-reference the core arguments of the verb. This is exemplified via K'iche' in Table 3.

Table 3. Index-sets and pronouns of K'iche' (Can Pixabaj 2017: 467).

	Set A		Set B	Independent
	preconsonantal	prevocalic		
1SG	*nu- / in-*[99]	*w- / inw-*	*in-*	*in*
2SG[100]	*a-*	*aw-*	*at-*	*at*
3SG	*u-*	*r-*	*ø*[101]	*ri are' (ra're')*
1PL	*qa-*	*q-*	*oj-*	*(ri) oj*
2PL	*i-*	*iw-*	*ix-*	*(ri) ix*
3PL	*ki-*	*k-*	*ee-, e-*	*ri e are' (ri a're', ra're')*

The first index-set is traditionally called set A. It is used to index the possessor of a noun as seen in (1) as well as the agent of transitive verbs as in (2). It has a preconsonantal variant for

[99] Can Pixabaj (2017) does not include the variant *in-* but other sources, e.g., Larsen (1988: 213–215), mention a dialectal preconsonantal form *in-* corresponding to *inw-*.

[100] There is also an honorific *=laa* (singular, independent pronoun: *laal*) and *=alaq* (plural, independent pronoun also: *alaq*) (Can Pixabaj 2017: 467). These are enclitics, not prefixes. Since they behave differently and are not uniformly attested in other Mayan languages, they have a different origin and are a later addition to the paradigm. Therefore, they are not discussed here.

[101] As already stated in Abbreviations and orthography, I do not use the zero morpheme in my glossing. The absence of person indexes, which in Mayan regular ergative-absolutive alignment means that S or O is a third person for intransitive or transitive verbs respectively, is glossed as part of the verb stem, e.g.:

CHR *ixin*
go.B3
'he went'.

verbs that begin with a consonant and a prevocalic variant for verbs that begin with a vowel. The second set of indexes, set B[102], is used to cross-reference the subject of intransitive verbs as in (3) as well as the object of transitive verbs as seen in (2). Set B indexes also index the subject of non-verbal predicates like 'doctors' in (4) in the so-called "stative construction", and they form the basis for the independent pronouns (cf. Table 3).[103]

(1) Possession (López Ixcoy 1997: 66)

k-aj

A3PL-elote

'their *elote* (corn cob)'

(2) Transitive verb (Can Pixabaj 2017: 466)

x-oj-k-il-o

COM-B1PL-A3PL-see-TS[104]

'They saw us.'

(3) Intransitive verb (Can Pixabaj 2017: 466)

x-oj-b'iin-ik

COM-B1PL-walk-IS

'We walked.'

(4) Stative construction (Larsen 1988: 107)

oj *kunaneel*

B1PL doctor

'We are doctors.'

Set A is prefixed to the root in all Mayan languages, while set B is used as a prefix in some languages like K'iche', as a suffix in others, e.g., Yucatec Maya, or even as both depending

[102] The two pronoun sets are also often called "ergative pronouns" (set A) and "absolutive pronouns" (set B). This is avoided here because of the arising confusion when it comes to split patterns. This way, nonsensical statements like "Nominative-accusative alignment is achieved through the use of ergative pronouns on intransitive verbs" are avoided.

[103] The paradigm of independent pronouns is based on set B in most Mayan languages; notable exceptions are Chontal (Cholan) and Tojolabal (Greater Q'anjob'alan) where the paradigm is formed with set A (Schumann Gálvez 2007: 133). In K'iche', the only difference between set B as a prefix and as an independent pronoun is the third person. The reason is simple: The third person of set B is unmarked and thus difficult to use for emphasis – an important function of independent pronouns. The third person is also usually preceded by a determiner like *ri* (as in Table 3) or *le*. The determiner is optional for the other persons and its use depends on the dialect, though it is commonly used with the first and second person plural (Can Pixabaj 2017: 466–467).

[104] Mayan status suffixes provide a mixture of information on aspect, mood and transitivity.

on certain conditions (as in, e.g., Tsotsil). The alignment portrayed in examples (2) and (3) is ergative-absolutive since the S in (3) and O in (2) are marked in the same way (by set B) whereas A in (2) is marked differently (by set A). Aside from this morphological ergativity, some Mayan languages also show syntactic ergativity. In those languages it is, e.g., impossible to put A in front of the predicate for focus, interrogation or relativization without a special construction, while the same is possible for S and O (Aissen, England & Zavala Maldonado 2017a: 7).

An interesting phenomenon happens with A1SG in Mayan languages. As Table 3 shows, there are two variants of it. The first, *nu-* ~ *w-* is considered the original form and reconstructed for Proto-Mayan while *in-* ~ *inw-* is assumed to be based on B1SG *in-* and is only used on verbs, not on nouns in the possessive function of set A (Kaufman 2015: 67).[105] Robertson (1999) offers a survey of the phenomenon and attempts to explain it but the scenario that he offers is not plausible.[106] However, for now, it is only important to understand that this variation exists in A1SG and that the newer morpheme very likely comes from set B.[107]

The use of the index-sets described above applies only to Mayan languages like K'iche' that show consistent ergative-absolutive alignment that is not split based on tense, aspect, clause type or anything else. Many Mayan languages do not fall into this category. They show various split phenomena where the ergative-absolutive alignment only remains in parts of the grammar while other alignment types, mostly nominative-accusative alignment, develop in other parts. These different patterns emerge when the choice of index-set for the three syntactic functions S, A and O is altered (Zavala Maldonado 2017: 226–227), usually as a result of the conventionalization of a new construction. The next chapter will give a rough overview of the various alignment split patterns that are attested in Mayan languages today.

---

[105] According to Larsen (1988: 102–103), there are K'iche' dialects where *in-* ~ *inw-* is also used on nouns but only on a handful, which are mostly kinship terms, e.g., Nahualá K'iche' *in-taat* 'my father' instead of Momostenango K'iche' *nu-taat*. This is curious as typologically, one would rather expect these to retain archaic marking than to be subject to an innovation before other types of nouns.

[106] He suggests that the phenomenon spreads from areas that are less marked to those that are more marked and considers this a sufficient explanation: "The general idea that the several grammatical functions of a given grammatical morpheme stand in an ordered relationship to each other […] also has explanatory consequences for language change" (Robertson 1999: 449). The change seems to happen "by the power of markedness itself", as Fought (2000: 339) said about another of Robertson's theories. I cannot investigate this in detail here but intend to do so in the future.

[107] Since the first person of A is marked differently from the others, one could argue that this constitutes a case of NP-based split ergativity (see 2.1.2). Larsen (1988: 214–215) argues against this, though not very convincingly. If it could be shown that this is a case of split ergativity, that would mean that K'iche' (and all other Mayan languages) do not show consistent ergative-absolutive alignment.

3.2 Split patterns in Mayan languages

All Mayan languages have ergative-absolutive alignment in at least some part of their grammar. This is why the alignment is usually described as "split-ergative" when there is a split (for a discussion of whether the use of this label is justified see chapter 3.4). The other alignment is then often nominative-accusative where S is marked like A and differently from O; this is attested in Yucatecan, Cholan, Poqom, Q'anjob'alan and Mamean (Zavala Maldonado 2017: 227). These splits usually involve an extension of the use of set A "that marks the A in the unmarked conditions to S in the marked structural conditions" (Zavala Maldonado 2017: 232) – that is, in most cases, set A appears outside of its original role as an A marker and also comes to mark S. In one case, tripartite alignment is attested alongside ergative-absolutive, where S, A and O are all marked differently (Ch'orti') and there is also neutral alignment, where S, A and O are all marked in the same way (Mam).

The conditions for alignment splits (see chapter 2.1.2) range from clausal dependency (Q'anjob'alan, Mamean) to inherent features of the arguments (Mocho') or the predicate (Chol, Chontal, Poqomchi' and Mopan) but the most widespread condition is aspect, which affects the whole of Yucatecan, Cholan and Poqom (Zavala Maldonado 2017: 227). It will be the topic of the following chapter and is thus not discussed here.

Splits conditioned by clausal dependency are attested in Q'anjob'al, Popti', Akatek, Chuj (all Q'anjob'alan) and Awakatek (Mamean), all spoken in the Huehuetenango area (Zavala Maldonado 2017: 235). Such a split was shown in chapter 2.1.2 in examples (26)–(29). There, in dependent clauses, which are not marked for aspect[108], S is marked by set A yielding nominative-accusative alignment whereas ergative-absolutive alignment is used elsewhere; transitive verbs receive a dependent status marker (Zavala Maldonado 2017: 235). In some cases (Mamean languages Mam, Teko and Awakatek[109]), this split yields neutral instead of nominative-accusative alignment in dependent, aspectless clauses meaning that not only S but also O is marked in the same way as A (Zavala Maldonado 2017: 239).

[108] Verbs in dependent clauses receive person marking, but no aspect marker.

[109] Note that Awakatek appears in both groups and has an especially complex system. According to Zavala Maldonado (2017: 240), neutral alignment in Awakatek is restricted to "transitive verbs with preverbal directionals". Transitive verbs with postverbal directionals or no directionals do not mark their O with set A but with set B, as is the norm.

Mocho' is usually named as the only Mayan language where an alignment split is conditioned by lexical features of the verbal arguments, that is, due to the arguments ranking higher or lower on the animacy hierarchy (Zavala Maldonado 2017: 236). It was likewise already discussed in 2.1.2 in (16)–(19). Additionally, some languages show an active-stative split alignment where S is marked either like A or like O depending on lexical, not grammatical criteria (Chol, Chontal, Poqomchi', Mopan). This is discussed in 3.4.2. Finally, there is an inverse alignment split in Huastec and Cajolá Mam, where marking of A and O is determined by their relative rank concerning some dimension (person, animacy, topicality) (Zavala Maldonado 2017: 247). Kondić (2012: 161) describes a special set of portmanteau[110] person markers for South Eastern Huastec which are used in inverse alignment depending on what person acts on what other person and how they rank relatively to each other, cf. Table 4.

Table 4. South Eastern Huastec portmanteau person markers (Kondić 2012: 161).

Marker	S	O
tin	2SG/PL, 3SG/PL	1SG
ti	3SG/PL, 1PL	2SG
tu	2SG, 3SG/PL	1PL
	1SG	2SG
tixi	1SG	2PL
tuxu	2PL	1PL

For example, when a second or third person acts on a first-person singular, *tin* is used to mark both A and O in one form as in (5):

(5) Huastec inverse marking

tin	*chuuj*	*ti*	*we'eel*	*nuu'*	*ti*	*plaasa*
2>1	see.COM	PREP	yesterday	there	PREP	market

'You saw me yesterday at the market.'

Though not all details about the development in Huastec have been worked out yet, the fact that the relative position of the arguments to each other plays a role in the marking makes it attractive to analyze this as an inverse alignment split though Kondić (2012: 169) states that it is not a "typically inverse" language. For example, there is no overt inverse marker that indicates that the marking is reversed. Clearly, more research is needed on Huastec and the changes that it went through after separating from Proto-Mayan (not only in terms of

[110] Portmanteau meaning that they include two meanings, both A and O, in a single morpheme.

alignment). As stated in chapter 2.2.4, it likely separated from the main group as the first branch, after which it went through a series of radical changes.

3.3 Aspect-based split ergativity

The languages that show aspect-based split ergativity mostly belong to the Greater Lowland Mayan Linguistic Area. They are either spoken in the Lowlands (Yucatecan, Western Cholan) or have been in contact with these languages (Tseltalan, Q'anjob'alan, Poqom, Q'eqchi' and Ixil) (Law 2017b: 116). The following chapters will describe different manifestations of aspect-based split ergativity from various Mayan subbranches or single languages starting with the classic example: Yucatecan.

3.3.1 Yucatecan split ergativity

3.3.1.1 Synchronic description

Yucatec Maya was the first Mayan language for which split-ergativity was described and analyzed historically in detail, namely by Bricker (1981) (Polian 2017: 210).[111] Yucatecan languages show ergative-absolutive alignment in the completive aspect, but a nominative-accusative alignment in the incompletive aspect (Zavala Maldonado 2017: 232). Although the Yucatecan subbranch is considered to be the second branch to have split from Proto-Maya after Huastecan (Campbell 2017: 54), it likely only started to diversify into different languages relatively recently (Hofling 2017: 685).[112] As a result, Yucatec Maya, Lacandón, Itzaj and Mopan are still very similar, though not identical, as we will see. The following examples are from Yucatec Maya.

Examples (6) and (7) below show that in the completive aspect the alignment is ergative-absolutive, just like in K'iche' (cf. chapter 3.1) since S and O are marked by set B while set A is used to mark A. However, in the incompletive aspect the alignment is different because S in (9) is marked by set A and thus aligns with A in (8). The forms vary slightly (A2SG *a-* ~ *aw-*), but this is due to the use of prevocalic and preconsonantal allomorphs. The separate marking of O in (8) with set B – or, as the object is in the third person, the absence of a marker – shows that the alignment is now nominative-accusative. Note that the change of alignment

[111] Larsen and Norman (1979: 355–356) briefly address this issue, as well.

[112] About a thousand years ago according to Hofling (2017: 685) although he references Kaufman (1991) and therefore this assessment is likely based on glottochronology, in which case it would not be reliable.

happens because a different construction is now used on the intransitive verb where S is marked by set A, not set B.

(6) Yucatec Maya completive transitive (Hofling 2017: 710)

t-aw-il-aj-en[113]

COM-A2SG-see-TS-B1SG

'you saw me'

(7) Yucatec Maya completive intransitive (Hofling 2017: 711)

(j) jóok'-ech

(COM) go.out-B2SG

'you went out'

(8) Yucatec Maya incompletive transitive (Hofling 2017: 710)

k-aw-il-ik-en

INC-A2SG-see-TS-B1SG

'you see me'

(9) Yucatec Maya incompletive intransitive (Hofling 2017: 711)

k-a-jóok'-ol

INC-A2SG-go.out-IS

'you go out'

However, these examples conceal the richness of the Yucatecan aspectual system where a wide variety of markers is used to provide further details about an action, e.g., its aspectual contour. In fact, there is no agreement among scholars when it comes to the exact number of aspect markers present in the language (Yoshida 2016: 39–41). Additionally, it is not only difficult to say how many of these markers there are – there are also differences in terminology and classification. The markers are often subsumed under "TAM" (i.e., "tense/aspect/mood") markers to stress their variety of meanings. This way, it is also possible

[113] Lehmann (2020: 12, fn. 10) argues strongly against regarding set A as a prefix or proclitic because in Yucatec, it is possible to use adverbs between set A and the stem, e.g. in the following example (Lehmann 2020: 13):

ts'=u *hach* *ka'n-al-o'b'*

TERM=A3 very tired-INC-PL

'they are already very tired'

This is a valid and important point but since it seems to be an unresolved issue among Mayanists, I will, for now, retain the segmentation proposed in the respective sources. Based on the example provided by Lehmann, it seems that one could regard set A as clitic – it would then show ditropic clitic behavior (see also footnote 287.

to reference as a whole a phenomenon that is likely diachronically heterogeneous (see following section) without committing to specific analyses.

Hofling (2017: 708–709) distinguishes "aspectual, mood, and modality markers" (Table 5) that mostly consist of single phonemes prefixed to a verb – like *k-* for general incompletive in the examples above – from "aspectual adverbs" (Table 6 and Table 7) and "aspectual auxiliaries" (Table 8). Aspectual adverbs can express more refined aspectual information, sometimes with an admixture of modal meanings, e.g., YUC *táan* 'durative', *yaan* 'obligative; about to' or *óolak* 'almost'. Likewise, aspectual auxiliaries, that is, aspect markers that developed out of verbs (see section 3.3.1.2), provide a variety of meanings like Itzaj *jop'* 'inceptive' for an action that has just begun or *tz'o'ok* 'terminative' for an action that has been finished.

Table 5, Table 6, Table 7 and Table 8 provide a general overview of the variety of markers in the four Yucatecan languages Yucatec Maya, Lacandón (Southern and Northern dialect), Itzaj and Mopan. They illustrate the significant differences between the aspect markers, which are one area where the languages seem to differ. Mopan especially shows that even the prefix markers in Table 5 – presumably the oldest layer of this system because it is most fused with the verb – are not identical. If the languages indeed only started to diversify recently as proposed by Hofling, this would mean that the TAM marking system is relatively young.

Table 5. Yucatecan aspectual/mood/modality markers, slightly modified from Hofling (2017: 708).

Yucatec	**S Lacandón**	**N Lacandón**	**Itzaj**	**Mopan**	**Gloss**
k-	*k-*	*ø*	*k-*	*walak*	'incompletive'
t-	*t-*	*t-*	*t-*	*ø*	'completive'
(j)-	*ø*	*(j)-*	*ø*	*ø*	'intrans. completive'
ma't-	*ma'*	*ma'*	*ma' (ta'ax)*	*ma'(-ta'ach)*	'negative'
ka'(aj)	*ka'*	*k(aj)*	*ka'*	*ka'*	'optative, dependent'
				ma'ax-to	'not yet, dependent'

The new construction of intransitive verbs that leads to aspect-based split ergativity in (9) requires the so-called incompletive status.[114] This concerns the marker *k-* from Table 5 as well as the adverbs listed in Table 6 and those auxiliaries in Table 8 that are marked with INC.[115]

[114] As a reminder: Status suffixes provide a mixture of information on aspect, mood and transitivity.

[115] They are sometimes subsumed under "incompletive aspects", although the one "true" incompletive aspect is formed with *k-* (MOP *walak*) and all other formations offer additional nuances or even have

Table 6. Yucatecan aspectual adverbs that occur with incompletive status, slightly modified from Hofling (2017: 708).

Yucatec	S Lacandón	N Lacandón	Itzaj	Mopan	Gloss
suuk	*suuk*	*suk*	*suk*		‘customarily’
táan	*táan*	*tan*	*ta(a)n*	*tan*	‘durative’
táan-t…e’	*táan-t*		*tan-toj*	*tan-to(j)*	‘immediate past’
				tan=tun	‘immediate future’
chich	*chich*		*chich*	*chich*	‘hard’
séeb’	*séeb’*		*seeb’*	*seeb’*	‘celeritive’
k’abéet			*k’ab’éet*		‘necessitative’
taak	*taak*		*tak*		‘desiderative’
yaan	*yaan*	*yan*	*yaan*	*yan*	‘obligative’
ta’ay-tak					‘about to’
óolak			*olak*		‘almost’
ko’ox	*kux*	*ko’ox*			‘hortative’
je’el…-e’	*je’r…-e’*	*je’*	*je’le’…-ej*	*jed’e’ek*[116]	‘assurative’
páat			*pat*	*paatal*	‘abilitative’
	taab’ar				‘immediate future’
	b’ik				‘abilitative’
				tzaj	‘necessitative’

Verbs in Yucatecan languages have five different statuses: the incompletive and completive status, the dependent (or subjunctive) status, the imperative status and the perfect status (Hofling 2017: 709–710). A glance at Table 7 shows that Mopan especially features the dependent status more often than the incompletive status compared to the other Yucatecan languages. What is more, the markers in Mopan have cognates in the other languages that are used differently: for example, while *suuk* ‘customarily’ is used with the incompletive status in Yucatec Maya, both dialects of Lacandón and Itzaj (as displayed in Table 6), it requires the

meanings that are harder to reconcile with the notion of “incompletive”. For example, the ‘terminative’ in Table 8 describes a completed action, though one that is still relevant for the present. It is important to understand that the similarities in these synchronic formations are not due to specific incompletive semantics but due to all of them being based on a specific construction diachronically (see following chapter). It still makes sense to group them together, both for analysis and for language learning, no matter what kind of label one wants to use for this category, precisely because they originate in the same kind of construction. However, “incompletive aspects” might not be an adequate name.

[116] Mopan has a unique sound change where the sound sequence [ʔl] became the implosive [ɗ], written <d’>, e.g. ITZ *te’lo’* ‘there’ but MOP *ted’o’* ‘there’ (Hofling 2017: 691).

dependent status in Mopan – thus, the construction is not the same although the same lexical item is used.

Table 7. Yucatecan aspectual adverbs that occur with dependent status, slightly modified from Hofling (2017: 708).

Yucatec	S Lacandón	N Lacandón	Itzaj	Mopan	Gloss
				suk	'customarily'
				patal	'abilitative'
kV$_2$n					'definite future'
sáan, sáam				*sam-i*	'anterior past'
b'íin					'indefinite future'
			ko'ox	*ko'ox*	'hortative'
				(paatal)	'abilitative'
				(tan)	'durative'
				(tan-toj)	'immediate past'
				(tan-tun)	'immediate future'

Table 8 demonstrates that, though there is relative uniformity in the aspectual auxiliaries that are used in Yucatecan, there are again differences as to what status suffix they are used with, incompletive or dependent status. Therefore, the construction is not necessarily identical, even where the same auxiliaries are used. This is, again, especially true for Mopan.

Table 8. Yucatecan aspectual auxiliaries, slightly modified from Hofling (2017: 708).

Yucatec	S Lacandón[117]	Itzaj	Mopan	Gloss
jo'op' (INC)		*jop'* (INC)		'inceptive'
káaj (INC)		*kaj* (INC)	*kaj* (DS)	'inceptive'
		jo'm (INC)	*job'* (INC/ DS)	'inceptive'
ts'o'ok[118] (INC)	*ts'okar* (INC)	*tz'o'ok* (INC)		'terminative'
n-u ka'aj (INC)	*b'in u-ka' (ti)* (INC)	*b'el u-ka'aj ti* (INC)	*b'el u-ka'a ti* (INC)	future (intr.)
n-u ka'aj u- (INC)	*b'in u-ka' u-* (DS)	*b'el-u-ka'aj u-* (DS)	*b'el u-ka'a u-* (DS)	future (tr.)

In general, there is significant variation in the details of these constructions even beyond the status suffix. This cannot be discussed here, but an initial overview can be found in Hofling

[117] No separate forms given for Northern Lacandón.

[118] Note that the orthography is <tz, tz'> for languages spoken in Guatemala while languages in Mexico commonly use <ts, ts'> for the same sounds. This is why the verb in Itzá is written *tz'o'ok*, while the Yucatec Maya form is *ts'o'ok*.

(2017). This once more is evidence that the TAM marking system cannot go back to Proto-Yucatecan but likely developed after the languages had split from each other. However, since they remained in touch, contact played a role in the somewhat parallel innovations (Hofling 2017: 685). Compared to this, the system of numeral classifiers (Hofling 2017: 717), for instance, shows remarkable uniformity.

3.3.1.2 Diachronic explanation

The reason why it is not easy to determine how many TAM markers Yucatecan languages have is that the system is the result of, as Lehmann calls it, "a highly productive grammaticalization channel for auxiliaries", where synchronically, a "heterogeneous" paradigm "unites forms that have entered it at different times and have advanced to different degrees on the scale" (Lehmann 1993: 315). The heterogeneity already becomes obvious in the different constructions discussed in the previous chapter, e.g., the use of different status suffixes.

By now, it is well established[119] that aspect-based split ergativity in Mayan may result from "biclausal structures that included a matrix verb with a nominalized intransitive verb functioning as a complement" (Zavala Maldonado 2017: 234). In all cases, the status suffixes used in the respective constructions have been "analyzed synchronically or diachronically as nominalizers, suggesting that the subordinated verbs are to be treated as possessed nouns whose possessor expresses the subject of the embedded nominalized verb" (Zavala Maldonado 2017: 235).

Indeed, in Yucatec Maya, the suffix of the intransitive status that is used in all "incompletive" constructions that lead to nominative-accusative alignment (i.e., with the elements from Table 5 and Table 6 above as well as with auxiliaries) is the originally nominalizing suffix *-Vl* (Law 2017b: 118). With the combination of nominalizing suffix and set A index, which is also used to mark possession on nouns, the construction in (10) can historically be understood as '(it is) incomplete (i.e., not finished) your going out'. This is how the set A marker is able to enter the paradigm of intransitive verbs: it does so not in its function as a marker of transitive subjects but in its other function as marker of possession on nouns (Zavala Maldonado 2017: 235).

[119] Comrie (1978), Larsen and Norman (1979), Robertson (1980), Bricker (1990), Kaufman (1990) and Coon (2010a; 2012) are only some examples of publications discussing or mentioning this (Zavala Maldonado 2017: 234).

(10) Yucatec Maya incompletive intransitive (Hofling 2017: 711)

k-a-jóok'-ol

INC-A2SG-go.out-IS/NMLZ

'you go out'

This means that just as in the classic examples of split ergativity described in Trask (1979) and just like in Cariban languages (for both see chapter 2.1.3), in Yucatecan, too, "nouns, especially nominalized forms of verbs, have been recruited for renovating[120] the imperfective aspect"[121] (Polian 2017: 210). The difference to the scenario discussed in Trask (1979), however, is that we are not dealing with a form with "stative force". However, this is also not expected since, after all, Trask's explanation referred to the genesis of ergative-absolutive alignment in a language with otherwise nominative-accusative alignment. The orientation of the verbal adjective (used to refer to S and O but not A) yields ergative-absolutive alignment in Indo-Iranian, which leads to a split because the original alignment happens to be nominative-accusative in those languages. In Mayan, we see the exact opposite: The original alignment is ergative-absolutive and the new construction leads to nominative-accusative alignment. The form responsible for this cannot be a form with the same function as in Trask because such a form would not have caused a nominative-accusative alignment but again an ergative-absolutive one. Instead, the form in *-Vl* is best described as an action noun **possessed** by S. The fact that possession plays a role makes the construction "oriented" towards S and A (since A is likewise marked by the "possessive" index-set A) even though no participle is involved.

We find confirmation for the idea that the incompletive aspect in Yucatecan is indeed based on biclausal structures with a matrix verb and an associated complement in the form of a

[120] The term "renovating" implies that there used to be a different incompletive construction that was replaced by the one we see in Yucatecan languages today. An aspectual distinction of completive-incompletive is generally reconstructed for Proto-Mayan (Kaufman 2015: 211). See, however, the introduction in chapter 7 on this.

[121] Imperfective and perfective on the one hand as well as incompletive and completive on the other are terminological pairs used by different authors to refer to the same thing in Mayan (Polian 2017: 210). According to Dayley (1982: 47), although they are practically synonymous, the labels "incompletive" and "completive" are preferable for Mayan since these languages have first of all an aspectual distinction between perfect and non-perfect and then, in the non-perfect category, there are aspects like completive and incompletive etc. "The terms 'completive/incompletive' should be reserved for two aspects typically found in the nonperfect systems in Mayan languages, and the terms 'perfective/imperfective' should be avoided altogether, so that no confusion arises between them and 'perfect/nonperfect'" (Dayley 1982: 47). Though no such confusion seems to exist in other language families, e.g., Indo-Iranian, which also feature "perfective" aspect alongside perfect, the terminological pair "completive" – "incompletive" is well-established in Mayan linguistics and therefore also used here.

nominalized intransitive verb in some of the aspect markers, which are attested as fully inflected verbs in Colonial Yucatec (Bricker 1981: 85).[122] I will demonstrate this for the terminative auxiliary YUC *ts'o'ok*. Example (11) shows that *ts'o'ok* used to be inflected with an aspect marker, a person prefix to mark S and a status suffix/nominalizer – just like any other intransitive verb in the language would.

(11) Colonial Yucatec (Bricker 1981: 85)

<cuɔocol cmeyah trapich>

k-u-ts'o'ok-ol	*k-meyah*	*trapich*
INC-A3-finish-IS/NMLZ	A1PL-work	trapich

'we finish working the grinding machine'

Today, the respective verbs are rarely used as main verbs of complement constructions – they have become auxiliaries (Bricker 1981: 85).

(12) Modern Yucatec (Bricker 1981: 85)

Ts'o'ok	*a*	*han-al?*
TERM	A2SG	eat-IS/NMLZ

'Have you finished eating?'

Note that the verb form that has been grammaticalized as the marker of terminative aspect is the completive one in the third person, *ts'o'ok*, without any affixes, not the inflected one presented in (11). This makes sense because the terminative semantics come into play when something is already finished – not when it is about to or will be finished in the future. In theory, since B3 is unmarked, *ts'o'ok* could also be analyzed as the mere root and not necessarily the verb in the third person. However, if the question in (12) were to be answered, the affirmative answer would be as in (13):

(13) Affirmative answer with *ts'o'ok* (Bricker 1981: 85)

ts'o'ok-ih.
TERM-B3SG

'He has.

[122] Some authors, including Bricker (1981), use "Classical Yucatec" to refer to the colonial variety of Yucatec. I find this misleading because of the parallel use of "Classic" for the Mesoamerican archaeological period from 250 CE – 909 CE (Martin & Grube 2008: 8). What Bricker discusses is not the Yucatec of the (archaeological) Classic period but that which is attested in colonial sources. In newer publications (e.g., Bricker 2019), she has abandoned the term following the recommendation of Smailus (1989: 1).

The suffix *-ih* "is a phonological terminal marker that occurs with intransitive verbs in the completive aspect if the subject is third person singular (ø) and nothing follows the verb in the phrase" (Bricker 1981: 85). Though it remains unclear how this distribution of the marker came about, the fact that it exists is without doubt. Its use in the affirmative answer proves that the verb in (12) is indeed a completive B3 one.

Still, a question presents itself. Example (11) shows that before *ts'o'ok* was grammaticalized as a terminative marker, it was itself already used in an incompletive aspect that showed split ergativity because it itself receives the suffix *-Vl* and is marked with incompletive aspect marker *k-*. The TAM markers that are transparently derived from verbs apparently developed at a time where the split already existed. Therefore, they cannot be used to explain how it arose; on the other hand, the origin of other types of markers remains more insecure.

Bricker supposes that all constructions involving aspect markers might be explainable through "complementation", i.e., that a verbal origin can be assumed for all of the markers. For example, she views *k-* as the remains of the verb <lic> 'to be accustomed to' because it still appears "in expanded form" (Bricker 1981: 86) in Colonial Yucatec. This would mean that the clitic particle *k-* developed out of an earlier auxiliary *lik.*[123]

> I suspect that the other aspects that govern nominative-accusative case marking[124] can also be analyzed as complement constructions, although I do not yet have enough data to support such an interpretation. Nevertheless, the fact that four aspect words seem to be verbs rather than particles or clitics suggests that the ergative split may result from complementation. (Bricker 1981: 86)

As far as is known to me, these claims have not been further investigated up to this point. This is a worthwhile task for future research because either a verbal origin must be demonstrated for all markers or the mechanics of adverbs taking nominal complements explained in another

[123] The implied sound change, a reduction of *lik* to *k* in the course of grammaticalization, may seem drastic if one believes in the regularity of sound change. It would be preferable not to resort to ad hoc explanations based on the belief that "anything goes" in grammaticalization when it comes to reduction of the involved forms and that drastic shortening therefore does not require further clarification. Nevertheless, a similar shortening is attested for other aspectual markers "at least since the 1960s" (Bricker 2019: 97), so that durative *táan* becomes *t* (e.g., *t in [< táan in] chan wenel* 'I'm having a snooze') and *ts'ó'ok* becomes *ts'* (e.g., *ts' u [< ts'ó'ok u] máan àanyos b'eyo* 'the years finished passing like that'). Lehmann (1993: 315) accepts this development as part of the – in his eyes – expected process of phonological erosion during grammaticalization.
However, the two reductions are not identical because in the latter case, only the first sound of the marker remains while in the case of *lik*, one would have to assume that it is the final sound that is grammaticalized. For this, I would like to point out the possibility that we are dealing with a separate aspect marker in Colonial Yucatec – one that did not survive into modern times. The fact that its semantics are similar to those of the prefix *k-* may even have played a role in its non-survival. A new corpus study of the use of the two markers from earliest attestation onwards might bring further clarification.

[124] Older sources refer to case marking in Mayan, though this is generally not done anymore.

way. In the latter case, this would mean that the "incompletive" aspects are the result of convergent evolution of not entirely identical constructions (as hinted at by Lehmann, see page 75).

Further issues with this theory that were already addressed in Bricker (1981) are the following:

- It seems that motion verbs in Classical Yucatec used the gerundial suffix *-el* when used as complements, while other verbs took the suffix *-Vl*. Bricker (1981: 87) sees this as proof of the suffix slowly harmonizing its vowel with the respective verbal roots. However, verbs like *tàal* 'to come', *b'in* 'to go', *máan* 'to pass (by)' and *u'ul* 'to arrive' usually do not take any suffix at all in their function as complements.[125] I suspect that the suffixes are of different origins and therefore the constructions might also be different.

- As Bricker points out, nominalization is a plausible explanation for intransitive complements but "not immediately obvious in the case of transitive complements" (Bricker 1981: 87). The transitive construction does not use the suffix *-Vl* but rather *-ik*. Bricker (1981: 87–89) tries to show that *-ik*, too, conceals a nominalizing suffix. However, this theory is not entirely convincing because it, too, requires spontaneous and unsystematic sound change. Perhaps the solution is simpler: the new construction requires set A marking, which is already present on transitive verbs. This may be the reason why the transitive construction did not change. More research is necessary to understand the origins of the Yucatecan status suffixes and also how transitive verbs are viewed by speakers synchronically. There may also be a different way to derive the transitive verb from a nominalized form.

- It seems that there are other forces at play when it comes to the development of aspect-based split ergativity, so that aspect-based split ergativity is not "automatically triggered by nominalization" (Bricker 1981: 89). These "forces" shall remain mysterious for now but will be addressed in chapter 7.4.4.

Remaining questions aside, Bricker's (1981; 2019) work has convincingly shown that Yucatecan split ergativity historically involves a change in syntactic class. The fact that the

[125] "Informants tell me that -el (but not -Vl) can be suffixed to these verbs (particularly when they are inflected for durative aspect with táan), but that the preferred form is without any suffix at all." (Bricker 1981: 87).

underlying construction used a possessed nominalized verb form is uncontroversial because this is witnessed by the suffix *-Vl* of which we know that it is a nominalizer. What remains less secure and needs to be investigated in the future is the reason for this nominalization.

In the following chapter, we will contrast the observations made for Yucatecan with split ergativity in Chol and Chontal.

3.3.2 Western Cholan split ergativity

The most extensive work on Western Cholan split ergativity has been done by Coon (e.g., 2008; 2010a; 2010b; 2012; 2013) on Chol. Therefore, I will also use Chol to illustrate it, even though there are a lot of differences between Chol and Chontal that would be worthy of discussion.[126]

Just as in Yucatec, ergative-absolutive alignment occurs in the completive aspect where O in (14) is marked by set B just like S in (15), while set A is used for A in (14). On the other hand, the incompletive aspect features the same marking of A in (16) and S in (17), for which set A is now used, while set B is reserved for O alone.

(14) Chol completive transitive (Coon 2008: 99)

Tyi *i-mek'-e-yoñ*[127].

COM A3SG-hug-TS-B1SG

'She hugged me.'

(15) Chol completive intransitive (Coon 2008: 99)

Tyi *wäy-i-yoñ.*

COM sleep-IS-B1SG

'I slept.'

(16) Chol incompletive transitive (Coon 2010a: 216)

Mi *a-mek'-oñ.*

INC A2SG-hug-B1SG

'You hug me.'

[126] For Chontal see, e.g., Osorio May (2016) and Vinogradov (2016).

[127] The intervocalically occurring *-y-* is segmented and treated as a glide by some authors (e.g., Vázquez Álvarez 2011) and as part of an allomorph B1SG *-yoñ ~ -oñ* by others like Coon. I use the latter strategy for all examples but this is not to be taken as a statement of theoretical analysis, just a convenient regularization.

(17) Chol incompletive intransitive (Coon 2010a: 216)

Mi *a-wäy-el.*

INC A2SG-sleep-IS/NMLZ

'You sleep.'

Precisely as in Yucatec, the new construction used for intransitive verbs in the incompletive aspect in (17) features a nominalizing suffix. However, note that the suffix is *-el* (Coon 2017: 657), not *-Vl*. This distinction sometimes goes unnoticed in the literature but it is important. Based on the similarities to Yucatec Maya, Coon (2008: 99) proposes the same analysis for Chol, namely that the nominalized forms are non-verbal complements while the aspect markers are the actual verbs that take these complements. The difference between her and Bricker's explanation is that Coon regards the nominalizations not simply as the diachronic sources of the constructions that lead to the new incompletive aspect. Instead, she proposes to analyze the incompletive stems with Set A indexes as seen in (17) as "formally possessed nominal arguments of a one-place predicate", namely the aspect marker *mi* with which the stems are obligatorily used. Thus, she implies an underlyingly nominal structure for this construction even synchronically. In her own words: "The fact that nonperfective aspect markers are verbs which must embed nominalized clauses gives rise to the apparent split." (Coon 2010b: 92)

To Coon (2008: 99) this means that the resulting nominative-accusative alignment and the split are just an "illusion" and that in fact "*all predicates* in Ch'ol show an ergative-absolutive pattern". What makes this illusion possible is the Pan-Mayan syncretism of possession and agent marking through set A, so that "ERGATIVE = GENITIVE", as Coon (2008: 99) states it. In examples (16) and (17), the actual predicate is *mi*, not *mek'* or *wäy*, and an additional third-person set B index is assumed (which, as we remember, is zero-marked) yielding an actual translation "Your hugging me occurs." and "Your sleeping occurs." (Coon 2010a: 215–216). *Mek'* and *wäy*, on the other hand, are then just nominalized complements to the actual verb. As B3 is unmarked, it is difficult to prove whether this analysis holds true synchronically or whether it is just the diachronic pathway that yielded an aspect "particle" that was originally a B3 form.

This view leads to some problems. For example, it is necessary to assume that the form of transitive verb 'to hug' in (16) is also nominalized, even though it does not have an overt nominalizer. Although Coon claims that at least in Chol "we find evidence for the nominality of transitives as well, despite the lack of any overt nominalizing morphology" (Coon 2010b:

230), her arguments supposed to support this are not convincing. She proposes to analyze sentences like (18) as compounds similarly to the English translation that she gives for the example. However, this still does not explain the absence of a nominalizer, especially since conversion generally does not occur in Mayan (Aissen, England & Zavala Maldonado 2017a: 6), and also this case only applies to transitive verbs with objects marked by an overt NP.

(18) Chol transitive sentences as compounds (Coon 2010a: 230)

Tyi[128] *majl-i* [*tyi* *kuch* *si'*] / *[tyi wäy-el].*

COM go-IS PREP carry wood PREP sleep-IS/NMLZ

'She went to wood-carry/sleep'.

This specific proposal of transitive verbs being underlyingly nominalized without overt suffix is rejected in Zavala Maldonado (2017: 235) who points out that transitive forms do not show the same properties that make an analysis of intransitive verbs as nominalizations convincing: They do not combine with determiners and adjectives and do not function as complements of a preposition. This is a convincing assessment.

The general view advocated by Coon, namely that there actually is no nominative-accusative alignment in Chol, may ultimately depend on one's point of view. Sometimes, "deep structures" as opposed to "surface phenomena" may in my view be better understood as diachronic layers. Coon's (2008: 102) conclusion that "[t]he true predicate is the aspectual marker *mi* (or its allomorph *muk'*)" could be tested, e.g., by examining whether the aspect markers still occur as full verbs, not only as auxiliaries, in Chol today. If not, then her statement may only be true concerning the diachronic source of the construction. Of course, this would only apply to aspect markers that derive from verbs, which has not yet been demonstrated for all of those that co-occur with nominative-accusative alignment, just as in Yucatec Maya. Vázquez Álvarez (2011: 193) argues that the aspect markers clearly "come from verbs", since they can receive person marking, but that in some contexts, when they do not receive person marking, they are auxiliaries. It is difficult to prove that an aspect marker indeed has "no marking", since B3 is unmarked. Furthermore, set B is used in Mayan to form non-verbal predicates (in the stative construction). Therefore, its presence does not yet prove that something is in fact a verb. The most conclusive proof would be a solid etymology.

[128] Note that the completive marker *tyi* is not the same as the preposition *tyi* – at least based on Coon's glossing. To decide whether the completive marker is historically the same as the preposition, more historical-comparative research is required.

Whether they are verbs or not, almost all of the aspect markers have developed allomorphs – supposedly due to "phonological reductions or morphological constraints" (Vázquez Álvarez 2011: 193). These are unique to Chol, because no other Cholan-Tseltalan language, not even the closely related Chontal, shows a similar variation (Vázquez Álvarez 2011: 70). In the completive aspect, the bare marker, which can be seen in (19), is *tyi*, but the form changes[129] to *tsa', ta', ts* or *t*[130] when clitics are added (Vázquez Álvarez 2011: 194) as in (20). A use of *tyi* with *=ku* would be ungrammatical (Vázquez Álvarez 2011: 70).

(19) Completive without additional material on auxiliary (Vázquez Álvarez 2011: 70)
tyi *k-il-ä-yety*
COM A1-see-TS-B2
'I saw you'

(20) Completive with additional material on auxiliary (Vázquez Álvarez 2011: 70)
tsa'=ku *k-il-ä-yety*
COM=AFF A1-see-TS-B2
'yes, I saw you'

The same applies to the incompletive aspect marker, which is *mi* in its bare form as in (21), but changes to *muk', mu'* or *mu*[131] when followed by clitics (22) or person markers (23) (Vázquez Álvarez 2011: 196). The latter cannot be used with *tyi* and its allomorphs (Vázquez Álvarez 2011: 194, fn. 5). Again, a hypothetical use of *mi* with set B as in **mi=oñ tyi toñ-el* is impossible (Vázquez Álvarez 2011: 71).

(21) Incompletive without additional material on auxiliary (Vázquez Álvarez 2011: 71)
mi *k-il-añ-ety*
INC A1-see-TS-B2
'I see you'

[129] This development does not even apply to all dialects of Chol, only to Tila Chol and not to Tumbalá Ch'ol, which only uses *tsa'* or sometimes *ta'* (Vázquez Álvarez 2011: 194).
[130] The allomorphs *ts* and *t* "occur preferably where the next element is a vowel" (Vázquez Álvarez 2011: 194).
[131] *Muk'* is used preferably before vowels, *mu'* before vowel or consonant and *mu* when the first vowel of the clitic is dropped as in *mu'=äch* (INC=AFFR) as opposed to *mu=ch* (INC=AFFR) (Vázquez Álvarez 2011: 196–197).

(22) Incompletive with *=ku* on auxiliary (Vázquez Álvarez 2011: 71)

mu'=ku *k-il-añ*

INC=AFF A1-see.B3-TS

'yes, I see it'

(23) Incompletive with person markers (Vázquez Álvarez 2011: 71)

muk'-oñ *tyi* *toñ-el*

INC-B1 PREP work-IS/NMLZ

'I work'

Further research into the distribution of the aspect markers and their allomorphs as well as their etymology will likely yield more insights into the source of the construction that led to aspect-based split ergativity in Western Cholan. It is theoretically possible that all aspect markers are verbs historically – but it is likewise possible that none of them are verbs or that some are, while others are adverbs. As it stands, I would argue that what can be observed in Chol synchronically is simply a nominative-accusative alignment in the incompletive (and progressive) aspect. The forms of the incompletive aspect would not be "formally *possessed nominals*" (Coon 2008: 103) but **formerly** possessed nominals.

It is generally assumed that Western Cholan "borrowed"[132] split ergativity from Yucatecan languages, with which it has been in contact, because, it is argued, the constructions are similar and Western Cholan even has the same nominalizing suffix *-Vl* on intransitive incompletive verbs (Law 2017b: 118). This is partly based on the fact that the split is reconstructed for Proto-Yucatecan, while this is not the case for Cholan-Tseltalan as a whole (Zavala Maldonado 2017: 233). It is considered more likely that the split spread from the branch where it is attested for all languages. In theory, this is reasonable. However, as we have seen, the Yucatecan languages show a considerable variation when it comes to the details of the TAM markers that are involved in the split, so that its reconstruction for Proto-Yucatecan needs to be reinvestigated. Furthermore, the most important difference between Yucatecan and Chol is the fact that the suffixes are not fully identical. As mentioned above, the constructions in Chol operate with *-el* exclusively, not *-Vl*, while *-el* is only used for motion verbs in Yucatec. Only further research into the diachrony of the respective branches will reveal how similar the origin constructions actually are.

[132] As noted in chapter 2.1.3, alignment per se is not something that is borrowed. Constructions may be borrowed and consequently lead to similar alignment change phenomena.

3.3.3 Poqom split ergativity

According to Zavala Maldonado (2017: 232, 234), both languages of the Poqom (Greater K'iche'an) subbranch, Poqomam and Poqomchi', show aspect-based split ergativity in the potential and progressive aspects. This is a key difference to Yucatecan and Western Cholan as the incompletive aspect is therefore unaffected in Poqom. Instead, the incompletive aspect, e.g., of Poqomam shows ergative-absolutive alignment because S in (24) and O in (25) are marked in the same way (set B) while A is marked differently.

(24) Poqomam incompletive intransitive (Santos Nicolás & Benito Pérez 1998: 183)

ti-tin-a

B2SG-bathe-IS

'You bathe.'

(25) Poqomam incompletive transitive (Santos Nicolás & Benito Pérez 1998: 183)

ti-ru-k'ul

B2SG-A3SG-meet

'He meets you.'

On the other hand, in the potential aspect, S in (26) and A in (27) are marked alike (by set A), while set B is used for O. This pattern corresponds to nominative-accusative alignment in Yucatecan and Western Cholan even though the elements employed in the constructions are not cognate.

(26) Poqomam potential intransitive (Santos Nicolás & Benito Pérez 1998: 183)

n-a-wur-a

POT-A2SG-sleep-IS

'You will sleep.'

(27) Poqomam potential transitive (José Francisco Santos Nicolás in Zavala Maldonado 2017: 234)

k-in-a-to'-om

INC-B1SG-A2SG-help-POT.TS

'You will help me.'

It is often pointed out that Poqom employs a functionally equivalent, but not cognate suffix in the respective construction, namely *-ik* (Law 2017b: 118). In the case of the Poqomam

potential, it is neither *-ik* nor *-Vl/-el* but something else entirely, as example (26) demonstrates. The mention of *-ik* is interesting, however, as that is the suffix used to inflect transitive verbs in the incompletive aspect of Yucatecan as shown in (8). It could be investigated whether Poqom can help us understand whether transitive verbs in the incompletive are also nominalized in Yucatec Maya. Yucatecan *-ik* and Poqomchi' *-ik* could be cognates based purely on what we know about Mayan sound correspondences (Campbell 1984: 6). All in all, there has been a lot less research on Poqomam aspect-based split ergativity than on that of Yucatecan or Western Cholan. I am especially not aware of any studies from a diachronic perspective. A study seems worthwhile if we want to understand whether split ergativity in fact evolved via contact as proposed in the literature and how similar the mechanisms are that gave rise to it.

3.3.4 Discussion

The investigation of aspect-based split ergativity in Yucatecan, Western Cholan and Poqom has shown that the constructions are less uniform than they may seem based on their portrayal in publications that emphasize the fact that they all lead to the same type of alignment split. Although the branches could only be exemplified by individual languages here, a careful study of all individual languages will likely yield even more differences. However, it seems to be secure that diachronically, we are dealing with constructions in which a possessed nominalized verb form is used as a complement to something else. In some cases, it is possible to demonstrate that the nominalized forms are complements to verbs while in others, the etymology of the aspect markers is less clear.

It is important to keep in mind, as I pointed out in my discussion of Yucatec, that these are seemingly parallel constructions with nevertheless some significant formal and especially functional differences. To me, this suggests that they arose independently, even though the tendency to form these types of constructions may go back to areal effects. More importantly, I would suggest that it is not helpful for our understanding of their history to subsume them under terms like "incompletive aspects" or emphasize their parallels to the point where significant differences are glossed over.

3.4 Active-stative alignment

3.4.1 Split ergativity or "split intransitivity"?

In chapter 2.1.2, "split ergativity" that is conditioned by tense or aspect was exemplified with Indo-Iranian languages. If we contrast this with Mayan, some differences become apparent. I give again the examples from Katë. Katë usually shows nominative-accusative alignment where S and A are marked by direct case and thus in the same way, while the oblique case is used for O as illustrated in the optative forms of (28) and (29) that are relics of the old present.

(28) Katë optative intransitive (old present) (Grjunberg 1980: 238)

(vúze)	*nargé-m*
1SG.DIR	run-OPT.1SG

'I shall run'

(29) Katë optative transitive (old present) (Grjunberg 1980: 129)

(vúze)	*suv*	*paṣyé (< paṣí-e)*	*yu-m.*
1SG.DIR	all	mountain.slope-OBL.M	eat-OPT.1SG

'I shall eat the whole mountain'

In the so-called perfective past of Katë (Grjunberg 1980: 218–219; Halfmann 2024: 533–534), on the other hand, the alignment is ergative-absolutive because S in (30) is marked by direct case just like the O of (31), while A is now marked by the oblique case.

(30) Katë perfective past intransitive (Mohammad 1991: 12)

Asád	*á-y-o*
Asad.DIR	come-PFV-PST.3SG.M

'Asad came'

(31) Katë perfective past transitive (Mohammad 1991: 12)

Yéme	*Asád*	*vř-y-o*
1SG.OBL	Asad.DIR	see-PFV-PST.3SG.M

'I saw Asad'

If we compare this to Mayan, several differences become apparent. In the Yucatec examples the marking of the intransitive verb in the innovative incompletive aspect form in (35) changes compared to the completive aspect while it remains the same in the transitive verb in (34). For Katë, it is the other way around. Nothing changes in the marking of the intransitive

verb in the innovative perfective past form in (30), while the marking of the transitive arguments A and O in (31) is reversed.

(32) Yucatec Maya completive transitive (Hofling 2017: 710)

t-aw-il-aj-en

COM-A2SG-see-TS-B1SG

'you saw me'

(33) Yucatec Maya completive intransitive (Hofling 2017: 711)

(j) jóok'-ech

(COM) go.out-B2SG

'you went out'

(34) Yucatec Maya incompletive transitive (Hofling 2017: 710)

k-aw-il-ik-en

INC-A2SG-see-TS-B1SG

'you see me'

(35) Yucatec Maya incompletive intransitive (Hofling 2017: 711)

k-a-jóok'-ol

INC-A2SG-go.out-IS/NMLZ

'you go out'

In Mayan aspect-based split ergativity, the innovation concerns intransitive verbs, while it concerns transitive ones in Indo-Iranian. Therefore, in theory, it is debatable whether it is really the same – or even just a similar – phenomenon that we are dealing with in Maya. This kind of doubt is expressed in Bohnemeyer (2004) who argued that Yucatec alignment is better termed "split-intransitive" precisely because changes that lead to the split affect intransitive verbs and not transitive ones like in Indo-Iranian[133] and proposed an analysis which

[133] "Yukatek and some other Mayan languages (as mentioned in the introduction) are the only languages in evidence, to the best of my knowledge, with split argument marking, controlled exclusively by aspect marking, occuring in all and only in intransitive clauses. There has been much controversy around this phenomenon. Most Mayanists (e.g. Bricker 1981; Hofling 2000a; Robertson 1992) consider the pattern a straightforward case of split ergativity, ignoring the fact that the split occurs only with intransitive verbs. In contrast, Kaufman (1990) stresses the restriction to intransitive verbs. He terms the pattern 'mixed-ergative.' Straight (1976) is the first to apply Sapir's (1917) 'active-inactive' to Yukatek. DeLancey (1985) points out that Yukatek indeed falls under Sapir's definition, but with aspect marking as the conditioning factor, not lexical class as in the example Sapir considers (Dakota) […]. Outside Mayan, only cases of more indirect correlations between aspect marking and split intransitivity are known. Thus, in Iroquoian, a split which is otherwise conditioned lexically or by construal is neutralized in certain aspect-mood forms

emphasized agentivity over aspect. However, he analyzes aspectual morphology together with status suffixes while I believe that the two are independent morphology best kept separate.[134]

I disagree with Bohnemeyer's proposal for several reasons. The first reason is that the Mayan phenomenon still has more in common with Indo-Iranian split ergativity than with what is canonically described by active-stative. As explained in chapter 2.1.2, an active-stative alignment split is when a language inflects intransitive verbs according to two classes: the S_A class (where S is marked like A) refers to an "activity" or an action with the attribute +control or +volition while the S_O class (where S is marked like O) refers to a "state or condition" or –control/ –volition (Dixon 1994: 71).

When the class membership is fixed, i.e., it is not possible to use a verb from the more "stative" class with marking that is usually used for the other class, this is called split-S. Whenever this is in fact possible and the semantics change accordingly, e.g., to express volition or control over an action, one is dealing with a "fluid-S". In "split-ergative" Mayan languages, we are not dealing with a split of this kind. We are dealing with two kinds of alignments that are distributed according to aspect. All verbs can be inflected in all aspects. This is exactly as in Indo-Iranian. Additionally, the constructions of both languages can be explained in a similar way diachronically because in both cases they include a nominalized form that is integrated into the verbal paradigm and, via a syncretism of possession (in the case of Mayan) or instrumental/dative case (in the case of Indo-Iranian) and argument marking, causes a change in alignment.

It would be possible to develop a completely new label for this type of Mayan alignment and the consequent choice would be "split nominativity" or "split accusativity" to exactly mirror what happened in Indo-Iranian. However, "split ergativity" seems to be well-established in

---

(Marianne Mithun, p.c.; cf., e.g., Merlan 1985 for Seneca, and Mithun 1991 for Mohawk). Georgian and other Kartvelian languages have a class of so-called 'medial' verbs which follow the case marking pattern of transitive verbs, including the aspect-induced split (1985). As already noted by DeLancey (1985), in one language of this family, Mingrelian, all intransitive clauses seem to follow the split (c.f. Harris 1991)." (Bohnemeyer 2004: 101, en. 11).

[134] On this and for a discussion of Bohnemeyer's analysis in general see also Bergqvist (2011: 248): "It is obvious that status markers have a grammatical connection to certain AM[NK: aspect-mood]-markers, but this connection is not a stable one, grammatically. A case in point is the AM-marker *7uhch* in Lakandon that has radically changed semantics and grammatical status when compared to Yucatec (cf. Bergqvist 2007). More importantly, status markers indicate the syntactic status of a verb phrase, i.e. both the plain and the dependent status are used in dependent phrases where neither is preceded by an AM-marker." To conclude, he states: "Although it appears at first glance that status marking in Yukatekan is inseparable from aspect-mood inflection, there are several arguments coming from syntax and lexical semantics that questions [sic] such a view" (Bergqvist 2011: 255).

the Mayanist literature and I do not believe there is reason to separate the Mayan phenomenon terminologically from splits such as the Indo-Iranian one. This is the way I understand split ergativity in this thesis, which, I believe, is not an idiosyncratic definition:

> **Split ergativity** refers to a language that shows different alignment in subparts of its grammar conditioned by grammatical features like aspect which apply to all verbs. One of the alignments involved is an ergative-absolutive one. It does not matter whether the construction that leads to a new alignment affects transitive or intransitive verbs.

That last sentence is crucial in understanding why I am discussing this in such detail. There may be some interesting parallels between the conditioning of aspect-based split ergativity and the active-stative alignment that operates based on actionality that may be worth discussing. However, I would like to stress that the simple fact that the split arises due to a construction change in intransitive, not transitive verbs does not warrant the establishment of a new label. I would argue that the connection between split ergativity and transitive clauses is simply something that we are used to because the phenomenon was first described for languages where the new construction led to a change in transitive verbs while not changing the person marking of intransitive ones. As pointed out in chapter 2.1.1, alignment is established based on the relationship between intransitive and transitive verb marking, not based on the marking of one of these alone.

The final reason why "split intransitivity"/"active-stative" is an unfortunate label for the aspect-based split in Mayan is that there are some Mayan languages that actually manifest a split for which the label "active-stative" is a better fit. We will examine this phenomenon in the following chapter.

3.4.2 Active-stative split alignment in Mayan

Some Mayan languages manifest an additional kind of split marking on top of aspect-based split ergativity. Until very recently, Mopan was the only language for which this kind of split was described (Zavala Maldonado 2017: 245). By now, it has also been documented for Chol and Chontal (Western Cholan) and Poqomchi', though not Poqomam (Zavala Maldonado 2017: 227).[135] The split manifests itself in a way where intransitive predicates can be used

[135] Interestingly, all these languages also show aspect-based split ergativity. However, this is not true the other way around. That is, not all languages with aspect-based split ergativity also have an active-stative split.

with either one or the other construction depending on the "nature of the involvement of the single participant of the event portrayed" (Zavala Maldonado 2017: 241). When the subject is a volitional entity, a light verb construction is used that involves set A marking; otherwise, set B is used (Zavala Maldonado 2017: 241).

Here, examples from Chol will be used but the split works similarly in Chontal (Osorio May 2016). For Poqomchi', a short overview is provided in Zavala Maldonado (2017); a detailed account of the active-stative alignment split in Mopan is found in Danziger (1996). The split in Mopan works similarly to what is described for Chol in the following (and also similarly to Poqomchi') insofar as it is also based on a light verb construction, but instead of the transitive verb 'to do'[136], Mopan has the intransitive light verb *uch* 'to occur' (Zavala Maldonado 2017: 246).

Let us now turn to Chol. Here are again the examples that demonstrate split ergativity with ergative-absolutive alignment in the completive in (36) and (37) and nominative-accusative alignment in (38) and (39):

(36) Chol completive transitive (Coon 2008: 99)

Tyi *i-mek'-e-yoñ.*

COM A3SG-hug-TS-B1SG

'She hugged me.'

(37) Chol completive intransitive (Coon 2008: 99)

Tyi *wäy-i-yoñ.*

COM sleep-IS-B1SG

'I slept.'

(38) Chol incompletive transitive (Coon 2010a: 216)

Mi *a-mek'-oñ.*

INC A2SG-hug-B1SG

'You hug me.'

(39) Chol incompletive intransitive (Coon 2010a: 216)

Mi *a-wäy-el.*

INC A2SG-sleep-IS/NMLZ

'You sleep.'

[136] Chol *cha'l-*, Chontal *che-* and Poqomchi' *b'an-*.

In addition to that, set A is used to mark S on a specific class of "agentive" intransitive verbs regardless of the aspect: it is entirely impossible to use verbs like *ajñel* 'to run', *oñel* 'to shout', *tse'ñal* 'to laugh', *pay* 'to call', *tyujb* 'to spit' with set B as in (40) (Vázquez Álvarez 2011: 27). Instead, a light verb construction as in (41) is necessary, where the actual inflection is found on the light verb *cha'l* 'do'. This results in a literal translation 'I did a song/singing' instead of 'I sang'.[137] The use of set A is unsurprising, as the light verb, which receives the person marking, is transitive.

(40) Intended: completive intransitive (agentive verbs) (Vázquez Álvarez 2011: 27)

**tyi* *k'ay-oñ*
COM sing-B1
intended meaning: 'I sang'

(41) Completive intransitive (agentive verbs) (Vázquez Álvarez 2011: 27)

tyi *k-cha'l-e* *k'ay*
COM A1-do.B3-TS sing
'I sang'

Meanwhile, there is another class of verbs, for which the opposite is the case: when it comes to intransitive verbs like *majl* 'to go', *lets* 'to climb', *chäm* 'to die', *tyojm* 'to explode' or *jil* 'to finish', S can only be indicated by set B as in (42) while it is impossible to use the light verb construction as in (43) (Vázquez Álvarez 2011: 27).

(42) Completive (Vázquez Álvarez 2011: 27)

tyi *majl-i-yoñ*
COM go-IS-B1
'I went'

(43) Intended: completive with light verb (Vázquez Álvarez 2011: 27)

**tyi* *k-cha'l-e* *majl-el*
COM A1-do.B3-TS go-IS/NMLZ
Intended meaning: 'I went'

[137] It remains unclear to me how the authors working on Cholan distinguish the verbal root *k'ay* from the noun *k'ay* 'song' in cases like these. From my impression, it seems as if the construction is used both with nominalized verbs that feature an overt nominalizing suffix (*wäy-el*; see Table 9 below) and with underived nouns (*k'ay*).

Vázquez Álvarez (2011: 28) classifies this kind of alignment as "split-S", which fits the definition from chapter 2.1.2, although it may be up to argumentation or analysis whether this constitutes a split, as, technically, we are dealing with a transitive verb still. It would depend on the extent of grammaticalization. There is, strictly speaking, no use of set A on an intransitive verb comparable to the construction in the incompletive aspect as portrayed in (39). Still, there are two classes of verbs and each must obligatorily use one kind of marking, while the same marking is impossible to use with the other class. In Mayan, the light verb construction is the "agentive"[138] (active) one that generally involves more volition on the parts of the subject, while the verbs that can never use this light verb construction are considered "non-agentive" (stative) with less volition. Since the classes are fixed, the distribution need not be semantically clear in all cases. For example, one might wonder why *ajñel* 'to run' is agentive, while 'to go' is not. However, the semantics and volitionality are in fact important for a third class that exists in Chol.

This third class is best described as a case of "fluid-S", because for some verbs, e.g., *wäy* 'to sleep', *uk'* 'to cry', *ts'äm* 'to bathe' or *tyijp'* 'to jump', both constructions are possible (Vázquez Álvarez 2011: 28). The difference between the two constructions then is that (44) expresses more volition on the parts of the subject (in that the person deliberately went to sleep), while the use of set B in (45) does not (i.e., they fell asleep accidentally). The distinction would be clearer if the author had provided context for the sentences. As it stands, we need to rely on his analysis.

(44) Completive with light verb (Vázquez Álvarez 2011: 28)

tyi	*k-cha'l-e*	*wäy-el*
COM	A1-do.B3-TS	sleep-IS/NMLZ

'I slept'

(45) Completive with set B (Vázquez Álvarez 2011: 28)

tyi	*wäy-i-yoñ*
COM	sleep-IS-B1

'I slept'

For lists of some verbs that belong to either of these verb classes see Zavala Maldonado (2017: 242–243). Table 9 provides an overview of how the two alignment splits, split-/fluid-S and split ergativity interact in Chol.

[138] Agentive and non-agentive are the labels employed for Mayan languages.

Table 9. Overview of split-/fluid-S and split ergativity in Chol.

	agentive (+VOL)	non-agentive (–VOL)	fluid
incompletive	light verb *mi k-cha'-añ k'ay* 'I sing'	split ergativity (set A) *mi k-majl-el* 'I go' (Coon 2017: 667)	both possible depending on volition a) *mi k-cha'-añ wäy-el* 'I sleep' (+VOL) b) *mi k-wäy-el* 'I sleep' (–VOL) (Vázquez Álvarez 2011: 24)
completive	light verb *tyi k-cha'l-e k'ay* 'I sang'	split ergativity (set B) *tyi majl-i-yoñ* 'I went'	both possible depending on volition a) *tyi k-cha'l-e wäy-el* 'I slept' (+VOL) b) *tyi wäy-i-yoñ* 'I slept (–VOL)

While the non-agentive predicate class still features the split-ergative pattern in the different aspects, aspect differences in the agentive classes are encoded solely in the aspect marking of the light verb while the lexical verb, *k'ay*, is always merely a complement. As stated above, the agentive construction is ultimately a transitive one because the light verb that is used is transitive.

Returning to the definitions discussed in chapter 2.1.2 and to the question of label for the two splits discussed here, it seems that following Zúñiga (2018) or Bohnemeyer (2004), we would need to call both the active-stative split and split ergativity "split intransitivity". The former would be a mixture of split- and fluid-S since we have two strict classes that can only use one kind of construction (light verbs vs. split-ergativity) and one that is flexible. The latter would also be considered fluid-S, but with aspect as the conditioning factor. It would be rather cumbersome to use the same label for very different things, especially if they interact as closely as they do in Chol. Therefore – and in addition to the arguments provided in 3.4.1 –, it is preferable to refer to the split that is conditioned by predicate classes as "active-stative" (with the subgroups "split-S" and "fluid-S"), while we should keep the label "split ergativity" for the pattern conditioned by aspect.

4 The verbal system of Ch'orti'

Now let us consider the third modern Cholan language, Ch'orti'. It is not part of the Lowland Mayan area today, but it used to be spoken across a wider territory before the Spanish conquest (cf. chapter 2.2.5.1). It is therefore possible in theory that its aspect-based alignment split is also due to contact effects of this larger Lowland area. However, as is also usually acknowledged in the literature, Ch'orti' lacks any trace of nominalization entirely, which, as we have seen in chapter 3.3, is an important component of the construction that leads to the aspect-based split in the other languages.[139] The supposed close relatives of Ch'orti', Colonial Choltí and Hieroglyphic Maya, also do not show classic Lowland Mayan aspect-based split ergativity. While Choltí only features nominative-accusative alignment in the progressive aspect – in that respect it is described as resembling Poqom more closely than Yucatecan or Western Cholan –, Hieroglyphic Maya does not show split ergativity at all (Law 2017b: 118).[140] This and other arguments are the reason why Law suggests that the "contact-induced changes of the Maya lowlands postdate the height of Classic Maya power" (Law 2017b: 123). We have seen in chapter 2.2.5.1 that Ch'orti' quickly became isolated and was increasingly pushed back during the Spanish conquest. Therefore, depending on how late after "the height of Classic Maya power" split ergativity arose, it need not necessarily be true that what happened in Yucatecan influenced Ch'orti'.

In this chapter, I offer an examination of aspect-based split ergativity in Ch'orti' in order to determine how similar or different it is compared to what we see in Lowland Mayan languages. I first introduce the phonological system of Ch'orti' as it will become important later on, then the three index-sets and their use, after which I discuss the ways aspect (and tense) are marked in Ch'orti'. The translation of the verbs (though not the glossing of examples) usually also includes the verb class according to Hull (2016) as this is sometimes useful to know.

[139] Interestingly, this is also the case for Ixil (Mamean), which also shows aspect-based split ergativity but does not seem to use a nominalized verb (Law 2017b: 118).

[140] This means that the split cannot be reconstructed for Proto-Cholan but must have happened independently in the Cholan languages (Law 2017b: 118). Another argument in favor of this is the different character of the constructions involved in the non-ergative-absolutive alignment in Chol, Chontal and Ch'orti'. If the split were in fact Proto-Cholan, one would expect to find more similarities among the three languages.

4.1 Ch'orti' phonological system

The following overview is based on Dugan (2013: 18–21), which in turn is based on an older description in Pérez Martínez (1994), and expanded by information provided in Hull (2016: 7–9).[141] The IPA symbols were added by me based on the authors' descriptions.

Consonants

b [b] rare, only (or, according to Hull (2016: 7), "primarily") in Spanish borrowings

b' [ɓ] glottalized voiced bilabial stop/implosive, sometimes also realized as the corresponding ejective [p']; Hull (2016: 7): strong plosive quality at the beginning/in the middle of a word, but only lightly articulated word-finally; according to Fought (1972: 14), *b'* is only realized as an implosive in Camotán, while it is closer to ejective [p'] elsewhere

ch [t͡ɕ] alveolo-palatal plain affricate

ch' [t͡ɕ'] alveolo-palatal glottalized affricate/ejective

d [d] rare, only (Hull (2016: 7) "primarily") in Spanish borrowings; status as a phoneme debatable according to Dugan (2013: 18)

f [f] rare, primarily in Spanish borrowings

g [g] Hull (2016: 8): a voiced velar stop; Dugan (2013: 18–19): only in Spanish borrowings and as variant of /w/

j [h] voiceless glottal fricative (Dugan (2013: 18) specifies: **not** velar fricative [x])[142]; between a vowel and a consonant it is sometimes pronounced with what Dugan (2013: 21) calls an "echo vowel", e.g., *k'ajk* 'fire', which is more often pronounced [k'ah^{a}k] then [k'ahk]. "It appears that the *j* is realized as a lengthened vowel that moves from voiced to devoiced and back to voiced, at least in some circumstances" (Dugan 2013: 21).

k [k] voiceless plain velar stop

k' [k'] voiceless glottalized velar stop/ejective

[141] More treatments of the exact phonetic representation can be found in Fought (1967; 1972) as well as Pérez Martínez (1994) and Pérez Martínez et al. (1996).

[142] Note that /j/ usually denotes [x] in Mayan languages that keep the distinction between PM *h and *j.

l [l] alveolar lateral sonorant; Hull (2016: 7): in free variation with /r/ depending on speaker and dialect

m [m] bilabial nasal sonorant

n [n] alveolar nasal sonorant; Hull (2016: 8): often assimilates to following consonants, e.g., > /m/ before bilabial stops like /b'/ and /p/ (*inb'utz* 'good' > *imb'utz*, *ak'unpa* 'is content' > *ak'umpa*); Dugan (2013: 19, fn. 1, 34): assimilated to [ŋ] before velars and the labiovelar approximant /w/; both observations are confirmed by Fought (1967: 102, 114).

The nasal consonants [n] and [m] often nasalize a preceding vowel (Dugan 2013: 21).

p [p] voiceless bilabial plain stop

r [ɾ] alveolar tap; Hull (2016: 7): in free variation with /l/ depending on speaker and dialect

s [s] voiceless alveolar fricative

t [t] voiceless alveolar plain stop

t' [t'] voiceless alveolar glottalized stop/ejective

tz [t͡s] voiceless alveolar plain affricate

tz' [t͡s'] voiceless alveolar glottalized affricate/ejective

x [ɕ] voiceless alveolo-palatal (Hull: palatal) fricative

' [ʔ] plain glottal stop, can be realized as creaky voice on a preceding vowel as in *e'nte'* [ḛẽteʔ] (Dugan 2013: 21); see section on vowels below

Hull (2016: 8) notes that there is considerable variation among speakers with certain lexical items: ejectives are commonly articulated as regular plosives in final position (*nak'* ~ *nak* 'stomach') and as the first component in a consonant cluster (*ak'b'ar* ~ *akb'ar* 'night', *ak'ta* ~ *akta* 'dance'), the latter especially when the preceding vowel is "rearticulated" (see discussion on vowels below).

Semivowels

w [w] voiced labio-velar approximant

y [j] alveolo-palatal approximant

There is an alternation *g ~ w* in Ch'orti' today, possibly due to Spanish influence where *w* is not phonemic. The two phonemes, once allophones, "have now become distinct but related phonemes" (Hull 2016: 8) and regularly substitute for each other before /o/ as in *kawori ~ kagori* 'we rolled it into a ball' based on speaker preference. Dugan (2013: 18) describes that [g] replaces initial [w] in some words, e.g., *wororoj ~ gororoj* 'round' (doubtless the same root as in the example by Hull, *kawori ~ kagori*). However, he reports that informants that had received linguistic training in the PLFM often avoided using the letter *g* even when they unambiguously pronounced it due to it being "not really Maya" (Dugan 2013: 28).

Vowels

Information on the vowels is based on Hull (2016: 7); IPA was added by me based on his descriptions. The basic information sketches a very simple vowel system:

a [a] short low central vowel

e [e] short mid-front vowel

i [i] short high front vowel

o [o] short mid-back vowel

u [u] short high back vowel

However, the situation is in fact more complex. It is true that, though some other Mayan languages like K'iche' have long vowels and they are reconstructed for Proto-Mayan as well as attested in the hieroglyphic corpus, modern Cholan languages have merged historical long and short vowels.[143] According to Hull (2016: 8), there are no long or geminate vowels in Ch'orti', only so-called rearticulated vowels (*a', e', i', o', u'*), which seem to consist of a vowel and a glottal stop. The contrast between simple and rearticulated vowels is phonemic since there are minimal pairs like *xex* 'dirty' and *xe'x* 'corn husk' (Hull 2016: 8) .[144]

[143] See Kaufman and Norman (1984: 85) for potential remains of the PM, PCH *a : *aa distinction in Western Cholan in the vowels *ä ~ a.*

[144] In other Mayan languages where they are also phonemic (e.g., Yucatec), these vowels are sometimes called "broken" vowels. Their origin is assumed to lie in combinations of long or short vowels plus glottal stop. They are usually written <V'V> because this is closest to what people hear and "[i]t is regarded as too abstract to write them as VV' or V'" (Aissen, England & Zavala Maldonado 2017a: 11). As we have seen,

The phonetic reality of these vowels today does not become clear from Hull (2016) or many other publications. He speaks of "rearticulated vowels (e.g., *xe'x*) and glottalized vowels (e.g., *ja'*)" (Hull 2016: 7) but does not explain whether there is a difference between the two or whether he uses the terms interchangeably. The exact phonetic nature of "rearticulated vowels" has been analyzed for Yucatec Maya (Avelino, Shin & Tilsen 2011; Frazier 2011). The findings suggest that "the descriptive category 'rearticulated' corresponds to non-modal laryngealized vowels" (Avelino 2011: xi). These are "not generally produced as a vocalic gesture with an intervening glottal stop. Instead, laryngealized vowels are consistently produced with falling tone and laryngealized phonation towards the end of the vowel." (Avelino 2011: xi–xii).

Fought's careful phonetic description of Ch'orti' suggests that here, too, we are dealing with long vowels that involve "creaky voice" (Fought 1972: 14). Based on Fought's recordings on AILLA, this seems to me a fitting analysis. A precise and adequate description of these vowels is essential to understand the phenomenon affecting set C (and partly set A) that will be discussed in chapter 6. Following Avelino, Shin & Tilsen (2011: 2, fn. 3), I will refer to this type of vowels as "laryngealized" vowels from now on.

Since, as Hull states, these vowels are phonemic, we need to correct the Ch'orti' vowel system sketched at the beginning of this section to include long counterparts with creaky voice:

[a] [a꞉̰]

[e] [e꞉̰]

[i] [i꞉̰]

[o] [o꞉̰]

[u] [u꞉̰]

Aside from these long vowels with creaky voice, nasal vowels also exist in Ch'orti' though Hull (2016) does not describe nasal vowels at all and they are also not written in the modern orthography. According to Fought (1967: 106), they are limited to first-person singular forms.

---

these vowels are in fact written <V'> in most sources on Ch'orti'. Aissen, England and Zavala Maldonado (2017a: 11) only state that they are considered part of the vowel inventory in Yucatec, though in general they mention their existence in Yucatec, Itzaj, Mopan, Lacandón and Teko. They clearly also exist in Ch'orti' though it would need to be investigated whether they are phonetically the same as the vowels in Yucatec.

On the other hand, Dugan (2013: 21) states that the nasal consonants [n] and [m] generally often nasalize a preceding vowel. It seems as if they are not phonemic.

Diphthongs are described as uncommon, although Dugan (2013: 18) counts cases like *me'yn* 'shadow' (resulting from vowel contractions, see chapter 6) as diphthongs. From his fieldwork, Dugan reports an interesting discrepancy when it comes to two consecutive vowels: "I sometimes felt I heard *tar.i.e'n*, which my informants insisted should be written *tarye'n.*" (Dugan 2013: 28). Thus, there seems to be the notion that two vowels should never be written together, but at least to Dugan's ears it is not entirely clear whether we are dealing with a vowel or an approximant in these cases.

Also worth noting is the fact that in (orthographically) vowel-initial words, the initial glottal stop is often clearly audible although it is obscured by the writing system, which does not record it word-initially, and frequently deleted in connected speech (Dugan 2013: 23).

Consonant clusters and syllable structure

jn Hull (2016: 8): at the end of the word, /n/ is reduced and practically inaudible, e.g., in *kojn* 'ravine', *k'ajn* 'bench, seat', *warajn* 'trap'

jr found in, e.g., *sujrku / sujrki* in Hull (2016: 377)

Ch'orti' features "consonant clusters that are relatively complex by the standards of Maya languages" (Dugan 2013: 22), e.g., *a-takr-y-o'n* 'you help us' (Dugan's analysis: A2SG-help-STEM.FORMATIVE-B1PL). Dugan claims that these consonant clusters are not really "permitted" by Ch'orti' phonotactics: "The root *takr* is not pronounceable, as the consonant cluster *kr* is not permitted by Ch'orti' phonotactics. A root-final cluster such as this is often simplified by suffixing a stem formative before any additional morphological processing." This teleological analysis is unlikely to be correct as stem formatives are not added deliberately by speakers to relieve pronunciation.

More detailed information on syllable structure can be found in Fought (1967; 1972) and Dugan (2013). In general, complex syllable margins are highly restricted. Onsets may consist of a consonant with a glide (*y/w*), e.g., *a-tyob'* 'they bathed', but there are no clear examples of other clusters (Dugan 2013: 23). Although orthographically vowel-initial words often begin with an actual glottal stop, this is not obligatory for syllables: Derivational and inflectional suffixes especially can lack an onset and thus be of the form V or VC (Dugan 2013: 25).

The only elements that appear between a nuclear vowel and the coda are glottal stops, *j* or "orthographic *y*, which might be counted as a vowel or a consonant" (Dugan 2013: 24) as in *me'yn* 'shadow', see the discussion above. The three elements could, according to Dugan, likewise be analyzed as being part of the coda itself. If they are not, the analysis yields the following possible syllable nuclei (Dugan 2013: 25):

a	*e*	*i*	*o*	*u*
aj	*ej*	*ij*	*oj*	*uj*
a'	*e'*	*i'*	*o'*	*u'*
	e'y			

The coda can then only consist of a C, which is optional; in fact, syllables are often open (CV) (Dugan 2013: 25). Dugan (2013: 24) claims that he found no cases of *a'y, i'y, o'y* or *u'y* in his texts. On the other hand, Hull's (2016) bigger corpus provides many examples like *ka'y* 'to begin', *b'ani'y* 'thus, it's true, truly', *k'o'y* 'to suffer the agony that precedes a sure death, a loss of hope before dying', *t'u'y* 'to run down (a liquid)'.

Stress

Stress falls "almost universally" on the last syllable of a word (Dugan 2013: 26). The only exceptions are loanwords like *ilama* 'wife' < Nahuatl *ilamatl* 'old woman' (Karttunen 1992: 103) or *kilis* 'eclipse' (likely from Spanish *eclipse* with simplification of the consonant clusters *cl* and *ps*[145]), which are both stressed on the penultimate syllable (Dugan 2013: 26).

A common reduction process concerning vowels and shift of stress is the syncope of the second vowel in two-syllable lexical roots of the form CVCVC when the word receives a suffix (Wichmann 1999: 17). Wichmann calls this "pre-tonic syllable reduction" because it results from the fact that stress always falls on the last syllable and therefore shifts from the second syllable of the lexical root to the suffix:

ak'ab' 'night' + *-ar*	>	*akb'ar* 'night'
ka-winik 'our man' + *-ir*	>	*kawinkir* 'our Lord'
wayan 'to sleep' + *-ij*	>	*waynij* 'sleeping, sleepiness'

[145] The deletion of the initial vowel present in Spanish is either due to a "preference" for two-syllable words or due to reanalysis of *e* as the definite article *e* (Dugan 2013: 26–27).

Wichmann's restriction to lexical roots is opposed to the regularity of sound change and it needs to be reinvestigated in which cases this syncope does not happen . For instance, when *wayan* is suffixed with a set B index for inflection, e.g., *wayan-en* 'I slept' or even with an additional suffixed clitic as in *wayan-en=to* 'I slept until', there is no syncope. It is possible that the vowel is restituted based on the third person form *wayan* which is unmarked for person and therefore always keeps the second *a*. A different possibility is that the sound change described by Wichmann operated at a time where the elements that do not lead to it (e.g., set B) were not affixes/clitics yet.

Orthography

Nowadays, Ch'orti' follows the official orthography established in Guatemala for all Mayan languages, although there is still much variation in orthography in practice (Hull 2016: 7). The conjunction 'and', e.g., is mostly written <yi> or <y> by the Ch'orti' today (no doubt due to Spanish influence, where 'and' is *y*). The form <yi> is an attempt to distinguish the form precisely from the same Spanish *y*, but it is deemed phonetically inaccurate by Hull (2016: 7), who rejects both forms and instead writes <i>. Cases where *i* 'and' "palatalizes to a /y/ in spoken Ch'orti' when preceding a vowel" (Hull 2016: 7) are due to assimilation and not due to the true articulation of the phoneme.

4.2 Index-sets and pronouns in Ch'orti'

The verbal system of Ch'orti' is unique among the Mayan languages in several ways. As Vinogradov (2016: 61) puts it, "Ch'orti' has an exceptional grammatical TAM system, not only among the Cholan-Tseltalan group, but also among the entire Mayan family". For this reason, the system is generally believed to be an innovation (Kaufman & Norman 1984: 102). The most obvious peculiarity is the presence of a third index-set where other Mayan languages only have set A and set B. The existence of the forms *inwayan* (C1SG *in-* + *wayan* 'AP12 to sleep') and *inchamay* (C1SG *in-* + *chamay* 'MPAS4 to die') in Galindo's (1834) wordlist seems to be evidence for the fact that set C already existed at this point. Related to the presence of set C is the absence of morphological aspect marking (Vinogradov 2016: 61), at least for transitive verbs. The following chapters will first present the forms of all three sets and then their use in the language.

4.2.1 Paradigms

In Table 10 and Table 11, the index-sets and pronouns of K'iche' (as a prototypical example of a Mayan language) and Ch'orti' are contrasted.

Table 10 Index-sets and pronouns of K'iche' (Can Pixabaj 2017: 467).

	Set A		Set B	Independent
	preconsonantal	prevocalic		
1SG	*nu- / in-*	*w- / inw-*	*in-*	*in*
2SG	*a-*	*aw-*	*at-*	*at*
3SG	*u-*	*r-*	*ø*	*ri are' (ra're')*
1PL	*qa-*	*q-*	*oj-*	*(ri) oj*
2PL	*i-*	*iw-*	*ix-*	*(ri) ix*
3PL	*ki-*	*k-*	*ee-, e-*	*ri e are' (ri a're', ra're')*

Table 11 Index sets and pronouns of Ch'orti' (synthesis of Dugan (2013: 34–38) and Pérez Martínez (1994: 45, 55, 66, 84).

	Set A		Set B	Set C[146]	Independent
	_C	_V			
1SG	*ni-* *in-*[147]	*(ni)w-*[148] *inw-*	*-en / -e'n*	*in-*	*ne'n*[149]
2SG	*a-*[150]	*aw-*	*-et / -e't*	*i-*	*ne't*
3SG	*u-*	*uy- / uw-*[151]	ø	*a-*	*ja'x(ir)*
1PL	*ka-*[152]	*kaw-*	*-on / -o'n*	*ka-*	*no'n*
2PL	*i-*	*iw-*	*-ox / -o'x*	*ix-*	*no'x*
3PL	*u- … (-ob')*	*uy- … (-ob')*	*(-ob' / -o'b')*	*a- … (-ob')*	*ja'x(ir)(ob')*

Aside from the presence of a third index-set (set C), which is also a prefix like set A, K'iche' and Ch'orti' still have much in common, though there are also further substantial differences. I already mentioned that set B is prefixed in some Mayan languages (e.g., K'iche') and suffixed in others (e.g., Ch'orti'). Minor phonological differences that are worth mentioning

[146] Set C has also been called "nominative set", e.g., by Dugan (2013: 33), or is considered a kind of "second ergative set". The latter term reflects underlying historical assumptions as to the origin of set C because most authors view it as a variant of set A historically. After reviewing the proposed theories in chapter 5, it will become clear why set C should not be called "second ergative set". "Nominative" is even less appropriate. Dugan justifies the choice by stating that the "term aptly describes the use of the third set of affixes to indicate the subject (only) of intransitive verbs" (Dugan 2013: 33). This reasoning appears to be based on a misunderstanding of the definition of "nominative". Nominative would rather be the use of the same marking for S **and** A, not for S exclusively. I only use the neutral labels "set A", "set B" and "set C".

