Coon, Jessica & Vázquez Álvarez, Juan Jesús. 2025.
Embedding, extraction, and clausal pied-piping in Ch'ol.
Glossa: a journal of general linguistics 10(1). pp. 1–28.
DOI: https://doi.org/10.16995/glossa.17171

Open Library of Humanities

Embedding, extraction, and clausal pied-piping in Ch'ol

Jessica Coon, McGill University, jessica.coon@mcgill.ca

Juan Jesús Vázquez Álvarez, CIMSUR-UNAM, juanvazquezalvarez@gmail.com

This paper examines clausal embedding and long-distance wh-dependencies in Ch'ol, a Mayan language of southern Mexico. We demonstrate that when it comes to finite embedded clauses, two options are available: (i) long-distance wh-extraction and (ii) clausal pied-piping with inversion of the wh-word. In clausal pied-piping, the embedded wh-word fronts to the edge of the embedded clause, and the entire embedded clause fronts to the left edge of the matrix clause. We show further that clausal pied-piping is possible only with realis embedded clauses, even when there is no obvious difference in syntactic clause size between these and the irrealis clauses which disallow it. While some work has discussed long-distance extraction in other Mayan languages (e.g. Craig 1977; Erlewine 2016; Can Pixabaj 2020; Mendes & Ranero 2021), we are unaware of descriptions of clausal pied-piping in elsewhere in the Mayan literature. Through the examination of complex clauses, we argue that the Ch'ol patterns provide evidence in favour of a QP approach to pied-piping (Cable 2007), and against an analysis involving feature percolation. This work thus contributes both to our understanding of embedding and extraction within the Mayan family, as well as to typological and theoretical discussions of clausal pied-piping and the syntactic mechanisms underlying it.

1 Introduction

Ch'ol allows long-distance wh-extraction from embedded clauses, as shown by the pair in (1).

(1) a. K-om [che' mi i-juch' ixim **li** **xk'aläl**].
A1-want that IPFV A3-grind corn DET girl
'I want the girl to grind corn.'
b. **Maxki**$_i$ aw-om [mi i-juch' ixim t_i]?
who A2-want IPFV A3-grind corn
'Who do you want to grind corn?'

However, from certain embedded clauses, such as the one shown in (2a), long-distance wh-extraction is one of two options for forming a long-distance wh-dependency, shown in (2b) and (2c). While long-distance extraction as in (2b) is grammatical, an additional (and sometimes preferred) option for a wh-word in the embedded clause in (2a) involves *clausal pied-piping with inversion*, previewed in (2c).[1] Here, the embedded wh-word appears at the left edge of the embedded clause, which in turn appears at the left edge of the matrix clause.

(2) a. Tyi k-äl-ä [che' tyi i-juch'-u ixim **li** **xk'aläl**].
PFV A1-say-TV that PFV A3-grind-TV corn DET girl
'I said that the girl ground corn.'
b. **Maxki**$_i$ tyi aw-äl-ä [tyi i-juch'-u ixim t_i]?
who PFV A2-say-TV PFV A3-grind-TV corn
'Who did you say ground corn?' (long-distance extraction)
c. [**Maxki**$_i$ tyi i-juch'-u ixim t_i]$_k$ tyi aw-äl-ä t_k?
who PFV A3-grind-TV corn PFV A2-say-TV
'Who did you say ground corn?'
(lit. ≈ 'Who ground corn did you say?') (clausal pied-piping)

Though described as rare cross-linguistically (Heck 2008), clausal pied-piping has been documented in other languages, including Quechua (Cole 1982; Hermon 1985), Basque (Ortiz de Urbina 1993; Arregi 2003), Wolof (Torrence 2013), and Tlingit (Cable 2007). Clausal pied-piping has been used as evidence for the successive cyclic nature of long-distance wh-movement, as it requires what Heck terms "secondary movement"—that is, movement of the wh-word to the edge of the embedded clause, shown for *maxki* 'who' in (2c), in addition to the "primary" movement to the matrix clause. Where both long-distance extraction and clausal pied-piping options are available in other languages for which clausal pied-piping has been described, the choice between

[1] We assume that in (2b) the embedded wh-word moves through the edge of the embedded clause, though for the sake of simplicity, we do not represent this in all of the examples.

the two has been described as optional and semantically vacuous (e.g., Arregi 2003; Cable 2007 on Basque and Tlingit, respectively), a finding we corroborate for Ch'ol below.