[147] Dugan (2013: 34) gives A1SG as *iN- ~ inw- / -n-*. The capital N in *iN-* is Dugan's way of describing the phenomenon that *n-* assimilates to a following consonant (> [m] before bilabials, > [ŋ] before velars) as described in chapter 4.1. On the other hand, *-n-* is meant to depict the phenomenon where A/C1SG *in-* morpheme is "metathesized" into the verb root. This happens with vowel-initial verbs, e.g., *ati* 'IV25 to bathe' – *a'nti* 'I bathe' (*in-ati*) and will be discussed in detail in chapter 6.

[148] Pérez Martínez (1994: 44–45) lists *ni- ~ w-* but in fact only uses *niw-* as a prevocalic form in examples, not *w-*.

[149] Dugan (2013: 36) additionally provides unlaryngealized forms for the independent pronouns but I could not confirm this with Hull (2016).

[150] Ch'orti' does not seem to use separate honorific pronouns like, e.g., K'iche' (compare footnote 100).

[151] The glide sometimes seems to be *-w-* in the third person instead of the expected *-y-*. Dugan explains this as the speaker "reanalyzing the root as if it began with an underlying *w* instead of with an underlying glottal stopped [sic] realized as a *w*" (Dugan 2013: 34). Since this seems to be the case with verbs beginning in *i-* (Fought 1967: 110; Schumann Gálvez 2007: 133–135), a purely phonetic explanation is more attractive than morphological reanalysis.

[152] Some languages like Chol differentiate between inclusive and exclusive first person plural forms. This is not the case in Ch'orti'.

include the regular sound correspondences of KCH *q* ~ CHR *k* as well as KCH *r* ~ CHR *y* (Campbell 1984: 6).

Just like K'iche', Ch'orti' has a preconsonantal and a prevocalic[153] form of set A. The fact that the prevocalic version of set A is mainly the preconsonantal one with an additional glide is also a common Mayan feature. However, note the innovation in the variation of CHR *u-* ~ *uy-*, where *u- ~ *y- would be the original distribution.[154] Likewise, *ka- ~ kaw-* is observed instead of the expected *ka- ~ *k-. Compared to K'iche', the paradigm in Ch'orti' looks more regularized.

For CHR A1SG, two different forms (with respective prevocalic variants) exist, which is also parallel to K'iche', both formally and in terms of distribution: the original form of set A is still used to mark possession on nouns and the new form that is based on B1SG *in-* is used to mark A on verbs (see the discussion on page 67).

Nouns: *ni-chitam* 'my pig' *niw-ajtzo'* 'my turkey' (Pérez Martínez 1994: 44–45)

Verbs: *in-jatz'e't* 'I hit you' *inw-ire't* 'I see/saw you' (Pérez Martínez 1994: 62, 64)

Although synchronically, B1SG is *en* and not *in* in Ch'orti', *in/iin* is what is attested in most Mayan languages and what is currently reconstructed for Proto-Mayan in *iin (Kaufman 2015: 161). The alternative form of set A must have developed before CHR B1SG changed the vowel to *e*. As the phenomenon is so widespread, it likely happened early on in Mayan history.

Set B in Ch'orti' shows an unexpected variation of forms with laryngealized and simple vowels. Laryngealization is usually pronounced in careful speech according to Dugan (2013: 39) but disappears in connected or rapid speech. He was unable to identify a grammatical function for this alternation (Dugan 2013: 36) but suggests that it simply represents a (perhaps

[153] There are differing opinions on the question of whether there are "truly" vowel-initial words in Mayan or not. Kaufman argues against it; however, Campbell (2017: 46) points out that at least some stems must begin with a vowel and not a glottal stop. Dugan attempts to explain the presence of the glides in set A through an interaction of set A and the (unwritten) glottal stop, which is then "realized as a glide" (Dugan 2013: 34), more precisely as *w* in the first and second person and as *y* in the third. According to Edmonson (1988: 115), Huastec does not distinguish prevocalic and preconsonantal forms for set A. This is significant for our understanding of Mayan history as Huastec is thought to be the first language to have split off from the others. Therefore, it may not be possible to reconstruct the distinction for Proto-Mayan.

[154] Forms like these are common in Lowland Mayan languages but deemed incorrect in others like K'iche'. López Ixcoy (1997: 100) calls this "double possession" and points out that it happens often with vowel-initial words, but that the "correct" form is one that uses the vowel-initial prefix (*w-ixiim, r-ixiim*) instead of both combined (**nuw-ixiim, **ur-ixiim).

dialectal) variation in the language. Schumann Gálvez (2007: 61–62, 135) reports that laryngealized vowels are a feature of the speech of "los ancianos" (the elderly), while younger people tend to eliminate it. However, he himself admits that even in young people's speech, laryngealization sometimes reappears, although the general tendency is for it to get lost. The origin of this distinction will be discussed in 6.1. The laryngealized forms result from an interaction of the suffixes with different kinds of stem endings. This process follows specific phonological rules.

Ch'orti' does not have any special forms for the third person plural compared to K'iche' (e.g., KCH A3PL *ki-* ~ *k-*). Instead, the third person singular is used with an optional plural morpheme *-ob'* as in other Cholan languages and Yucatecan, among others. The same morpheme marks plural on nouns, e.g., CHR *chij-ob'* 'horses' (Pérez Martínez 1994: 43). There is some uncertainty as to whether the plural morpheme *-ob'*, which is not mandatory, should be considered part of the bound person paradigm. Dugan (2013: 33) argues against this and rather considers *-ob'* a "discourse marker" precisely because it is not mandatory. Though this label does not seem appropriate since discourse markers are not defined by being optional, the optionality of *-ob'* can lead to ambiguous sentences especially in transitive verbs, e.g., when both subject and object are in the third person and the plural marker could thus apply to either of them as in *uy-ust-ob'*, which could mean 's/he prepared them' or 'they prepared it' or 'they prepared them' (Dugan 2013: 83).

Independent pronouns (used for emphasis) are morphologically based on set B, just like those of K'iche' and most other Mayan languages, although in Ch'orti', they are added to a stem *n-* in the first and second persons with a special form *ja'x(ir)* for the third person. *Ja'x* and *ja'xir* are "largely interchangeable, sometimes both used in the same sentences to refer to the same person" (Dugan 2013: 39). A subtle distinction sometimes seems to be that *ja'xir* is used more often "[w]hen referring to an abstract concept or an indistinct group of individuals, while *ja'x* is more likely to refer to a specific individual or object" (Dugan 2013: 39). This is plausible since *-ir* is a suffix used to form abstract nouns in the language (see the Appendix E chapter c.4). He also describes cases where both appear near to each other, which gives them a 'the one' – 'the other' or 'the former' – 'the latter' meaning. The plural marker can attach to both forms. It follows the *-ir* if one is present.

Though the other authors also point out that there is some phonological interaction between set C and the verb that it is prefixed to or at least record the variation in their examples, Schumann Gálvez (2007: 135) is the only one among the Ch'orti' grammarians who presents

set C as having a full-fledged distinction of preconsonantal and prevocalic forms just like set A.[155] His version of the paradigms is given in Table 12. If one compares set A to set C, it becomes immediately obvious that the two sets do not interact in the same way with prevocalic stems – in fact, there is not a single overlap. The most drastic interaction is shown by C3 *a-*: it is described as assimilating to the initial vowel of the verb followed by a glottal stop, which is in turn sometimes followed by a "rearticulated" vowel (Schumann Gálvez 2007: 135) – likely the author's description of a laryngealized vowel. He does not provide an example there but there is one in his recorded texts: *a'(a)-xin* (< *a-ixin*) 'he goes' (Schumann Gálvez 2007: 255).

Table 12. Set A from Table 11 and set C according to Schumann Gálvez (2007: 135).

	Set A		Set C	
	_C	_V	_C	_V
1SG	*ni-* *in-*	*(ni)w-* *inw-*	*in-*	*-n-*
2SG	*a-*	*aw-*	*i-*	*y-*
3SG	*u-*	*uy- / uw-*	*a-*	*V'(V)-*
1PL	*ka-*	*kaw-*	*ka-*	*k-*
2PL	*i-*	*iw-*	*ix-*	
3PL	*u- … (-ob')*	*uy- … (-ob')*	*a- … -ob*	*V'(V)- … -ob*

In chapter 6, I will discuss the conditions of this variation. One could in fact argue that set C has a preconsonantal and a prevocalic set – however, the situation is a little more complicated than that and there are reasons why the parallel to set A should not be emphasized too strongly. That is why I choose not to record the variation in the paradigm overview. One should keep in mind, however, that it exists and that it differs from the variation of set A.

4.2.2 The use of the three sets

The following examples demonstrate the use of the three sets. Set A marks possession as in example (1) as well as A on transitive verbs as in (3). Set B marks S in the completive aspect (referring to an action that is finished from the point of view of the speaker[156]) as in example (2) as well as O in (3). Set C is used exclusively to mark S in the incompletive aspect

[155] Quizar (2023) also makes this distinction in a recent paper.

[156] Tense anchors an action to some temporal reference point while aspect rather describes the viewpoint of the speaker, i.e., whether the action is viewed from outside, in which case it is completed, or from the inside, in which case it is, e.g., in progress, habitual or repeated.

(referring to an action that is habitual or ongoing) as seen in (4). When comparing the intransitive completive in (2) with the transitive verb in (3), S and O are marked alike while A is marked separately – this constitutes ergative-absolutive alignment as we already know it from other Mayan languages. However, when we contrast the intransitive incompletive in (4) with the transitive verb in (3), each argument receives its own marking. This constitutes a case of tripartite alignment.

(1) Ch'orti' possessed noun (Hull 2016: 75)

u-jun

A3-book

'his book'

(2) Ch'orti' intransitive completive (Quizar 1979: 44)[157]

k'axi-en *ta* *ch'en*

fall-B1SG PREP hole

'I fell into the hole.'

(3) Ch'orti' transitive (Quizar 1979: 44)

e *winik* *u-ira-en*

DEF man A3SG-see-B1SG

'The man sees/saw me.'

(4) Ch'orti' intransitive incompletive (Quizar & Knowles-Berry 1988: 79)

a-k'otoy

C3SG-arrive

'He arrives.'

Ch'orti' also possesses the stative construction that forms copula-like sentences with non-verbal predicates through suffixing of set B to nouns or adjectives. It is illustrated in example (5).

(5) Stative construction (Quizar 1979: 45)

pakar-en

upside.down-B1SG

'I am upside down.'

[157] Ch'orti' has a great variety of "thematic suffixes". Unless their exact analysis is relevant for a specific argument, I do not gloss them separately as IS or TS and only provide the translation for each verb stem as a whole. I do, however, separate some derivational affixes like the causative.

Just like Yucatecan, Western Cholan and Poqom, Ch'orti' has split ergativity that is conditioned by aspect with ergative-absolutive alignment in the completive. What is different, however, is that the incompletive aspect does not feature nominative-accusative but tripartite alignment (Zavala Maldonado 2017: 237). Furthermore, all existing aspectual contrasts usually apply to both types of verbs equally in the other languages, even if the morphology involved is not identical. In Ch'orti' on the other hand, transitive verbs have no means of marking aspect morphologically. A transitive sentence without additional aspect markers like adverbs is ambiguous in whether the action it describes is ongoing (incompletive) or not (completive). The sentence in (3) can mean both 'The man sees me.' and 'The man saw me.' depending on the context. Therefore, we have to compare the two intransitive aspects with the same single transitive clause to establish the alignment types.

Aside from the somewhat peculiar marking of completive (set B) or incompletive (set C) aspect on intransitive verbs exclusively, there is another strategy of expressing aspect in Ch'orti', which I will describe in the following chapter.

4.3 Aspect (and tense) in Ch'orti'

Languages that do not possess morphological means of marking aspect can, of course, still express aspect via lexical material like adverbs. Ch'orti' possesses a number of aspectual or temporal markers, usually vaguely called "particles", e.g., the progressive marker *war* which stresses that the action is in progress. Contrary to the marking via set B/C, the aspect/tense markers/particles are used with both intransitive (6) and transitive (7) verbs.

(6) Intransitive progressive (Hull 2016: 80)

E	*b'utu'pat*	*chitam*	*war*	*a-wayan.*
DEF	curved.back	pig	PROG	C3-sleep

'The pig with the curved back is sleeping.'

(7) Transitive progressive (Hull 2016: 478)

E	*winik*	*war*	*u-k'uxi*	*e*	*we'r.*
DEF	man	PROG	A3-eat.B3	DEF	meat

'The man is eating the meat.'

The progressing action can be explicitly situated in the past when another particle, *ani*, is added between *war* and the verb as in (8). *Ani* is analyzed as an irrealis marker by Hull (2016:

52–53) that is "also used to show perfective aspect in the past or future".[158] Note that the meaning 'yesterday' is likewise explicitly conveyed by *akb'i* in addition to *ani*.

(8) *war ani* (Hull 2016: 499)

Akb'i war ani in-xich'i u't pojp.					
Akb'i	*war*	*ani*	*in-xich'i*	*u-ut*	*pojp.*
yesterday	PROG	PST	A1SG-spread.out.B3	A3-face	tule

'Yesterday I was laying out tule strands.'

The two strategies to encode aspect (the morphological one for intransitive verbs and the one with particles such as *war* for both intransitive and transitive verbs) are best viewed as separate phenomena because set B and C do not obligatorily co-occur with specific particles. In fact, there are even cases of *war* occurring with the completive aspect, e.g., in (9), which we would not necessarily expect if set C had evolved together with the use of aspect particles *war*. Set B here expresses that the action is completed from the point of view of the speaker, though the verb includes the lexical information that the action is repeated. *War* encodes that the action of repeatedly cutting is in progress.

(9) *war* without set C (Hull 2016: 115)

War ch'akruma e sitz' turu to'r ti'naj.						
War	*ch'akruma*	*e*	*sitz'*	*turu*	*ta-u-jor*	*ti'naj.*
PROG	cut.repeatedly.B3	DEF	boy	be.seated.B3	PREP-A3-head	patio

'The boy is seated on the patio cutting repeatedly.'

Further evidence for the fact that *war* and set C are separate phenomena may be seen in the fact that other elements can occur between *war* and the verb as in the phrase *war tokto* 'to have just done sth., just have started doing sth., just now' in (10):

(10) *war tokto* (Hull 2016: 93)

E	*winik*	*war*	*tokto*	*a-b'ixk'a*	*kumja'=to*
DEF	man	PROG	right.now	C3-wake.up	swallowing.water=still
war	*u-che.*				
PROG	A3-do.B3				

'The man just woke up and still has water in his mouth.'

[158] More details on *ani* can be found in Appendix E chapter b.5.

Before we can begin examining hypotheses concerning the origin of set C in Ch'orti', it is necessary to clarify that we are in fact dealing with a split that is conditioned by aspect. In chapter 3.4.2, we have seen that split ergativity and split-/fluid-S can co-occur in Mayan languages. Since two of the languages with split-/fluid-S are Cholan ones, it is justified to ask whether this might also be the case for Ch'orti' or whether Ch'orti' is not split-ergative at all. This is especially true because in Ch'orti', the arising aspect distinction only affects intransitive verbs while transitive verbs remain morphologically ambiguous as to aspect (or tense).[159] In addition to that, Schumann-Gálvez (2007: 195–196) in his grammar mentions that the split might be best described in terms of agentivity. However, he does not provide further information on his reasoning and, interestingly, he still ultimately labels the split as aspect-based.

If the split that we observe in Ch'orti' is in fact a split- or fluid-S one, we would expect tripartite alignment used with agentive verbs but ergative-absolutive alignment with non-agentive verbs or the other way around. If the classes were lexically determined, this would constitute a split-S system. Alternatively, if all or most verbs could in theory be used with either marking, the difference would then need to be a semantic one of volition or control over the action etc. constituting a fluid-S split. Of course, Ch'orti' could also have two rigid classes and one fluid class, just like Western Cholan, and thus be both split-S and fluid-S.

At least for examples (2) and (4) above, the semantics of the verbs fit into this theory. In terms of active-stative alignment, this could be conditioned in two ways: the example for the completive aspect, 'I fell into the hole', could be argued to describe the state of having fallen into a hole whereas the example for the incompletive aspect would describe the action of arriving. On the other hand, one could also argue that falling into a hole is an action with –volition or –control while the arrival is +volition or +control. Based on the Western Cholan distinction, one would expect the latter for Ch'orti'.

However, if the split is lexically based, one would expect (2) to have another valid translation, namely 'I fall into the hole', because the forms should not code for any specific aspect. Likewise, (4) would be expected to also mean 'I have arrived.' This is not what we see in the corpus. For 'to fall' in the incompletive aspect, set C is usually used, e.g., in (11). It is not possible to argue that this example describes an action as opposed to the state in (2) or that it

[159] Dugan (2013: 87) claims that "the default aspect of transitive verbs is completive" but gives no evidence or justification for this claim.

involves control or volition. More importantly, no potentially confounding aspect/tense particles are present in this example or the ones below.

(11) *k'axi* with set C (Hull 2016: 61)

Tama	*e*	*mayo*	*a-k'axi*	*e*	*b'ajxan*	*jaja'r.*
PREP	DEF	may	C3-fall	DEF	first	rain

'In May the first rain falls.'

Examples (12) and (13), (14) and (15) and (16) and (17) respectively show various verbs used both with set B and set C. The fact that most if not all verbs can be used with both index-sets makes it impossible that we are dealing with split-S where class membership would be rigid.[160] If the alignment in Ch'orti' were active-stative, we would therefore have to assume fluid-S where verbs change their semantics depending on what type of marking is used. However, the following examples cannot be interpreted in terms of state/action or –/+ control or volition in any meaningful way. Instead, anytime set C is employed, the verbs denote something that is still ongoing and not completed. In (13), vomiting is described as something that starts to affect women when they are pregnant therefore referring to a repeated action, in (15), we follow a crab that seems to go for a walk along the bank of a river on a regular basis, and in (17), the man cannot manage to fall sleep. On the other hand, the remaining three examples describe actions that have been completed: in (12), the watermelons have already begun to produce fruit, in (14), the girl has already left and in (16), the drunk's sleeping on his hat is also already completed since the consequences of it (the hat being folded over) are already visible.

(12) *ka'y*: intransitive with set B(Hull 2016: 524)

Ka'y yutiri e sandía.

ka'y	*yutiri*	*e*	*sandía*
begin.B3	produce.fruit.B3	DEF	watermelons

'The watermelons have begun producing fruit.'

[160] There is some verbal stem suppletion based on aspect like COM *tari*, INC *watar* 'to come', but this is not the same as one verb only ever being used with one kind of marking. Therefore, split-S is off the table.

(13) *ka'y*: intransitive with set C (Hull 2016: 241)

Aka'y uxe' ub'a.

a-ka'y	*u-xe'*	*u-b'a*
C1SG-begin	A3-vomit.B3	A3-REFL

'They begin to vomit.'

(14) *lok'oy*: intransitive with set B (Hull 2016: 258)

Lok'oy	*xana*	*e*	*ijch'ok.*
leave.B3	walk.B3	DEF	girl

'The girl left to take a walk.'

(15) *lok'oy*: intransitive with set C (Hull 2016: 425)

Akb'ar alok'oy axana e jopop tu'ti' e xukur.

akb'ar	*a-lok'oy*	*a-xana*	*e*	*jopop*	*ta-u-ti'*	*e*	*xukur*
night	C3-leave	C3-walk	DEF	crab	PREP-A3-mouth	DEF	river

'At night the crab goes out along the bank of the river.'

(16) *wayan*: intransitive with set B (Hull 2016: 257)

E	*aj-karer*	*wayan*	*tama*	*u-b'itor […]*
DEF	AGT-drunk.VN	sleep.B3	PREP	A3-hat

'The drunk slept on his hat [and it made it folded over.]'

(17) *wayan*: intransitive with set C (Hull 2016: 316)

Ma'chi a'ktana awayan.

ma'chi	*a-aktana*	*a-wayan*
NEG	C3-be.left	C3-sleep

'He isn't able to sleep.'

Where the other examples may be taken as ambiguous, (16) and (17) explicitly exclude the possibility that the marking via set B or set C is one depending on control or volition because both examples refer to the absence of control or volition. Finally, *wayan* can refer both to 'sleeping' and to 'falling asleep'. Examples (18) and (19) show *wayan* with the semantics of 'falling asleep' used both with set B and set C. This excludes the possibility of the sets referring to differences in actionality (e.g., telic/atelic etc.).

(18) *wayan*: intransitive with set B (Hull 2016: 521)

Tama uyujkna'r e ch'urkab' wayanen.

tama	*u-yujkna-ar*	*e*	*ch'urkab'*	*wayan-en*
PREP	A3-swing-VN	DEF	baby	sleep-B1SG

'In rocking the baby I fell asleep.'

(19) *wayan*: intransitive with set C (Hull 2016: 522)

Yujku-n	*e*	*ch'urkab'*	*twa'*	*a-wayan.*
swing.IMP	DEF	baby	so.that	C3-sleep

'Swing the baby so that it'll fall asleep.'

While the examples in Hull (2016) are useful, I had to rely on his translations[161] to pinpoint the semantic details because we lack information on the context of the action. Here are some more examples, this time from the Ch'orti' legends compiled by Pérez Martínez (1996) with more context. Examples (20) and (21) show the verb *lok'oy* 'MPAS4 to leave, go out; come out, result in, turn out' in use with set B while (22) and (23) represent cases of the same verb with set C.

(20) […] *konda lok'oy, e ixik tza'yi.* ("E noxib'ob'" in Pérez Martínez 1996)

konda	*lok'oy*	*e*	*ixik*	*tza'yi*
when	leave.B3	DEF	woman	be.happy.B3

'[…] when he left, the woman became happy. [Span. "se puso contenta"]'

(21) […] *muktz'a taka e masa' i ma'chi'x lok'oy.* ("E jaja'r" in Pérez Martínez 1996)

muktz'a	*taka*	*e*	*masa'*	*i*	*ma'chi'x*	*lok'oy*
disappear.B3	PREP	DEF	deer	and	no.longer	go.out.B3

'[…] he disappeared with the deer and he did not come out again.'

(22) […] *sib'i e xukur xe' alok'oy Hondura* […] ("E jaja'r" in Pérez Martínez 1996)

sib'i	*e*	*xukur*	*xe'*	*a-lok'oy*	*Hondura*
inflate.B3	DEF	river	REL	C3-leave	Honduras

'[…] it [the rain] had inflated the river that leaves Honduras […]'

[161] It is helpful that he provides both a Spanish and an English translation. Still, too much remains open when we know nothing about the story.

(23) […] *konda alok'oy ani,* […] ("E noxib'ob'" in Pérez Martínez 1996)

konda *a-lok'oy* *ani*

when C3-leave PST

'[What was happening was that] whenever he left [someone saw that come night another man arrived there to sleep with the woman.]'

With the use of set B, the examples have a clearly completive meaning. This is especially clear if we contrast (20) with (23). In (20), the story starts with a single occurrence of the man leaving. This action is already finished – this is why the woman became happy, because her husband is gone and she can engage in her extramarital affair. In (23), the narrator explains that this actually happens on a regular basis. Therefore, set C with incompletive aspect is used. Likewise, (21) features the use of set B to describe a completed action because the man did not reemerge from the water and was never seen again whereas (22) describes something that is not completed because the river is still flowing out of Honduras. The distribution of completive and incompletive in Ch'orti' furthermore fits the usual distribution in languages that have a grammatical distinction of aspect: in narration, actions, especially sequences of completed actions, are expressed using completive/perfective aspect, while background information is usually given in the incompletive/imperfective (Hopper 1979).

For the sake of completion, I would like to demonstrate that the distinction also applies to the special class of intransitive verbs that are derived from positional roots (on positionals see chapter b.1 in Appendix E), *tak'wan ~ ta'k'wan* 'POS to be pregnant, produce children'. In (24), the woman has succeeded in falling pregnant and thus set B is used to mark the completive. On the other hand, in (25), the action of 'being pregnant' has not been achieved yet; therefore, incompletive marked by set C is used.

(24) *tak'wan ~ ta'k'wan*: completive positional (Hull 2016: 391)

E *ixik* *ta'k'wan=ix* *uy-ar.*

DEF woman be.pregnant.B3SG=already A3-child

'The woman is pregnant.'

(25) *tak'wan ~ ta'k'wan*: incompletive positional (Hull 2016: 391)

E *ixik* *ma'chi* *a-tak'wan.*

DEF woman NEG C3-be.pregnant

'The woman can't get pregnant.'

Finally, one more grammatical phenomenon needs to be discussed. In chapter 3.4.2, it was mentioned that making a terminological distinction between split ergativity and active-stative split is especially important because both exist in Western Cholan, Mopan and Poqomchi'. Active-stative split alignment in Western Cholan arises via a light verb construction with the verb 'to do'. A comparable light verb construction consisting of a light verb *che* 'to do' and a complement is attested in Ch'orti', as well (described in detail in Appendix E, chapter h.6). However, contrary to Western Cholan, here, it is used almost exclusively with genuine nouns and adjectives as well as Spanish infinitives but not with nominalized Ch'orti' verb forms. It does not seem to be used in Ch'orti' with verbs that require it in both Chol and Chontal. Therefore, it is not comparable to the construction in Western Cholan and does not constitute split-/fluid-S marking.[162]

To conclude, the alignment split in Ch'orti' is indeed 1) best described as conditioned by aspect and 2) arises through the use of set B versus set C on intransitive verbs while the aspect marking strategies that employ aspect (or tense) particles are an independent phenomenon. Therefore, in order to understand how the split came about, we need to investigate the origin of set C, not that of the aspect markers.

[162] Schumann Gálvez himself explicitly states that one should not call Ch'orti' an agentive language because of this construction, not only because we are dealing with a very limited phenomenon but also because there is no "representative interference in its syntax" ["por no tener injerencia representativa en su sintaxis"] (Schumann Gálvez 2007: 141).

5 The origin of Set C: Traditional explanations

For reference, the different index-sets of Ch'orti' are displayed again in Table 13.

Table 13 Index-sets and pronouns of Ch'orti' (synthesis of Dugan (2013: 34–38) and Pérez Martínez (1994: 45, 55, 66, 84).

	Set A		Set B	Set C	Independent
	_C	_V			
1SG	*ni-* *in-*	*(ni)w-* *inw-*	*-en / -e'n*	*in-*	*ne'n*
2SG	*a-*	*aw-*	*-et / -e't*	*i-*	*ne't*
3SG	*u-*	*uy- / uw-*	*ø*	*a-*	*ja'x(ir)*
1PL	*ka-*	*kaw-*	*-on / -o'n*	*ka-*	*no'n*
2PL	*i-*	*iw-*	*-ox / -o'x*	*ix-*	*no'x*
3PL	*u- … (-ob')*	*uy- … (-ob')*	*(-ob' / -o'b')*	*a- … (-ob')*	*ja'x(ir)(ob')*

The following chapters will discuss the theories that have been proposed to explain the origin of set C in chronological order.

5.1 Robertson (1998)

The most prominent hypothesis on the origin of set C is the one proposed in Robertson (1998), a short paper with far-reaching consequences. It is most of all this paper and its claim that "Classic Maya, Ch'olti', and Ch'orti' form a lineage" that the famous article[163] by Houston, Robertson and Stuart (2000a)[164] concerning the language recorded in the hieroglyphic texts builds on.

In the following section, I will first sketch those features of Cholti morphosyntax that are necessary to understand Robertson's proposal. In section 5.1.2, I will discuss the actual proposal and whether it is plausible to assume that the system of Cholti evolved into what we see in Ch'orti'.

[163] The authors specifically ascribe an Eastern Cholan affiliation to Hieroglyphic Maya making Ch'orti' its only surviving descendant and arguing that Cholti forms a link between the two (Houston, Robertson & Stuart 2000a: 337–338). Though the paper was initially met with some criticism (e.g., Fought 2000; Grube 2000; Hofling 2000b), the proposal has by now been broadly accepted in the community (though see the ongoing debate between the authors and, e.g., Mora-Marín mentioned on page 32).

[164] Some of the ideas were apparently also discussed in an earlier unpublished manuscript (Houston, Robertson & Stuart 1998).

5.1.1 Choltí morphosyntax

The Choltí grammar handed down in two versions in the Morán Manuscript (MM; see chapter 2.2.6 for a general introduction) attests five inflectional categories for intransitive verbs[165]: "present" I, "preterite", "future", "future 'en -rus'" and "present" II.[166] Table 14 provides an overview including the Spanish names of the categories used in the grammar, a transcription, transliteration and translation. The green color in the table highlights the forms with a preverbal *a* that represent the parallel to Ch'orti' that Robertson builds his theory on.

Table 14. Summary of intransitive verb categories in Choltí with formations involving *a* highlighted in green (Morán (1695: 30–32, 54–55) in Robertson, Law and Haertel (2010: 245–246)).

	Transcription	Transliteration[167]	Translation
"presente" I	AI[168] <yual[169] inpaɛxiel> AII <Yual in pacxiel>	*i-wal in-pakxi-el*[170]	'I return (now)' (<actualmente me buelbo>)
"preterito"	AI <paɛxien> AII <pacxien>	*pakxi-en*	'I returned' (no translation in manuscript, just "preterito")
"futuro"	AI <xpaɛxic en> AII <xpacxicen>	*x-pakxi-k-en*	'I will return' (no translation in manuscript, just "futuro")
"futuro en -rus"[171]	AI <apaɛxiel en> AII <apacxielen>	***a** pakxi-el-en*	'I must return' (AI <tengo de uol//berme> / AII <tengo de bolberme>)
"presente" II	AI <auixieˇ>[172] AII <a Vixi en>[173]	***a** bixi-en*	'I go' (AI <uoi> / AII <boi>)

[165] Intransitive verbs are called "neuter verbs [Arte I: <uerbos neutros>]" (Morán (1695: 29) in Robertson, Law and Haertel (2010: 244)).

[166] I only translate the categories once and continue to use the Spanish names in quotation marks in my discussion because they are not accurate descriptions of the phenomena – Choltí very likely shows aspect, not tense marking (aside from future) and therefore there is no "present" or "preterite". By simply employing the manuscript's terminology, I try to stay as neutral as possible when it comes to my own interpretation. The manuscript has two formations called "presente". To distinguish them, I use Roman numerals I and II.

[167] I always use my own transliteration and glossing unless stated otherwise.

[168] I give both versions (Arte I and Arte II) whenever they are not identical.

[169] <Yual> is more often written as <iual> in Arte I, though there is variation with <yual>. Arte II uses <yual> exclusively. Robertson, Law and Haertel (2010: 169–170) transliterate this as *iyuwal*, which I consider inaccurate. With *iwal*, I stay closer to what is written in the source.

[170] Deriving the status of the pronoun sets as affixes, clitics or independent elements from the MM is difficult. I treat both set A and set B as affixes based on the situation in most Mayan languages but this is not to be taken as a definite statement, rather a unifying convention.

[171] Arte I: <futuro en Ruz>, Arte II <futuro en rrus> (Morán (1695: 32, 55) in Robertson, Law and Haertel (2010: 246)).

[172] The sign <ˇ> is often used instead of <n> in Arte I.

The MM notes one "curiosity" that only concerns intransitive verbs, namely that in the "preterito" and "futuro", these are inflected with set B suffixes while for the "presente" I, set A prefixes are used.[174] The three paradigms are contrasted in Table 15:

Table 15. Full paradigms for 'to go' from Arte II (Morán (1695: 54) in Robertson, Law and Haertel (2010: 246)).

	"Presente" I	"Preterite"	"Future"
1SG	<Yual in Vixnel>[175] 'now I go/am going [actualmente me boi]'	<Vixi en> 'I went [yo fui]'	<xchicen>
2SG	<Yual a Vixnel>	<Vixi et>	<xchicet>
3SG	<Yual u Vixnel>	<Vixi Po. [=Pedro]>	<xchic Po.>
1PL	<Yual ca Vixnel>	<Vixion>	(no example given)
2PL	<Yual i Vixnel>	<Vixiox>	(no example given)
3PL	<Yual u Vixnelob>	<Vixiob>	(no example given)

The parallels to aspect-based split ergativity as discussed in chapter 3.3 for, e.g., Yucatecan, are striking. In the "preterito", which is likely a completive aspect, the construction consisting of the verb stem and a set B suffix in (1) is the same as in Yucatec in (2).[176]

[173] I use the verb *pakxi* to illustrate the paradigm because unlike 'to go', it does not have suppletive forms. However, a "presente" II form of *pakxi* is unattested. It would have been **a pakxi-en.

[174] Arte I: <Nota una // curiosidad para todos los uerbos neu//tros […] y es que en el preterito // y futuro la declinõn es por los casos // y personas posponiendo Siempre los // pronombres primitiuos y en el pnte. ante//poniendo los posesiuos.> (Morán (1695: 31) in Robertson, Law and Haertel (2010: 245–246))

Arte II: <Nota una curiosidad para todos los Veruos ne//utros […] Y es que // en preterito i futuro. se declinan por los casos y // personas posponiendo siempre los pronom//bres primitiuos. Y en el presente anteponiendo, las particulas que significan posesion. in. a. u.> (Morán (1695: 54) in Robertson, Law and Haertel (2010: 245–246))

"Note one curiosity for all neuter verbs and it is that in the preterite and future, the declension for case and person is by suffixing always the primitive pronouns and in the present by prefixing the possessives (the particles that indicate possession. *in. a. u.*)."

[175] Arte I only provides one example for present <iual inuix//nel> and only the first three persons of the preterite paradigm: <uixien Yo fui Vixi et // Vxi Pedro>. Future forms are written as follows: <xɛhiɛen Yo ire xɛhi//cet xɛhiɛ Po.> (Morán (1695: 31) in Robertson, Law and Haertel (2010: 246)).

[176] Minor differences include the sound correspondence Yucatec *ch* ~ Cholti *t* as well as the fact that in Yucatec, the bare root of the verb *jóok'* is used as a stem while the verb in Cholti contains an additional vocalic stem element *-i*.

(1) Choltí "preterito"/completive intransitive (Morán (1695: 54) in Robertson, Law and Haertel (2010: 246))

<vixi et>

b'ixi-et

go-B2SG

'you went'

(2) Yucatec Maya completive intransitive (Hofling 2017: 711)

(j) jóok'-ech

(COM) go.out-B2SG

'you went out'

If we contrast the transitive verbs, Choltí and Yucatec differ more (Yucatec Maya has an aspect marker *t-* and a different status suffix) but the argument marking is the same: set A is used to mark A and set B is used to mark O (unmarked B3 in Choltí). Examples (1) and (3) show that the alignment in the completive aspect in Choltí is ergative-absolutive just as in Yucatec in (2) and (4) because S and O are marked by set B and A by set A.

(3) Choltí "preterito"/completive transitive (Morán (1695: 17, 45) in Robertson, Law and Haertel (2010: 233))

AI: <aɛale auotot>, AII: <acale a Votot>

a-kal-e *aw-otot*

A2SG-make.B3-TS A2SG-house

'You made your house. [Span. AII: <hisiste tu casa>]'

(4) Yucatec Maya completive transitive (Hofling 2017: 710)

t-aw-il-aj-en

COM-A2SG-see-TS-B1SG

'you saw me'

Turning now to "presente" I in Choltí in (5), best understood as an incompletive or progressive[177], we again see parallels to Yucatec in (6). Though the aspect markers are not the

[177] Robertson, Law and Haertel translate the present of the MM as a progressive construction, but this is not necessarily the correct interpretation. I find it striking that, although the progressive exists as a category in Spanish, the MM authors chose not to translate the construction with <yual> this way. Instead, they add a meaning of 'now/at the moment [Span. actualmente]' and in the vocabulary list, <yual> is glossed with 'actualidad' (Robertson, Law & Haertel 2010: 294). This could imply a progressive or durative meaning, but it is equally possible that we are dealing with a more general present. However, since there are two

same (YUC *k-* versus CHT <yual>), the basic construction is identical. An aspect marker is used with an intransitive verb that is prefixed with set A and receives a nominalizing suffix (YUC *-Vl* ~ CHT *-el*). Again, Cholti appears to have an additional stem-forming element *-n-*, but the strategies behind the constructions are otherwise parallel.

(5) Cholti incompletive intransitive (Morán (1695: 54) in Robertson, Law and Haertel (2010: 246))

<Yual a Vixnel>

iwal *a-b'ix-n-el*

"actualidad" A2SG-go-IS-NMLZ

'You go now. [Span. <actualmente te bas>]'

(6) Yucatec Maya incompletive intransitive (Hofling 2017: 711)

k-a-jóok'-ol

INC-A2SG-go.out-IS/NMLZ

'you go out'

If we now contrast the marking of transitive verbs with that of intransitives in the incompletive, we see that the alignment is nominative-accusative both in Cholti in examples (5) and (7) and in Yucatec Maya in examples (6) and (8): A and S are marked alike (set A) and O separately (set B).

(7) Cholti incompletive transitive (Morán (1695: 181, 44) in Robertson, Law and Haertel (2010: 231))

<yual achohben>

iwal *a-chohben*

"actualidad" A2SG-love.B3

(8) Yucatec Maya incompletive transitive (Hofling 2017: 710)

k-aw-il-ik-en

INC-A2SG-see-IS-B1SG

'you see me'

I will not discuss the "futuro" because it is not relevant to this discussion and instead skip to the other two categories, "presente" II and "futuro en -rus". The "presente" II displayed in (9)

"presente" constructions in the language, the analysis of the first as progressive and the second as a more general present is attractive.

only exists for intransitive verbs. For transitive verbs, only the formation with the particle *iwal* is mentioned (Morán (1695: 17–18a, 44) in Robertson, Law and Haertel (2010: 231)).

(9) Choltí "presente" II (Morán (1695: 32, 55) in Robertson, Law and Haertel (2010: 246))
AI: <auixie˘>, AII: <a Vixi en>
a *b'ixi-en*
? go-B1SG
'I go [Span.: AI: <uoi>, AII: <boi>]'

The particle *a* is used to "form present from preterite".[178] If we compare (9) to (1) above, we see that the verb form is the same. The only difference consists in the preverbal particle *a*. The difference between "presente" I and II seems to be the notion of "actualidad" that accompanies "presente" I, whereas II refers to a more general present like 'I go' in (9).

Finally, there is the "futuro en -rus" which is formed by putting *a* in front of a verb form with the suffix *-el* followed by set B[179] as in (10):

(10) Choltí "futuro en -rus" (Morán (1695: 32, 55) in Robertson, Law and Haertel (2010: 246))
AI: <auixnel on ti manche > / AII <a Vixnel on ti manche>
a *b'ix-n-el-on* *ti* *manche*
? go-IS-NMLZ-B1PL PREP Manché
'we shall go to Manché [<emos de ir al manche>]

To my knowledge, it is not stated anywhere explicitly where the name of this future formation comes from, so it is worth mentioning here. "Futuro en -rus" is a label used in Spanish late

[178] Arte I: <tambien con esta a se haçe // de preterito presente […] Y lo mesmo haçen los preteritos de // los Verbos pasiuos.> (Morán (1695: 32) in Robertson, Law and Haertel (2010: 246))
Arte II: <tambien con esta. a. se hace de preterito. presen//te […] y lo mismo ha//se en los preteritos de los Veruos pasibos> (Morán (1695: 55) in Robertson, Law and Haertel (2010: 246))
'Also with this a is made from preterite present […] and the same do (AII: it does in) the preterites of the passive verbs.'

[179] Arte I: <Nota q˘el presente de los Verbos neutros // se haçe futuro en Ruz anteponiendole // una a.> (Morán (1695: 32) in Robertson, Law and Haertel (2010: 246))
Arte II: <Nota mas que el presente de los Veruos neutros se hase // futuro en rrus anteponiendole una. a. y posponiendo las particulas en, et, etc. de los pronobs. primitibos> (Morán (1695: 55) in Robertson, Law and Haertel (2010: 246))
'Note (further) that the present of the neuter verbs is turned into a 'futuro en -rus' by prefixing an a. +(and suffixing the particles en, et, etc. of the primitive pronouns'.

medieval[180] sources for Latin future participles (Baldischwieler 2004: II, 169). The function of these participles when used with the copula is described as follows and illustrated in example (11):

> From Early Latin onwards, combinations of the future participle in *-urus* and present indicative forms of the verb *sum* ('to be') are used to assert that someone is at this very moment about to or fated to do something and, in the case of controlled events, has the intention and determination to do or not to do something. With respect to this value of immediacy, fate, or intention, the *-urus* + *sum* expression differs from the simple future, which only asserts that something will take place. Some scholars speak of the 'modal' value of the periphrastic expression. In this syntax the term PROSPECTIVE will be used instead. (Pinkster 2015: 429–430)

(11) Latin periphrastic future with future participle (Plautus *Pseudolus* (105–106) in Pinkster (2015: 430))

Atque	*id*	*futurum*	*undeunde*
and	DEM.NOM.SG.N	be.PTCP.FUT.ACT.NOM.SG.N	wherever.from
dicam		*nescio, /*	*nisi* *quia*
say.1SG.PRS.SBJV.ACT		not.know.1SG.PRS.IND.ACT	except that
futurum		*est*	
be.PTCP.FUT.ACT.NOM.SG.N		be.3SG.PRS.IND.ACT	

'And I don't know where I should say it'll come from, except that come it will.'

It is not used "to express assumptive and deductive epistemic modality, nor is it used in declarative sentences with a directive illocutionary force" (Pinkster 2015: 431) – that is the realm of the simple future. However, the construction is used in "interrogative sentences with a binding directive illocutionary force" (Pinkster 2015: 431), e.g., (12):

(12) Latin periphrastic future with future participle interrogative use (Plautus *Poenulus* (432) in Pinkster (2015: 431))

Abiturus-n'	*es?*
go.away-PTCP.FUT.ACT.NOM.SG.M-Q.NEG	be.2SG.PRS.IND.ACT

'Will you not go away?'

According to Pinkster (2015: 431), this latter use in questions, direct and indirect, is especially common because they often address the intentions of the person they are directed at.

Knowing this is important to be able to understand the function of the form that is described in the MM. Since the Latin periphrastic future with the future active participle has a modal connotation – not unlike the English *shall* or Icelandic *skulu* – and since the authors of the

[180] Baldischwieler (2004) examines a source from 1486.

manuscript found this term fitting to describe the use of the Choltí "futuro en -rus", we can assume that the latter also has some modal connotations and this is what sets it apart from the other future in Choltí that is formed with a prefix *x*- (Robertson, Law & Haertel 2010: 173, 175).

To sum up, according to the MM, the particle *a* can both

1. form present from preterite as stated above as well as
2. turn a "presente" I into a "futuro en -rus" <iual in pacxiel> 'I return [<yo me buelbo>]' <apacxielen> 'I need to return [<tengo de bolberme>]'.

This exact phrasing of a form "turned into" another (<el presente […] se hase // futuro en rus>, see fn. 179) is misleading because it implies a relationship between them, which is a base for Robertson's argument in the following section. This is especially unfortunate with the "futuro en -rus" because it differs from the "presente" I not only in the tense/aspect marking but also in the employed index-set. The use of set B in the "futuro en -rus" in (10) implies that the nominalized form in *-el* is used as a non-verbal predicate in the Mayan stative construction 'I (am) returning' (similar to 'I am a teacher/small/etc.'). It does not involve possession. On the other hand, the use of set A in "presente" I in (5) implies a possessed nominalized complement ('now (is) my returning') parallel to the cases discussed in chapter 3.3 for Yucatecan, Western Cholan and Poqom. The two categories are therefore best understood as different constructions that involve a nominalized verb in *-el* instead of a form that is somehow transformed and becomes another.

To conclude, the preverbal particle *a* that appears in Choltí is used with forms inflected with set B and does not involve any kind of possessive morphology. As to its syntactic status, it is sometimes written as a prefix, sometimes separately. It might be a clitic, but this is difficult to ascertain from the single source available to us. At least no other elements seem to occur between it and the verb.

With this background we can now turn to the evaluation of Robertson's hypothesis.

5.1.2 Clash of paradigms

Robertson proposes to explain set C as the outcome of the "collapse of the two INCOMPLETIVES" (Robertson 1998: 8) of Cholti[181] based on the following observations (Robertson 1998: 6):

1. The Cholti particle *a* resembles CHR C3 *a-*.
2. The use of set C in Ch'orti' corresponds to the pattern in Mayan languages with aspect-based split ergativity where intransitive verbs are used with set A in the incompletive but set B in the completive[182] while no such strategy exists for transitive verbs. Set C in Ch'orti' fulfills the same function as set A does in the Lowland Mayan split: it marks S on intransitive verbs in the incompletive aspect, while set B is used to do the same in the completive. Therefore, set C must have originated in set A.
3. There are two kinds of incompletive paradigms documented in the colonial grammar of Cholti but only one incompletive in Ch'orti': "Morán's Ch'olti' grammar describes two types of the INCOMPLETIVE for intransitive verbs. These two paradigms apparently fell together, yielding the hybrid pronominal set attested in modern Ch'orti'" (Robertson 1998: 6).

Robertson's understanding of the function of the Cholti particle *a* as having a "temporalizing effect of moving the 'préterito' [sic] to the 'presente', and the 'presente' to the 'futuro-en-ruz'" (Robertson 1998: 6–7)[183] is perhaps based on the misleading wording in the MM

[181] Robertson makes the strong claim that set C can **only** be derived from the verbal system that is described for Cholti: "One of the unresolved linguistic questions in Mayan linguistics is the fact that Ch'orti' has three pronominal sets, whereas Mayan languages generally have only two such sets. The provenance of Ch'orti's newly formed, third pronominal set has never been successfully explained. It is precisely the explanation, given herewith, that secures the Ch'olti'-Ch'orti' relationship." (Robertson 1998: 5). And, even more strongly phrased: "A genuine understanding of where this innovative, aspectual pronominal series came from depends crucially on Ch'olti' grammar; it alone contains the clarifying information necessary to explain the new Ch'orti' pronouns." (Robertson 1998: 6).

[182] Note that Robertson applies the labels "ergative" and "absolutive pronominal set" instead of the more neutral labels set A and set B, but this is avoided here for reasons explained earlier.

[183] Robertson further states that this is "surely an innovation" of Cholti, "but it is not far removed semantically from its use in Acalán Ch'ontal [sic], where the translation of *a* + VERB into Spanish is almost invariably present perfect" (Robertson 1998: 7). The parallel between the semantics in Chontal and the semantics ascribed to the Cholti *a* by Robertson is that "present perfect has an effect similar to what has already been described: bringing the effect of a past action into the present, where it is currently felt" (Robertson 1998: 7). The formation in Chontal will be discussed in chapter 9.

discussed in the previous section.[184] However, he presents a simplified view of the situation in Choltí. As we have seen above, saying that "the secondary incompletive comes from *a-* prefixed to the completive, while the secondary, 'futuro en ruz' comes from the very same *a-* prefixed to the incompletive" (Robertson 1998: 6) is not the whole picture. It works for the first pair, which only differs in the absence or presence of *a*:

"Preterito"/completive *b'ixi-et* "Presente" II *a b'ixi-et*

It does not, however, work for the second pair because the "presente" I and the "futuro en -rus" are different constructions that simply share a verbal noun in *-el*. Deriving one from the other requires a motivation for the use of set B instead of set A as well as the loss of *iwal*. Assuming that they are not connected to each other historically, as I argued, is the more elegant solution.

"Presente I"/incompletive *iwal in-b'ix-n-el* "Futuro en -rus" *a b'ix-n-el-on*

This calls into question the proposed unified function of *a*. Additionally, if *a* were completely grammaticalized, one would expect a clearer semantic difference between the two "presente" forms described in the MM. However, Robertson himself admits that "the semantic difference […] is difficult to assess completely", though he speculates that "it seems to be the difference between definiteness versus an indefinite statement of some general truth (e.g., 'water boils at 212 degrees')" (Robertson 1998: 7). The distribution of the two forms in the text is such that the present with <yual> is "almost always" (Robertson 1998: 7) – but not exclusively – translated as 'actualmente', whereas the present with <a> seems to express gnomic truths.

Taking all this into account, it becomes clear that the first step of Robertson's (1998: 8) account of the genesis of set C from the "two incompletives" in Choltí simply would not have happened:

1. The prefix *a-* in the incompletive (i.e., "presente" II) was reinterpreted as a morpheme for the third person, which is originally unmarked (in set B). This created C3 *a-* but led to an "unacceptable homonymy" with the A2SG *a-*. See Table 16:

[184] Robertson (1998: 7) explicitly states that Morán's analysis of deriving the second future from the first incompletive (i.e., first present) is correct.

Table 16. Choltí > Ch'orti': step 1. Color marks "homonymous" forms.

	Set A	Set B	Set C
1SG	*in(w)-*	*-en*	
2SG	*a(w)-*	*-et*	
3SG	*u(y)-*	*ø*	*a-*
1PL	*ka(w)-*	*-on*	
2PL	*i(w)-*	*-ox*	
3PL	*u(y)-…-(ob')*	*-(ob')*	

Robertson suggests that a reanalysis of *a* as a morpheme of the third person in "presente" II forms like *a b'ixi* but this is unlikely to have happened because *a* is synchronically also used with other persons in forms like *a b'ixi-en* in example (9) above.[185] Had the reanalysis happened in set B after all, there would still not have been homonymy with a form of set A because the formations differ in other respects, too: *a-b'ixi* (3SG "presente" II) is not

[185] This, among many other points is also what Quizar (2023) criticizes about Robertson's theory. She points out that Choltí *a* does not behave like the historical incompletive in Ch'orti' because it can occur on intransitive verbs with other absolutive markers but "[f]or the Ch'olti' incompletive (habitual) proclitic *a* to become the Ch'orti' prefix *a-*, the proclitic must be limited to third person only" (Quizar 2023: 271).
In Robertson and Law (2009), the theory concerning the origin of concerning set C from Robertson (1998) is discussed again and reaffirmed. There, the authors claim that they found traces of a Choltí-like *a* used in Ch'orti' with other persons than the third. They quote "*war a-in-boroh ni maxtak* 'I am increasing in family (said by a man with a pregnant wife); *sahmi a-in-xin in-chonoh* 'today I shall sell [lit., 'today I go, I sell']" (Robertson & Law 2009: 310) from Wisdom's data where *a* clearly appears together with C1SG *in-*. This would indeed be significant if that were what Wisdom wrote. The authors must have used Stross's (1992: 25, 49) transcription because the mistake originates there. Scans of Wisdom's (1950: 589, 706) original show that he wrote <war inp'oroh nimaštak> and <sahmi iinšin inčonoh>. The long *i* in the second case is likely Wisdom's way of recording the phenomenon discussed in chapter 6.3. In any case, there is no *a* in the original manuscript though based on Wisdom's handwriting it is understandable why the confusion happened.
The third piece of evidence for the author's claim is from Galindo (1834) who recorded <aingüe> 'to eat' (Robertson & Law 2009: 310–311). However, Galindo only records four verbs (see Appendix B), all of them in the first person, and the other three do not show an <a> although they are also intransitive (except for *u'nchi* which could be intransitive or transitive based on morphology alone):
<unchi> 'to drink' (*u'nchi*; for this special development see again chapter 6.3)
<inguaian> 'to sleep' (*inwayan*)
<inchamai> 'to die' (*inchamay*).
In addition to that, *inwayan* (C1SG *in-* + *wayan* 'AP12 to sleep') and *inchamay* (C1SG *in-* + *chamay* 'MPAS4 to die') seem to be evidence for the fact that set C already existed at this point, which makes it even less likely for the *a* to be identical to the Choltí one. In any case, it is impossible to use this single instance of an unclear *a* to argue the point that Robertson and Law try to make, namely that Ch'orti' used to use *a* with other grammatical persons, as well.

homonymous to *iwal a-b'ixnel* (2SG "presente" I). The fact that the morphemes are homophonous does not suffice for confusion and reanalysis to occur.

Still, let us assume that it did happen to examine the following arguments. This is the next step (Robertson 1998: 8):

2. To avoid the homonymy, A/C2PL *i-* replaced A/C2SG *a-*. Unfortunately, this created another homonymy between A/C2PL and A2SG. See Table 17:

Table 17. Choltí > Ch'orti': step 2. Color marks "homonymous" forms.

	Set A	Set B	Set C
1SG	*in(w)-*	*-en*	
2SG	*a(w)-*	*-et*	*i-*
3SG	*u(y)-*	*ø*	*a-*
1PL	*ka(w)-*	*-on*	
2PL	*i(w)-*	*-ox*	
3PL	*u(y)-…-(ob')*	*-(ob')*	

Homonymy avoidance is not a strong argument. Though some studies argue that it does play a role in language change (see, e.g., Baerman 2011; De Smet & Rosseel 2021) there are likewise many cases of homonymous person markers in the languages of the world, e.g., in Icelandic (2/3SG) and Lithuanian (3SG/PL) and this does not seem to bother the speakers. There are even examples of this for Mayan as many Mayan languages, e.g., Yucatec, differentiate between A/B3SG and A/B3PL with the general plural suffix *-ob'*, but the suffix is often optional – even in Ch'orti' itself. Additionally, I find it difficult to imagine that speakers would resort to using A2PL *i(w)-* to avoid the homophony of *a(w)-* even though this form then leads to another homophony – this would completely counter the supposed purpose of the change. In theory, they would have more likely adopted the suffix *-ob'* to differentiate A2SG and A2PL as just described. As Baerman (2011: 1) aptly states: "In most cases there is no way to resolve the question, since the assumption that something is being avoided is itself a theoretical construct." I would add that it implies that language change has a teleological component, a view that I reject in my thesis.

The next step in Robertson's theory is another homonymy avoidance:

3. To avoid this new homonymy, A2PL borrowed a *-x* from B2PL *-ox* and thus the form A2PL became *ix-*. See Table 18:

Table 18. Choltí > Ch'orti': step 3.

	Set A	Set B	Set C
1SG	*in(w)-*	*-en*	
2SG	*a(w)-*	*-et*	*i-*
3SG	*u(y)-*	*ø*	*a-*
1PL	*ka(w)-*	*-on*	
2PL	*i(w)-*	*-ox*	*ix-*
3PL	*u(y)-…-(ob')*	*-(ob')*	

There is an ongoing debate whether morphological analogy replaces whole forms or morphemes (e.g., Hill 2020). However, what nobody is arguing for is the fact that morphemes can somehow suck up segments from other morphemes and fuse them with their existing material. Much remains mysterious about morphological analogy, but, surely, there is consensus that this is not how it works. It implies an ability to directly manipulate individual sounds for morphological reasons that speakers likely do not have. The *-x* that is supposed to have been added is, after all, not a known plural morpheme but just a random sound from an unrelated bound person form. This is even less plausible because B2PL is already *-ox* in Choltí, not *-ex/-ix, as it is reconstructed for Proto-Mayan (Kaufman 2015: 161). That means that A2PL *i-* would have received a segment *-x* from a form with which it does not even share the vowel.

To finish, Robertson (1998: 8) discusses the remaining persons and the "collapse" of the "two incompletives" of Choltí:

4. The C1SG and C1PL correspond to A1SG and A1PL respectively, so no explanation is needed there. Likewise, C3PL is just C3 with the common plural suffix. See Table 19:

Table 19. Choltí > Ch'orti': step 4.

	Set A	Set B	Set C
1SG	*in(w)-*	*-en*	*in-*
2SG	*a(w)-*	*-et*	*i-*
3SG	*u(y)-*	*ø*	*a-*
1PL	*ka(w)-*	*-on*	*ka-*
2PL	*i(w)-*	*-ox*	*ix-*
3PL	*u(y)-…-(ob')*	*-(ob')*	*a-…-(ob')*

5. And finally: "[T]he old nominalizing form *-el* […] was lost with the collapse of the two incompletives, and furthermore split-ergativity was marked simply by the bare, newly-devised pronominal set, with no trace of the historical nominalizer" (Robertson 1998: 8).

Needless to say, the final step, which essentially consists of "and then it disappeared", does not suffice as an explanation for the loss of the nominalizer.

To sum up:

1. There is no context where the reanalysis of *a* as a third person morpheme could have happened.
2. If it had happened then there would not have been homophony with an index from set A as these are separate paradigms without overlapping uses.
3. If the forms from different paradigms would have been perceived as homonymous, the homonymy would not have necessarily been eliminated.
4. If it had been eliminated, it would be rather peculiar if a mechanism had been employed that led to another homonymy.
5. If homonymy had been eliminated a second time, there is no mechanism in language change that could account for the modification of C2PL from *i- to *ix-*.
6. If the modification had happened nonetheless, still an ad hoc assumption of spontaneous loss is necessary to explain the missing nominalizing suffix.

Additionally, Robertson does not explain 1) why the resulting forms do not likewise use prevocalic glides like virtually all set A paradigms do and 2) how it is possible that set A only developed these specific new forms in the incompletive aspect of intransitive verbs but not in all other domains where set A is used. If a paradigm develops several variants, a plausible scenario must be provided for why that happened. As we see in other languages with aspect-based split ergativity, set A is never differentiated in that way.

For further criticism of this paper, though with a different emphasis, see Quizar (2023) who develops her own theory for the origin of set C. This will be discussed in section 5.3. In the next section I will first address a second explanation that is very similar to Robertson's.

5.2 Wichmann (1999)

The second attempt at explaining set C is a brief discussion in a footnote in Wichmann's unpublished sketch of Ch'orti' morphology, introduced with the words "I currently work with the idea […]" (Wichmann 1999: 20–21, fn. 1). Though it is similar to Robertson' proposal, it has apparently been developed independently. I will only examine in which ways Wichmann's proposal differs from Robertson's. He, too, assumes that set C developed out of set A, but in his view the changes were triggered by the aspect marker *war* which is used to form the progressive aspect. Thus, he assumes the exact same development in Ch'orti' as in the other Mayan languages with aspect-based split ergativity: A new construction involving an aspect marker arose and this caused set A to also be used on intransitive verbs. These are the steps that set A went through to become set C according to Wichmann:

1. The base for all change is the Ch'orti' progressive that is formed with the aspect particle *war*. A3SG then changed from *war u-* to *war a-* due to vowel harmony. Thus, C3SG *a-* arose. A/C3PL changed accordingly. It is only logical according to Wichmann (1999: 20, fn. 1) that "it is the most frequently used form that changes first".

2. Further changes happen to avoid homophony:

    a. A2SG *war a-* changes to C2SG *war i-* to avoid homophony with the C3SG. This happens through an "analogy" with A2PL *i-*.

    b. A2PL *war i-* changes to C2SG *war ix-* to avoid homophony with the C2SG. An additional *-x* is added. Although Wichmann does not elaborate on this, it can be assumed that he also motivates this with B2PL just like Robertson does.

3. At some point, C1SG developed from a metathesis of A1SG *war ni-* > *war in-*. Wichmann does not situate this change in the relative chronology of the others.

4. A/C1PL *war ka-* did not change and thus does not require an explanation.

Since the two theories are so similar, the criticism brought forward against Robertson's proposal also applies here. The only thing that Wichmann's theory explains better is why the change was triggered in the first place: He clearly states that it is tied to the aspect marker *war*. However, as we have seen in chapter 4.3, aspect marking via set B/set C is independent from aspect marking with particles like *war*, so it is not really convincing to attribute the genesis of set C to the presence of the aspect marker.

Wichmann does not comment on the lack of nominalizer at all. If he indeed assumes a construction parallel to split ergativity in Yucatecan or Western Cholan – and it seems like he does – then either the lack of a nominalizer or the use of set A without a nominal form requires an explanation. Additionally, the change from *u-* to *a-* in C3SG that he ascribes to vowel harmony seems somewhat ad hoc. This does not mean that it could not have happened, just that it is a shaky foundation to build a theory on. For the metathesis of *ni-* to *in-*, he provides a parallel of the causative suffix, which is mostly *-se* but sometimes *-es*, although he states that this still needs to be investigated. Judging from what is so far known about the variation *-se* ~ *-es* (on that cf. chapter b.7.6 in Appendix E), it is purely phonological and thus one would expect a productive variation of *ni-* ~ *in-* in the language today. However, no verb forms with *ni-* are attested, especially not for set C. In fact, a better explanation would have been available: As I described in chapter 4.2.1, Ch'orti' shares the Mayan phenomenon of *ni-* ~ *in-* variation in the first person singular and there is consensus that *in* comes from B1SG.

To conclude, the proposal discussed here is just as unsatisfying as Robertson's though it perhaps deserves more lenience as it is just a footnote in a manuscript. It is worth noting that even though Wichmann envisions a similar scenario for the development of set C as Robertson does and in fact he later states that he finds Robertson's account "entirely convincing"[186] (Wichmann 2002: 3), he rejects the strong claim that Ch'orti' is necessarily the daughter of Cholti, "when the only requirement for the scenario to work is to assume that the ancestor of Ch'orti' is *like* Ch'olti' in having a preverbal *a* incompletive marker".

5.3 Quizar (2023)

A different origin for C3 *a-* has been proposed very recently in Quizar (2023) who observed that a neighbor of Ch'orti', Xinka, shares both the form *a-* for the third person and the aspect marking pattern of prefixation vs. suffixation of indexes. I will first present the language family, then summarize what we know about contact between Xinka and Mayan and afterwards lay out the theory as proposed by Quizar and discuss it.

[186] Whether this means that he finds Robertson's account more convincing than his own and has abandoned the idea based on *war* remains unclear.

5.3.1 Xinkan

Xinkan is a small, practically extinct language family spoken in the southeast of Guatemala at the border to El Salvador (Sachse 2010: 29). It therefore occupies an area that would have allowed for long-term contact with Ch'orti'. So far, no affiliation of Xinkan to other languages or language families could be established, although many have been proposed (Sachse 2010: 49). Therefore, Xinkan and Mayan languages are most likely unrelated and thus any similarity between them must be due to chance or contact.

The goal of Quizar's publication is not only to offer an explanation for the peculiar similarity observed between Xinkan and Ch'orti' but also to provide "a plausible resolution to the current controversy regarding the linguistic relationship between Ch'orti' and Ch'olti'" (Quizar 2023: 256) and to ultimately show that the verbal system of Ch'orti' cannot be explained from one that was similar to the one in Choltí, contrary to claims by Robertson (e.g., 1998).

5.3.2 Mayan-Xinkan Contact

Mayan loans in Xinkan are extensively discussed in Campbell (1972; 1976; 1978; 1984; 1997). Key findings from these studies are summarized in Quizar (2023: 259):

- There is strong evidence of language contact before the colonial period with more than 135 Mayan loanwords in Xinkan having been identified that belong to various semantic fields like agriculture, food preparation, commerce, religion, animals, plants, disease and material culture.
- There is evidence both for early loans and later loans, which suggests long-term contact.
- Campbell (1984: 8–9) identifies many loans to come specifically from Cholan or at least Greater Tseltalan.
- Considering the number and type of loans, at least some of the contact must have occurred during the Classic Mayan period where Cholan culture is assumed to have been at its height and "the center of a highly complex culture" (Quizar 2023: 259).
- Numerous toponyms that are possibly of Xinkan origin overlap with the Ch'orti' region: according to Campbell (1997: 90) "from the Motagua Valley in the north to the Pacific Ocean in the south and from near Guatemala City in the west to Honduras

and El Salvador in the east" (Quizar 2023: 259), which supports the idea of Xinkan-Ch'orti' contact.[187]

- There is no evidence of significant Ch'orti'-Xinkan contact after the conquest.

The most important conclusion, however, is the following:

> "As of the present, no loans from Xinkan have been definitively identified in Mayan. The unidirectionality suggests that Xinkan speakers were bilingual in Mayan language(s) but that Mayans were not learning Xinkan." (Quizar 2023: 259)

Therefore, it would be unlikely for Ch'orti' to have borrowed anything from Xinkan – one would expect borrowing to go the other way around. Additionally, if there were indeed proof of Xinkan influence on the index-set and aspect marking of Ch'orti', this would have occurred before colonial times making a descent of Ch'orti' from Choltí impossible because set C does not exist in Choltí (Quizar 2023: 256).[188]

5.3.3 Set C as a contact phenomenon

Quizar's paper addresses many of the wrong analyses made by Robertson and colleagues[189] but I will here focus on her ideas concerning set C and aspect marking on intransitive verbs and how these dispute Robertson's theory.

As a starting point, Quizar (2023: 267) assumes that the aspectual split of Ch'orti' used to be between ergative-absolutive and nominative-accusative alignment just like in Western Cholan, not between ergative-absolutive and tripartite as today. She further assumes that Ch'orti' used to have the exact same "redundant" aspect marking that is present in Chol and Chontal. What she means by "redundant" is that the languages indicate aspect through affixes on the one hand but also through the use of set A versus set B on intransitive verbs, which in

[187] Note, however, the discussion in Metz, McNeil and Hull (2009) and in chapter 2.2.5.1 of what exactly constitutes the Ch'orti' region and how the assumption of continuity is problematic.

[188] Contrary to what has been claimed by Fought (1984: 49–50) and Storniolo (2008: 82–100). Storniolo (2008: 220–222, 238) also claims to have found evidence for set C in the hieroglyphic corpus of Copan, but her few examples are not convincing.

[189] Examples include their claims that Ch'orti' *mix*, one of several ways to express negation, shows a reflex of the Choltí future marker *x-* (Robertson 1998: 8; Robertson & Law 2009: 306–307). Quizar (2023: 262) rather analyzes this as *ma* + *ix* 'NEG + already = no longer, not yet, not', which had already been proposed by Wichmann (2002: 4) and Vinogradov (2016: 63) and possibly others. Though the expected form for such a contraction would theoretically be *me'yx or at least *mi'x based on the data discussed in chapter 6, *=ix* often does not show laryngealization when it is expected (see footnote 201). Quizar further argues that the suffixes *-n* and *-ik* that Robertson and Law (2009: 306) identify as future tense are better viewed as optative.

theory would be unnecessary. This "redundancy" is the base of her theory.[190] Quizar assumes that set C then replaced set A in the incompletive aspect of intransitive verbs creating a tripartite instead of nominative-accusative alignment in this aspect. Tense/aspect markers were lost and only the portmanteau use of person markers remained as the "no longer redundant but now the primary and only" (Quizar 2023: 267) indication of aspect for intransitive verbs. Transitive verbs, on the other hand, lost aspect marking altogether because no special aspectually conditioned index-set was developed for them. She claims that this development was facilitated by contact with speakers of Xinkan. Let us therefore first examine how aspect is marked on Xinkan verbs and then proceed to the ways in which this could have influenced the development of Ch'orti'.

Xinkan verbs also mark tense/aspect "redundantly" – as seen by Quizar – "by changes in the verb stem involving glottalization and vowel length, plus a further redundant marker in a verb class suffix" (Quizar 2023: 268). The following examples are from Guazacapán Xinka, but the other Xinkan varieties follow this pattern, as well (Rogers 2016: 233–237). Two classes of intransitive verbs are distinguished by Rogers (2016: 89, 93)[191], an agentive one where the subject controls the action and an affective one where the subject is the undergoer. In both classes, the third person is marked through *a-* in the incompletive as in examples (13) and (15) and unmarked in the completive in (14) and (16).

(13) Incompletive agentive (Rogers 2016: 233)

a-yan'a

3SG(D).INC-be.ashamed.AGT.INC

's/he is ashamed'

(14) Completive agentive (Rogers 2016: 233)

ø-yana-lha'

3SG(C).COM-be.ashamed.AGT.COM-AGT.COM

's/he was ashamed'

[190] As to Choltí, she argues that it "no longer had a clear and distinct incompletive-completive contrast, and […] was replacing its aspectual system with a tense system" (Quizar 2023: 267) thus accepting the analysis that Robertson, Law and Haertel (2010: 169–178) offer for Choltí.

[191] It should be noted that Sachse (2010: 554) analyzes intransitive verbs differently: she assumes that the glottal stop at the end is an aspectual suffix that marks the completive and that incompletive verbs have no aspectual marking, while Rogers regards the glottal stop as part of the root and views *-lha'* as a verb class suffix (Quizar 2023: 272, fn. 7).

(15) Incompletive affective (Rogers 2016: 233)

a-saka'

3SG(D).INC-be.lifted.AFV.INC

's/he is lifted'

(16) Completive affective (Rogers 2016: 233)

ø-saaka'

3SG(C).COM-be.lifted.AFV.COM

's/he was lifted'

The similarities between Ch'orti' and Xinkan are therefore (Quizar 2023: 268):

1. The use of a third person incompletive prefix *a-* for intransitive verbs that contrasts with an unmarked form in the completive.
2. The portmanteau usage of person markers to indicate aspect with no specific aspect markers involved.

Similarity #1 is especially striking to Quizar since "no other Mayan language besides Ch'orti' has a morpheme *a-* that combines both incompletive aspect and third person" (Quizar 2023: 268). As to similarity #2, the stems of the Xinkan verbs do seem to differ when one compares completive and incompletive aspect, even if no designated aspect marker is present. Therefore, the typological parallel to Ch'orti' is not very strong.

Quizar also acknowledges the following differences between Ch'orti' and Xinkan:

1. Ch'orti' shows no change in verb stem (Quizar 2023: 269).
2. The assumed zero morpheme is a prefix in Xinkan as seen in examples (14) and (16) but must be assumed to be a suffix in Ch'orti' since the other persons are also suffixed in the completive aspect (Quizar 2023: 270). Of course, arguing about the position of non-existent marking is a bit futile.
3. The prefixed index-sets involved in incompletive and completive marking in Xinkan are identical except for the third person (Quizar 2023: 270), while in Ch'orti' we are dealing with two distinct index-sets.

Though there are more differences than similarities, one could argue that the latter concern structural phenomena on a larger scale, while the differences only consist in marginal details.

But in what ways exactly could these superficial similarities have been created by contact between the two languages?

First, Quizar establishes which language is likely to have first developed the pattern. The prefix *a-*, a marker of third person incompletive, can be reconstructed for Proto-Xinkan because this pattern appears in all Xinkan varieties as noted above (Rogers 2016: 233–237), but it is an innovation in Ch'orti' (Quizar 2023: 270). However, as was shown in chapter 5.3.2 on Xinkan-Mayan contact, we have no evidence for Xinkan influence on Mayan, only the other way around. Therefore, we need a plausible scenario for the intrusion of a Xinkan morpheme into Ch'orti' grammar.

Quizar proposes that the morpheme *a-* entered the Ch'orti' verbal paradigm "through substratum interference, with Xinkan speakers mistakenly using the Xinkan prefix *a-* instead of the Ch'orti' ergative prefix *u-* for the identical function when speaking Ch'orti'" (Quizar 2023: 278). Xinkan *a-* then replaced A3 *u-*, whereby both grammatical meanings (i.e., 'third person' + 'incompletive') were transferred unchanged and the new prefix became integrated into the already existing split ergative system (Quizar 2023: 270).

A grammaticalized marking of incompletive vs. completive as the sole tense/aspect indicated on verbs is considered by Quizar to be inheritable by both languages, but it could have been "a significant factor during language contact" (Quizar 2023: 274), as well. Under Rogers' (2016: 213) reconstruction, there were no tense/aspect markers in Proto-Xinkan because the attested Xinkan languages also do not have any.[192] According to Quizar, speakers of both languages would have noticed that the aspectual markers are "unnecessary for maintaining the meaning" (Quizar 2023: 275). Contact to Xinkan would then have encouraged both the preservation of the inherited incompletive-completive contrast and the loss of the "redundant" overt aspect marking through aspectual particles. Quizar imagines that bilingual Xinkans would have omitted the overt tense/aspect markers when speaking Ch'orti' because they did not have them in their native language and either simply forgot to use them or viewed them as unnecessary. In theory, Ch'orti' could have lost the overt aspect marking independently, but Xinkan might also well have influenced it – Quizar points out that the other languages (Chol, Chontal) did not lose the overt marking even though it was redundant.

[192] Note that this depends on the analysis: cf. the divergent analysis of Sachse (2010) mentioned above.

Further similarities[193] are discussed in the paper but they do not provide further evidence for the theory concerning the genesis of set C.

5.3.4 Discussion

There are a number of methodological flaws in this paper, of which I can only point out a few examples. For example, the close affinity that Quizar sees between the aspectual system of Ch'orti' and that of Western Cholan (Chol and Chontal) as well as Tsotsil is problematic since the latter, unlike Ch'orti', all have overt aspect markers and also use this marking both on intransitive and transitive verbs. Quizar (2023: 264) acknowledges this but seems to weigh the fact higher that the languages have an aspectual distinction at all. Additionally, the idea of "redundant" marking (Quizar 2023: 267) and its subsequent loss for reasons of economy expresses a kind of notion of language change that is overly teleological. Whether one believes in efficiency as a motivation in language change or not, "redundant" marking is not a strong argument to motivate change since there are numerous examples of languages where some feature is marked "redundantly". The simple dropping of bound morphemes (without a trace outside of phonological attrition caused by regular sound change) is such a radical change that one would first need to provide evidence for the fact that it indeed happens in the languages of the world.

Of course, it is striking that there are similarities in the marking of aspect and especially in the marking of the third person morpheme precisely in the one Mayan language that is spoken in

[193] Two of them are universally common and might have developed independently due to "internal pressures" (Quizar 2023: 276), but the author also considers contact a possibility. The first similarity is "an overabundance of person marker sets" to indicate aspect: Xinkan languages have four different sets and Ch'orti' has an additional third one compared to other Mayan languages. The second similarity is "analytic marking of tense/aspect with auxiliaries, particles, and adverbs". Both patterns are not only universally common, but they are also common in Mayan languages. Ch'orti' and Xinka also share several traits that are in general common to the Mesoamerican language area like relational nouns, ejectives or antipassive constructions (Quizar 2023: 276–277). These typological or areal parallels are neither exclusive to Xinka and Ch'orti' nor rare in general.
Additionally, Quizar (2023: 277) points out two lexical similarities, namely the Ch'orti' 1SG independent pronoun *ne'n* which resembles Xinkan *nen* (already given in the early *Arte* of 1770 (Sachse 2010: 267, 272)) and the preposition and complementizer *ti*, "common to the Lowland Mayan languages, as well as the language of the hieroglyphs" (Quizar 2023: 277). It is unclear why Quizar excludes Highland Maya since the preposition exists there as well (e.g., KCH *chi* with the characteristic sound correspondence of KCH *ch* ~ YUC, CHL etc. *t* (Campbell 1984: 6)) and likely goes back to Proto-Mayan. Quizar assumes that Xinkan loaned it from Mayan (Quizar 2023: 278). For *ne'n*, she does not explicitly state in which direction the loaning process went but since she points out that Ch'orti' independent pronouns cannot have come from Proto-Cholan, it seems she suggests that Ch'orti' loaned them from Xinkan (Quizar 2023: 277).

direct vicinity to Xinka. In fact, the similarity was already pointed out in Sachse (2010: 725, fn. 180), who, however, dismisses a contact explanation for this parallel as unlikely because the similarity is only superficial and the details of the marking are not the same:

> It may seem in some way tempting to regard tense/aspect-based tripartite alignment in Xinka as an influence from Western Mayan split-ergative languages. However, there is little evidence for such a scenario given that in WM languages it is the absolutive pronoun marking S and O that takes the suffix position. It needs to be mentioned that the neighbouring WM language Ch'orti' also exhibits tripartite alignment (see e.g. Dixon 1994:100); an influence from Xinka may, however, be doubted for the same reason. (Sachse 2010: 725, fn. 180)

Overall, the question remains whether it is helpful to explain the peculiar state of Ch'orti' by moving the problem to a different language family (Xinkan) where it might likewise pose problems for the history of the languages. Furthermore, if aspect marking was dropped due to being unnecessary as argued by Quizar – why does it still (or again) exist in Ch'orti', e.g., in *war* 'progressive'?

Quizar argues that her proposal is preferable to "the more complicated scenario under the direct descent proposal" (Quizar 2023: 270). Though I certainly agree with her on that, her proposal still only explains that part of the set C enigma which is the least problematic one. It is not difficult to find a valid source for C3 *a-* in Mayan. There is a deictic element *a* in many Mayan languages (see discussion in chapter 7.3.2) and this can easily be assumed to have been recruited as a third person. On the other hand, her theory explains little else. It does not explain why this development only affects intransitive verbs. Most importantly, it does not explain where the other problematic morphemes of set C come from. She only states that the newly introduced difference "became a catalyst for the replacement of the ergative second person markers, while the first person markers remained the same as the original ergative ones" (Quizar 2023: 270). She does not go into detail as to how exactly the second person indexes were replaced, where the forms come from and why the first person remains as it was. She also does not comment on C2PL *ix-*. Perhaps she accepts the proposals in Robertson (1998) (or Wichmann 1999), but she does not state this.

However, Quizar's paper proves that Robertson's (1998; Houston, Robertson & Stuart 2000a) claims have not been accepted unanimously among Mayan linguists even though little has been published to challenge them in the last 25 years, while Robertson and colleagues have reaffirmed their ideas in numerous publications presenting them as established fact (see, e.g., Robertson, Houston & Stuart 2004; Law, Robertson & Houston 2006; 2009; Robertson & Law 2009; Robertson, Law & Haertel 2010).

5.4 Discussion

The explanation of set C proposed in Robertson (1998: 199) has been uncritically adopted by many authors (e.g., Schweitzer 2006: 135; Becquey 2014: 334–336; Law 2014: 136; Vinogradov 2016: 33; Lois et al. 2017: 104). Law especially states:

> Even a superficial investigation of the so called 'Set C' person markers […] shows that they are historically related to the Set A markers. They are prefixes, like the Set A markers, and the first person markers (both singular and plural) for both sets are identical. (Law 2014: 136)

A similar view is expressed in Schweitzer who states that

> a tripartite system arose in the imperfective aspect through further development of the A series […]. However, even from a folk-etymological point of view, this further development is transparent and recognizable as a unit with the series A, so that from a typological point of view this detailed differentiation can be ignored. In contrast to the distinction between the series A, B, C [...], it is also better to speak terminologically of the series A1, B and A2, since A2 can be clearly related to A1 […] (Schweitzer 2006: 135)[194]

However, superficial similarities are precisely **not** how the comparative method operates: "It is a frequent misconception that reconstruction is based on similarities" (Campbell & Harris 2002: 603). Schweitzer even acknowledges the difficulties in deriving set C from set A but does not question the idea in itself:

> […] although the phonetic derivation is somewhat difficult since no difference in phonological position can be recognized between the related series, i.e., no phonological conditioning but only morphological conditioning. However, morphological conditioning cannot be accepted in a classical reconstruction technique. Particular care must be taken with hypotheses, especially in the case of divergent development, since no compensation by analogy is possible there either. (Schweitzer 2006: 135)[195]

Sometimes, glimpses of an alternative view are offered. Kaufman and Norman (1984: 90) describe set C as "neither ergative nor absolutive". In passing, Polian (2017: 209) states that "prefixed absolutive markers in Tsotsil and Ch'orti' are exclusively verbal, whereas suffixed absolutive markers are used both on verbs and non-verbal predicates". Schumann Gálvez (2007) calls the index-sets "ergative" (set A) and "absolutive" (set B and set C). It is unclear

[194] German: "[…] durch eine Weiterentwicklung der Reihe A ein Tripartite-System im imperfektiven Aspekt entstanden ist […]. Allerdings ist diese Weiterentwicklung selbst aus volksetymologischer Sichtweise durchsichtig und als Einheit mit der Reihe A erkennbar, so daß man aus typologischer Sicht diese Detaildifferenzierung übergehen kann. Im Unterschied zu der Unterscheidung der Reihen A, B, C […] sollte man auch terminologisch besser von den Reihen A1, B und A2 sprechen, da sich A2 klar auf A1 beziehen läßt […]".

[195] German: "[…] wobei die lautliche Herleitung etwas schwierig ist, da sich zwischen den verwandten Reihen kein Unterschied in der phonologischen Position, also keine phonologische Konditionierung sondern nur eine morphologische Konditionierung erkennen läßt. Morphologische Konditionierungen sind aber bei einer klassischen Rekonstruktionstechnik nicht zu akzeptieren. Gerade bei einer Auseinanderentwicklung muß man mit Hypothesen besonders vorsichtig sein, da dort auch keine Ausgleiche durch Analogie möglich sind."

whether they do so because they assume that set C developed out of set B or whether this is just a synchronic parallel because both are used on intransitive verbs.

I believe that this chapter has demonstrated why none of the approaches that derive set C from set A offer satisfying explanations. Problems with them include the following:

- The proposed analogies and reanalyses lack contexts where they could have happened.
- Some ad hoc assumptions are required to bend the theories into shape.
- The rather weak argument of homophony avoidance is overused.
- More problematic, however, is the fact that all three theories fail to address or fail to explain to satisfaction the following key issues of the Ch'orti' construction:
    - If the underlying construction is the same in Ch'orti' as in Yucatecan, Western Cholan and Poqom, why is there no trace of the missing nominalizer?
    - If set C comes from set A, why does it only undergo the specific changes with intransitive verbs, but keeps all its earlier forms with transitive verbs where the conditions are otherwise identical?
    - Finally, and perhaps most crucially, why does the resulting aspectual distinction only apply to intransitive verbs?

In the remaining section of this chapter, I will recount an interesting exchange of opinions that took place in 1982, i.e., long before Robertson and Wichmann published their ideas on set C. This exchange deals precisely with the origin of set C and points out an important difference between set C and set A.

5.5 Dayley vs. Fought (1982)

A debate was kicked off when Dayley (1981a: 44–46) wrote a journal article on voice and ergativity in Mayan that also had a section on Ch'orti'.[196] This publication led to a heated discussion with John Fought, who had done extensive fieldwork on Ch'orti' and did not agree at all with how Dayley had portrayed the language.

[196] A publication with the same title that is more easily accessible is Dayley (1983). The relevant information is found on pages 66–70.

One of the points of contention was Dayley's assumption that set C is just a variation of set A. In his response to Dayley (1981a), Fought (1982: 34) pointed out that the morphemes of set A and set C cannot be the same, though some of them might look alike when used before verbs beginning with a consonant (except *j*).[197] However, while set A ends in the usual, well-known Mayan glide before verb-initial vowels, set C has no final glide and instead C3 *a-* usually assimilates with the following vowel (Fought 1982: 34). He therefore concluded that Dayley's musings on set A being the source of set C are "unfounded" (Fought 1982: 34).

The response by Dayley, named somewhat humorously "A hard Fought repartee", in which he defended his basic points even though "Fought apparently believes this is all hogwash" and stated that his own "speculations are not as unfounded as Fought would have us believe" (Dayley 1982: 48), was considerably more heated in tone. He pointed out that what he suggested had already been described for other Cholan as well as Yucatecan languages and that the parallels to Ch'orti' are remarkable since there, it is also the incompletive aspect of intransitive verbs that is affected by the new development (Dayley 1982: 48–49). His strongest argument seems to be the fact that

> the putative difference in 'underlying form' of the first person plural ergative prefix and the first person plural IV incompletive subject prefix is based solely on Fought's notational device '&' used on ergative prefixes (cf. ka&- ergative vs ka- IV incompletive subject); likewise, for second person plural ergative and second person singular IV incompletive subject (cf. i&- ergative vs i- IV incompletive). (Dayley 1982: 49)

Here, some methodological flaws become apparent that many publications in Mayan linguistics and Ch'orti' specifically suffer from. Dayley (1982: 48) put the two paradigms next to each other in the following way:

Set C			Set A	
S1	in-	P1 ka-	S1 in-	P1 ka-
S2	i-	P2 ix-	S2 a-	P2 i-
S3	a-	P3 a-…-ob'	S3 u-	P3 u-…-ob'

Presented like this, it is not surprising that the parallels seem more striking than they actually are, especially in the first person singular and plural. This led Dayley to believe that "Chorti displays a tendency toward accusative marking in the incompletive (more precisely, toward accusative marking via an 'extended ergative system' in Dixon's (1979) terms)" (Dayley 1982: 48). In fact, Dayley himself argued that Fought's disagreement is solely based on the

[197] Ch'orti' *j* is lost between vowels. This will be discussed in detail in chapter 6.2.

notation that he uses, which he finds idiosyncratic (deviations from Dayley marked in yellow; plural marker is optional):

Set C		Set A	
vn-	ka-	in&-	ka&-
i-	ix-	a&-	i&-
a-		u&-	

The "&" in Fought's notational system expresses the glide seen when set A attaches to a verb that begins with a vowel. The form of C1SG is *vn-* instead of *in-* because the vowel varies when attached to verbs that begin with a vowel: *v* becomes the vowel that the verb begins with and is laryngealized (see section 4.1). As chapter 6 will show, Fought's presentation of the index-sets is justified because it unveils the differences in how set A and C interact with a verbal stem.

Curiously, Dayley (1982: 1949) seemed to be aware of the different behavior of the two sets. This makes it all the more difficult to understand why, despite acknowledging the different behavior, he did not think that this points to different origins. Instead, he stated that "&" is simply used by Fought "to indicate differences between the combinatorial properties of the ergative prefixes and IV incompletive subject prefixes" and that "[d]ifferences in combinatorial properties of the ergative prefixes and other person marking prefixes (e.g. absolutive prefixes) are common in Mayan languages and they have less to do with underlying form than with purely morphological factors" (Dayley 1982: 49). He did not elaborate further, leaving the readers alone with their astonishment – how could the "combinatorial preferences" be different if the employed index-set is the same?

In the end, it seems that there was less reason for conflict than both parties thought, since Dayley ultimately classified set C as "neither absolutive nor ergative" (Dayley 1982: 45). In a subsequent publication, the connection between set A and C is formulated in a rather neutral way where it is described that Ch'orti' "displays certain accusative-like features" (Dayley 1983: 69) because the first-person prefixes are identical in set A and set C.[198]

[198] However, he also suspects that C2SG *i-* may be an extension of the A2PL *i-* and points out that C3 *a-*, while not being identical to A3, still is not unmarked as in other Mayan languages, which leads him to conclude that "there is a tendency towards accusative marking in the incompletive, especially with 1st person" (Dayley 1983: 69–70).

In the following chapter I will discuss the different ways in which set A and C interact with the verbal (and nominal in the case of set A) stems and explore what this can tell us about their relationship.

6 Morphophonological alternation of set A vs. set C

The morphophonological variation discussed in the following chapters has also been treated or at least mentioned in Fought (1967; 1972), Wichmann (1999), Schumann Gálvez (2007), Storniolo (2008), Becquey (2014) and Marcos García (2015), though never systematically and with a historical perspective as it will be done here.[199] As a result, the allomorphs of set C have gone unmentioned or been reduced to simple vowel assimilation in some treatments of Ch'orti'[200] while in fact, the interaction of the indexes with vowel- or *j*-initial stems is a more complex phenomenon. Among the Cholan languages, it is most pervasive in Ch'orti' and a systematic, in-depth study is therefore needed (Becquey 2014: 204).

To date, the most in-depth theoretical treatment of Ch'orti' phonology and morphophonology is Fought (1967; 1972) who transcribed his recorded texts while paying attention to the exact phonetic representation. This helps in understanding what sounds precisely we are dealing with, e.g., the nature of the laryngealized vowels that are the topic of this chapter. For easier comparability, I decided to adapt Fought's notational conventions to the modern orthography used in Hull (2016) but will use his exact notation whenever it is necessary for the argumentation.

One of the most important domains where morphophonological alternations happen is at the border of index-set and verb (or noun) stem, which is why the question is relevant for this thesis. Two different phenomena must be distinguished, though they are interconnected. The interaction between index-sets and the nouns or verbs they attach to is apparently obligatory and leads to regular results based on, e.g., the data in Hull (2016). Rapid speech can yield the same results in phrases across lexical boundaries, but this is not obligatory.[201] For instance, *ta*

[199] Wichmann (1999), e.g., is difficult to use because he formulates the rules of these morphophonological alternations in a counterintuitive way. He states that he uses "a not too formal version of the standard format for such rules" (Wichmann 1999: 13); however, the rules at least do not seem to follow the conventions used in historical-comparative linguistics. For instance, rule 1d says "*i* -> *y*/__V, where V *i*" which would imply that *i* changes to *y* before another *i* but then the examples clearly show that this is not what Wichmann meant (*ajn-i-en* > *ajnye'n*, *ni-otot* > *nyo'tot*). This is not the only case – the phrasing of the rules is frequently misleading in various ways. Therefore, they are not useful to anyone who does not already have some knowledge of Ch'orti' contractions. Wichmann also states that his study is preliminary, especially when it comes to contractions involving a disappearing *j* (Wichmann 1999: 18), which makes a more in-depth study necessary.

[200] Quizar (2023: 261) is a notable counterexample though she also does not discuss the conditions of the changes in detail.

[201] Not unexpectedly, the more complex morphophonemic variation has been observed in informal situations (Fought 1967: 96). Fought's principal informant, Isidro González, used "the same narrative style for texts as he used for sermons. This style is characterized by slow, solemn diction with frequent pauses

uyotot ‘to his house’ (PREP *ta* + A3 *uy-* + *otot* ‘house’) can become *tu’uyotot* or even *tu’yotot* (Fought 1967: 95). The latter will not be investigated in depth in this chapter, but it is noteworthy that the results of both processes are apparently the same as witnessed by the data in Fought (1967). I will sometimes supplement my evidence with cases of contractions due to rapid speech and also with etymological/comparative evidence. By demonstrating that the results of the morphophonological alternations are the same as what happened in the course of the history of Ch’orti’ with lexemes that used to feature the respective vowel combinations, I intend to demonstrate that the alternations are the result of a sound change in the past.

As we will see, set A and set C result in different outcomes when a sequence *-V’V-* is created by their attachment to a lexeme. The traditional variation of set A shared by most Mayan languages[202] depends on whether the stem it is prefixed to begins with a consonant or a vowel. Table 20 is an overview of the forms of set A on consonant- and vowel-initial verbs or nouns. An example is provided for each vowel. Some forms, especially of verbs beginning with *u-*, are conspicuously unattested in my corpus. For A1SG with lexemes beginning with *o-* or *u-*, I only found cases of nouns with the nominal version of set A *niw-*, none of verbs.

Table 20. Set A interaction with stem depending on initial sound.

	_C	_a	_e	_i	_o	_u
A1SG	*in-na’ta* ‘I know’	*inw-a’re* ‘I say’	*inw-ejtz’u* ‘I try’	*inw-ira* ‘I see’	*niw-otot* ‘my house’	*niw-uch’nib’* ‘my cup’
A2SG	*a-na’ta* ‘you know’	*aw-a’re* ‘you say’	*aw-ejta* ‘you try’	*aw-ira* ‘you see’	*aw-ojroner* ‘your words’	–
A3	*u-na’ta* ‘s/he knows’	*uy-a’re* ‘s/he says’	*uy-eb’ta* ‘s/he sends’	*uw-ira* ‘s/he sees’	*uy-ojroner* ‘his/her words’	*uy-usre* ‘he wants’
A1PL	*ka-na’ta* ‘we	*kaw-a’re*	*kaw-eroj* ‘our	*kaw-ira* ‘we	*kaw-ojroner*	–

and infrequent reduction of vowel sequences” (Fought 1967: 96). To determine the presence or absence of the glottal stop, Fought (1967: 97) thus deems it necessary to carry out a detailed dialectal or stylistic analysis, which he has not done himself and to my knowledge has not been done by others so far either. Cases that would actually require laryngealization like *una’ta* ‘he knows’ or *ma ani* ‘it wasn’t’ are occasionally attested as /unaata/ and /maani/ with a long vowel without glottal stop (Fought 1967: 98). This confirms the observations made by Hull (2016) concerning the variation of the actual realization of glottal stops in the language: He described that laryngealization can be absent especially often with the plural suffix *-ob’* and the clitic *=ix* ‘already’ (2016: 8). These are both highly frequent forms. It is also not consistently present in medial vowels in closed syllables, e.g., *pak’-* ‘to plant’ ~ *pa’k’ma’r / pak’ma’r* ‘planting’ (Hull 2016: 8). The large variation in the glottal articulation could be explained by the different oral articulations that accompany the glottal one (Fought 1972: 14).

[202] It is, apparently, absent in Huastec (Edmonson 1988: 115).

	know’	‘s/he says’	faces’	see’	‘our language’	
A2PL	*i-na’ta* ‘you (pl.) know’	*iw-akta* ‘you (pl.) dance’	*iw-ermano*[203] ‘your (pl.) brother’	*iw-ira* ‘you (pl.) see’	*iw-ojroner*[204] ‘your (pl.) words’	–

When set A combines with vowel-initial verbs, a glide is usually inserted between the index and the stem. The glide is *-w-* except for the third person, which mostly shows *-y-*.[205]

If we now compare the alternation of set A with vowel-initial stems to what happens to set C in the same environment, we make a somewhat surprising discovery already hinted at in chapter 5.5. Although both are prefixes and they even supposedly share homophonous forms like A2SG/C3SG *a-*, they behave differently when prefixed to verbs that start with vowels. While we observe the usual prevocalic glide for set A in example (1) in *aw-akte’nix*, in (2), *a-ati* with set C is contracted to *a’ti* instead of **awati.

(1) Prevocalic set A (Hull 2016: 50)

Awakte’nix taka nipatna’r.

aw-akta-en=ix *taka* *ni-patna-ar*

A2SG-surpass-B1SG=already PREP A1SG-work-VN

‘You have been promoted above me.’

(2) Prevocalic set C (Hull 2016: 233)

E ajmok ixik a’ti tama e k’ijna’.

e *ajmok* *ixik* *a-ati* *tama* *e* *k’ijna’*

DEF sick woman C3-bathe PREP DEF hot.water

‘The sick woman bathes in hot water.’

[203] Prevocalic *iw-* seems to be unattested in Hull (2016), likely due to the size of the corpus and the fact that the second person plural occurs more rarely than the others. Source for *iw-ermano*: Ch’orti’ New Testament San Mateo 18,35 (Wycliffe Bible Translators 2012: 52).

[204] Just like *iw-ermano*, this form is attested in the Ch’orti’ New Testament San Mateo 5,33 (Wycliffe Bible Translators 2012: 11).

[205] However, Table 20 shows that verbs beginning in *i-* show the glide *-w-* in all persons, including the third: *awira* ‘you see it’ ~ *uwira* ‘he sees it’ – something already noted in Fought (1967: 110). Examples from Hull (2016) of unexpected *w* in A3 with stems that do not start with *i-* include: *uw-ej* ‘her mouth’ and *uw-un e max* ‘fruit of a kind of tree’.

The opposite case is recorded with the verb *ejta* ‘TV9 to try, tase, test, take a whiff of’, where we see the glide *y* even in the 1SG with *iny-ejta* ‘I try’ (Hull 2016: 134) instead of the expected *inw-ejta*. In *ejtz’u* ‘TV20 try, attempt, test’, the form is as expected: *inw-ejtz’u* ‘I am trying’ (Hull 2016: 134).

set A: a + akta > awakta

set C: a + ati > a'ti

The contraction that happens with set C is part of a widespread phenomenon in Ch'orti' while glide insertion occurs exclusively with set A and occasionally in Spanish loanwords containing vowel sequences like Spanish *día* > CHR *diya* (Hull 2016: 259). Two further examples of contractions are *patna'r* 'work' and the contraction between the vowel stem and B1SG *-en* in *awakte'nix* in example (1).

In this chapter, I will provide a detailed overview of how vowels interact with each other in sequences. There are two different sources for these sequences which I will discuss separately in 6.1 and 6.2. I also discuss the interactions of A/C1SG *in-* (and C2PL *ix-*) separately in 6.3. Though these affixes do not end in vowels, they still interact with the following stem in somewhat unexpected ways, which, however, can easily be understood when one pays attention to Ch'orti' phonetics.

I do not limit my investigation to vowels that occur in the index-sets. This way, this chapter is not only useful for the question at hand but also for future studies of Ch'orti' where morphophonological alternations play a role.

6.1 Vowel contractions with glottal stop

6.1.1 a + V

For vowels that are not represented in the index-sets, data can be scant. They are especially abundant for *a* because many indexes end in *-a* (A2SG/C3 *a-*, A/C1PL *ka-*).

a + a > a'

a'ti 'he bathes' (C3 *a-* + *ati* 'IV25 to bathe')
a'rob'na 'you are called' (C3 *a-* + *arob'na ~ aro'b'na* 'PAS2 to be told, ordered')
twa'ch 'one's own' (*twa'* 'for/of (him/her/it)' + *-ach* 'intensifier, own'[206])
eb'tana'r 'sending, commanding' (*eb'tana* 'PAS2 to be ordered, commanded' + *-ar* 'verbal noun')

[206] *-ach* is an intensifying particle with the rough meaning 'own, oneself, only, alone, one's own' (Hull 2016: 27).

The result seems to be a straightforward *a'*. Compare this again to the result of homophonous A2SG *a-* on transitive verbs or nouns which shows the expected glide for A2SG:

awa're 'you say (it)' (A2SG *a-* + *a're* 'TV10 to say')
awarib' 'your daughter-in-law' (A2SG *a-* + *arib'* 'daughter-in-law')

a + e > e'

e'btana[207] 'he is ordered' (C3 *a-* + *eb'tana* 'PAS2 to be ordered, commanded')
e'kmay 'he is coming down' (C3 *a-* + *ekmay* 'MPAS4 to go down, descend')
e't'oki 'he accompanies' (C3 *a-* + *et'oki* 'IV7 to accompany')
inwire't 'I saw you' (A1SG *in-* + *ira* 'TV to see' + B2SG *-et*)

Again, the result is straightforward: *e'*. Meanwhile, A2SG *a-* shows the expected glide:

aw-ejta 'you try' (A2SG *a-* + *ejta* 'TV9 to try')

a + i > e'y, e', a', a

More variation is observed when *a* interacts with *i*. It seems like the most common result is *e'y*:

e'yxna 'he takes long to …' (C3 *a-* + *ixna* 'IV to last, take time')
pe'yr 'bitter parts' (< *pa'-ir* according to Hull (2016: 332) likely with abstractive suffix *-ir* but he does not state what *pa'* is in his opinion)
je'yr 'juice of a fruit' (*ja'* 'water' + *-ir* 'abstractive suffix')
nojte'yr 'bigness' (A3 *u-* + *nojta'* 'big' + *-ir* 'abstractive suffix')
nojte'yx 'he is already big' (*nojta'* 'big' + *=ix* 'already')
e'ksijb'a '(her eyes) went dark' (C3 *a-* + *iksijb'a* 'IV12 to get dark, become dark')

It is interesting that a glide may be involved when *i* is one of the vowels. Though one could argue that this resembles set A, the situation is in fact very different. The glide in this case is not simply inserted between two colliding vowels; instead, a contraction happens and the result is still laryngealized. But the glide can also be missing, as the last example shows. Here,

[207] An ejective may lose its glottalization when laryngealization occurs before it, thus *a + **eb'**tana* > *e'**b**tana* (Hull 2016: 8).

this might be due to the otherwise resulting cluster **yks, which is unattested in the language and perhaps always simplified.

The numeral *in-* and the verb *ixin* 'IV to go' show special developments. Here, it almost looks as if (contrary to the cases seen until now) the first vowel is more prominent than the second because the result is *a'* or even *a*. This is especially surprising for *ixin* as *ixna* above, likely a derivation of *ixin*, behaves as expected.

a'xin 'he goes' (C3 *a-* + *ixin* 'IV to go')
ka'xin 'we go' (C1PL + *ixin* 'IV to go')
tamante'/tama'nte' 'in one, on one' (*tama-inte'* (Hull 2016: 392); PREP *tama* + *in-* 'one' + CL *-te'*)
tante' 'in a …, on a …' (*ta-inte'* (Hull 2016: 393); PREP *ta* + *in-* 'one' + CL *-te'*)
tanyajr 'once upon a time' (*ta-in-yajr* (Hull 2016: 393); PREP *ta* + *in-* 'one' + CL *-yajr*)

Set A shows its usual glide when A2SG combines with a verb starting in *i-*:

aw-ira 'you see (it)' (A2SG *a-* + *ira* 'TV to see')
kaw-ira 'we see (it)' (A1PL *a-* + *ira* 'TV to see')

Though *a* + *i* shows us that the results of the contractions are not as uniform as *a* + *a* and *a* + *e* made us believe, at least none of the results are the same as the one that set A produces under otherwise identical conditions.

a + *o* > *o'*

o'jri 'he falls' (C3 *a-* + *ojri* 'IV25 fall, fall down')
o'b'okna 'it smells good' (C3 *a-* + *ob'okna* 'AFV2 to smell good')
war o'jronob' 'they are conversing' (C3 *a-* + *ojron* 'AP2 to speak, talk' + PL *-ob'*)
o'cho'b 'they enter' (C3 *a-* + *ochoy*[208] 'MPAS4 to enter' + PL *-ob'*)
o'jronob' 'they speak' (C3 *a-* + *ojron* 'AP2 to speak, talk' + PL *-ob'*)

With *a* + *o*, the result is again a straightforward *o'*. Ideally, we would now contrast this with the glides that are inserted when homophonous A2SG is prefixed to a transitive verb starting in *o-*. However, what we see in fact are forms like the following:

[208] On why the mediopassive class 4 suffix *-Vy* disappears here see chapter 7.4.2.

o'jres 'you dropped (her)' (A2SG *a-* + *ojres* 'CAUS2 to drop') instead of **awojres
o'jtz'ika 'do you smell (it)?' (A2SG *a-* + *ojtz'i* 'TV4 to smell' + *=ka* 'interrogative') instead of **awojtz'i(ka)
o'jmes 'you boil (the water)' (A2SG *a-* + *ojmes* ~ 'CAUS2 to lather; boil') instead of **awojmes

The same happens with A1PL *ka-*:

ko'se 'we put (it) in' (A1PL *ka-* + *ose* 'TV10 to put in') instead of **kawose

It seems as if no set A forms that end in *-a* (A2SG *a-*, A1PL *ka-*) appear with glides when prefixed to verbs starting with *o-* at all – at least not in the corpus.

On the other hand, this does not apply to A3 *u-* as the following examples show:

uy-ob'okres 'it perfumes (= makes smell good)' (A3 *u-* + *ob'okres* 'CAUS16 to perfume')
y-ojmes[209] 'she lathers (the soap)' (A3 *u-* + *ojmes* 'CAUS2 to lather; boil')
uy-ojres 'she dropped (it)' (A3 *u-* + *ojres* 'CAUS2 to drop')
uy-ose 'he put (it) (A3 *u-* + *ose* 'TV10 to put in')

This behavior of set A also extends to the use as possessive marker on nouns: there is some variation with forms that still have the original glides (see *aw-ojroner* 'your words' in Table 20) but many forms participate in contractions identical to those that set C (and other vowel contractions in the language) produces:

twam o'k 'between your legs' (A2SG *a-* + *ok* 'foot, leg') instead of **aw-ok
but: *uy-ok* 'his leg(s)' (A3 *a-* + *ok* 'foot, leg')

It may seem that set A is starting to behave like set C in those forms that the two sets share, even though they do not mark the same grammatical person, A2SG/C3 *a-* and A/C1PL *ka-* (and theoretically A2PL/C2SG *i-*). One could argue that this was made possible by verb roots that exist both in intransitive and transitive stems. For example, *ojmes* has an intransitive counterpart *ojmay* 'MPAS4 to lather, soap', which becomes *o'jmay* when used with C3 *a-*. It would then be natural that those forms of set A homophonous to the ones of set C that are already participating in the contraction would lead the way in this new development. However, the following section will show that this phenomenon has a purely phonetic background.

[209] This looks like an old form with prevocalic *y-* instead of doubly marked *uy-*. Perhaps a mistake or an archaic form.

a + u > o'y, o', u'

o'sta 'he is getting better' (C3 *a-* + *usta* 'IV49 to arrange, fix, get better')
o'b'yan 'he understands' (C3 *a-* + *ub'yan* 'AP11 to understand')
o'ch'pa 'it is drunk' (C3 *a-* + *uch'pa* 'MPAS1 to be drunk, be consumed (liquid)')
o'jya'n 'he is breathing heavily' (C3 *a-* + *ujya'n* 'AP12 to breathe heavily from exhaustion')
o'tzerna 'it is desired' (C3 *a-* + *utzerna* 'PAS2 to be desired')

It seems at first glance as if set C yields the same result both with verbs that start with *o-* and those that start with *u-*, namely *o'*. There are even examples of these contractions having been lexicalized, in both cases formations including C3 *a-* and the verb *ub'i* 'to hear':

o'b'an 'really? are you sure? or do you…?'
o'b'tz'a 'MPAS2 it is heard'

Aside from *o'*, we also see the result *o'y*:

ko'ych'i 'we drink' (C1PL *ka-* + *uch'i* 'IV25 to drink')
o'yk'i 'he cries' (C3 *a-* + *uk'i* 'IV to cry')
o'ych'er 'drunk, drunkard' (C3 a + *uch'i* 'IV25 to drink' + *-er* 'verbal noun')
perhaps also *cho'ynen* ~ *cho'wnen* 'step-son, step-daughter' (*cha'-* 'two' + *unen* 'son/daughter (of a man)')

Perhaps this is conditioned by the following *-Ci*, though more examples would be needed to prove this and the suggested *cho'ynen* would not be covered by this conditioning. As in the first group of examples, all verbs would yield a complex consonant cluster of three consonants (or in the case of *tz* at least an affricate + another consonant) if the result were *o'y-* instead of *o'-*, I rather suspect that we are dealing with cluster simplification and that *o'y-* would indeed be the "basic" result of *a + u*. If we take the evidence from *a + i* and *a + u* together, it seems as if high vowels tend to also produce a glide in addition to the laryngealization when contracted. A bigger corpus study will reveal clearer patterns of when the glide occurs and when it does not.

A third result of the contraction *a + u* is *u'*. Just like with *a + i*, it, too, happens when the preposition *ta* is involved.

tu'tamir 'in the midst of, into the very center of, in the middle of' (PREP *ta* + A3 *u-* + *tamir* 'deepness')

tu'jam 'between, in the middle of' (PREP *ta* + A3 *u-* + *jam* ~ *am*[210] 'space (between things)') with variations *tu'wam*, *twam*[211]
tu'k'a'pa'r 'at the end' (PREP *ta* + A3 *u-* + *k'a'pa'r* 'destiny')[212]

Examples of a transitive verbs or nouns with set A participating in the "set C" pattern include:

ko'sre 'we desire (her)' (A1PL *ka-* + *usre* 'TV12 to desire')
ko'jtz'i 'we worship (God)' (A1PL *ka-* + *ujtz'i (ut)* 'TV4 to worship')
mato'yt 'it still isn't enough (for you)' (NEG *ma* + *=to* 'still' + A2SG *a-* + *ut* 'face')

The third person of set A features the familiar old glide: *uy-ujtz'o'b* 'they are worshipping (God)'.

The key observation to understanding this phenomenon is one that was already hinted at in Table 20 above. There, we saw that some set A forms (A2SG, A1PL and A2PL) with glides are unattested for verbs with *u* in my corpus. Though the lack of examples for A2PL may simply be due to the fact that this form is generally less common and therefore has fewer attestations in my corpus, this cannot be the explanation for the remaining two forms, both of which end in *a-*.

Since both *a* + *o* and *a* + *u* are affected while we do not see this phenomenon with other vowels, a plausible hypothesis is that the origin of this behavior is the resulting sequence *wo/wu* (*w* + rounded vowel). These sequences, especially *wu*, are barely attested in Hull (2016) and completely unattested in the translation of the New Testament (Wycliffe Bible Translators 2012). The few forms that do attest the glide can be explained as new formations, e.g., *uw-un e ma'x* 'the un ma'x fruit' (Hull 2016: 468), which is a unique attestation of A3 with *w* instead of *y* as a glide, or the Spanish name *Juana* rendered as *Jwuana* in Ch'orti' (Hull 2016: 111).

It is best not to describe this as a "restriction" of the phonological system; one may rather consider that in [wo] and [wu], the semivowel is not particularly distinct and could have been lost in a sound change of *wo, wu > CHR *o, u*, which might have happened before the contractions so that cases of set A used before words in *o-* or *u-* formed part of the *-VV-* sequences that rendered laryngealized vowels. Therefore, the fact that some forms of set A start to behave like set C is not due to analogical extension but due to a historical phonetic

[210] The variation of stems starting with *j* or with vowel will be discussed in 6.2.
[211] The variation is interesting but there is little material to make any generalizations about it.
[212] There are many more examples in Hull (2016: 424–426).

accident.[213] The reason why we do not see the development with A3 *u-* is that this form usually appears with the glide *y*.

6.1.2 e + V

There is fewer material that shows morphophonological alternations with *e* as the first vowel of the contraction because no set A or set C forms end in *-e* and there are also fewer verbs that end in *-e*. Nevertheless, some examples exist. For a more elaborate study, one could include the contractions that Fought described for rapid speech because the definite article is *e*. His recordings are available on AILLA.[214] Diachronic data will likely yield more cases, as well.

e + a > ya'

chamsya'n 'AP5 to kill' (*chamse* 'CAUS1 to kill' + *-an* 'antipassive suffix')
intya'ch 'so much, over and over' (*in-* 'NUM one' + CL *-te'* + *ach* 'intensifier, own')

The data suggest that *e* and *a* contract to *ya'*.

e + e > ye', e'

uk'eche'n 'it took me a long time' (A3 *u-* + *k'eche* 'TV8 to take so. a long time' + B1SG *-en*)
chanle'n 'I stare' (*chanle* 'TV12 to stare' + B1SG *-en*)
chanlye'n 'stare!' (*chanle* 'TV12 to stare' + IMP *-V/en*)
atakrye'nix 'you've already helped me' (C3 *a-* + *takre* 'TV12 to help' + B1SG *-en* + *=ix* 'already')
uchamsye'tob' (A3 *u-* + *chamse* 'CAUS1 to kill' + B2SG *-et* + PL *-ob'*)
uya'rye't 'he tells you' (A3 *u-* + *a're* 'TV10 to say, tell' + B2SG *-et*)

Based on the data of *a + V*, the expected reflex of a contraction of *e + e* would be *e'* and this is in fact what we find in some cases. But there are more cases of *ye'*. The pair *chanle'n* 'I stare' and *chalye'n* 'stare!' is especially interesting. According to Dugan, the imperative is

[213] Wichmann (1999: 18), on the other hand, explains this as special cases of reductions involving medial glides *y* and *w*, for example by assuming an underlying *ni-y-otot* for *nyo'tot* 'my house'. However, the expected glide with *ni* is *w*, not *y*. Therefore, I would argue that my explanation of *ni-w-otot* > *ni-otot* > *nyo'tot* is preferable, especially since it covers all other cases.
[214] https://ailla.lib.utexas.edu/collections/822/ (last accessed: 2025-03-13; requires registration).

formed with *-Vn* with the vowel often being *e* (see Appendix E chapter b.6). This does not explain why a glide would appear in one form but not in the other. The imperative form likewise contradicts Wichmann's proposed distribution of imperative morphology (also discussed in chapter b.6) because the morpheme cannot simply be *-n,* either. Since the opposite distribution is attested for *tzunle* 'TV12 to guard or watch over sth. jealously' (see below), the explanation perhaps is not morphological but one of synchronic variation, though this is an unsatisfying assumption. It is clear that the formation of the imperative in Ch'orti' requires further, especially historical, analysis.

utzunlye'n 'he guards me jealously' (A3 *u-* + *tzunle* + B1SG *-en*)
tzunlen 'guard it jealously!' (*tzunle* + IMP *-V/en*)

Historical evidence includes the following lexemes:

PCH *teʔ-el 'woods'[215] > CHR *te'rar* 'the tree of, the wood of' with an additional suffix *-ar*
PCH *weʔ-el 'meat'[216] > CHR *we'r* 'muscle, fleshy part'

<u>*e + i > i', e'y, e', e*</u>

ch'epi'k 'scratch!' (*ch'epe* 'TV8 to scratch' + IMP/OPT *-ik*)
jeb'i'k 'open (a little)!' (*jeb'e* 'TV8 to open a little' + IMP/OPT *-ik*)
eri'x 'right now' (*ere'/e* 'just' + *=ix* 'already')
otronte'yx ~ otro'nte'yx 'a different thing, an altogether different thing' (*otro* 'another [Span. *otro*]' + *in-* 'one' + CL *-te'* + *=ix* 'already')
e'n-kojt ~ en-kojt 'the other (of living things)' (DEF *e* + *in-* 'one' + CL *-kojt*)

[215] Reconstructed by Kaufman and Norman (1984: 132) based on Chol and Chontal but clearly attested in Ch'orti', as well. As mentioned in chapter 2.2.4, they do not provide lexical material but only name the languages that they used. Though they state which resources they used in general, I decided to use more modern sources since great resources have become available. I provide lexical data from the oldest sources whenever possible: data for Chol (1789–1935) are published in Hopkins, Josserand and Cruz Guzmán (2011); the best source for Acalan Chontal (ACN) (1610/12) is Smailus (1973). However, the corpus of Acalan Chontal especially is very limited.
CHL *te'el* 'forest' (Hopkins, Josserand & Cruz Guzmán 2011: 218)
CHN *te'e* 'mountains, jungle, forest'. Modern Chontal regularly loses word-final *l* (Smailus 1973: 130) but it is retained in compounds like *te'el animá* 'mountain/forest animal' (with *animá* being Spanish *animal*, again with loss of final *l*) (Keller & Luciano G. 1997: 235).
[216] Reconstructed by Kaufman and Norman (1984: 135) based on Chol, Chontal, Ch'orti'.
CHL *we'el* 'food' (Hopkins, Josserand & Cruz Guzmán 2011: 264)
CHN *we'e* 'meat' (Keller & Luciano G. 1997: 281)

The result *i'* seems to be slightly more common than *e'y* or *e'* (or *e*) but in general, we again see most variation with the involvement of *i*.

e + o > yo', o'

uwarajsyo'n '(it) shines on us' (A3 *u-* + *warajse* 'CAUS12 to visit' + B1PL *-on*)
uchamsyo'n 'he kills us' (A3 *u-* + *chamse* 'CAUS1 to kill' + B1PL *-on*)
uchamsyo'b' 'they kill (them)' (A3 *u-* + *chamse* 'CAUS1 to kill' + PL *-ob'*)
uwe'syo'b' 'they feed (them)' (A3 *u-* + *we'se* 'CAUS1 to feed' + PL *-ob'*)
inwa'ryo'x 'I'm telling you' (A1SG *inw-* + *a're* 'TV10 to say, tell' + B2PL *-ox*)
tyo'b' 'trees' (*te'* 'tree' + PL *-ob'*)
uchyo'b', ucho'b' 'they do (it)' (A3 *u-* + *che* 'TV10 to make, do' + PL *-ob'*)

Most cases of the contraction *e + o* yield *yo'* where *y* is what remains of the mid-high vowel *e*. An example for the result *o'* is *ucho'b'*, though here, too, *uchyo'b'* is more frequently attested.[217]

*e + u > *yu'?*

There is no grammatical combination where a first vowel *e* would combine with a second vowel *u*. With better understanding of Ch'orti' etymology and historical phonology, we might identify cognates that are results of a hypothetical **e + u*. Alternatively, one could examine Fought's (1972 and AILLA) data for contractions of *e + u* in rapid speech (definite article *e* + verb beginning in *u-*). The sequence is completely unattested in the New Testament (Wycliffe Bible Translators 2012).

6.1.3 i + V

The vowel *i* is involved in A2PL/C2SG *i-* but these forms are not very frequent in the corpus because the second person is generally used less frequently in texts than the third.

When *i* is the first vowel of the sequence V'V, it always turns into semi-vowel *y*:

[217] This contradicts Wichmann's (1999: 15) rule 1g that *y* does not appear after a *ch.*

i + a > ya'

ya'ti 'you bathe' (C2SG *i-* + *ati* 'IV25 to bathe')

onya'n 'former, of old' (*oni'* 'years ago' + *-an* 'suffix')

akb'ya'n b'u'r 'beans sitting out since yesterday on the fire' (*akb'i* 'yesterday' + *-an* 'suffix' + *b'ur'* 'beans')

ajtisya'r 'farter' (AGT *aj-* + *tisi* 'TV1/IV7 to fart'[218] + VN *-ar*)

tya'ch[219] 'from so much, after so much' (PREP *ti* + *ach* 'intensifier, own')

Historical evidence is found in the following lexeme:

PCH *ni?äl[220] 'son-in-law' > CHR *nya'r* 'son-in-law'

Even though only one example could be given for C2SG *i-* above, it matches the reflexes of contractions of *i + a* in other contexts and differs from what happens with A2PL when it attaches to a verb beginning with *a-*:

iwakta 'you (pl.) leave' (A2PL *iw-* + *akta* 'TV9 to leave')

i + e > ye'

ye'ra'ch 'you are true' (C2SG *i-* + *e'ra'ch* 'IV to be true')

ink'anye't 'I love you' (A1SG *in-* + *k'ani* 'TV1 to like, love, want' + B2SG *-et*)

k'axye'r 'fall (noun)' (*k'axi* 'IV1 to fall' + VN *-er*)

chye't 'who are you' (*chi'* ~ *chi* 'who' + B2SG *-et*)

ab'chin 'urinate!' (*ab'chi* 'TV1 to urinate' + IMP *-V/en*)

---

[218] If the verb is intransitive, it likely had a final **-j* and belongs in the following chapter (see discussion on *j loss* there).

[219] Hull (2016: 423) suggests *tya'* + *ach* as the source of *tya'ch*, which is possible. However, the basic preposition is *ti* ~ *ta* in Ch'orti'. The vowel *i* is the older form because it is *ti* (e.g., in Yucatecan) or *chi* (e.g., in Greater K'iche'an). Therefore, *ti* + *ach* is equally possible.

[220] Reconstructed by Kaufman and Norman (1984: 127) based on Chol, Chontal, Choltí and Ch'orti'. *Ä* is a sound that only exists in Western Cholan (Chol and Chontal) and goes back to an earlier short *a* etymologically (Kaufman & Norman 1984: 85). Otherwise, vowel quantity is not distinguished in modern Cholan.

CHL *nij'al* 'son/father/child/patent-in-law' (Hopkins, Josserand & Cruz Guzmán 2011: 160) with *j* does not seem to support this reconstruction.

CHN *ni'ä* 'son-in-law' (Keller & Luciano G. 1997: 172)

CHT <nial> 'father/mother-in-law' (Robertson, Law & Haertel 2010: 346).

There is again only one case of C2SG *i-* in the corpus and the verb already has a laryngealized vowel *e'*, so it is difficult to tell what the result of this contraction is. However, other contexts support a resulting *ye'*. It seems that cases where the expected vowel is missing often involve palatal sounds like *ch* in *ab'chi.*

Unfortunately, there are no examples of A2PL used either with a verb beginning with *e-* (there are not many transitive verbs that do) or with a noun beginning with *e-* (because the second person plural is rarely used altogether). We would expect a result **iwe-* parallel to, e.g., first person singular *niw-et'ok* 'my wife'.

i + i > i'

i'xin 'you (sg.) go' (C2SG *i-* + *ixin* 'IV to go')
sajmi'x 'recently' (*sajmi* 'today, this morning, the part of the day that has already passed' + =*ix* 'already')
lichi'k 'poke!' (*lichi* 'TV1 to poke' + IMP/OPT *-ik*)

We expect an outcome *i'* and this is in fact what we see in the data, even though there are not many examples. On the other hand, A2PL shows the usual glide *iw* when attaching to verbs or nouns starting in *i-*:

iwixner 'your (pl.) exits' (A2PL *i-* + *ixner* 'exit, leaving')
iwirsi'k 'you (pl.) make (it) known' (A2PL *i-* + *irse* 'CAUS1 to show, teach' + IMP/OPT *-ik*)

i + o > yo', o'

uyub'yo'b' 'they ask (it)' (A3 *u-* + *ub'i* 'TV to ask[221]' + PL *-ob'*)
uk'uxo'n 'it stings us' (A3 *u-* + *k'uxi* 'TV1 to bite, sting' + B1PL *-on*)

Again, we unfortunately do not have examples of C2SG *i-* with verbs beginning in *o-*. However, based on contractions from other contexts and the data discussed so far, we would generally expect *i + o* to contract to *yo'*. There is also some evidence for *o'* in *uk'uxo'n.* We are again dealing with a palatal vowel in whose environment an expected *y* is missing.

[221] Hull classifies this verb as intransitive class 25 but this is incorrect. The verb is clearly inflected as transitive morphologically and must at least historically be transitive.

Set A shows its characteristic glides as in *niw-ocher* ‘my entrance’ or *niw-otot* ‘my house’. However, we also observe forms that show contractions instead, just like in the case of *a* + *o*:

ta nyo’k ‘on foot’ (*ta -ok*[222] ‘on foot’ with PREP *ta* + A1SG *ni-* + *-ok* ‘foot’)
nyo’tot ‘my house’ (A1SG *ni-* + *otot* ‘house’)

This can, again, be explained by a sound change *wo > *o*.

i + u > yu’

yu’b’yan ‘you understand’ (C2SG *i-* + *ub’yan* ‘AP11 to understand’)
yu’sre ‘you (pl.) wanted (it)’ (Dugan 2013: 96) (C2SG *i-* + *usre* ‘TV12 to desire, want’)

The few examples that can be found show a result *yu’*. The same is seen with A1SG *ni-* on nouns:

nyu’nen ‘my son’ (A1SG *ni-* + *unen* ‘son (of a man), daughter (of a man)’)

This is explained by the proposed sound change of *uw > *u*: **niw-unen > **ni-unen > *nyu’nen*.

6.1.4 o + V

Finding relevant data on contractions where *o* is the first vowel is difficult since neither set A nor set C end in *-o*.

o + a > wa(’)?

ujenywach ‘habitually, customarily’ (A3 *u-* + *jenio* ‘custom [Span. *genio*]’ + *ach* ‘intensifier, own’)
u-tantwach ‘her portion’ (A3 *u-* + *tanto* ‘so much [Span. tanto]’ + *ach* ‘intensifier, own’) The examples suggest a result *wa* although from the following examples involving other vowels as well as from the evidence so far, we would expect a laryngealized vowel to be involved. Both examples, unfortunately, involve Spanish loanwords. However, there are but few transitive verbs that end in *-o* and these are not attested with suffixes starting in *-a* in my corpus.

[222] Obligatorily possessed *-ok*: requires a possessive suffix (because a foot usually belongs to somebody).

Intransitive verbs, on the other hand, diachronically end with a *-j* and are therefore discussed in chapter 6.2.

o + e > we'?

Theoretically, there should be cases of B1SG *-en* or B2SG *-et* suffixed to verbs ending in *-o* like classes TV3, TV23, TV25. They are not attested in the corpus but may be found in a larger dataset.[223] Based on all data of *o + V* and *-ojo-* contractions (the latter discussed in the next chapter) as well as the fact that their results generally seem to overlap, we would expect *o + e* to yield *we'*.

o + i > wi', i', i, o', o

Following the general trend, we would expect *o + i* to yield *wi'*. This is in fact what verbs ending in *-o* show when the clitic *=ix* is suffixed, though we also see *i'* or even *i*:

k'ojkwi'x -jab' 'to be X years old' (*k'ojko* 'TV3 to have' + *=ix* 'already' + [numeral +] *jab'* 'year')
warti'x (PROG *war* + *=to* 'still, yet' + *=ix* 'already')
ja'xtix 'unique, being the only one' (3SG *ja'x* + *=to* 'still, yet' + *=ix* 'already')

However, *warti'x* also has a variant *wartwi'x*, which fits better with the expected outcome *wi'*.

We again also have less clear examples involving a Spanish loan word:

otronte'yx ~ otro'nte'yx 'a different thing, an altogether different thing' (*otro* 'another [Span. *otro*]' + *in-* 'one' + CL *-te'* + *=ix* 'already')
otron- 'one more of (+NC)' (*otro* 'another [Span. *otro*]' + *in-* 'one')

o + o > o'

Theoretically, there should be cases of B1PL *-on* or B2PL *-ox* suffixed to verbs ending in *-o* like classes TV3, TV23, TV25. They are not attested in the corpus but may be found in a larger

[223] Wichmann's (1999: 15) rule 1h states that if the first vowel of a contraction is rounded, it will become *w* in most cases. Like me, he did not find examples of *o + e* yielding this result but since he also did not find counterexamples, he assumes that *o > w* is to be expected there, too.

dataset. The wordlist compiled by Súarez (1892) offers two examples which are, however, inconclusive due to unclear orthography and reconstruction:[224]

1892 <bohóz> ‘garbage [Span. basura]’ > *b’o’s* ‘brushwood’
1892 <pohó> ‘sore, wound [Span. llaga]’ > *po’(w)* ‘pus’

Still, there is some historical evidence of what happens when *o* + *o* contract in the thematic suffixes of Ch’orti’. Mediopassive verbs of class 4, e.g., *lok’oy* ‘MPAS4 leave, go out’, *k’otoy* ‘to arrive there’ and *ochoy* ‘to enter’, are formed with a suffix -*Vy*. This suffix is not used in the completive aspect:

1SG *lok’e’n* ‘I left’
2SG *k’ote’t* ‘you arrived there’
1PL *lok’o’n* ‘we left’
2PL –
3PL *lok’o’b’* ‘they left’

The forms of set B used with these verbs are commonly laryngealized. Based on the data I discussed so far, it is possible to assume that this is always the case when it is suffixed to a stem that ends in a vowel. Therefore, the verb stem in these cases cannot simply be *lok’* or *k’ot* or *och* – otherwise the result would have been **lok’en, not *lok’e’n* as well as **lok’on, not *lok’o’n* etc. The stem cannot be *lok’oy* either because there is no evidence that -*VyV*- sequences also contract[225] just like -*V’V*- or (as we will see in section 6.2) -*VjV*-. I instead assume that mediopassive class 4 verbs are formed to transitive stems like **lok’o* (Hull’s class TV3). The results in the paradigm above are just as expected for *o* + *o* because we would assume the result is **wo’* which then becomes *o’*. The fact that forms with B1SG, B2SG do not show *we’* but *e’* requires an explanation, however. It could be that we are dealing with a different kind of stem vowel in the assumed transitive verb – in that case, this would not be evidence for *o* + *o*.

The diachrony of this development will be discussed in chapter 7.4.2.

[224] It is not entirely clear to me whether he used <h> to record only glottal stops/laryngealization or whether there indeed is a *j*.

[225] HGM <ma-yu-yu>, <ma-yu> *mayu[y]* ‘mist, fog’ (Kettunen & Helmke 2020: 111) > CHR *mayuy* ‘smog, contaminated air, fog, smoke, blight (for corn)’.
HGM <MUYAL>, <mu-MUYAL-la>, <MUYAL-ya-la>, <MUYAL-la>, <MUYAL-li> *muyal/muyaal* ‘cloud’ (2020: 85, 112) > CHR *muyarir* ‘thickness of things grown together tightly’.
Lower-case letters indicate syllabograms in transliteration of Hieroglyphic Maya while upper-case letters refer to logograms.

o + u > u'

matu't 'it still isn't enough (for him)' (*ma* 'NEG' + =*to* 'still' + A3 *u-* + *ut* 'face')

Though we only have one example, the result *u'* fits the pattern. Theoretically, one could expect *wu' here as Wichmann (1999: 15) does according to his rule 1h, though he states that he did not find examples for *o + u*. This example is a counterexample. As already discussed, we generally expect *w* to be lost in front of *o* or *u*.

6.1.5 u + V

u + a > wa?[226]

b'ijnwach 'TV13 to imagine' (*b'ijnu* 'TV20 to think of, guess' + *ach* 'intensifier, own)

It is interesting that here, the laryngealization is missing in a lexeme that is not a Spanish loanword. However, since we only have a single example, it is difficult to tell whether this is truly the general outcome. Examples of A3 include:

uyajk'u 'he hits (it)'(A3 *u-* + *ajk'u* 'TV20 to hit')
uya're 'he says (it)'(A3 *u-* + *a're* 'TV10 to say, warn, advise, order, command')
uyalma 'her stomach' (A3 *u-* + *alma* 'stomach, chest [Span. *alma* 'soul']')

I have found no evidence of A3 *u-* participating in "set C"-type-like contractions, which we also do not expect since the glide employed with *u* is usually *y*. Though it sometimes becomes *w* before *i*, as previously discussed, this still does not produce a sequence *wo/wu* and therefore, the *w* has no reason to disappear to leave two contracting vowels.

u + e > we', e'

utakswe'n 'he pushed me' (A3 *u-* + *taksu* 'CAUS29 to push' + B1SG *-en*)
ink'ajb'we't 'I was remembering you' (A1SG *in-* + *k'ajb'u* 'TV.POS to think about, remember' + B2SG *-et*)

[226] Wichmann's (1999: 15) rule 1h states that if the first vowel of a contraction is rounded, it will become *w* in most cases. He did not find examples of *u + a* yielding this result but since he also did not find counterexamples, he assumes that *u > w* is to be expected there, too. This example proves it, especially taken together with those discussed under *uja* contractions, which yield the same result.

ulatz'we'n 'it obstructed me' (A3 *u-* + *latz'u* 'TV5 to obstruct, hinder' + B1SG *-en*)
utz'u'we'nix 'it has emaciated me' (A3 *u-* + *tz'u'* 'TV20 to suck, absorb' + B1SG *-en* + *=ix* 'already')
inch'ab'we't 'I lay you down' (A1SG *in-* + *ch'ab'u* 'TV.POS to lay down, put to bed' + B2SG *-et*)
unik'che'n 'it gave me a strong pain' (A3 *u-* + *nik'chu* 'TV6 to tilt, twist; strong pain from an illness' + B1SG *-en*)
ch'ab'un 'lay it down!' (*ch'ab'u* 'TV.POS to lay down, put to bed' + IMP *-V/en*)

In almost all cases, the result of *u* + *e* is *we'*. *Utz'u'we'nix* would theoretically be expected to be **utz'we'nix. The divergent result is perhaps due to the fact that the verb ends in a glottal stop, not merely a vowel. *Ch'ab'un* is likewise problematic but the imperative is a less reliable morpheme because, as the data so far show (and as discussed in Appendix E b.6), it is not *-en* in all cases but the exact conditions of the respective allomorphs remain obscure.

Adjectives belonging to Hull's class 26 that are derived from color terms with *-pwe'n* like likely include a sequence of *u* + *e* or *o* + *e* though it is unclear what suffix -pu/-po might be:

sakpwe'n 'having a light color' (*sak* 'mild, partial, semi-'[227] + *-pu/o* '?' + *en* 'derivational affix')

With A3, the usual glides are employed:

uyeb'ta 'he sent (it)' (A3 *u-* + *eb'ta* 'TV19 to send')
uyej 'its edge' (A3 *u-* + *ej* 'edge')
uyeroj 'his image' (A3 *u-* + *eroj* 'image, face')

u + i > wi', i'

ukejri'x 'he drags himself along the ground' (*kejru* 'TV20 to drag on the ground' + *=ix* 'already')
inlat'b'i'x 'I already support it' (A1SG *in-* + *latb'u* 'TV.POS to support, hold up' + *=ix* 'already')
ak'ajb'iken 'you have to remember me' (A2SG + *k'ajb'u* 'TV.POS to think about, ponder,

[227] Ch'orti' has developed new color terms based on reduplication of the Proto-Mayan forms. For example, 'white' is now *saksak*, while the original color terms have a vaguer meaning *sak* 'mild, partial, semi-, of medium intensity'. Interestingly, the derived verb *saka* 'IV2 to turn white' still attests to the original meaning.

remember' + IMP/OPT *-ik* + *=ix* 'already')
utajchwi'k 'let him cut it up' (A3 *u-* + *tajchu* 'TV20 to cut into pieces, cut up' + IMP/OPT *-ik*)

With *u + i*, there seems to be less of a tendency for a *w* to occur but I suspect this is due to the fact that *wi'* would result in too complex consonant clusters. In any case, the result is never the same as with set A, where the glide *w* occurs between the vowels:

uwira 'he sees (it)' (A3 *u-* + *ira* 'TV to see')
uwixka'r 'his wife' (A3 *u-* + *ixka'r* 'wife')

u + o > o'

usakro'n 'he orders us' (A3 *u-* + *sakru* 'TV20 to order, command' + B1PL *-on*)
uyajk'o'b' k'ewer 'they whip' (*ajk'u k'ewer* 'TV20PHR to whip'; A3 *u-* + *ajk'u* + PL *-ob'* + *k'ewer*)
uyak'o'n 'it provides (it)' (A3 *u-* + *ajk'u* 'TV20 'to give, deliver'+ B1PL *-on*)[228]
uch'ab'o'b' 'they laid (it) down' (A3 *u-* + *ch'ab'u* 'TV.POS to lay down, put to bed' + PL *-ob'*)
usakro'n 'he orders us' (A3 *u-* + *sakru* 'TV20 to order, command' + B1PL *-on*)

The outcome of *u + o* is *o'*. Based on the proposed sound change of *wo > o*, no *w* is expected. Set A has glides:

uyob'okres 'it perfumes (it)' (A3 *u-* + *ob'okres* 'CAUS16 to perfume')
uyose 'he puts (it) in' (A3 *u-* + *ose* 'TV10 to put in')
uyotot 'his house' (A3 *u-* + *otot* 'house')
uyok 'his leg' (A3 *u-* + *ok* 'leg')

u + u > u'

PCH *b'u'ul[229] 'bean' > CHR *b'u'r* 'beans'

The only evidence for the result of *u + u* is historical. A3 + verb beginning in *u-* always yields a glide:

[228] *J* sometimes disappears from consonant clusters in Hull (2016) or participates in metatheses.
[229] Kaufman and Norman (1984: 117) reconstruct PCH *b'uʔul 'bean' based on all Cholan languages.
CHL *b'u'ul* 'bean(s)' (Hopkins, Josserand & Cruz Guzmán 2011: 25)
ACN <buul> (Smailus 1973: 136); CHN *bu'u* (with final loss of *l* as already discussed in previous examples)
CHT <bul> 'bean [Span. *frijol* <frisol>]' (Robertson, Law & Haertel 2010: 320).

uyub'i 'he listened (to him)' (A3 *u-* + *ub'i* 'TV4 to hear, listen to')
uyuch'i 'he drinks (it)' (A3 *u-* + *uch'i* 'TV4 drink, drink alcohol')
uyunen 'his son' (A3 *u-* + *unen* 'son/daughter (of a man)')

In the following chapter we will investigate cases where similar contractions with verbs and nouns that begin with *j-* as well as the reason for why this happens.

6.2 Vowel contraction after intervocalic loss of -j-

Fought was to my knowledge the first to observe that in Ch'orti', *j* is "regularly lost adjacent to a morphological boundary between vowels" (Fought 1967: 112). As mentioned in section 4.1, Dugan describes that "the j is realized as a lengthened vowel that moves from voiced to devoiced and back to voiced, at least in some circumstances" (Dugan 2013: 21). Without going into detail, Hull states in the introduction to his dictionary that "many terms (noun and verbs) with an initial /j/ are regularly pronounced with /Ø/, i.e., the consonant is deleted" (Hull 2016: 8). Common examples where *j-* regularly drops are *jut ~ ut* 'surface; face; eye; appearance' and *jor ~ or* 'peak, high point; head; surface; boss, leader' (Hull 2016: 10). As Hull describes, the initial vowel is laryngealized when A3 *u-* is prefixed and the initial /j/ is dropped yielding forms like *u't* and *o'r*. According to him, the respective two forms, *ut ~ jut* and *or ~ jor*, are "in free variation".

I propose that this variation is the result of several sound changes. To understand my point, it is necessary to sketch the history of CHR *j. Proto-Mayan had two distinct sounds *h [h] and *j [x] (Campbell 2017: 46). This context was retained (in the correct etymological distribution) in Hieroglyphic Maya, as has been shown convincingly by Grube (2004).[230] The following examples show that 1) this distinction is lost in Ch'orti'[231] and 2) that Ch'orti' *j* (< PCH *h, *j) is mostly retained word-initially and word-finally:

[230] Grube tested the hypothesis that the hieroglyphic script uses distinct signs for syllables with *h* and with *j* by comparing the attested lexemes to cognates from languages that "either still have the *h : j* contrast or for which the contrast has been reconstructed from morphophonemic evidence" (Grube 2004: 68). This is important to determine whether the two sounds distinguished in the script really correspond to PM *h and *j etymologically. His data show that this is indeed the case: the distinction was upheld during most of the Classic period until the Late Classic, from where on the distinction stopped to be made word-finally. In the latest hieroglyphic texts, the Postclassic codices, it is only upheld in "a few contexts" (Grube 2004: 72).

[231] This preliminary list is based on Kaufman and Norman (1984). As already mentioned, they do not give the data that their reconstructions are based on. Because of the number of examples, I did not look up the data from the individual languages but this will be remedied in future publications dedicated to an updated

Initial *h, *j

PCH *haʔ > CHR *ja'* 'water'
*hab' > *jab'* 'year'
*häch > *jachi ~ achi* 'TV1 to raise, lift'
*hiʔ > *ji'* 'sand'
*hil > *jiri ~ iri* 'IV1 to rest'
*huj > *juj* 'iguana'
*hun > *jun* 'paper, book'

*jäb' > *jab'-i* 'to open'
*jit' 'to tie crossbars of structure' > *jit'-i* 'TV1 to tie tightly, cinch, squeeze with a rope'
*joj > *joj* 'heron'
*jol > *jor ~ or* 'head'
*jawän> *awan* 'woman's sister-in-law'

Here, we see sporadic cases of synchronic variation between *j-* and vowel-initial forms as in *jiri ~ iri* or even forms that are missing the initial *j-* altogether where we must assume based on related languages that it used to be there.

Final *h, *j

PCH *b'ah > CHR *b'aj* 'gopher'
*eh > *ej* 'tooth'
*neh > *nej* 'tail'

*chij 'deer' > *chij* 'horse, mule, beast of burden'[232]
*joj > *joj* 'heron'
*huj > *juj* 'iguana'

reconstruction of Proto-Cholan phonology. Where I do not provide a separate Proto-Cholan translation, it is the same as in Ch'orti'.

[232] This is a common semantic change among languages in the Americas. With the arrival of the Spaniards, the indigenous peoples were also confronted with horses – a new animal that needed a name and was apparently perceived as deer-like or, interestingly, tapir-like by the Maya (Kaufman & Norman 1984: 569, 583–584).

J likewise does not disappear in clusters:

PCH	*ahn	> CHR	*ajni* ‘IV6 to run’
	*b’ahläm/*bähläm	>	*b’ajram* ‘jaguar’
	*k’ahk	>	*k’ajk’* ~ *k’ajk* ‘fire’
	*sahb’in	>	*sajb’in* ‘weasel’
	*sahm-i ‘earlier today’	>	*sajmi* ‘today, this morning, the part of the day that has already passed’
	*b’ah-	>	*b’aj-xan* ‘first’
	*sih ‘gift’; *sih-i ‘to give a gift’ > *sij-pa* ~ *tzij-pa* ‘IV38 to give as a gift, share’		
	*tah	>	*taj-wi* ‘TV4 to find’
	*toj$_1$	>	*toj-b’ir* ‘straight’
	*toj$_2$	>	*toj-ma* ‘AP3 to pay’

1892 data from Suárez (Korovina 2021) and 1930 data from Wisdom (1950) provide evidence that we see the same phenomenon with word-final *j* as with word-initial *j* – it sometimes disappears where it is expected:

1892 <shanaj>	1930 <šanah>	2016 *xana* ‘IV2 to walk, travel’
1892 <guehej>	1930 <we’eh>	2016 *we’* ‘IV3 to eat’
1892 <guarachuhuj>[233]	1930 <ču’uh>	2015 *chu’* ‘IV3 to nurse’

This is easily explained because in all cases, we are dealing with forms that frequently receive affixes (verbs and nouns) . Once a prefix or suffix is added to a stem that begins or ends with *j* respectively, the affix moves *j* into a vowel-medial environment (e.g., *xanaj-en ‘I walked’) because prefixes commonly end in vowels and suffixes often begin in vowels. We know that vowel-medial *j* indeed historically disappeared in Ch’orti’:

CHL *pajäy*[234] 1892 <pajay> 1930 <p’a’i>[235] 2016 *pa’y* ‘skunk’

HGM <a/AJ-AJAN-na>[236] *ajan* ‘Ajan. Name of God E (“he of *elote*”)’ > CHR *a’n* ‘green ear of

[233] This is an inflected form *war a-chu’uj* ‘she is nursing’.

[234] Early Chol data based on (Hopkins, Josserand & Cruz Guzmán 2011). Also PM *pahar (Kaufman 2003: 575–576) with final *-r based on K’iche’an evidence, e.g., KCH paar.

[235] Wisdom recorded ejective *p’* for Ch’orti’, which is doubtful, even if one assumes, like Kaufman and Norman (1984: 87) that Ch’orti’ used to have it. As we see from the Chol evidence, this is not a lexeme where *p’* would be expected etymologically. Mistakes of this type are, unfortunately, frequent in Wisdom’s data.

[236] Alternative spelling <AJAN-na>, <AJAN> (Kettunen & Helmke 2020: 97).

corn’

HGM <k’u-hu-lu> *k’uhul* ‘god-like’[237] > CHR *ch’u’r* in *Uch’u’r e Witzir.* ‘Lord/God of the Hills’ (Hull 2016: 466)

In this chapter, I will demonstrate that the results of contractions happening after intervocalic loss of *j* are the same as those of contractions without *j*. This made it possible to reanalyze *j*-initial verbs as beginning in vowels (parallel to vowel-initial verbs):

j-initial:	V-verb	*ati* ‘IV25 to bathe’	:	*a’ti* (*a-ati*) ‘he bathes’
	j-verb	*jachpa* ‘MPAS1 to get up’	:	*a’chpa* (*a-jachpa*) ‘she gets up’
			↓	
	V-verb	*ati* ‘IV25 to bathe’	:	*a’ti* (*a-ati*) ‘he bathes’
	**j*-verb	***achpa*** ‘MPAS1 to get up’	:	*a’chpa* (*a-achpa*) ‘she gets up’

Likewise, verbs ending in *-j* were reanalyzed as ending in vowels (parallel to transitive verbs which historically end in vowels[238]).

j-final:	TV	*ira* ‘TV to see’	:	*ire’n* (*ira-en*) ‘he sees me’
	IV	*xanaj* ‘IV2 to walk’	:	*xane’n* (*xanaj-en*) ‘I walk’
			↓	
	TV	*ira* ‘TV to see’	:	*ire’n* (*ira-en*) ‘he sees me’
	IV	***xana*** ‘IV2 to walk’	:	*xane’n* (*xana-en*) ‘I walk’

Evidence for the fact that these verbs indeed used to have a final *-j* can sometimes be found in further derivations, e.g., *b’oro* ‘AP1 to abound, multiply, increase, produce, yield’ > *b’orojse* ‘CAUS12 to increase sth.’. Because we know that intransitive verbs used to end in *-j*

[237] Alternative spelling <k’u-ju-lu>, <K’UH-HUL>, <K’UH-JUL-lu>, <K’UH-JUL>, <K’U’-u-lu>; epithet of Maya kings (Kettunen & Helmke 2020: 110). It seems that only the Yucatecan form of this word is attested in the hieroglyphic texts; in Cholan languages, we expect PM *k to become *ch*. Kaufman and Norman (1984: 119) reconstruct this as *ch’uh ‘god, holy thing’ (*-ul* is a derivational suffix). According to them, the word is unattested in Ch’orti’ but I find it likely that *Uch’u’r e Witzir* is indeed a cognate given the semantics of ‘God’. Otherwise, they used CHL, ACN, CHN and CHT.
CHL *ch’ujul* ‘spirit’ (Hopkins, Josserand & Cruz Guzmán 2011: 55)
ACN <chuhul> ‘worship [Germ. verehren]’ (Smailus 1973: 143)
CHN *ch’u’ulchere, ch’u’ul ayan* ‘holy’ (Keller & Luciano G. 1997: 397).
CHT <chu> ‘idol’ (Robertson, Law & Haertel 2010: 326) 326.

[238] Wisdom seems to consistently record intransitive verbs in all inflectional classes and derivations (antipassives, passives etc.) with a final *-h*. Likewise, he seems to be consistent in writing transitive verbs ending with vowels, e.g., *ira* ‘see or look’ (Wisdom 1950: 484) or *chamse* ‘kill, murder, butcher’ (Wisdom 1950: 693).

historically, I treat all contractions that take place when elements are suffixed to intransitive verbs as contractions involving *j*. However, this is not necessarily the case. One could likewise assume that the forms are formed synchronically from the now vowel-final verbs. Most likely, in these contexts, intervocalic *j* first disappeared leaving a vowel sequence to which then *ʔ* was inserted to avoid hiatus. Transparent compounds like reduplicated *jaja'r* 'rain' or forms involving the agentive prefix *aj-*[239] are not subject to loss of *j* vowel contractions.

It is even possible to date these changes. Intervocalic *j* apparently begins to disappear between 1892 and the 1930s because it is still recorded in Suárez's data but not anymore in Wisdom's (<pahay> versus <p'a'i>). Since Wisdom diligently recorded final *-j* of intransitive verbs in the 1930s but they are not found anymore in Fought's data from the 1960s, this stem reanalysis must have happened sometime in between. Wisdom transcribes most cases of *j* with <h>, not <x> (although he uses that sign, as well) which means that by then, Ch'orti' *j* likely was phonetically [h] in most positions. This change apparently happened in all intransitive verbs because no variation is recorded. Finally, the variation of initial *j-* versus initial vowel in some verbs and nouns suggests that a similar process is on the way with stem onsets but it has not been established throughout the language yet.

The sound change that led to the disappearance of *j* in word-final position is the same as the one that affects *j* in word-initial position – both happen because *j* finds itself in a vowel-medial position following the attachment of affixes. What is not the same, however, is the analogy that happened afterwards and is responsible for the fact that all verbs are synchronically vowel-final and have lost *j completely while the same is only true for *j* in initial position in some verbs and, apparently, not in all speakers equally. This provides the perfect situation to study when analogies do and do not happen in the future.

In the following chapters, sequences involving different vowels will be discussed. The goal is 1) to find out how set C and set A interact with stems beginning in *j-* and 2) to compare these results to the contractions that happen with sequences of *V'V.*

[239] Fought (1967: 113; 1972: 29) reports a form *a'ter* 'bather' (AGT *aj-* + *ati* 'IV25 to bathe + VN *-er*) but this is unattested in modern data (Hull 2016; Pérez Martínez 1994). Instead, we see a variation of, e.g., *ajakta'r ~ ajkta'r* 'dancer' (AGT *aj-* + *akta ~ ak'ta* 'IV to dance' + VN *-er*).

6.2.1 ajV

aja > a'

a'b'tz'a 'it is free' (C3 *a-* + *jab'tz'a ~ ab'tz'a* 'MPAS2 to be free, unoccupied')
a'chpa 'she gets up' (C3 *a-* + *jachpa ~ achpa* 'MPAS1 to get up')
a'ktz'a 'it decreases' (C3 *a-* + *jaktz'a ~ aktz'a* 'MPAS2 to decrease, diminish')
a'tz'arna 'he is afflicted by a sickness' (C3 *a-* + 'PAS2 to be afflicted by a disease with shakes')
a'ytz'ix 'she is already hungry' (C3 *a-* + *jaytz'a ~ aytz'a* 'MPAS2 to be hungry' + =*ix* 'already')
a'tz'i 'you (sg.) beat it' (A2SG *a-* + *jatz'i ~ atz'i* 'TV1 to hit, strike, whip') (Fought 1967: 111)

When C3 *a-* is prefixed to verbs beginning in *ja-*, the result is always *a'*, exactly as was the case with contractions of *a* + *a*. This variation is recorded for both intransitive and transitive verbs. For transitive verbs, there are fewer examples of contracted forms in the corpus (because A2SG is used less frequently than C3) but Hull describes variation for a handful of stems, e.g., *jachi ~ achi* 'TV1 to pass so. over for a promotion, out-perform so. at work' or *jajpi ~ ajpi* 'TV1 to fit, hold'. We see that the reanalysis indeed must have happened in forms with A3 where, e.g., *uyachi* and *uwachi* are attested alongside more regular *ujachi*. The glides suggest that the verbs are viewed by some speakers (or in some dialects) as vowel-initial.

ujachi ~ uyachi ~ uwachi 'she lifted (it) up' (A3 *u-* + *jachi ~ achi* 'TV1 to lift up, raise, build')
ujajpye'n ~ uwajpye'n 'it seized me' (A3 *u-* + *jajpi ~ ajpi* 'TV1 to fit, hold' + B1SG *-en*)
uwajya'r 'the hanging of (the *petate*)' (A3 *u-* + *jajya'r ~ ajya'r* 'hanging, laying out')
uwatz'i 'he whipped (him)' (A3 *u-* + *jatz'i ~ atz'i* 'TV1 to hit, strike, whip')
uwayi ~ ujayi 'she spreads (it)' (A3 *u-* + *jayi ~ ayi* 'TV1 to spread out')

This synchronic variation is also attested for nouns:

jam ~ am 'space, space between things'

As to the historical evidence, we have already discussed *pa'y* 'skunk' and *a'n* 'green ear of corn'. Further historical evidence based on Suárez (Korovina 2021) and Kaufman and Norman (1984) is:
1892 <sajay> *tzajay* 'to make happy; happy [Span. *alegrar*; *alegre*]' > 2016 *tza'y* 'IV be happy, be joyful'
1892 <lajar> *lajar* 'same [Span. *igual*]' > 2016 *la'r* 'similar to, like, having characteristics of, same'

1892 <chajac> *chajak* ‘storm [Span. *tempestad*]’; <chahac> *cha’(a)k* ‘lightning bolt, ray [Span. *rayo*]’ > 2016 *cha’k’ ~ cha’k* ‘slug, snail’, *cha’k’-tun ~ cha’ktun* ‘obsidian, flint (*tun* ‘stone’: i.e., thunder stone)’[240]

PCH *kaj ‘to begin’[241] > CHR *ka’y* ‘IV to begin’

PCH *majan ‘borrowing’[242] > CHR *ma’n* ‘hired hand/field laborer’

PCH *ajaw ‘king, lord’ > CHR *a’aw=chan* ‘rattlesnake [Span. *víbora de cascabel*, “*culebra rey*”]’[243]

aje > e’ (a’?)

e’kcha ‘it opens up’ (C3 *a-* + *jek’cha ~ ek’cha* ‘INCH to open up’)[244]

e’k’o ‘they are hanging down’ (C3 *a-* + *jek’o ~ ek’o* ‘AP1 to hang down (of branches of palms)’)

e’k’pa ‘it spreads out’ (C3 *a-* + *jek’pa ~ ek’pa* ‘MPAS1 to open out, stretch out open’)

a’b’tz’a ‘they slide open’ (C3 *a-* + *jeb’tz’a ~ je’b’tz’a ~ eb’tz’a* ‘MPAS2 to open up, rip open’)

The last example shows *a’*, which is unexpected. More data is needed to determine the conditions of this peculiar change (e.g., possibly, though in my opinion unlikely, due to implosive *b’*).

[240] I will discuss this etymology in a separate upcoming paper. It was also one of the hypotheses that I discussed during my defense in May 2024.

[241] This word is unattested in Ch’orti’ according to Kaufman and Norman (1984: 122) and reconstructed based on Chol, Acalan and modern Chontal and Choltí.
CHL *kaj-* ‘to begin’ (Hopkins, Josserand & Cruz Guzmán 2011: 93)
ACN <cah-el> ‘to begin’ (Smailus 1973: 153); CHN *kaye* ‘to begin’ (Keller & Luciano G. 1997: 345)
CHT <cahez> ‘to begin’ (Robertson, Law & Haertel 2010: 304).
I suggest that *ka’y* is the Ch’orti’an cognate – the semantics fit perfectly and the result of the contraction is the same as in the other cases and, most importantly, the development is parallel to that of *tza’y* in the line above. This makes the verb a member of mediopassive class 4 in *-Vy*. The root also might survive in *kajyes* ‘CAUS2 to begin, start’.

[242] Reconstructed by Kaufman and Norman (1984: 125) based on Chol and external evidence (PTS *majan, PM *majaan). CHL *majan* ‘loaned’ (Hopkins, Josserand & Cruz Guzmán 2011: 138).
The semantic change of ‘borrowing’ to ‘hired hand’ in Ch’orti’ is made plausible by the verb, which underwent the same change: PCH *majn-ä ‘to borrow’ > CHR *majna/majnu* ‘TV9 hire, employ’.

[243] This word is unattested in Ch’orti’ according to Kaufman and Norman (1984: 115) and Hull (2016) although most other Mayan languages have it and it is an important title in Hieroglyphic Maya, but apparently present in this compound (Kaufman 2003: 639).

[244] Ejective *k’* may lose its glottalization as the first member of consonant clusters as described by Hull (2016: 9). Note that this is not an obligatory phenomenon because the following example keeps *k’*.

aji > e'y, e', i'

ajiri ~ e'yri 'he rests' (C3 *a-* + *jiri ~ iri* 'IV1 to rest')
ke'yni 'we prune (them)' (A1PL *ka-* + *jini ~ ini* 'TV1 to prune, pick fruit, fell (a tree)')
e'yxpa 'they are degrained' (C3 *a-* + *jixpa* 'MPAS1 to be degrained')
chokme'yb' 'anus' (*chokma*[245] 'to cast out' + *-ib'* 'instrumental suffix')
ajinpa ~ ajimpa ~ e'mpa 'they fall down' (C3 *a-* + *jinpa ~ inpa* 'MPAS1 to fall down')
ketpi'x 'it became frayed' (*ketpa*[246] 'MPAS1 to remain, stay' + *=ix* 'already')

In those cases where *aji* sequences do contract, the result is *e'y* except for the case of *e'mpa* where the glide is perhaps prevented by the otherwise resulting consonant cluster of *-ynp/ymp-*. There, we only see *e'.* As mentioned in footnote 201, *=ix* sometimes is not laryngealized even when a vowel contraction took place. In the case of *ketpi'x*, it is laryngealized but shows an unexpected vowel quality. The following forms of set A are evidence for the fact that the verb is sometimes reanalyzed as beginning with a vowel because a prevocalic glide is used:

inw-ini 'I cut (it) down' (A1SG *in-* + *jini ~ ini* 'TV1 to prune, throw down, pick fruit, fell (a tree)') instead of *injini*.
ujixi ~ uwixi 'he shells (it)' (A3 *u-* + *jixi ~ ixi* 'TV1 to shell (corn)')

Historical evidence is found in the lexeme *pe'ych* 'tomato'. This is similar to but not etymologically identical with another Mayan word for tomato, *pix*, which is most of all used in Greater K'iche'an languages. The Morán Manuscript also records <pix> for tomato alongside <paac> *pak?* (Robertson, Law & Haertel 2010: 350) but apparently, this lexeme is unattested in Ch'orti', which has the lexeme *pe'ych* instead. Kaufman (2003: 1133) suggests a Western Mayan (WM+) reconstruction *paj 7ik based on said Ch'orti' form and also CHJ *paj'ich*, QAN *pajich*, AKA, IXL *pa7ich*. Though nothing is stated on the semantics of the two elements, it does show the expected development in Ch'orti' compared to the other languages.

ajo > o'

[245] Though only *chokmayan* 'AP10 throw, throw away; abandon; hand over to another' is attested, apparently a form *chokma* used to exist, as well. According to Quizar (2020), the antipassive suffix *-ma* is a Ch'orti' innovation. A parallel construction seems to exist in Huastec (Kondic 2012: 52) where the vowel of *-Vm* is lost when other aspect suffixes are attached. Wisdom's (1950: 705) data show that the verbs in *-ma* used to have an additional final *-j*, e.g., in this case, *chokmaj*.
[246] *Ketpa* used to have a final *-j* according to Wisdom (Korovina 2019: 34).

o'k'o 'he scratches' (C3 *a-* + *jok'o ~ ok'o* 'AP1 to scratch')
o'nokno'b' 'they buzz' (C3 *a-* + *jononka ~ ononka* 'AFV2 to growl, buzz' + PL *-ob'*)
ub'o'b' 'each other' (A3 *u-* + *-b'a*[247] 'self' + PL *-ob'*)
takpo'x 'you (pl.) have made progress' (*takpa*[248] 'MPAS1 to make progress' + B2PL *-ox*)

Though there are not many examples of verbs in *jo-* contracting with set C in the corpus, the result is *o'* as expected. Additionally, there is some evidence of some of the verbs being reanalyzed as starting with a vowel because set A sometimes appears featuring prevocalic forms with glides:

ujochi ~ uwochi 'he poured (it)' (A3 *u-* + *jochi ~ ochi* 'TV1 to empty (liquid, or otherwise), pour out by turning a container upside down')
ujok'i ~ uwok'i 'he scratched (it)' (A3 *u-* + *jok'i ~ ok'i* 'TV1 to dig, scratch')

It is interesting that the prevocalic form of A3 *u-* occurs more often with *w* as a glide than *y*, regardless of what vowel follows. This is a point where this phenomenon seems to differ from original set A glides. The reason for this is that the paradigm of set A synchronically is being regularized by the generalization of the glide *w* even for A3 *u-*. While most of the originally vowel-initial verbs are still used with *uy-*, those transitive verbs that only become vowel-initial recently (via the loss of *j* and vowel-contraction discussed in this chapter) receive the new, more regularized allomorph of set A *uw-*.

aju > o'y

o'ych'u 'she grinds ' (C3 *a-* + *juch'u ~ uch'u* 'IV22 to grind corn, scrape')
o'jyi 'he breathes heavily' (C3 *a-* + *jujyi ~ ujyi* 'IV25 to sigh, breathe heavily from being tired')

[247] Originally *b'aj* (Wisdom 1950: 577).
[248] Originally *takpaj* according to Wisdom (1950: 661) although the semantics are different: 'dry of itself, become dry (as the weather), die or wither (as a plant)'. Still, even if the verbs are different, they belong to the same inflectional class (MPAS3) and these verbs, as *ketpa* discussed above shows, apparently all used to have a final *j*.

If *j* is lost and the remaining vowels contract, the result is *o'y* or *o'* as with *a* + *u*. Among nouns, this involves body parts like *(j)ut* 'face'[249] which are mostly used with a set A prefix and therefore the initial *j* is found especially frequently in intervocalic position.

ti ko'yt 'before us' (PREP *ti* + A1PL *ka-* + *(j)ut* 'face')
o'yt 'your face' (A2SG *a-* + *(j)ut* 'face')

There is also a historical example from Galindo (1834).

1834 <Unacaut> 'eye' > 2016 *unak'u't* 'your eye(s)' (A3 *u-* + *nak'* 'inside part' + A2SG *a-* + *ut* 'eye')

6.2.2 ejV

Since *e* is not part of either set A or set C, I will only discuss a handful of examples of contractions with *e* as the first vowel.

The results are apparently identical to those of "simple" vowel contractions where no *j* is involved. For instance, an *-eje-* becomes *e'* while *-eji-* contracts to *e'y*. The *u'* in *inxejb'u't* is either proof for *a* + *u* or *-aju-* (see discussion above):

wya'r 'food' (*we'*[250] 'IV3 to eat' + VN *-ar*)
k'uxe'yr 'toothache' (compound of *k'ux-* 'TV1 to bite, sting' + *ej* 'tooth' + *-ir* 'abstractive suffix')

---

[249] It is unclear whether *ut* 'face' used to begin with a vowel or with *j. On the one hand, Kaufman (2003: 324) reconstructs it with a "weak h" (PM *Haty and PCH *(h)ut) which he defines in the following way:
"The Mayan etymology for 'eye/face' is one of the most complicated that exist. I reconstruct its etymon as a *-Haty. *H ("weak h") is a morphophoneme that is */h/ word-initially, but disappears, like /7/ does, after ergative [NK: = set A] prefixes. Two other common Mayan roots beginning with *H are *Ha7 'water' and *Ho7 '5'. Do not make the mistake of believing that all initial */h/s are weak in this way". (Kaufman 2015: 755).
Hull gives the form invariably as *ut* in the dictionary entries but notes a variation *jut ~ ut* in the introduction (Hull 2016: 10). Already in Hieroglyphic Maya, the word is spelled without *h*, e.g., <u-ti> for *uut* in Kettunen & Helmke (2020: 119), though it is given as the lemma *hut ~ ut* elsewhere (Kettunen & Helmke 2020: 104). Since we expect the same result for *a* + *u* and for *-aju-*, *o'y* could theoretically point to either onset. On the other hand, *u't* 'his face' (A3 *u-* + *ut* 'face') rather points to an initial *j-* because the contraction of A3 *u-* does not happen with historically prevocalic words like *ok*, while they do happen with *jor* > *o'r* (*u-jor*). Furthermore, Suárez records 1892 *ujut* <ujut> 'almond, kernel [Span. almendra]' which must be cognate because, aside from 'face, eye' or 'surface', Ch'orti' *ut* also means 'grain, seed' today. As he writes it with <j>, not <h>, this seems to suggest that there is indeed a *j* between the vowels. However, this then does not match the following example from Galindo.
[250] Originally *we'ej* according to Wisdom (1950: 756).

akb'aryo'b' 'it got dark' (*akb'are*[251] 'IV9 to be night, get dark at night' + PL *-ob'*)
inxejb'u't 'I can't stand (him)' (A1SG *in-* + *xejb'e (ut)* 'TV10 to dislike, despise' + A3 *u-* + *ut* 'face')

There is also historical evidence for *-eje-*:

PCH *nejep'[252] 'aged' > CHR *ne'p* 'half-ripe'.

6.2.3 ijV

The result of an earlier *-ija-* is *ya'*:

ya'tzi 'you (pl.) beat it' (A2PL *i-* + *jatz'i* ~ *atz'i* 'TV1 to hit, strike, whip') with loss of glottalization on *tz'* (Fought 1967: 111)
takya'r 'dryness' (*taki*[253] 'IV1 to dry up, wither' + VN *-ar*)
pujk'ya'r 'birth, hatching' (*pujk'i*[254] 'IV6 to be born' + VN *-ar*)

The result of an earlier *-ije-* is *ye'*.

pujk'ye'n 'I was born' (*pujk'i* < *pujk'ij* (see above) 'IV6 to be born' + B1SG *-en*)
tarye'n 'I came' (*tari*[255] 'to come' + B1SG *-en*)
k'axye'r 'fall (noun)' (*k'axi*[256] 'IV1 to fall, fall down' + VN *-er*)
ab'che'n (*ab'chi*[257] 'IV25 to urinate' + IMP *-V/en*)

-Iji- yields *i'*:

k'axi'x 'they have now befallen' (*k'axi* < *k'axij* (see above) 'IV1 to fall, fall down' + *=ix* 'already')

A contraction of the sequences *-ijo-* yields *yo'*:

---

[251] Originally *ak'b'arej* according to Wisdom (1950: 450).
[252] Reconstructed by Kaufman and Norman (1984: 126) based on Chol and Ch'orti' as a lexeme without outside cognates.
CHL *nejep'* 'old person' (Hopkins, Josserand & Cruz Guzmán 2011: 158).
[253] Originally *takij* (Wisdom 1950: 660).
[254] Originally *pujk'ij* (Wisdom 1950: 1972). Note that Wisdom transcribed it <puhq'ih> but his identification of the uvular *q* in Ch'orti' is considered incorrect because the sound is not found where it would etymologically be expected.
[255] Originally *tarij* (Wisdom 1950: 664). Somehow, Hull (2016) is missing this lemma. However, the verb is clearly still used in the language. The class is likely IV1, Hull's root intransitive verb in *-i*.
[256] Originally *k'axij* <q'aših> (Wisdom 1950: 611)
[257] Originally *ab'chij* (Wisdom 1950: 456).

yo'bi[258] 'you (pl.) wrap (it)' (A2PL *i-* + *job'i* 'TV1 to cover and wrap') (Fought 1967: 111)
yu'ch'ru 'you (pl.) grind (it)' (A2PL *i-* + *juch'ru* 'ITER to grind corn continually') (Fought 1967: 111)
uyub'yo'n 'we are able to [lit. "it is possible to us"]' (A3 *u-* + *ub'i* 'IV25 to be able to'[259] + B1PL *-on*)
pujk'yo'n 'we were born' (*pujk'i* < *pujkij* (see above) 'IV6 to be born' + B1PL *-on*)
ch'yo'n 'we grew up' (*ch'i'*[260] 'IV3 to grow, develop' + B1PL *-on*)
nyo'r 'my hair' (A1SG *ni-* + *jor* 'head, hair')

-Iju- yields *yu'* or *u'*:

o'jtz'u't 'you worship (God)' (C3 *a-* + *ujtz'i (ut)* 'TV4 to worship by bowing down' + A3 *u-* + *ut* 'face')
nyu't 'my face' (A1SG *ni-* + *ut* 'face')
umayu't 'she confounds (him)' (*mayi jut* 'TV1PHR to deceive, bewitch, confound': A3 *u-* + *mayi* + A3 *u-* + *ut* 'face')

Overall, the results match those of *i + V*.

6.2.4 ojV

Just like *e*, *o* is not included in the forms of set A or set C. Some examples can nevertheless be found. *-Oja-* becomes *wa'*:

k'inwa'r 'divination' (*k'ino*[261] 'AP1 to divine' + VN *-ar*)
lajchwa'r 'scratching' (*lajcho*[262] 'AP1 to scratch, scrape' + VN *-ar*)

-Oje- yields *we'*:

chorwe'n 'I worked in the cornfield' (*choro*[263] 'AP1 to work in the cornfield' + B1SG *-en*)
ch'amwe'n 'harvest!' (*ch'amo*[264] 'AP1 to harvest' + IMP *-V/en*)

---

[258] Fought writes <yo'pi> with loss of glottalization and *p* instead of b since this is according to him the phonetic reality of the sound in most dialects (see also chapter 4.1).
[259] Originally *ub'ij* (Wisdom 1950: 747). It is interesting that *ub'ij* synchronically is a homophone of the transitive verb *ub'i* 'to hear'.
[260] Originally *ch'i'j* (Wisdom 1950: 720).
[261] Originally *k'inoj* (Wisdom 1950: 504).
[262] I did not find this form in Wisdom (1950) but would expect it to also have been *lajchoj like the antipassive above it.
[263] Originally *choroj* (Wisdom 1950: 707).

The result of *-oji-* is *wi'*:

ach'okwi'x 'it is sprouting' (C3 *a-* + *ch'oko*[265] 'AP1 have shoots (a plant), to be sprouting' + *=ix* 'already')
apewi'x (C3 *a-* + *peto* 'AP1 have branches hanging down' + *=ix* 'already')

We lack clear data for the result of an *-ojo-* or *-oju-* contraction.