For Ch'ol, we show further that while both clausal pied-piping and long-distance extraction are possible from embedded *realis* complements, as in (2), long-distance extraction is the only possibility from embedded *irrealis* complement clauses, like the one in (1)—even when there is no obvious difference in syntactic size between the two. This contrast is particularly clear with certain verbs, such as *al* 'say, tell', which may take either type of complement, shown by the two possible translations in (3).[2]

(3) Tyi k-äl-ä [che' mi i-päk' ixim li alob].
PFV A1-say/tell-TV that IPFV A3-plant corn DET boy
'I said that the boy plants corn.' →*realis complement interpretation*
'I told the boy to plant corn.' →*irrealis complement interpretation*

The choice of strategy to form a long-distance wh-dependency disambiguates these readings. Clausal pied-piping is possible only for realis complements, as in (4), while long-distance extraction favours an irrealis interpretation, as shown in (5).

(4) *Realis context*: I said that one of my three sons plants corn, but you don't remember who it was. You ask me…
[**Maxki**$_i$ mi i-päk' ixim t_i]$_k$ tyi aw-äl-ä t_k?
who IPFV A3-plant corn PFV A2-say-TV
'Who did you say plants corn?'

(5) *Irrealis context*: You know I've sent one of my three sons to plant corn. You ask me…
Maxki$_i$ tyi aw-äl-ä [mi i-päk' ixim t_i]?
who PFV A2-say-TV IPFV A3-plant corn
'Who did you tell to plant corn?'

We will see below that *focussed* elements—taken to occupy the same position as wh-question words—may also trigger clausal pied-piping. Throughout this paper, we use "wh-movement" to include both wh-questions and focus movement.

This paper aims to establish the empirical generalizations around long-distance extraction and clausal pied-piping in Ch'ol, and to examine the differences between these two options for embedded clause types. We motivate a syntactic analysis for clausal pied-piping of realis clauses in which wh-movement always involves a relationship between the interrogative C and a projection,

[2] The vowel of the verb *al* raises to the mid-high unrounded vowel *ä* in most inflectional contexts, a regular phonological process for certain verbs in the language.

QP, and differences between long-distance extraction and clausal pied-piping arise from different possible merge sites for Q, following Cable's (2007) analysis for Tlingit.

The remainder of this paper is organized as follows: section 2 provides relevant background on the language, with a focus on wh-extraction and an overview of embedding strategies. Section 3 examines patterns of subextraction versus clausal pied-piping and their relation to clause type. Section 4 delves into additional details of clausal-pied-piping in Ch'ol, comparing it with discussion of clausal pied-piping cross-linguistically. We propose a syntactic analysis of clausal pied-piping, providing evidence in favour of a QP analysis and against feature percolation, and offer suggestions for the impossibility of pied-piping with irrealis clauses. Section 5 concludes.

2 Ch'ol extraction and embedding

Ch'ol is a member of the Cholan-Tseltalan branch of the Mayan family, and is spoken in Mexico in the states of Chiapas, Tabasco, and Campeche by around 250,000 speakers (Vázquez Álvarez 2011). The two main dialects of Ch'ol are Tila and Tumbalá; the data in this paper come from contextually-driven elicitation with three speakers of the Tila dialect, as well as native-speaker intuitions of the second author. This section provides an overview of relevant grammatical properties and wh-movement (§2.1) before turning to a specific look at embedding strategies (§2.2).

2.1 Basics and wh-movement

Ch'ol is a morphologically ergative language in which grammatical relations are head-marked via two sets of morphemes: Set A (ergative, possessive) and Set B (absolutive); all core arguments may be pro-dropped. As in other Mayan languages, predicate stems can be divided into "verbal" (generally eventive) and "non-verbal" (generally stative) classes. Verbal stems typically appear with a "status suffix" which varies based on transitivity, and require one of three aspects: perfective (*tyi, tsa', ta'*), as in (6a); imperfective (*mi, muk', mu'*); or progressive (*choñkol, woli*). Non-verbal predicates, as in (6b), may not appear with aspectual inflection and do not have status suffixes.

(6) a. **Tyi** jul-i-yoñ.
PFV arrive-ITV-B1
'I arrived.' (perfective)

b. K-om waj.
A1-want tortilla
'I want tortillas.' (non-verbal predicate)

Basic constituent order is VOS/VS, with clause-initial topic and focus positions (Vázquez Álvarez 2011; Clemens & Coon 2018). Vázquez Álvarez (2011) argues that Ch'ol topics are *external* topics in the sense of Aissen (1992), preceding the clause-internal focus position, as in (7).