6.2.5 ujV

The result of a contraction of *-uja-* is *wa' ~ a'*:

tura'ch 'there are some' (*turu*[266] 'to exist' + *ach* 'intensifier, own')
arwa'r ~ arwar 'crying, howling whining, screaming (*aru*[267] 'IV10 to cry, scream' + VN *-ar*)

There is evidence that some verbs have been reanalyzed as beginning with a vowel because forms with glides appear:

ujachi ~ uyachi ~ uwachi 'she lifted (it) up' (A3 *u-* + *jachi ~ achi* 'TV1 to lift up, raise)

The sequence *-uje-* yields *we' ~ e'*:

arwe'n 'cry!' (*aru* < *aruj* (see above) 'IV10 to cry, scream' + IMP *-V/en*)
ture'n 'I live' (*turu* < *turuj* (see above) 'IV22 to exist, be alive' + B1SG *-en*)

The difference of *we' ~ e'* does not seem to be conditioned by the preceding consonants because the sequence *ru-en* becomes *rwe'n* in one case and *re'n* in another.

-Uji- becomes *wi'*:

amukwi'x '(the moon) is waning' (C3 *a-* + *muku*[268] 'IV22PHR to wane (the moon)' + *=ix* 'already')
turikon imb'utz 'let's behave' (*turu imb'utz* < *turuj imb'utz (see above, though not attested

[264] Originally *chamoj* (Wisdom 1950: 715).
[265] This precise form is not found in Wisdom (1950). However, I expect the verb to have had a *-j* like the others of that antipassive class.
[266] Originally *turuj* (Wisdom 1950: 678).
[267] Originally *aruj* (Wisdom 1950: 457). Note that Wisdom records the form with laryngealized first vowel: <a'ruh>.
[268] I did not find this verb in Wisdom (1950) but assume that it also used to end in *-j* like the other intransitive verbs seen so far.

as a phrase in Wisdom) 'IV22PHR to behave; be well'; *turu* + IMP/OPT *-ik* + B1PL *-on* + *imb'utz*)

-Ujo- yields *wo'*:

aku'rwo'b 'they are having sex' (C3 *a-* + *kuru*[269] ~ *ku'ru* 'IV22 to have sex' + PL *-ob'*)
turo'nach 'we live' (*turu* < *turuj* (see above) 'IV22 to exist, be alive' + B1PL *-on* + *ach* 'intensifier, own')
turo'x 'you (pl.) are there' (*turu* < *turuj* (see above) 'IV22 to exist, be alive' + B2PL *-ox*)
o'r 'his head' (A3 *u-* + *jor* 'head')
to'r upix 'on bent knee, on one's knees' (*ta* 'preposition' + A3 *u-* + *jor* ~ *or* + A3 *u-* + *pix* 'knee(cap)')

The second example even involves a contraction of three vowels *a* + *u* + *jo*. Interestingly, though one could have expected this, no cases of contractions of *a-jor* 'your head' (**o'r) are attested in the corpus at all (therefore the form was not discussed in chapter 6.2.1). Perhaps this is due to the corpus size. In any case, as long as the formation is still transparent synchronically, forms like *ajor* can be reintroduced into the language. The fact that *o'r* would be the expected reflex of both 'your' and 'his/her head' may have contributed to the arisal of the new formation *ajor* in situations where disambiguation between the two meanings was needed.

There are some examples of *-uju-* contractions. The result is *u'*:

u'ch'ru 'he grinds (it)' (A3 *u-* + *juch'ru* 'ITER to grind corn continually') (Fought 1967: 111)
u'ryob' 'they shoot it' (A3 *u-* + *juri* ~ *uri* 'TV1 to throw, shoot') (Story "E noya i e sitz'b'ir" in Pérez Martínez 1996))
u't 'his face' (A3 *u-* + *ut* 'face')

For the historical evidence, recall the following dataset and its discussion in 6.2 (with the same caveat that only the Yucatecan form seems to actually be attested in the hieroglyphic corpus):

HGM <k'u-hu-lu> *k'uhul* 'god-like'[270] > CHR *ch'u'r* in *Uch'u'r e Witzir.* 'Lord/God of the Hills' (Hull 2016: 466) (A3 *u-* + *ch'u'r* '~divine'+ DEF *e* + *witz* 'mountain' + *ir* 'abstractive')

[269] Originally *kuruj* <quruh> (Wisdom 1950: 603).
[270] Alternative spelling <k'u-ju-lu>, <K'UH-HUL>, <K'UH-JUL-lu>, <K'UH-JUL>, <K'U'-u-lu>; epithet of Maya kings (Kettunen & Helmke 2020: 110). In Cholan languages, we expect PM *k to become *ch*,

In the following chapter I will describe some interesting interactions of vowel- or *j*-initial verbs with A/C1SG *in-* and C2PL *ix-*.

6.3 Vowel contractions with A/C1SG *in-* and C2PL *ix-*

We would not expect C1SG *in-* or C2PL *ix-* to participate in any contractions because these forms do not end in a vowel and thus do not produce a *-V'V-* sequence when attached to vowel-initial verbs. However, contractions do happen at least with C1SG *in-* as well and they merit a closer look. In the following examples, it seems as if the vowel in *in-* assimilates to the stem vowel and becomes laryngealized:

a'nti 'I bathe' (C1SG *in-* + *ati* 'IV25 to bathe')

This reminds us of the contractions discussed for *V + V* and *VjV*. However, additionally, *n* somehow ends up on the wrong side of the vowel through a kind of metathesis:

in-aC > a'nC

The phonetic reality of these contractions can easily be understood if we recall the discussion on vowels from chapter 4.1 where it was mentioned that Ch'orti' has nasalized vowels. Quizar (2023: 261) also describes the prevocalic allomorphs of C1SG *in-* as nasalized and laryngealized vowels *ã'- / ẽ'- / ĩ'- / õ'- / ũ'*. Fought transcribes forms inflected with *in-* as follows:

C1SG *in* + *ojri* 'IV25 to fall, fall down'	→	<õõ'ŋjri> 'I fall' (Fought 1967: 106)
C1SG *in* + *irna* 'PAS2 to be seen'	→	<ĩĩ'rna> 'I am seen' (Fought 1967: 106)

Fought (1967: 106) claims that the nasalization part of this is optional while assimilation of *i-* to the stem initial vowel of the verb as well as insertion of a glottal stop/use of creaky voice is "automatic". In the case of nasalization, *n* can disappear entirely but this, too, is optional. If it remains, it assimilates to the following consonant as already mentioned (see chapter 4.1). I suggest that historically speaking, we are dealing with nasalized vowels in all cases with

---

which is why Kaufman and Norman (1984: 119) reconstruct this as *ch'uh 'god, holy thing' (*-ul* is a derivational suffix). According to them, the word is unattested in Ch'orti' but I find it likely that *Uch'u'r e Witzir* is indeed a cognate given the semantics of 'God'. Otherwise, they used CHL, ACN, CHN and CHT.
CHL *ch'ujul* 'spirit' (Hopkins, Josserand & Cruz Guzmán 2011: 55)
ACN <chuhul> 'worship [Germ. verehren]' (Smailus 1973: 143)
CHN *ch'u'ulchere, ch'u'ul ayan* 'holy' (Keller & Luciano G. 1997: 397)
CHT <chu> 'idol' (Robertson, Law & Haertel 2010: 326) 326.

sporadic reemergence of a consonantal *n* in particular phonetic environments[271] that would need to be worked out in a bigger study. Assuming nasalized vowels gives us a plausible scenario for the otherwise somewhat peculiar sound change with metathesis:

**ĩ(ĩ)'VC > *ĩ(ĩ)'ṼC > Ṽ(Ṽ)'C*[272] instead of *in-VC > V'nC.*

Of course, if this is a sound change, we expect it to happen in other areas of the language, as well. Further research is required to identify potential cases.

I here present some examples (Fought 1967 with updated orthography; Hull 2016) that demonstrate the phenomenon for all vowels. It affects both C1SG and A1SG, though not to the same extent. The discussion in chapters 6.1 and 6.2 and the examples assembled in Fought (1967: 108–111) clearly suggest that transitive vowel-initial verbs may only participate in the contraction with A1SG when they begin with *o-/u-* while all *j*-initial transitive verbs[273] may participate in it. The following examples will confirm this distribution.

in + C-

inpatna 'I work' (C1SG *in-* + *patna* 'IV to work, labor)
ink'ani 'I want (it)' (A1SG *in-* + *k'ani* 'TV1 to like, love, want')
imb'isi 'I measure (them)' (A1SG *in-* + *b'isi* 'TV1 to measure')

With words that begin with consonants (except for *j*), no contractions take place. However, as we have seen a couple of times by now, *n* tends to assimilate to the following consonant, e.g., > *m* before *b'*.

in + (j)a- > a'n

a'nti 'I bathe' (C1SG *in-* + *ati* 'IV25 to bathe')
a'nchpa 'I wake up' (C1SG *in-* + *achpa* 'MPAS1 to get up')
a'nsi 'I play' (C1SG *in-* + *asi* 'TV4 to play')

[271] A typological parallel is found in Polish where nasal vowels *ę* and *ǫ* are realized as *e/o* + nasal consonant before specific consonants, e.g., as [em] / [om] before labials (Birnbaum & Molas 2012: 147).
[272] I put each second vowel in brackets because the sources are inconclusive as to the quantity of the vowels.
[273] Fought has no data on *je-* but examples are found in Hull, e.g., *e'nk'pa* 'I split open' (see below).

injatz'i ~ a'ntz'i 'I hit (it)' (A1SG *in-* + *jatz'i ~ atz'i* 'TV1 to hit, strike, whip')
injajpi ~ a'njpi 'I get (it)' (A1SG *in-* + *jajpi ~ ajpi* 'TV1 to grab, capture, get')

Compare this to set A on a verb starting with a vowel:

inwajk'u <iŋguahk'u> (Fought 1967: 110) 'I give (it)' (A1SG *in-* + *ajk'u* 'TV20 to give')

The interaction can be a little more complex depending on the verb structure, e.g., when *y* is involved:

a'yntz'i'x (C1SG *in-* + *jaytz'a ~ aytz'a* 'MPAS2 to be hungry' + =*ix* 'already')

in + (j)e- > e'n

e'nra'ch 'I am right' (C1SG *in-* + *e'ra'ch* 'IV to be true')
e'njk'a 'I am spread out' (C1SG *in-* + *ejk'a* 'to be spread out'[274])
e'nron 'I can no longer see' (C1SG *in-* + *e'ron* 'AP2 to see; it seems'[275])
e'nk'pa 'I split open' (A1SG *in-* + *jek'pa ~ ek'pa* 'MPAS1 to open out, stretch out open')

The use of set A produces a glide before verbs starting with a vowel:

inwerpes <iŋguerpes> (Fought 1967: 110) 'I worsen it' (A1SG *in-* + *erpes* 'CAUS6 to make worse, infect')

in + (j)i- > i'n

i'nri 'I rest' (C1SG *in-* + *jiri ~ iri* 'IV1 'to rest')
i'njt'a 'I am bound' (A1SG *in-* + *jit'i ~ injit'i* 'TV1 I tied up') (Fought 1967: 108)

Set A with glide before vowel-initial verbs:

inwira <iŋguira> (Fought 1967: 110) 'I see (it)' (A1SG *in-* + *ira* 'TV to see')

The following example demonstrates that the verb was reanalyzed as vowel-initial because it is used with a set A form with glide.

inwirijse 'I bring (it) to rest' (A1SG *in-* + *jirijse* 'CAUS1 to rest')

[274] Unattested in Hull but would likely be a mediopassive of class 3 with *-k'a* suffix.
[275] This stem must be a derivation from *ira* 'TV to see'. The initial vowel *e'* is likely a result of a reanalysis of the root as beginning with *e'* due to the frequent use of the word with C3 *a-* (**a-iron 'he sees' > *e'ron* 'he sees' > *e'ron* 'to see, it seems').

Curiously, we do not observe the expected contraction in *ixin* ‘to go’:

inxin ‘I go’ (C1SG *in-* + *ixin* ‘IV to go’)

The first person singular *inxin* never has a laryngealized vowel in Fought (1967; 1972) and Hull (2016). Perhaps this can be explained by looking at the verb’s prehistory: Schumann Gálvez (2007: 175–176) recorded special dialectal forms of this verb with older speakers in the eastern part of the *municipio* of Camotán:

b’ix-i-en	‘I went’
b’ix-i-et	‘you went’
b’ix-i	‘s/he went’
b’ix-i-on	‘we went’

Comparative Cholan evidence[276] also suggests that this verb used to have a *b’. This would mean that initially, the 1SG of the verb would have been *inb’ixin*. However, the circumstances of how *b’* was lost as well as the exact development that enables a 1SG form *inxin* still need to be worked out. This will likely be possible once we have gained a better understanding of the prehistory of Ch’orti’.

Alternatively, it could be that the reflex of *in* + *i*-initial verbs is never laryngealized since we otherwise only have examples that start with *ji-*. The only intransitive verbs that begin with *i-* in Hull (2016) aside from *ixin* are *iksijb’a* ‘to become dark’, *ik’ari* ‘to fart’, *irna* ‘to be seen’ and *ixna* ‘to last, take time’. Unfortunately, none of these are attested in the first person in incompletive aspect.

in + o- > o’n

[276] Compare CHN *b’ixe* (Keller & Luciano G. 1997: 45) and CHT <bixi>. Chol has a new verb, *majl*, for ‘to go’ but still has *b’ix(el)* ‘suddenly’ (Hopkins, Josserand & Cruz Guzmán 2011: 23, 139–140). Kaufman and Norman (1984: 117) reconstruct PCH *b’ix and even mention that the *b’ is lost in Ch’orti’. They claim that the word has no “direct outside cognates” but suspect a connection to PCH *b’ih ‘road’, which does appear in other Mayan languages (e.g., MOP *b’ej*; TSE *b’e*; KCH, MAM *b’e*(*e*)) and is reconstructed as PM *beeh (Kaufman 2003: 66–68). The sound change of PM *ee > PCH *i is one of the key innovations that define the Cholan branch (see discussion in chapters 2.2.4 and 2.2.5). Finally, and most importantly, Hieroglyphic Maya attests the verb *bixan- / bih-xan* (Kettunen & Helmke 2020: 89), which must be the form behind the Cholan root and would only require vowel assimilation to arrive at Ch’orti’ *(b’)ixin*. The authors suggest an analysis of ‘to “road-go”/“road-walk”’ which fits the idea in Kaufman and Norman. The second part still survives as an independent verb in Ch’orti’ today (*xana* ‘to walk’). However, the orthography <[bi]XAN?-na> *bi[h]-xan*; <[bi]XAN?-ni-ya> *bi[h]-xan-iiy* seems to leave some questions open.

o'njri 'I fall' (C1SG *in-* + *ojri* 'IV25 to fall')

o'njron 'I speak' (C1SG *in-* + *ojron* 'AP2 to speak, talk')

o'nchoy 'I enter' (C1SG *in-* + *ochoy* 'MPAS4 to enter')

o'mb'i <o'mb'i> (Fought 1967: 111) 'I wrap it' (A1SG *in-* + *job'i* 'TV1 to cover and wrap')

No forms of set A with glide (**inwo-) are attested.

in + u- > u'n

u'nch'u 'I grind' (C1SG *in-* + *juch'u ~ uch'u* 'IV22 to grind')

u'njta 'I am cooled' (C1SG *in-* + *ujta? juta?* (unattested in Hull (2016)[277])

u'nch'ru 'I grind (continually)' (A1SG *in-* + *juch'ru* 'ITER to grind corn continually')

u'nch'i 'I drink' (A/C1SG *in-* + *uch'i* 'TV4/IV25 'to drink')

u'njtz'i 'I am praising' (A1SG *in-* + *ujtz'i* 'TV4 to worship by bowing down')[278]

u'nsre 'I want (it)' (A1SG *in-* + *usre* 'TV12 to desire, want, like to do sth.')

u'njtz'i 'I smell (it)' (A1SG *in-* + *ujtz'i* 'TV4 to smell')

u'nsta 'I arrange (it)' (A1SG *in-* + *usta* 'TV9 to arrange')

No forms of set A with glide (**inwu-) are attested.

Interestingly, *u'nch'i* is already attested in Galindo (1834) as <Unchi> 'to drink' (the verb can be inflected both as transitive and intransitive) which means that this sound change had already happened by that time.

Contractions with C2PL *ix-*

Marcos García (2015: 72–76) also describes an allomorph C2PL *ixy-* in front of vowels:

[277] Perhaps related to the participle *jutb'ir ~ ujtb'ir* 'blown'?

[278] For this verb, Dugan described that the expected "metathesis" of *-n-* has the unexpected result of also deleting the root-internal *-j-*: *u'ntz'i* (in-ujtz'i) 'I smelled him' (Dugan 2013: 93). He attributes this to the "complex coda (CC rather than just C)" (Dugan 2013: 92): the infixation of *-n-* would create a cluster *-njtz'-*, which is "not permitted" by Ch'orti' phonotactics. Thus, *j* is deleted to get an "acceptable" cluster *-ntz'-*, "possibly because that is the 'weakest' phoneme in the cluster (compared to *n* and *tz'*)" (Dugan 2013: 92). In a footnote on the same page, Dugan admits that he is missing phonological data to "define a strength hierarchy of consonants or even to be sure there is one" (Dugan 2013: 92, fn. 6). Likewise, he can imagine that *-n-* "has to be retained because it is a complete morpheme". Deleting *tz'* on the other hand is not possible because the cluster *-nj-* is also not allowed in Ch'orti'. Note that the form *u'njtz'i* is in fact attested in Hull (2016: 422).

ixya'chpa 'you (pl.) get up' (C2PL *ix-* + *achpa* 'MPAS1 to get up')
ixya'jni 'you (pl.) run' (C2PL *ix-* + *ajni* 'IV6 to run')
ixyo'jri 'you (pl.) fall' (C2PL *ix-* + *ojri* 'IV25 to fall, fall down')

The inserted glide *y* is not the only change: the vowel is additionally laryngealized. The resulting *ixya'* or *ixyo'* resembles the result of contractions of *i* + *a* and *i* + *o*. One could argue that this change is due to *x* being a palatal phoneme, which apparently has the same result as if there were in fact an *i* on *ixi*. In that case we would expect *x* to have the same effect everywhere else in the language, which is not the case.

We could also think of a morphological solution: in theory, the forms could consist of a C2SG form *i-* + verb, which would be *ya'chpa, ya'jni* and *yo'jri* respectively. The prefix *ix-* would then be added additionally to form a second person plural from the singular. However, in that case we would expect the same with verbs that begin with consonants, i.e., **ix-i-lok'oy 'you (pl.) leave', which is unattested. As so far, Marcos García is the only source on this. For the time being, the origin of these interesting forms must remain a topic for further research.

6.4 Discussion

In this chapter, we have seen that those forms of set C that end in vowels interact with vowel-initial verbs in a way that suggests vowel contractions.[279] This is radically different from set A, which has allomorphs with glides as in most Mayan languages. Though set A partly is attested without glides (in all forms except for A3 and when used on verbs beginning with *o-* or *u-*) and instead with results mirroring those of set C, this is purely due to phonetic reasons because of a sound change *wo, wu > *o, u* that made the original glide disappear thereby producing a vowel + vowel environment that then underwent vowel contraction.

With verbs that begin with *j-*, the situation is a little different. *J* [h] disappears intervocalically when vowel-final forms of set C and set A are attached. Therefore, inflected forms without *j* and with laryngealized vowel occur for both set C and set A. Speakers are beginning to reanalyze some intransitive and transitive verbs as vowel-initial. However, many verbs have commonly unprefixed nominal derivations which retain the initial *j-* of the roots and perhaps act as a kind of counterbalance to this change. If contractions happen, the results are quite

[279] Although Mayan languages are generally described as being agglutinative, one may almost argue that the phonological complexity that results from the interaction of morphemes in Ch'orti' rather befits an inflectional system – an observation also made by Dugan (2013: 32).

similar, though not identical to those of *V + V* (see tables below). I suspect that remaining differences are likely due to the limited data.

Finally, C1SG *in-* also contracts with the verbs that it is prefixed to, which is best explained with it being a nasal vowel phonetically. Here, too, there is a parallel development for set A under exactly the same conditions described for vowel- and *j*-initial verbs. The interaction of C2PL *ix-* with vowel-initial verbs remains somewhat mysterious.

Table 21 gives an overview of the results of all vowels as well as *in-* interacting with each other in contractions without the involvement of *j*. Table 22 does the same for contractions with *j*.

Table 21. Morphophonological alternations without *j*. Overview. IPA based on description in Fought (1967).

First vowel → ↓ Second vowel	a	e	i	o	u	in
a	*a'*	*ya'*	*ya'*	*wa(')* ?	*wa(')* ?	*a'n* [ãą̃]
e	*e'*	*ye' ~ e'*	*ye'*	*we'* ?	*we' ~ e'*	*e'n* [ẽę̃]
i	*e'y ~ e' ~ a' ~ a*	*i' ~ e'y ~ e' ~ e*	*i'*	*wi' ~ i' ~ i ~ o' ~ o*	*wi' ~ i'*	**i'n [ĩį̃] / **in ?
o	*o'*	*yo' ~ o'*	*yo' ~ o'*	*o'*	*o'*	*o'n* [õǫ̃]
u	*o'y ~ o' ~ u'*	**yu' ?	*yu'*	*u'*	*u'*	*u'n* [ũų̃]

Table 22. Morphophonological alternations with *j*. Overview. Differences to Table 21, consisting of either additions or missing variation, are marked in yellow.

First vowel → ↓ Second vowel	a	e	i	o	u	in
a	*a'*	*ya'*	*ya'*	*wa'*	*wa' ~ wa*	*a'n* [ãą̃]
e	*e' (~ a')* ?	*e'*	*ye'*	*we'*	*we' ~ e'*	*e'n* [ẽę̃]
i	*e'y ~ e' ~ i'*	*e'y*	*i'*	*wi'*	*wi' ~ i*	*i'n* [ĩį̃]
o	*o'*	*yo'*	*yo'*	**o'	*o' ~ wo'*	*o'n* [õǫ̃]
u	*o'y*	**yu' ?	*yu'*	**u'	*u'*	*u'n* [ũų̃]

The differences between the two kinds of contractions are minor and mostly consist in the contractions with *j* showing less variation. This could either be due to insufficient data or due to genuinely different outcomes. Since in the cases of *-VjV-* loss *j* likely disappears first and then leaves two colliding vowels to contract, we would not expect there to be any difference in the outcomes.

In general, it seems as if a sequence of V1 + V2 usually leads to laryngealized V2'. This may presuppose a phase of assimilation (V1 + V2 > V2 + V2) as proposed by Fought (1967: 117). The vowels *i* and *u* are often realized as semi-vowels *y*, *w* when directly preceding or following *a*, *e* or *o* (Fought 1967: 85). *E* and *o* sometimes have the same effect. Not all allomorphs are necessarily attested as the results of contractions with set C but since we have demonstrated that set C just follows contractions that happen everywhere else in the language, it is to be expected that all of them can also occur as set C allomorphs in the respective environments.

If we contrast the patterns established by me for vowel contractions with those given in Quizar (2023: 261) for the various indexes of Ch'orti', the results are almost identical. The differences concerning C1SG *in-* are purely orthographical. Otherwise, my corpus study yielded a greater variety of allomorphs especially for the vowels *i* and *u*. Also, Quizar does not mention that set A partly participates in the same contractions instead of appearing with glides.

				Result (corpus)	Quizar (2023)
C1SG *in-*	+	a	>	*a'n-*	*ã'-*
		e	>	*e'n-*	*ẽ'-*
		i	>	*i'n-* ~ *in ?*	*ĩ'-*
		o	>	*o'n-*	*õ'-*
		u	>	*u'n-*	*ũ'-*
C2SG *i-*	+	a	>	*ya'-*	*ya'-*
		e	>	*ye'-*	*ye'-*
		i	>	*i'-*	*i'-*
		o	>	*yo'-* ~ *o'-*	*yo'-*
		u	>	*yu'-*	*yu'-*

C3 *a-*[280]	+	a	>	*a'-*	*a'-*
		e	>	*e'- (~ a'?)*	*e'-*
		i	>	*e'y- ~ e'- ~ i'- ~ a'- ~ a-*	*a'- ~ e'-*
		o	>	*o'-*	*o'-*
		u	>	*o'- ~ o'y- ~ u'-*	*oy-* (without '!)

Sometimes a contraction does not happen when it is expected in theory. Fought (1972: 24) already noted that "variation", as he calls it, does not apply to all morphemes. For example, the causative morpheme *-es* does not seem to produce laryngealized forms: *k'apa* 'MPAS1 to finish (doing sth.)' but *k'apes* 'CAUS6 to finish (doing sth.)'. The only instances of the sequence *-e's-*[281] in the language are verbs that already end in *-e'* without any contraction, e.g., *leche', we'* 'to eat'. On the other hand, the allomorph *-se* does participate in contractions, e.g., *chamse* 'CAUS1 to kill, murder' ~ *uchamsyo'n* 'he kills us' (A3 *u-* + *chamse* + B1PL *-on*). The explanation for this is simple: causative forms in *-es* are likely older formations not derived from the mediopassive form in *-a* but from the root *k'ap.*

An interesting question is whether different dialects of Ch'orti' all use the same contraction rules. According to Wichmann (1999: 13), the Ch'orti' pedagogical coursebook by Lubeck and Cowie (1989: 92, 146) "has a good treatment of morphophonemics involving person markers, but this could not be used because it turns out that the rules in the Guareruché[282] dialect is different from those of the Jocotán dialect represented by most of the materials used for this study". However, it seems that the stem interactions recorded by Lubeck and Cowie are not substantially different apart from the fact that the glottal stop appears less in their notations. Additionally, Fought's fieldwork was conducted, among other places, in Guareruché so we would not expect the data to be different from his. According to Hull (2016: 3), "[i]nconsistent orthography and notation plague" Lubeck and Cowie's coursebook. Therefore, what Wichmann regards as dialectal difference may be just that – inconsistency.

6.5 Set C < set A?

Chapter 5 has shown very clearly that deriving set C from set A does not yield satisfying results. Every explanation that has been presented leaves many questions open and either

[280] First person plural C1PL *ka-* ends in *-a* and thus behaves in the same way as C3 *a-*.
[281] In theory, the result could also be *a's, i's, o's* or *u's* or variants with glides. However, this also does not seem to happen with causative *-es*.
[282] Guareruché is an *aldea* of Jocotán.

resorts to ad hoc assumptions to explain the individual person morphemes or does not offer explanations at all. If we add to this the conclusion of this chapter – namely that set A and C behave completely differently when interacting with vowel- or *j*-initial verbs, there is even less reason to derive set C from set A.

If we seriously consider the similarities between the paradigms of Ch'orti' indexes given again in Table 23 where I have placed set C in between the other two sets to facilitate comparability, set C has exactly one form that suggests an origin in set A (C1PL) and exactly one form that suggests the same for set B (C2PL). Two forms must be innovations and require separate explanations (C2SG, C3). The remaining form (C1SG) is inconclusive and could either be derived from set A or set B because of the (likely Pan-Mayan) innovation of a verbal A1SG allomorph based on set B *in-* (see discussion in chapter 3.1). In other words: since we know neither 1) when set C arose nor 2) when the change of A1SG happened nor 3) when CHR B1SG changed from *in to *en*, it is impossible to decide whether C1SG reflects the original set B form or the form of set A that was innovated based on set B.

Table 23. Index-sets of Ch'orti' repeated from Table 11 with Proto-Mayan reconstructions beneath (Kaufman 2015: 161).

		Set A	Set C	Set B[283]
Singular	1st	*in- / inw-* (verbs) *ni- ~ (ni)w-* (nouns) *nu-/ w-	*in-*	*-en* *-iin
	2nd	*a- / aw-* *aa-/aaw-	*i-*	*-et* *-at
	3rd	*u- / uy- (uw-)* *u-/r-	*a-*	ø *ø
Plural	1st	*ka- / kaw-* *qa-/q-	*ka-*	*-on* *-o'ŋ
	2nd	*i- / iw-* *ee-/eer-	*ix-*	*-ox* *-ex
	3rd	*u- / uy- … (-ob)* *ki-/k-	*a- … (-ob)*	(*-ob'*) *-eb'

For the question at hand this means that, from a purely formal viewpoint, we have equal reason to assume that set C is based on set A or on set B. In the first case, set C would have evolved after the A1SG innovation. However, since the latter occurs in virtually all Mayan

[283] I exclude the variation of set B and set C because I have demonstrated that it follows morphophonological rules that likely go back to recent sound changes.

languages, it may in fact be rather old.[284] If we explain it from set B, making a statement as to the age of the formation is more difficult. Either way, it is not an argument in favor of deriving set C either from set A or set B, which leaves us with one form that set C shares with the other two sets each.

Every single one of the proposals discussed so far suffers from the fact that aspect-based split ergativity in Ch'orti' is automatically equated with the similar yet not exactly equivalent split ergativity that can be observed in the other Cholan, Yucatecan and Poqom languages. I would like to abandon this idea together with all preconceptions that are tied to it and propose an entirely different explanation.

[284] It seems to be already attested in Hieroglyphic Maya (Kettunen & Helmke 2020: 122).

7 A new origin of set C

The advantage of deriving set C from set B is that problems like the missing nominalizer or why set A only develops this special subset with intransitive verbs but not in its use with transitive ones are avoided. On the other hand, set B brings with itself other issues, the most obvious one being that set B is suffixed in Ch'orti' while set C is prefixed.

The explanation that is proposed in this chapter for set C will address all its puzzling qualities:

1. its position as a prefix
2. the form of every single index of set C
3. the development of aspect marking without aspect markers
4. the fact that this new aspect distinction develops for intransitive verbs only

A preliminary remark is in order: Contrary to older research (e.g., Kaufman & Norman 1984; Robertson 1992: 63; Law, Robertson & Houston 2006: 443–444, among others)[285], I assume that Proto-Cholan (and perhaps even Proto-Mayan, though this needs a thorough study of the other subbranches) did not distinguish between completive and incompletive aspect morphologically, neither for transitive nor for intransitive verbs. This means that I do not search for ways to explain how Ch'orti' lost its aspect distinction on transitive verbs. Instead, I provide a scenario of how an aspect distinction could have arisen on intransitive verbs alone.

This point of view is preferable due to the following reasons: first, even among the closest Western Cholan languages Chol and Chontal, the morphology of aspect marking differs greatly (for an overview see Vinogradov 2016). We have seen considerable levels of diversity in Yucatecan in chapter 3.3.1. Generally, this can be observed across all Mayan languages. It is therefore highly questionable whether a morphological distinction between incompletive and completive aspect should be reconstructed for Proto-Cholan (and Proto-Mayan) at all.[286]

[285] Law, Robertson and Houston propose, however, that the Eastern Cholan branch "innovated by replacing that aspect system with a tense system". This claim needs to be reinvestigated in a future paper because it greatly depends on the analysis of Choltí. Suffice it to say that the evidence for a tense system is not strong.

[286] Kaufman justifies the reconstruction of an incompletive-completive distinction for Proto-Mayan with the fact that all languages show this distinction in some form or another: "Although *habitual/incompletive is not represented in a unique/single/uniform way, it seems that the category is present in all the Mayan languages. It contrasts directly with *punctual/completive. Since *completive is often unmarked with an aspect particle, *incompletive was most probably contrastively marked with one such." (Kaufman 2015: 194). The fact that a distinction exists synchronically means little for the reconstruction of a proto-language

The second reason is that we cannot securely confirm that the oldest Cholan language, Hieroglyphic Maya, showed a grammatical distinction between completive and incompletive aspect. There are but few forms that are not written in the completive aspect and wherever there are any, the reading is mostly unclear or insecure (Gronemeyer 2014: 154, fn. 320). In the past, it was assumed, e.g., in Houston (1997) or Robertson, Houston and Stuart (2004) "that texts are generally written in an incompletive aspect […] and earlier or terminated actions are marked by a completive suffix" (Gronemeyer 2014: 24). Others like Wald (2000; 2004a; 2007) or Wald and MacLeod (1999) thought that the texts are written in the completive aspect while "[a]nteriority or futurity relative to the completive aspect are marked with temporal deictic enclitics" (Gronemeyer 2014: 24). The fact that it is apparently possible to justify both views, which are the exact opposite of each other, is telling. Newer research agrees that aspect, at least when it comes to finite verbs, is unmarked grammatically in hieroglyphic texts (Law 2016; Law & Stuart 2017: 168). Therefore, I base the following argumentation on the assumption that Pre-Ch'orti' started without a grammatical distinction of aspect or tense.

7.1 Position: Suffix > prefix

A derivation of set C from set B will not be successful if no explanation of the switch of position from suffix to prefix is provided. Therefore, I will address this issue first.

7.1.1 Reconstruction

Set B appears both as a prefix and a suffix in individual Mayan languages. Example (1) shows that it is prefixed to the verb in K'iche', while it is suffixed to the verb in Yucatec Maya in (2):

(1) K'iche' (Can Pixabaj 2017: 466)
x-oj-b'iin-ik
COM-B1PL-walk-IS
'We walked.'

if there is no consensus on the actual form. Whether or not it is possible to agree on a form for the PM completive and incompletive marking will be discussed in a separate publication.

(2) Yucatec Maya (Hofling 2017: 711)

(j) *jóok'-ech*

(COM) go.out-B2SG

'you went out'

For Proto-Mayan, set B is reconstructed as a suffix (Campbell 1979: 976; Robertson 1980) or as "enclitic to the first word of the predicate" (Kaufman 2015: 161). When it comes to the question of suffix vs. enclitic, Vinogradov summed the situation up perfectly when he commented that the exact morphological properties of inflectional morphemes in most Mayan languages "usually seem to be established based on some intuitive knowledge of an author, which sometimes proves to be inaccurate after the critical examination of a problem" (Vinogradov 2017: 106). Or, in Lehmann's words, "information on the clitic or affixal status of grammatical formatives in the Mayan languages is notoriously unreliable" (Lehmann 2020: 12, fn. 10). This is especially true for set B markers which show a high degree of diversity among the Mayan languages. The recent assessment of the situation by Grinevald and Peake referenced in Vinogradov (2017: 106) bears witness to this confusion when they conclude that set B markers "may be prefixed and/or suffixed" but also later state that they "may in fact be clitics or free morphemes that occur (semi)independently of the verb form" (Grinevald & Peake 2012: 23). A recent contribution by Holtmann (2023) argues that they are in fact best reconstructed as second-position enclitics in Proto-Mayan.

For now, I will not focus on whether set B is enclitic or a suffix but rather on the position: it is reconstructed as always following, not preceding the word it attaches to. This is based on the observation that in all cases where it is prefixed, an aspect marker is present.[287] Therefore, it is generally assumed that set B went through a stage where it was not actually prefixed to the verb but suffixed to the aspect marker which came before the verb (Campbell 1979: 976; Robertson 1980: 16).[288] Q'eqchi' (Greater K'iche'an) demonstrates this especially well because the position of set B before the verb as opposed to after the verb relies on the

[287] In footnote 113, I mentioned Lehmann's objection that set A, as well, is not a prefix or a proclitic but can in fact be used as an enclitic of aspect markers that precede the verb with adverbs coming between it and the verb stem. It is an interesting question whether Yucatec retains a state where set A was not fully affixed to the verb yet (as it is in many other languages) or – and I find this more likely – that set A used to be proclitic to the stem in principle but encliticized to the preceding aspect particle in some cases instead. This kind of behavior is known as that of "ditropic clitics" (Hill et al. 2019). It will be discussed in chapter 7.3.3.3.

[288] This is sometimes taken to be evidence of a clitic stage of set B because it is argued that it attached to the first element in the clause. Set B would then have originally been a Wackernagel or "second position clitic" (Kaufman 2015: 298). Contrary to that, Robertson (1980: 52) reconstructs set B as affixes.

presence of tense-aspect prefixes (Vinogradov 2017: 109). Example (3) demonstrates the use of set B with non-verbal predicates in the stative construction where set B is found behind the verb. On the other hand, set B is placed between the aspect marker *x-* and the verb root *war* 'to sleep' in (4). Furthermore, it cannot occur at the beginning of a word form, which becomes obvious in cases where verbs are used without tense-aspect-prefix, e.g., in optative forms as in (5). It would be ungrammatical for set B to occur before the verb in (3) and (5).

(3) Q'eqchi' stative construction (Stewart 2015: 76)
winq=at[289]
man=B2SG
'You are a man.'

(4) Q'eqchi' completive aspect (Stewart 2015: 66)
x-at-war
COM-B2SG-sleep
'You slept.'

(5) Q'eqchi' optative (Stewart 2015: 60)
q-il-aq=at
A1PL-see-IRR=B2SG
'We would see you.'

The strong correlation between the position of set B before the verb and the presence of tense-aspect-markers can readily be explained by assuming that it is indeed encliticized to the marker.[290] Further examples of such a distribution of set B based on the absence or presence of aspect markers is seen in Poqom (Greater K'iche'an) (Vinogradov 2017: 118), as well as Chuj and Tsotsil (Campbell 1979: 976).

Still, this explanation based on the presence or absence of tense-/aspect-markers leaves some open questions. First, in Yucatecan or Western Cholan languages set B does not get suffixed to aspect particles although they are present as demonstrated in (6).[291]

[289] Note that according to Vinogradov, set B is clearly an affix when placed in between aspect marker and verb whereas he analyzes it as a clitic in its position following the verb.
[290] Based on the Q'eqchi' data, Kaufman's assessment of set B as a second-position enclitic in Proto-Mayan seems more appropriate than the analysis of it as a suffix: if set B can freely move to encliticize to the first element of a clause, this is strong evidence that it is a clitic rather than an affix.
[291] One could investigate whether set B did not move to that position because it was already "taken" by set A enclitics (see footnote 287).

(6) Yucatec completive transitive (Hofling 2017: 710)

t-aw-il-aj-en

COM-A2SG-see-TS-B1SG

'you saw me'

Here, one could assume that the explanation only works one way: If set B appears before the lexical verb, this is always due to it being suffixes/encliticized to a preceding aspect marker. On the other hand, the presence of a marker does not necessarily require set B to encliticize to it – in Yucatecan, e.g., it can apparently stay where it is. This would, however, pose a problem for the analysis as a second-position enclitic.

The data from K'iche' also do not entirely match this theory. K'iche' forms the stative construction in the way portrayed in (7). Although they both belong to the Greater K'iche'an branch, K'iche' differs from Q'eqchi' in this regard (cf. example (5) above).

(7) K'iche' stative construction (Larsen 1988: 107)

oj *kunaneel*

B1PL doctor

'We are doctors.'

There is no element that set B might be encliticized to and yet, it is found in front of the verb. This problem will be discussed in more detail in chapter 7.3.3.2.

This neat distribution of set B markers also would not work as an explanation for Ch'orti'. As we have seen in chapter 4.3, there is no obligatory aspect (or tense) marker that could explain why set B appears in front of the verb. Although it is theoretically possible that the structure arose in connection with aspect particles such as PROG *war* and later it became possible to use set C as marker of incompletive aspect without them, this is an additional assumption. It would be preferable to find a different explanation that does not rely on a scenario of this kind, which is impossible to prove considering the available data.

When looking for a fitting source of this "misplacement" of set B, further valuable evidence comes from C2PL *ix-*. The generalization of the vowel *o* in the plural of set B in B1PL *-on*, B2PL *-ox*, B3PL *-ob'* as opposed to Proto-Mayan reconstructed by Kaufman (2015: 161) as PM B1PL *o'ŋ, B2PL *ex/ix, B3PL *eb' is considered to be a relatively old development since it is reconstructed for Proto-Cholan (Kaufman & Norman 1984: 91). However, this analogical levelling is not an exotic development. It may even have happened repeatedly and separately in the individual languages. For Hieroglyphic Mayan, there are unfortunately no clear

attestations of B2PL so that we do not know whether it was still **-ix or already **-ox. If set C *ix*- is indeed derived from set B, its vowel would point to the fact that the form (but not necessarily set C itself) must be older than Proto-Cholan, since it conserves the old vowel *i* of PM B2PL *ex/ix.[292]

To sum up all the evidence: for our new explanation of set C, we are looking for a special use of set B in a construction that must be relatively old. The special use of set B is necessary to explain why it underwent a different phonological development as a suffix (synchronically set B) and a prefix (synchronically set C). I propose that this special use of set B is that as independent pronouns.

The fact that it is possible for set B to develop into two different paradigms is already evidenced by Hieroglyphic Maya where the forms of set B as part of the independent pronouns are not identical to the also attested set B suffixes (see Table 24):

Table 24. Set B as reconstructed for Proto-Mayan by Kaufman (2015: 161) contrasted with its use in the independent pronouns of HGM with base *h*- (Lacadena 2013: 18; reconstructed forms mine) and as a suffix (Kettunen & Helmke 2020: 28).

	Proto-Maya set B	HGM independent pronouns	HGM set B suffix
1SG	*-iin	*hiin* <hi-na>[293] *hin* <hi-ni>	*-en / -een* <Ce-na>
2SG	*-at	*hat* <ha-ta>	*-at/ -et* <-ta> / <-te>?
3SG	*ø	*haa'* <ha-i> *ha'* <ha-a>[294]	ø

[292] The pronoun is reconstructed as **-ex* for Proto-Mayan, but Kaufman (2015: 161) points out that it is equally possible that the second person plural set B marker was in fact originally **ix* and developed into **ex* "by analogy with **ee(r)* and **eb*". Some languages like K'iche' show *ix*. This is just one of many areas where the phonological reconstruction of Proto-Mayan still needs some refinement (see chapter 2.2.4).

[293] Robertson instead suggests to interpret this reading as *hi'n* (Hull, Carrasco & Wald 2009: 40). The Mayan script apparently used final syllables with disharmonic vowels to indicate vowel length. Whether one interprets the form as *hiin* or *hi'n* depends on what rules for disharmonic spelling one follows (Houston, Stuart & Robertson 2004; Lacadena & Wichmann 2004). However, the long vowel has an advantage over the laryngealized one because one could then consider the spelling with the longer vowel an earlier variant, whereas <hi-ni> *hin* would then be a younger form already after the merger of long and short vowels that took place in all Cholan languages (except for *aa and *a in Western Cholan). Lacadena (2013: 18) also seems to imply that he assumes a diachronic relationship of the two forms by writing "hiin > hin".

[294] Kettunen and Helmke (2020: 123) additionally list a reading of <ha-i> as *ha'i'* but this seems to contradict the orthographic rules of the Maya script if one takes them seriously. Additionally, the pair *haa' ~ ha'* is parallel to 1SG *hiin ~ hin*. This is further evidence for the general merger of long and short vowels as already suggested in footnote 293 above.

1PL	*-oŋ	*hon[295]	*-on / -o'n* <Co-na>
2PL	*-ix (or -ex)	*hix	-? / *-ox ?
3PL	*-eb'	*ha'o'b'* <ha-o-b'a> *ha'ob'* <ha-o-b'o>[296]	*-ob'/o'b'* <-Co-b'a>

Also, a parallel development of set B is described by Vinogradov. In Modern Q'eqchi, the prefixed and encliticized versions of set B are very similar, but not entirely identical, as Table 25 shows, which is why Vinogradov (2017: 108) suggests considering them as two different paradigms synchronically and not as one paradigm that just exhibits both sets of properties.

Table 25. Set B in Modern Q'eqchi' (Vinogradov 2017: 107).

	1SG	2SG	3SG	1PL	2PL	3PL
enclitic	*=in*	*=at*	*ø*	*=o*	*=ex*	*=eb'*
prefix	*in-*	*at-*	*ø*	*oo-*	*ex-*	*e'-/eb'-*

This proves that it is possible for set B to undergo different sound changes when the two uses are distinct in terms of position and/or function. Now we need to establish what kind of special function we might be dealing with in the prehistory of set C.

A syntactic position documented for Mayan languages that can be filled by independent pronouns, which contain set B, and can account both for their position before the verb and for the special use that enables them to develop differently from suffixed set B is that of focus. Chapters 7.1.2, 7.1.3 and 7.1.4 describe key features of focus marking in Mayan that will be relevant for the argument presented later. Chapter 7.1.5 deals with focus marking in Cholan and 7.1.6 presents a possibility of deriving set C from an earlier focus construction.

[295] This form is likely cognate to the first element of Chol *joñ-oñ* '1SG independent pronoun'. It may even be a direct descendant if it turns out that Hieroglyphic Maya is ancestral to all Cholan languages, not just Eastern Cholan ones. Note that Chol has generalized the 1PL form for the first person and the 2SG form for the second. Plural is marked through additional affixes.

[296] Kettunen and Helmke (2020: 123) only list the spelling <ha-o-ba>, for which they suggest a reading *ha'ob.* This again goes slightly against orthographic rules. However, the interpretation by Lacadena (2013: 18) as *ha'o'b'* also might profit from a minor adjustment: if we interpret <ha-o-b'a> as *ha'oob'*, not *ha'o'b'*, we again get an earlier form with a long vowel and a later form *ha'ob'*, which would be further evidence for the long and short vowel merger.

7.1.2 Focus

As this section deals with information structure, which has not been mentioned before, a few terminological remarks concerning the central notions of "focus" and "topic" are necessary before we start. Usually, "focus" is understood as that part of the utterance which contains the new or most important information as opposed to the "background" while "topic" is the entity about whom or which information ("comment") is provided (Aissen 2017b: 293). It is important to stress that focus and topic are not a complementary pair – they each have their own counterpart (focus vs. background, topic vs. comment) and "belong to different dimensions of information structure" (Aissen 2017b: 309).

"Focus" can be further differentiated into "new information focus"[297], which is not well-studied in Mayan as opposed to "contrastive focus", which has been studied extensively (Aissen 2017b: 296–309). The former is a more basic notion of focus as it does not involve explicit emphasis and delimitation from other options while this is true for the latter, i.e., while the answer to a question already constitutes an instance of new information focus (Aissen 2017b: 296–297), contrastive focus commonly puts emphasis on the fact that it is one entity, not another that is affected or has done something (Aissen 2017b: 298). The latter is especially commonly realized via cleft constructions such as "It is xy who/which…" (Güldemann 2010: 72).

Now we can return to the specifics of topic and focus marking in Mayan. Though the notion of "basic word order" is problematic and some authors have pointed out that it is not easy to determine what the basic word order of Mayan languages is (Brody 1984; England 1991), there seems to be consensus that they are largely verb-initial (see chapter 2.2.2). Although there is some variation among the individual languages[298], there are generally two possible positions for non-verbal elements before the verb. The first position is reserved for topics while the second position is reserved for contrastive focus.[299] This was initially proposed by Norman (1977) who had observed that topics occur sentence-initially, while the focus position is found directly before the verb (Aissen 2017b: 296).

[297] Other labels in use are, e.g., "assertive" or "completive" focus (Güldemann 2010: 72).

[298] Notable publications that are concerned with Mayan word order and information structure in individual languages are, among many others, Quizar (1979) for Ch'orti', Datz (1980) for Popti', Brody (1982) for Tojolab'al, Dayley (1985) for Tz'utujil. Recently, a lot has been published on K'iche' (Can Pixabaj & England 2011; Velleman 2014) and Yucatec (Gutiérrez-Bravo 2011; Gutiérrez-Bravo & Monforte y Madera 2011; Verhoeven & Skopeteas 2015). A general overview is given in Aissen (1992; 2017b).

[299] It is apparently not (always) obligatory for the focused constituent to appear in this position (Aissen 2017b: 299). This will become relevant in chapter 8.

In example (8), we observe "basic" word order in Tz'utujil (K'iche'an), where the verb is indeed found at the beginning of the sentence. On the other hand, in (9) a topic occurs before the verb and in (10) a focused constituent. Example (11) shows that both can occur together, as well, and that it is in fact the focused constituent that appears immediately before the verb.

(8) Tz'utujil "basic" word order (Dayley 1985: 302)

X-pi	*jun*	*aachi*	*Xelaju'.*
COM-come.B3	one	man	Quetzaltenango

'A man came from Quetzaltenango.'

(9) Tz'utujil (internal) topic (Dayley 1985: 321)

Ja	*ch'ooy*$_T$[300]	*ma*	*x-uu-tij*	*ta*	*ja*	*kéeso.*
DEF	rat	NEG	COM-A3SG-eat.B3	IRR	DEF	cheese

'The rat didn't eat the cheese.'

(10) Tz'utujil focus (Dayley 1985: 355)

Machat$_F$	*x-a-choy-b'e-j*	*chee'.*
machete	COM-A2SG-cut.B3-APPL-TS	tree

'It was a machete that you cut wood with.'

(11) Tz'utujil (internal) topic and focus (Dayley 1985: 309)

Ja	*tzyaq*$_T$	*ch'ooyaa'*$_F$	*x-ee-tij-ow-i.*
DEF	clothes	rats	COM-B3PL-eat-AF-TS

'Rats were the ones who ate the clothes.'

Aissen (1992) further developed Norman's observations and established that these two positions are "linearly indistinguishable when only a single constituent precedes the verb, but may be distinguished in the presence of other elements, e.g., negation" (Aissen 2017b: 296). In fact, without context, it would be difficult to determine for (9) and (10) whether we are dealing with topic or focus. In (9), however, the position of the negation suggests that we are in fact dealing with a topic because it is sentence-initial, whereas focus would occur after the negation because it is preverbal. The constructions work the same way with independent pronouns as with nouns as seen in (12):

[300] I adopt Aissen's method of marking topic and focus with subscript $_T$ or $_F$ respectively whenever it helps to differentiate the two or understand the sentences.

(12) Tz'utujil focus with independent pronoun (Dayley 1985: 386)

Jar[301]	*iinin*$_F$	*jar*	*iin*	*k'o*	*waawe'*.
DEF	1SG	who	B1SG	be	here

'It's me who's here.'

It is worthwhile investigating whether for set C in Ch'orti' we might not historically be dealing with set B indexes in their function as independent pronouns in topic or focus position. However, we must establish first which of the two is the more likely candidate for this. This is where the other peculiarity of set C becomes relevant: assuming a use of set B in topic position would not explain why this only affected intransitive verbs. On the other hand, as we will see in the following chapter, set B in a focus position can readily explain this.

7.1.3 Agent focus construction (AFC)

It is a known feature of some Mayan languages that S and O can be focused and put in the preverbal focus position, while this is only possible for A under certain conditions. In order to focus A in those Mayan languages that have this restriction, a special construction called "agent focus construction" (AFC) is needed that involves an intransitivized form of the transitive verb, e.g., an antipassive (Aissen 2017b: 306). It is not the intention of this thesis to explore why this restriction exists though note that this kind of syntactic ergativity is known from other languages/language families as well, e.g., Dyirbal (Dixon 1972: 100).[302] The construction is not used in all Mayan languages, but is commonly found in Eastern Mayan, Q'anjob'alan and in individual languages and dialects of the other branches, e.g., Yucatec Maya and the Zinacantec dialect of Tsotsil (Aissen 2017b: 306).[303] An overview is provided in Stiebels (2006).

[301] *Jar* or *ja* according to Dayley (1985: 386).

[302] According to Aissen, the reasons for the existence of this special construction may lie in the observation first made by DuBois (1987) that A commonly expresses known information, i.e., topics while S and O express new information and are therefore part of the comment, "not topic but possibly focus" (Aissen 2017b: 295).

[303] Kaufman (2015: 338) reports that the agent focus construction exists in all Mayan languages except for Huastecan, Yucatecan and Greater Tseltalan. Huastecan and Yucatecan do use a special agentive antipassive verb form to focus A but it is not intransitive. In some languages, there are restrictions on when the AFC is used. For instance, in Q'anjob'alan, it is only used when a third person A acts on an O (any person) (Zavala Maldonado 2017: 230). It seems that the cognate morpheme in Huastec marks a middle voice (Kondic 2011: 138–139).

The AFC is illustrated below with examples from K'iche'. They are paired to show sentences with an unfocused and a focused S, O and A respectively. If we compare the unfocused S in (13) with the focused one in (14), we observe that in the latter, *ri tata'* is used with the focus marker *aree* and put in front of the verb, but the verb form itself does not change.[304] The same is true for the focused O *ri ak'* in (16) compared to the unfocused one in (15).[305] When A is focused in (18), however, the verb form does change. While the transitive verb is inflected with set A to mark A in (17) as would be expected in the language, the verb in (18) is unmarked (B3SG) and receives an antipassive suffix. This means that in (18), the verb is intransitive. It only has one syntactic argument, S, which is marked by set B or, as we are dealing with a third person, by lack thereof.

(13) K'iche' unfocused S (López Ixcoy 1997: 381)

Xwar	*ri*	*tata'.*
COM-sleep.B3	DEF	old.man

'The old man went to sleep. [Span. Se durmió el anciano.]'

(14) K'iche focused S (López Ixcoy 1997: 381)

Aree[306]	*[ri*	*tata']$_F$*	*x-war-ik.*
FOC	DEF	old.man	COM-sleep.B3-IS

'It was the old man who went to sleep [Span. Fue el anciano quien se durmió.]'

(15) K'iche' unfocused O (López Ixcoy 1997: 382)

X-u-tij	*ri*	*ak'*	*ri*	*utiiw.*
COM-A3SG-eat.B3	DEF	chicken	DEF	coyote

'The coyote ate the chicken. [Span. Aquel coyote se comió a aquel pollo.]'

[304] The suffix *-ik* that occurs in (14) is a status suffix that is used with intransitive verbs. These suffixes are often phrase-final in Mayan, which means that they are omitted if the verb is not at the end of the phrase (see, e.g., Can Pixabaj 2017: 476 for K'iche'). Therefore, it is missing in (13) because the verb is not at the end of the phrase and not because there is different morphology due to the focus construction.

[305] Some further restrictions apply in K'iche' both with focusing O and with focusing A (see López Ixcoy 1997: 381–382) but I will not address them here because we are only interested in the basics of the AFC.

[306] The focus particle looks almost like the form of the third person independent pronoun *ri are' (ra're')* and this is likely its origin.

(16) K'iche' focused O (López Ixcoy 1997: 382)

Aree	*[ri*	*ak']$_F$*	*x-u-tij*	*ri*	*utiiw.*
3SG	DEF	chicken	COM-A3SG-eat.B3	DEF	coyote

'It was the chicken that the coyote ate. [Span. Fue aquel pollo el que se comió aquel coyote.]'

(17) K'iche' unfocused A (López Ixcoy 1997: 382)

X-u-loq'	*ri*	*jaa*	*ri*	*ali*	*Ixkaaj.*
COM-A3SG-buy.B3	DEF	house	DEF	CL	Ixkaaj

'Ixkaaj bought the house. [Span. Ixkaaj compró aquella casa.]'

(18) K'iche' focused A (López Ixcoy 1997: 382)

Aree	*[ri*	*ali*	*Ixkaaj]$_F$*	*x-loq'-ow*	*r-eech*[307]	*ri*	*jaa.*
3SG	DEF	CL	Ixkaaj	COM-buy.B3-AP	A3SG-RN	DEF	house

'It was Ixkaaj who bought the house. [Fue Ixkaaj quien compró aquella casa.]'

The AFC often requires a derivational morpheme that is otherwise used to form antipassives in Mayan languages (Holtmann 2022: 1). The morpheme is not cognate across all languages, i.e., the languages do not all use the same kind of antipassive for this construction. Mayan languages have several ways of forming antipassives depending on whether the object is deleted entirely, incorporated into the verb or expressed obliquely (Aissen, England & Zavala Maldonado 2017a: 7). According to Holtmann (2022: 1), it is generally the suffix that forms the absolutive antipassive[308] that is used for the AFC. In the case of K'iche', it seems the suffixes *-ow/uw* (for root transitive verbs as in (18)) or *-n* (for derived transitive verbs) are used instead, which are identical to those of the incorporated antipassive (while the absolutive antipassive would be *-Vn*[309]) (Can Pixabaj 2017: 486–487).

If we again review the Tz'utujil examples in the previous chapter, we will see that the sentence where A is focused in (19) also has the agent focus construction and that in Tz'utujil a cognate suffix *-ow* is even used:

[307] Can Pixabaj (2017: 487) writes that in some dialects, the patient must be introduced by a relational noun as in the grammar by López Ixcoy (1997), while in other dialects this is optional when the patient is an NP (third person), as Mondloch(1981) reports.

[308] The absolutive antipassive either has no overtly expressed patient or an oblique patient.

[309] However, the absolutive antipassive suffix for derived transitive verbs is also *-n*, just like for the incorporated antipassive (Can Pixabaj 2017: 486). Therefore, it is strictly speaking not possible to differentiate whether the absolutive or incorporated antipassive suffix is used in the AFC for derived transitive verbs in K'iche'.

(19) (Internal) Topic and focus (Tz'utujil) (Dayley 1985: 309)

Ja	*tzyaq$_T$*	*ch'ooyaa'$_F$*	*x-ee-tij-ow-i.*
DEF	clothes	rats	COM-B3PL-eat-AF-TS

'Rats were the ones who ate the clothes.'

Aissen (2017b: 307) points out that "agent focus" is not an ideal terminology for the phenomenon at hand since, technically, "it is not peculiar to focus of A per se". The construction is not only used in contrastive focus but also with interrogatives, relative clauses and certain indefinite constructions. This is not very surprising as these constructions, together with negation, may be regarded as "inherently focused" (see discussion in 7.2). Example (20) demonstrates this for a question in K'iche'. Though the suffix is not the same as in (18) above (because root transitive verbs receive different suffixes than derived transitive verbs), in both cases an originally transitive verb is intransitivized when A is put into a position before the verb. In example (21), we observe the same with negation. Can Pixabaj (2017: 492) confirms that for relative clauses, too, an antipassive form must be used, though she does not demonstrate this with an example.

(20) K'iche' AFC with interrogative (Mondloch 1981: 227)

Jachiin	*x-paxi-n*	*lee*	*laq?*
who	COM-break.B3-AF	DEF	bowl

'Who broke the bowl?'

(21) K'iche' AFC with negation (Can Pixabaj 2017: 490)

na	*ak'al-aab'*	*ta*	*k-e-tzuku-n*	*r-eech*	*ri*	*ixiim*
NEG	child-PL	IRR	INC-B3PL-seek-AF	A3SG-RN	DEF	maize

'It is not the children who are looking for the maize.'

An interesting point is that since the AFC involves an intransitive verb, it can only reference one of two semantic arguments morphologically. However, the languages where it is still in use differ in which of the two semantic arguments is cross-referenced on the verb, A or O. This will be the subject of the following chapter.

7.1.4 Original pattern of indexation

Synchronically, three indexation patterns occur in Mayan languages in the AFC (Holtmann 2022: 2–3; Stiebels 2006: 254):

1. cross-referencing of (semantic) agent (e.g., Greater K'iche'an: Q'eqchi, Poqomchi')
2. cross-referencing of (semantic) patient (e.g., Mamean: Ixil; Greater Q'anjob'alan: Akatek, Popti', Q'anjob'al, Chuj; Tseltalan: Tsotsil)
3. cross-referencing of either (semantic) agent or patient based on their positions on a person hierarchy[310] (Greater K'iche'an: e.g., Kaqchikel, Tz'utujil, K'iche', Sakapulteko, Sipakapense)

All three patterns have been proposed as the original pattern for Proto-Mayan (Holtmann 2022: 5–6). Based on data from colonial Q'eqchi', Holtmann (2022) describes a fourth pattern where the verb shows no overt indexation at all, neither of the semantic agent nor of the patient, and argues that this is most likely the original one.

Examples (22) and (23) show the same sentence, 'I love you', once for colonial Q'eqchi' and once for 19th century Q'eqchi'. In both cases, the first-person singular independent pronoun is in front of the verb and the verb shows the AFC suffix *-n*. However, in (22) the verb *loq'o* 'to love' is neither marked for A, which would have been the first person, nor for O, which would have been the second. Instead, it is unmarked (B3). On the other hand, in the later example from the 19th century, agent indexation has developed because in (23), *loq'o* is no longer unmarked but instead features the first-person singular prefix *in-*.

(22) Colonial Q'eqchi' AFC (Berendt (1875: 39) in Holtmann (2022: 6))
<hain naloɛonɛ auech>
Ha'in *na-loq'o-n-k* *aw-ech.*
1SG PROG-love.B3-AF-TS A2SG-RN
'I love you.'

(23) 19th century Q'eqchi' AFC (Berendt (1875: 39) in Holtmann (2022: 6))
<lain nin loḳonḳ auetx>
La'in *n-in-loq'o-n-k* *aw-ech.*
1SG PROG-B1SG-love-AF-TS A2SG-RN
'I love you.'

Since a third person is nowhere involved in the expression 'I love you', it is possible to state with confidence that the original AFC indeed does not agree with either of the verbal

[310] Nonthird person > third person plural > third person singular (Dayley 1981b: 482) That is, the AFC always agrees with the argument that is highest according to this hierarchy. For instance, if the patient is a first or second person and the agent is a third person, the AFC will agree with the patient.

arguments (Holtmann 2022: 7). This is even explicitly described in the colonial grammar copied by Berendt: "The verb (is) always in the third person singular."[311] (Berendt (1875: 39) in Holtmann (2022: 7)). I agree with the conclusion in Holtmann (2022: 8) that this is likely to be the original AFC indexation pattern. Though it is true that this pattern is attested less frequently across Mayan languages, it is equally true that the direction of the change suggests an original state where no indexation of either semantic agent or patient was present at all. This is also the best way to account for the variety seen in the languages today (Holtmann 2022: 8). Unmarked forms often invite all kinds of reanalyses.

All of this is significant because Holtmann likewise proposes to view the AFC of Hieroglyphic Maya as not indexing either agent or patient. This will be treated in the following chapter.

7.1.5 Focus in Cholan

As pointed out in chapter 7.1.3, Cholan languages are usually named among those Mayan languages that do not use the agent focus construction. The following example (24) from Ch'orti' shows that A can be focused without additional morphology (Quizar 2020: 263). The verb is exactly the same in (24) as the transitive verb in a non-focused sentence in (25).

(24) Ch'orti' A focus (Quizar 2020: 263)

Ja'x uwinkir e tumin k'ani uwajpyon.

ja'x	*[u-wink-ir*	*e*	*tumin]$_F$*	*k'ani*	*uw-ajp-i-on*
3SG	A3-man-POSS	DEF	money	want	A3-grab-TS-A1PL

'it is the owner of the money who wants to grab us'

(25) Ch'orti' *ajpi* without focused A (Hull 2016: 356)

Uwajpyo'n[312] *e sakojpa'r.*

Uw-ajp-i-on	*e*	*sakojpa'r.*
A3-grab-TS-B1PL	DEF	dawn

'Dawn is already upon us. [Literally: It grabs/grabbed us the dawn.]'

However, the AFC is not absent from all Cholan languages – we do in fact find it in the hieroglyphic texts. Even though not many examples are known so far, some authors

[311] Span. "Siempre el verbo en 3a persona de singular."

[312] Note that Quizar gives this form without laryngealization in the example above. This is likely a case of the variation described by Hull (2016: 8).

(Lacadena 2000; Hull, Carrasco & Wald 2009) describe specific agent focus constructions that use an intransitivized verb form when referring to an A as in (26):

(26) AFC in Hieroglyphic Maya (Kerr 1398) (Hull, Carrasco & Wald 2009: 37)[313]
<hi-na PAT-ta b'u-ni-ya "Jaguar-throne"-na TE'-BAH TOK'-BAH>[314]
hiin patb'uniiy "jaguar throne" te' baah took' baah
hiin$_F$ *pat-b'u-n=iiy*[315] *"jaguar throne"*
1SG overturn.B3-CAUS.POS-AF=ADV throne
te' *baah* *took'* *baah*
wood image/strike flint image/strike
'It is I who overturned the "jaguar throne". There were images/strikes of wood, images/strikes of flint.'

The sentence structure of fronted independent pronoun *hiin* combined with a verb in the third person with antipassive suffix *-n* strongly resembles an AFC. Further examples are presented in (27), (28), (29) and (30). Examples (27) and (28) show instances of a fronted third person independent pronoun, while (29) and (30) provide two further examples with a first person singular. Example (29) is especially interesting since a negation is present and therefore the independent pronoun is indeed in the focus and not the topic position since it occurs after the negation (see chapter 7.1.2).

(27) AFC in HGM (Pomona Panel 4) (Hull, Carrasco & Wald 2009: 38)
<ha-i ILA-ni-ya>
haa'$_F$ *ila-n=iiy*
3SG see-AF=ADV
'It is he who witnessed.'

[313] Lacadena (2000) was the first to identify this sequence as an AFC, though he did not provide the context of the full phrase in his paper. Hull, Carrasco and Wald (2009: 38) state that he simply identified it as an antipassive and that they consider this not specific enough. This is incorrect: Lacadena (2000: 170) explicitly refers to the construction as "agent-focusing antipassive construction" following a hint given to him by Robertson in a personal communication.

[314] Conventions for transliteration of Maya glyphs: Upper case for logograms (signs referring to words), lower case for syllabograms (signs that denote syllables).

[315] Davletshin (2013b: 80) classifies *=iiy* as a "deictic clitic, mostly referring to events in the past". According to Lacadena (2013: 60), *=iiy* has a meaning of 'already' or 'in the past' while *=ij-iiy* refers to 'in the future'. However, the former entry has a note saying "full form -ijiiy?" so the difference of past versus future may not be encoded via *ij*. For the time being, I gloss it as a general adverb without specifying its meaning.

(28) AFC in HGM (Copan Stela A) (Hull, Carrasco & Wald 2009: 38–39)

<ha-o-b'a pa-sa-no-ma "portal"-ya ma-ka-no-ma "portal"-ya>

ha'o'b'$_F$	*pas-n-o'm*	"portal"	*mak-n-o'm*	"portal"
3PL	open-AF-RES	portal	close-AF-RES	portal

'It is them who opened the portal and closed the portal.'

(29) AFC in HGM (K0793) (Hull, Carrasco & Wald 2009: 40)

<mi-hi-na che-ke-na ? ?-b'a ya-la-ji-ya ?-EK?>[316]

mi-hiin$_F$	*chek-een*	*?*	*?y-al-aj=iiy*	*?*	*ek'?*
NEG-1SG	appear-AP	?	?-say.B3-RES=ADV	?	Ek'

'"It is not me who appeared", said ? Ek'.' [Hull, Carrasco & Wald: 'I am the one who did not appear' with *chek-een* as B1SG]

(30) AFC in HGM (Kerr 1440, B4'–F2') (Hull, Carrasco & Wald 2009: 37)

<hi-na ?-?-ya-si tz'i?-na u-? che-he-na u-TZ'IB? li ?-?-ya-si tz'i?-na u-?-?-na sa?-ja-la>

hiin ??yaas tz'iin u? cheheen utz'ihbil ??yaas tz'iin u?n sajal

hiin$_F$	*?-yaas*	*tz'iin*	*u-?*	*cheh-een*	*u-tz'ihb-il*
1SG	"name"	Tz'iin	A3-?	say-AP	A3-writing-POSS

?-yaas	*tz'iin*	*u-?-n*	*sajal*
"name"	Tz'iin	A3-?	"title"

'It is I, [name] Tz'iin ?, who said: "It is the writing of [name] Tz'iin, [title]."'

I propose slightly different readings for (29) and (30) than the authors do. This is based on the idea that in both cases the verbs (*chekeen* and *cheheen*) are in fact transitive verbs with an antipassive suffix *-en*, not intransitive ones with a B1SG suffix. Grube (1998) had proposed to understand *cheheen* as a quotative particle unmarked for person. Hull, Carrasco and Wald (2009) instead offer an analysis of *cheh* 'to say' + B1SG *-een*. Their argument is that the verb would then agree with the agent, the independent pronoun *hiin*. However, Houston in (2017) still states that he prefers the interpretation by Grube for this verb because "on ceramic texts and Ceibal Hieroglyphic Stairway 1, among other places, a switch from an involved declaration ('I say') to the statement of a name may require too many pivots in point-of-view."

[316] Logograms or syllabograms with a question mark indicate insecurity on the authors' part. Independent question marks indicate completely unreadable signs.

In (26), the verb does not agree with the agent. If the analysis proposed by the authors is correct, the verbs in (29) and (30) would agree with the agent. This difference would then need to be explained. On the one hand, the fact that we have two cases of agent indexation versus one case of non-agent indexation could be taken as evidence that the former is more correct. On the other hand, it is difficult to find an explanation for the single deviating case where the verb does not seem to agree with the agent.[317] The position of the arguments on the person hierarchy cannot be used as an explanation because in all three cases, the agent is a first-person pronoun while the object seems to be either third person (the "jaguar throne") or a more abstract 'it' as in 'I said [it]'.[318] In any case, the object is not a first or second person and thus not higher on the hierarchy.

On the other hand, if we understand *-(ee)n* as an antipassive suffix, a reflex of the PM *-*Vn* antipassive marker as in examples (26)–(28) or as from Ch'orti' (*way-an* 'AP12 to sleep', *xur-on* 'AP2 to cut'), all three examples show unanimous B3 indexation. Though it seems as if Ch'orti' only attests antipassives in *-on, -o* and *-yan* as well as *-ma*, which is an innovation (Hull 2016; Quizar 2020), I suspect that other verbs may be explained in the same way. For example, though the quotative particle today is mostly used as *che* without additional suffixes, it is also attested as *che'n*.[319] Other Mayan languages suggest that the suffix could indeed feature an *e* vowel, e.g., in K'iche' in example (31):

(31) K'iche' AFC with *-en* (López Ixcoy 1997: 369)

Ix$_F$	*x-oj-k'ak'al-en*	*pa*	*ri*	*nimaq'iij.*
2PL	COM-B1PL-control-AF	PREP	DEF	party

'It was you (pl.) who controlled us at the party.'

Another example from Ch'orti' could be the additional detransitivizing suffix *-V$_1$n* in which *V$_1$* mirrors the root vowel (see Appendix E section b.7.5). Its semantics are described by Dugan as giving the "sense that the grammatical subject is emitting or producing a particular

[317] I am confident that a big corpus study of hieroglyphic texts will uncover more examples where no agreement takes place. Unfortunately, most sentences I found (e.g., there are two more examples in Gronemeyer 2014: 164, fn. 351) involve a third person and are therefore inconclusive as to the agreement.

[318] 'I appeared' is more problematic. However, there seems not to be consensus on the semantics of the verb anyway: Boot (2009: 50) gives the meaning as 'to clear (sth.)' instead. The whole sentence in general seems not to be well understood.

[319] See, e.g., Fought (1972: 215) *ka-che'n* 'we say' (C1PL-say), "E noxib'ob'" in Pérez Martínez (1996) *che'nob'* 'they say'. *Che'n* is exactly what we expect from an earlier form *chehen following the contraction rules discussed in chapter 6.

kind of sensory experience" (Dugan 2013: 76) and I suspect that it goes back to the antipassive historically.

If, then, the verb is unmarked for person (B3) in all cases, it would theoretically be possible to assume that it shows patient indexation since we do not have a sentence without a third person in the hieroglyphic corpus. This is how the authors analyze the situation (Hull, Carrasco & Wald 2009: 39) – interestingly, despite their proposal of analyzing *cheheen* as featuring a B1SG suffix *-een*, which would mean that it agrees with the agent.

However, this assumption poses problems for the interpretation of the agent focus construction in another Cholan language, Choltí. In Choltí, only one case of the AFC[320] (explicitly identified as an antipassive in the manuscript) is reported (Holtmann 2022: 11). It is given in example (32):

(32) Choltí AFC (Morán (1695: 17, 43) in Robertson, Law and Haertel (2010: 230))
AI: <Dios εoεian taba>; AII: <dios coquian taba>
Dios$_F$ *kohk-yan* *t=a-ba*
God protect-AF PREP=A2SG-self
'It is God who protects you.'

If in Hieroglyphic Maya it might have seemed as if the AFC agreed with the patient, this is impossible here since the patient is in the second person, whereas the verb is in the third (Holtmann 2022: 11). If the verb agrees at all, it does so with the agent. Since it is generally assumed that Choltí is the direct descendant of Hieroglyphic Maya, one would then need to explain how the language lost patient indexation and developed agent indexation.

Compared to that, the solution proposed by Holtmann is much simpler: If we assume that the verb did not agree with either argument in Proto-Mayan, we can assume the same for Hieroglyphic Maya and even for Choltí, though it would be equally possible for Choltí to have developed agent indexation from a stage without indexation – as Q'eqchi' has done (Holtmann 2022: 11). In the cases of both languages, unfortunately, we lack examples without any third person (as we had for Q'eqchi'), which would conclusively prove that there is in fact no indexation of either semantic agent nor patient.

[320] The manuscript explicitly identifies this sentence as containing an antipassive (called "absolute [Span. absoluto]" by the colonial authors) and states that this is a rare formation that is not commonly used anymore in Choltí (Morán (1695: 16–17, 43–44) in Robertson, Law and Haertel (2010: 230)).

At first it might seem counter-intuitive to have an intransitive verb not agree with its subject (i.e., with the semantic agent). However, observe that this is the same way that focus is formed in other languages like English. After all, the construction of Hieroglyphic Maya can be reproduced without difficulties in the English translation: 'It is I who has overturned the jaguar throne.' On the other hand, indexation of the agent would be ungrammatical: *It is I who have overturned the jaguar throne.

If there is evidence for the use of the AFC in Hieroglyphic Maya on the one hand but it does not exist in Ch'orti' anymore on the other, this means that it was either lost or that it grammaticalized into something else instead. In the next chapter, I will explain how this focus construction could have become set C in Ch'orti'.

7.1.6 Agent focus construction > set C

Starting from a verbal system like the one in Hieroglyphic Maya, we assume that at the "Pre-Ch'orti'" stage, there was no grammatical distinction of completive and incompletive. As demonstrated in examples (33) and (34), it would have been possible to understand the sentences both in a completive and in an incompletive sense. To better illustrate this, I keep the forms exactly as they are in present-day Ch'orti' phonologically.[321]

(33) Pre-Ch'orti' transitive verb
*uw-ira
A3-see.B3
'he sees/saw it'

(34) Pre-Ch'orti' intransitive verb
*k'axi-et
fall-B2SG
'you fall/fell'

The focus position puts emphasis on the element that is focused. In general, independent pronouns are used for this in Mayan instead of purely fronted set B. Emphasis is in fact their

[321] For the schematic representation of how an agent focus construction could have become set C, I use reconstructed Pre-Ch'orti' forms that basically correspond to Ch'orti' except for those elements that do not exist in the language anymore, namely the older paradigm of independent pronouns that became set C (see Table 24 for an overview and discussion of Hieroglyphic Mayan independent pronouns). I use the youngest attested forms of Hieroglyphic Mayan pronouns, i.e., the ones with the short vowels (*hin* instead of *hiin* and *ha'* instead of *ha*).

main function: they often occur in the context of a syntactically "marked" construction because they are not obligatory (Hull, Carrasco & Wald 2009: 38). If S or O were put in the focus position, the verb would not have been used in a form with special morphology:

Pre-Ch'orti' S focus (no AFC): *k'axi* 'IV1 to fall':

1SG	*hin	k'axi	'It is I who falls/fell.'
2SG	*hat	k'axi	'It is you who falls/fell.'
3SG	*ha'	k'axi	'It is him/her who falls/fell.'
1PL	*hon	k'axi	'It is us who falls/fell.'
2PL	*hix	k'axi	'It is you (pl.) who falls/fell.'
3PL	*ha'(ob')[322]	k'axi	'It is them who falls/fell.'

Pre-Ch'orti' O focus (no AFC): *ira* 'TV to see':[323]

1SG	*hin	uy-ira-en	'It is me whom s/he sees/saw.'
2SG	*hat	uy-ira-et	'It is you whom s/he sees/saw.'
3SG	*ha'	uy-ira	'It is him/her whom s/he sees/saw.'
1PL	*hon	uy-ira-on	'It is us whom s/he sees/saw.'
2PL	*hix	uy-ira-ox	'It is you (pl.) whom s/he sees/saw.'
3PL	*ha'(ob')	uy-ira(-ob')	'It is them whom s/he sees/saw.'

On the other hand, the agent focus construction would have still existed in the language because it is attested in Hieroglyphic Maya. For agent focus, the verb would have been intransitivized, e.g., with the agent focus suffix *-(V)n* that we have seen used in Hieroglyphic Maya in the previous chapter.

Pre-Ch'orti' A focus AFC: *ira* 'TV to see':

1SG	*hin	ira-n	'It is I who sees/saw.'
2SG	*hat	ira-n	'It is you who sees/saw.'
3SG	*ha'	ira-n	'It is him/her who sees/saw.'
1PL	*hon	ira-n	'It is us who sees/saw.'

[322] It seems that *-ob'* is optional in the third-person plural form even today.

[323] There is some additional uncertainty in the case of O focus as I cannot provide any actual examples from the hieroglyphic corpus. Theoretically, this is what the forms would look like. The independent pronoun would be used to focus the object of the clause. The verb itself would be transitive with the regular indexation of both A and O via set A and B respectively.

2PL	*hix	ira-n	‘It is you (pl.) who sees/saw.’
3PL	*ha’(ob’)	ira-n	‘It is them who sees/saw.’

My proposal is that the pronouns employed in the S-focus construction and AFC construction, i.e., pronouns that placed **emphasis** on the subject, were reanalyzed as **markers** of subject and became set C indexes in Ch’orti’. The functional change from focus marking to aspect marking will be addressed in the following chapter (7.2). I will approach the question from a purely formal point of view first.

Generally speaking, it is well known that bound person forms often originate in corresponding independent forms (Cristofaro 2013: 78). Therefore, the idea as such practically suggests itself. For focus of S and A, the fact that the verbs are intransitive in both cases would have provided the environment to reanalyze the independent pronouns as S markers instead. This was additionally facilitated by the fact that the verb did not agree with either S or A (or O) as we have seen in chapter 7.1.4.[324] That is, the verb was unmarked and this made it possible for the independent pronouns to become inflectional prefixes. Perhaps this development was even facilitated by the fact that the suffixes of the independent pronouns and the set B suffixes already showed some differences as early as in Hieroglyphic Maya (cf. Table 24 in chapter 7.1.1). On the other hand, the focus of O would have been viewed as something different entirely because it was used with a transitive verb, therefore the construction was not parallel to the other two. Since the employed verb form is not unmarked but already has set A prefixes, there was less room for a reanalysis of the independent pronoun.

S (intransitive):	1SG	*hin	k’axi	‘It is I who falls/fell.’
AFC (intransitive):	1SG	*hin	ira-n	‘It is I who sees/saw.’
↕				
O (transitive):	1SG	*hin	uy-ira-en	‘It is me whom s/he sees/saw.’

Table 26 gives an overview of independent pronoun paradigms of K’iche’ (as a typical representative for Mayan independent pronouns) and all Cholan languages. It shows that the independent pronouns of Ch’orti’ do not coincide at all with those of Hieroglyphic Maya. Only the third person looks slightly similar superficially, though in that case it would feature

[324] Ideally, one would be able to show that in an S-focus construction, the verb also does not agree with the S but is always in third person. So far, I have not found any such examples in the hieroglyphic corpus. However, it makes sense to assume that a focus construction works the same way with all intransitive verbs, be they root intransitive or derived stems like antipassive. Therefore, it is to be expected that the verb in an S-focus construction also does not agree with the S.

additional morphology (*-x*(*ir*)). Choltí on the other hand looks neither like Hieroglyphic Maya nor like Ch'orti'. With the theory laid out in this chapter, it makes sense that Ch'orti' innovated the paradigm of independent pronouns because the old paradigm would have been grammaticalized to set C. The paradigms of Ch'orti' (as well as Choltí) must therefore be interpreted as innovations. This is confirmed by the other Cholan languages, Chol and Chontal, where it is also not immediately obvious what development they went through, though some forms remain reminiscent of Hieroglyphic Maya.[325] Hieroglyphic Maya itself, on the other hand, strongly resembles the paradigms of other Mayan languages and consequently also the one of Proto-Mayan.

Table 26. Independent pronouns of K'iche' (Can Pixabaj 2017: 467), Hieroglyphic Maya (Lacadena 2013: 18; reconstructed forms mine), Ch'orti', Choltí (Moran (1695: 10–11, 39) in Robertson, Law and Haertel (2010: 224)), Chol (Vázquez Álvarez 2011: 153) and Chontal (Smailus 1973: 218).

	K'iche'	HG Maya	Ch'orti'	Choltí	Chol	Chontal[326]
1SG	*in*	*hiin* <hi-na> *hin* <hi-ni>	*ne'n*	<natzen>	*joñoñ*	<nadzon>
2SG	*at*	*hat* <ha-ta>	*ne't*	<natzet>	*jatyety*	–
3SG	*ri are'* *(ra're')*	*haa'* <ha-i> *ha'* <ha-a>	*ja'x(ir)*	<ne>	*jiñ*	<hain>
1PL	*(ri) oj*	**hon	*no'n*	<natzon>	INCL *joñoñla* EXCL *joñoñloñ*	–
2PL	*(ri) ix*	**hix	*no'x*	<natzox>	*jatyetyla*	–
3PL	*ri e are' (ri a're', ra're')*	*ha'o'b'* <ha-o-b'a> *ha'ob'* <ha-o-b'o>	*ja'x(ir)-(ob')*	<natzob>	*jiñob*	<hainob>

As the table demonstrates, the third person singular and often also the plural deviate from the rest of the forms. Mayan languages show a wide variety of elements that are used instead to refer to the third person in the paradigm of independent pronouns. However, most (if not all) of them have in common that they use demonstratives or other deictic material for this. For instance, in K'iche', they are "strengthened" by the far-deixis demonstrative *ri*, which is also optionally used for 1PL and 2PL (see chapter 3.1). However, the second element, *are'*, is not

[325] A reconstruction was attempted in Mora-Marín (2009a) but it needs to be revised because he reconstructs a Proto-Cholan pronoun stem for each deviating form where in my view some of them are innovations.

[326] For Chontal, I use the oldest source (Acalan Chontal from 1610/12) because Modern Chontal differs even more from both Hieroglyphic Maya and Acalan Chontal.

immediately transparent to me, though the *a* is reminiscent of another Mayan deictic element that will be discussed later in chapter 7.3.2.

If we contrast modern-day set C of Ch'orti' (e.g., C1SG *in-*, C3 *a-*) with the independent pronouns of Hieroglyphic Maya above, we see that the *h-* is missing from set C. I will discuss whether this can be explained by initial loss of **h* in Ch'orti' later in chapter 7.3.3, but will, for now, focus on the morphosyntactic derivation of set C from the AFC.

The base with *h-* is well-known from other Mayan languages: Kaufman (2015: 804) reconstructs a "generic determiner" PM *ha that is the base for many Mayan independent pronoun paradigms. We have seen in the examples in chapter 7.1.2 that the focus construction not only involves an argument in the preverbal focus position but also often a specific focus particle. In K'iche', this is *aree*, which strongly resembles the third person independent pronoun element *are'*. In other languages, e.g., Popti' (Q'anjob'alan), it is *ha'* as in example (35).

(35) Focused A in Popti' (Craig 1977: 213)

Ha'	*naj*$_F$	*x-'il-ni*	*ix.*
FOC	3SG.M	COM-see-AF	3SG.F

'It was him who saw her.'

The Popti' focus particle is certainly cognate to 3SG *haa'/ha'* of Hieroglyphic Maya. It seems that in Hieroglyphic Maya, no designated focus particle was necessary: focus was expressed by putting the independent pronouns in the focus position. On the other hand, in modern languages like Popti', the third person independent pronoun became a designated focus particle that is then used with new independent pronouns in cleft-like constructions.

It is likely no coincidence that the focus particle is identical or very similar to demonstratives or deictic elements in the languages. The development of focus marking from demonstratives is well established cross-linguistically, especially in the following grammaticalization chain with or without the intermediate stage of personal pronoun (Kuteva et al. 2019: 140):

demonstrative (> personal pronoun) > copula > focus

Returning now to the grammaticalization of set C, it is worth pointing out how the proposed hypothesis interacts with Holtmann's ideas aside from supporting the fact that there likely was no indexation of either agent or patient initially. Holtmann (2022: 9) reconstructs a stage of "(optional) pragmatic agreement" in between the AFC without indexation of either semantic

agent or patient and the one with obligatory indexation either of agent, patient or both. He proposes that this optional agreement would have followed the already mentioned person hierarchy based on the prominence of the arguments yielding a pattern as the one attested in, e.g., K'iche'. For Ch'orti', there is no evidence for this intermediate stage. Since Holtmann (2022: 9) himself admits that this stage is not attested in any diachronic layer of any Mayan language, it is perhaps not a necessary step in the development. In a final step, either agent or patient indexation or indexation following the person hierarchy became obligatory in the individual Mayan languages, likely based on which argument agreed with the verb more frequently (Holtmann 2022: 9).

For Ch'orti', this generalization of one or the other indexation patterns apparently never happened. Instead, the focus semantics of the construction was lost altogether because it was functionally reanalyzed as something different. This reanalysis will be the topic of the following chapter.

7.2 From focus to aspect

We now need to provide evidence for the fact that it is possible for a focus construction to develop into aspect marking – more specifically, into marking of incompletive aspect since this is the supposed outcome in Ch'orti'.

7.2.1 Typological parallels

A systematic correlation between information structure and grammatical aspect has already been observed by Hopper (1979: 213) who reported a strong correlation for the perfective aspect to be used when recounting events (foreground), while the imperfective aspect is used to convey additional information and comments (background). He even went as far as viewing "aspectual distinctions […] as DERIVING from discourse, rather than as ready-made devices 'deployed' in discourse because they happen already to exist" (Hopper 1979: 217). That is, he proposed to understand the aspectual distinction as a result of discourse strategies: "Aspect considered from a discourse perspective is a device or set of devices which exists in order to guide the language user through a text." (Hopper 1979: 219).

7.2.1.1 Languages of Africa

A specific interaction of aspect and focus has been reported for various African languages (e.g., Hyman & Watters 1984; Güldemann 2003; De Kind et al. 2015). Hyman and Watters (1984: 233–234) describe that in many related and unrelated African languages, there are two sets of tense-aspect markers. Though the details vary across languages, the authors demonstrate that in all cases some property of focus determines the choice of tense-aspect markers from the two corresponding sets. In examples (36) and (37) from Aghem (Grassfields Bantu), the statement is essentially the same. P_1, synchronically the marker of past tense, has the allomorphs *mɔ̀* and *máà* and according to Hyman (1979) *máà* is simply a focused variant of *mɔ̀*. This is reflected in the respective semantics because (36) is a neutral proposition whereas in (37) the same proposition is focused:

(36) Aghem unfocused (Hyman & Watters 1984: 234)

m̀	*mɔ̂*	*zɨ́*	*kɨ́-bɛ́*	*nɛ́*
I	PST$_1$	ate	fufu	today

'I ate fufu today'

(37) Aghem focused (Hyman & Watters 1984: 234)

m̀	*máà*	*zɨ́*	*bɛ́-ˈkɔ́*	*nɛ́*
I	PST$_1$/FOC	ate	fufu	today

'I DID eat fufu today'

In this case, the choice of the aspect marker depends on whether focus is pragmatically intended or not, i.e., the use of one or another morpheme is at the discretion of the speaker. However, sometimes the choice is not a matter of pragmatics but one of grammar so that the use of one over another tense-aspect marker is obligatorily dictated by grammatical rules.

Efik (a Cross River language of Nigeria) also has two sets of tense-aspect markers that are chosen depending on whether focus is intended. However, as Table 27 shows, the progressive aspect only exists in the focused subset of the present tense (Hyman & Watters 1984: 246):

Table 27. Tense-aspect paradigm of Efik (Hyman & Watters 1984: 245).

	[–FOC]	[+FOC]
past	*-kV-*	*-ma*
present	*ø*	*-mV-* [–PROG] *kɛ́-* [+PROG]
future	*dì*	*-yɛ́-`*

Unfortunately, the authors do not provide an actual example of a progressive sentence in their paper. However, as becomes clear from Table 27, the choice of a [+FOC] marker becomes obligatory when the intention is to express something in the progressive aspect. Additionally, the same is demonstrated for a different language, Haya (Bantu). The following two examples from Haya contrast a present tense sentence in (38) with a progressive aspect sentence in (39). The only difference between the two is the prefix *ni*, which is otherwise used in the language as a focus marker in cleft constructions as in (40) as well as an identificational copula as in (41):

(38) Haya present (Hyman & Watters 1984: 260)
ba-mu-kóm-a
'they tie him up'

(39) Haya progressive (Hyman & Watters 1984: 260)
ni-ba-mu-kóm-a
'they are tying him up'

(40) Haya cleft focus (Bennett and Sterk (1977: 182) in Güldemann (2003: 324), glossing simplified)

ní	*ḿbwá*	*ky'*	*éy'*	*ómu-sháíj'*	*a-hail'*	*éŋkoni*
FOC	dog	which	REL	man	PST-give	stick

'which dog did the man give a stick to' [lit.: it is which dog …]

(41) Haya identificational copula (Hyman & Watters 1984: 260; Güldemann 2003: 324)

ní	*Káto*
COP	Kato

'It's Kato.'

The match between progressive marker and focus marker is not a matter of random homophony; the two are in fact etymologically related and share the same morphosyntactic structure: the progressive marker, too, historically goes back to 'it is (the case that) X' (Güldemann 2003: 324). The scope extends to "the immediately following constituent, which can be a nominal or a predicate" (Güldemann 2003: 324). In the case of the progressive, it is the whole predication, i.e., the entire action, that is focused.

Building on the work by Hyman and Watters (1984), Güldemann (2003) presents more evidence for the isomorphic marking of focus and progressive in Bantu languages where focus markers occur as progressive markers either in related languages or even in the same language but in different contexts or with the possibility of both readings "which indicates that this isomorphism can be the direct result of a historical change from one to the other" (Güldemann 2003: 331–332).

7.2.1.2 Standard Sino-Tibetan Nominalization

A second typological parallel[327] is described in Bickel (1999), which primarily deals with nominalizations in Kiranti (Sino-Tibetan) languages. The so-called "Standard Sino-Tibetan Nominalization" (SSTN) describes a morphological convergence of syntactic functions that is common to many Sino-Tibetan languages where relative clause and attributive/genitive markers are identical to nominalization devices. Bickel discusses the languages Belhare, Limbu and Athpare. In Belhare, the SSTN marker is *-hak ~ -khak,* and in Limbu, it is *-pa ~ -ba*, while "number-differentiating articles" fulfill the same function in Athpare (Bickel 1999: 271–272).

Sentences that bear such markers can function as non-embedded independent utterances in which case, Bickel (1999: 280) argues based on a parallel from Yup'ik Eskimo, they have a "vividness and sometimes exclamatory force" (Woodbury 1985: 76) that essentially makes them focus constructions. By contrasting these focus constructions from Belhare with parallel constructions from the closely related languages Athpare and Limbu, Bickel shows that there, they have been grammaticalized to take on some additional functions. In Athpare, all questions seem to be obligatorily focused (Bickel 1999: 287) as example (42) shows. Keep in mind that Athpare does not use nominalizations but articles that are inflected for number instead (see above).

[327] Further typological parallels named in Güldemann (2003: 347) include: Kambera (Central Malayo-Polynesian, Austronesian; Klamer (2000)), Albanian (Indo-European; Hans-Jürgen Sasse p.c.; Orel (1998: 336–337)) and, curiously, Yucatec Maya (Lehmann 2000). Apparently, Lehmann reported a grammaticalization change in Colonial Maya in which a cleft construction with the auxiliary verb *ka'h* 'to do' and expressing predication focus developed to a progressive. To my knowledge, this talk has not appeared in published form.

(42) Athpare question (Ebert (1997: 111–112) in Bickel (1999: 287))

khan-na	*hit-na-ga*	*suga-ci*	*a-nis-u-c-e-g-i?*
2-ERG	DISTR-DEM-NSG.ART	parrot-NSG	2-see-3U-NSG.U-PST-NSG.ART-Q

'Did you see those parrots?'

This is not at all a surprising development, "since question words are inherently focused" (Bickel 1999: 287). However, since the marking became obligatory, "the discourse function of stimulating controversy fades away" (Bickel 1999: 287) and the focus marker is now "no longer under pragmatic but under grammatical control" – a parallel to what happens with focus marking in African languages as discussed above. In other words, the former function of the focus construction, namely, to draw the attention of the listener to new and potentially controversial key information in discourse, is lost with time when the construction is no longer employed for emphasis but obligatorily. This leaves room for it to become a simple grammatical marker, e.g., a marker of questions.

If in questions, the focus lies on the "truth value or 'polarity' of the proposition" (Bickel 1999: 288), it is not very surprising that statements with reversed polarity, i.e., negations as in (43), are also almost always focused:

(43) Athpare negation (Ebert (1997: 132) in Bickel (1999: 288))

ni-natni-ŋ-na

see-NEG.PST-1SG-ART

'I didn't see it.'

Just like questions, negative statements also have an inherent potential for focus-marking because they reject "an alternative variable-instantiation" (Bickel 1999: 288), i.e., something could have been the case but is not and this is pointed out in a somewhat emphatic manner. Note that in English, too, the auxiliary verb *do* has become obligatory in negation and questions, while it remains in use as a focus marker in positive clauses.

Athpare apparently has broadened the functional range of this device in other ways as well. In Belhare, it is possible to focus complete propositions to (re)instantiate the proposition as a given fact, "as something 'that is the case'" (Bickel 1999: 287) as in (44) where the nominalizer *-ha* refers to the complete sentence and asserts its truth value.

(44) Belhare focus in constative function (Bickel 1999: 283–284)

[Before, gods and humans used to talk to each other and to marry each other on an equal basis. But then, after all we do the same things as the gods have been doing forever (lit., up from [the past]). Well, in earlier times, they used to do the hunting ritual *ramaŋ*, right?]

abo	*maŋ-lo*	*manua-lok-phu*	*biha*	*bari*	*n-cog-a-ha.*
now	deity-CMT	human-CMT-RPT	marriage	ECHO	3NSG-do-SBJV-NMLZ

'Now, it's that gods and humans inter-married.'

[So, a girl of a god was married (lit., brought) by a human and then...' [follows the account of how the human went hunting with the gods].]

This use is also a prominent use of the focus device (article) *n(a)* in Athpare as illustrated in (45):

(45) Athpare focus in constative function (Ebert (1997: 131) in Bickel (1999: 288))

khan-na	*aŋgreji*	*riŋ*	*a-nis-u-t-u-n-i?*
2-ERG	English	language	2-know-3U-NPT-3U-ART-Q

'Do you know English?'

nis-u-ŋ-na.

know-3U-1SG.A[328]-ART

'I do (it's the case).'

This function of focus constructions is, according to Bickel (1999: 288), very similar to the "denotative" or "constative" function of aspect where something is reported that is a matter of fact. There are many languages where the constative function can be covered by the same morphology as the imperfective aspect, e.g., Russian as in (46) (Bickel 1999: 289) or in example (45) above.

(46) Russian (Bickel 1999: 289)

Vy	*čita-l-i*	*'Vojnu*	*i*	*mir?'*
2PL	read.IPFV-PST-PL	war.ACC	and	peace.ACC

'Have you [ever] read 'War and Peace'?

[328] In Bickel's examples, A denotes the agent (as opposed to his undergoer U), not set A (as in Mayan).

The choice of imperfective over perfective aspect expresses that it is irrelevant whether the listener actually finished the book; instead, the inquiry aims at establishing a general assertion (Bickel 1999: 289).

Therefore, it is not surprising that a third Kiranti language, Limbu, has generalized the focus construction even further. Here, it appears to also mark imperfective aspect as examples (47) and (48) demonstrate (Bickel 1999: 289–290). When used with telic verbs, *-pa ~ -ba* induces a conative reading (Bickel 1999: 289): the speaker intended to complete a goal-oriented action but did not manage to do it to completion and thus imperfective aspect is used.

(47) Limbu imperfective aspect (Bickel 1999: 289–290)

phoːks-u-ŋ-ba	*mɛ-boːg-ɛ-n*
wake.up.TV-3U-1SG.A-IPFV	NEG-wake.up.IV-PST-NEG

'I tried to wake him up, but he didn't wake up. [I'll wake him up later.]'

(48) Limbu imperfective aspect (Bickel 1999: 290)

phaːks-u-ŋ-ba	*mɛ-baːks-ɛ-n.*
untie-3P-1SG.A-IPFV	NEG-come.undone-PST-NEG

'I tried to untie [the knot], but it didn't come undone.'

It seems then that Sino-Tibetan Limbu attests to the development of specifically imperfective aspect marking from a focus construction. Bickel argues that in Limbu, this conflation of focus and imperfective arose in the same way as suggested for African languages by Hyman and Watters (1984). In their paper, the authors introduced the notion of "auxiliary focus" as opposed to "constituent focus". This means that the scope of focus can not only fall on constituents like subject, object etc. but can also extend to the predicate or auxiliaries. Additionally, they argued that certain features of language are inherently focused[329]: these include negation, WH-questions or relative clauses but also apparently the progressive aspect (Hyman & Watters 1984: 260–261). Bickel concludes that Limbu likely developed imperfective aspect by shifting the focus from the polarity of the predication as in 'he <u>does</u> work' to the internal time structure of the predication 'he is (<u>now</u>) working' (Bickel 1999: 291).

The following is important for the discussion of Ch'orti': based on the parallel cases, the relationship between grammaticalized focus marking and aspect seems to be such that it is the

[329] This apparently becomes obvious in the different behavior of the tones in the languages (Hyman & Watters 1984: 260). I will not discuss this in detail here.

imperfective, not the perfective aspect which receives obligatory focus marking if such a grammaticalization takes place.

7.2.2 Application to Ch'orti'

We have sufficient evidence for a grammaticalization pathway of predicate focus to progressive/imperfective aspect. However, when we try to apply the specifics of the explanation to Ch'orti', a problem arises. Güldemann (2003: 344) reports that some languages exclude progressive constructions from inherently focused contexts, e.g., negation, or that their occurrence leads to a kind of double focus marking that is "at least more marked". The occurrence of two inherently focused constructions seems to be dispreferred. Likewise, Bickel's examples of generalized focus marking on negation and questions refer to the other two languages Belhare and Athpare, not to Limbu, which is the one that has a grammaticalized imperfective aspect from focus. Contrast this with Mayan where we know that the agent focus construction is obligatorily used with negation, questions and relative clauses (see chapter 7.1.3).

However, the bigger problem lies in the scope of the focus. All Mayan examples discussed so far were cases of constituent, not predication focus, i.e., it is always either a subject or an object that is emphasized by being put in the focus position. Example (49) repeated here from chapter 7.1.5 is usually understood as '(It is) I (who) overturned…' though in fact the sentence just consists of an emphatic pronoun in focus position and a verb that is unmarked for person.

(49) AFC in Hieroglyphic Maya (Kerr 1398) (Hull, Carrasco & Wald 2009: 37)
<hi-na PAT-ta b'u-ni-ya "Jaguar-throne"-na TE'-BAH TOK'-BAH>
hiin patb'uniiy "jaguar throne" te' baah took' baah

hiin	*pat-b'u-n=iiy*		*"jaguar throne"*
1SG	overturn.B3-CAUS.POS-AF=ADV		throne
te'	*baah*	*took'*	*baah*
wood	image/strike	flint	image/strike

'It is I who overturned the "jaguar throne". There were images/strikes of wood, images/strikes of flint.'

Unfortunately, we know too little about the pragmatics of the focus position in Hieroglyphic Maya to determine whether we are dealing with subject focus or perhaps predication focus. The difference would be the following:

Subject constituent focus: [(It is) I (who)]$_F$ overturned the jaguar throne.

Predication focus: I [overturned]$_F$ the jaguar throne.
or even I [overturned the jaguar throne]$_F$.

I can offer three solutions to this problem.

7.2.2.1 Subject focus > predication focus

First, in the case of predication focus, we could confidently propose a parallel development to the Sino-Tibetan and African languages. The studies on African and Sino-Tibetan languages at least seem to suggest that incompletive aspect marking can in fact develop from predication focus. Though very little is known so far about the typologically common developments of focus marking, it is conceivable that predication focus can develop from subject focus. Returning to Ch'orti', this would mean a pathway like the following:

Subject focus > predication focus > incompletive[330]

At this point, this is pure conjecture and therefore somewhat weak as an explanation.

7.2.2.2 Agentive marking > imperfective aspect

The second possibility is a pathway of agentive marking > imperfective aspect. Some Indo-Iranian languages, e.g., Katë (Nuristani), feature a construction where an active participle (or agent noun) is used to form imperfective aspect as demonstrated in (50) (Fries et al. 2023: 62).

(50) Katë imperfective with agent noun (Grjunberg 1980: 244)

vúze	*Kobél*	*é-l=as-um*	(or *élë asum*)
1SG.DIR	Kabul	go-IPFV.PTCP.M=COP-1SG	

'I (regularly) go to Kabul'

[330] Based on his data, Güldemann (2003: 343) claims that, as is claimed for many grammaticalization chains, the change from predication focus to imperfective/incompletive goes precisely in this direction, not the other way around, and is unidirectional.

The form employed in Katë and the other languages discussed in the paper, the *l*-participle, is an active participle/agent noun that expresses "properties/actions typical of referents" (Fries et al. 2023: 79). An agent noun may be defined as a deverbal nominal form that refers to an entity that performs the action expressed by a verb. A construction such as the Mayan one that consists of a pronoun and a verb with semantics focusing on the S (or semantic A) argument of a verb has a very similar function to an agent noun. Therefore, the following development is possible:

Subject focus: [(It is) I (who)]$_F$ overturned the jaguar throne.

↓

Agentive: I (am the one who) overturns/ed (the jaguar throne).

↓

Set C	I-overturn	(the jaguar throne)
	you-overturn	
	s/he-overturn(s)	
	etc.	

The agent focus construction puts the A in a more prominent or salient position by fronting it and deleting the O while emphasizing the parallel between A and S, which is their function as a subject. This subject salience then turns into a marker of some inherent property ("I am somebody who overturns…") functionally while formally it eventually becomes a subject marker in what is known today as set C. The fact that the new semantics are that of incompletive, not completive aspect is easily understandable from the fact that an inherent property refers to a general truth. Even though the object could still be expressed obliquely, the fact that the use of an intransitive, i.e., antipassive form for agent focus was obligatory explains the restriction of the grammaticalization of this construction to intransitive verbs.

7.2.2.3 Theticity

For the third explanation, I again turn to languages of Africa for help. In a study of cleft constructions in the Tuu language family spoken in Africa Güldemann (2010: 78) found that while it seems to be possible to account for many examples of this construction in his corpus with an analysis as contrastive focus with specific constituents as scope, a number of cases

remain where this analysis does not work. Specifically, this involves constituents that "would in unmarked sentences be the subject" (Güldemann 2010: 78–79) such as in (51).

(51) West !Xoon cleft construction (Güldemann 2010: 83)[331]

tuu	*k[u*	*si*	*quye]*
people.4	?:<4	IPFV	dance.initiation:3i

'The people (women) dance for her the initiation dance.'

The element in question, *kV*, cannot according to Güldemann have the function of contrastive focus here. There is no special emphasis on the constituent that is clefted (the women). Discourse contexts where an analysis as constituent focus does not fit in his corpus include, e.g., the beginning of a narration, exclamations, utterances of surprise, descriptions providing explanations and background to a story – essentially, discourse situations where all information is new and there is no topic (Güldemann 2010: 80–81).

For an explanation, Güldemann (2010: 85–92) calls upon the distinction between categorical and thetic statements treated extensively in Sasse (1987). Categorical statements are those where the participants and the event are clearly categorized ("logically analyzed into two successive mutually related judgements, one naming an individual and one naming an event" in Sasse's (1987: 554) terms) whereas this is not the case in thetic statements. Instead, thetic statements include all-new information without a topic (which would encode already known information, see section 7.1.2) or, in other words, sentences where "various parts of the communicated state of affairs remain unanalyzed" (Sasse 1987: 554). Sasse (1987: 512) provides the following examples:

Categorical statement (two parts):	John	is intelligent.
Thetic statement (one part):	It is raining.	

The first sentence contains an entity (John) and a statement about this entity (is intelligent) while the second one only consists of a statement. Güldemann suggests that categorical statements

> can be conceived of in terms of a sentence-internal focus-background configuration, namely a subject-topic vs. an assertive focus which can be a predicate, a predicate + non-subject terms, or just a non-subject term. Under this analysis, a marked 'thetic statement' neutralizes this very constellation [...]. (Güldemann 2010: 86)

[331] Additional glossing conventions: Arabic numerals refer to agreement class while < indicates the position of the agreement trigger (Güldemann 2010: 92). I am not entirely sure what "i" in 3i refers to. The use of square brackets is different from the one employed by me in the thesis – the cleft here is found outside the brackets.

He explicitly rejects to view thetic statements as expressing focus on the whole sentence because, as Sasse (1987: 572–573) likewise pointed out, if our definition of "focus" presupposes a background component, the focus cannot, by definition, lie on the whole utterance (Güldemann 2010: 86).

Güldemann demonstrated that the cleft construction in Tuu languages is polyfunctional as it is both used to form contrastive constituent focus and express thetic statements. More specifically, we are dealing with entity-central, not event-central thetic statements[332] as this functional interpretation of the cleft construction affects constituents with the subject role (Güldemann 2010: 87). Therefore, what the two uses share is the salience of the subject constituent (Güldemann 2010: 88).

As Aissen (2017b: 307) pointed out, the agent focus construction is used in Mayan languages not only for actual contrastive focus but also in contexts of negation, questions, relativization etc. Its use in Hieroglyphic Maya does not necessarily constitute contrastive focus either, even though I emphasized the contrast in my translations when introducing the construction in section 7.1.5. Think about the context of most Hieroglyphic Maya inscriptions: they are royal inscriptions detailing the life and achievements of Maya rulers. In that, they are more likely to constitute thetic than categorical statements in general, but perhaps they are even more likely to be thetic than contrastive. What is especially interesting about thetic statements is their typical context given in Sasse (1987: 566–567):

- existential statements
- explanations
- surprising or unexpected events
- general statements
- background descriptions
- weather expressions
- statements relating to body parts

[332] "Both share the property of 'positing something […], but they differ crucially as to what is posited: an entity-central thetic statement is a type of utterance stating the existence of an entity, while an event-central thetic statement is one which states the existence of an event" (Sasse 1987: 526).

The majority of these conspicuously closely approach the semantics of an imperfective aspect. Therefore, I propose that the agent focus construction, polyfunctional as its Tuu counterpart, was reanalyzed. Güldemann (2010: 91–92) himself suggests that a reanalysis where a thetic sentence is understood as a categorical one may occur: the thetic function becomes lost in sentences where the clefted constituent would be a subject and

> the subject role constituent is reinterpreted as a real subject-topic. Such a process might well be triggered by an overuse of the thetic structure whereby its pragmatic markedness is likely to decrease. Its final result would be the full reanalysis of a subject-central thetic statement as a normal categorical statement. (Güldemann 2010: 92)

With the fusion of independent pronouns and the verb and the emergence of set C, the possibility arose to form categorical statements for intransitive verbs. Based on the typical context thetic statements occur in and its parallels with imperfective aspect, I further suggest that the newly emerging set C was reanalyzed as expressing imperfective (or incompletive in Mayan terms) aspect.

The fact that this is not a peculiarity of the Tuu language family but in fact reflects "a universal theme" is witnessed by the number of typological parallels that Sasse (1987) names for this "possible structural parallelism between term focus and entity-central theticity" (Güldemann 2010: 87–88). Therefore, it is not unreasonable to propose this explanation for Ch'orti'.

As also suggested by Güldemann (and Bickel) above, I would argue that what makes a functional reanalysis of the polyfunctional agent focus construction as incompletive aspect possible, both in its use as a device of contrastive focus as well as the expression of thetic statements, is the fact that it is obligatory in "inherently focused" contexts like negation or questions so that its original pragmatics become bleached leaving the way open for all kinds of reanalyses. Its reanalysis as incompletive aspect in Ch'orti' would have given rise to an aspectual distinction that hitherto had not existed in the language. As is often the case, the completive then would have developed as a "leftover" – in fact, Bybee, Perkins and Pagliuca (1994: 90–91) report that this a very common way of how completive aspect develops.

7.3 Completing the set

If we compare the independent pronoun forms proposed in chapter 7.1.6 as the base for the development of set C, we see that they do not match set C in all cases (see Table 28). I will first address the development of C1PL *ka-* in chapter 7.3.1 and then that of the remaining

persons in chapter 7.3.2 because they are connected. In chapter 7.3.3, I will discuss the initial *h-* from Hieroglyphic Maya.

Table 28. Pre-Ch'orti' focused independent pronouns contrasted with set C and B.

	Pre-Ch'orti' independent pronouns	Ch'orti' set C	Ch'orti' set B
1sg	*in	*in-*	*-en*
2sg	*at	*i-*	*-et*
3sg	*ø	*a-*	ø
1pl	*on	*ka-*	*-on*
2pl	*ix	*ix-*	*-ox*
3pl	*ob'	*a- … (ob')*	*(-ob')*

7.3.1 C1PL *ka-*

As already mentioned in chapters 3.1 for K'iche' and chapter 4.2.1 for Ch'orti', there is a tendency for B1SG in- to replace A1SG *nu- ~ w-* (K'iche') or *ni- ~ (ni)w-* (Ch'orti') at least when used on verbs. Based on this, I propose a simple proportional analogy to explain how A1PL *ka-* replaced Pre-Ch'orti' B1PL *on.

Due to the intrusion of B1SG *in-* into set A, there was an overlap in the language in the marking of the subject of transitive verbs (*in-kani 'I learn (it)') marked by set A and that of intransitive verbs (*in-wayan 'I sleep') marked by set B. This overlap was then extended to the first person plural:

Pre-Ch'orti': Set A Set B

*ni-chitam 'my pig'

*in-kani 'I learn (it)' *in-wayan 'I sleep'

*ka-kani 'we learn (it)' **on-wayan 'we sleep'

↓

analogy:	A1SG *in-	:	C1SG *in-	*in-kani	:	*in-wayan
	A1PL *ka-	:	C1PL **on-	*ka-kani	:	**on-wayan

↓

Ch'orti':	A1SG *in-*	:	C1SG *in-*	*in-kani*	:	*in-wayan*
	A1PL *ka-*	:	**C1PL *ka-***	*ka-kani*	:	***ka-wayan***

One might wonder why C1PL did not likewise adopt the glides of A1PL *ka- ~ kaw-*. Theoretically, one would expect the whole construction to be replicated and thus C1PL should also show the variation *ka- ~ kaw-* as A1PL does today. However, it is plausible to assume that the analogy happened based on consonant-initial verbs as these are more frequent than the others. Therefore, no glide was involved and the usual contractions then happened when the pattern was extended to vowel- and *j*-initial verbs.

7.3.2 C3 *a-* and C2SG *i-*

The remaining forms that must be explained are C2SG *i-* and C3 *a-* (see Table 29). The third person plural is simply C3 with an optional plural suffix *-ob'* exactly as in set A and set B. I will discuss the position of *-ob'* in chapter 7.4.1.

Table 29. Pre-Ch'orti' focused independent pronouns contrasted with set C and B.

	Pre-Ch'orti' independent pronouns	Ch'orti' set C	Ch'orti' set B
1sg	*in	*in-*	*-en*
2sg	*at	*i-*	*-et*
3sg	*ø	*a-*	ø
1pl	*on (analogy > *ka)	*ka-*	*-on*
2pl	*ix	*ix-*	*-ox*
3pl	*ob'	*a- ... (ob')*	*(-ob')*

Of course, it is a bit risky to attempt to find an etymology for elements as tiny as one-vowel affixes. Nevertheless, a promising candidate suggests itself. Many Mayan languages use different kinds of combinations of the vowels *a, e, i* or (more rarely) *o* or *u*[333] to refer to deictic dimensions of near vs. far etc. Kaufman (2015: 806) calls them "second position demonstratives" and reconstructs them as follows:

*=a 'this/here'; in all major branches but Huastecan. "In comparison with *+i, which also means 'this/here', *+a probably/apparently refers to a position closer to the speaker, thus '(this) right here'." (Kaufman 2015: 806)

[333] "In deictics, /u/ and /o/ seem to be equivalent in value. I reconstruct only *o, on the assumption that *o can shift to /u/ in unstressed environments. In the same environment *e easily shifts to /i/, but *i is also etymologically distinct from *e." (Kaufman 2015: 804).

*=i ‘this/here’; in all major branches (but “ambiguously”[334] in Huastecan and Yucatecan)

*=e both ‘not near the speaker’ and ‘definite/in mind/in context’; in all branches but Huastecan. “In the absence of *+o, it covers all distal meanings.” (Kaufman 2015: 806)

*=o ‘yon(der)’; “[…] not widely attested […]. Since it has been widely lost, it probably marks the most extreme distal position” (Kaufman 2015: 806)

Many languages use these elements for a three-dimensional deictic contrast, e.g., Yucatec Maya where they are suffixed to nouns that are used with a determiner *le* (Hofling 2017: 720):

le wíinik-o’ ‘that man’
le wíinik-a’ ‘this man’
le wíinik-e’ ‘the man’

Kaufman (2015: 804) further reconstructs a “generic determiner” *ha that is the base for many Mayan independent pronoun paradigms. He describes that these can combine with the “second position demonstratives” described above to encode “whatever deictic values are signaled by the latter morphemes. Combined with *+e, it may have functioned as a definite article”. This formation must be the source of the 3SG independent pronoun of Hieroglyphic Maya, *ha(a)’* with the generic determiner in first position and the deictic element in second. Obviously, then, since I propose to derive set C from the focus construction present in Hieroglyphic Maya, the most logical step is to derive CHR C3 *a-* from HGM *ha(a)’*.

Since Ch’orti’ also has a definite article *e* (which is very likely identical with Kaufman’s *=e[335]), two of the second position clitics are attested in the language, *=e and *=a.[336] This brings up the question whether C2SG *i-* could not also go back to a deictic element, namely Kaufman’s *=i meaning ‘this/here’ (but not immediately next to the speaker because that is *=a). To do that, we need to prove that such a change can indeed take place. Though deictic elements are one of the main sources for person forms, this change may be most common with the third person. However, here, we are dealing with a second person.

[334] It is unclear to me what Kaufman means by that.
[335] Kaufman (2015: 772) himself suggests that CHR *e* comes from an earlier *he though he does not address how *h was lost.
[336] Whether or not the Ch’orti’ *i* ‘and’ continues this clitic is up for debate. Though *i* ‘and (then)’ occurs extensively at the beginning of clauses in Hieroglyphic Maya, its use in Ch’orti’ is more limited. As Ch’orti’ has extensively borrowed conjunctions and adverbs from Spanish (e.g., *o* ‘or’, *ke’* ‘that’ from Span. *que*, *konde* ~ *konda* ‘when’ from Span. *cuando*), synchronic *i* may likewise not be identical to the one attested in Hieroglyphic Maya, especially given that *o* was borrowed.

A similar case has been discussed in Olander (2022) for Italo-Romance. In Italian, the adverbial clitics *ci* 'here' and *vi* 'there'[337] have apparently replaced first and second person plural pronominal clitics 1PL *no* and 2PL *vo*[338] via the following steps:

1. In early Italo-Romance, *no* and *vo* first developed to *ne* (*ni*) and *ve/vi* by analogy the other persons 1SG *me/mi*, 2SG *se/si*, 3SG *te/ti*.
2. This vowel change led to a homophony with *ne* 'from there' in the case of 1PL and with *vi* 'there' in the case of 2PL.
3. On the way from early Italo-Romance to Italian, *ne* 'us' was then replaced by *ci* 'here' due to the homophony and reinterpretation of *vi* as meaning 'there'.

If we try to envision a similar scenario for Ch'orti', it could follow these steps:

1. A deictic element **=a*, probably attached to a base **h-*, is used to refer to the third person in the independent pronoun paradigm because B3 would otherwise have been unmarked.
2. The form is semantically transparent in that it still expresses the deictic dimension of 'directly near the speaker'. A verb with (incipient) set C in the third person is understood as (*h*)*a-wayan* 'this one (here) sleeps'.
3. The second person singular is innovated analogically to the third to (*h*)*i-wayan* 'you (there) sleep'.

Compared to Italian, we are missing a stage similar to when *no* changed the vowel and became homophonous with 'from there' thus already inviting a reinterpretation based on deixis. However, I would argue that we do not need to rely on homophony because in Ch'orti', the new third person morpheme is the deictic element itself. An extension of the formation 'this one here sleeps' to a second person 'this one there sleeps' may therefore even be more plausible than in the case of Italian.[339]

Now it would be ideal to also have cases where demonstratives not only develop into third-person forms OR first- and second-person forms, but where these two categories are mixed, ideally in a mix of second- and third-person forms as would be the case in Ch'orti'. In fact,

[337] Latin (*)*hince* > *inci* > *ci* 'here'; Latin *ibi* > *ivi* > *vi* 'there' (Olander 2022: 8).
[338] Latin *nōs*, v*ōs* > tonic *nói*, *vói* > clitic *no*, *vo* (Olander 2022: 8).
[339] To name another parallel, Cristofaro (2013: 76) lists "cislocative elements meaning 'hither', 'here' or 'this way'" as one common source for markers of inverse alignment. Though this is a rather broad parallel to the case at hand, it nevertheless proves that such a change is common.

some typological parallels can be found. In Archaic Chinese, there is a set of markers *er/nai/ruo/ru* which are demonstratives for medial distance ('neither near nor far') that developed into second-person pronouns (Long, Gu & Hong 2012).[340] Further evidence comes from Japanese. The modern pronoun *anata* 'you' used in neutral register originally formed part of the following demonstrative paradigm (Frellesvig 2010: 246–247):

- *konata* 'this direction (speaker)' (< *ko-no kata* 'this side')
- *sonata* 'that direction (hearer)'
- *(k)anata* 'that direction (distal)' (*anata* was the less frequent form at first)

The use of *kanata* or *anata* to refer to the third person is attested since Early Middle Japanese (Frellesvig 2010: 247). It is only from well into New Japanese on, towards the end of the eighteenth century, that *anata* began to be used to refer to the second and no longer the third person (Frellesvig 2010: 247). It is easy to imagine how a third person is used to refer to a second person indirectly for reasons of politeness, especially in a society that holds politeness in high esteem. This means that we are dealing with a possible pathway of how a third-person form becomes a second-person form due to the desire to be especially polite. One might speculate that the same played a role in the Ch'orti' example. Of course, this would first require a sociolinguistic and pragmatic study on politeness in Mayan.

7.3.3 The problem with *h-

The final problem concerning the forms of set C pertains to the absence of initial *h- while the independent pronouns of Hieroglyphic Maya clearly attest to the presence of it.[341] As was

[340] Kuteva et al. (2019: 142) note, however, that according to Alain Peyraube (p.c. to them), the direction was the other way around, namely, that *er/nai/ruo/ru* were first used as second-person pronouns and their infrequent use as demonstratives is a secondary development. I find this less plausible. Demonstratives inherently refer to the third person. The infrequent use as demonstratives should rather be seen as a remnant of the ancient function.

[341] It could be argued that the third-person independent pronoun *ja'x(ir), ja'x(ir)ob'* is a descendant of the HGM independent pronouns *ha(a)'*, which would then attest to the survival of *j- in Ch'orti'. Several other features would then require an explanation. First, what is the additional suffix *-'x* that likely resulted from a contraction involving a sequence *-V'V- or *-VjV? It might be the enclitic *=ix* 'already' but this would need to be justified semantically. More importantly, we need to explain the optional addition of the abstractive suffix *-ir*. Grammars seem to struggle with determining an exact function of its presence or absence: *ja'x ~ ja'xir* seem to be "largely interchangeable, sometimes both used in the same sentences to refer to the same person" (Dugan 2013: 39). A subtle distinction sometimes seems to be that *ja'xir* is used more often "[w]hen referring to an abstract concept or an indistinct group of individuals, while *ja'x* is more likely to refer to a specific individual or object" (Dugan 2013: 39).

discussed in chapter 6.2, CHR *j* (< HGM *h, j*) is retained in initial (e.g., *ja'* 'water') and final position (e.g., *chij* 'horse'). The only exception is when *j* finds itself in a vowel-medial position due to the presence of affixes, especially on verbs, and is lost due to analogical processes illustrated in the same chapter.

If it could be shown that independent pronouns were used most often with preposed particles in Hieroglyphic Maya, e.g., the discourse particle *i* 'and (then)' (Kettunen & Helmke 2020: 122) that is often found at the beginning of sentences, it would be possible to apply the explanation of vowel-medial *j* loss to this case, too. However, as far as our corpus of independent pronouns in HGM goes, so far, the examples show the opposite. In the examples presented in chapter 7.1.5, the independent pronouns most commonly occur at the very beginning of the sentence and *i* is not used. Only a single case is attested where the negation *mi* is used before *hiin*, given here again in (52).

(52) K0793 (Hull, Carrasco & Wald 2009: 40)

<mi-hi-na che-ke-na ? ?-b'a ya-la-ji-ya ?-EK?>

mi-hiin	*chek-een*	*?*	*?y-al-aj-iiy=*	*?*	*ek'?*
NEG-1SG	appear-AP	?	?-say.B3-RES=ADV	?	'name'

'"It was not me who appeared", said ? Ek'.' [Hull, Carrasco & Wald: 'I am the one who did not appear' with *cheh-een* again as B1SG]

Another point to consider about this explanation is what the most likely outcome is. Based on the results presented in chapter 6.4 in Table 21 and Table 22, the outcome would be as follows:

i + a > ya' *i + e > ye'* *i + i > i'* *i + o > yo' ~ o'* *i + u > yu'*

These are then the possible results for the independent pronouns:

CHR *-ir* has two functions. 1) It is added to body parts if these are possessed (see explanation of possession classes in Appendix E a.2). It is possible that *ja'x* etymologically goes back to a lexeme for a body part which came to denote the third person by way of a metaphor. In that case, however, we would expect a set A prefix, which is missing from the form. Furthermore, in this use *-ir* alternates with *-er*.

2) The suffix *-ir* is likewise used to derive abstract nouns from nouns and adjectives (see Appendix E c.4). In this case, set A is not required and the abstractive semantics fits the description of the use of *ja'x* vs. *ja'xir* above. It seems a plausible hypothesis to assume that *ja'x* goes back to a noun (or adjective) that is sometimes derived to an abstract noun. However, I cannot propose any specific lexeme yet.

As it stands, *ja'x* remains problematic but in any case, the presence of the suffix *-ir* strongly suggests that we are dealing with a noun at least etymologically. Therefore, it is not a reflex of an earlier independent pronoun.

	Pre-Ch'orti'		Ch'orti'
1SG	*(C)i + *hin	>	**(C)i'n
2SG	*(C)i + *hat / *hi	>	**(C)ya't / *(C)i'
3SG	*(C)i + *ha	>	**(C)ya'
1PL	*(C)i + *hon	>	**(C)(y)o'n
2PL	*(C)i + *hix	>	**(C)i'x
3PL	*(C)i + *ha'(ob')	>	**(C)ya' / **(C)yo'b'

None of these forms are what is attested in Ch'orti'. Theoretically, one could assume that the laryngealized character of the independent pronouns is what makes set C interact differently with vowel-initial verbs. However, then we would expect traces of the glottal stop in the forms that are used with verbs beginning in consonants. This is not the case. Alternatively, we could assume that a separate sound change happened before the contractions discussed in chapter 6 that did not lead to laryngealization. Further problems include the fact that if the first element is not simply *i but *Ci, e.g., NEG *mi*, this would have been expected to leave traces and we would need to explain the loss of another consonant. Elements other than *i* or *mi* are theoretically imaginable but this would need to be justified through their occurrence in the Hieroglyphic corpus. All in all, this scenario is not very likely.

There are then three possible solutions.

1. We are dealing with a very specific conditioned sound change.
2. Ch'orti' did not use the base *h- for its independent pronouns but either
    a. a different sound, which then disappeared through regular sound change or
    b. pure set B.
3. Set B changed its enclitic status to proclitic via a stage of "ditropic" behavior.

7.3.3.1 A highly conditioned sound change

If initial *j generally is retained in Ch'orti', we could still be dealing with a lost *j in the case of the independent pronouns if the sound change behind the process is highly conditioned phonetically, e.g., if it only happens in front of specific consonants or in forms with a specific number of syllables. Aside from the independent pronouns and deictic elements discussed in the preceding chapters, we have evidence for one additional lexeme where initial *j is lost in

Ch'orti' without the involvement of intervocalic position. I will describe this second case and then discuss what it has in common with independent pronouns and deictic particles.

The numeral '1' is reconstructed as PM *juun based on evidence from all Mayan languages, which mostly retain it either as *jun* or *juun* depending on whether the language retains the vowel quantity contrast or not, see, e.g., MOP *jun*, CHL *jun=*, KCH *juun*, MAM *juun* (Kaufman 2003: 1454–1458). While, e.g., the Chol form is ambiguous as to whether the initial consonant is PM *j or PM *h, whether the vowel is short or long or whether the final consonant is PM *n or PM *ŋ because these each merge in the Cholan branch, K'iche' alone seems enough to prove that PM *juun is in fact what we must reconstruct because K'iche' retains all three distinctions. In Hieroglyphic Maya, numerals are generally written logographically and therefore not spelled out in syllables. However, since a phonetic complement is attested as <-na> (Kaufman 2003: 1454) rendering the spelling disharmonic (<JUN-na>), this seems to prove that the vowel was also long in the Cholan branch initially.

Among all Mayan forms of '1', the Ch'orti' form *in=* is very unusual. As we have seen, initial *j* should not disappear. Also, the vowel is unexpected. Theoretically, this could mean that this is something else etymologically. Replacement of the numeral '1' is attested from, e.g., Russian where 'one' is replaced by 'time' (as in 'one time, two times, …') when counting: *raz, dva, tri* instead of *odín, dva, tri*. A formally obvious candidate for this would be A1SG/C1SG *in-*. However, the semantic change would be unusual; furthermore, the numeral does not show the same contractions as the person prefixes when it comes to interaction with the following stem.

A second option would be the demonstrative PM *ʔin reconstructed by Kaufman (2015: 1515) based on HUA *in-* 'A3', CHT <ha-ine> 'that [Span. aquel]', CHL *il-; jini* 'this [Span. este]' and older QEQ *ha'in* 'this (Span. esto)'. Mora-Marín (2009a) also reconstructs a demonstrative enclitic *=in for Cholan. I doubt that such a demonstrative ever existed: the Huastec index arose through inverse marking where original A3 *u-* came to mark the first person and original A1SG *in-* came to mark the third person in specific contexts (see chapter 3.2). That is, Huastec *in-* is in fact historically the first person and not some demonstrative *in that already existed in Proto-Mayan. The other cases are less clear but it should be investigated whether they can be explained through reanalyses of independent first-person pronouns as third-person ones in

focus constructions.[342] Therefore, we are left without a plausible alternative lexical origin for the numeral '1'.

Instead, we can consider a specific sound change of the sequence PRE-CHR **jun* > CHR *in*, i.e., loss of *j* and a change *u* > *i* before *n* (or perhaps generally before nasals). The sequence *jun* practically does not occur in Ch'orti' today (Hull 2016). One exception is *junio* 'june' – a direct and likely late loan word from Spanish. We can easily assume that the sound change happened before *junio* was loaned. The existence of CHR *jun* 'book, paper, ficus' is more difficult to explain. This is a culturally significant lexeme attested in all Mayan languages and reconstructed as Proto-Mayan *hu'uŋ (Kaufman 2003: 1107–1108) based on, e.g., TSE *hun*, CHR *jun*, CHJ *hu'um*, KCH *wuuj*, MAM *u'j*. Here, again, the K'iche' form *wuuj* is sufficient to demonstrate that the initial consonant is *h and not *j because *w* is the expected reflex of PM *h before *-u* in K'iche'. Likewise, K'iche' shows that the final consonant is *ŋ, not *n as the other languages, which do not distinguish between PM *n and *ŋ, might suggest. Some languages also hint at a laryngealized vowel but I consider this less secure because such a vowel would be expected to be preserved in Ch'orti' (HGM *b'u'ul / b'u'l* <b'u-la> '(black and brown) beans' (Kettunen & Helmke 2020: 100) > CHR *b'u'r* 'bean'). The spelling of HGM <HUN-na> 'book' in fact receives ambiguous readings as *hun / hu'n / huun* in Kettunen and Helmke (2020: 104). Based on the disharmonic spelling of the phonetic complement *na*, we expect either a laryngealized or a long vowel.

Be that as it may, although the reconstruction of PM *juun '1' and PM *hu'uŋ 'paper, ficus, book' is not identical, we do not expect this to play a role in Ch'orti'. PM *n and *ŋ are already merged in Hieroglyphic Maya. In Ch'orti', both HGM *h* and *j* and long and short vowels are merged, as well. According to Grube (2004: 72), this process begins in the Late Classic in word-final position. As discussed above, a laryngealized vowel would be expected to be retained, so it is unlikely that Hieroglyphic Maya had it. Therefore, the different treatment of Pre-Ch'orti' *jun '1' and *jun 'book' cannot be explained etymologically.

[342] Since the original focus construction involved an unmarked verb, the first person independent pronoun could have been reinterpreted as a third person. Keep in mind that Chol generalized the first person plural pronouns for the first person and uses special plural suffixes to distinguish 1SG from 2SG (see Table 26). It is therefore possible that the original 1SG evolved into something else. A parallel development may be seen in Chontal where the same generalization happened and the determiner *ni* (see, e.g., its use in chapter 9.2) likely developed from former A1SG *ni-*. However, this explanation does not cover Choltí and Q'eqchi' and the whole theory requires further research. What is clear though is that we cannot securely reconstruct this pronoun for Proto-Mayan.

In Ch'orti', numerals are not used on their own but depend on a numeral classifier. This is already attested in the earliest Ch'orti' source by Galindo (see Appendix B) with <Inté>, <Chaté>, <Uxté> etc. Today, too, numerals are combined with different classifiers, *-te'* being the most basic one. Since stress lies on the final syllable in Ch'orti', this would put *jun= in an unstressed position. In addition to that, **jun* apparently became a productive prefix in the formation of adjectives and adverbs, which are all stressed on the final syllable (Hieroglyphic Maya data from Kettunen & Helmke 2020):

HGM	*ch'ah* 'bitter'	>	CHR	*inch'aj* 'bitter'
HGM	*tat* 'thick, fat'	>	CHR	*intat* 'thick (of liquids), viscous, muddy'
HGM	*utz* 'good'	>	CHR	*inb'utz ~ imb'utz*[343] 'well, good, precious, …'

This produced even more contexts where *jun* was in an unstressed position. On the other hand, *jun* 'book', is mostly used without suffixes and is thus always stressed.

Taken together with the fact that Hieroglyphic Mayan independent pronouns are attested with initial *j-*, which is not present on set C in Ch'orti', it is possible to propose a conditioned sound change where initial *h- is lost in unstressed position in multisyllabic forms. Additionally, when *h was followed by *u + n/nasal, the vowel was raised to *i* (*ju > *i* |_*n*). The independent pronouns neither confirm nor contradict the special development of *jun* > *in* (with vowel change) because *u* does not occur in the pronoun forms. A remaining problem for this theory is the fact that multisyllable words starting with *j-* indeed exist in Ch'orti'. Perhaps the conditioning needs to be narrowed down further to forms where *h is part of a proclitic.

A look at a similar phenomenon of initial consonant loss in Australian languages may help. Blevins discusses the loss of consonants that "have intrinsically weak perceptual cues in word-initial position" (Blevins 2001: 482). This involves sounds like the voiced velar fricative [ɣ], the voiceless bilabial fricative [ɸ] or sequences like *wu* or *ji* [yi] that in some languages are phonologically indistinguishable from *u* or *i* respectively (Blevins 2001: 483). Certainly, Ch'orti' *j* [h] also qualifies as "intrinsically weak". Another parallel is that initial consonants

[343] Apparently with insertion of *b'* because PM *ʔutz is reconstructed as beginning with a glottal stop, not PM *b', since no language attests a different initial sound than *u* (Kaufman 2003: 199–200). A parallel case is *aj-b'uch* 'possum' < WM+LL *ʔuch 'possum' with unclear relationship to PM *huhty' 'possum' (Kaufman 2003: 577–578).
The agentive prefix *aj-* is often used with animal names in Mayan yielding a form like 'He of the possum' or 'Mr. Possum'. As a parallel feminine form with prefix *ix-* 'female' only *ixk'anan* 'earwig' (a bright yellow bug) is attested in Hull (2016), which one could understand as 'She of the yellow color' or 'Mrs Yellow'. This word is attested as <iš q'ana'n> 'female spirit of beans' in Wisdom (1950: 486).

can be lost when stress shifts from the initial syllable to a following syllable yielding unstressed/reduced initial syllables (Blevins 2001: 484). Blevins argues that reduced syllables are shorter than "normal" syllables and this shortening can lead to underarticulation of initial consonants.

Most importantly, just as in Ch'orti', there are some Australian languages (e.g., Warumungu, Warlpiri, Burarra) for which the explanation of intrinsically weak onset and/or unstressed position is not sufficient to explain their initial consonant loss. There, consonant loss happens specifically utterance- or phrase-initially, e.g., in Warumungu where initial loss affects all pronouns, which synchronically appear either initially or in second position, or in Arabana where initial consonants disappear in exclamations, vocatives and pronouns (Blevins 2001: 486). For example, Archaic Arabana *ŋama* yields both *ama* 'mother (vocative)' and *ŋama* 'milk' in Modern Arabana. What makes the phrase-initial position special compared to phrase-medial or -final position is the fact that there, initial consonants never appear after vowels, which could strengthen the pronunciation producing an intervocalic context, so that "the utterance-initial position provides fewer perceptual cues to consonant identity than other positions, all else being equal" (Blevins 2001: 486).

Blevins herself admits that there must be more to this explanation because it, too, does not explain, e.g., cases of loss of consonants that are not intrinsically weak. Perhaps a more careful phonetic study of Ch'orti' could yield interesting insights that might contribute to the investigation of phrase-initial consonant loss as a whole. A recent contribution by Culhane, Peck and Reinöhl (2023) describes a peculiar case of initial consonant loss in Kera'a (Sino-Tibetan) that cannot be explained by any of the explanations proposed by Blevins. There, it is also shown that the change is not as rare as it might seem because it has been described for Algonquian, Sogeram, Greek and other Sino-Tibetan languages (Culhane, Peck & Reinöhl 2023: 8).

However, Ch'orti' seems to be readily explainable using Blevins' suggestions. So far, the contexts that have been identified for this proposed change are independent pronouns, a definite article and the numeral 'one'. All these occur at the beginning of phrases in Ch'orti'. Additionally, the consonant that is supposed to have been lost is in fact "intrinsically weak". Furthermore, the respective morphemes are found in unstressed position because words in Ch'orti' are generally stressed on the final syllable.

As to the question of when exactly this change would have taken place, there is some conflicting evidence from Galindo: <Jacatiniti> 'thirst' likely contains an accidental metathesis of *Ataki niti'*. 'I am thirsty' (C3 *a-* + *taki* 'IV1 to dry up' + A1SG *ni-* + *ti'* 'mouth'). If we take the recorded initial <J> seriously, it represents evidence 1) for the fact that the morpheme indeed used to have an initial *h-, which strengthens the hypothesis of set C coming from the focus construction and 2) for the fact that the specific sound change discussed in this chapter happened after Galindo's wordlist was recorded. However, the evidence is hardly conclusive in this single occurrence. One may consider <Joté> '5' further evidence; however, precisely <Inté> '1' already shows the sound change as completed (see Appendix B).

7.3.3.2 Alternative base vs. no base

The second alternative is somewhat radical. The assumption that perhaps a different base than *h was used in Ch'orti' which consisted of a sound (or multiple sounds) that were subsequently lost through regular sound change is not a valid option because to our knowledge, no Proto-Mayan sounds simply disappear in Ch'orti' without leaving any trace. Is it perhaps possible that Ch'orti' did not use any base at all and that set B was not a clitic but an independent lexeme in its prehistory? The reason why I label this option as radical is the following: if it could be proven that it was indeed possible to use set B in focus position in front of the verb without additional material, this could be considered counterevidence for the theory that Ch'orti' is a descendent of Hieroglyphic Maya.[344]

In chapter 7.1.1, we discussed the fact that the status of set B as a clitic or independent form in Proto-Mayan is an unresolved question in Mayan historical linguistics. We also discussed ambiguous evidence from K'iche'. Example (53) once again shows that set B is used in front of the verb in K'iche' without obvious elements that it could be suffixed/encliticized to.

(53) K'iche' stative construction (Larsen 1988: 107)

oj *kunaneel*

B1PL doctor

'We are doctors.'

Now, Kaufman assumes that these forms were attached to a base *h-:

[344] We could then hypothesize that Ch'orti' is rather a sister of Hieroglyphic Maya. They would share large parts of morphology due to the accidental fact that Ch'orti' remained relatively archaic until the present.

> The independent pronouns of K'ichee7 do not reveal a fact which is obvious in most Mayan languages: that they are formed by attaching the Abs agreement markers to a demonstrative base, usually one derived from proto-Mayan *ha7 'it'. In fact, the Kch forms *7in, 7at, 7oj, 7ix* could be derived from pre-K'ichee7 *h-in, *h-at, *h-oj, and *h-ix. (Kaufman 2015: 443)

However, this does not work out neatly. PM *h usually does not simply disappear in K'iche' but leaves different reflexes depending on which vowel follows. In his reconstruction of Proto-Greater-K'iche'an, Campbell (1977: 36) observed the following developments:

PGK *h > *w* before *u, o* and also after *u*

- KCH *wonon* 'bumblebee' < PGK *honon (QEQ[345] *honon*)
- KCH *wuuj* 'paper' < PGK *huuj̲ (QEQ *huuh*)
- KCH *chuu(w)* 'stinking' < PGK *chuuh (QEQ *chuh*)

y before *i, e*

- KCH *jeeh* 'tail' < PGK *j̲eeh[346] (QEQ *heeh/yeeh*)

y/h/j before *a*

- KCH *ja'/ya'* 'water' < PGK *ha' (QEQ *ha'*)

h elsewhere

- KCH *eeh* 'tooth' < PGK *eeh (QEQ *eeh*)
- KCH *ch'ooh* 'mouse rat' < PGK *ch'ooh (QEQ *ch'ooh*)

Though there is counterevidence for Campbell's reconstruction in his own data, this cannot be discussed here in detail.[347] Ultimately, the exact reflexes are not important for the question at hand. What is important is the fact that, following his results, we would not expect *h to disappear without a trace before all forms of set B. Were the paradigm of independent pronouns really built upon a base *h-, we would expect the following forms:

Table 30. Expected independent pronouns for K'iche' based on Campbell's (1977) reconstruction contrasted with the attested ones (Can Pixabaj 2017: 467).

	Expected	Attested
1SG	**yin	*in*
2SG	**yat/hat/jat	*at*

[345] Q'eqchi' forms are given as a substitute for all Greater K'iche'an languages and their forms, which Campbell's reconstruction is based on. Q'eqchi' seems to retain PM *h in all contexts (Campbell 1977: 36).

[346] Alongside *h, Campbell (1977: 34; 2017: 48) reconstructs two kinds of *j for Proto-Greater-K'iche'an, 1) a velar *j [x] < PM *j and 2) *j̲, which is closer to a uvular [χ]. The latter comes from the Proto-Mayan sound *ŋ that yields *j* in some Mayan languages and *n* in others.

[347] Observe, e.g., KCH *paar* 'skunk' < PGK *pahar (QEQ *pahr*), where *h* disappears completely leaving a long vowel *aa* as a reflex of **aha* (Campbell 1977: 48), where it should not do so according to Campbell's conditioned sound change.

3SG	?	*ri are' (ra're')*
1PL	**woj	*oj*
2PL	**yix	*ix*
3PL	?	*ri e are' (ri a're', ra're')*

Here, one would need to assume that *h was lost in one form for whatever reason, perhaps in front of one of the vowels, and that the paradigm got leveled afterwards. Alternatively, one could follow Kaufman (2015: 134) who claims that some K'iche' dialects in fact do lose *h without a trace, though he only gives little evidence for this (PM *huhch 'possum' > KCH *wuuch', ʔuuch'*). His etymological dictionary (Kaufman 2003: 1506–1509) gives dialectal variants from Chikaj (KCHk) for the independent pronouns which start with *y-*: *yin*, *yet* (not **yat!), *yoj* and *yix*. Unfortunately, this dialect is missing from the entry of determiner *ha7 (Kaufman 2003: 1534). The expected form would be **ya' but I was unable to find it. Note that this would then be a generalized base *y-* without regard to what vowel follows after *h.

Now it becomes obvious why it is unfortunate that Campbell decided to standardize the dialectal data for the individual languages in his reconstruction (as discussed in chapter 2.2.4). To settle the question of whether set B could be used on its own without attaching to a base *h- (or any other base), the work basically needs to be redone while taking all dialectal variation into account.

7.3.3.3 Ditropic clitics

There is a third possible explanation. It is imaginable that both set B and the deictic elements *=a, *=i and *=e were encliticized to some base, e.g., *h-, but changed from being enclitic to this base to being proclitic to the following verb as sketched here:

*h=set B/deictic + verb

↓

*h + set B/deictic=verb

A similar process has been described in Hill et al. (2019) for Baltic, Germanic and Armenian where originally postposed local adverbs have become prepositions or prefixes. If the base was *h- as witnessed in Hieroglyphic Maya, ditropic clitic behavior would even provide a plausible scenario for the loss of *h- if the set B enclitics changed their behavior to proclitic

so that *h- would be the only thing left of the independent pronouns. It is easy to imagine that a form like that is then lost.

7.3.3.4 Discussion

Although it is impossible to settle the debate at this point, we are left with an optimistic conclusion. The attested definite article *e* in Ch'orti' is beneficial for the theory proposed in this dissertation either way: it proves that, whatever has happened, did in fact happen. At least synchronically the deictic elements do appear without *h* in Ch'orti'. If we assume that the elements are clitics and needed to be suffixed to a base *h, *e* proves that there is a conditioned sound change where initial *j can be lost (likely through a phrase-initial consonant loss as discussed above), which was perhaps facilitated by the ditropic clitic behavior. On the other hand, *e* can likewise be the result if we assume that the elements could stand for themselves without being suffixed.

7.4 Useful side effects

This chapter discusses some features of Ch'orti' (and, in the last chapter, Yucatec Maya) that I have not taken into account for the development of the argument but that may or may not be explained through the proposed theory and therefore provide further evidence for it.

7.4.1 Plural particle *ob'*

We have seen that the third person plural form of independent pronouns in Hieroglyphic Maya is attested as *ha'oob' / ha'ob'*, e.g., in example (54):

(54) Copan Stela A (Hull, Carrasco & Wald 2009: 38–39)

<ha-o-b'a pa-sa-no-ma "portal"-ya ma-ka-no-ma "portal"-ya>

ha'oob'	*pas-n-o'm*	*"portal"*	*mak-n-o'm*	*"portal"*
3PL	open-AF-RES	portal	close-AF-RES	portal

'It is they who opened the portal and closed the portal.'

Theoretically, we would then expect this independent pronoun to have yielded a C3PL prefix **ob'-. This, however, is not the case: *-ob'* only occurs as an optional suffix on the verb, not as a prefix, as (55) shows.

(55) Ch'orti' C3PL (Hull 2016: 310)

War *a-patna-ob' […].*

PROG C3-work-PL

'[They] are working.'

As I have argued previously, *ha'oob' / ha'ob'* is actually the third singular *ha'* with an added and likely optional plural morpheme *ob'* (or *oob'*) as in many Mayan languages, plural marking is optional in this form. If *-o(o)b'* were the real third-person plural form, we would instead expect a form **ho(o)b' for Hieroglyphic Maya, which is not attested. It is to be expected that Hieroglyphic Maya followed the same strategy as other Mayan languages (mostly Lowland languages like Yucatecan) in that it only marked plural optionally in the third person.[348] This strategy is continued in Ch'orti', which has the same kind of marking for the 3PL of all three index-sets:

A3SG	*u-XXX*	A3PL	*u-XXX(-ob')*
B3SG	*XXX*	B3PL	*XXX(-ob')*
C3SG	*a-XXX*	C3PL	*a-XXX(-ob')*

However, a phrase-initial *ob'* is in fact attested in Ch'orti', though not in Fought (1967; 1972) or Hull (2016). I found it in the legends recorded by Pérez Martínez (1996), e.g., examples (56) and (57):

(56) "Ob' e cha'te' compagre" (Pérez Martínez 1996)

Ob'	*e*	*cha'-te'*	*compagre*
PL	DEF	two-CL	friend

'The two friends'

(57) "E noya i e sitz'b'ir" (Pérez Martínez 1996)

ob'	*e*	*ch'o'k*	*uy-a'ryob'*	*ayi*	*e*	*mwon*
PL	DEF	rat	A3-say.PL	RPT	DEF	hawk

'the rats said to the hawk, it is said'

The same phenomenon is discussed in Wichmann (1999: 117) who confirms that he, too, only found it in PLFM materials. He calls it a "focalizing plural pronominal" and considers it likely to be cognate with HGM *ha'ob'* though he does not elaborate on this and especially leaves it to the imagination of the reader to explain the missing *ha'*.

[348] There are likely contexts in the Hieroglyphic Mayan corpus where the text refers to multiple entities but the pronoun still remains in the singular.

I am unsure whether the position of *ob'* and the fact that it is used independently and not as a suffix can be derived from its occurrence in the focus position in Hieroglyphic Mayan times. In all cases that I found it is not used preverbally, only prenominally, though one could argue that the fact that it also survived as an independent lexeme is due to its use in the focus position.

However, I do not believe that *ob'* has a "focalizing" function. In the corpus, I found only one example (58) where it is used in a focus context because it is paired with modern Ch'orti' *ja'xto* consisting of the third-person pronoun *ja'x* and *=to* 'still, yet', which means, among other things, 'it is s/he who'. Therefore, the focus semantics likely come from the use of the independent pronoun, not *ob'*. Otherwise, *ob'* is not necessarily used in detectably focalizing contexts and simply seems to be a means of forming plural, though a bigger corpus study would be needed to determine how to classify this construction with respect to other plurals.

(58) "Uyojroner inte' winik i e tiw" (Pérez Martínez 1996)

ja'xto	*ob'*	*e*	*b'ik'it*	*t'iw*	*a-tza'yob'*	*uw-irob'*
3SG.FOC	PL	DEF	small	eagle	C3-be.happy.PL	A3-see.PL

e	*winik*	*i*	*u-tu'*	*e*	*t'iw, […]*
DEF	man	and	A3-mother	DEF	eagle

'(…) (it was) the little eagles (that) were happy to see the man and the eagle mother (…)'

An interesting parallel is found in Q'eqchi', where Vinogradov (2017: 108) reports that the cognate morpheme *=eb'*, also a "default transcategorial plural marker used both with verbs and nouns" that has no personal reference and only indicates the number, can both precede and antecede its head as in (59) and (60) (Vinogradov 2017: 108).

(59) Plural marker before head (Vinogradov 2017: 108)

eb'	*li*	*q-as*
PL	DET	A1PL-older.brother

'our older brothers'

(60) Plural marker after head (Vinogradov 2017: 108)

li	*r-ochoch=eb'*
DET	A3SG-house=PL

'their houses'

The relationship with the focus position would be clearer if we had examples of the use of *ob'* in front of verbs, as well. As it stands now, we can only speculate how *ob'* came to be used in this way. A parallel to the hypothesis of set C evolving out of independent pronouns in a focus position is the fact that *-ob'/-Vb'*, too, is a form that is considered either an affix or a clitic across Mayan languages, not an independent lexeme. Therefore, when trying to understand whether it initially was suffixed or encliticized to, e.g., the same *h- base as the independent pronouns or whether it could indeed be used on its own, everything discussed in chapter 7.3.3 applies to *ob'*, as well. The HGM form *ha'o(o)b'* seems to contradict the existence of a **ho(o)b', however. As an alternative explanation, one would need to assume degrammaticalization, i.e., the extraction of the morpheme from its position as a suffix and its independent use in an unusual position.

When it comes to the synchronic function of the preposed *ob'* in Ch'orti', no clear conclusion can be reached. A hypothesis is that it might be a new way of marking plural on human referents. The information in Pérez Martínez (1994: 43) seems to imply that in Ch'orti' *tak* is usually used to form plurals of words referring to humans like *ixik-tak* 'women', while the suffix *-ob'* is used elsewhere, e.g., with *chij-ob'* 'horses'. However, *-tak* as a plural suffix does not occur in the legend corpus except for the lexeme *maxtak* 'children, family', which is likely an old collective form that is synchronically sometimes enhanced with an additional *-ob'* as in *maxtakob'* 'children, family' (see Appendix E chapter a.1). Since this lexeme is a known exception and the *-tak* in it can at best be considered a plural marker historically, not synchronically, it seems that *-tak* as a plural marker does not occur in the corpus of legends. Perhaps the cases of *ob'* (mostly with following *e*) + noun are a new way to mark plural on human referents? In fact, in the examples the plural is formed to a human referent in one case ('friends') and to animals otherwise. Though plurals like *e chayob'* 'the fish' are also attested in the corpus, in all cases where *ob'* is found before the noun, the animals share human traits like the ability to speak as in example (57).

Still, human referents also form plurals with *-ob'* in the corpus, among others, even the same *konpágryob'* 'friends (Span. compadres)' in example (56). This Spanish loan word is spelled differently in various parts of the text but is nevertheless the same lexeme. Further research is needed to securely determine the function of this construction. However, since the independent and preposed *ob'* does not occur in Hull (2016) at all, only in the material collected by Pérez Martínez (1996), this development might be a dialectally restricted feature.

7.4.2 Mediopassive class 4

As already briefly discussed in chapter 6.1.4 in the contractions of *o* + *o*, certain thematic suffixes/status suffixes of Ch'orti' show an interesting behavior that differs from the other thematic suffixes. The phenomenon concerns (at least) mediopassive class 4 (MPAS4), which forms mediopassive intransitive verbs with the suffix *-Vy* that usually involve some kind of motion according to Hull:

chamay 'to die, pass away'
ejmay 'to go down, descend'
k'otoy 'to arrive there'
lok'oy 'to leave'
numuy ~ lumuy 'to pass, happen, experience, go through, come to pass'
ochoy 'to enter'
puruy 'to burn (down)'
tob'oy 'to fly, jump; float (up), swim; rise'
t'ab'ay 'to go up, ascend'[349]

These verbs seem to "lose" the suffix *-Vy* in the completive aspect or if other suffixes attach to the root. Thus, *lok'oy* 'MPAS4 to leave' forms a completive *lok'i'x* 'it has already sprouted (i.e., come out)' with *=ix* 'already' and an imperative *lok'i'k* 'he should leave' with optative/imperative suffix *-ik*.

Now, let us contrast the behavior of this type of verbs in set B and set C. I use *lok'oy* whenever the forms are attested in the corpus, otherwise I use other verbs from the same class. The 2PL is, unfortunately, very rare and therefore unattested in my data.

	Set B	Set C
1SG	*lok'-e'n* 'I left'	*in-lok'oy* 'I leave'
2SG	*num-e't* 'you (sg.) passed by'	*i-lok'oy* 'you (sg.) leave'
3SG	*lok'-oy* 'he left'	*a-lok'oy* 'he leaves'
1PL	*lok'-o'n* 'we've been gone'	*ka-lok'oy* 'we leave'

[349] One could add more examples when taking into account the vowel contractions discussed in chapter 6: *ka'y* 'to begin' (< earlier *kajay), *tza'y* 'to be happy' (< *tzajay); perhaps also *ch'o'y-nak'i* 'to growl (the stomach)' (< *ch'ojoy with incorporated *nak'i*), *t'u'y* 'to run down a liquid' (< *t'ujuy). However, an etymological study is needed to confirm this securely.
B'osmay 'to foam up' (and *sak-b'ot'oy* with the same semantics), *je'ray* 'to leak liquid (a wound)', *k'ab'ey* 'it is rumored, said', *k'ajkray* (containing *k'ajk* 'fire') 'to shine' and *tz'ab'ray* 'to shine' are classified by Hull as irregular but may feature the same formation with *-Vy*.

2PL	–	–
3PL	*lok'-o'b'* 'they went (outside)'	*o'choyob'* 'they enter' (*a-ochoy-ob'*)

The base for the forms of set B seems to be a stem without the suffix *-Vy* (*lok'o* or *lok'oj* based on the contraction rules discussed in chapter 6[350]) while the whole paradigm of set C, even C3PL, which does have another suffix, is built upon the form with the suffix *-Vy*. I take this as strong evidence for the fact that set C is in fact based on a third person verb form that was reanalyzed as unmarked for person and paired with a new set of person affixes (independent pronouns in the focus position). This fits the focus construction remarkably well, which I suggest is the origin of set C. It is difficult to decide whether *lok'o* or *lok'oj* is the stem that the forms are based on. Both would yield the following results:

lok'o + en > *lok'e'n*
lok'oj + en > *lok'e'n*

Interestingly, we also find alternative forms for the third person plural systematically in Pérez Martínez (1996) like *ochoy-ob'*, *lok'oy-ob'*, *k'otoy-ob'*, *ekmay-ob'*. This variation does not affect the other persons. As we have seen multiple times, the plural morpheme *-ob'* has a unique position in the verbal paradigm because it is optional and therefore not strictly speaking a person affix but a plural marker that is used freely on both verbs and nouns. It is also homophonous for all index-sets. I suggest that the completive forms attested in Pérez Martínez can be explained by assuming a proportional analogy where the stem of incompletive forms like *a-t'ab'ay-ob'*, *a-numuy-ob'* is *xxx-Vy* with an attached plural morpheme, which was reanalyzed as a stem for the completive, as well, though for now only for the form that is homophonous in both paradigms (3PL *-ob'*):

3SG *num-uy*	:	*a-numuy*	>	*numuy*	:	*a-numuy*
3PL *num-o'b'*	:	*a-numuy-ob'*		***numuy-ob'***	:	*a-numuy-ob'*

I suggest that the same kind of analogy is behind sporadically attested forms like *a-lok'ob'* (Hull 2016), only the other way around:

3SG *lok'-oy*	:	*a-lok'oy*	>	*lok'-oy*	:	*a-lok'oy*
3PL *lok'-o'b'*	:	*a-lok'oy-ob'*		*lok'-o'b'*	:	***a-lok'-o'b'***

In the future, it should be investigated whether other stem formations of Ch'orti' can confirm this theory.

[350] It cannot just be the root because this would not explain the laryngealized vowel in the set B suffixes.

7.4.3 Ch'orti' word order

It has been known for a long time that Ch'orti' is unique among Mayan languages in terms of word order because only there, SV(O) can be found as a basic word order (Quizar 1979; 1994b; Dayley 1983: 102; England 1991). Though focused constituents are still found before the verb, as (61) demonstrates, this does not work the other way around, i.e., the fact that a constituent appears in front of the verb does not mean that it is focused. Instead, this seems to be the most natural word order in the language as in (62):

(61) Ch'orti' A focus (Quizar 2020: 263)

ja'x	*u-wink-ir*	*e*	*tumin*	*k'ani*	*uw-ajp-i-on [uwajpyon]*
3SG	A3-man-POSS	DEF	money	want	A3-grab-TS-A1PL

'it is the owner of the money who wants to grab us'

(62) "Basic" SVO (Hull 2016: 68)

Ne'n	*war*	*in-tziki*	*u-b'e'ykir*	*e*	*pojp.*
1SG	PROG	A1SG-count	A3-smallness	DEF	tule

'I am counting the young tule plants.'

If the focus position lost its focus semantics in the course of the history of Ch'orti', as the grammaticalization of focalized independent pronouns to subject affixes suggests, this might also explain how it was possible for SVO word order to be generalized. If bleaching of the focus semantics of elements in focus position took place, it makes sense that it would also extend to NPs and other elements, not only to independent pronouns. Therefore, I would argue that the worder order of Ch'orti', unique among Mayan languages, in fact can be seen as further evidence for the grammaticalization of the focus construction into set C.

7.4.4 What about Yucatec Maya?

Bricker's investigation of aspect-based split ergativity in Yucatec Maya yielded the result that possessed nominalized complements of aspect markers alone might not be enough to explain how the ergative split arose in the language (see page 79).

One problem she tried to solve was the fact that transitive verbs do not seem to be overtly nominalized. In her investigation of the transitive incompletive status suffix *-ik*, Bricker tried to find a hidden nominalizer in it. She expected it to be there because the intransitive verb also

features a nominalizing suffix (*-el/-Vl*) in the incompletive. I argued in chapter 3.3.1.2 that an overt nominalizer might not have been necessary if transitive verbs were viewed as nouns or if the distinction between nouns and verbs is not as strong in Mayan as we are used to from Indo-European languages. After all, A is marked with set A, which is also used as possessive marking on nouns. Nevertheless, let us examine Bricker's arguments.

Bricker (1981: 87–89) proposed that the transitive status suffix *-ik* is in fact a contracted form of *-il-ak with an intermediate stage of vowel harmony *-il-ik consisting of a nominalizing *-il* and the intransitive subjunctive/irrealis suffix *-ak*. This idea is based on dialectal difference illustrated in (63) and (64):

(63) Hocaba dialect (Bricker 1981: 88)
má'alob' *mentá'ab'-ik*
'it was well made'

(64) Ticul dialect (McClaran (1972: 119, 148) in Bricker (1981: 88))
má'alob' *mentá'ab'-il-ak*
'it was well made'

The question is whether we are dealing with parallel constructions in this case at all or whether the two represent different formations. To prove this, one would need to demonstrate that this contraction of *ilVk* > *ik* happens in other lexemes or grammatical contexts of the language. Otherwise, this is just an ad hoc assumption. More importantly, the forms in (63) and (64) are passive: *mentá'ab'* 'to be made'. As passives are an intransitive formation, these data cannot be used to explain the active transitive suffix *-ik*.[351]

However, her investigation found another interesting pattern:

> Not only can aspect words be replaced by focused temporal and manner adverbs, but they govern the same suffixes as these adverbs, and, because they carry both tense and aspectual information, they combine their functions. This implies either that focused adverbial constructions should be treated as aspectual complement constructions or that what are called 'aspects' in Yucatec Maya are simply focused adverbs or adverbs that have become grammaticalized as aspect particles. (Bricker 1981: 95)

The parallel consists in the fact that an adverb in focused position, i.e., in front of the verb, requires the same kind of construction as an incompletive aspect. Therefore, (66) features the

[351] Instead, we may be dealing with *-Vk* in both cases, which is the dependent status suffix of intransitive verbs according to Hofling (2017: 710). In the Hocaba dialect, the suffix would then be attached to the passive form directly, while in the Ticul dialect, there is an additional suffix *-il*. This suffix is unlikely to be the incompletive intransitive status suffix *-Vl* because this is usually harmonic to the preceding root vowel and also, the incompletive aspect of the passive is formed differently (Hofling 2017: 712).

same kind of intransitive verb form with *-Vl* and set A marking as the incompletive construction with the grammaticalized durative marker *táan* in (65). If an aspect marker is present, adverbs like *sáamal* can only occur after the verb as in (65), they cannot both occur together in front of the aspect marker (Bricker 1981: 95).

(65) Incompletive (durative) aspect (Bricker 1981: 95)

táan *u-k'uch-ul* *sáamal*

PROG A3-arrive-IS/NMLZ tomorrow

'he is arriving tomorrow'

(66) Focused time adverb (Bricker 1981: 95)

sáamal *u-k'uch-ul*

tomorrow A3-arrive-IS/NMLZ

'he will arrive tomorrow'

Bricker concludes that the nominalized suffixes in Yucatec

> have two major functions: They mark the verbs of several kinds of focus constructions […], and they give verbal complements a gerundial meaning.
> Still to be resolved is the question of whether nominalization was the source of the ergative split in Yucatec Maya or whether it was an independent development. (Bricker 1981: 100)

While examining data from the other Yucatecan languages Lacandón, Itzaj and Mopan, as well as from the Cholan branch, she found nominalizations to be "much less pervasive" (Bricker 1981: 101) in the latter specifically because, e.g., transitive verbs do not receive nominalizers in the incompletive in Chol and nominalization is even missing from Ch'orti' incompletive entirely. Thus, her conclusion is that "nominalization was not the source of the ergative split, but developed after the split had begun" (Bricker 1981: 110). Though I would not agree with her diachronic analysis of the transitive suffix *-ik* as argued above and therefore nominalizations are not more "pervasive" in Yucatec Maya, the parallel between aspect markers and focused constituents is significant.

Perhaps the underlying construction could be described as 'it is ongoing$_F$ his arrival tomorrow' for (65) or 'tomorrow$_F$ (is) his arrival' for (66) because aspect markers that derive from verbs like *táan* 'to be ongoing' or *ts'o'ok* 'to finish' and focused adverbs like *sáamal* both seem to require a possessed nominalized complement. However, we are left with an open question as to why a focused adverb as the one in (63) and (64), *má'alob'*, does not result in set A marking and nominalization of the intransitive verb even though the clause structure is parallel ('well$_F$ it was made'). This requires further research.

It was already argued in early research on Mayan ergativity that split ergativity conditioned by focused constituents immediately before the verb (as in the Yucatec examples above) may in fact be the same as split ergativity conditioned by subordination because in both cases, verb forms with dependent status suffixes are used (Larsen & Norman 1979: 355). The authors even argue that tense/aspect-based split ergativity may be a variant of the same process:

> The tenses or aspects which trigger split ergativity are always morphologically marked by the presence of some auxiliary verb or particle preceding the verb. Some of these aspect markers are historically verb roots. This suggests that such constructions are to be analysed diachronically as higher verbs with sentential subjects, that is, as instances of subordination. (Larsen & Norman 1979: 355)

On the one hand, Ch'orti' aspect-based split ergativity fits the pattern well because it arose through a focused constituent before the verb. On the other hand, Ch'orti' cannot be argued to use dependent status suffixes just as it cannot be argued to use nominalized verbs. The verbs are finite synchronically and continue finite forms from Hieroglyphic Maya.

Splits conditioned by clausal dependency are attested in Q'anjob'al, Popti', Akatek, Chuj (all Q'anjob'alan) and Awakatek (Mamean), all spoken in the Huehuetenango area (Zavala Maldonado 2017: 235). Such a split was shown in chapter 2.1.2 in examples (26)–(29). There, in dependent clauses, which are not marked for aspect, S is marked by set A, which leads to nominative-accusative alignment, whereas ergative-absolutive alignment is used elsewhere. Transitive verbs receive a dependent suffix that is cognate to the agent focus marker *-on*. Although Comrie (1978: 377–378), among others, tried to explain this in the same way as aspect-based splits, namely through nominalizations that thus require a possessive prefix, this only works for Awakatek, where a nominalizer is present. The other languages lack an overt nominalizer, which makes the analysis questionable according to Zavala Maldonado (2017: 236). Perhaps taking into account my results for Ch'orti' may help to understand this kind of split diachronically, as well.

7.5 Discussion

In this chapter, I proposed to explain set C as a remnant of the original Cholan independent pronouns in the focus position. Since independent pronouns are based on set B, not set A, the premise is a completely different one than in the theories of Robertson (1998), Wichmann (1999) or Quizar (2023).

A specific property of Mayan made this change possible, namely the fact that in order for the agent to be focused, many languages need to use an intransitivized verb form, usually an

antipassive. As Holtmann (2022) suggests, this agent focus construction originally used an unmarked verb in the third person (B3) which did not agree with either of the semantic arguments of the verb, neither the agent nor the (obliquely expressed or completely deleted) object. Modern Cholan languages are known to have lost the specific agent focus construction but it is attested in Hieroglyphic Maya. I suggest that the focus semantics gradually faded so that the independent pronouns could be reanalyzed as S markers of intransitive verbs instead.

Since intransitive verbs also mark S by set B suffixes, the construction could only develop in a special functional niche. This niche, I suggest, is the incompletive aspect. I have given several typological parallels for the development of incompletive aspect from focus constructions, e.g., in Bantu and other African languages, and discussed three possible functional derivations for Ch'orti', of which the last one is in my view the most plausible. Antipassives per definition delete or demote the object (Harris & Campbell 1995: 245), therefore shifting the focus to the subject of the action. This subject salience, when coupled with the use of the AFC and S focus in thetic statements alongside contrastive focus allowed for its functional reanalysis as incompletive aspect marking. By default, the old, originally aspectless construction then became the completive. Bybee, Perkins and Pagliuca (1994: 90) argue that this is often how the completive/perfective aspect arises – as a kind of leftover from the grammaticalization of an incompletive.

My proposal explains everything that the other theories leave open (see discussion in chapter 5.4):

1. Set A and set C show different behavior when interacting vowel-initial (or *j*-initial) verbs because they are not the same diachronically.

2. There is no need to assume zero-derivation or the loss of a nominalizer because we no longer need to justify the use of set A on intransitive verbs.

3. Contrary to Robertson (1998), Wichmann (1999) and Quizar (2023), I can easily explain why set B developed into two separate, synchronically dissimilar paradigms. A divergent evolution of set B as a suffix vs. as part of independent pronouns is even attested for Mayan (Q'eqchi', see chapter 7.1.1). My explanation does not run into the problem of having to explain how the *a* as it is attested in, e.g., Choltí, was reanalyzed as a third-person marker while still being used with other persons, which are overtly marked, as well – a specific problem of Robertson's approach.

4. The fact that agent focus required the use of a special intransitivized verb (as witnessed in Hieroglyphic Maya) can account for the fact that the innovation of set C only affected intransitive verbs. As a consequence, a morphological aspect distinction only evolved for intransitive verbs.

Additionally, I offer an explanation for every form that deviates from the one in the original set B paradigm. For this, I limit my use of analogies to those cases where I can make explicit every step of the reanalysis that needed to have happened for the analogy to be possible. I argue that C1PL *ka-* is based on a proportional analogy that was enabled by the homophony of A/C1SG *in-*. Since we may assume that C3 *a-* goes back to one of Kaufman's reconstructed deictic elements (near deixis), I propose another analogy to account for C2SG *i-* based on deictic dimensions where, e.g., *a-wayan* 'this one here sleeps' was extended to the second person using a deictic element for farther deixis as in *i-wayan* 'the one there sleeps'.

The only questions that are left open pertain to the fact that HGM initial *h-* is seemingly lost even though this sound is in general retained in Ch'orti'. For this, I offered three possible explanations.

(1) We might be dealing with a highly conditioned sound change of phrase-initial loss of *j-* well-known from (but not limited to) Australian languages.

(2) Ambiguous data from K'iche' offer the possibility that set B could have been used as independent pronouns without being attached to any other base. This option, however, would have strong implications for the relationship of Hieroglyphic Maya and Ch'orti'.

(3) A phase of ditropic clitic behavior where set B was initially enclitic to a base *h-* and then changed to being proclitic to the following verb may have facilitated the development in 1) though this is not necessary.

I currently tend to favor the first option, possibly facilitated by the third, because it is less radical in its implications. However, the second proposal is also a valid choice, though the situation in K'iche' would need to be clarified to pursue this further.

Finally, I also discussed some other phenomena like the peculiar word order of Ch'orti' or open questions of split ergativity in Yucatec Maya, which may be viewed in a different light considering my new proposal. What now remains to be done is to see whether, given the

parallels that exist between Ch'orti' and Cholti, my theory can likewise provide an explanation for the mysterious *a* found in the Morán Manuscript.[352]

7.6 Wald (2007)

The idea presented by me in this chapter has in fact already been proposed some 20 years before me as a sidenote in Wald's (2007) unpublished dissertation. It is not mentioned in any of the publications dealing with set C (e.g., Quizar 2023). Here, I will briefly sketch how Wald's proposal differs from mine.

Just like me, Wald found Robertson's (1998) and Wichmann's (1999) proposals[353] lacking based on the same problems that I have discussed and that have been mentioned in other publications such as Quizar (2023): the complicated chain of changes resembles a "domino effect" (Wald 2007: 913) and requires the assumption of ad hoc sound changes due to metathesis, vowel harmony or homophony avoidance. He further raises the question of a) why certain developments affected Set A when used on intransitive verbs, but not when used on transitive ones (Wald 2007: 902), b) why, if *war* played a significant role in the development of set C, the set likewise occurs without it and c) why the development only affected intransitive verbs when *war* is likewise used with transitive ones (Wald 2007: 913–914).

Like me, he comes to the conclusion that Ch'orti' split ergativity differs fundamentally from split ergativity in other Cholan or Yucatecan languages because it does not involve nominal forms (Wald 2007: 895).[354] He explicitly states that he finds it unlikely that

> the Ch'orti' verb system went through the same stages as Ch'olti'. Ch'olti' preserved the Classic Ch'olan morphological forms in the completive, but then, along with the other two Ch'olan languages, innovated by employing nominals to use as incompletive inflection. It seems highly unlikely that Ch'orti' would have reverted back again to the previous Classic Ch'olan forms after having switched away from them earlier in its history. (Wald 2007: 859).

[352] The following section, in which I discuss Wald's parallel proposal, is a later addition (see Introduction).
[353] Of the two, he also prefers Wichmann's proposal as "it doesn't require an interpretive mistake" (Wald 2007: 902) on behalf of the speakers since "mistaking a temporal adverbial proclitic for an ergative pronominal prefix by native speakers seems highly unlikely" (Wald 2007: 901).
[354] He seems, however, to imply that Ch'orti' cannot feature split ergativity as "ergative pronouns" are not used in the new intransitive incompletive construction. This, of course, goes back to the confusion between "ergative" index-set and ergative-absolutive alignment and is precisely the reason why one should use "set A" instead of "ergative set". As established in detail in chapter 2.1, ergative-absolutive alignment is defined by the **relationship** between the marking of S, A and O. Whether or not **forms** with the (in that case unfortunate) label "ergative" are actually employed does not matter.

He likewise assumes that there is a "lack of distinction between incompletive and completive aspect, not to mention between present and past tense, by verbal morphological means" (Wald 2007: 856) in Hieroglyphic Maya, which means that it is not the loss of aspect marking on transitive verbs that needs to be explained but its genesis on intransitive ones.[355]

Most importantly, he was the first scholar to propose to derive Ch'orti' set C from the paradigm of independent pronouns instead by way of the agent focus construction due to its limitation to intransitive verbs (Wald 2007: 903–909). In his account, the grammaticalization of set C was facilitated by the fact that the original construction enabled speakers to put special emphasis on the S arguments as agents. The functional change of focus to aspect marking is briefly addressed in the following way: "the morphemes recruited for Set C had been used for the same general purpose before the set's development, that is, to stress the agency or dynamism of the intransitive subject" (Wald 2007: 913).

Aside from the fact that Wald's presentation of this idea is rather a brief sketch as it is only a minor part of his thesis, the biggest differences are found in the explanation of the individual indexes of set C. The forms C1SG, C2PL and C3(PL) are unproblematic. For C1PL, he does agree that it comes from set A "by analogy" though he does not demonstrate the actual process (Wald 2007: 912). For C2SG, he provides an explanation that is closer to Robertson's and Wichmann's reasoning, therefore all criticism given in section 5.1.2 applies here, too:

> The original **2nd sg.** *-at* from the independent pronoun *ha'at* would have likely changed to *-i* by analogy along with the 2nd pl. in *-ix,* perhaps at the same time that its original *-ex* changed to *-ix*. In addition, the *-x* of the 2nd sg. was dropped to differentiate from the 2nd pl. (Wald 2007: 912)

As Wald does not elaborate on the nature of the analogy, I can only state that the suggestion is not plausible. An *-x* would also not just disappear without motivation.

It is especially the fact that there is less "necessity to switch among" the person markers that in his view makes the proposal so elegant:

> Only in the case of the 2nd sg. is it necessary to argue from analogy with the 2nd pl. of the same set and then to drop the final consonant. […] Alternatively, the change to /i/ might have been influenced by the /i/ of the 2nd pl. of Set A […]. Otherwise, one form was clearly borrowed from Set A, the 1st pl. pronoun, and it was taken from its exact counterpart in that set. (Wald 2007: 912)

In chapter 5.1.2, I emphasized the necessity to demonstrate the actual contexts where proposed changes could have happened. Wald, unfortunately, as many scholars do, uses "by analogy" as if it were in itself an explanation, which is not the case. It is simply a term for a

[355] His claim that the Ch'orti' incompletive form "is closely related to that of a progressive and likely developed from it into an incompletive" (Wald 2007: 894) is not really clear to me.

language change process that requires an elaborate justification of how forms came to be mixed up and ended up influencing each other. The fact that we are dealing with "the exact counterpart" in C1PL does not matter if the forms are not used in the same context where a reanalysis and influence may occur. The only context he provides is a relative chronology of changes that allows for the emerging set C and set A to get mixed up because they were both used before the predicate:

> A very important advantage of this reconstruction consists in allowing for a possible intermediate stage during which both the *at* of the 2nd sg. and the *on* of the 1st pl. could have easily served as Set C pronouns. The *ex* of the 2nd pl. could have also served well during that time. This would have provided time for the 1st and 2nd pl. of the ergative Set A, *ka*- and *i*- respectively, to influence the original 1st and 2nd pl. of Set C because of their new frontal position and the increased connotations of agency, animism, and dynamism that came from their use in progressive and incompletive constructions. Then, at the same time or later, the /i/ of the 2nd pl. *ix*- would have influenced, by analogy, the original *at*- of the 2nd sg. to be replaced by *i*-. (Wald 2007: 912–913)

Functionally, I would argue that our ideas are compatible but as he barely dedicates any space to the change from focus to aspect, that is perhaps not surprising. Finally, he does not discuss the difference in initial *h*- between Hieroglyphic Maya and Ch'orti' aside from noting that the forms are "a direct match […] provided one allows the /h/ to elide or be dropped as is the case for the whole set when used in this context. Glottal /h/ often elides in morphemes even during the Classic Period." (Wald 2007: 910) Studies of Hieroglyphic Maya orthography have suggested that "underspelling" sometimes happens in the text, i.e., that, especially in cases of "grammatical predictability" (Zender 1999: 139), some "weak" sounds (especially word-final l, m, n), were not always explicitly written by the scribes although they were to be pronounced (Zender 1999: 130–142). A *h* would certainly qualify as weak and is also subject to underspelling.

Though, as Zender points out, there are parallel cases of underspelling in writing systems from the "Old World" (e.g., Mycenaean Greek written in Linear B), I do wonder whether the idea of underspelling can be upheld if we start taking the hieroglyphic texts seriously as historical linguistic sources as Lacadena (2011) demanded and do not just try to match what we read to the closest attested forms that we recognize from attested languages today. In chapter 7.3.3, I offer three other possible theories of how initial *h may have been lost.

Though it does not have immediate implications for the theory discussed here, it is worth mentioning that throughout chapter 9 of his dissertation, Wald points out a number of wrong assumptions and analyses in works by Robertson and colleagues that are worth investigating

in the future. For instance, he proposes a different analysis for CHT, ACN *a* (discussed by me in chapters 5.1.1, 8 and 9) than Robertson (1998) does based on the MM:

> It moves the incompletive or future back closer to the present. It also moves the completive or past forward closer to the present in Chontal […]. This *a-* neither derives present (incompletive) forms from preterites (completives) nor future forms from the present (incompletive) forms. Instead, the effect in Ch'olti' appears to be the same as it is in Chontal, thereby moving the action closer to the present whether forward or backward." (Wald 2007: 898)

Wald further decidedly rejects the analysis of this *a-* as C3 in Ch'olti' as done by, e.g., Fought (1984: 49) and apparently also MacLeod (1987: 27–29) and Bricker (Wald 2007: 898–899, fn. 378), though in the latter case the publication he quotes (Bricker 1987: 23) is not in his bibliography and I am not sure which one he refers to.

As a final sidenote, I want to defend the historical-comparative method and disagree with Wald that the conclusion in Robertson (1998) in any way represents "the best one according to the usual comparative-linguistic standards" (Wald 2007: 858–859). This is based on the observation that it would be more economical to assume that Ch'orti' innovated once than to view the languages that synchronically feature new grammatical verb forms with nominalizations as innovative as this would include a number of parallel independent innovations. As Wald's perspective is that of a Classic Maya philologist, it is understandable that he calls for taking the hieroglyphic data seriously and points out that under his analysis, there is no way but to accept the uneconomical solution. However, by no means does historical-comparative linguistics simply operate on the assumption that the pathway with the least innovations must necessarily be the correct one.

Even though my and Wald's proposals are far from being identical, I hope that the fact that we developed such similar ideas independently from each other will lend credence to the theory itself. What we can all learn from my experience is that maybe we should actually read the dissertations that people spend years of their life working on, even if they remain unpublished.

8 Choltí

Though it seems to be scholarly consensus that Ch'orti' and Choltí are closely related, there is disagreement as to the exact relationship of the two. Kaufman (1976) and Wichmann (2002) see Ch'orti' and Choltí as sisters, not daughter and mother language. Robertson, Law and Haertel (2010: xiii) argue, based on Robertson (1998), that, although the two must have shown at least some dialectal differences in colonial times already, set C can only be understood if one assumes that Ch'orti' went through a stage where the verbal system looked exactly like the one in Choltí – an assessment that is untenable in view of the discussion in chapter 5. Now the theory of Choltí-to-Ch'orti' descent lacks a vital argument, which ultimately has implications for how Hieroglyphic Maya relates to Choltí and Ch'orti' respectively.

In the preceding chapter, I provided an alternative explanation of set C. Since Choltí and Ch'orti' share the restriction of the *a* particle and set C to intransitive verbs, it would, of course, be desirable for the explanation proposed for Ch'orti' to account for the state of things in Choltí, as well. A thorough exploration of this question would involve a complete analysis of the Choltí corpus, which goes beyond the scope of my dissertation. However, I intend to briefly discuss the various possibilities we are presented with.

8.1 Formal derivation from focus

Formally, it is possible to derive Choltí *a* from an earlier *ha' given the fact that <h> is not consistently written in the Morán Manuscript. It is, e.g., generally omitted in consonant clusters (see *kojko* 'TV6 to guard, protect' below in example (7) written as <εoco> or <coco>) and in some lexemes, e.g., *o'* '5'[356], where other Cholan languages retain the Proto-Cholan *j-: CHL *jo'* (Hopkins, Josserand & Cruz Guzmán 2011: 81), ACN <ho> (Smailus 1973: 147).[357]

However, by attempting to explain the origin of the Choltí particle *a* as coming from the focus particle (including the agent focus construction to account for the fact that *a* is only used with intransitive verbs), we are confronted with two problems. First, and most importantly, in

[356] Lemma <ote> according to Robertson, Law and Haertel (2010: 312) or <o. ote. oote> according to Boot (2004: 18). The suffix -<te> is the numeral classifier *-te'* attested in all Cholan languages.

[357] A more thorough future study of Proto-Cholan historical phonology should contain a detailed phonetic investigation and interpretation of Choltí orthography, which was not conducted in Robertson, Law and Haertel (2010).

Choltí, the *a* is not restricted to the third person but used with set B suffixes indicating other persons, e.g., in (1)–(3).

(1) Choltí "presente" II with B1SG *en* (Morán (1695: 32, 55) in Robertson, Law and Haertel (2010: 246))
AI: <auixie˘>, AII: <a Vixi en>
a *b'ixi-en*
? go-B1SG
'I go [Span.: AI: <uoi>, AII: <boi>]'

(2) Future "en -rus" with B1PL *on* (Morán (1695: 32, 55) in Robertson, Law and Haertel (2010: 246))
AI: <auixnel on ti manche>, AII: <a Vixnel on ti manche>
a *b'ix-n-el-on* *ti* *Manche*
? go-IS-NMLZ-B1PL PREP Manche
'we shall go to Manche [Span. AI/AII <emos de ir al manche>]'

(3) Future "en -rus" with B2PL *ox* (Morán (1695: 32, 55) in Robertson, Law and Haertel (2010: 246))
AI: <atzatpael[358] ox>, AII: <atzatzpael ox>
a *sat-pa-el* *ox*
? lose-MPAS-NMLZ B2PL
'you shall get lost [Span. AI: <aueis os de perder>, AII: <os abeis de // perder>]'

It is possible that we are dealing with the same phenomenon that was discussed for Popti' on page 213, example given again in (4). There, *ha'* is a focus marker that needs to be coupled with the new independent pronoun *naj*, which must be an innovation just like many independent pronoun paradigms in Mayan are. The focus marker *ha'* is not restricted to third person. If the use of HGM *ha(a)'* (> CHT *a*) changed in the same way and had become the general marker of focus, this would explain the possibility of it being combined with all persons, not just the third.

[358] Note that in AI, the second *tz* is written <t>. It was not possible for me to confirm whether this is in fact how it is written in the manuscript or whether this is a typo in Robertson, Law and Haertel (2010: 246). In the vocabulary list, the lemma is given as <zata, act^{o}; zatpael, neu^{o}> (Boot 2004: 35) or <zata; de zatpael neu^{o}> (Robertson, Law & Haertel 2010: 338) meaning that the "active (act^{o})", i.e., transitive verb 'to lose' is <zata> and there exists a "neutral (neu^{o})", i.e., intransitive derivation <zatpa(el)>. These are attested in Ch'orti' as *sata* (transitive) and *satpa* (mediopassive) respectively. I am not sure why Robertson, Law and Haertel (2010: 269) found it more plausible to transcribe <tzatz>.

(4) Focused A in Popti' (Craig 1977: 213)

Ha'$_F$ *naj* *x-'il-ni* *ix.*

FOC 3SG.M COM-see-AF 3SG.F

'It was him who saw her.'

If we assume that *a* is a focus-marking particle, we are left with the second problem, namely that it is not used with focused constituents as in Popti' but instead occurs directly before the verb. Though it does occur in questions and relative clauses and we know that the "presente" II expresses gnomic truths, there is no direct correlation in Choltí between sentences containing fronted elements and the presence of *a*. To solve this problem, we need to once again review the possibilities of Mayan languages to express focus.

Aissen (2017b: 298) describes a contrast of "ex situ" and "in situ" focus. Ex situ focus is formed by placing the focused element before the verb and using special morphology in the case of agent focus. It is considered to express a more contrastive kind of focus. This is the kind of focus discussed in chapter 7 as the origin of Ch'orti' set C. However, "[i]n several languages, including at least Tsotsil, Tseltal, and Tojolab'al, a contrastive focus can remain in situ and be flagged by F. In all three languages, F has the form *ja'*" (Aissen 2017b: 300).

With in situ focus, there is no change in word order but a focus particle *ja'* must appear before the verb. I propose that in Choltí, we might be dealing with in situ focus as the origin of the constructions with *a*. The use of in situ focus is attested for Tseltalan and Tojolab'al, which is an unclassified language that is thought to belong either to the Tseltalan or the Q'anjob'alan branch (Campbell 2017: 45). Cholan languages are thought to form a group with Tseltalan, so it is not too bold to assume that Choltí might share some syntactic behavioral properties with Tseltal, Tsotsil and Tojolab'al. Likewise, if we compare the maps in Figure 2 (chapter 2.2.1) and Figure 6 (chapter 2.2.6), we see that Choltí was spoken in the vicinity of the other three so contact may also be taken into account when discussing shared constructions.[359]

Example (5) demonstrates how in situ focus works in Tojolab'al. It is important to point out that the sentence is semantically ambiguous as to what component is focused – the father, the money or the person who gives the money. This ambiguity is also there in Tsotsil and Tseltal. Aissen (2017b: 301) argues that, based on word order conventions in the languages, it is not possible to understand *ja'* as a predicate, which is why she analyzes it as a focus particle.

[359] On the other hand, since the construction in Tseltal, Tsotsil and Tojolab'al on the one hand and Choltí on the other hand is not entirely identical, as we will see below, it is perhaps not necessary to invoke relationship or contact for justifying its postulation in Choltí.

(5) Tojolab'al in situ focus (Aissen 2017b: 301; Curiel Ramírez del Prado 2017: 593)

Ja'	*y-a'-a*	*y-i'*	*tak'in*	*ja=j-tat=i.*
FOC	A3-give-TS	A3-DAT	money	DET=A1-father=TOP

'It was my father to whom he gave the money. / It was money that he gave my father. / It was he who gave the money to my father.'

If we compare (5) to (6), we see that the construction could be analyzed in a parallel way in Choltí: the sentence begins with a particle *a*, which is followed by a verb and then the NPs 'the bread' and 'into the beloved flesh of our Lord Jesus Christ'. The translation given by Robertson, Law and Haertel could equally be changed into 'It is the bread that changes into…' or, since the construction is ambiguous, 'It is the beloved flesh (…) that the bread changes into'.

(6) Choltí in situ focus? (Morán (1695: 63, l. 26) in Robertson, Law and Haertel (2010: 120))

<Aquexpa nepa ti chohuia bactalka Vahauil Jesuχρto>

a-k'ex-pa	*ne*	*pa'*	*ti*	*chohbya*	*baktal-ka*[360]
?-change-MPAS	DEF	bread	PREP	love	flesh-A1PL
w-ahaw-il	*Jesucristo*				
A1PL-lord-POSS	Jesus.Christ				

'The bread changes into the beloved flesh of our Lord Jesus Christ […]' (Robertson, Law & Haertel 2010: 50)

It is difficult to study information structure with resources as limited as they are at our disposal for Choltí. Clear cases of focus in Choltí include those sentences where elements appear before verbs as in (7). However, the parallel to Tojolab'al would suggest that the construction in (6) may have been intended to transport a more focused meaning.

(7) Choltí A focus (Morán (1695: 17, 44) in Robertson, Law and Haertel (2010: 230–231)

AI: <Dios xuεocon et.>, AII: <Dios xucoconet.>

Dios$_F$	*x-u-ko(j)ko-n*[361]	*et*
God	FUT-A3-guard-TS	B2SG

'God will guard you.'

[360] Cf. footnote 113.

[361] The Ch'orti' cognate is *kojko* 'TV6 to guard, protect'. <H> is frequently omitted in the manuscript and it is unclear whether it was not pronounced or just is not written.

With the results from Güldemann discussed in chapter 7.2.2.3, however, we might not even need to push for the semantics to reflect contrastivity. I believe it is not a stretch to argue that a sentence such as "the bread changes into flesh" could be understood as a thetic statement.

A remaining problem concerns the fact that, as (5) demonstrated, the in situ focus construction can focus A without requiring special morphology in Tseltalan and Tojolab'al. In this case, the construction would not explain the fact that *a* only occurs with intransitive verbs in Choltí, which would defeat the whole purpose of this chapter. However, recent research on K'iche' (Velleman 2014) and Yucatec Maya (Verhoeven & Skopeteas 2015) suggests that in these languages, an A that remains in situ cannot be "felicitously interpreted as new information focus" (Aissen 2017b: 297). Examples (8)–(10) demonstrate this for K'iche':

(8) K'iche' in situ S focus, context: Which of them is going to eat? (Velleman 2014: 186)

Aree	*ka-wa'*	*[le*	*al*	*Mari'y]$_F$.*
FOC	INC-eat.B3	DET	CL	María

'[María]$_F$ will eat.'

(9) K'iche' in situ O focus, context: What does María want to eat? (Velleman 2014: 186)

Aree	*k-u-tij*	*[le*	*ichaj]$_F$*	*le*	*al*	*Mari'y.*
FOC	INC-A3SG-eat.B3	DET	vegetable	DET	CL	María

'María will have [the vegetables]$_F$.

(10) **K'iche' in situ A focus, context: Who is going to eat the vegetables? (Velleman 2014: 186)

Aree	*k-u-tij*	*le*	*ichaj*	*[le*	*al*	*Mari'y]$_F$.*
FOC	INC-A3SG-eat.B3	DET	vegetables	DET	CL	María

Intended: '[María]$_F$ will eat the vegetables.'

It is possible that in Choltí the same restriction on in situ focus existed as in K'iche' and Yucatec. In order to focus A, it would then automatically have required an intransitivized verb form. Perhaps Tseltalan and Tojolab'al likewise had this restriction but lost it eventually. This, however, needs to be further investigated in in-depth studies of Mayan historical pragmatics.

8.2 Functional derivation from focus

Functionally, a pathway similar to the one in Ch'orti' is conceivable for Choltí, i.e., the development of incompletive aspectual marking from a construction used for focus and thetic statements by way of subject salience. Alternatively, the agentive explanation sketched in 7.2.2.2 is also appealing: The manuscript itself even gives us a hint that this might be worth investigating: not only does it compare the formation of *a* + verb-*el* + set B to the Latin periphrastic future by calling it "futuro en -rus" (see page 122 and example (11) below), it also describes that *a* can form agent nouns from verbal nouns: "(And) with this *a* by putting it in front are made present participles by putting it on verbal nouns and it refers to the person that does the thing that the noun means."[362] This is demonstrated in (12):

(11) "Futuro en -rus" with B1PL *on* (Morán (1695: 32, 55) in Robertson, Law and Haertel (2010: 246))
AI: <auixnel on ti manche>, AII: <a Vixnel on ti manche>

a	*b'ix-n-el-on*	*ti*	*Manche*
?	go-IS-NMLZ-B1PL	PREP	Manche

'we shall go to Manche [Span. AI/AII <emos de ir al manche>]'

(12) Choltí agent noun with *a* (Morán (1695: 32, 55) in Robertson, Law and Haertel (2010: 247))
AI: <aɛhohbia>, AII: <achohbia>
a-chohb-ya
?-love-VN
'he who loves [Span. <el que ama>]'

An interesting addition to this explanation is made in Arte II which often provides further explanation for the original in Arte I: "although it is more secure to add to this *a* – for this mode of speaking – an *h* always to avoid misunderstandings".[363] The author is referring to the agentive prefix <ah> which is also attested as *aj-* in close relative Ch'orti' and other Mayan languages, e.g., K'iche'. Indeed, most cases of agent nouns in the Choltí glossary seem to

[362] AI: <Y con esta a antepo//niendola se haçen participios de // presente poniendole a los nombres // Verbales y denota la persona que haçe // la cosa que significa el nombre.> (Morán (1695: 32) in Robertson, Law and Haertel (2010: 246))
AII: <[…] con esta. a. // anteponiendola se hasen partisipios de presen//te. poniendola a los nombres Veruales. y denota // la persona que ase la cosa que significa el nom//bre.> (Morán (1695: 55) in Robertson, Law and Haertel (2010: 246)).

[363] AII: <aunq es mas seguro añadirle a esta. a. para este modo de hablar, una. h. siempre para quitar equivocasiones> (Morán (1695: 55) in Robertson, Law and Haertel (2010: 247)).

feature a prefix *ah-* with <h>, regardless of whether it is followed by a vowel or consonant, e.g.:

<ahacut> ‘dancer’ (Robertson, Law & Haertel 2010: 314)
<ahcantezia> ‘he who teaches’ (Robertson, Law & Haertel 2010: 316)
<ahzib> ‘writer’ (Robertson, Law & Haertel 2010: 318)

However, in the texts, it becomes clear that not all agentive forms feature the expected <h>. Examples (13) and (14) feature cases where <h> was not written. Example (14) can be directly contrasted with (15) where the <h> is present on an identical form.

(13) Choltí agentive prefix *aj-* (Morán (1695: 72) in Robertson, Law and Haertel (2010: 136)
<s.ta Ysabel ailia tuba>
santa Isabel *ah-il-ya* *t-u-ba*
Saint.Isabel AGT-see-VN PREP-A3-self
‘[…] Saint Isabel, who bore witness of her [= Holy Mary]’

(14) Choltí agentive (Morán (1695: 71) in Robertson, Law and Haertel (2010: 134)
<xauaulun Dios ticahol atahnalon>
x-aw-awlu-n *Dios* *ti* *ka-hol* *ah-tahnal* *on*
FUT-A2SG-pray-TS God PREP A2PL-head AGT-sin B1PL
‘May you pray to God for the sake of us who are sinners.’ (Robertson, Law & Haertel 2010: 77)

(15) Choltí agentive (Morán (1695: 69) in Robertson, Law and Haertel (2010: 130))
<ah-tahnal winik>
ah-tahnal *winik*
AGT-sin man
‘the sinful men’ (Robertson, Law & Haertel 2010: 66, 200)

This inconsistency of the manuscript led Fought to conclude that most forms featured in the *Arte* in the part where the agentive construction is discussed “appear to be agentives, with or without *h*” (Fought 1984: 49) while he (mistakenly) believed that the forms without *h* in the doctrinal texts constitute cases of a third index-set in Choltí parallel to Ch’orti’.

It is possible that we are in fact dealing with two kinds of “agentive” constructions in Choltí. The first would consist of a noun or verbal noun prefixed with *ah* cognate to the agentive

prefix *aj-* in Ch'orti' and other Mayan languages. The second would have developed out of the focus marker similarly to the development discussed for Ch'orti' in chapter 7.2.2[364] given here again:

CH'ORTI'

Subject focus:	(It is) I (who) overturns/ed the jaguar throne.	
	↓	
Agentive:	I (am the one who) overturns/ed (the jaguar throne).	
	↓	
Set C	I-overturn	(the jaguar throne)
	you-overturn	
	s/he-overturn(s)	
	etc.	

However, instead of all independent pronouns developing into S marking, which was assumed as the final step in Ch'orti', I suggest that the third person was used as a designated focus marker with the possibility of the focused constituent remaining in situ after the verb:

CHOLTÍ

Subject focus:	(It is) I (who) overturns/ed the jaguar throne.
	↓
Agentive:	I (am the one who) overturns/ed (the jaguar throne).
	↓
"Presente" II	I overturn (the jaguar throne).
	(general property, gnomic truth)
"Futuro en -rus"	I (will/shall) overturn (the jaguar throne).
	(future with modal connotations)

The changes that Choltí went through seem to be less rigid – after all, no separate index-set developed. It is therefore not surprising that the *a* is still only a focus particle and occurs in

[364] It is even possible that the two influenced each other or got confused by speakers, since, if the orthography of the manuscript is taken seriously, the *h* in *ah* apparently was weak enough to sometimes be left out from the spelling.

two different constructions, one being an incompletive and one a future. Following the development of the "presente" II construction (*a-b'ixi-en* 'I go'), the general verbal form *b'ixi* with set B suffixes and without *a* came to designate the completive aspect as a leftover, just as in Ch'orti'. On the other hand, the modal meaning with connotations of obligation gives the "futuro en -rus" its own niche as opposed to the general future formed by a *x-* prefix.

8.3 Discussion

The pathways discussed for Cholti and Ch'orti' are not necessarily identical. If we accept the theticity hypothesis for Ch'orti' but the agentivity hypothesis for Cholti, I would even go as far as saying that they are mutually exclusive.[365] Had Ch'orti' gone through a stage where *ha' was used as a general focus marker with all persons, it would not have been possible for it to be reanalyzed as third person S marking. On the other hand, we have no reason to assume that *ha' was ever grammaticalized as a third person marker in a new paradigm in Cholti, even though it ultimately derives from a third person in Hieroglyphic Maya. If we conjecture that Cholti went through a stage similar to that of Ch'orti' only to then degrammaticalize the third person *a* and to use it with all persons, this would be an unnecessary additional assumption, which we should try to avoid. What is certain is that it is possible to derive both from the Hieroglyphic Mayan focus construction (subject + agent focus) – nothing more. At the moment, I would tend to use the theticity explanation to cover both languages for not only because it is an economical and elegant solution if we do not need two different pathways for a phenomenon that shows remarkable similarities but also because it means that we do not need to overly rely on the analyses of the authors of the MM as they are not necessarily trustworthy (see chapter 2.2.6).

It is important to stress that the proposal sketched in this chapter can only constitute that – a sketch. The proposal would need to be substantiated in a detailed study that would need to account for the fact that another – and, one might say, more likely – candidate for the descendant of HGM *ha' exists in Cholti[366], namely the demonstrative particle *ha'i*, which is frequently used with the definite article *ne* or another deictic particle, *ila* (Robertson, Law & Haertel 2010: 196). Functionally, it would be unproblematic to assume that this is identical to

[365] There are many more differences between Ch'orti' and Cholti that need to be explored in future studies and that make the direct descent of one from the other ever more unlikely.

[366] Just as we have a *ja'x* of unknown origin as third person independent pronoun in Ch'orti', see footnote 341.

the focus particle (as these commonly come from deictic elements) were it not for the fact that *h-* is missing in *a* but present in *ha'-*. I currently cannot propose a solution to the problem as precisely the question of the reconstruction of PM, PCH *h and *j is a remaining problem in the historical phonology of Mayan that would need to be investigated first.

In view of this, we may consider the possibility that Choltí *a* is simply a reflex of Kaufman's second position demonstrative *a* used on its own without being encliticized to anything instead. A demonstrative with a meaning such as 'this' or 'here' developing into incompletive aspect marking, perhaps via a stage of progressive aspect, is not at all unusual: Bybee, Perkins and Pagliuca (1994: 128–129) name 'here' as a source of the progressive among the languages in their corpus study and they assume that the progressive is a common source for incompletive aspect and/or future tense.

However, in this case we are essentially left with the same problems that were discussed for Ch'orti' in chapter 7.3.3 – if PM *a is indeed to be reconstructed as a second-position clitic, the loss of the material that it was once attached to, whatever this may have been, would require an explanation. Otherwise, we must assume that Kaufman's clitics were in fact independent elements and could be used at the beginning of clauses on their own. Only a thorough study of the historical phonology of Cholan languages will be able to give us satisfactory and reliable results on that.

Furthermore, if we give up focus marking as a source of Choltí *a*, we must find another explanation to account for the fact that it is only used on intransitive verbs. Perhaps with a more thorough understanding of Mayan information structure in the future, especially from a historical perspective, we may find other ways to explain this restriction along the line.

I tried to present the situation in Choltí involving the particle *a* in as neutral a way as possible and to give ideas for a more detailed future study. In the following chapter, I will do the same for the Western Cholan language Chontal.

9 Chontal

Today, Chontal (Western Cholan) is spoken in the Mexican state of Tabasco by about 37.000 speakers (Osorio May 2016: 19). Prior to the Spanish conquest, Chontal used to also be spoken further east at the rivers Río Pedro Mártir[367] and Río Candelaria in an area called "Acalan" by the Aztecs, which was an important trade hub between Yucatán, Mexico and eastern Péten (Smailus 1973: 3); see Figure 2 (chapter 2.2.1) and Figure 6 (chapter 2.2.6). The speakers of Chontal of this region were displaced to Tixchel on the southern coast of Campeche in 1557 where the language eventually died out[368] (Smailus 1973: 3).

From this region, a single source in the Chontal language survives, a document dated to 1610/1612[369] commonly referred to as the "Maldonado-Paxbolon Papers" (MPP) (Smailus 1973: 4). The text contains a history of the people of Acalan-Tixchel with a focus on the deeds of the ruling family, especially governor Don Pablo Paxbolon and his son-in-law Francisco Maldonado (Smailus 1973: 5). For the complex genesis of the manuscript see Smailus (1973: 4–7) as well as Scholes and Roys (1968: 359–366). The latter in general provides a detailed ethnographic treatment of the Chontal Maya.

Since the Acalan variant of Chontal died out, Acalan and Tabasco Chontal can be considered dialectal variants of each other but not mother and daughter. A particle *a*, possibly with aspectual meaning, is attested in both variants. I will first discuss the older attested variant and then proceed to modern Chontal.

9.1 Colonial (Acalan) Chontal

A facsimile of the MPP with Spanish and English translation is found in Scholes and Roys (1968). Furthermore, Smailus (1973) offers a Spanish and German translation with morphemic segmentation, a sketch of the Acalan Chontal grammar and a word index with the locations where they are found in the document. The Spanish translation is the same in both publications and already forms part of the MPP. The particle *a* is used abundantly in this text.

[367] Also called Río San Pedro.

[368] Smailus reports a high influx of speakers of Yucatec Maya into the region with the Chontal only making up 10% of the population in 1688. Chontal disappeared "in the course of the following centuries [German: "im Verlauf der folgenden Jahrhunderte"]" (Smailus 1973: 3).

[369] Hruby and Child (2004) also claim to have found traces of Chontal specifically in the hieroglyphic corpus. Aside from that, the document from 1610/12 is the earliest attestation of Chontal.

To facilitate the understanding of the examples, Table 31 offers an overview of Chontal set A and B indexes:

Table 31. Acalan (Smailus 1973: 193–194) and Tabasco (Osorio May 2016: 34) Chontal index-sets.

	Set A				Set B	
	Acalan		Tabasco		Acalan	Tabasco
	_C	_V	_C	_V		
1[370]	<ca>-[371]	<c>-	*kë-*	*k-*	<-on>	*-on*
2	<a>-	<au>-	*a-*	*aw-*	<-et>	*-et*
3	<u>-	<(u)y>-	*u-*	*(u)y-*	ø	ø

Examples (1) and (2) contrast the same verb, once used with *a* and once without. The semantic difference does not become immediately obvious and is often described along the lines of: "[*a*] is attached to the completive to give a sense of 'already', or even present perfect" (Robertson, Houston & Stuart 2004: 263) or "[a]n action is considered completed from the temporal point of view of the context"[372] (Smailus 1973: 132). For the examples at hand, this analysis fits because in (1), a sequence of events is related while in (2), the arrival of the license is expressed as a condition for the rest of the action described in the sentence ("And thus as the aforementioned license arrived, Pablo, our governor, indicated it to all villages."[373]).

[370] The Proto-Mayan distinction between singular and plural pronouns does not exist in Western Cholan anymore. Instead, the plural form was generalized for the first person and the singular form for the second. Plural is marked via special enclitics. They are the same for set A and set B: *=la* for the first (inclusive) and second person plural, *=doko'b* for the exclusive first person plural, *jo'b ~ o'b* for the third person plural (Osorio May 2016: 34). Plural forms attested for Acalan Chontal (Smailus 1973: 194) are <la> for the first and second person and <ob> for the third with two instances of a special morpheme <tac>, once with an additional <ob>, i.e., <tacob>. In both cases, the plural is formed to human referents.

[371] A1 can also be <c> if it is preceded by a vowel: <ca-cab> 'my land' but <ta-c-cab> 'in my land' (Smailus 1973: 193). This is reminiscent of Lehmann's analysis of set A as enclitic to the word that precedes the verb instead of proclitic/prefix to the verb (see footnote 110).

[372] German: "Eine Handlung gilt vom Zeitstandpunkt des Kontextes her als abgeschlossen."

[373] Spanish: "Y así como llegó la dicha licencia, Don Pablo, nuestro gobernador, lo dió a entender a todos los pueblos" (Smailus 1989: 114).

(1) Acalan Chontal intransitive <huli> without *a* (MPP (1610: 167) in Scholes and Roys (1968)[374], analysis based on Smailus (1973: 101))

<lahun yuxkal ti soltadosob huli // viticah>

lahun	*y-ux-kal*	*ti*	*soltados-ob*[375]	*hul-i*	*wi*	*ti*	*kah*
10	A3-3-20[376]	PREP	soldier-PL	arrive-IS.B3	here	PREP	village

'50 soldiers arrived here to the village (of Tixchel)'

(2) Acalan Chontal intransitive <huli> with *a* (MPP (1610: 169) in Scholes and Roys (1968), analysis based on Smailus (1973: 114))

<bache ahuli liçensia>

bache	*a-hul-i*	*licensia*
how	?-arrive-IS.B3	license

'as the license arrived'

It is significant that *a* seems to express a completed action because this is the exact opposite of the *a* in Choltí from a functional perspective. On the other hand, the use with other persons than the third is parallel to Choltí. Example (3) shows the use of ACN *a* with a first person. Smailus (1973: 132) only mentions the use of *a* with finite intransitive verbs in the first and third person. The reason why it does not occur in the second person may be due to the limited corpus.

(3) Acalan Chontal *a* with first person (MPP (1610: 160) in Scholes and Roys (1968), analysis based on Smailus (1973: 49))

<uixachmeahulon tacab>

wi	*xach=me*	*a-hul-on*	*t-a-kab*
here	now	?-arrive-B1	PREP-A2-land

'Here now I have arrived to your lands.'

Although Smailus states that *a* is only used with intransitive verbs, there are many instances in the corpus, e.g., example (4), that demonstrate that it can be employed with transitive verbs, as well. This is a significant difference to Choltí and Ch'orti'.

[374] The facsimile is found on unnumbered pages in between page 366 and 367.

[375] Synchronically, Chontal *b* is written <b> without implosion even though there is a great deal of dialectal variation in terms of its actual pronunciation, which sometimes includes a "superimposed glottal quality" (Knowles 1984: 43). *B'* is also usually simply <b> in colonial sources.

[376] Mayan languages have vigesimal counting systems. The sequence '10 his 3-20' is to be understood as 'ten missing from three twenties (i.e., 60)'.

(4) Acalan Chontal *a* with transitive verb (MPP (1610: 159) in Scholes and Roys (1968), analysis based on Smailus (1973: 43))

<chan tzuculcab[377] acathane>

chan *tzuk-ul kab* *a-ka-t'an-e*

four village.quarter ?-A1-say.B3-TS

'(the) four village quarters that I named'

Similarly to Cholti, some confusion seems to exist between the aspectual <a> and the agentive prefix <ah>. Smailus (1973: 132, 210–211) describes an orthographic variation of <a> ~ <ah> though he states that the latter is less frequent and, since it is used with Yucatec toponyms, it may be a loan altogether. This would mean that the Chontal aspectual <a> and the agentive <ah> would have merged at least orthographically and possibly also phonologically. Though it seems that in Cholti as well as in Acalan Chontal preconsonantal <h> is often not written at all (Kaufman & Norman 1984: 86), we, unfortunately, lack a substantial study of the historical phonology of Chontal to determine what this information means. Cognates like CHN *sami* ~ CHR *sajmi* 'today' or ACN, CHN *ak'-e'* ~ CHR *ajk'-u* 'to give' suggest that *j* indeed disappears before other consonants. However, the prefix *aj-* remains *aj-* in modern Chontal (Keller & Luciano G. 1997: 4–5) and only alternates with *a'-* before vowels (e.g., in rapid speech), not before consonants: *aj-mis* 'cat'[378], *aj-tz'ib* 'writer, scribe'.[379]

According to Smailus, <a> has agentive semantics when used with nouns or verbal nouns and aspectual semantics (as described above) when used with finite verbs. Therefore, though example (5) seems to at first glance represent an instance of *a* used with <hul> 'to arrive' in the incompletive aspect (formed with <-el> as in close relative Chol and modern Chontal, see chapter 3.3.2), Smailus rather analyzes it as an agentive 'those who came'.

[377] The final letter looks very much like a <b> in the manuscript. However, Smailus (1973: 137) reports <tzucul cah>, not <cab> as a term referring to the four village parts. Perhaps this is a mistake on the scribe's part.

[378] See footnote 343 on the use of *aj-* and *ix-* for animal names.

[379] Note that the same applies to *aj-* in Ch'orti': there, too, *aj-* keeps the *j* even though it is often found in intervocalic position (see chapter 6.2).

(5) Acalan Chontal *a* with incompletive aspect (MPP (1610: 168) in Scholes and Roys (1968), analysis based on Smailus (1973: 105))

<tali yubi don pablo uthan ahulelob>

tal-i	*y-ub-i*	*don pablo*	*u-t'an*	*a-hul-el-ob*
come-IS.B3	A3-hear.B3-TS	Don.Pablo	A3-word	AGT-arrive-NMLZ-PL

'And so Don Pablo heard the word of those who came […]'

On the other hand, *a* in example (4) above cannot reflect agentive <ah> instead of aspectual <a> with a transitive verb because 1) the verb is finite and 2) even if one may argue that transitive verbs have some kind of nominal character because of their use with the possessive prefix set to mark A, <ah> always attaches directly to the stem and would be expected to yield **ka-a-t'an-e[380] at best, not the attested *akat'ane.*

In the following chapter I will describe the use of *a* in modern Tabasco Chontal.

9.2 Modern (Tabasco) Chontal

In modern Chontal, *a* can be used both with intransitive (6) and transitive (8) verbs and both with completive (6) and incompletive aspect (7) as the following examples show. Vinogradov analyzes it as a kind of "'proximative', referring not only to the temporal proximity, but also to the discursive relevancy of events" (Vinogradov 2018: 276).

(6) Chontal *a* with completive (Vinogradov 2018: 276)

ni	*chimaya*	*a*	*num-i*	*äk'bi*
DET	deer	PROX	pass.by-COM.B3	yesterday

'The deer passed by yesterday.'

(7) Chontal *a* with incompletive (Keller & Luciano G. 1997: 451; Vinogradov 2018: 276)

a	*kä=k'ux-e'*[381]
PROX	1A=eat.B3-INC

'I am going to eat it.'

A may co-occur with other "peripheral prepositive particles" like *sam* in (8).

[380] This kind of form is, to my knowledge, not attested anywhere across Mayan languages.

[381] I follow Vinogradov's segmentation here but note that Keller and Luciano segment it <acä c'uxe'> with *kä* being enclitic/suffixed to the preceding *a* instead of proclitic/prefixed to *k'uxe'*. Note that Lehmann also proposes this as the correct segmentation for Yucatec Maya (see fn. 110).

(8) Chontal proximative with aspect particle (Vinogradov 2018: 276)

sam *a* *kä=tsäm-s-i=t'oko'*

PRF PROX A1=die-CAUS-COM.B3=EXCL

'We have already killed it.'

Compared to Acalan Chontal, *a* has a broader functional range: it does not refer to recent past anymore but can apparently also describe imminent actions. This difference could either be due to diachronic development, dialectal difference or a misanalysis by Smailus, possibly due to the corpus size. Most labels that *a* has been given in the literature that are discussed by Vinogradov share the assumption that *a* conveys a kind of immediacy or relevance for the present; some emphasize a perfect meaning. An especially fitting functional parallel is found in the Spanish *ya* 'already', which can also account for the fact that *a* can be used both in incompletive and completive aspect. It is possible that *ya* has influenced CHN *a* functionally; however, it is equally possible that its non-perfect uses are either unattested in the Acalan Chontal corpus or were not recognized by Smailus as stated above. Since he also seems to have overlooked the use of *a* with transitive verbs, a new corpus study with a focus on the use of *a* seems to be desirable for the future.

It is interesting that *a* has a special restriction in that it cannot be used before A2/3 (*a-/u-*). This restriction is purely phonological according to Vinogradov because using *a* in front of vocalic indexes is "impossible in order to avoid hiatus" (2018: 276). Already for Acalan Chontal, Smailus (1973: 196) had remarked upon the fact that aspectual *a* is substituted by *t-* before the third person set A prefix, which is *u-*. Like Vinogradov, he suspects the reason behind this to be phonological but considers the material too little to say anything further on the matter. At least for Acalan Chontal, it is likely that *a* and *t-* only form a paradigm synchronically but have different origins. Therefore, the variation is unlikely to be purely phonological.

9.3 Discussion

A lot more research is needed to understand what we are dealing with in the case of Chontal *a* both synchronically in terms of its function and diachronically in terms of its origin. Based on the hypothesis proposed for Ch'orti', one might expect that the presence or absence of *a* correlates with focus or thetic semantics, the presence of relative clauses, negation or question verbs or other fronted material. However, just like in Choltí, this cannot be conclusively

shown to be the case with the studies available to me now. For example, in (9), there is fronted matter before the verb but *a* is not used.

(9) Chontal <huli> with fronted <yuual> (MPP (1610: 155) in Scholes and Roys (1968), analysis based on Smailus (1973: 20))
<hainchanpeli hainxach ahauyu//ual huli>
hain chanpeli hain xach ahau yuual hul-i
DEM Chanpel DEM now king then arrive-IS.B3
'This Chanpel, this (one) now (was) king when he then arrived (to conquer Tatenam, which is today called Términos).'

It should be pointed out that, while most authors (e.g., Robertson 1998: 7; Houston, Robertson & Stuart 2000b: 350; Robertson, Houston & Stuart 2004: 263) consider the Acalan Chontal *a* to be a cognate to the Choltí *a*, Mora-Marín (2021: 31) claims that they are unrelated. According to Quizar (2023: 260, fn. 2), he instead believes that the Chontal morpheme goes back to existential *ayan* 'there is, there are', which is known from Greater Tseltalan and occurs in Eastern Cholan as well. It is unclear how this would work since *ayan* is attested in Acalan Chontal alongside *a*. For lack of reasons to consider the opposite, I agree with Robertson and colleagues that the *a* is cognate in both languages, though it is always risky to try to establish an etymology or cognacy for elements that consist of only one single sound.

Just as in Choltí, the Chontal *a* likely is the deictic element expressing 'this' or 'here' reconstructed by Kaufman as *a (see chapters 7.3.2 and 7.3.3). Therefore, just like Choltí *a*, it either was suffixed to an element (possibly *h-?), in which case we need to explain how this element was lost, or it was used independently, in which case we need to reject Kaufman's reconstruction of it as a second-position clitic.[382] Since it is not the case in Chontal that *a* is restricted to intransitive verbs (contrary to the statement by Smailus as discussed above), we are free to assume that Chontal simply grammaticalized a deictic element to express recent past. This as well as the functional difference of incompletive and future in Choltí versus recent past/perfect in Chontal suggests that the constructions developed separately from each other, i.e., the languages were not necessarily in contact while this happened. This scenario would be able to account for why the cognate elements become grammaticalized in slightly different ways.

[382] Here, too, we need to take into account the demonstrative form <hain> which seems to consist of HGM *ha'* and an additional element *-in* (see discussion for Choltí in chapter 8.3).

Bybee, Perkins and Pagliuca's (1994: 55) data also yield an example of a construction consisting of a proximal demonstrative with copula developing into an anterior aspect (Tahitian). Based on the authors' persuasion that the lexical source of a construction "uniquely determines the path that grammaticalization follows and, consequently, the resulting grammatical meanings" (Bybee, Perkins & Pagliuca 1994: 9), it may be worthwhile to study in which ways the constructions in Choltí or Chontal differ that may account for the different pathways taken in the grammaticalization.

On the other hand, if we want to explain Chontal from the construction proposed for Choltí, this, too, is possible. As soon as *a* got grammaticalized as an aspectual marker, it is easy to imagine that the restriction to intransitive verbs is lifted and speakers extend its use to transitive verbs. However, I believe that a new corpus study of the MPP will yield a much clearer picture. Therefore, I refrain from further analyses now.

For the sake of completeness, I will discuss the use of an aspectual *a* in one more language, Poqomchi' (Greater K'iche'an), in the following chapter.

10 Poqomchi’

As already mentioned in Houston, Robertson and Stuart (2000b: 350), a suspicious *a* morpheme is also present in a Greater K’iche’an language. According to the authors, Poqomam, “being the only K’iche’an language that possesses this element, likely borrowed it from Ch’olan” (Houston, Robertson & Stuart 2000b: 350). Since they provide no further information, it is unclear whether they really meant Poqomam and not Poqomchi’. In any case, the phenomenon is discussed by Vinogradov (2022) for Poqomchi’.

In colonial Poqomchi’, future was formed with a prefix *a-* as demonstrated in (1). As described in Vinogradov, a process similar to the hypoanalysis described in Haspelmath (1998) for old presents, which may evolve into futures because they are restricted in their use by new present formations, took place in the language because a different construction came to denote future in modern Poqomchi’ restricting the old *a* future to the domain of the optative as seen in (2).

(1) Colonial Poqomchi’ future tense (Zuniga 1614: f. 314v; Vinogradov 2022: 385)
<quijb ixib nucoral ancor aue>

ki’-ib’	*ix-ib’*	*nu-q’or-al*	*a-n-q’or*	*aw-eh*
two-NUM	three-NUM	A1SG-word-ABST	FUT-A1SG-say.B3	A2SG-DAT

‘I will tell you two or three words.’

(2) Modern Poqomchi’ optative (Vinogradov 2022: 385)

a-nw-is-aj	*peet*	*cho*	*ni-timiin*	*chi*	*ru-paam*
OPT-A1SG-take.out.B3-TS	first	DIRL	A1SG-money	PREP	A3SG-in
niw-ihq					
A1SG-load					

‘First let me take the money out of my load.’

In Colonial Poqomchi’, *a* could be combined with incompletive TAM markers and in the case of intransitive verbs also with the irrealis status suffix *-oq* (Vinogradov 2022: 385). This resembles the broad use of CHN, CHT *a*. As to the diachrony of the construction, Vinogradov simply states: “The prefix *a-* is an innovation in Poqomchi’. As pointed out by an anonymous reviewer, this marker was apparently borrowed from Ch’olti’, a lowland Mayan language.” (Vinogradov 2022: 386).

Unfortunately, Poqomchi' has been studied a lot less diachronically than other Mayan languages, which means that foundational studies on many aspects of it need to be conducted before questions such as the one discussed in this chapter can be addressed. There is little information on the historical phonology of Poqomchi', the diachrony of its alignment split (see chapter 3.3.3) or its information structure. What can be said based on the limited information available at the moment is that the independent pronouns of Poqomchi' displayed in Table 32 are strikingly similar to what is either attested or reconstructed for Hieroglyphic Mayan (and reconstructed for Proto-Mayan by Kaufman): *hin* and *hat* correspond to the Hieroglyphic Mayan forms while *hoj* is the Greater K'iche'an counterpart to what we would expect to find if an independent 1PL form were attested in the hieroglyphic corpus (*hon). *Hat taq* seems to be the second person singular with an added plural *taq*. As to the third person, we do not find forms akin to HGM *ha'* in Poqomchi' independent pronoun paradigm. Instead, the forms consist of a definite article/demonstrative/focus particle *re'* and a so-called "relational noun" *-eh* 'for' with prefixed set A (Mó Isém 2006: 135–138, 278), a type of possessed substantive that is a common way in Mayan languages to form prepositions.

Table 32. Poqomchi' pronouns after Mó Isém (2006: 68).

	Independent pronouns
1sg	*hin*
2sg	*hat*
3sg	*re' reh*
1pl	*hoj*
2pl	*hat taq*
3pl	*re' keh*

The fact that *ha'* is not found in the paradigm nor, apparently, anywhere else in the language, at least according to Kaufman (2003: 1534–1535), offers the possibility that it might have become the future marker *a-* seen in Colonial Poqomchi'. However, this analysis is not strictly necessary:

1. According to Campbell (1977: 36), Poqomchi' retains Proto-Mayan *h in all cases except before *i* where it apparently becomes *yi*. This is also witnessed by the other forms in the independent pronoun paradigm. Therefore, we would not expect an earlier **ha to have lost the initial **h-. In this case, we would need to assume that this is the deictic element *a* as an independent form, not as a clitic.

2. As stated for Chontal in chapter 9.3, there is nothing about the construction in Poqomchi' that would require it to be derived from a focus construction – the construction is not restricted to intransitive verbs as example (1) demonstrates. We an therefore imagine a wealth of different pathways which we will only be able to narrow down once the prehistory of Greater K'iche'an is sufficiently reconstructed.

This leaves us with two possibilities for now:

1. Poqomchi' developed this future construction independently, either based on the deictic element *a* (which would prove that Kaufman's reconstruction of PM *=a, *=i etc. as clitics is wrong) or an *a* of unknown origin. As discussed in chapter 8.3, it is entirely possible for a deictic element to eventually turn into future marking.

2. Poqomchi' developed the future construction under the influence of a language that used *a* in a similar way, possibly Cholti. Additional research into the history of Poqomchi', Greater K'iche'an and colonial contact of it to other languages is needed to further investigate the genesis of this form. I find it unlikely that *a* was simply borrowed as a morpheme without the according constructions also being similar. Borrowing of morphemes is often and, I would argue, sometimes too willingly, assumed in Mayan linguistics (see chapter 2.2.3). If contact is in fact the origin of this construction, as argued in Vinogradov (2022), clerics devoted to evangelization and their command of the various indigenous languages may have played a role in this.

The formation discussed in this chapter likely has very little in common with the origin of set C in Ch'orti' except for, possibly, the shared etymology of PM *a 'this; here'. If this origin can be confirmed in both cases via further historical study, a future investigation may address the question of how the element came to be grammaticalized in these different ways and what parts of the involved constructions ultimately determined the respective outcomes.

11 Hieroglyphic Maya

The first attestation of a hieroglyphic script demonstrably recording a Mayan language was found at San Bartolo in Petén, Guatemala and is dated to the so-called Late Preclassic period in the third century BCE (Saturno, Stuart & Beltran 2006). Featuring a rapid decline in use after the beginning of the Spanish conquest, its large-scale use likely ceased entirely with the conquest of the Itzaj capital Tayasal in the last independent Mayan kingdom at the end of the 17th century (Lacadena 2011: 343). Most of the texts are found throughout the Lowlands of the Yucatecan peninsula reaching into the highlands of what is today Chiapas (Mexico), Guatemala and Honduras (Lacadena 2011: 343). It is still not entirely clear whether the hieroglyphic texts represent a single language or multiple. According to Lacadena (2011: 344), while traces of at least five different languages can be found in the hieroglyphic script, one of them, the one classified as Eastern Cholan and hypothesized to be the ancestor of Cholti and Ch'orti', seems to be the most prominent. The other languages are:

1. a form of Western Cholan (probably the ancestor of colonial and modern Chontal),
2. a form of Yucatecan (the ancestor of the four modern Yucatecan languages)
3. a form of Tseltalan (probably the ancestor of colonial and modern Tseltal) and
4. a language that must belong to the Highland Mayan languages but cannot yet be further specified, although it is probably related to Greater K'iche'an (Beliaev 2005).

Law and Stuart (2017: 133) hold a slightly different view:

> "The language recorded in the Classic script displays some regional and temporal variation. It is clear, however, that the corpus of texts, with a couple of possible isolated exceptions, records a single, remarkably uniform language."

Thanks to many decades of work on its decipherment, most of the more than 15.000 texts can be understood today, though many details remain to be worked out in terms of the actual orthography and phonological detail (Lacadena 2011: 343–344). Although one would expect a language to change drastically within almost 2000 years, Hieroglyphic Maya remains surprisingly consistent and is thought to be a highly formalized and archaic prestige language (Law & Stuart 2017: 128).

> The extent and nature of the Cholan (and Cholan-Tzeltalan) loan words found in other Mayan languages, both neighboring and non-neighboring, attest the extreme importance of Cholan(-Tzeltalan) in the formation of Maya civilization […]. [There are] a few cases of loan words into Yucatecan which are either attested hieroglyphically (in the codices), or have ritual/calendrical content, or both. Culturally comparable loans from Yucatecan into Cholan are not found. (Campbell 1984: 7)

Phonological evidence points to the fact that this single prestige language must belong to the Cholan-Tseltalan branch[383], since, e.g., it shows <ch> as a reflex of PM *k (e.g., <chi-ji> *chij* 'deer' < PM *keej based on YUC *kéej*, KCH *keej*, TSE *chij*, CHL *chijmay*, CHR *chij*; though see Table 2 and discussion of the evidence in chapter 2.2.5) – a shared innovation that is said to only affect languages of this branch (Law & Stuart 2017: 129–130). The lexeme *chij* also shows the second criterion for classifying the language as Cholan, namely the raising of PM *ee to PCH *i (see also introduction to chapter 2.2.5). The word for bat, HGM <su-tz'i> *suutz* (Kettunen & Helmke 2020: 116), demonstrates that the rounded counterpart of this sound change, PM *oo > PCH *u, also applies to Hieroglyphic Maya. Furthermore, Law and Stuart (2017: 130) cite evidence from the domain of inflectional verbal morphology for the fact that Ch'orti' (and its potential ancestor Choltí) are the closest relatives of the language of the inscriptions. Since Robertson's proposed explanation of set C is a significant part of this, it cannot be regarded as evidence anymore after the discussion in chapter 5.

In this chapter, I would like to briefly discuss what my proposal for the origin of set C in Ch'orti' means for the question of alignment in Hieroglyphic Maya as this is still a controversial topic. Chapter 11.1 deals with general views on alignment in the language, chapter 11.2 discusses presumed traces of split ergativity in the hieroglyphic corpus and 11.3 offers a discussion with a summary of the results.

11.1 Alignment in Hieroglyphic Maya

The question of whether alignment in Hieroglyphic Maya is "straight"- or "split"-ergative is unresolved to this day. Law, Robertson and Houston (2006: 444) argued that split ergativity can neither be reconstructed for Proto-Cholan nor is it present in Hieroglyphic Maya texts written mainly in a Cholan language because the alignment systems in the daughter languages are too different to be reconciled in a shared variant of split ergativity in the proto language. Lacadena (2011: 360) seems to consider the matter settled, apparently precisely because of the authors' argumentation.

Though I do not agree with all the details of the analyses presented by the authors[384], I agree with them on this point. I believe that my treatment of Ch'orti' has shown without doubt that

[383] Different views on the question of whether it is specifically Eastern Cholan, i.e., the exclusive ancestor of Choltí and Ch'orti', or closer to Proto-Cholan were already briefly presented in chapter 2.2.1.

[384] This cannot be discussed here in detail but their reconstruction of syntactic processes like "raising" without having any cognate substance to work with (e.g., constructions or at least cognate lexemes or

it has even less in common with aspect-based split-ergative alignment in Western Cholan than was believed until this point. Kaufman and Norman (1984) reconstructed Proto-Cholan as split-ergative but this cannot be upheld. On the one hand, they lacked some of the data that have become available today, on the other, some of their analyses in general require a reevaluation.

It is worth investigating how the belief that Hieroglyphic Maya may have had split-ergative alignment came about at all – because it is, I would argue, ultimately based on a series of misconceptions from earlier periods of reconstruction and glyphic research. In her treatment of Hieroglyphic Mayan verbs, Schele (1982: 9) reported that

> "Victoria Bricker […] has proposed that the Classic writing system reflects not an ergative morphology, but rather that of a split-ergative system, and Barbara MacLeod and Will Norman (personal communication, 1980) have both detected split-ergative morphology in the Cholan languages."

Bricker (1985) even made a prognosis of what the indexical glyphic pattern would look like if the language were split-ergative including the use of set A on intransitive verbs. Schele admits that this prognosis "does not precisely match" the distribution of verbal affixes in the glyphic corpus, however, "it explains major anomalies in which clearly intransitive verbs (i.e., the auxiliary verb T757) appear consistently with a third-person ergative pronoun" (Schele 1982: 10). She concludes: "The Classic writing system may well reflect a split-ergative language."

Unfortunately, the sign that Schele refers to as an example, T757, has by now been identified to be the glyph for *-b'ah* (Montgomery 2002: 40), an obligatorily possessed noun meaning 'image, self' (Kettunen & Helmke 2020: 99), which explains its common use with set A

---

morphemes) is highly problematic since they also do not use a specific framework that would at least try to put such a reconstruction on methodologically sound grounds.

Schweitzer (2006: 54) likewise criticized precisely this kind of argumentation, as already partly noted in footnote 55: "What is particularly problematic with many of Robertson's arguments is that he places too much emphasis on the theoretical postulation of systems and filling in gaps, while using the instrument of analogy in a very free manner. While Robertson still reconstructs pure ergativity for Proto-Mayan in his work from 1992, Robertson, Houston and Stuart (2004) assume a split for Proto-Mayan that shows a nominative-accusative behavior in the progressive but not in the incompletive; this, however, only in a syntactic and not morphological way, which is typologically similar to the progressive formations in the K'iche'an languages, but cannot be directly reconstructed." [German: "Problematisch bei vielen Argumentationen von Robertson ist insbesondere, daß er zu sehr das theoretische Postulieren von Systemen und das Ausfüllen von Lücken in den Vordergrund stellt und dabei das Instrument der Analogie in sehr freier Weise einsetzt. Während Robertson in seinem Werk von 1992 noch reine Ergativität für das Proto-Maya rekonstruiert, setzen Robertson, Houston und Stuart (2004) für das Proto-Maya einen Split an, der im Progressiv, nicht aber im Inkompletiv, ein nominativisch-akkusativisches Verhalten zeigt, allerdings nur auf eine syntaktische und nicht morphologische Weise, die den Progressivbildungen in den k'iche'ischen Sprachen typologisch ähnelt, aber nicht direkt rekonstruiert werden kann."

prefixes. A reexamination of the other cases of "clearly intransitive verbs" may likewise reveal them to be nominal. However, these would need to be collected first because Schele only gives one example (the abovementioned T757). As to the evidence found by MacLeod and Norman, this cannot be judged since it is cited from "personal communication".

A short chapter on the subject of split ergativity in modern Cholan and Yucatecan languages written by MacLeod (1982) can be found in the appendix of Schele (1982). This description, unfortunately, is based on severe misunderstandings both of Ch'orti' morphology specifically and of what constitutes an alignment split in general. For example, she lists possessed verbal nouns like *uw-irna'r* 'their being seen' as evidence for split ergativity (MacLeod 1982: 421) without paying attention to whether they are actually integrated into the verbal domain or simply used as nouns. A comparable situation is the parallel light verb construction with *che* that exists both in Western Cholan (see chapter 3.4.2) and in Ch'orti' today – however, in Ch'orti', the construction does not lead to split-S/fluid-S alignment (cf. Appendix E chapter h.6). The presence of a cognate form or construction is not enough evidence for a specific alignment. Based on this problematic understanding of the modern languages, one may argue that it would have been difficult for the authors to evaluate the situation that presented itself to them in the hieroglyphic texts correctly.

Fourty years have passed since then and our knowledge of both the Maya script and alignment theory has grown substantially since the 80s. Probably most scholars would agree at this point that we can simply leave the idea of split ergativity in Hieroglyphic Maya in the past. Nevertheless, I want to address a couple of texts that are often named as attesting to split ergativity in Hieroglyphic Maya in the following section.

11.2 Traces of split ergativity?

There are some examples that are often referred to as (potential) evidence for split ergativity in Hieroglyphic Maya. The shell in Figure 7 below, which is unfortunately missing its lower part, has been pointed out to me by Albert Davletshin (p.c. 2020) as one such case.

The incised shell is held and on display (as of January 2024) at the Humboldt Forum in Berlin. The shell has a height of only 5,5 cm and is likely from the Late Classic (600–900 CE) (Grube & Gaida 2006: 208). Aside from the text, we see a parrot sitting in a characteristic Mayan window which has the form of the glyph IK' 'wind'; the text that accompanies it must be taken as a commentary to the depicted scene, of which, unfortunately, a part has broken off

(Grube & Gaida 2006: 209). Remnants of a hand and a headdress at the breaking edge below suggest that the lower part contained an illustration of a person (Grube & Gaida 2006: 208). A remarkable trait of the text is that "it seems to be composed entirely of phonetic syllables" (Zender in Houston et al. 2004) without any logograms.

Figure 7. Incised shell K8885. Unknown provenance, 600–900 CE.

Following general conventions in glyphic research, I divide the shell into four lines (A–D) and number the glyph blocks from left to right. This yields the following pattern:

A1

B1 B2

C1 C2 C3

D1

The following analysis of the glyphic text on the shell is based on Grube and Gaida (2006: 208–211). Further discussions of it can be found in Davletshin (2013b: 87) and an unofficial exchange of opinions between Stephen D. Houston, Barbara MacLeod, Peter Mathews, David Stuart and Marc Zender archived online (Houston et al. 2004).

(1) K8885 A1

<a-wu?-le-li-ya>

aw-[h]ul-el=iiy

A2SG-arrive-NMLZ-ADV

‘“you arrived/your arrival?’

(2) K8885 B1 + B2

<ti-ni-?[385]-la me-te-ya-a-la-ni>

ti *ni-?l* *met* *y-a’al-aan*[386]

PREP A1SG-?[387] nest[388] A3-say-PST.PTCP[389]

‘at my … nest”, his said [=words]

[385] A new reading of JOM has recently been proposed for this glyph (T650; Lopes, MacLeod & Sheseña 2023). Together with the following phonetic complement <-la> they read *jomal*. Following the information provided to them by Kerry Hull, the authors connect this word with Ch’orti’ *jomi* ‘to burn’ and *jomir* ‘heat’ but choose to translate it as ‘window’ in this case following Tzotzil *jom* ‘embrasure, loophole, tronera’ because a T-shaped window is depicted on the shell. The full expression would be *ti ni-jomal met* ‘to/at my window-nest’ (Lopes, MacLeod & Sheseña 2023: 84).

[386] Davletshin reads both instances of the verb as *a’l* without reduplicated vowel. This would fit better to a Cholan verb, while *a’al* would be expected to be Yucatecan. The spelling of <ya-a> for *ya’* would constitute a case of what Lacadena and Wichmann (2004: 111) call “orthographic doubling” in their treatment of the orthography of the glottal stop in Hieroglyphic Maya.

Houston is wondering about the presence of the glottal stop in the verb (“Why the glottal stop, I wonder? Why the [ni]?” (Houston et al. 2004)) but in fact, the form is attested with glottal stop in modern Ch’orti’, so we may expect Hieroglyphic Maya to be *a’l*, as well, even though it is commonly read without glottalized vowel because of the harmonic spelling <ya-la> (Kettunen & Helmke 2020: 88; Boot 2009: 23).

[387] This word likely modifies *met*, perhaps another noun or an adjective (Grube & Gaida 2006: 210). Grube and Gaida mention that the same illegible sign appears in Yaxchilan with a suffix *-si*, of which it is known that it modifies body parts in their “absolutive”, i.e., unpossessed form (Zender 2004). Therefore, the word may be a body part. In the discussion online, Zender first suggested that the word may be <ho-la> ‘head’ but then pointed out that this would be expected to be written with <jo>, not <ho>.

(3) K8885 C1–C3

<?-po-ya-a-la-ji-ya[390]	hu-b'i-u-ju-chi			˙po-(po-)lo-tz'i-i[391]>
?p[392]	*y-a'al-aj=iiy*	*huub'*	*u-juuch*	*Popol Tz'i'*
?	A3-say-TS=ADV	shell.trumpet	A3-shell	Mat.Dog

(4) K8885 D1

<b'a-che-b'u>

b'aah chehb'[393]

first quill

'first of the quill'

In total, though most of the signs can be read and most of the words determined, the sentence still is not entirely intelligible and "remains frustratingly opaque overall" (Zender in Houston et al. 2004).

Grube and Gaida translate: '"You arrived/your arrival at my … nest", are its words, ?p, said the shell trumpet. (This is the) shell of Popol Tz'i' ("Mat Dog"), first scribe."[394]

Davletshin's proposed translation is as follows: "Then? you are coming to my (such-and-such) nest, Lorikeet says, as it has been said by Shellfish. It is the shell of Pohpol-Tz'iʔ, the first quill (palace title of chief scribe)."

---

[388] The meaning 'nest' is attested in Chol. There is also the meaning 'crown' (based on CHT *met* 'crown') or 'trivet, circular base, stand' (based on YUC *me'et*) (Grube & Gaida 2006: 211, fn. 2).

[389] According to Grube and Gaida (2006: 211, fn. 3), this formation is not used anymore (in modern languages?) and is "best understood as a participial construction [German "am besten als Partizipialkonstruktion zu verstehen"]".

[390] Of <ya-a-la-ji-ya>, only the beginning and ending syllable <ya> are really clear. Stuart transcribes it as "ya-[la]a-ji-ya" (Houston et al. 2004). All authors apparently assume that <a-la-ji> is written in the single sign in between the two <ya> syllables – and in fact, it is likely that the sign contains a <la> in its top right corner.

[391] Two dots above a glyph are a means of reduplicating signs (Houston et al. 2004). Therefore, <po> must be read twice here.

[392] The first sign evades a reading but has been hypothesized to be <o->. Davletshin proposes to read this as the plural morpheme *-ob'* despite the fact that it is written with <p>, not <b'>. Zender points out that "[s]ome of the languages have /op/ and /ohp/ as 'anona' (sour sop [sic]); but it's very unclear how this would relate to the rest of the text" (Houston et al. 2004). Perhaps *op'* is simply be onomatepoetic, i.e., just an imitation of the sound of a shell trumpet. Depending on what shell was used, it may have sounded like this: "The Maya Conch" by Penn Museum, https://www.youtube.com/watch?v=2Ke7GpsCM7A (last accessed: 2025-03-13).

[393] Both *chehb'* and *b'aah chehb'* are attested as scribal titles with *chehb'* actually denoting a writing tool (Grube & Gaida 2006: 211, fn. 7).

[394] "'Du bist angekommen auf meinem …-Nest' ist das Gesagte, …p', hat gesagt die Muscheltrompete. (Dies ist) die Muschel von Popol Tz'i', dem ersten Schreiber" (Grube & Gaida 2006: 211).

Since the verb *a'al* 'to say, speak' occurs twice on the shell, we may assume that it records direct speech, which is notoriously difficult to understand in Hieroglyphic Maya (Grube & Gaida 2006: 210).

The form in question, the one potentially showing split ergativity, is the first one, <a-wu?-le-li-ya>, analyzed as *aw-[h]ul-el-iiy*. Stuart stated: "This could be a split-ergative form, but this is something we'll all need to discuss and consider carefully." MacLeod wondered:

> "Is split ergativity being manifested in 'a-w-ul-el (if this proves correct)? It might just be a nominalization (my preference)-- not more nor less than that puzzling 'u-lok'-ob? on Copan St. 11, meaning 'it is their departure (YKM and YP)'. But these sorts of constructions were likely among the antecedents to split ergativity." (Houston et al. 2004)

There are a couple of problems with this example. According to the quoted discussion and Grube & Gaida (2006: 211, fn. 1), it is not entirely clear whether *awuleliiy* is a verb – in which case it would be an intransitive verb inflected with set A, which would constitute a case of split ergativity – or just a possessed noun phrase 'your arrival'. Since Hieroglyphic Maya, like most Mayan languages, is verb-initial, we would theoretically expect a verb at the beginning of the sentence. However, we cannot even be sure that the verb in question is in fact *jul* – no securely identified syllabic sign for <wu> is attested to this day. For example, it is missing from the most recent syllabary in Kettunen and Helmke (2020: 77). In Boot (2009: 81, 203), it is only attested twice, both times with a question mark, and one of the cases is this shell. If the syllable is in fact <wu>, it is unlikely that we are dealing with a Cholan language since they generally retain initial *j*.

If the verb is in fact *jul*, the presence of *=iiy* 'already, in the past' (Lacadena 2013: 60) is puzzling because the adverb is generally used with the completive, while the form here would be an incompletive one with a nominalizer *-el* like in Western Cholan. On the other hand, based on data from Gronemeyer (2014) it seems to be possible to use *=iiy* with nominalized verbal forms. A convincing translation would be "after it was your coming" as proposed by Gronemeyer (2014: 196), although he also considers the possibility that "[d]espite the problematic reading and the occurrence of the temporal enclitic, there is a possibility that this is an incompletive 'after you come'." Gronemeyer (2014: 196, fn. 439) states that despite the argumentation in Law, Robertson and Stuart (2006), he believes in split ergativity in both Hieroglyphic Maya and Proto-Cholan but chooses not to relay his argumentation because he only states this as a side remark in a footnote. His reference of Coon's analysis of split ergativity as an "illusion" and all split-ergative forms as underlyingly nominal (Gronemeyer 2014: 196, fn. 440) shows that the latter may cause confusion and reinforce the idea that

nominalized verb forms are representative of split-ergativity even if they are not integrated into verbal paradigms – whether this was Coon's intention or not.

Especially since the shell seems to record direct speech, we need to consider the possibility that at least the first part is not written in the Cholan language that the hieroglyphic texts otherwise are written in. The Maya in the Lowlands lived in a state of diglossia: the ones that could write, did so in the prestige language but in their daily life they probably spoke a number of different Mayan languages (Kelly 2017). Therefore, if the shell indeed features split ergativity, we would still need to prove that the language it is written in is Cholan if we want to use it as proof of how split ergativity developed in the Cholan branch. The fact that we do not know anything about the provenance of the shell does not aid in its interpretation. However, whether the form belongs to a Cholan or a Yucatecan language, the current hypothesis concerning the genesis of split ergativity in both languages operates with the assumption that the split was caused by some other element (verbs, adverbs, perhaps something else) that occurred before the verb and required a nominal complement (see chapter 3.3). Here, we have nothing in front of the verb and therefore, it is not likely that we are dealing with a precursor of split ergativity. We are probably just dealing with a nominal form.

The second example that I want to briefly touch on is Copan Stela 11[395], which was mentioned in MacLeod's quote above with the proposed reading "'u-lok'-ob?" for the verb in question. Though the text on that stela is enigmatic (Martin & Grube 2008: 212), the verb in question is more likely to be read <u-lo-k'o-ma> rendering *u-lok'-oom* in (5).

(5) Copan Stela 11 (date: 9.19.10.0.0, i.e., 801 CE)
<u-lo-k'o-ma>
u-lok'-oom
A3-leave.AGT
'his/their leaver (?)'

The suffix *-oom* is given with a short vowel in Kettunen & Helmke (2020: 124) though the disharmonic spelling would suggest either length or glottalization. The authors describe it as an agentive suffix. Based on the attested forms in their document, it can be used to derive agentive forms from both intransitive and transitive verbs as well as from other nouns:

[395] A drawing by Linda Schele can be found here: http://research.famsi.org/schele_list.php?_allSearch=stela+11&hold_search=stela+11&tab=schele&title=Schele+Drawing+Collection&x=0&y=0 (last accessed: 2025-03-13).

Noun: *k'ay* 'song' > *k'ay-om* 'singer' (Kettunen & Helmke 2020: 109)
IV: *uht* 'to happen' > *u[h]t-om* '?' (Kettunen & Helmke 2020: 95)
TV: *jatz'*- 'to strike' > *jatz'-om* 'striker' (Kettunen & Helmke 2020: 91)

The semantics remain a bit unclear with the intransitive verb but that does not matter for the argument I want to make: these are **nominal** forms. They are not embedded in any kind of special verbal paradigm. Though they are not used with a prefixed set A in the cases mentioned by Kettunen & Helmke, it is not difficult to imagine a meaning like 'my singer'.

I need to stress that statements like "these sorts of constructions were likely among the antecedents to split ergativity" as MacLeod phrased it (see above) are not helpful in understanding the genesis of split alignment. We need to be very clear in our analysis: the fact that nominalized forms exist in a language has no bearing on whether or not a split will develop. As she herself says, the split-ergative alignment in modern languages today develops based on a specific construction that affected intransitive verbs. Here, there is no construction, just an agent noun. Of course, a single form can already constitute a construction when it is used in a special new way but as already stated, stela 11 more or less evades a clear reading and interpretation for now. Therefore, we have no reason to assume that that is the case especially since this nominal form has an ending (*-o(o)m*) that does not occur in the incompletive aspect of Cholan languages with split ergativity.[396] How can "these sorts of constructions" then be precursors to it?

Exactly the same criticism applies to Mora-Marín's "evidence suggestive of split ergativity" as it is based on the same kind of form, again from Copan (Monument 157):

(6) Copan Monument 157 (Mora-Marín 2005: 78; date: 783 CE)
<u-CHOK-no-ma>
u-chok-n-oom
A3-scatter-AP-AGT
'his/their scatterer'

In (6), we again have a verb, transitive but intransitivized via the agent focus antipassive marker *-n*, which is further derived to an agentive form. Mora-Marín pointed out that "the dual presence of u- and -n can only be reconciled here if one assumes a type of split ergativity at work" (Mora-Marín 2005b: 78) but he analyzed *-om* as potential aspect, which likely is the

[396] I should note that it does occur in the potential aspect in Poqomam, where it is, however, the suffix of the transitive verb, not the intransitive one (see chapter 3.3.3 for details).

source of this misinterpretation because then the form would indeed be an inflected intransitive verb with set A marking of S. Though he reaffirmed this statement in 2009 (Mora-Marín 2009b: 121), he may have changed his mind by now, especially since *-om* is generally analyzed differently today.

Concerning example (6), Mora-Marín states: "This is the best example of a split ergative construction in any text known to me" (Mora-Marín 2005b: 78). If this is the case, the matter barely merits further discussion. Even if the potential examples of split ergativity were valid and clear, they would still be so scarce that they could not be considered evidence for a systematic alignment split in Hieroglyphic Mayan. I would consider the shell the most conspicuous piece of evidence but with it, too, a lot remains unclear.

11.3 Discussion

In principle. the debate on whether Hieroglyphic Maya shows aspect-based split ergativity is easily settled considering that it depends on the presence of aspect. As discussed in the introduction to chapter 7, there seems to be consensus that Hieroglyphic Maya did not distinguish between completive and incompletive aspect morphologically. Even if it makes sense to assume that most Hieroglyphic Mayan texts are historical and historical texts are frequently written in the completive aspect, especially when they recount sequences of completed events, it would be extremely unlikely that a corpus as large as the one of Mayan hieroglyphic texts would not have a single case of incompletive aspect (Law & Stuart 2017: 168), especially since there are cases of direct speech. Therefore, there is likely no morphological marking of aspect and hence there can be no alignment split between the two aspects.

Theoretically, it is possible that Hieroglyphic Maya shows split alignment that is conditioned by something other than aspect, e.g., by independent vs. dependent clauses. However, as a whole, this seems rather unlikely for the following reasons:

1. Earlier researchers presumed the existence of a split precisely because the daughter languages show one. If we cannot assume that Hieroglyphic Maya has an aspect system comparable to those of the daughter languages and if the daughter languages differ significantly in how the split alignment operates, this argument becomes superfluous.

2. We have seen that the assumption that split ergativity should be found in Hieroglyphic Mayan texts rests to a large extent upon a number of misconceptions about modern Cholan languages, e.g., Ch'orti', as well as the incorrect interpretation of some of the glyphs. Today, we know the correct reading, e.g., for the glyph denoting *b'aah* and I need to again emphasize that the simple presence of nominalized verb forms does not constitute evidence of split alignment.

3. Examples that have been claimed to show split ergativity like the shell discussed above commonly turn out to be inconclusive or can be refuted easily.

If we conjecture that Hieroglyphic Maya features a different type of split alignment but also want to hold onto the theory that Ch'orti' is its descendant, we need to assume that one alignment split developed just to be lost again and that another split alignment then developed on the way to Ch'orti' that is synchronically conditioned by aspect. This seems to be an unnecessary assumption given that the reasons why we ever assumed that Hieroglyphic Maya may show ergativity in the first place have either become obsolete or are hardly conclusive.[397]

However, as Hieroglyphic Maya does not feature clear cases of split ergativity on the one hand, but modern Cholan languages do feature it on the other hand, it is a valid hypothesis to assume that the genesis of split ergativity is recorded in the hieroglyphic corpus somewhere. Of course, this ties into the debate of whether the language of the hieroglyphs is the ancestor of Eastern Cholan or of all Cholan languages. In this thesis, I demonstrated which Hieroglyphic Maya construction the type of split ergativity that is featured in Ch'orti' arose from. The type of split ergativity of Choltí, which depends on the presence of aspect marking and a suffix *-el*, seems to be parallel to that of Western Cholan languages; therefore, I would expect that in all three cases, it goes back to the same kind of construction, which is different from the one in Ch'orti'. The examples discussed in this chapter cannot be seen as evidence for the incipient development of these constructions, so the search must go on. The later attested hieroglyphic texts exceedingly feature a Yucatecan variety, e.g., the Dresden Codex, which was written around the 13th century in Northern Yucatan, partly in Yucatecan and partly in Cholan (Wald 2004b). This enables us to look for source constructions of both Cholan and Yucatecan *-el/-Vl*-type split ergativity in the corpus while paying close attention to which of the languages we are dealing with at every step. A detailed analysis of the whole

[397] Notably, Lacadena (2011: 360) reaches the same conclusion concerning split ergativity in Proto-Yucatecan based on hieroglyphic evidence written in a Yucatec variant.

text with this goal has not been carried out yet, neither for the Dresden Codex nor for any other hieroglyphic Maya source. This is a very promising project for the future.

12 Typological considerations

The subtitle of my dissertation is "A contribution to a diachronic typology of alignment change". The pathway of change described in this dissertation for set C of Ch'orti' is found at the intersection of discourse and grammar and demonstrates how pragmatic communicative principles can become manifest in a language's grammar or, more specifically, how incompletive/imperfective aspect can arise from a focus construction.

Although there has been substantial discussion of how information structure and grammatical aspect correlate and interact (including the paper by Hopper (1979) which I discussed in chapter 7.2.1), and although we have some knowledge of the sources of focus elements (e.g., deictics), it is difficult to find information on what focus may then further grammaticalize into. From my impression, research on this has at least in the past been largely restricted to languages of Africa – as Güldemann remarks, the work by Hyman and Watters (1984) "has received relatively little recognition outside certain circles in the field of African linguistics" (Güldemann 2003: 323). Consequently, focus constructions as potential sources for, e.g., incompletive aspect are not taken into account even though Hyman and Watters' analysis is "a groundbreaking treatment of the interaction between the information structure of a clause and the morphosyntactic marking of predication-operator functions like polarity, modality, tense, aspect, etc." (Güldemann 2003: 324). Linguists working on languages of Africa have produced such a wealth of literature on information structure and grammar that I was barely able to scratch at the surface when looking for inspiration and data for my analysis of Mayan.

Most importantly, Güldemann (2010) has proven very helpful to me in understanding how Ch'orti' incompletive aspect likely developed by way of subject salience from the polyfunctional focus construction that encoded both contrastive focus and could likely also be used to express thetic statements. Following the discussion in chapter 7.2.2, we now can imagine three possible pathways from focus to aspect:

1. focus > predication focus > progressive/imperfective (> future)
2. focus > agentivity > imperfective (> future)
3. focus/thetic statements > imperfective (> future)

The final step from imperfective to future, a kind of hypoanalysis akin to what is described by Haspelmath (1998) for old presents that become futures because they are restricted in their

use by new present formations, is possibly witnessed in Cholti and Poqomchi', though this requires a more in-depth historical study of the respective languages.

Functional reasoning aside, I would like to focus on the formal side of things and to return to the correlation of ergative-absolutive alignment with past/perfective on the one hand and nominative-accusative alignment with non-past/imperfective on the other (see chapter 2.1.3). It is insightful to review Anderson's (1988: 340–349) suggestion to understand it as a secondary effect of the contrasting pathways of passive vs. antipassive reanalysis, which seem to often lead to ergative-absolutive alignment and nominative-accusative alignment respectively. Harris and Campbell conclude that, while the correlation is not universally valid, "the kernel of truth it still contains is due to the independent associations of passive with perfectivity and object demotion with imperfectivity" (Harris & Campbell 1995: 246). Their emphasis of independent association is very important. The usual statement that in the case of a split imperfective mostly going together with nominative-accusative alignment while perfective shows ergative-absolutive alignment in my view invites a semantic explanation as well as a direct connection between alignment type and aspect. However, it is only by way of analyzing the formal side, i.e., the construction that the alignment goes back to, that we can begin to understand why this correlation exists.

As I pointed out in chapter 2.1.3, the source of Indo-Iranian tense-/aspect-based split ergativity is not a reanalyzed passive because the participle that is used in the construction is a resultative one that denotes a state and therefore, one could argue, stresses the affected participant, either S or O. What the resultative participle and a passive have in common is their absolutive or resultative orientation.[398] The Indo-Iranian *ta-participle can be made to reference S or O but never A. Similarly, the passive formation highlights O of transitive sentences while A is not expressed at all or obliquely, turning O into an S. If either of the formations is integrated into a verbal paradigm, the result will be ergative-absolutive alignment because S and O will be marked in the same way, e.g., by the same case as in Indo-Iranian, while A will be marked in some other way. Therefore, the pathway is best described as "absolutive orientation > ergative[399]", not "passive > ergative" because the latter will by far not cover all cases. The correlation of ergative-absolutive alignment and past/perfective is

[398] Participial orientation describes "different possible relations between the participle, which is a verb form, and the nominal it modifies, which is a participant related to this verb" (Shagal 2019: 51). Participles may be "inherently oriented" when they only ever modify one (or two) particular participant(s) of a verb or "contextually oriented" when they can modify various participants depending on the context (Shagal 2019: 51).

[399] "Ergative" here meaning ergative-absolutive alignment, not ergative case.

precisely that: a correlation. It exists because past and perfective both stress the result of the action (since it is already completed/in the past) and because resultative forms often develop into past/perfective forms (Bybee, Perkins & Pagliuca 1994: 68, 81).

If we apply this to the opposite case, namely the grammaticalization of formations yielding nominative-accusative alignment and the correlation of this alignment type with non-past tense or imperfective aspect, we may conclude that here, the forms that enter the verbal paradigms are oriented towards S and A but never O.[400] In the Georgian examples discussed in chapter 2.1.3, nominative-accusative alignment developed by way of an antipassive in a language that otherwise attests ergative-absolutive alignment. Functionally, the construction in (1) was reanalyzed as imperfective aspect parallel to imperfective aspect of non-derived intransitive verbs in (2) that already existed in the language.

(1) Modern Georgian object demotion (Harris & Campbell 1995: 245)

deda	*p'erang-s*	*recxavs*
mother.ABS	shirt-DAT	washes

'Mother is washing the shirt.'

(2) Modern Georgian intransitive imperfective (Harris & Campbell 1995: 246)

tamar	*didi*	*mepe*	*aris*
Tamar.ABS	great	monarch.ABS	is

'Tamar is a great monarch.'

Since an antipassive construction contains a syntactically intransitive verb, absolutive case originally marked S in both examples. However, the obliquely expressed object in dative case was then reanalyzed as a direct object. As a result, synchronically, the old antipassive is now a transitive verb again and with S and A being marked alike whereas O is marked differently, we are left with nominative-accusative alignment.

Returning to the functional analysis, Harris and Campbell argue that "[t]he object demotion construction is associated with imperfective aspect or incomplete effect on the object" (Harris & Campbell 1995: 246), i.e., just as the passive construction emphasized S and O while demoting A, the antipassive construction emphasizes S and A while demoting O. I would argue that here, too, we are dealing with a mere correlation of antipassive and nominative-accusative alignment while the cause of both is the inherent orientation of the employed constructions. I propose to see agent nouns and similar formations (as discussed in 7.2.2) as

[400] When these forms are participles, they may be called "agent participles" following Shagal (2019: 63).

the counterpart to the *ta-participle, i.e., a single form with a specific orientation that may be grammaticalized to yield a specific alignment as opposed to whole constructions such as passive and antipassive as visualized in Table 33:

Table 33. Sources of alignment.

	Ergative-absolutive alignment	Nominative-accusative alignment
Orientation	**S + O ↔ A**	**S + A ↔ O**
Form, e.g.,	Iir. *ta-participle	agent noun
Construction	passive/resultative	antipassive

It is important to stress that these are just examples and many more specific sources may be identified in the future. I hypothesize that what they will all have in common will be their inherent orientation towards either S and O or S and A.

However, we should not forget that there are other basic alignment types (tripartite, neutral or horizontal alignment) alongside the most common ones (ergative-absolutive and nominative-accusative). The former are severely understudied and the variables that lead to their development are not well understood. For Ch'orti', I argued that a special antipassive used in Mayan focus constructions when the agent was put into the focus position together with S focus was grammaticalized to denote S specifically in the incompletive aspect. In other words, there was a similar treatment of S and A while O focus used transitive verb forms. Still, the resulting alignment in the new incompletive aspect is not nominative-accusative. It seems, then, that antipassive constructions can apparently also lead to a tripartite alignment[401] given the right circumstances.

To conclude, the resulting alignment of a newly grammaticalized construction strongly depends on the orientation of the formations or constructions that are incorporated into the verbal paradigms. This orientation may lie in individual forms as in the case of the Indo-Iranian *ta-participles or in complete constructions like passive or antipassive. However, many small variables can influence the precise path a grammaticalization takes. Were it not for the agent focus construction and its requirement to demote transitive verbs to intransitive ones, the outcome in Ch'orti' may have been different from the tripartite alignment that

[401] Historically, we may still consider Ch'orti' a "straight" ergative language because S in the incompletive continues to be marked by forms that ultimately go back to set B, but synchronically, set C and B differ so radically that the label "tripartite" is certainly more appropriate. Additionally, set C is not simply set B but set B added to an independent pronoun base *h.

ultimately resulted. Understanding the details of the grammaticalization processes in individual languages will aid us in understanding why seemingly parallel constructions may yield different results – or, conversely, how identical results may arise from different source constructions.

13 Conclusion

In this thesis, I investigated the alignment split of Ch'orti' and the associated third index-set ("set C") from a diachronic perspective. Conventionally, this split alignment has been treated as comparable to that of Lowland Mayan languages that also show split ergativity. Consequently, set C has been explained as originating in set A. In chapter 3, I demonstrated that Yucatecan and Western Cholan split ergativity are not identical and that there must be something about the origin of it in Yucatecan that we have not accounted for yet. If the two are not similar, then this is all the more true if we compare them to Ch'orti' (chapter 4) with its aspectual distinction that only applies to intransitive verbs and the optionality of aspect markers.

I further demonstrated that the conventional explanation of the origin of set C offered by Robertson (1998) and Wichmann (1999) cannot stand up to scrutiny both on methodological grounds (e.g., liberal use of analogy without appropriate contexts for reanalysis, ad hoc assumptions) and because it ultimately leaves many questions unanswered (chapter 5). Quizar's (2023) theory only differs from Robertson's (1998) and Wichmann's (1999) in the origin that she proposes for the C3 *a-* morpheme while the development of the other forms that are not identical to that of set A is not discussed at all. All three theories do not adequately account for the puzzling features of set C and split ergativity in Ch'orti': 1) lack of nominalizer, 2) special development of set A in the form of set C while at the same time preserving the original paradigm, 3) an aspectual distinction that only applies to intransitive verbs and 4) different morphophonological behavior (chapter 6).

Therefore, I suggested that a derivation of set C from set B may be preferable because this enables us to account for all peculiarities of the morphosyntax of Ch'orti' (chapter 7). I argued that set B, which is originally reconstructed as enclitic, occurred in front of the verb as part of the independent pronoun when it was used in the focus position. Some Mayan languages have a restriction when it comes to focusing the A of transitive verbs: the verb must be intransitivized, which is frequently done with an antipassive form. The original agent focus construction is attested in the hieroglyphic corpus while it is lost in the modern Cholan languages. Since both A and S were encoded in the same way when focused, i.e., as S, they were reanalyzed as a new set of S markers of intransitive verbs in Ch'orti'.

As to the semantic development of focus to incompletive aspect, one could imagine three different possible paths. For Ch'orti' specifically, it is the fact that the focus construction

(including agent focus for A) was used to for thetic statements alongside contrastive focus that ultimately made it possible for the newly emerging set C to occupy the functional niche of incompletive aspect as the latter shares many properties with thetic statements. This was facilitated by the fact that the pragmatic force of this construction faded as it was obligatorily used in certain contexts. Since the construction was only used with intransitive verbs, aspect marking only developed for intransitive verbs.

For the two indexes that did not match the independent pronoun paradigm (C2SG and C1PL), I proposed two proportional analogies with clear contexts of where the reanalysis may have happened. The remaining problem concerns the loss of the initial *h- element that set B was suffixed to in Hieroglyphic Maya. In general, CHR j (< PM, PCH *j and *h) is retained in word-initial and -final positions and only elides with subsequent vowel contraction when it occurs intervocalically (chapter 6). For this, I proposed three possible explanations:

1. Evidence from the numeral '1' hints at the fact that a specific sound change made initial *j- disappear utterance-initially.

2. Alternatively, we could assume that set B was not suffixed to anything in Ch'orti' and that it was used in the focus position as an independent pronoun on its own. Somewhat conflicting data from K'iche' suggest that this may have been possible in Proto-Mayan. However, this assumption would potentially have far-reaching consequences for the relationship of Hieroglyphic Maya and Ch'orti'. The reconstruction of the syntactic status of set B as affix (unlikely), clitic or independent form is still a matter of debate but set B does not seem to have been used independently in Hieroglyphic Maya and if this was in fact possible in Ch'orti', this would suggest a more archaic state. It may then be unlikely that Hieroglyphic Maya is the ancestor of Ch'orti'.

3. A behavioral stage of "ditropic clitics" may have been the reason for the prefixation of set C to the verb instead of being suffixed to the preposed *h-. This may even have facilitated the utterance-initial loss of *h-.

A phenomenon that has often related to set C in Ch'orti' or at least to C3 *a-* are the two constructions of Colonial Choltí involving a particle *a* that form 1) a general present and 2) a future with modal connotations of obligation. In chapter 8, I discussed what kind of assumptions need to be made to apply a similar explanation to the *a* in Choltí as I proposed for Ch'orti'. Though the initial situation is not exactly parallel and some open questions remain to be addressed, one may assume that in Choltí, the third person singular independent

pronoun developed into a designated focus marker and was henceforth used in in situ focus, i.e., a construction that indicates that an argument is focused by putting a focus particle in front of the verb while the focused constituent itself remains in its usual position in the sentence. In some languages, e.g., K'iche', the in situ focus likewise can only focus S (or O) but never A, which would account for the use of *a* only with intransitive verbs.

I also discussed similar aspectual/temporal markers for Chontal and Poqomchi' where, however, it is not necessary to assume that the focus construction was involved at all because the formation is used with transitive verbs, too, and therefore we do not need to justify the same restriction as for Ch'orti' and Cholti. An aspectual *a* can easily derive from a deictic element, which the *a* originally is. However, if we wanted to derive the *a* in these languages from a focus construction similar to Cholti, this would be possible.

The derivation of Ch'orti' alignment from the Hieroglyphic Mayan focus construction also has implications for Hieroglyphic Maya itself, more specifically for the question whether it features an alignment split or not. I demonstrated that the idea of split ergativity in Hieroglyphic Maya is partially based on misconceptions about both modern Cholan languages and the interpretation of specific glyphs and otherwise rests on inconclusive evidence. From our viewpoint today, one would not even consider the idea that split alignment may be present in the language. Since there is no grammatical distinction of aspect in the hieroglyphic texts at all, there cannot be an alignment split conditioned by aspect, either. If, on the other hand, we assume that Hieroglyphic Maya features an alignment split conditioned by something other than aspect (again, without there being any reason for it), we have to make the assumption that said split disappeared in between Hieroglyphic Maya and Ch'orti' only for Ch'orti' to then develop another, this time aspectual, split – an unnecessary additional assumption that should be avoided.

From a typological standpoint, this dissertation contributed to understanding the intersection of pragmatics and grammar on one hand and clarifying the mechanisms behind certain correlations seen in alignment theory on the other. I discussed different pathways of how incompletive/imperfective aspect marking can develop from focus: in addition to the pathways via predication focus and agentive marking, my proposal that is based on Güldemann (2010) provides evidence that the use of a construction for thetic statements may facilitate the development of said construction into incompletive/imperfective aspect marking as theticity and imperfectivity are very similar.

I further hope that my terminological discussion of split ergativity versus active-stative split (in general but also specifically in Mayan, see chapter 3.4) clarifies why it is not a useful categorization to separate Mayan split ergativity from the same phenomenon in, e.g., Indo-Iranian languages only because the new construction that gave rise to the alignment split affected intransitive verbs. Rather, it is the conditioning of the split or its diachronic origin that should be used as a base for classification and comparison.

Finally, I pointed out many areas where research needs to be carried out in the future: if we take the proposal of this thesis seriously, in-depth research, both in the form of reconstructive work and corpus studies, is needed on the languages which I could only touch on superficially (Cholti, Chontal and Poqomchi'). Furthermore, more research is needed into the diachrony of split ergativity in Yucatecan and the role that focus played in its genesis. More specifically, we should evaluate whether this can account for the fact why set B is not automatically suffixed to the aspect markers though they are present in Yucatecan. Hopefully, more examples of focus constructions will turn up in Hieroglyphic Maya so that we can gain more insight into the formation. Insights will likewise come from a more detailed comparative study of Mayan languages in general. Most importantly, a cooperation with scholars studying Bantu and other languages of Africa may turn out to be especially fruitful as they clearly have made inspiring advances in the study of information structure.

As a more general outlook, I believe to have clearly demonstrated that Ch'orti' specifically but also Mayan languages in general are in need of further methodically sound historical-comparative study and reconstruction. A better understanding of the grammar and history of Ch'orti', especially of its status suffixes, the precise rules of the morphophonological interaction and its information structure, may aid in the further decipherment of Hieroglyphic Maya and in understanding unclear text passages like the discussed shell. Because of their wealth of attested alignment features, Mayan languages also present an excellent case study for examining the subtle nuances in constructions that determine the outcome in alignment change. I hope that I have inspired others working on alignment to take these languages into account in the future.

Appendix A – Language names and classification

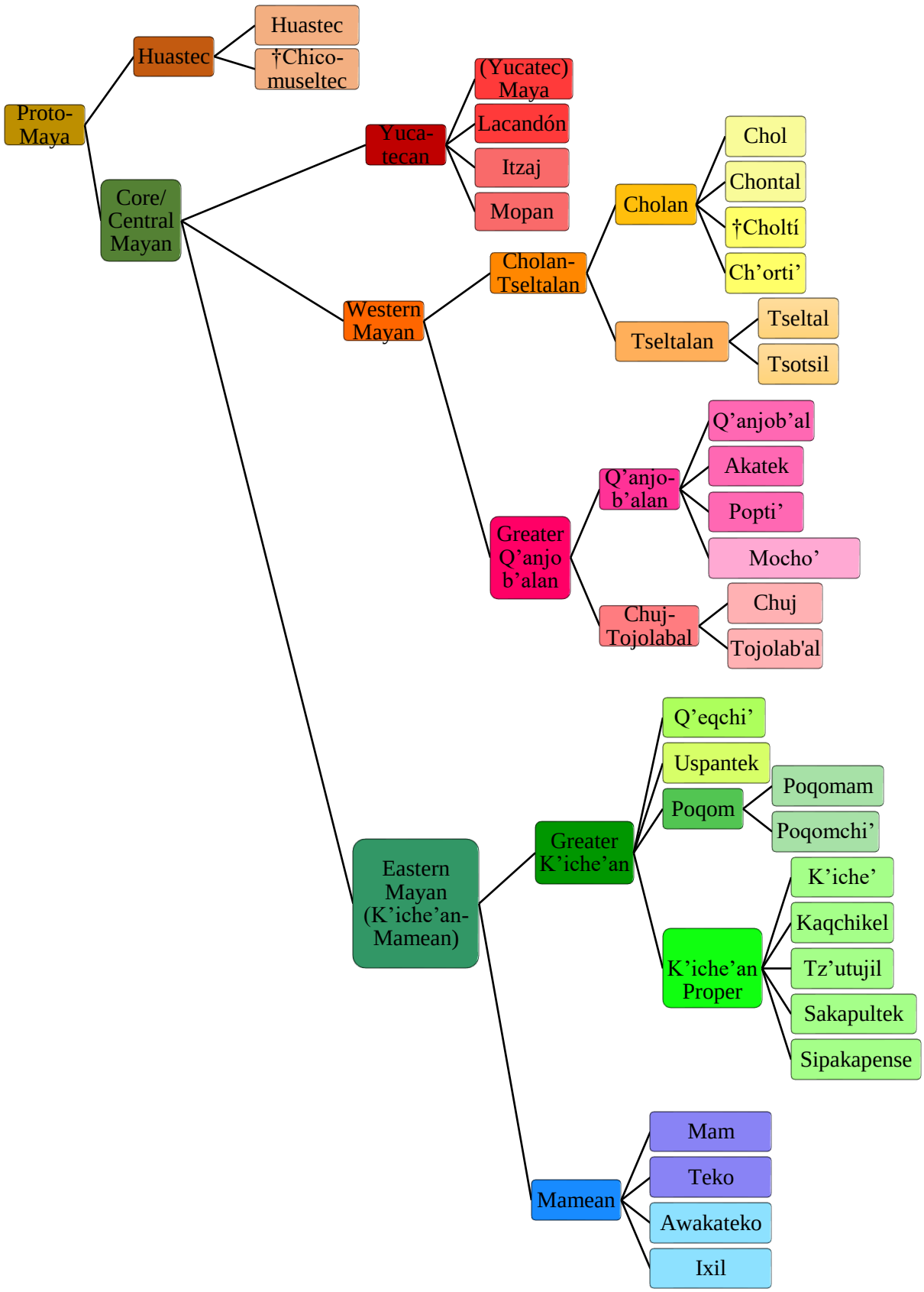
Table 34. Current and former Mayan language names with variants, mostly based on conventions of the ALMG for languages spoken in Guatemala and of the INALI for languages spoken in Mexico (Aissen, England & Zavala Maldonado 2017a: 8–9).

Current	Older (if different)	Additional Variants
Achi	Achí	
Akatek	Acatec	
Awakatek	Aguacatec	
Chicomuseltec	Chicomuceltec	Kabil
Chontal (de Tabasco)		Yokot'an
Chuj		
Ch'ol/Chol (dialectal)		
Ch'olti'	Choltí	Cholti'
Ch'orti'	Chortí	
Huastec		Wastek, Teenek
Itzaj	Itzá	
Ixil		Ixhil
Kaqchikel	Cakchiquel	
K'iche'	Quiché	K'ichee'
Lacandón[402]		Lakantun
Mam		
Mocho' (+Tuzantec)	Mochó, Motozintlec	Cotoque, Kotoke (for both varieties)
Mopan	Mopán	
Poqomam	Pocomam	Pokomam
Poqomchi'	Pocomchí	Pokomchi'
Popti'	Jacaltec	Jakaltek
Q'anjob'al	Kanjobal	
Q'eqchi'	Kekchí	
Sakapultek	Sacapultec	
Sipakapense	Sipacapense	Sipacapeño, Sipakapenyo
Tojol-ab'al	Tojolabal	Tojolab'al
Tseltal	Tzeltal	
Tsotsil	Tzotzil	

[402] I use the form "Lacandon" without an accent following Hofling (e.g., 2017).

Tz’utujil	Tzutujil	Tzutuhil, Tz’utujiil
Tektitek	Teco	Teko (Cakchiquel *sic*)
Uspantek	Uspantec	
Yucatec Maya		Maya, Yucatec

Figure 8. Classification of Mayan languages based on Campbell (2017: 44). Converted from list to color-coded graphic. The length of the lines does not imply relative time depths.



Appendix B – Word list Juan Galindo (1834)

Quoted after Galindo (1834) and Morley (1920: 601–602).

1834 Orth.	Translation	Modern Ch'orti' (Hull 2016)
Quin	Sun.	*k'in*
Uj	Moon (a month).	†*uj*[403]
Ek	Star.	*ek'*
Kak	Fire.	*k'ajk' ~ k'ajk*
Ja	Water.	*ja'*
Tokar	Cloud.	*tokar*
Uinik	Man.	*winik*
Ixik	Woman.	*ixik*
Tegerom	Boy.	*tejrom* 'boy, young man'
Ikchok	Girl.	*ijch'ok* 'girl, daughter, female'
Tatá	Father.	*tata'*
Tu	Mother.	*tu'*
Unen	Son, Daughter.	*unen* 'son/daughter (of a man)'
Sacun	Elder brother.	*sakun*
Uitsin	Younger brother.	*uwijtz'in* (A3 *u-* + brother *ijtz'in*)
Jor	Head.	*jor*
Sutsernijor	Hair.	*-tzutzer nijor* '(its) hair my head'[404]
Unacaut	Eye.	*unak'u't* 'your eye(s)' (A3 *u-* + *nak'* 'inside' + A2SG *a-* + *ut* 'face')
Chiquin	Ears.	*chikin*
Ti	Mouth.	*ti'*
Caab	Hand.	*k'ab'*
Tigere	Tapir.	**tijr?[405]

[403] Still attested in HGM *u'h/uh* (Kettunen & Helmke 2020: 118). According to Wisdom (1950: 751), <ux> was "apparently not often" used anymore in the 1930s. He otherwise only attests <ka.tu> 'our mother' (see discussion in Appendix E chapter a.2 on why the moon is referred to as a female relative). *Katu'* is the only attested lexeme for 'moon' today according to Hull (2016).

[404] *Tzutzer* would usually be used with a possessive set A prefix today.

[405] No word for 'tapir' is attested in Hull (2016). Wisdom (1950: 730) gives <cimin>, which must be *tsimin* based on comparative evidence (MOP *tzimin*, TSE *tzemen*) but is not the expected lexeme for Cholan, which is reconstructed as PCH *tihl (Kaufman 2003: 569), though the forms this reconstruction is based on are not provided except for the Hieroglyphic Maya form, which is *til* (Kettunen & Helmke 2020: 117). According

Masá	Deer.	*masa'*
Mut	Bird.	*mut*
Chai	Fish.	*chay*
Guiná	Hunger.	likely *wi'na* 's/he is hungry' ('IV50 be hungry') rather than *wi'na'r ~ wi'nar* 'hunger'
Jacatiniti	Thirst.	*Ataki niti'.* 'I am thirsty' (C3 *a-* + *taki* 'IV1 to dry up' + A1SG *ni-* + *ti'* 'mouth')
Aingüe	To eat.	*inwe'* (C1SG *in-* + *we'* 'IV3 to eat') with unclear <A> (see footnote 185)
Unchi	To drink.	*u'nch'i* (A/C1SG *in-* + *uch'i* 'TV4/IV25 to drink)
Inguaian	To sleep.	*inwayan* (C1SG *in-* + *wayan* 'AP12 to sleep')
Inchamai	To die.	*inchamay* (C1SG *in-* + *chamay* 'MPAS4 to die')
Te	Tree.	*te'*
Ucabte	Branch of a tree.	*uk'ab'te'* (A3 *u-* + *k'ab'* 'arm' + *te'* 'tree')
Tajte	Ocote-pine.	*tajte'*
Ixim	Maize.	*ixim*
Ajan	Ear of green corn	*a'n*
Uchigtun-cha	Grindstone.	?[406]
Tun	Stone.	*tun*
Otot	House.	*otot*
Uitsir	Hill.	*witzir*
Nojá	River.	*xukur*[407]
Inté	1.	*inte'*
Chaté	2.	*cha'te'*
Uxté	3.	*uxte'*
Chanté	4.	*chante'*
Joté	5.	*inmojy/jo'te'*[408]

---

to Kaufman and Norman (1984: 132), the form is otherwise only attested in Choltí as <tiil>. The expected reflex of *tihl in Ch'orti' would be **tijr. Perhaps this is precisely what is rendered by Galindo's orthography.

[406] Unclear formation containing *tun* 'stone' and *cha'* 'grinding stone' with <Uchig> possibly being *juch'i* 'TV1 to grind' or perhaps intransitive *juch'ij still attested in Wisdom (1950: 475) as <huč'uh> though the second vowel does not match. This would then be proof that Galindo does not consequently write initial *j*.

[407] <Nojá> likely consists of *noj* 'big' and *ja'* 'water'.

[408] Modern Ch'orti' uses a new counting system that is based on *inmojy* 'five' as well as Spanish numerals (see discussion in Appendix E chapter g). Higher numerals are unattested in their original Mayan form in

Uakté	6.	*wakte'*
Uaxikté[409]	7.	*ukte'*
Ukté	8.	*waxikte'*
Boronté	9.	*b'oromte'*
Launté	10.	*la'inte'*

Ch'orti' according to Hull (2016) though the ALMG lists some forms in the descriptive and pedagogical grammars of Ch'orti' (Morwa'r ojroner Ch'orti'/Comunidad lingüística Ch'orti' 2004: 67; Morwa'r ojroner Ch'orti'/Comunidad lingüística Ch'orti' 2008: 83). It is unclear whether these are an attempt at language revitalization or whether the numerals really survived in Ch'orti'. The latter is unlikely since Hull does not list them.

[409] From comparative evidence we know that Galindo mixed up '7' and '8'.

Appendix C – Mayan linguistic terminology

Completive – usually preferred term for perfective aspect to avoid confusion with the category → **perfect** (Dayley 1982: 47); see also footnote 121

Dependent status – the difference in status suffixes depending on whether the verbs are found in independent or dependent clauses is reconstructed for Proto-Mayan (Kaufman 2015: 193); a very salient feature of Mayan languages

Derived verb – verbs may require different morphology depending on whether their stem is formed from the root or derived, e.g., as passive forms to transitive verbs or as causative forms to intransitive verbs

Imperfective – mostly not used in Mayan linguistics → **incompletive**

Incompletive – usually preferred term for imperfective aspect as a counterpart to → **completive**

Perfective – mostly not used in Mayan linguistics → **completive**

Positional – special class of verbal roots denoting bodily positions or states that requires different morphology; see chapter b.1 in Appendix E

Relational noun – "a special category of always possessed nouns in Mayan languages which primarily introduce NPs showing case and locative relationships" (England 1983a: 4); a productive class of prepositions built on possessed (mostly) body part terms that express a relationship metaphorically

Root verb – verbs may require different morphology depending on whether their stem is formed from the root or derived, e.g., as passive forms to transitive verbs or as causative forms to intransitive verbs

Set A – index-set used to mark A and possession in a language without alignment split or A + S, e.g., in incompletive aspect of languages with aspect-based split ergativity

Set B – index-set used to mark S and O in a language without alignment split as well as to form non-verbal predicates

Set C – index-set only used in Ch'orti' to mark S in incompletive aspect

Stative construction – forms non-verbal predicates by attaching a set B morpheme to, e.g., a noun, adjective, etc.; comparative to copula

Status suffix – a specific type of suffix that may contain a mixture of information such as verb type (transitive/intransitive), status (dependent/independent) and aspectual or modal nuances (Aissen, England & Zavala Maldonado 2017a: 5)

Thematic suffix – a term used by Kaufman and Norman (1984: 81–82, 94) to denote the variety of suffixes used on intransitive verbs in Ch'orti' and Choltí (see chapter b.2)

Appendix D – List of Ch'orti' morphemes[410]

Prefixes[411]

a- C3

aj- agentive prefix attached to nominal forms

a(w)- A2SG

i- C2SG

in- C1SG

in(w)- A1SG on verbs

ix- C2PL

i(w)- A2PL

ni(w)- A1SG on nouns

u(y)- A3

ka- C1PL

ka(w)- A1PL

w- according to some authors the prevocalic form of A1SG used on nouns (and also the older one); does not really occur in the corpus

Infix

-j- passive-deriving root infix used in transitive roots that are not derived and do not already have *-j-* as part of their root; → see also *-na*

Suffixes

-an$_1$ variant of positional suffix → *-wan* used after positional roots ending in *-r*

-an$_2$ antipassive → see also *-ma, -san, -yan*

-ar suffix to derive verbal nouns from verbs

-b'a derives transitive verbs from positional roots → see also *-b'u*

-b'ir$_1$ suffix of class 3 substantives used when they are employed without a set A prefix

-b'ir$_2$ derives verbal adjectives → see also *-em*

-b'u derives transitive verbs from positional roots → see also *-b'a*

-em derives verbal adjectives → see also *-b'ir*$_2$

-en$_1$ B1SG

[410] Another list is found in Wichmann (1999). More information on all affixes is found in the grammar sketch in Appendix E.

[411] For stem interaction especially of set A and C, but also set B and other morphemes, see chapter 6.

-en₂ imperative mood marker → see also *-V₁* and *-Vn*

-es causative suffix → see also *-se*

-er₁ possessive suffix of class 2 substantives used when they receive a set A prefix, see also → *-ir₁*

-er₂ derives abstract nouns from verbs

-et B2SG

-ib' derives instrumental nouns from verbs

-ik imperative/optative marker

-ir₁ possessive suffix of class 2 substantives used when they receive a set A prefix, see also → *-er₁*

-ir₂ derives abstract nouns from adjectives (~ English *-ness*)

=ix 'already'

=ka particle that forms yes-or-no questions

-k'a mediopassive; → see also *-pa, -tz'a, -V₁y*

-ma antipassive; → see also *-an₂, -san, -yan*

-na passive for derived transitives and root transitives that already have *-j-* as part of their root; → see also *-j-*

-ob' plural suffix, sometimes used independently (see chapter 7.4.1), generally optional

-on B1PL

-ox B2PL

-pa mediopassive; → see also *-k'a, -tz'a, -V₁y*

-r suffix that often derives verbs from adjectives (Dugan 2013: 73)

-ru repetitive aspect marker, added to the root (one action repeated serially to a single object or once over multiple objects or simultaneously over several objects): *u-tek'e* 'she struck him' ~ *u-tek'-ru* 'she struck him repeatedly' (but also possibly that she struck several persons one after the other or simultaneously) (Dugan 2013: 69)

-san antipassive suffix; → see also *-an₂, -ma, -yan*

-se causative suffix → see also *-es*

-tak plural suffix used on some nouns that denote human referents

-te' most common numeral classifier

=to 'yet, still'

-tz'a mediopassive; → see also *-k'a, -pa, -V₁y*

-V₁ imperative mood marker; → see also *-Vn* and *-en₂*

-*V₁n* detransitivizing suffix that adds "a semantic sense that the grammatical subject is emitting or producing a particular kind of sensory experience" (Dugan 2013: 76), V_1 = vowel harmonic with root

-*V₁r* derives stative participles from positional roots; V_1 = vowel harmonic with root

-*V₁y* mediopassive; → see also -*k'a, -pa, -tz'a*; V_1 = vowel harmonic with root

-*(V)n* imperative mood marker; → see also -*V₁* and -*en₂*

-*wan* derives intransitive verbs from positional roots; → see also -*an*

-*(y)aj* derives abstract nouns from verbs

-*yan* antipassive; → see also -*an₂, -ma, -san*

Appendix E – Grammatical sketch of Ch'orti'

For the convenience of the reader, I include in this dissertation my personal grammatical sketch, which is an amalgamation of all the existing literature on Ch'orti'. This sketch cannot substitute a detailed and modern descriptive grammar of Ch'orti' as it only includes the basics. Nevertheless, it may be useful to provide a quick overview of the basic features of the language to resolve any questions that might arise during the reading of this dissertation, especially to readers not familiar with Mayan.

Some features of Ch'orti' are not discussed here as they are treated extensively in the dissertation itself. These include a sketch of the phonology (found in chapter 4.1) as well as the index-sets (set A, B and C) and their use, which are first presented in chapter 4.2 and discussed extensively in the following chapters.

a. Nouns

a.1 Plural

Plural marking is often not obligatory in Mayan languages. In Ch'orti', it is also not regularly used (Hull 2016: 11). This applies to verbs in third person and nouns. The optional plural suffix is *-ob'*:

winik 'man'	*winik-ob'* 'men'
otot 'house'	*otot-ob'* 'houses'
a-patna 'he works'	*a-patno'b'* (*a-patna-ob'*) 'they work'

From my general impression of Hull (2016), it seems as though *-ob'* is used more frequently on human referents, though they also occur without it. On the other hand, non-human referents, including animals, seem to rather not mark plural, though occasionally they do. A corpus study may yield more concrete patterns.

A few nouns, which all denote human referents, have the plural suffix *-tak* instead of *-ob'* (Dugan 2013: 102). However, as *winikob'* and *uyarob'* above show, this is not obligatory for human referents. In Dugan's corpus, *-tak* appears with the following nouns:

pax 'boy'	*pax-tak* 'boys'
ijch'ok 'girl'	*ijch'ok-tak* 'girls'
ixik 'woman'	*ixik-tak* 'women'

Interestingly, *maxtak* ‘son’ or ‘child’ is, at least synchronically, a singular form that forms a plural *maxtakob’* (Dugan 2013: 102). In Hull (2016), *maxtak* ‘children, family’ is mostly used as a plural without the additional *-ob’*. Theoretically, it is likely that *maxtak* is an original plural form to *max, which was then reanalyzed as singular and now receives the plural marker *-ob’*. However, neither the earlier sources nor Kaufman’s etymological dictionary yield an immediately obvious result for *max. Since the word primarily refers to children, one may consider that PM *maax ‘monkey’ is behind this formation parallel to how children are sometimes called ‘little monkeys’ in English, but this is an ad hoc proposal.

a.2 Possession

Possession is marked by prefixing set A to nouns that are possessed (see Table 35). Just as with verbs, special preconsonantal or prevocalic forms of set A are used depending on whether the noun begins with a consonant or a vowel. As in the other index-sets, *-ob’* is optional in A3PL.[412] In Mayan languages, A1SG is different for nouns from that used with the verbs (*in- ~ inw-*; see discussion on page 67).

Table 35. Set A morphemes on nouns with examples from Hull (2016).

	_C	_V
1SG	*ni-si’* ‘my firewood’	*niw-et’ok* ‘my wife’
2SG	*a-chor* ‘your cornfield’	*aw-ijtz’in* ‘your little sister’
3SG	*u-b’aker* ‘her bones’	*uy-otot* ‘his house’
1PL	*ka-mama’* ‘our uncle’	*kaw-ej* ‘our teeth’
2PL	*i-lok’er* ‘your (pl.) entrance’	*iw-ermano*[413] ‘your (pl.) brother’
3PL	*u-patna’r-ob’* ‘their work’	*uy-ej* ‘their teeth’

According to some authors (e.g., Pérez Martínez 1994: 44–45), an allomorph *w-* exists as the prevocalic A1SG but it is difficult to find examples of it in the data. Pérez Martínez himself uses *niw-* in his examples despite naming *w-* as the regular allomorph. Since *w-* is closer to PM

[412] A corpus study may reveal some patterns as to when *-ob’* does in fact appear. Since it is used to mark both the plural of A (set A) and O (set B) on a transitive verb, it would be especially interesting to investigate which of the two arguments receives it if both A and O are plural entities. Double marking (*ob’ob’) does not seem to be attested.

[413] Prevocalic *iw-* seems to be unattested in Hull (2016), likely due to the size of the corpus and the fact that the second person plural occurs less frequently than the others. Source: Ch’orti’ New Testament San Mateo 18,35 (Wycliffe Bible Translators 2012: 52).

*w-, one may assume that this is an older form that has (almost?) disappeared in modern Ch'orti'.

In a possessive expression, the possessor follows the possessed item, which is marked with set A:

u-ti' e chiki' (A3-mouth DEF basket) 'the rim of the basket ("its mouth the basket")'
u-jor ni-noya (A3-head A1SG-grandmother) 'my grandmother's hair ("her hair my grandmother")'
u-sakinir e ab'ich (A3-urine smell DEF urine) 'the urine smell ("its urine smell the urine")'

Possession classes

Case or gender is not marked morphologically in Mayan languages[414] but there are various noun classes based on possessive marking. Although the morphology involved in this marking is not always cognate across the individual languages, they all tend to distinguish between, e.g., nouns that are usually possessed and require a special suffix when used without set A on the one hand and nouns that are never possessed and then need a special suffix when one intends to use them with set A indicating possession. In Ch'orti', too, some nouns require special derivational suffixes when they are used in their possessed or unpossessed form (Dugan 2013: 99). These are the classes that exist in Ch'orti':[415]

Class 1 Invariable substantives

[414] See, however, non-canonical cases of gender marking as described in Contini-Morava & Danziger (2018) for Mopan.

[415] Other languages, e.g., K'iche', have an additional class that changes vowel quantity when possessed called "sustantivos que cambian vocal", substantives that change vowel (López Ixcoy 1997: 102). Since Ch'orti' has lost the Proto-Mayan vowel quantity contrast, no such class exists in the language. It would be interesting to investigate whether there are any nouns that underwent the Cholan sound change of PM *ee > PCH *i and PM *oo > PCH *u that belong to this class based on comparative evidence and whether we see some traces of this (variation of *e ~ i* or *o ~ u*) is the language today.
Another class that is missing in Ch'orti' is that of suppletive noun paradigms where two unrelated words combine to an unpossessed-possessed paradigm. For K'iche', only two cases are reported (López Ixcoy 1997: 104):
k'uul 'Span. *chamarra* (jacket)' – *nu-q'uu'* 'my chamarra'
jaa 'house' – *w-ochooch* 'my house'
Hieroglyphic Maya had *nah* 'unpossessed house' – *-otot* 'possessed home' (Kettunen & Helmke 2020: 112, 114) but *nah* does not exist anymore in Ch'orti'. For the other word pair, no cognates are known from Ch'orti' either. Therefore, it is perhaps not surprising that this class is unattested in Ch'orti'.

The most basic class consist of nouns that do not change in any way, whether they are used with or without set A, e.g., *wakax* ‘livestock’ – *ni-wakax* ‘my livestock’.

Class 2 Substantives that receive suffix *-er/-ir* when possessed, often body parts

Dugan (2013: 102) was unable to determine either the semantic or phonological criteria for the class or for the use of one or the other suffix. According to Pérez Martínez (1994: 46), the substantives that belong to this class (he names *b’ak* ‘bone’, *ch’ich’* ‘blood’, *b’ajk* ‘joint (the connection between bones)’ and *chich* ‘vein’) are all body parts. The following sentence in (1) from Dugan (2013: 103) shows the use of *ch’ich’* ‘blood’, a lexeme that belongs to this class, both without and with a suffix.

(1) *Kay* *u-moroj-sy-ob’* *e* *ch’ich’,* *yi* *jaxto* *ayi*
GER A3-gather.B3-CAUS-PL DEF blood and this then
u-ch’ich’-er *e* *ka-tata’.*
A3-blood-POSS DEF A1PL-father
‘They were gathering the blood, and this was the blood of God.’

First, *ch’ich’* is used in a general sense without stating whose blood it is. When the possessor is expressed and a set A prefix used, the suffix *-er* is required. Another member of this class is *me’yn – u-me’yn-ir* ‘spirit’.

Class 3 Substantives that require a suffix *-b’ir* when unpossessed, many kinship terms

Class 3 works in exactly the opposite way in requiring a suffix, *-b’ir*, not when the noun is possessed but when it is not possessed. Many kinship terms belong to this class (Dugan 2013: 103; Pérez Martínez 1994: 46):

sakun-b’ir ‘older sibling’ *ni-sakun* ‘my older sibling’
tata-b’ir ‘father’ *u-tata / u-tata’* ‘his/her/its/their father’
tu-b’ir ‘mother’ *i-tu’* ‘your (pl.) mother’
pixam-b’ir ‘father-/mother-in-law’
sitz’-b’ir ‘grandchild’
mama’b’ir ‘uncle’
nar-b’ir ‘corncob’
ixim-b’ir ‘maize’

The inclusion of maize in a semantic field that seems to otherwise be mostly reserved for kinship terms is interesting, especially with the knowledge that maize is an extremely important part of Mayan culture – and of "Mesoamerican" culture in general. This is by no means unique to Ch'orti' – for example, K'iche' also includes maize in the respective possession class.

Class 4 Always possessed substantives

Pérez Martínez (1994: 47) provides the following examples:

u-k'anar 'yolk'
u-xejxar 'corncob stripped of its kernels'

Other always possessed nouns are *-ok* 'foot, base' or *-ar* 'child'. It is unclear whether it is possible to use them without set A but with the *-b'ir* suffix – they do not occur in unpossessed form in Hull (2016) at all. Hull (2016: 9) reports that in some cases the obligatory set A possessive marking has been reanalyzed as part of the actual lexeme: *-ar* nowadays has a variant *yar* (from *y-ar* 'his/her child'), which must be an independent noun because it can be preceded by a definite article and the combination of DEF + set A usually does not occur in Ch'orti'.

The reflexive pronoun *-b'a* also belongs in this class. In Hieroglyphic Maya, one of its meanings was 'image, self'; it was always possessed (Kettunen & Helmke 2020: 99). Today, it is only a grammatical marker: *xab'i u-b'a* 'he scratches himself'.

Class 5 Never possessed substantives

Words that are never possessed include, according to Pérez Martínez (1994: 47):

k'in 'sun'
utk'in 'sky ("face-sun")'
tokar 'cloud'
makchan 'rainbow'
katu' 'moon ("our mother")' (though see discussion below)

This class has a parallel in K'iche', where it is described in an instructive way. López Ixcoy (1997: 103–104) points out that using these nouns in a possessed form simply yields illogical

results. Most of the words from this class belong to the realm of nature and thus cannot be possessed by anyone. An exception to this is the use in poetic or ceremonial contexts, where it is indeed acceptable to refer to, e.g., 'my sun'. Another possibility to use these nouns with possessive marking is to combine them with a kinship term as a metaphor. The kinship term is then possessed and forms a compound with the nature term: KCH *q-ati't iik'* 'our grandmother moon' or KCH *qa-taat q'iij* 'our father sun'. Interestingly, the Ch'orti' term *katu'* 'moon' goes back to such a formation because it literally means *ka-tu'* 'our mother'.

Sometimes, relational nouns are discussed in the context of possession classes, e.g., in Pérez Martínez (1994: 47). Here, they are found in chapter d.2 Relational Nouns.

a.3 Compounding

Compounds can be formed either by combining an adjective with a noun or by combining two nouns, of which the second is often derived from a verb (Dugan 2013: 106).

nojjor 'leader' < *noj* 'big' + *jor* 'head'
ajyumpatna'r 'overseer, supervisor' < ajyum 'boss' (with agentive prefix *aj-*) + *patna'r* 'work' (*patna* + *-ar*).

In Hull (2016), we also find compounds of verbal roots (both intransitive and transitive) and nouns:

k'ux-e'yr (*k'ux-ej-ir*) 'toothache ("bite-tooth-ABST")'
lok'-k'in 'east ("exit-sun")'
man-ixim 'maize-buying ("buy-maize")'

b. Verbs

There are three types of verb roots in Ch'orti' (and in Mayan languages in general) – transitive, intransitive and positional. For each type of root, distinct morphology is used in inflection.

There is only limited information available on so-called "affect" verbs in Ch'orti'. They are considered to be a special root class in Mayan languages which can also derive verbs and is used "to add expressivity to speech" (Polian 2017: 218), e.g., MAM *ni'm* 'umph!' or KCH *puuq'* 'splash'. According to Polian, the concept is comparable to what is called "ideophones" in

African linguistics or "mimetics" in Japanese. I will not discuss them here because more research is needed on their morphology in Ch'orti'.

b.1 Positionals

Positional roots describe bodily positions or states. They do not constitute lexemes themselves but can be used to derive stative participles, intransitive verbs or transitive verbs (Pérez Martínez 1994: 76–77; Hull 2016).

As Table 36 demonstrates, stative participles are derived with *-V_1r*, intransitive verbs with *-wan*[416] and transitive verbs (with causative semantics) are formed with *-b'u ~ -b'a* in general, though other morphemes may be involved in forming iterative transitive verbs. In his overview of grammatical classes, Hull (2016: 19) notes that iterative causatives are formed with *-r-es ~ -l-es* but *chej-ru* in the table below demonstrates that they can also be formed with *-ru*.

Table 36. Positional paradigms based on data from Hull (2016).

Positional	Stative participle	Intransitive	Transitive
wa'-	*wa'r* (*wa'-ar*) 'standing, set up'	*wa'-wan* 'to stand'	*wa'-r-es* 'to make sth. stand up'
chej-	*che'r-* (*chej-er*) 'disorganized, out of order'	(**chej-wan – unattested)	*chej-b'u* 'to put one thing on top or above another haphazardly without any thought, put sth. out of order or in a disorganized fashion' ITER *chej-ru* 'to disorganize, place in a disorganized fashion'
jin-	*jin-ir (-jor)* 'tangled up/messed up (hair)'	*jin-wan* 'to fall over'	*jin-b'u ~ jim-b'u* 'to cut down indiscriminately, put out of order'
kot-	*kot-or* 'kneeling, on bent knee, on one's knee'	*kot-wan* 'to be kneeling, kneel down'	*kot-b'a* 'to turn upside down'
b'uk-	*b'uk-ur* 'piled up'	*b'uk-wan* 'to be piled up'	*b'uk-b'a* 'to pile up'

[416] After verb stems ending in *-r*, a variant *-an* is used (Dugan 2013: 75–76).

b.2 Intransitive and transitive suffixes

Compared to other Mayan languages, Ch'orti' is often described as having a wealth of suffixes that derive intransitive or transitive verbs from roots. In their reconstruction of Proto-Cholan, Kaufman and Norman introduced a distinction of status suffixes and "thematic" suffixes. The former are a term used in Mayan linguistics for a specific type of suffix that may contain a mixture of information such as verb type (transitive/intransitive), status (dependent/independent) and aspectual or modal nuances (Aissen, England & Zavala Maldonado 2017a: 5). The latter are a supposed innovation of Eastern Cholan and only found on intransitive verbs (Kaufman & Norman 1984: 81–82, 94).

To illustrate the difference between the two, Kaufman and Norman (1984: 94) provide two examples. First, in the intransitive verb *nijk-i* 'he/it moved', *-i* is a thematic suffix that occurs before status suffixes like the imperative *-en* in *nijkye'n* (nijk-i-en). However, the *i* is absent when a causative with *-es* is formed to the root *nijk-* as in *nijk-es* 'to cause to move'. Likewise, in *tajpa* 'to be extinguished', *-a* is a thematic suffix because stems derived from the root, e.g., the perfect participle *tajp-em* 'extinguished' or the causative *tajp-es* 'to extinguish', do not use it before their respective suffixes. In other words: thematic suffixes are taken to be something different because they are sometimes present in further derivations and sometimes not. Whether this distinction is really a useful one needs to be investigated in the future.

A thorough historical study including both the available data from Hieroglyphic Maya and comparative data from other Mayan languages will likely enable us to recognize more regular patterns and to derive the synchronic diversity from a more reduced morphology in the past. For the time being, I use Hull's (2016: 17–26) classification of verb inflectional classes without analyses as to which of them may derive from the same class historically.

b.3 Suppletive and irregular paradigms

There are some verbs in Ch'orti' which show some irregularities in their inflection and or/formation, i.e., *ixin* 'IV to go' or *ira* 'TV to see', for which Hull was apparently unable to establish an inflectional class. There is even at least one suppletive paradigm consisting of INC *watar-* and COM *tar-* (Dugan 2013: 98–99):

	INC	COM
1SG	*wate'n* 'I come'	*tarye'n* (tari-en) 'I came'
2SG	*wate't* 'you come'	*tarye't* (tari-et) 'you came'
3SG	*watar* 's/he comes'	*tari* 's/he came'
1PL	*wato'n* 'we come'	*taryo'n* (tari-on) 'we came'
2PL	*wato'x* 'you (pl.) come'	*taryo'x* (tari-et) 'you (pl.) came'
3PL	*wato'b'* 'they come'	*taryo'b'* (tari-et) 'they came'

Both verbs mean 'to come' with *watar* being used in the incompletive aspect and *tari* in the completive. Note that *watar* is used with set B suffixes to mark S despite expressing incompletive aspect. Given that *tar* is included in the form *watar*, it is possible that the form developed out of a construction consisting of *tar-* and an additional element, perhaps PROG *war* (*war tar > watar).

Another point worth mentioning is the fact that *watar* only occurs in the third person while in the other persons, the root seems to be *watV(j)-* (since the vowel of set B is laryngealized, we must be dealing with some kind of contraction (see chapter 6) and not simply the root *wat-* suffixed with set B). The variation is reminiscent of mediopassive class 4 (see chapter 7.4.2) and it must be relatively old because we already find *wat- ~ watar* in Wisdom (1950) and, more significantly, in the Morán Manuscript for Choltí with *watal ~ tali* (Robertson, Law & Haertel 2010: 91, 276). It is also highly likely that it attested in ACA <vat-el> *watel* 'to set off, leave (German "aufbrechen (Span. Despacharse")' and <tal-el> *talel* 'to come' (Smailus 1973: 173, 179). Dugan reports that one of his informants thought that some speakers use *watar* in all forms (i.e., *watar-en*, *watar-et* etc.), even when a suffix is added, but none of Dugan's informants produced such forms spontaneously.

b.4 Existential statements

Mayan languages generally do not use a copula. Instead, they form non-verbal predicates with the so-called stative construction consisting of a nominal form and an affixed set B, e.g., *winik-en* (man-B1SG) 'I am a man', *b'ixir-et* (alive-B2SG) 'you are alive'.

Formally, this corresponds to verbal inflection in the completive aspect: *patne'n* (*patna-en*: work-B1SG) 'I worked', *tarye't* (*tari-et*: come-B2SG) 'you came'.

For existential statements, the defective verb/particle *ayan* ‘it exists, there exist/are etc.’ is used as in (2). It is invariable across person, number or any other verbal category (Dugan 2013: 140).

(2) *Ayan e maxtak xe’ a-we’ me’yra.*
exist DEF child REL C3-eat much
‘There are children who eat too much.’

For negative existential statements, *matuk’a* ‘nothing’ is used, likely a compound of *ma* ‘NEG’ and *tuk’a* ‘what’ (Dugan 2013: 140):

(3) *Matuk’a e wya’r.*
matuk’a e we’-ar
exist.NEG DEF eat-NMLZ
‘There is no food.’

b.5 Aspect, tense and mood particles

For a thorough discussion of aspect in Ch’orti’, see chapters 4.2.2 and 4.3. Here, I will present special constructions involving particles (in alphabetical order) that convey a more specific meaning than simply completive or incompletive, including temporal nuances. Some of the particles “appear to specify past tense” (Dugan 2013: 43) but there is no morphological means of expressing tense.[417] The list is not exhaustive and only includes the most common cases.

ani

Ani is used after the verb that expresses the action that *ani* refers to. Hull classifies *ani* as an ‘irrealis particle’.

(4) *ani* as irrealis in conditional sentence (Hull 2016: 52)
In-k’eche-nik ani ni-tumin, in-mani ani e we’r.
A1SG-carry.B3-CNTF IRR A1SG-money A1SG-buy.B3 IRR DEF meat
‘If I had brought my money, I would have bought the meat.’

[417] Interestingly, Dugan (2013: 44) describes that those of his informants that had received some linguistic training from the PLFM, used the labels “present tense” and “past tense” instead of speaking of aspect, but he sees no clear evidence for the action being anchored in time. It is likely that this rather reflects a Spanish-influenced terminology than a conscious decision for tense over aspect.

However, it "is also used to show perfective aspect in the past or future" (Hull 2016: 53). Furthermore, it describes habitual or repeated actions that can often be translated as 'used to' or 'would' but while *kay* (see below) does not state whether the action is still occurring, *ani* emphasizes that it is long since over (Dugan 2013: 161–162):

(5) *ne'n* *in-ket-pa* *ani* *in-kojko*
I C1SG-remain-MPAS IRR A1SG-watch.B3
'I used to stay and watch them (my siblings).'

<u>*kay*</u>

Kay is analyzed by Hull as a "imperfective gerund" which can be translated by English 'was + gerund' on the one hand (6) and as a "pluperfect" (7) on the other. The latter use does not become clear from the example, which rather seems to be a perfect form. A more detailed corpus study is needed to confirm this meaning for *kay*.

(6) *kay* as imperfective gerund (Hull 2016: 192)
Kay *pok'cha* *e* *k'ajk* *tu'jor* *e* *k'ech'uj yaja'.*
GER fall.B3 DEF fire on.top.of DEF Ciguanaba
'(Balls of) fire were falling on top of that Ciguanaba[418].'

(7) *kay* as "pluperfect" (Hull 2016: 192)
Kayka yo'pe't tara?
kay=ka *yopa-et* *tara*
GER=Q come-B2SG here
'Have you come here?'

Dugan only describes *kay* as indicating a continued or repeated action similar to *war*. To Dugan, the two seem to primarily differ in tense because *kay* emphasizes that the action took place in the past. Thus, in (8), the meaning 'she <u>is</u> crying' should not be possible.

(8) *e* *ixik* *ira* *kay* *aru* *tama* *yer* *e* *ch'urkab'*
DEF woman this GER cry.B3 PREP little DEF baby
'This woman was crying (and crying) about the baby.'

[418] According to Dugan (2013: 50, fn. 2) a monster that eats babies and leads men astray in the wilderness.

k'ani + (ani) + verb

Synchronically, *k'ani* is both a productive transitive verb 'TV1 to like or love sth. or so., want' which can be used with a nominal object (9) and it can be combined with another finite (i.e., non-nominal) verb in a serial verb construction[419] as in (10).

(9) *k'ani* with object (Hull 2016: 220)

Ne'n	*in-k'ani*	*e*	*pa'.*
1SG	A1SG-like.B3	DEF	tortilla

'I like tortillas.'

(10) auxiliary *k'ani* with finite verb (serial verb construction) (Hull 2016: 220)

Ne'n	*in-k'ani*	*in-xin*	*ta*	*chinam.*
1SG	A1SG-want.B3	C1SG-go	PREP	town

'I want to go to town.'

The same serial verb construction can also be understood as a future construction as in example (11). Futures frequently develop from a verb expressing wishes or plans, e.g., the English *will* future.

(11) inflected *k'ani* with finite verb (serial verb construction) (Hull 2016: 220)

No'n	*ka-k'ani*	*ka-we'*	*b'ajxan.*
1PL	A1PL-want.B3	A1PL-eat.B3	first

'We will eat first.'

The combination of *k'ani* (here apparently in its uninflected form, i.e., a particle, as otherwise set A would have been obligatory) and *ani* yields an optative meaning, cf. example (12):

(12) *k'ani ani* (Hull 2016: 460)

K'ani ani in-tz'ojye't, Mama' Zope.

k'ani	*ani*	*in-tz'oji-et*	*Mama'*	*Zope*
want	IRR	A1SG-bother-B2SG	uncle	buzzard

'I wanted to bother you, Uncle Buzzard.'

war + (*ani*) + verb 'PROG'

[419] Mayan verbs are generally not analyzed as forming serial verb constructions. According to Quizar (1994a), this seems to be a valid analysis at least for Ch'orti' because many combinations of two finite verbs exist in the language.

Depending on which grammar you consult, *war* is used to express either a kind of present progressive (Pérez Martínez 1994: 55) or an action that is "repeated, habitual, or drawn out over a long period of time, but […] doesn't anchor that action relative to any particular reference point in time" (Dugan 2013: 160).

When used with intransitive verbs, *war* usually is combined with an incompletive form with set C, though it also occurs with set B. Of all particles, *war* is "the least one loaded with tense" (Dugan 2013: 160). Instead, it emphasizes that an action is repeated, performed habitually or drawn out over a long period of time, often describing background action in narratives. When combined with *ani*, the action that is in progress is situated in the past, cf. example (13):

(13) PROG *war* + *ani* (Hull 2016: 508)

war *ani* *a-xana*

PROG IRR C3-walk

'he was walking'

b.6 Mood

The mood system in Ch'orti' is not very rich. Aside from the indicative, which is used for actions believed to be factual but also for, e.g., probability or hopes, there is an optional subjunctive mood in *-ik* used to express exhortation, mild command, wishes or discuss hypothetical situations as in *U-b'an-ik-e'n!* 'I hope he releases me!/May he release me!' (Dugan 2013: 45–46).

Imperatives are formed either through reduplication of the root vowel (*chon-o* 'sell (it)!' or through addition of *-Vn* (often, but not always, *-en*) as in *Ch'ar-en tara!* 'Lie down here!' depending on the verb (Dugan 2013: 47). The precise conditions of the suffixal distribution need to be investigated. Curiously, Hull (2016: 10) describes *-ik* as a plural imperative: for all verbs, he lists the imperative forms, first singular and then plural, if attested. The second form commonly has a suffix *-ik.*

Wichmann (1999: 39–42) describes more precise rules, which need to be confirmed in a corpus study and compared with the imperative forms in Hull.

-ik used on specific verb classes when the verbs take "non-third person and/or non-singular subject or object" (Wichmann 1999: 39)

-n non-CVC transitive verbs

-en intransitive verbs

-V' CVC transitive verbs taking third person singular object

ø causatives – unmarked

Some irregular imperative forms are also found in Wichmann (1999: 42), e.g., of *ixin* 'to go':

kiki' 'go (sg.)!'
kiki'k 'go (pl.)!' (with suffixed *-ik*)
chik 'that he may go'

Prohibitives are formed with negation (see chapter h.3) and the regular verb form (Wichmann 1999: 43):

ira i-cham-se (NEG A2PL-kill.B3-CAUS) 'Don't kill it!' (directed at multiple people)
ma'chi i'xin (NEG C2SG-go) 'Don't go!'

The following example (14) is worth discussing:

(14) Imperative warning (Wichmann 1999: 43)
mix sutpaket
ma=ix *sutpa-ik-et*
NEG=already return-OPT-B2SG
'You won't come back (said as a warning)!'

The proposed contractions of *ma + =ix* and *sutpa + ik* do not entirely match the results of the contractions discussed in chapter 6. Though we have seen that this frequently happens when *=ix* in involved, the result of *sutpa + ik* requires an explanation – perhaps the morpheme is rather *Vk* or *k* in some cases.

b.7 Voice

b.7.1 Active voice

Compared to tense, aspect and mood and also compared to Indo-European languages, Ch'orti' – as other Mayan languages' – is relatively rich in voice categories because it not only has active and passive, but also middle and antipassive voice, often including several formations depending on verb type. Derivation always requires designated markers.

b.7.2 Passive voice

Passive voice promotes an object to the subject position with the original subject either being deleted completely or demoted to an oblique phrase outside of the core verb form, often introduced by a relational noun like *u-men* ‘by him’ (Dugan 2013: 49). Example (15) shows that passives are derived through infixation of *-j-* if the verb is underived itself and the root does not already contain a *j*; in all other cases, *-na* is used as in (16) (Dugan 2013: 50).

(15) *K’ujxa* *ayi* *u-men* *e* *k’ech’uj*[420].
eat.PASS.B3 there A3-by DEF k’ech’uj
‘It was eaten by the k’ech’uj.’

(16) *Ma’chi* *u-k’ani* *twa’* *a-lok’-es-na* *u-men*
NEG A3-want.B3 PREP C3-leave-CAUS-PASS A3-by
e *apostol-ob’.*
DEF apostle-PL
‘He didn’t want to be set free by the apostles.’

b.7.3 Mediopassive

Although both passive and mediopassive delete the subjects of transitive constructions and promote the original object into subject position, the mediopassive/middle voice is rather used

> “to reduce the concept of agency itself, not just to emphasize one role over another. Actions in the middle voice are not carried out by a specific agent, but happen of their own accord or are carried out by vague and unknown agents. In many instances, middle voice emphasizes that the agent and patient are a single entity, or at least that there is no agent other than the patient.” (Dugan 2013: 51)

In Ch’orti’, the morphemes *-k’a, -pa* or *-tz’a* are used to form the middle voice (Dugan 2013: 52). Dugan (2013: 52–53) gives the following examples, but does not specify when either of the suffixes is used:

(17) *Ma’chi* *a-pas-k’a.*
NEG C3-open-MPAS
‘It didn’t open.’

[420] According to Dugan (2013: 50, fn. 2) a monster that eats babies and leads men astray in the wilderness. Sometimes translated as Spanish *ciguanaba*.

(18) *Ka-yori* *e* *ch'en* *tya'* *a-kux-pa* *watar* *e* *k'in*.
A1PL-dig DEF hole where C3-bear-MPAS come.B3 DEF sun
'We dug a hole where the sun is born.'

(19) *Ka-yori* *e* *ch'en* *tya'* *a-nam-tz'a.*
A1PL-dig DEF hole where C3-disappear-MPAS
'We dug a hole where it disappears (sets).'

Judging from (17), *-k'a* could be a result of dissimilation of *-pa* because the verb already contains a *p*. Still, the distribution of *-pa* as in (18) and *-tz'a* as in (19) remains unclear.

Hull (2016: 22) distinguished four mediopassive classes and provides the following information:

MPAS1 *-p* mediopassive, usually involving motion (unspecified)
MPAS2 *-tz'* mediopassive (unspecified)
MPAS3 *-k* mediopassive, involving change of state
MPAS4 CVC intransitive in *-Vy*, usually involving motion

The fourth class is discussed in chapter 7.4.2. Further research is needed to determine the synchronic or diachronic reasons for the distribution of the classes. Class 4 already exists as a mediopassive in Hieroglyphic Maya (Kettunen & Helmke 2020: 68).

b.7.4 Antipassive

The function of an antipassive is to put more emphasis on an agent and the action that is carried out than on the object that is being acted upon. Although the resulting structure is morphologically intransitive, it remains semantically transitive and subject and object remain distinct from each other (Dugan 2013: 53), which distinguishes the antipassive from the mediopassive. For instance, in (20) it is not important what the man is planting specifically, just that he is. But it remains clear that he is not planting himself. Likewise, in (21) the agent and the teaching are emphasized.[421]

[421] "Since Spanish and English both allow many verbs to shift between transitive and intransitive meanings without derivational markings, translating a Ch'orti' antipassive into either of those languages often requires nothing more than using a potentially transitive verb without an object (for example, 'he teaches', rather than 'he teaches English'). In some cases, though, the semantics may require a careful choice of verb or change of voice in English or Spanish." (Dugan 2013: 55).

(20) *War a-pak'-ma in-kojt winik tara.* (Dugan 2013: 54)
PROG C3-sow-AP one-CL person there
'A man was sowing/planting there.'

(21) *U-k'eche ayi in-te' u-jun twa' a-kan-se-yan.* (Dugan 2013: 54)
A3-bring.B3 then one-CL A3-book PREP C3-learn-CAUS-AP
'He brought a book in order to teach.'

The suffixes *-o, -on, -yan* and *-ma* are used in Ch'orti' to form the antipassive (Hull 2016; Quizar 2020). Apparently, it is possible to use multiple antipassive suffixes at the same time. Example (22) with both *-ma* and *-yan* has been recorded by Dugan (2013: 74):

(22) Ch'orti' antipassive with "double" marking:
Ja'x war uyaryob'[422] *twa' axek'mayan e ciego.*
Ja'x war uy-are-ob' twa' a-xek'-ma-yan e ciego.
they PROG A3-tell- PL PREP C3-pierce-AP-AP DEF blind.person
'They were ordering the blind man to do the piercing.'

Dugan's corpus does not have sufficient examples of this kind of double marking to determine any semantic distinction.

Mayan languages commonly have an absolutive antipassive, an incorporated antipassive (in which the object is incorporated into the verb) and a special agent focus antipassive. As discussed in chapter 7.1.5, the latter is absent in modern Cholan languages. The formation discussed above is an absolutive antipassive "in which the antipassive suffix is attached to verb roots/stems to create derived intransitive verbs" (Quizar 2020: 238) for all four suffixes. Dugan (2013: 56) states that he found no evidence of an incorporated antipassive in Ch'orti' but points out that Quizar and Knowles-Berry give evidence for it. Their evidence is presented in (23) with a contrast sentence without antipassive given in (24):

(23) Ch'orti' incorporated antipassive (Quizar & Knowles-Berry 1988: 90)
e winik war a-pak-nar-i
DEF man PROG C3-double.over-cornstalks-SS
'The man is doubling over cornstalks.'

[422] We would expect a laryngealized *-o'b'* here because of the underlying vowel contraction but this might be a case of the variation in pronunciation mentioned by Hull.

(24) Ch'orti' transitive clause (Quizar & Knowles-Berry 1988: 90)

e	*winik*	*war*	*u-paki*	*e*	*nar*
DEF	man	PROG	A3-double.over.B3	DEF	cornstalks

'the man is doubling over the cornstalks'

Ch'orti' antipassive morphemes have recently been studied historically by Quizar (2020).

I suspect that some antipassive verbs have become the regular form of intransitive verbs, e.g., *wayan* 'to sleep' – historical evidence suggests that this is built upon a root *way-* with an antipassive suffix *-an* and Hull even classifies it as an antipassive. However, there is no corresponding intransitive, underived verb in the language anymore contrary to, e.g., KCH *war-ik* (with regular correspondance CHR *y* ~ KCH *r* and intransitive status suffix *-ik*).

b.7.5 Other detransitivizing suffixes

Another detransitivizing suffix that does not form a special voice (or tense, aspect or mood) according to Dugan but provides the "sense that the grammatical subject is emitting or producing a particular kind of sensory experience" (Dugan 2013: 76) is *-V_1n* with *V_1* reflecting the root vowel. Thus, *k'ux-* 'to eat' is derived to *k'uxun* 'to hurt' in the sense of 'to emit pain' and *ujtz'-* 'to perceive an odor' becomes *ujtz'un* 'to emit an aroma'. Example (25) demonstrates the use of such a form in context (Dugan 2013: 77):

(25) *k'uxun* 'IV to hurt' (Hull 2016: 245)

Ni-chek	*k'uxun*	*me'yra*	*akb'i.*
A1SG-wound	hurt.B3	very	yesterday

'My wound hurt a lot yesterday.'

This suffix looks suspiciously like an antipassive suffix (PM *-Vn, see discussion in chapter 7.1.5) and a careful historical study may reveal that it is in fact that.

b.7.6 Transitivizing suffixes

Intransitive verbs can be transitivized through a general transitivizing suffix with the forms *-b'a* or *-b'u* with the choice of either allomorph being determined lexically (Dugan 2013: 77). According to Dugan's data, this suffix can only attach to root intransitives, not derived intransitives (like, e.g., passives). Aside from examples like *u-tur-b'a* 'she seated him', there are also cases where this suffix derives transitive verbs from nouns as in *ch'ich'* 'blood' ~ *u-*

ch'ich'b'a 'he made her bleed' (Dugan 2013: 77–78). It is also used with positional roots (see section b.1).

A specifically causative suffix is *-se* or *-es*, which derives transitive stems from intransitive roots or stems or adds a sense of causation to a root/stem that is already transitive (Dugan 2013: 78). The choice of one or the other variant is apparently purely phonological with *-se* being the base form that attaches to CVC roots as in *ulok'se*. Otherwise, the form is *-es* (Dugan 2013: 78–79):

k'anpa – kak'ampes 'we need it/use it'
u-noj-r-es 's/he enlarges it'

c. Derivation

c.1 Verbs < adjectives

Aside from deriving intransitive from transitive verbs and the other way around as discussed in chapter b.7, it is also possible to derive verbs from other parts of speech.

The suffixes *-ran* or *-res* derive verbs from adjectives. More precisely, *-r-* is the derivation marker that is then combined with *-an* (antipassive?) or *-es* (causative) (Dugan 2013: 85). While *-ran* creates an intransitive verb that indicates that someone or something acquires the attribute expressed by the adjective (like *kanoj-ran* 'we get big'/'we gain weight' or *mok-ran* 's/he gets sick'), *-res* creates a transitive verb that indicates that the acquisition of the adjective characteristics by the patient has specifically been caused by an agent like *u-rax-res* 'he made it smooth' (Dugan 2013: 85).

c.2 Agentive prefix *aj-*

Aj- is an agentive prefix that is used to indicate where a person comes from, that they possess animals or that they work in a specific field[423] (Pérez Martínez 1994: 51). It is possible to summarize its meaning by 'the one from/of x':

aj-chinam 'the one from the pueblo'
aj-wakax 'the one who possesses cattle [lit. the one of the cattle]'
aj-patna'r 'worker'

[423] "1) procedencia, 2) personas que poseen animales domésticos, 3) persona trabajadora o agentivo".

Aj- can modify both nouns and adjectives (Pérez Martínez 1994: 58).

c.3 Nouns < verbs (verbal nouns)

Ch'orti' has a wide variety of derivational affixes to form verbal nouns (Hull 2016: 10). Their distribution as well as their historical origin need further study.

A suffix *-ar* is used to derive nouns from verbs. The result is a "noun that describes the typical object of a verb or a product resulting from the action of the verb" (Dugan 2013: 108) as in *patna* 'to work' > *patna'r* (*patna-ar*) 'work, task' (Dugan 2013: 109).

Another suffix, *-er*, is used to derive abstract nouns from verbs. The resulting noun refers "to an indistinct group of objects or an abstract concept" (Dugan 2013: 108), e.g., *kar* 'be drunk' > *kar-er* 'drunkenness' (Dugan 2013: 109).

The suffix *-ib'* is used to derive nouns that describe the instrument with which or the place where the action of the verb is carried out (Dugan 2013: 108), e.g., *majk'* 'to enclose (passive stem)' > *majk'ib'* 'enclosure, jail' (Dugan 2013: 109).

Finally, a suffix *-(y)aj* derives abstract nouns (*che* 'do' > *cheyaj* 'deed, action'). When used with the causative suffix *-se*, the *-yaj* nominalizer may remain a distinct syllable (*kanse* > *kanseyaj*) or merge with *-se* to either *-saj* or *-syaj*, e.g., *chamse* > *chamsaj*, *k'ajtse* 'cause to tell' > *k'ajtsyaj* 'tale, prayer' (Dugan 2013: 109).

c.4 Nouns < adjectives, nouns (abstract nouns)

The *-ir* suffix derives nouns from adjectives yielding a sense similar to English *-ness*: *mok* 'sick' > *mokir* 'sickness' (Dugan 2013: 110). It also derives nouns from nouns with a more abstract meaning "either making it refer to a less distinct instance of the noun or altering the semantic sense of the noun" (Dugan 2013: 104) as in example (26). According to Pérez Martínez (1994: 74), these abstract nouns in *-ir* are usually used in their possessed form, i.e., with set A: *u-chakchak-ir* 'the redness (lit. "its redness")'.

(26) *Kay uchob' uyototob' yaja', uchob' uyototir e katata' tya' ojronob' taka e katata'*
(Dugan 2013: 104–105)

Kay	*u-che-ob'*	*uy-otot-ob'*	*yaja',*	*u-che-ob'*	*uy-otot-ir*	*e*
GER	A3-do.B3-PL	A3-house-PL	there	A3-do.B3-PL	A3-house-ABST	DEF
ka-tata'	*tya'*	*ojron-ob'*	*taka*	*e*	*ka-tata'.*	
A1PL-father	where	speak.B3-PL	with	DEF	A1PL-father	

'They made their houses there, (and) made the house of God where they prayed with God.'

Here, the house of God is a church or a village *adoratorio* – thus a more abstract meaning of house is expressed than the home of a person.

It is likely that there is some historical connection between this abstractive *-ir* and the one employed in possessive marking (see chapter a.2 above). For a future study of Ch'orti' vowel syncope in affixed forms it is worth mentioning that it happens in *saksak* 'white' > *u-saksk-ir* 'its whiteness' but not in *k'ank'an* 'yellow' > *u-k'ank'an-ir* 'its yellowness' (Pérez Martínez 1994: 74).

c.5 Adjective < verb (verbal adjectives)

Dugan lists two verbal adjectives with the morphemes *-b'ir* and *-em*. They are sometimes called participles because their function is similar to that of Spanish past participles in *-ado* or *-ido* (like *pintado* 'painted'), although they do not participate in periphrastic verb forms compared to their Spanish counterparts (Dugan 2013: 79). They only indicate "a quality that results from the application or completion of the action" (Dugan 2013: 80) of the verb. Here are some examples from Hull (2016):

lo'pa 'MPAS1 to become loose'	–	*lo'pem* 'loose'
pochi 'TV1 to peel, shell, husk'	–	*pochem* 'peeled, husked, shucked'
		pochb'ir 'peeled, husked, shucked'
b'ut'i 'TV1 to fill, fill up'	–	*b'ut'b'ir* 'full'

As to the distribution of the two, verbs commonly occur with one or another, but there is no evidence that there is a real restriction to use any root with either suffix or that they yield different meanings (Dugan 2013: 80). As the examples above show, the same verb can in fact be used with both suffixes without a difference in semantics.

Furthermore, *-b'ir* is attached to roots without the need of a thematic suffix and both suffixes can be attached to either roots or derived stems, but they preclude any further derivation or inflection (Dugan 2013: 81). Pérez Martínez (1994: 57) also names both suffixes and also does not provide a semantic or functional difference between them. According to him, both mark the result of an already finished action.

As discussed in chapter b.1, adjectives to positional intransitive roots are formed with *-V_1r* with a reduplicated root vowel (Dugan 2013: 81). Adequate English translations are often past participles like 'stood' or gerunds like 'standing' as in *nob'-or* 'holding (with hand)' *kuch-ur* 'it was carried'. This suffix seems to be exclusively used with positional roots and also precludes further derivation and inflection (Dugan 2013: 82).

d. Prepositions

d.1 Basic prepositions

Compared to other Mayan languages,[424] Ch'orti' has a relatively rich repertoire of prepositions that are not built upon relational nouns (see following chapter), e.g.:

- *ti ~ ta* 'in, on, at, to, about, for, from, with, according to' (Hull 2016: 383, 402)
- *tama ~ tamar* 'in, on, at, with; about' (Hull 2016: 392)
- *taka ~ takar* 'with; against' (Hull 2016: 387)
- *twa'* 'from, for, of; before' with dialectal variant *kwa'* (Hull 2016: 214, 421)
- *maku'*, *makwi'r* 'inside' (Hull 2016: 270)

Although Dugan considers the possibility that these are all the basic preposition *ta-* reduced to *t-* with another morpheme, he considers them "lexicalized as independent prepositions" (Dugan 2013: 129) synchronically. Dugan further claims that *ti* is a result of a contraction of *ta* and definite article *e*. Since *ti* is etymologically the expected form,[425] this is unlikely. A variation of *ti ~ ta* is already attested for Hieroglyphic Maya (Kettunen & Helmke 2020: 122). It is likely that in Ch'orti' *ta* is simply used more frequently than the original *ti*. Otherwise, the prepositions need further study.

[424] K'iche' only has two basic prepositions *chi* and *pa*; more specific relationships between entities are expressed with relational nouns (López Ixcoy 1997: 211, 231–245).

[425] HUA *ti* (Edmonson 1988: 491), YUC *ti'* (Hofling 2017: 717), CHL *ti* (Hopkins, Josserand & Cruz Guzmán 2011: 222), KCH *chi* (López Ixcoy 1997: 211).

d.2 Relational nouns

England (1983a: 4) defines relational nouns as "a special category of always possessed nouns in Mayan languages which primarily introduce NPs showing case and locative relationships". This is more evident in other Mayan languages that build enormous paradigms around body part metaphors (e.g., K'iche', see footnote 424). There are also some cases in Ch'orti' but the group is not as big as in K'iche'.

A relational noun is a noun that consists of a set A prefix indicating who the possessor is and a noun that specifies the kind of relationship. Therefore, *u-men* 'A3 + by' is used to express the agent obliquely in passive constructions (Dugan 2013: 107). When used without an overt NP possessor and only with indexes, *-men* requires the suffix *-er*. Compare *u-men e k'ech'uj* 'by the k'ech'uj' with *u-men-er-ob'* 'by them' (Dugan 2013: 107–108).

Dugan (2013: 125) mentions that many nouns, especially body-related terms, can be used to express spatial relationships between actors in sentences, but most of them require an explicit preposition, while *-men* does not: *tujor* 'above' is literally 'at the head of' and consists of the preposition *ta*, the set A prefix *u-* and *-jor* 'head'. Another example is *t-u-pat* 'behind (literally: at the back of)' (Dugan 2013: 126). The above-mentioned *-b'a* (see chapter a.2) fulfills reflexive function. A more systematic list of relational nouns can be found in Pérez Martínez (1994: 47–49).

d.3 "Prepositional pronouns"

An interesting phenomenon that Dugan (2013: 41–42) calls "prepositional pronouns" is the use of the prepositions *taka* 'with' and *tama* 'to' with set B suffixes. From a typological point of view, this is strongly reminiscent of "conjugated prepositions" known from Celtic or Semitic languages.

The difference between these prepositions and relational nouns is that the former are stative constructions while the latter use set A prefixes indicating that they are actually possessed nouns. Interestingly, the forms imply that there was either another vowel or even a *-Vj-* in between the preposition and the set B suffixes. If they were simply affixed to *takar* or *tamar*, one would expect *takaren*, *takaret* etc. Instead, the forms are as follows:

Table 37. "Prepositional pronouns" (Dugan 2013: 42).

taka ~ takar		*tama ~ tamar*	
takare'n	with me	*tamare'n*	to me
takare't	with you (sg.)	*tamare't*	to you (sg.)
takar	with him/her/it/them	*tamar*	to him/her/it/them
takaro'n	with us	*tamaro'n*	to us
takaro'x	with you (pl.)	*tamaro'x*	to you (pl.)

However, there could be an easy explanation: Dugan does not seem to believe that the glottal stop follows regular rules and admits that the recordings he works with are not of the best possible quality. Hull mostly writes the corresponding forms without glottal stop, e.g., *takaren* 'with me' (Hull 2016: 56). It is possible that it is in fact not there.

e. Adjectives and adverbs

Neither adjectives nor adverbs are inflected in any way in Ch'orti'. Adjectives are used before the noun that they describe: *nojta' winik* 'tall man' (Pérez Martínez 1994: 73) though according to Dugan (2013: 114) adjectives and adverbs can precede or follow the words or phrases they modify. If true, a corpus study would uncover a meaningful distribution. Alternatively, the use of the adjectives after the referent may be conditioned by Spanish influence where this is the usual adjectival position.

According to Dugan (2013: 110), Ch'orti' has an intensifying suffix *in-* that is added to adjectives to give them a more intense sense akin to 'quite' or 'very': *intuj* 'very smelly', *ink'ijn* 'very hot', *inlatz* 'quite narrow'. It cannot be used on adjectives derived from verbs. A different method of intensification is reduplication: *saksak* 'very clear' (from *sak* 'white'), *chakchak* 'very red' (from *chak* 'red') (Dugan 2013: 110). This analysis of *in-* as an intensifying prefix is not confirmed in Hull (2016). Instead, it seems as if most adjectives simply include an *in-* prefix and it remains unclear how this came about.

f. Articles

The numeral *in-*, always paired with an adequate numeral classifier (e.g., *inte'* 'a thing', *inkojt* 'a being', *ingojr* 'a round thing'), can be used as an indefinite article (Dugan 2013:

116). However, *e*, usually described as a definite article, can, according to Dugan, often also be used in a sense that would correspond to indefinite in English.

When *e* and *in-* + CL are used together, the result is 'the/that one x' as in (27) (Dugan 2013: 117):

(27) *Nujb'i-Ø* *e* *in-kojt* *ijch'ok.*
marry-B3 DEF one-CL girl
'The one girl got married.'

E can refer to a whole clause, as well (Dugan 2013: 121), e.g., in (28)

(28) *taka* *e* *ja'xir* *ajk'u-na*
PREP DEF 3SG give.B3-PASS
'with what they are given'

g. Numerals

g.1 Cardinal numerals

The rich Maya vigesimal numeral system has mostly been lost in Ch'orti'. According to Pérez Martínez (1994: 52), only the following numerals are still in use, otherwise, Spanish numerals are used.

1	*inte'*	1-CL
2	*cha'te*	2-CL
3	*uxte'*	3-CL
4	*chante'*	4-CL
5	*inmojy*	1*5
10	*cha'mojy*	2*5
15	*uxmojy*	3*5
20	*chanmojy*	4*5

Pérez Martínez likewise cites an "approximately 300-year-old manuscript" that passes down Ch'orti' numerals from 1 to 10 that are cognate to other Common Mayan numerals. It is completely unclear what manuscript he is referring to because the numerals are neither identical with the first attestation of numerals from Galindo (1834) nor with those of Cholti (Robertson, Law & Haertel 2010: 58, 62, 82, 191, 330, 335). For the overview, I give K'iche' numerals as an example for the original Mayan numerals (López Ixcoy 1997: 134–135).

Numerals between 2 and 9 receive a plural suffix *-V(V)b'* instead of numeral classifiers as is the case in Cholan.

	Ch'orti'[426]	K'iche'	Galindo (1834)	Choltí
1	*in-te'*	*jun*	<Inté>	*hun-*
2	*cha'-te'*	*ka'-iib'*	<Chaté>	*cha'-*
3	*ux-te'*	*ox-iib'*	<Uxté>	*ux-*
4	*chan-te'*	*kaj-ib'*	<Chanté>	*chan-*
5	*jo'-te'*	*jo'-oob'*	<Joté>	*o'-*
6	*wak-te'*	*waq-iib'*	<Uakté>	–
7	*uk-te'*	*wuq-uub'*	<Uaxikté>[427]	*uk-* <vc>
8	*waxik-te'*	*waqxaq-iib'*	<Ukté>	*wakxik-*
9	*b'orom-te'*	*b'elej-eeb'*	<Boronté>	*b'olom-*?[428]
10	*la'in-te'*	*lajuuj*	<Launté>	*lahun-*

Note the difference between the numeral '5' that is used nowadays and the former one, *jo'te. Inmojy* is used today to form higher numerals like '10' and '15', i.e., modern Ch'orti' numerals are built upon a subbase '5'. This is highly unusual for Mayan numerals, which generally use the subbase '10'.

g.2 Ordinal numerals

Only four Ch'orti' ordinal numerals are reported by Pérez Martínez (1994: 53):

b'aj-xan	'first'
nak-pat	'second'
ux-yajr	'third or three times'
chan-yajr	'fourth'

They use different classifiers.

[426] These are the forms that the ALMG lists in the descriptive and pedagogical grammars of Ch'orti' (Morwa'r ojroner Ch'orti'/Comunidad lingüística Ch'orti' 2004: 67; Morwa'r ojroner Ch'orti'/Comunidad lingüística Ch'orti' 2008: 83). It is unclear whether these are an attempt at language revitalization or whether the numerals really survived in Ch'orti'. The latter is unlikely since Hull does not list them.

[427] Based on comparative evidence, we know that Galindo mixed up '7' and '8' in his account.

[428] The word <bolomac> is translated as 'Nahual of the Chol [Span. "Nagual de choles"]' (Robertson, Law & Haertel 2010: 335). *Bolom-* cannot be anything else than '9'.

h. Syntax

Ch'orti' syntax can be described as "clearly head first, meaning that the left-most word of a phrase defines its type" (Dugan 2013: 113). Signs for this are the use of prepositions instead of postpositions and the fact that determiners appear before their nouns. The phrase *tama e otot-ob'* 'among the houses' demonstrates both.

h.1 Word order

Though it is often stated that Ch'orti' is the only Mayan language that has generalized basic SVO word order over verb-initial word order, Dugan claims that SVO is not obligatory in Ch'orti'. Instead, the word order is relatively flexible, particularly in allowing deviating order for emphasis, fronting, subordination of clauses or even stylistic choice (Dugan 2013: 138).

h.2 Enclitics and particles

There are a couple of enclitics that can be added to the verb. The interrogative enclitic *=ka* is used to form questions.

(29) *I* *tama* *e* *ya'x* *war* *i-tz'e'ne=ka?*
and PREP DEF this PROG C2SG-laugh=Q
'And it's about this that you're laughing?'

According to Dugan (2013: 83), stress is placed on *-ka* rather than the final syllable of the preceding verb, which he takes as an indicator that the interrogative is "really attached to the verb complex".

Further clitics are aspectual in nature. The clitic *=ix* expresses a sense of 'already' and is often rendered as 'ya (= already)' in Spanish. It indicates that an action is either very recent or that it is from the past but relevant to a subsequent action (Dugan 2013: 84). The clitic *=xix* 'exactly', likely a reduplication of *=ix*, often adds a temporal sense of 'right now' or 'right away' as in *wab'u-n-xix!* (set.out-IMP-right.now) 'Set them out right now!' (Dugan 2013: 84).

The clitic *=to* 'yet, still' indicates an action that continues from the past into the present, e.g., *ka-k'an=to* (A1SG-want-still) 'we still want it' (Dugan 2013: 84).

Che is an "evidentiary particle that saturates storytelling" (Dugan 2013: 172) and marks information as hearsay. The speaker points out that he or she has no first-hand knowledge of

events. It can be translated as ‘they say’, ‘it is said’ etc. but it can also often be omitted (Dugan 2013: 173).

h.3 Negation

There are various particles that express negation in Ch’orti’ (Hull 2016). Their distribution still needs to be studied.

ma/ma’	‘no, not’
inma	‘no’
jola	‘no’
la	negative imperative
li	‘no, don’t’
ma’chi	‘no’, ‘but no’
mix	‘never’, ‘now there isn’t/aren’t’, negative future
ira + COM	negative imperative

Negation of copula sentences is often achieved by negation plus third-person independent pronoun (Dugan 2013: 141), cf. example (30):

(30) *Ma-ja’x* *ch’ok.*
NEG-3SG young
‘It isn’t new/young.’

Negation immediately precedes the element that is negated (Dugan 2013: 142) as in (31):

(31) *Tunor* *e* *akb’ar yaja’* *ma’chi* *wayan-o’n*[429].
all DEF night that NEG sleep-B1PL
‘All that night we did not sleep.’

Negation in questions is achieved with the interrogative clitic *=ka* for “yes or no” questions (Hull 2016: 184) with no change in word order (Dugan 2013: 413), e.g. in (32):

(32) *K’ani=ix=ka* *a-tak’a* *e* *pa’?*
want-already-Q C3-be.cooked DEF tamale
‘Aren’t the tamales done yet?’

[429] Based on Hull (2016), the laryngealization in the vowel is likely an error here.

h.4 Conditional clauses

Jay 'if, whether' is used to form conditional clauses. It is used directly before the condition it creates; an explicit optative marker (*-ik*) is optional (Dugan 2013: 144). In (33), *toyi* is used without *-ik*, while in (34), *ira* is used with it. The two sentences differ in whether the conditional clause is placed at the beginning or the end of the sentence. This might be a conditioning factor worth investigating when it comes to when *-ik* is used and when it is not, considering that status suffixes often disappear in Mayan languages when they are not used phrase-finally (Polian 2017: 210).

(33) Conditional clause without *-ik* (Dugan 2013: 145)

Katoye't jay ache yax.

Ka-toyi-et	*jay*	*a-che*	*yax*
A1PL-pay-B2SG	if	A2SG-do	that

'We'll pay you if you do that.'

(34) Conditional clause with *-ik* (Dugan 2013: 145)

Jay	*a-k'ani*	*inko'*	*kaw-ir-ik.*
if	A2SG-wish.B3	let's.go	A1PL-see.B3-OPT

'If you want to, let's go see it!'

Jay does not only appear before verbs but before any syntactic category. It is common before nouns and adjectives as in (35) (Dugan 2013: 145):

(35)

Jay	*tun*	*war*	*a-pak'i...*
if	stone	PROG	A2SG-plant.B3

'If it is stones you are planting, …'

h.5 Subordination

Complementizers: *ke', tuk'a, xe'*

The complementizer *ke',* clearly a loan from Spanish *que,* introduces phrases that are direct objects of verbs, usually verb of thinking, speaking or perceiving (Dugan 2013: 149), cf. example (36):

(36) Complementizer *ke'* (Dugan 2013: 150)

Aren	*ke'*	*ne'n*	*num-en*	*tya'*	*war*	*a-pak'i*	*atrigo*
say.IMP	that	1SG	pass-B1SG	when	PROG	A2SG-plant.B3	wheat

'Say that I passed (by) when you were planting your wheat.'

Subordination is also possible with the relative pronouns *xe'* and *tuk'a*. The two cannot be used interchangeably – according to Dugan, *tuk'a* is used when the NP introduced by it is the object of a verb "located in the immediately higher-level phrase" (Dugan 2013: 150) and *tuk'a* is the subject of the verb in the subordinate clause as in (37). Otherwise, *xe'* is used as in (38).

(37) Relative clause with *tuk'a* (Dugan 2013: 151)

K'ani	*in-chek-su*	*tuk'a*	*numuy (...).*
want	A1SG-appear.B3-CAUS	REL	pass.B3

'I want to reveal what happened (with those two people).'

(38) Relative clause with *xe'* (Dugan 2013: 152)

Ayan	*e*	*maxtak*	*xe'*	*a-we'*	*me'yra.*
exist	DEF	child	REL	C3-eat	much

'There are children who eat a lot.'

h.6 Light verb construction with *che*

Ch'orti' possesses a cognate construction to the one that led to split-S/fluid-S alignment in Western Cholan (see chapter 3.4.2). However, in Ch'orti', no alignment split results from this construction.

In a short subsection of his grammar, Schumann Gálvez (2007: 140–141) describes a light verb construction for Ch'orti'. Light verbs are, according to his definition, verbs "that have the structure or body of a verb that only occur to mark aspect and person of the verb that follows them".[430] The light verb that he found to be used in Ch'orti' is *che* 'to do' (although he gives the root as *ch-*), which is the same as the light verb used in Chol/Chontal split-/fluid-S. He reports that the use as a light verb is attested for only three verbs: *k'ay* 'to sing', *ajk'ut* 'to dance' and *xambar* 'to go for a walk/stroll'. A notable feature is that he describes the use of the construction in both the incompletive aspect as in (39) and in the completive one as in

[430] "[…] que tiene toda la estructura o cuerpo de un verbo, que solamente sucede para marcar el aspecto, y la persona del verbo que va después del mismo" (Schumann Gálvez 2007: 140–141).

(40) with a variant *chi*. Since *chi* is otherwise unattested (e.g., in Wisdom's ample earlier data), further research is required to confirm that this is, e.g., a dialectal form and not simply a mistake. Be that as it may, the light verb construction as such with *che* is confirmed by Hull (2016) though with other complements than described by Schumann Gálvez (see below).

(39) Incompletive light verb (Schumann Gálvez 2007: 141)

in-ch-e	*k'ay*
A1SG-do.B3-SS.INC	sing
'I sing'	

(40) Completive light verb (Schumann Gálvez 2007: 141)

in-ch-i	*k'ay*
A1SG-do.B3-SS.COM	sing
'I sang'	

Although he investigated other verbs, Schumann Gálvez did not find any others that were used in this construction. However, he states that all Spanish loan verbs are used in this way. It is striking that the construction is so similar to that of Western Cholan discussed in chapter 3.4.2 and that there, too, the construction is used to incorporate Spanish loan words into the verbal paradigm of the language (cf., e.g., Osorio May (2016: 120) with *benerar* 'to worship').

It seems that in Ch'orti', the construction never was grammaticalized further to denote either a specific aspect or +volition. According to Hull (2016), it is almost exclusively used with synchronically nominal forms, not with inflected forms. It is possible that the fact that Western Cholan incorporated nominal forms into the verbal paradigm facilitated the grammaticalization of the light verb construction with *che*, as well.

Hull (2016: 21, 25–26) identifies one intransitive and four transitive verb classes formed with this light verb *che*:

- IV5 comprised of *che* + Ch'orti' noun (7 cases)
    - *che k'in* 'IV5 to be sunny (literally: "to make sun"; compare Spanish *hacer sol*)'
    - *che nak'* 'IV5 to defecate, go poop (literally: "to make inner part")'
    - *che komo ke'* 'IV5 to act as if (Span. *hacer como si*)'

In *che komo ke'* 'to act as if', *che* does not combine with a noun but with a whole expression. The categorization of verbs as transitive or intransitive in Hull (2016) is often not historically or morphologically motivated but based on the semantics. Thus, examples like 'to be sunny' or 'to defecate' are listed as intransitive verbs even though *che* is inflected with set A for A, e.g., *u-che k'in* 'it is sunny ("it does (it) sun")'. I would instead classify this as TV15 (see below) according to the morphology.

- TV14 'transitive verb composed of *che* + a Spanish infinitive' (3 entries)
    - *che amansar* 'TV14 to tame, domesticate (Span. *amansar* 'to tame')'
    - *che faltar* 'TV14 to lack, be in need of (Span. *faltar* 'to lack' constructed with dative)'
    - *che ganar* 'TV14 to earn, win (Span. *ganar* 'to earn, gain, win')'
- TV15 'transitive verb composed of *che* + Ch'orti' noun' (20 entries)
    - *che ch'en* 'TV15 to make a hole, drill' (noun: hole)
    - *che pejk-jun* 'TV15 to read' (compound *pejka* 'TV9 to invite; call (upon)' + *jun* 'book')
    - *che chi'r* 'TV15 to sweeten' (lit. "to make sweet"; *chi'* 'sweet' + NMLZ *-ir*)
- TV16 'transitive verb composed of *che* + Ch'orti' adjective' (2 entries)
    - *che yab'ajn* 'TV16 to mix together, entwine two things' (lit. "to make mixed up" with a deverbal adjective < *yab'i* 'TV1 to mix together')
    - *che tzaj-tzaj* 'TV16 to make suffer' (lit. "to make sad")
- TV17 'transitive verb composed of *che* + Ch'orti' verb' (2 entries)
    - *che ch'am-nar* 'TV17 to harvest' (compound: *ch'ami* 'TV1 to accept, grab, grasp, take, receive' + *nar* 'corn' = *ch'am-nar* 'corn harvesting')
    - *che uk'i* 'TV17 to make cry, drive one to tears'

Hull's classification is not entirely consistent. For example, a compound consisting of a verbal root and a noun is considered TV15 in the case of *che pejk-jun* while the same kind of compound is classified TV17 in the case of *che ch'am-nar*. It is interesting that we have only

one clear case of *che* combining with another finite verb, namely *che uk'i*, which literally translated to 'he makes he cries', likely a serial verb construction.[431]

All in all, *che* seems to combine almost exclusively with nominal forms. It is striking that there are no cases of nominalized Ch'orti' verbs. All complements are originally nominal forms like nouns or adjectives. Spanish verbs are used in the infinitive, which is a nominal form, too. From the viewpoint of Ch'orti', the analysis of Spanish infinitives as nominal makes even more sense because the ending *-ar* looks like a nominalized form as this is the way Ch'orti' forms verbal nouns, e.g., *patna'r* < *patna* 'to work' + *-ar* 'verbal noun'. Still, the actual nominalized verbs in *-Vr* are not used in the light verb construction.

[431] Serial constructions in Ch'orti' are a complex and understudied phenomenon. I cannot treat them adequately here. An initial overview is found in Quizar (1994a). Dugan (2013: 146–147, 162–165) also provides some information, though the analyses are less reliable.

Bibliography

Adell, Eric. 2016a. Coronal contrasts in Chajul Ixil and the development of the Mamean Languages. Thesis. https://www.researchgate.net/publication/332870372_Coronal_Contrasts_in_Chajul_Ixil_and_the_Development_of_the_Mamean_Languages (last accessed: 2025-03-12).

Adell, Eric. 2016b. Q'anjob'alan evidence for a reevaluation of Proto-Mayan *t and *ty. Article. https://www.researchgate.net/publication/332870082_Q'anjob'alan_Evidence_for_a_Reevaluation_of_Proto-Mayan_t_and_ty (last accessed: 2025-03-12).

Aissen, Judith. 1992. Topic and focus in Mayan. *Language* 68(1). 43–80.

Aissen, Judith. 2017a. Correlates of ergativity in Mayan. In Jessica Coon, Diane Massam & Lisa Travis (eds.), *The Oxford Handbook of Ergativity*, 737–758. Oxford: Oxford University Press.

Aissen, Judith. 2017b. Information structure in Mayan. In Judith Aissen, Nora England & Roberto Zavala Maldonado (eds.), *The Mayan Languages* (Routledge Language Family Series), 293–324. London/New York: Routledge.

Aissen, Judith., Nora England & Roberto Zavala Maldonado. 2017a. Introduction. In *The Mayan Languages* (Routledge Language Family Series), 1–15. London/New York: Routledge.

Aissen, Judith., Nora England & Roberto Zavala Maldonado (eds.). 2017b. *The Mayan Languages* (Routledge Language Family Series). London/New York: Routledge.

Aldridge, Edith & Yuko Yanagida. 2021. Two types of alignment change in nominalizations. *Diachronica* 38(3). 314–357.

Academia de Lenguas Mayas de Guatemala (ALMG). 2001. *Ojronerob' Ch'orti': Vocabulario Ch'orti'*. Guatemala: K'ulb'il Yol Twitz Paxil.

Anderson, Stephen. 1988. Morphological change. In Frederick Newmeyer (ed.), *Linguistics: the Cambridge Survey*, vol. I, 324–362. Cambridge: Cambridge University Press.

Avelino, Heriberto. 2011. Preface. In Heriberto Avelino (ed.), *New Perspectives in Mayan Linguistics*, xi–xvi. Newcastle upon Tyne: Cambridge Scholars Publishing.

Avelino, Heriberto, Eurie Shin & Sam Tilsen. 2011. The phonetics of laryngealization in Yucatec Maya. In Heriberto Avelino (ed.), *New Perspectives in Mayan Linguistics*, 1–20. Newcastle upon Tyne: Cambridge Scholars Publishing.

Baerman, Matthew. 2011. Defectiveness and homophony avoidance. *Journal of Linguistics* 47(1). 1–29.

Baldischwieler, Thomas. 2004. *Antonio de Nebrija: Las introducciones latinas contrapuesto el romance al latin (1486)*. Düsseldorf: Heinrich-Heine-Universität Düsseldorf PhD Dissertation. https://docserv.uni-duesseldorf.de/servlets/DerivateServlet/Derivate-3293/1293.pdf (last accessed: 2025-03-12).

Barrett, Rusty. 2002. The Huehuetenango Sprachbund and Mayan language standardization in Guatemala. In Mary Andronis, Erin Debenport, Anne Pycha & Keiko Yoshimura (eds.), *Proceedings of the 38th Chicago linguistics society*, 309–318. Chicago: Chicago Linguistics Society.

Becquey, Cédric. 2012. Quelles frontières pour les populations cholanes ? *Ateliers d'anthropologie* 37.

Becquey, Cédric. 2014. *Diasystème, diachronie: Études comparées dans les langues cholanes*. Amsterdam: Vrije Universiteit PhD Dissertation.

Beliaev, Dmitri. 2005. Epigraphic evidence for the highland-lowland interaction in the Classic Period. Presented at the 10th European Maya Conference, Leiden.

Bennett, Patrick & Jan Sterk. 1977. South Central Niger-Congo: a reclassification. *Studies in African Linguistics* 8. 241–273.

Bennett, Ryan, Robert Henderson & Meg Harvey. 2023. Phonetic variability in the realization of glottalized stops in Uspanteko (Mayan). In Radek Skarnitzl & Jan Volín (eds.), *Proceedings of the International Congress of Phonetic Sciences (ICPhS) 2023*, 3301–3305. Prague: Guarant International.

Berendt, Carl Hermann. 1875. Arte de lengua Cacchí para bien comun / traslado de uno que tovo el Pe. Pdor. Grl. Fray Joseph Ruiz que de Dios gose San Juan Chamelco, 1741, MSS 279 Series 8 Sub-Series 2 Sub-Series 10, Box: 43, Folder: 9. William Gates papers, MSS 279. L. Tom Perry Special Collections, ms.. https://archives.lib.byu.edu/repositories/ltpsc/archival_objects/ref733_wla (last accessed: 2025-03-12).

Bergqvist, Henrik. 2011. Agentivity and status in Yukatekan languages. In Heriberto Avelino (ed.), *New Perspectives in Mayan Linguistics*, 242–256. Newcastle upon Tyne: Cambridge Scholars Publishing.

Bickel, Balthasar. 1999. Nominalization and focus constructions in some Kiranti languages. In Yogendra Yadava & Warren Glover (eds.), *Topics in Nepalese Linguistics*, 271–296. Kathmandu: Royal Nepal Academy. https://doi.org/10.5167/uzh-76651.

Bickel, Balthasar. 2011. Grammatical relations typology. In Jae Jung Song (ed.), *The Oxford Handbook of Linguistic Typology*, 399–444. Oxford: Oxford University Press.

Bickel, Balthasar, Giorgio Iemmolo, Taras Zakharko & Alena Witzlack-Makarevich. 2013. Patterns of alignment in verb agreement. In Dik Bakker & Martin Haspelmath (eds.), *Languages Across Boundaries. Studies in Memory of Anna Siewierska*, 15–36. Berlin/Boston: De Gruyter.

Bickel, Balthasar & Johanna Nichols. 2009. Case marking and alignment. In Andrej L. Malchukov & Andrew Spencer (eds.), *The Oxford Handbook of Case*, 304–321. Oxford: Oxford University Press.

Birnbaum, Henrik & Jerzy Molas. 2012 Das Polnische. In Peter Rehder (ed.), *Einführung in die slavischen Sprachen (mit einer Einführung in die Balkanphilologie)*, 147–164. 7th edn. Darmstadt: WBG.

Blevins, Juliette. 2001. Where have all the onsets gone? Initial consonant loss in Australian Aboriginal languages. In Jane Simpson, David Nash, Mary Laughren, Peter Austin & Alpher, Barry (eds.), *Forty years on: Ken Hale and Australian languages*, 481–492. Canberra: Pacific Linguistics.

Bohnemeyer, Jürgen. 2004. Split intransitivity, linking, and lexical representation: the case of Yukatek Maya. *Linguistics* 42(1). 67–107.

Boot, Erik. 2004. Vocabulary in the Ch'oltí' language. A transcription of the "Bocabulario Grande" by Fray Francisco Morán (1695). http://www.famsi.org/mayawriting/dictionary/boot/cholti_moran1695_revised.pdf (last accessed: 2025-03-12).

Boot, Erik. 2009. The Updated Preliminary Classic Maya - English, English - Classic maya Vocabulary of Hieroglyphic Readings. Mesoweb Resources. https://www.mesoweb.com/resources/vocabulary/Vocabulary-2009.01.pdf (last accessed: 2025-03-12).

Brewer, Stewart. 2009. The Ch'orti' Maya of eastern Guatemala under imperial Spain. In Brent Metz, Cameron McNeil & Kerry Hull (eds.), *The Ch'orti' Maya Area. Past and Present*, 137–147. Gainesville: University Press of Florida.

Bricker, Victoria. 1981. The source of the ergative split in Yucatec Maya. *Journal of Mayan Linguistics* 2(2). 83–127.

Bricker, Victoria. 1985. A Morphosyntactic Interpretation of Some Accession Compounds and Other Glyphs in Mayan Hieroglyphs. In Merle Greene Robertson & Elizabeth Benson (eds.), *Fourth Round Table of Palenque, 1980. Vol. 6*, 67–86. San Francisco: Pre-Columbian Art Research Institute.

Bricker, Victoria. 2004. Mayan. In Roger Woodard (ed.), *The Cambridge Encyclopedia of the World's Ancient Languages*, 1041–1070. Cambridge: Cambridge University Press.
Bricker, Victoria. 2019. *A Historical Grammar of the Maya Language of Yucatan. 1557–2000*. Salt Lake City: The University of Utah Press.
Brody, Jill. 1982. *Discourse processes of highlighting in Tojolabal Maya morphosyntax*. Saint Louis: Washington University PhD Dissertation.
Brody, Jill. 1984. Some problems with the concept of basic word order. *Linguistics* 22. 711–736.
Bybee, Joan, Revere Perkins & William Pagliuca. 1994. *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World*. Chicago/London: The University of Chicago Press.
Campbell, Lyle. 1972. Mayan loan words in Xinca. *International Journal of American Linguistics* 38(3). 187–190.
Campbell, Lyle. 1976. The linguistic prehistory of the southern Mesoamerican periphery. Las fronteras de Mesoamerica. *Fronteras de Mesoamérica. 14a Mesa Redonda* 1. 157–184.
Campbell, Lyle. 1977. *Quichean Linguistic Prehistory* (University of California Publications in Linguistics 81). Berkeley: University of California Press.
Campbell, Lyle. 1978. Quichean prehistory: Linguistic contributions. In Nora England (ed.), *Papers in Mayan Linguistics* (Miscellaneous Publications in Anthropology 6), 25–54. Columbia: University of Missouri.
Campbell, Lyle. 1979. (Review of) Pronominal Inflection in the Mayan Languages. Victoria Bricker. *American Anthropologist* 81(4). 976.
Campbell, Lyle. 1984. The implications of Mayan historical linguistics for glyphic research. In John Justeson & Lyle Campbell (eds.), *Phoneticism in Mayan Hieroglyphic Writing* (Institute for Mesoamerican Studies 9), 1–16. Albany: State University of New York.
Campbell, Lyle. 1997. Linguistic contributions to Guatemalan prehistory. In Jane Hill, P. J. Mistry & Lyle Campbell (eds.), *Papers in Honor of William Bright*, 183–192.
Campbell, Lyle. 2013. *Historical Linguistics. An Introduction*. Third. Edinburgh: Edinburgh University Press.
Campbell, Lyle. 2017. Mayan history and comparison. In Judith Aissen, Nora England & Roberto Zavala Maldonado (eds.), *The Mayan Languages* (Routledge Language Family Series), 43–61. London/New York: Routledge.
Campbell, Lyle & Una Canger. 1978. Chicomuceltec's last throes. *International Journal of American Linguistics* 44(3). 228–230.
Campbell, Lyle & Alice Harris. 2002. Syntactic reconstruction and demythologizing "Myths and the prehistory of grammars." *Journal of Linguistics* 38(3). 599–618.
Campbell, Lyle & Terrence Kaufman. 1985. Mayan linguistics: Where are we now? *Annual Review of Anthropology* 14. 187–198.
Campbell, Lyle, Terrence Kaufman & Thomas Smith-Stark. 1986. Meso-America as a linguistic area. *Language* 62(3). 530–570.
Can Pixabaj, Telma. 2017. K'iche'. In Judith Aissen, Nora England & Roberto Zavala Maldonado (eds.), *The Mayan Languages* (Routledge Language Family Series), 461–499. London/New York: Routledge.
Can Pixabaj, Telma & Nora England. 2011. Nominal topic and focus in K'ichee'. In Rodrigo Gutiérrez-Bravo, Line Mikkelsen & Eric Potsdam (eds.), *Representing Language: Essays in Honor of Judith Aissen*, 15–30. Santa Cruz: Linguistics Research Center, University of California, Santa Cruz.
Casaretto, Antje. 2020. Roots of ergativity in Africa (and beyond). *Studies in African Linguistics* 49(1). 111–140.

Coghill, Eleanor. 2016. *The Rise and Fall of Ergativity in Aramaic: Cycles of Alignment Change*. Oxford: Oxford University Press.
Commission for Historical Clarification (CEH). 1999. *Guatemala: Memory of Silence. Tz'inil na'tab'al. Report of the Commission for Historical Clarification. Conclusions and Recommendations*. Guatemala: Historical Clarification Commission.
Comrie, Bernard. 1973. The ergative: Variations on a theme. *Lingua* 32. 239–253.
Comrie, Bernard. 1978. Ergativity. In Winfred Lehmann (ed.), *Syntactic typology: Studies in the phenomenology of language*, 329–394. Hassocks, Sussex: The Harvester Press.
Comrie, Bernard. 2013. Ergativity: Some recurrent themes. In Edith Bavin & Sabine Stoll (eds.), *The Acquisition of Ergativity* (Trends in Language Acquisition Research 9), 15–34. Amsterdam: John Benjamins.
Contini-Morava, Ellen & Eve Danziger. 2018. Non-canonical gender in Mopan Maya. In Sebastian Fedden, Jenny Audring & Greville Corbett (eds.), *Non-Canonical Gender Systems*, 129–146. Oxford: Oxford University Press.
Coon, Jessica. 2008. When ergative = genitive. Nominals and split ergativity. In Natasha Abner & Jason Bishop (eds.), *Proceedings of the 27th West Coast Conference on Formal Linguistics*, 99–107. Somerville: Cascadilla Proceedings Project.
Coon, Jessica. 2010a. Rethinking split ergativity in Chol. *International Journal of American Linguistics* 76. 207–253.
Coon, Jessica. 2010b. *Complementation in Chol (Mayan): A theory of split ergativity*. Cambridge, MA: Massachusetts Institute of Technology PhD Dissertation.
Coon, Jessica. 2012. Split ergativity and transitivity in Chol. *Language* 122. 241–256.
Coon, Jessica. 2013. *Aspects of Split Ergativity* (Oxford Studies in Comparative Syntax). Oxford: Oxford University Press.
Coon, Jessica. 2017. Ch'ol. In Judith Aissen, Nora England & Roberto Zavala Maldonado (eds.), *The Mayan Languages* (Routledge Language Family Series), 648–684. London/New York: Routledge.
Coon, Jessica, Diane Massam & Lisa Travis (eds.). 2017. *The Oxford Handbook of Ergativity* (Oxford Handbooks). Online edition, Oxford Academic, 10 Aug. 2017.
Cortés y Larraz, Pedro. 1771. *Descripción Geográfico-Moral de la Diócesis de Goathemala*. Vol. 1. 3 vols. https://es.wikisource.org/wiki/Descripci%C3%B3n_Geogr%C3%A1fico-Moral_de_la_Di%C3%B3cesis_de_Goathemala (last accessed: 2025-03-12).
Craig, Colette. 1977. *Jacaltec. The Structure of Jacaltec*. Austin/London: University of Texas Press.
Cristofaro, Sonia. 2013. The referential hierarchy: reviewing the evidence in diachronic perspective. In Dik Bakker & Martin Haspelmath (eds.), *Languages Across Boundaries. Studies in Memory of Anna Siewierska*, 69–94. Berlin/Boston: De Gruyter.
Culhane, Kirsten, Naomi Peck & Uta Reinöhl. 2023. The mystery of word-initial consonant loss in Kera'a. SocArXiv. https://doi.org/10.31235/osf.io/zvfdw (last accessed: 2025-03-12).
Curiel Ramírez del Prado, Alejandro. 2017. Tojolabal. In Judith Aissen, Nora England & Roberto Zavala Maldonado (eds.), *The Mayan Languages* (Routledge Language Family Series), 570–609. London/New York: Routledge.
Dahl, Eystein (ed.). 2021. Diachronic dimensions of alignment typology. *Special Issue of Diachronica* 38(3).
Dakin, Karen, Claudia Parodi & Natalie Operstein (eds.). 2017. *Language Contact and Change in Mesoamerica and Beyond* (Studies in Language Companion Series 185). Amsterdam/Philadelphia: John Benjamins.

Danziger, Eve. 1996. Split intransitivity and active-inactive patterning in Mopan Maya. *International Journal of American Linguistics* 62(4). 379–414.
Datz, Margaret. 1980. *Jacaltec syntactic structures and the demands of discourse*. Boulder: University of Colorado PhD Dissertation.
Davletshin, Albert. 2013a. Historical linguistics helps to read Maya glyphs. "Advanced workshop 'Grammar of Hieroglyphic Maya' Brussels, October 29–31 2013" workbook, 69–77.
Davletshin, Albert. 2013b. Hieroglyphic Mayan interlinear glossed texts. "Advanced workshop 'Grammar of Hieroglyphic Maya' Brussels, October 29–31 2013" workbook, 78-. Brussels.
Dayley, Jon. 1981a. Voice and ergativity in Mayan languages. *Journal of Mayan Linguistics* 2(2). 3–82.
Dayley, Jon. 1981b. *Tzutujil Grammar*. Berkeley: University of California PhD Dissertation.
Dayley, Jon. 1982. A hard Fought repartee. *Journal of Mayan Linguistics* 3(2). 38–52.
Dayley, Jon. 1983. Voice and ergativity in Mayan languages. In Alice Schlichter, Wallace Chafe & Leanne Hinton (eds.), *Studies in Mesoamerican Linguistics: Report No 4*, 5–119. Survey of California and Other Indian Languages.
Dayley, Jon. 1985. *Tzutujil Grammar* (University of California Publications in Linguistics 107).
De Kind, Jasper, Sebastian Dom, Gilles-Maurice de Schryver & Koen Bostoen. 2015. Event-centrality and the pragmatics-semantics interface in Kikongo: From predication focus to progressive aspect and vice versa. *Folia Linguistica* 49(36). 113–163.
De Smet, Isabeau & Laura Rosseel. 2021. Who's afraid of homophones? A multimethodological approach to homophony avoidance. *Language and Cognition: An interdisciplinary journal of language and cognitive science* 13(3). 1–24.
DeLancey, Scott. 1985. On active typology and the nature of agentivity. In Frans Plank (ed.), *Relational Typology*, 47–60. Berlin: Mouton.
DeLancey, Scott. 2005. The blue bird of ergativity. In Francesc Queixalós (ed.), *Ergativity in Amazonia III (Proceedigs of the Workshop on "Ergatividade na Amazônia", Centre d'Études des Langues Indigènes d'Amérique (CELIA))*, 1–15. Paris: Centre Nacional de la Recherche Scientifique.
Dixon, Robert. 1972. *The Dyirbal Language of North Queensland* (Cambridge Studies in Linguistics 9). Cambridge: Cambridge University Press.
Dixon, Robert. 1979. Ergativity. *Language* 55(1). 59–138.
Dixon, Robert. 1994. *Ergativity* (Cambridge Studies in Linguistics 69). Cambridge: Cambridge University Press.
Donohue, Mark & Søren Wichmann (eds.). 2008. *The Typology of Semantic Alignment*. Oxford: Oxford University Press.
DuBois, John. 1987. The discourse basis of ergativity. *Language* 63(4). 805–855.
DuBois, John. 2017. Ergativity in discourse and grammar. In Coon, Jessica, Diane Massam & Lisa Travis (eds.). *The Oxford Handbook of Ergativity*, 23–58. Oxford: Oxford University Press.
Dugan, James. 2013. *The grammar of Ch'orti' Maya folktales*. New Orleans: Tulane University PhD Dissertation.
Ebert, Karen. 1997. *A Grammar of Athpare*. München: LINCOM.
Edmonson, Barbara. 1988. *A descriptive grammar of Huastec (Potosino dialect)*. New Orleans: Tulane University PhD Dissertation.
England, Nora. 1983a. Ergativity in Mamean (Mayan) languages. *International Journal of American Linguistics* 49. 1–19.
England, Nora. 1983b. *A Grammar of Mam, a Mayan language*. Austin: University of Texas Press.

England, Nora. 1991. Changes in basic word order in Mayan languages. *International Journal of American Linguistics* 57(4). 446–486.
England, Nora. 1992. *La autonomía de los idiomas Mayas. Historia e identidad*. Guatemala: Cholsamaj.
England, Nora. 2013. Marking aspect and mood and inferring time in Mam (Mayan). *Berkeley Linguistics Society* 33. 119–140.
Everett, Caleb. 2009. A reconsideration of the motivations for preferred argument structure. *Studies in Language* 33. 1–24.
Feldman, Lawrence. 2009. Some data and reflections on the demographic dynamism and continuity of the colonial Ch'orti' population: The many Copans and San Juan Ermita. In Brent Metz, Cameron McNeil & Kerry Hull (eds.), *The Ch'orti' Maya Area. Past and Present*, 148–156. Gainesville: University Press of Florida.
Fernández Guardia, León. 1883. *Coleccion de documentos para la historia de Costa Rica. Volume 1*. Barclona: Imprenta Nacional. https://bibliotecadigital.aecid.es/bibliodig/es/consulta/registro.cmd?id=559 (last accessed: 2023-07-02).
Fisher, William. 1973. *Towards the reconstruction of proto-Yucatec*. Chicago: University of Chicago PhD Dissertation.
Fought, John. 1967. *Chorti (Mayan) Phonology, Morphophonemics, and Morphology*. New Haven: Yale University PhD Dissertation.
Fought, John. 1972. *Chortí (Mayan) Texts 1*. (Ed.) Sarah Fought. Philadelphia: University of Pennsylvania Press.
Fought, John. 1982. A note on voice and ergativity in Chortí. *Journal of Mayan Linguistics* 3(2). 33–37.
Fought, John. 1984. Choltí Maya: A sketch. In Victoria Bricker & Munro Edmonson (eds.), *Supplement to the Handbook of Middle American Indians*, vol. 2, 43–55. Austin: University of Texas Press.
Fought, John. 2000. Comments on "The Language of Classic Maya Inscriptions" by Stephen Houston, John Robertson and David Stuart. 2000. Current Anthropology 41(3), 321–338. *Current Anthropology* 43(3). 339.
François, Alexandre. 2014. Trees, waves and linkages: Models of language diversification. In Claire Bowern & Bethwyn Evans (eds.), *The Routledge Handbook of Historical Linguistics*, 161–189. London/New York: Routledge.
Frazier, Melissa. 2011. Tonal Dialects and Consonant-Pitch Interaction in Yucatec Maya. In Heriberto Avelino (ed.), *New Perspectives in Mayan Linguistics*, 21–55. Newcastle upon Tyne: Cambridge Scholars Publishing.
Frellesvig, Bjarke. 2010. *A History of the Japanese Language*. Cambridge: Cambridge University Press.
Fries, Simon, Jakob Halfmann, Eugen Hill & Denise Hübner. 2023. From noun to future tense: The functional diachrony of the *l*-future in the Nuristani languages and its typological background. *STUF - Language Typology and Universals* 76(1). 53–85.
Galindo, Juan. 1834. Copan, le 19 Juin 1834. In *Mémoires, lettres, cartes, plans, dessins et objets d'antiquités adressés à la Société de géographie par M. le colonel Juan Galindo [1831–1843]*, 94–129 [pdf 103–138], ms. https://gallica.bnf.fr/ark:/12148/btv1b525153172 (last accessed: 2025-03-12).
García de Palacio, Diego. 1576. *Descripción de la Provincia de Guatemala*. https://es.wikisource.org/wiki/Descripci%C3%B3n_de_la_Provincia_de_Guatemala (last accessed: 2025-03-12).
Garrett, Andrew. 1990. The origin of NP split ergativity. *Language* 66. 261–296.
Gildea, Spike. 1992. *Comparative Cariban morphosyntax: On the genesis of ergativity in independent clauses*. Eugene: University of Oregon PhD Dissertation.

Gildea, Spike. 2004. Are there universal cognitive motivations for ergativity? In Francesc Queixalós (ed.), *L'ergativité en Amazonie*, vol. 2, 1–37. Brasília: CNRS, IRD and the Laboratório de Línguas Indígenas, UnB.

Gildea, Spike & Francesc Queixalós (eds.). 2010. *Ergativity in Amazonia* (Typological Studies in Language 89). Amsterdam/Philadelphia: John Benjamins.

Girard, Rafael. 1949a. *Los chortis ante el problema maya. Historia de las culturas indígenas de America, desde su origen hasta hoy*. V vols. Mexico: Antigua Librería Robredo.

Girard, Rafael. 1949b. *Los chortis ante el problema maya. Historia de las culturas indígenas de America, desde su origen hasta hoy*. Vol. Tomo I. Mexico: Antigua Librería Robredo.

Girard, Rafael. 1977. *Origen y desarrollo de los civilizaciones antiguas de América*. Mexico City: Editores Mexicanos Unidos.

Gould, Stephen. 1983. *Hen's Teeth and Horse's Toes. Further reflections in Natural History*. New York: Norton.

Grinevald, Colette & Marc Peake. 2012. Ergativity in Mayan languages: a functional-typological approach. In Katharina Haude & Gilles Authier (eds.), *Ergativity, Valency and Voice*, 15–50. Berlin: De Gruyter.

Grjunberg, A. L. 1980. *Jazyk Kati. Teksty, grammatičeskij očerk*. Moscow: Nauka.

Gronemeyer, Sven. 2014. *The Orthographic Conventions of Maya Hieroglyphic Writing: Being a Contribution to the Phonemic Reconstruction of Classic Mayan*. Bundoora, Australia: La Trobe University Dissertation.

Grube, Nikolai. 1998. Speaking through stones: A quotative particle in Maya hieroglyphic inscriptions. In Sabine Dedenbach-Salazar Sáenz, Carmen Arellano Hoffmann, Eva König & Heiko Prümers (eds.), *50 Años de Estudios Americanistas en la Universidad de Bonn: Nuevas contribuciones a la arqueología, etnohistoria, etnolingüística y etnografía de las Américas / 50 Years Americanist Studies at the University of Bonn: New Contributions to the Archaeology, Ethnohistory, Ethnolinguistics and Ethnography of the Americas* (Bonner Amerikanistische Studien (BAS) 30), 543–558. Markt Schwaben: Verlag Anton Saurwein.

Grube, Nikolai. 2000. On Classic Maya inscriptions. *Current Anthropology* 41(5). 837–838.

Grube, Nikolai. 2004. The orthographic distinction between velar and glottal spirants in Maya Hieroglyphic writing. In Søren Wichmann (ed.), *The Linguistics of Maya Writing*, 61–82. Salt Lake City: The University of Utah Press.

Grube, Nikolai & Maria Gaida. 2006. *Die Maya: Schrift und Kunst*. Berlin/Cologne: SMB-DuMont.

Güldemann, Tom. 2003. Present progressive vis-à-vis predication focus in Bantu. *Studies in Language* 27. 323–360.

Güldemann, Tom. 2010. The relation between focus and theticity in the Tuu family. In Ines Fiedler & Anne Schwarz (eds.), *The Expression of Information Structure. A documentation of its diversity across Africa* (Typological Studies in Language 91), 69–93. Amsterdam/Philadelphia: John Benjamins.

Gutiérrez-Bravo, Rodrigo. 2011. External and internal topics in Yucatec Maya. In Rodrigo Gutiérrez-Bravo, Line Mikkelsen & Eric Potsdam (eds.), *Representing Language: Essays in Honor of Judith Aissen*, 105–119. Santa Cruz: Linguistics Research Center, University of California, Santa Cruz.

Gutiérrez-Bravo, Rodrigo & Jorge Monforte y Madera. 2011. Focus, agent focus and relative clauses in Yucatec Maya. In Heriberto Avelino (ed.), *New Perspectives in Mayan Linguistics*, 257–274. Cambridge: Cambridge Scholars Publishing.

Haig, Geoffrey. 2008. *Alignment Change in Iranian Linguistics. A Construction Grammar Approach* (Empirical Approaches to Language Typology [EALT] 37). Berlin/New York: De Gruyter Mouton.

Haig, Geoffrey & Stefan Schnell. 2016. The discourse basis of ergativity revisited. *Language* 92(3). 591–618.
Halfmann, Jakob. 2024. *A Grammatical Description of the Katë Language (Nuristani)*. Cologne: University of Cologne PhD Dissertation. https://kups.ub.uni-koeln.de/73732/ (last accessed: 2025-03-12).
Halpern, Abraham. 1942. A theory of Maya ts-sounds. *Notes in Middle American Archaeology and Ethnology* 13. 51–62.
Harris, Alice. 1991. Mingrelian. In Alice Harris (ed.), *The Indigenous Languages of the Caucasus. Vol. 1: The Kartvelian Languages*, 313–394. Delmar: Caravan Books.
Harris, Alice. 1997. Review of Ergativity. By R.M.W. Dixon (Cambridge Studies in Linguistics, 69). Cambridge: Cambridge University Press, 1994. *Language* 73(2). 359–374.
Harris, Alice. 2008. On the explanation of typologically unusual structures. In Jeff Good (ed.), *Linguistic Universals and Language Change*, 54–76. Oxford: Oxford University Press.
Harris, Alice & Lyle Campbell. 1995. *Historical Syntax in Cross-Linguistic Perspective* (Cambridge Studies in Linguistics 74). Cambridge: Cambridge University Press.
Haspelmath, Martin. 1998. The semantic development of old presents: New futures and subjunctives without grammaticalization. *Diachronica* 15(1). 29–62.
Haspelmath, Martin. 2011. On S, A, P, T, and R as comparative concepts for alignment typology. *Linguistic Typology* 15(3). 535–567.
Haspelmath, Martin. 2013. Argument indexing: A conceptual framework for the syntactic status of bound person forms. In Dik Bakker & Martin Haspelmath (eds.), *Languages Across Boundaries. Studies in Memory of Anna Siewierska*, 197–226. Berlin/Boston: De Gruyter.
Herranz, Atanasio. 2000. *Estado, sociedad y lenguaje. La política lingüística en Honduras*. Tegucigalpa: Editorial Guaymuras.
Hill, Eugen. 2020. Analogy in inflectional change: Modification or whole-word replacement? *Language* 96(1). e38–e58.
Hill, Eugen, Daniel Kölligan, Corinna Scheungraber & Michael Frotscher. 2019. The development of prefixation in time and space. *Transactions of the Philological Society* 117(2). 157–198.
Hofling, Charles. 2000a. *Itzaj Maya grammar*. Salt Lake City: The University of Utah Press.
Hofling, Charles. 2000b. Comments on "The Language of Classic Maya Inscriptions" by Stephen Houston, John Robertson and David Stuart. 2000. Current Anthropology 41(3), 321–338. *Current Anthropology* 43(3). 339–341.
Hofling, Charles. 2006. A Sketch of the History of the Verbal Complex in Yukatekan Mayan Languages. *International Journal of American Linguistics* 72(3). 367–396.
Hofling, Charles. 2017. Comparative Maya. In Judith Aissen, Nora England & Roberto Zavala Maldonado (eds.), *The Mayan Languages* (Routledge Language Family Series), 685–759. London/New York: Routledge.
Hofling, Charles. 2018. Cambio diacrónico en la familia lingüística yucatecana. The diachronic change of the Yucatecan linguistic family. *Cuadernos de lingüística de El Colegio de México* 5(1). 6–46.
Holtmann, Momme. 2022. On agent focus agreement in Proto-Mayan. *LIAMES* 22. 1–15, e022009.
Holtmann, Momme. 2023. The position of set B markers in Proto-Mayan. *Cuadernos de lingüística de El Colegio de México* 10. e278. https://doi.org/10.24201/clecm.v10i00.278 (last accessed: 2025-03-12).

Hopkins, Nicholas, Kathryn Josserand & Ausencio Cruz Guzmán. 2011. *A Historical Dictionary of Chol (Mayan): The Lexical Source from 1789 to 1935*. Tallahassee: Jaguar Tours.
Hopper, Paul 1979. Aspect and foregrounding in discourse. In Talmy Givón (ed.), *Discourse and Syntax* (Syntax and Semantics), vol. 12, 213–241. New York/San Francisco/London: Academic Press.
Hopper, Paul & Sandra Thompson. 1980. Transitivity in grammar and discourse. *Language* 56(2). 251–299.
Houston, Stephen. 1997. The shifting now: Aspect, deixis, and narrative in Classic Maya texts. *American Anthropologist* 99(2). 291–305.
Houston, Stephen. 2017. Puzzle Writing. Academic site. *Maya decipherment. Ideas on Maya writing and iconography – Boundary End Archaeological Research Center*. https://mayadecipherment.com/2017/02/13/puzzle-writing (last accessed: 2025-03-12).
Houston, Stephen, Barbara MacLeod, Peter Mathews, David Stuart & Marc Zender. 2004. Commentary on Kerr 8885. https://www.mayavase.com/com8885.html (last accessed: 2025-03-12).
Houston, Stephen, John Robertson & David Stuart. 1998. The language of Classic Maya inscriptions. Unpublished paper. Brigham Young University/Harvard University.
Houston, Stephen, John Robertson & David Stuart. 2000a. The language of Classic Maya inscriptions. *Current Anthropology* 41(3). 321–356.
Houston, Stephen, John Robertson & David Stuart. 2000b. Reply to comments on "The Language of Classic Maya Inscriptions" by Stephen Houston, John Robertson and David Stuart. 2000. Current Anthropology 41(3), 321–338. *Current Anthropology* 41(3). 346–356.
Houston, Stephen, David Stuart & John Robertson. 2004. Disharmony in Maya hieroglyphic writing: Linguistic change and continuity in Classic society. In Søren Wichmann (ed.), *The Linguistics of Maya Writing*, 83–99. Salt Lake City: University of Utah Press.
Hruby, Zachary & Mark Child. 2004. Chontal linguistic influence in ancient Maya writing: Intransitive positional verbal affixation. In Søren Wichmann (ed.), *The Linguistics of Maya Writing*, 13–26. Salt Lake City: The University of Utah Press.
Hull, Kerry. 2003. *Verbal art and performance in Ch'orti' and Maya Hieroglyphic writing*. Austin: University of Texas PhD Dissertation.
Hull, Kerry. 2016. *A Dictionary of Ch'orti' Mayan-Spanish-English*. Salt Lake City: University of Utah Press.
Hull, Kerry. 2018. Semantic generalization in Ch'orti' Mayan numeral classifiers. In William McGregor & Søren Wichmann (eds.), *The Diachrony of Classification Systems*. Amsterdam/Philadelphia: John Benjamins.
Hull, Kerry, Michael Carrasco & Robert Wald. 2009. The first-person singular independent pronoun in Chlassic Ch'olan. *Mexicon* XXXI(2). 36–43.
Hyman, Larry (ed.). 1979. *Aghem Grammatical Structure* (Southern California Occasional Papers in Linguistics No. 7). Los Angeles: University of Southern California.
Hyman, Larry & John Watters. 1984. Auxiliary focus. *Studies in African Linguistics* 15. 233–273.
Jones, Grant. 1998. *The Conquest of the Last Maya Kingdom*. Stanford: Stanford University Press.
Jügel, Thomas. 2015. *Die Entwicklung der Ergativkonstruktion im Alt- und Mitteliranischen. Eine korpusbasierte Untersuchung zu Kasus, Kongruenz und Satzbau* (Iranica 21). Wiesbaden: Harrassowitz.
Karttunen, Frances. 1992. *An Analytical Dictionary of Nahuatl*. Norman: University of Oklahoma Press.

Kaufman, Terrence. 1964. Materiales lingüísticos para el estudio de las relaciones internas y externas de la familia de idiomas Mayanos. In Evon Vogt (ed.), *Desarrollo cultural de los mayas*, 81–136. Mexico: Centro de Estudios Mayas.
Kaufman, Terrence. 1969. Teco – a new Mayan language. *International Journal of American Linguistics* 35. 154–174.
Kaufman, Terrence. 1972. *El Proto-Tzeltal-Tzotzil: Fonología comparada y diccionario reconstruido* (Cuaderno 5). Mexico: Centro de Estudios Mayas.
Kaufman, Terrence. 1975. New Mayan languages in Guatemala: Sacapultec, Sipacapa, and others. *Mayan Linguistics* 1. 67–89.
Kaufman, Terrence. 1976. Archaeological and linguistic correlations in Mayaland and associated areas of Meso-America. *World Archaeology* 8(1). 101–118.
Kaufman, Terrence. 1990. Algunos rasgos estructurales de los idiomas mayances con referencia especial al K'iche'. In Nora England & Stephen Elliott (eds.), *Lecturas sobre la lingüística maya*, 59–114. Antigua Guatemala: CIRMA.
Kaufman, Terrence. 1991. Notes on the structure of Yukateko and other Yukatekan languages. Unpublished manuscript.
Kaufman, Terrence 2003. A preliminary Mayan etymological dictionary by Terrence Kaufman with the assistance of John Justeson. October 5, 2003.
Kaufman, Terrence. 2015. Mayan Comparative Studies. Albany, New York (Institute for Mesoamerican Studies). Unpublished manuscript.
Kaufman, Terrence. 2017. Lexicon of Proto-Mayan. In Judith Aissen, Nora England & Roberto Zavala Maldonado (eds.), *The Mayan Languages* (Routledge Language Family Series), 62–111. London/New York: Routledge.
Kaufman, Terrence & William Norman. 1984. An outline of Proto-Cholan phonology, morphology and vocabulary. In John Justeson & Lyle Campbell (eds.), *Phoneticism in Mayan Hieroglyphic Writing*, 77–166. Albany: Institute for Mesoamerican Studies.
Keller, Kathryn & Plácido Luciano G. 1997. *Diccionario Chontal de Tabasco*. Tucson: Summer Institute of Linguistics.
Kelly, Mary Kate. 2017. The Scribe's hand betrays his tongue: Diglossia among the Ancient Maya. In *SALSA XXV Proceedings* (Texas Linguistics Forum), vol. 60. Austin: University of Texas Press.
Kettunen, Harri & Christophe Helmke. 2020. Introduction to Maya Hieroglyphs. Seventeenth Revised Edition. Wayeb. https://www.mesoweb.com/resources/handbook/IMH2020.pdf (last accessed: 2025-03-13).
Klaiman, M. H. 1992. Inverse languages. *Lingua* 88. 227–261.
Klamer, Marian. 2000. Continuative aspect and the dative clitic in Kambera. In Ileana Paul, Vivianne Phillips & Lisa Travis (eds.), *Formal Issues in Austronesian Linguistics* (Studies in Natural Language and Linguistic Theory 49), 49–63. Dordrecht/Boston/London: Kluwer Academic Publishers.
Knowles, Susan. 1984. *A descriptive grammar of Chontal Maya (San Carlos dialect)*. New Orleans: Tulane University PhD Dissertation.
Kondic, Ana. 2011. Middle Voice in South Eastern Huastec. In Heriberto Avelino (ed.), *New Perspectives in Mayan Linguistics*, 114–143. Newcastle upon Tyne: Cambridge Scholars Publishing.
Kondic, Ana. 2012. *A grammar of South Eastern Huastec, a Mayan language from Mexico*. Sydney/Lyon: University of Sydney/Université Lyon 2 Lumière PhD Dissertation. https://ses.library.usyd.edu.au/handle/2123/29258?show=full (last accessed: 2025-03-12).
Korn, Agnes. 2009. The ergative system in Balochi from a typological perspective. *Iranian Journal of Applied Language Studies* 1. 43–79.

Korobzow, Natalie. 2022. Directions of alignment change. Presented at the International Conference on Historical Linguistics 1–5 August 2022 (ICHL25), Oxford.
Korobzow, Natalie. 2023. The significance of open data in Historical-Comparative Linguistics. Presented at the Lautwandel und morphologische Analogie (IG-AT2023) 12–14 September 2023, Cologne.
Korovina, Evgeniya. 2019. Typescript of the Ch'orti' dictionary in Wisdom 1950. https://www.academia.edu/38223647/Typescript_of_the_Chorti_dictionary_in_Wisdom_1950 (last accessed: 2025-03-12).
Korovina, Evgeniya. 2021. Preliminary typescript of the Ch'orti' dictionary in Ruano Suarez 1892 Vocabularios de los idiomas Chiorti y Pocomane. https://www.academia.edu/71155169/Preliminar_typescript_of_the_Chorti_dictionary_in_Ruano_Suarez_1892_Vocabularios_de_los_idiomas_Chiorti_y_Pocomane (last accessed: 2025-03-12).
Kuteva, Tania, Bernd Heine, Bo Hong, Haiping Long, Heiko Narrog & Seongha Rhee. 2019. *World Lexicon of Grammaticalization. Second, extensively revised and updated edition*. Cambridge: Cambridge University Press.
Lacadena, Alfonso. 2000. Antipassive constructions in the Maya glyphic texts. *Written Language and Literacy* 3(1). 155–180.
Lacadena, Alfonso. 2011. Mayan hieroglyphic texts as linguistic sources. In Heriberto Avelino (ed.), *New Perspectives in Mayan Linguistics*, 343–373. Newcastle upon Tyne: Cambridge Scholars Publishing.
Lacadena, Alfonso. 2013. Basic grammar of Hieroglyphic Maya. "Advanced workshop 'Grammar of Hieroglyphic Maya' Brussels, October 29–31 2013" workbook, 5–68.
Lacadena, Alfonso & Albert Davletshin. 2013. "Advanced workshop 'Grammar of Hieroglyphic Maya' Brussels, October 29–31 2013" workbook.
Lacadena, Alfonso & Søren Wichmann. 2004. On the representation of the glottal stop in Maya writing. In Søren Wichmann (ed.), *The Linguistics of Maya Writing*, 100–162. Salt Lake City: The Universty of Utah Press.
Lardé y Larín, Jorge. 1955. *Monografías históricas del departamento de Santa Ana*. San Salvador: Ministerio del Interior.
Larsen, Thomas. 1988. *Manifestations of ergativity in Quiché grammar*. Berkeley: University of California PhD Dissertation.
Larsen, Thomas & William Norman. 1979. Correlates of ergativity in Mayan grammar. In Frans Plank (ed.), *Ergativity. Towards A Theory of Grammatical Relations*, 347–370. London/New York/Toronto/Sydney/San Francisco: Academic Press.
Law, Danny. 2013. Mayan historical linguistics in a new age. *Language and Linguistics Compass* 7(3). 141–156.
Law, Danny. 2014. *Language Contact, Inherited Similarity and Social Difference. The Story of Linguistic Interaction in the Maya lowlands* (Current Issues in Linguistic Theory 328). Amsterdam/Philadelphia: John Benjamins.
Law, Danny. 2016. The (re)invention of incompletive aspect in Cholan. Invited talk presented at The 2016 Maya Meetings, University of Texas at Austin.
Law, Danny. 2017a. Language mixing and genetic similarity. The case of Tojol-ab'al. *Diachronica* 34(1). 40–78.
Law, Danny. 2017b. Language contacts with(in) Mayan. In Judith Aissen, Nora England & Roberto Zavala Maldonado (eds.), *The Mayan Languages* (Routledge Language Family Series), 112–127. London/New York: Routledge.
Law, Danny, John Robertson & Stephen Houston. 2006. Split ergativity in the history of the Ch'olan Branch of the Mayan language family. *International Journal of American Linguistics* 72(4). 415–450.

Law, Danny, John Robertson, Stephen Houston & Robbie Haertel. 2009. Most Maya glyphs are written in Ch'olti'an. In Brent Metz, Cameron McNeil & Kerry Hull (eds.), *The Ch'orti' Maya Area. Past and Present*, 29–42. Gainesville: University Press of Florida.

Law, Danny & David Stuart. 2017. Classic Mayan. An overview of language in ancient hieroglyphic script. In Judith Aissen, Nora England & Roberto Zavala Maldonado (eds.), *The Mayan Languages* (Routledge Language Family Series), 128–172. London/New York: Routledge.

Lehmann, Christian. 1993. The genesis of auxiliaries in Yucatec Maya. *Proceedings of the International Congress of Linguistics* 15(2). 313–316.

Lehmann, Christian. 2000. Focalización del verbo en maya yucateco. Tordesillas.

Lehmann, Christian. 2020. Grammaticalization in Yucatec Maya (Earlier presented at the Symposium on Grammaticalization, Universität Mainz, 12–15 March, 2015). In Walter Bisang & Andrej Malchukov (eds.), *Grammaticalization Scenarios. Areal Patterns and Cross-Linguistic Variation*, 803–852. Berlin/Boston: De Gruyter Mouton.

Liljegren, Henrik. 2014. A survey of alignment features in the Greater Hindukush with special references to Indo-Aryan. In Pirkko Suihkonen & Lindsay Whaley (eds.), *On Diversity and Complexity of Languages Spoken in Europe and North and Central Asia*, 133–174. Amsterdam/Philadelphia: John Benjamins.

Lois, Ximena, Valentina Vapnarsky, Cédric Becquey & Aurore Becquelin. 2017. Polycategoriality and hybridity across Mayan languages. In Valentina Vapnarsky & Edy Veneziano (eds.), *Lexical Polycategoriality. Cross-linguistic, cross-theoretical and language acquisition approaches* (Studies in Language Companion Series 182), 101–154. Amsterdam/Philadelphia: John Benjamins.

Long, Haiping, Feng Gu, Xiaoping Xiao & Bo Hong. 2012. *Translation of World Lexicon of Grammaticalization*. Beijing: Beijing World Publishing Corporation.

Lopes, Luís, Barbara MacLeod & Alejandro Sheseña. 2023. Of Heat, Holes, and Hollow Places: The Semantics and Phonetic Value of T650. *Estudios de cultura maya* LXII. 75–114.

López Ixcoy, Candelaria Dominga. 1997. *Ri ukemiik ri K'ichee' chii'. Gramática K'ichee'*. Ciudad de Guatemala: Cholsamaj.

Lubeck, John & Diane Cowie. 1989. *Método moderno para aprender el idioma chorti'*. Guatemala: Instituto Lingüístico de Verano.

MacLeod, Barbara. 1982. Appendix 5. Split-ergativity in the Cholan and Yucatecan languages. In Lina Schele, *Maya Glyphs: The Verbs*, 420–422. Austin: University of Texas Press.

MacLeod, Barbara. 1987. *An Epigrapher's Annotated Index to Cholan and Yucatecan Verb Morphology* (Monographs in Anthropology 9). Columbia: University of Missouri, Museum of Anthropology.

Macri, Martha & Matthew Looper. 2003. Nahua in ancient Mesoamerica. *Ancient Mesoamerica* 14. 285–297.

Maldonado, Francisco & Pablo Paxbolon. 1610. Maldonado-Paxbolon Papers, ms.

Marcos García, William Arnaldo. 2015. *Alomorfos de los marcadores de personas y números gramaticales en el idioma maya chorti'*. Guatemala: Universidad de San Carlos de Guatemala MA thesis.

Marhenke, Randa. 2015. Typescript of Ch'orti' Texts from Wisdom's Materials on the Chorti language. https://www.academia.edu/41737996/Typescript_of_Chorti_Texts_from_Wisdoms_Materials_on_the_Chorti_language (last accessed: 2025-03-12).

Marhenke, Randa. 2020. A Note on Robertson and Law (2009), "From Valency to Aspect in the Ch'olan-Tzeltalan Family of Mayan", IJAL75-3, pp. 293–316. https://www.academia.edu/43699121/A_Note_on_Robertson_and_Law_2009_From_Valency_to_Aspect_in_the_Cholan_Tzeltalan_Family_of_Mayan_IJAL75_3_pp_293_316 (last accessed: 2025-03-12).

Martin, Laura. 1998. Irrealis constructions in Mocho (Mayan). *Anthropological Linguistics* 40. 198–213.

Martin, Simon & Nikolai Grube. 2008. *Chronicle of the Maya Kings and Queens. Deciphering the Dynasties of the Ancient Maya*. Second Edition. London: Thames & Hudson.

Mateo Toledo, Eladio. 2008. *The family of complex predicates in Q'anjob'al (Maya); their syntax and meaning*. Austin: University of Texas PhD Dissertation.

Mateo Toledo, Eladio. 2022. *Complex Predicates in Q'anjob'al (Maya). Resultatives, End-states Positionals, Causatives, Monitorings, Ditransitives* (Brill's Studies in the Indigenous Languages of the Americas 10). Boston: Brill.

McClaran, Marlis. 1972. *Lexical and syntactic structures in Yucatec Maya*. Cambridge, MS: Harvard University PhD Dissertation.

McGregor, William. 2009. Typology of ergativity. *Language and Linguistics Compass* 3(1). 480–508.

McGregor, William. 2017. Grammaticalization of ergative case marking. In Jessica Coon, Diane Massam & Lisa Travis (eds.), *The Oxford Handbook of Ergativity* (Oxford Handbooks in Linguistics), 447–464. Oxford: Oxford University Press.

McQuown, Norman. 1955. The indigenous languages of Latin America. *American Anthropologist* 57. 501–569.

McQuown, Norman. 1956. The classification of the Mayan languages. *International Journal of American Linguistics* 22. 191–195.

Membreño, Alberto. 1897. *Hondureñismos. Vocabulario de los provincialismos de Honduras. Segunda edición*. Tegucigalpa: Tipografia Nacional.

Merlan, Francesca. 1985. Split intransitivity: Functional oppositions in intransitive inflection. In Johanna Nichols & Anthony Woodbury (eds.), *Grammar inside and outside the clause. Approaches to theory from the field*, 324–362. Cambridge: Cambridge University Press.

Metz, Brent. 2006. *Ch'orti' Maya Survival in Eastern Guatemala: Indigeneity in Transition*. Albuquerque: University of New Mexico Press.

Metz, Brent. 2009a. The "Ch'orti Area." In Brent. Metz, Cameron McNeil & Kerry Hull (eds.), *The Ch'orti' Maya Area. Past and Present*, 1–13. Gainesville: University Press of Florida.

Metz, Brent. 2009b. Ch'orti' Maya indigenousness in Guatemala, Honduras, El Salvador. In Brent E. Metz, Cameron L. McNeil & Kerry M. Hull (eds.), *The Ch'orti' Maya Area. Past and Present*, 161–172. Gainesville: University Press of Florida.

Metz, Brent. 2022. *Where did the Eastern Mayas go? The Historical, Relational, and Contingent Interplay of Ch'orti' Indigeneity* (IMS Studies on Culture and Society Series 11). Louisville: University Press of Colorado.

Metz, Brent, Cameron McNeil & Kerry Hull (eds.). 2009. *The Ch'orti' Maya Area. Past and Present*. Gainesville: University Press of Florida.

Metz, Brent, Teo Ramírez, Fred García & Mundo García. 1992. Español-Chorti-English-Dictionario [sic]. Unpublished manuscript.

Mithun, Marianne. 1991. Active/agentive case marking and its motivation. *Language* 67. 510–546.

Mithun, Marianne & Wallace Chafe. 1999. What are S, A, and O? *Studies in Language* 23(3). 569–596.

Mó Isém, Romelia. 2006. *Fonología y morfología del Poqomchi' occidental*. Guatemala de la Asunción: Universidad Rafael Landívar Licenciatura.
Mohammad, Jan. 1991. *Causative constructions in Kati, a Nuristani language of Afghanistan*. Athens: Ohio University MA thesis.
Mondloch, James. 1981. *Voice in Quiché-Maya*. Albany: State University of New York PhD Dissertation.
Montgomery, John. 2002. *Dictionary of Maya Hieroglyphs*. New York: Hippocrene Books.
Mora-Marín, David. 2003. Historical reconstruction of Mayan applicative and antidative constructions. *International Journal of American Linguistics* 69(2). 186–228.
Mora-Marín, David. 2005a. The Proto-Ch'olan positional status marker *-täl and additional comments on Classic Mayan positional morphology. *Wayeb Notes* 17. 1–5.
Mora-Marín, David. 2005b. Proto-Ch'olan as the standard language of Classic Lowland Mayan texts. Unpublished manuscript. https://davidmm.web.unc.edu/wp-content/uploads/sites/400/2025/02/ProtoCholanHypothesis.pdf (last accessed: 2025-03-12).
Mora-Marín, David. 2009a. Reconstruction of Proto-Ch'olan independent pronouns: Grammaticalization and evidence for sociolinguistic variation. *Transactions of the Philological Society* 107(1). 98–129.
Mora-Marín, David. 2009b. A test and falsification of the "Classic Ch'olti'an" hypothesis: A study of three Proto-Ch'olan markers. *International Journal of American Linguistics* 75(2). 115–157.
Mora-Marín, David. 2021. The "Initial Sign" of epigraphic Mayan as ʔAY 'existential particle': Evidence from predicative possession and evidential preverbal functions. Unpublished manuscript.
Mora-Marín, David, Nicholas Hopkins & Kathryn Josserand. 2009. The linguistic affiliation of Classic Lowland Mayan writing and the historical sociolinguistic geography of the Mayan lowlands. In Brent Metz, Cameron McNeil & Kerry Hull (eds.), *The Ch'orti' Maya Area. Past and Present*, 15–28. Gainesville: University Press of Florida.
Morán, Fray Francisco. 1695. Arte y vocabulario de la lengua Cholti. Manuscript. Catalog No. 497.4 M79 in the American Philosophical Society. Philadelphia (PA), ms.
Morley, Sylvanus. 1920. *Inscriptions at Copán* (Publication no. 219). Washington, DC: The Carnegie Institution of Washington.
Morwa'r ojroner Ch'orti'/Comunidad lingüística Ch'orti'. 2004. *U'tirach e wojroner Maya Ch'orti'. Gramática descriptiva Ch'orti'*. Jocotán: Academia de Lenguas Mayas de Guatemala.
Morwa'r ojroner Ch'orti'/Comunidad lingüística Ch'orti'. 2008. *Uyokir e kanseyaj ojroner Ch'orti'. Gramática pedagógica Ch'orti'*. Jocotán: Academia de Lenguas Mayas de Guatemala.
Munro, Pamela. 2017. The Mesoamerican linguistic area revisited. In Karen Dakin, Claudia Parodi & Natalie Operstein (eds.), *Language Contact and Change in Mesoamerica and Beyond* (Studies in Language Companion Series 185), 336–353. Amsterdam/Philadelphia: John Benjamins.
Norcliffe, Elisabeth. 2003. *The Reconstruction of Proto-Huastecan*. Christchurch, NZ: University of Canterbury MA thesis.
Norman, William. 1977. Topic and focus in Mayan. Paper presented at Mayan Workshop II, San Cristóbal de las Casas (Chiapas, México).
Norman, William & Lyle Campbell. 1978. Toward a Proto-Mayan syntax: A comparative perspective on grammar. In Nora England (ed.), *Papers in Mayan Linguistics*, 136–156. Columbia: University of Missouri.
Oakley, Helen. 1965. Chorti. In Marvin Mayers (ed.), *Languages of Guatemala*, 235–250. The Hague: Mouton & Co.

Olander, Thomas. 2022. The etymological substance of the Italian first- and second-person oblique clitic pronouns. *Revue Romane*. Preprint. https://curis.ku.dk/ws/portalfiles/portal/280611818/Olander_The_etymological_substance_of_the_Italian_first_and_second_person_oblique_clitic_pronouns.pdf (last accessed: 2025-03-12).
Orel, Vladimir. 1998. *Albanian Etymological Dictionary*. Leiden: Brill.
Osorio May, José del Carmen. 2016. *Temas de la sintaxis del yokot'an "chontal" de Tecoluta, Nacajuca, Tabasco*. México: Centro de Investigaciones y Estudios Superiores en Antropología Social PhD Dissertation.
Payne, John. 1980. The decay of ergativity in Pamir languages. *Lingua* 51. 147–186.
Payne, Thomas. 1997. *Describing Morphosyntax. A guide for field linguists*. Cambridge: Cambridge University Press.
Pérez Martínez, Vitalino. 1994. *Gramática del idioma Ch'orti'*. La Antigua: Proyecto Lingüístico Francisco Marroquín.
Pérez Martínez, Vitalino. 1996. *Leyenda maya Ch'orti'*. Antigua: Proyecto Lingüístico Francisco Marroquín.
Pérez Martínez, Vitalino, Fred García, Felipe Martínez & Jeremías López. 1996. *Diccionario Ch'orti', Jocotán, Chiquimula*. Guatemala: Proyecto Lingüístico Francisco Marroquín.
Pinkster, Harm. 2015. *The Oxford Latin Syntax. Volume I: The Simple Clause*. Vol. 1. 2 vols. Oxford: Oxford University Press.
Polian, Gilles. 2017. Morphology. In Judith Aissen, Nora England & Zavala Maldonado (eds.), *The Mayan Languages* (Routledge Language Family Series), 201–225. London/New York: Routledge.
Quizar, Robin. 1979. *Comparative word order in Mayan*. Boulder: University of Colorado Dissertation.
Quizar, Robin. 1994a. Motion verbs in Ch'orti'. *Función* 9(15–16). 211–229.
Quizar, Robin. 1994b. Split ergativity in word order in Ch'orti'. *International Journal of American Linguistics* 60(2). 120–138.
Quizar, Robin. 2020. Tracing the Ch'orti' antipassive system: A comparative/historical view. *International Journal of American Linguistics* 86(2). 237–283.
Quizar, Robin. 2023. Xinkan influence on Ch'orti' verbs. *International Journal of American Linguistics* 89(2). 255–281.
Quizar, Robin & Susan Knowles-Berry. 1988. Ergativity in the Cholan languages. *International Journal of American Linguistics* 54(1). 73–95.
Reinöhl, Uta. 2018. Review of Eystein Dahl and Krzysztof Stroński (eds.). Indo-Aryan ergativity in typological and diachronic perspective (Typological studies in language 112). Amsterdam & Philadelphia: John Benjamins, 2016. *Journal of South Asian Languages and Linguistics* 5(1). 111–121.
Reinöhl, Uta. n.d. From nouny to verby – The emergence of split ergativity in Indo-Aryan. Unpublished manuscript.
Robertson, John. 1980. *The Structure of Pronoun Incorporation in the Mayan Verbal Complex*. New York: Garland.
Robertson, John. 1992. *The History of Tense/Aspect/Mood/Voice in the Mayan Verbal Complex*. Austin: University of Texas Press.
Robertson, John. 1998. A Ch'olti'an explanation for Ch'orti'an grammar: A postlude to the language of the Classic Maya. *Mayab* (11). 5–11.
Robertson, John. 1999. The history of first-person singular in the Mayan languages. *International Journal of American Linguistics* 65(4). 449–465.
Robertson, John. 2010. From Common Cholan-Tzeltalan to Classical Ch'olti'. The identification of the language of Mayan hieroglyphs. *Mesoweb Articles*.

www.mesoweb.com/articles/robertson/Robertson-2010.pdf (last accessed: 2025-03-12).
Robertson, John & Stephen Houston. 2015. The Huastec problem. A linguistic and archaeological perspective. In Katherine Faust & Kim Richter (eds.), *The Huasteca. Culture, History, and Interregional Exchange*, 19–36. Norman: University of Oklahoma Press.
Robertson, John, Stephen Houston & David Stuart. 2004. Tense and aspect in Maya hieroglyphic script. In Søren Wichmann (ed.), *The Linguistics of Maya Writing*. Salt Lake City: University of Utah Press.
Robertson, John & Danny Law. 2009. From valency to aspect in the Ch'olan-Tzeltalan family of Mayan. *International Journal of American Linguistics* 75(3). 293–316.
Robertson, John, Danny Law & Robbie Haertel. 2010. *Colonial Ch'olti'. The Seventeenth-Century Morán Manuscript*. Norman: University of Oklahoma Press.
Rogers, Christopher. 2016. *The Use and Development of the Xinkan Languages*. Austin: University of Texas Press.
Sachse, Frauke. 2010. *Reconstructive description of eighteenth-century Xinka grammar*. Leiden: Universiteit Leiden PhD Dissertation.
Santos Nicolás, José Francisco & José Gonzalo Benito Pérez. 1998. *Rukorb'aal Poqom Q'orb'al. Gramática Poqom (Poqomam)*. Guatemala: Cholsamaj.
Sapir, Edward. 1917. (Review of) Het passieve karakter van het verbum transitivum of van het verbum actionis in talen van Noord-Amerika [The passive character of the transitive verb or of the active verb in languages of North America] by C.C. Uhlenbeck. *International Journal of American Linguistics* 1(1). 82–86.
Sapper, Karl. 1893. Beiträge zur Ethnographie der Republik Guatemala. *Dr. A. Petermanns Mitteilungen aus Justus Perthes' geografischer Anstalt* 39. 1–14.
Sapper, Karl. 1897. *Das nördliche Mittel-Amerika nebst einem Ausflug nach dem Hochland von Anahuac*. Braunschweig: Friedrich Vieweg und Sohn.
Sapper, Karl. 1912. Über einige Sprachen von Südchiapas. In *Proceedings of the Seventeenth International Congress of Americanists (1910)*, 295–320.
Sasse, Hans-Jürgen. 1978. Subjekt und Ergativ: Zur pragmatischen Grundlage primärer grammatischer Relationen. *Folia Linguistica* XII(3–4). 219–252.
Sasse, Hans-Jürgen. 1987. The thetic/categorical distinction revisited. *Linguistics* 25. 511–580.
Saturno, William, David Stuart & Boris Beltran. 2006. Early Maya Writing at San Bartolo, Guatemala. *Science* 311(5765). 1281–1283.
Schele, Linda. 1982. *Maya Glyphs: The Verbs*. Austin: University of Texas Press.
Schmitt, Rüdiger. 2009. *Die altpersischen Inschriften der Achaimeniden. Editio minor mit deutscher Übersetzung*. Wiesbaden: Reichert.
Scholes, France V & Ralph L. Roys. 1968. *The Maya Chontal Indians of Acalan-Tixchel. A Contribution to the History and Ethnography of the Yucatan Peninsula (with the assistance of Eleanor B. Adams and Robert S. Chamberlain)*. Second edition. Norman: University of Oklahoma Press.
Schuchardt, Hugo. 1896. Ueber den passiven Charakter des Transitivs in den kaukasischen Sprachen. *Sitzungsberichte der philosophisch-historischen Classe der Akademie der Wissenschaften zu Wien* 133. 1–91.
Schumann Gálvez, Otto. 2007. *Introducción a la morfología verbal del Chortí*. Graz: Academic Publishers.
Schweitzer, Jürgen. 2006. *Das Große im Kleinen. Zur diachronen Typologie des Maya-Verbalkomplexes, unter besonderer Berücksichtigung des Ch'olotzeltalischen Zweiges*. München: Ludwig-Maximilians-Universität PhD dissertation.

Shagal, Ksenia. 2019. *Participles. A Typological Study* (Empirical Approaches to Language Typology [EALT] 61). Berlin/New York.
Sharer, Robert. 2009. The Ch'orti' past: An archaeological perspective. In Brent Metz, Cameron McNeil & Kerry Hull (eds.), *The Ch'orti' Maya Area. Past and Present*, 124–133. Gainesville: University Press of Florida.
Shosted, Ryan. 2011. Towards a Glottalic Theory of Mayan. In Heriberto Avelino (ed.), *New Perspectives in Mayan Linguistics*, 80–113. Cambridge: Cambridge Scholars Publishing.
Siewierska, Anna. 2003. Person agreement and the determination of alignment. *Transactions of the Philological Society* 101. 339–370.
Silverstein, Michael. 1976. Hierarchy of features and ergativity. In Robert Dixon (ed.), *Grammatical Categories in Australian Languages*, 112–171. Canberra: Australian Institute of Aboriginal Studies.
Smailus, Ortwin. 1973. *Das Maya-Chontal von Acalan. Sprachanalyse eines Dokumentes aus den Jahren 1610/12*. Hamburg: Universität Hamburg PhD Dissertation.
Smailus, Ortwin. 1989. *Gramática del maya yucateco colonial*. Hamburg: WAYASBAH.
Stewart, Stephen. 2015. *Q'eqchi' Grammar*. Lafayette: SOS Publishing Co.
Stiebels, Barbara. 2006. Agent focus in Mayan languages. *Natural Language & Linguistic Theory* 24. 501–570.
Stoll, Otto. 1884. *Zur Ethnographie der Republik Guatemala*. Zurich: Orell Füssli und Co.
Stoll, Otto. 1885. Supplementary remarks to the grammar of the Cakchiquel language, ed. by Danial G. Brinton. *Proceedings of the American Philosophical Society* 22. 255–268.
Storniolo, Judith. 2008. *A Comparative Study of Eastern Cholan*. Philadelphia: University of Pennsylvania PhD Dissertation.
Straight, Stephen. 1976. Decompositional structure in Yucatec verbs. In Marlis McClaran (ed.), *Mayan Linguistics. Vol. 1*, 189–198. Los Angeles: UCLA American Indian Studies Center.
Stross, Brian. 1992. Ch'orti' Mayan Lexicon. Transcribed and transliterated from handwritten fieldnotes of C. Wisdom. Published in full online. Language Laboratory, Dept. of Anthropology, University of Texas., ms. https://utdirect.utexas.edu/apps/student/coursedocs/nlogon/download/689830 (last accessed: 2025-03-12).
Suárez, Ruano Alberto. 1892. Vocabularios pocoman y chorti de Honduras. Garrett-Gates Mesoamerican Manuscripts, no. 251, Manuscripts Division, Department of Rare Books and Special Collections, Princeton University Library, ms.
Terga Cintrón, Ricardo. 1980. El valle bañado por el río de plata. *Guatemala Indígena* 15(1–2). 1–110.
Thompson, Eric. 1990. *Maya History and Religion* (The Civilization of the American Indian Series 99). Norman/London: University of Oklahoma Press.
Trask, Larry. 1979. On the origins of Ergativity. In Frans Plank (ed.), *Ergativity. Towards a Theory of Grammatical Relations*, 385–404. London: Academic Press.
Uhlenbeck, Christianus Cornelius. 1916. Het passieve Karakter van het Verbum transitivum of van het Verbum actionis in Taalen von Noord-Amerika. *Verslagen en Mededeelingen der Koninklijke Akademie van Wetenschappen: Afd. Letterkunde* 5. 187–216.
Vázquez Álvarez, Juan Jesús. 2011. *A grammar of Chol, a Mayan language*. Austin: University of Texas PhD Dissertation.
Velleman, Leah. 2014. *Focus and movement in a variety of K'ichee'*. Austin: University of Texas PhD Dissertation.
Verhoeven, Elisabeth & Stavros Skopeteas. 2015. Licensing focus constructions in Yucatec Maya. *International Journal of American Linguistics* 81. 1–40.

Vinogradov, Igor. 2016. *Tense/Aspect/Mood Systems of Cholan-Tseltalan Mayan Languages*. München: LINCOM.
Vinogradov, Igor. 2017. From enclitic to prefix: diachrony of personal absolutive markers in Q'eqchi'. *Morphology* 27. 105–122.
Vinogradov, Igor. 2018. Grammatical periphery of Chontal Maya verb. In *Proceedings of the Workshop on the Structure and Constituency of Languages of the Americas 21* (University of British Columbia Working Papers in Linguistics 46), 266–280. Vancouver: University of British Columbia.
Vinogradov, Igor. 2022. Evangelization and language change: A transition from the progressive aspect to the future tense in two Mayan languages. *Transactions of the Philological Society* 120(3). 375–396.
Wald, Robert. 2000. Temporal deixis in Colonial Chontal and Maya Hieroglyphic narrative. *Written Language and Literacy* 3(1). 123–153.
Wald, Robert. 2004a. Telling time in Classic-Ch'olan and Acalan-Chontal narrative: The linguistic basis of some temporal discourse patterns in Maya hieroglyphic and Acalan-Chontal texts. In Søren Wichmann (ed.), *The Linguistics of Maya Writing*, 211–258. Salt Lake City: University of Utah Press.
Wald, Robert. 2004b. The Languages of the Dresden Codex: Legacy of the Classic Maya. In Søren Wichmann (ed.), *The Linguistics of Maya Writing*, 27–58. Salt Lake City: University of Utah Press.
Wald, Robert. 2007. *The verbal complex in classic-period Maya hieroglyphic inscriptions: Its implications for language identification and change*. Austin: University of Texas PhD Dissertation.
Wald, Robert & Barbara MacLeod. 1999. Narrative time in the Classic-period inscriptions. In Nikolai Grube (ed.), *Notebook for the XXIIIrd Maya Hieroglyphic Forum at Texas*, vol. II, 88–96. Austin: University of Texas.
Wichmann, Søren. 1999. A Ch'orti' Morphological Sketch. Unpublished manuscript. https://www.academia.edu/37358899/A_Ch_orti_morphological_sketch (last accessed: 2015-03-12).
Wichmann, Søren. 2002. Hieroglyphic evidence for the historical configuration of Eastern Ch'olan. *Research Reports on Ancient Maya Writing* 51. 1–35.
Wichmann, Søren. 2006. Mayan historical linguistics and epigraphy: A new synthesis. *Annual Review of Anthropology* 35. 279–294.
Wisdom, Charles. 1940. *The Chorti Indians of Guatemala*. Chicago: University of Chicago Press.
Wisdom, Charles. 1950. Material on the Chorti language. Microfilm Collection of Manuscripts on Meso-american Cultural Anthropology, no. 28. Chicago: University of Chicago, ms. https://www.lib.uchicago.edu/mca/mca-05-028.pdf (last accessed: 2015-03-12).
Witzlack-Makarevich, Alena. 2010. *Typological variation in grammatical relations*. Leipzig: Universität Leipzig.
Woodbury, Anthony C. 1985. Noun phrase, nominal sentence, and clause in Central Alaskan Yup'ik Eskimo. In Johanna Nichols & Anthony C. Woodbury (eds.), *Grammar inside and outside the clause*, 61–88. Cambridge: Cambridge University Press.
Wycliffe Bible Translators. 2012. *E ojroner xe' imb'utz tua' cawinquirar Jesucristo. El nuevo testamento en Chortí*. Second. https://www.scriptureearth.org/data/caa/PDF/00-WNTcaa-web.pdf (last accessed: 2015-03-12).
Ximénez, Francisco. 1702. *Arte de las tres lenguas cakchiquel, quiche y tzutuhil* (Biblioteca Guatemala 31). Guatemala: Academia de Geografía e Historia de Guatemala.
Yoshida, Shigeto. 2016. Un análisis morfo-sintáctico sobre el tiempo y el aspecto en la lengua maya yucateca. *Latin American and Caribbean Studies* 23. 39–51.

Zavala Maldonado, Roberto. 2017. Alignment Patterns. In Judith Aissen, Nora England & Roberto Zavala Maldonado (eds.), *The Mayan Languages* (Routledge Language Family Series), 225–258. London/New York: Routledge.

Zender, Marc. 2004. On the morphology of intimate possession in Mayan languages and Classic Mayan glyphic nouns. In Søren Wichmann (ed.), *The Linguistics of Maya Writing*, 195–210. Salt Lake City: The University of Utah Press.

Zender, Marc. 1999. *Diacritical Marks and Underspelling in the Classic Maya Script: Implications for Decipherment*. Alberta: University of Calgary MA thesis.

Zuniga, Dionisio de. 1614. Diccionario pocomchi-castellano y castellano-pocomchi de San Cristobal Cahcoh. University of Pennsylvania Libraries, Berendt-Brinton Linguistic Collection, item 61.

Zúñiga, Fernando. 2018. The diachrony of morphosyntactic alignment. *Language and Linguistics Compass* 12(9). e12300.