(7) Topic [Focus V O S]

Focussed constituents, wh-elements, relativized nominals, and other focus-sensitive operators all compete for the single position clause internal left-periphery position ("Focus" in (7); Norman 1977; Aissen 1992; Velleman 2014). Wh-movement is obligatory for question formation in Ch'ol and multiple wh-questions or wh-in situ are impossible in the language; see Vázquez Álvarez & Coon (2020) for examples and further details. While some Mayan languages restrict the extraction of transitive subjects, known in the Mayanist literature as the Ergative Extraction Constraint (Aissen 2017b) (an instance of "syntactic ergativity"), Ch'ol does not. All core arguments freely extract (Vázquez Álvarez 2011).

Finally, we note that while clausal pied-piping has not been described in Mayan (though see §4 for a comparison with "inversion" in Popti'), pied-piping with inversion has long been documented in the nominal domain for wh-possessors (see e.g., Aissen 1996; Coon 2009; Little 2020; Aissen & Polian 2024), as well as with the wh-complements of relational nouns and prepositional phrases (Aissen 1996; Ewing 2022). Ch'ol permits both possessive phrase and PP pied-piping; see Coon (2009) and Little (2020) for wh-possessors, and Cable (2007: ch. 5) on PPs.

2.2 Ch'ol embedded clauses

Vázquez Álvarez (2013) describes three types of embedded complements in Ch'ol: (i) non-finite; (ii) "less-finite"; and (iii) fully finite. Both non-finite and less-finite complements obligatorily lack aspect marking (taken in Aissen 1992 and subsequent work to occupy finite Infl^0), are dependent on the matrix clause for their temporal interpretation, and exhibit obligatory control of their subjects by a matrix argument. Non-finite clauses, like the one in (8a), are obligatorily intransitive (including examples of transitive roots with incorporated objects), while less-finite clauses, like the one in (8b) are transitive. While the "less finite" form requires Set A subject marking on the embedded clause, it is obligatorily identical to the subject marking in the matrix clause. We take these clauses to be smaller than IP, but remain agnostic about their exact structure; following Aissen's (2017a) discussion of complementation in Mayan more generally, below we refer to these two clause types together as "non-finite".

(8) a. Tyi k-tyech-e [wäy-el].
PFV A1-start-TV sleep-DEP
'I began to sleep.' (non-finite complement)

b. Tyi k-tyech-e [**k**-mel k-otyoty].
PFV A1-begin-TV A1-make A1-house
'I began to make my house.' (less-finite complement)

Finite embedded clauses, in contrast, are distinguished from both types of non-finite clauses in (8) by: (i) the obligatory appearance of an aspectual marker; (ii) the possibility of a complementizer (optional for some speakers, preferred for others); and (iii) the absence of any dependency between matrix and embedded arguments (see Aissen 2017a). These properties are shown in (9).

(9) Tyi k-il-ä [**che'** *(**tyi**) i-mäñ-ä ixim ajMaria].
PFV A1-see-DTV that PFV A3-buy-TV corn Maria
'I saw that Maria bought corn.' (finite complement)

The relationship between complement type and a control interpretation between matrix and embedded arguments can be seen clearly with the verb *om* 'want', which may combine with either non-finite or finite complements. As shown in (10), coreference with the matrix subject is obligatory in the non-finite complements in (10a). For disjoint reference between matrix and embedded subjects, a finite embedded clause is required, as in (10b).

(10) a. Y-om [i-jap i-kajp'ej] li x'ixik.
A3-want A3-drink A3-coffee DET woman
'The woman wants to drink her coffee.'
b. Y-om li x'ixik [$_{CP}$ **che' mi i**-jap i-kajp'ej ajMaria].
A3-want DET woman that IPFV A3-drink A3-coffee Maria
'The woman wants Maria to drink her coffee.'

The examples in (10) illustrate an additional contrast: while the aspectless non-finite complements appear in typical object position (recall that word order is VOS), finite clauses are obligatorily *extraposed*, appearing after the matrix subject (Vázquez Álvarez 2011). We remain agnostic about the nature of this extraposition, but take this to further motivate a difference in syntactic clause size between non-finite clauses, on the one hand, and fully finite embedded clauses on the other.

Only fully finite embedded clauses permit clausal pied-piping, the focus of the remainder of this paper. As shown in (11), for example, a wh-object of a non-finite complement must front to initial position; pied-piping of the dependent verb form is ungrammatical. We connect the impossibility of clausal pied-piping of non-finite clauses to the absence of higher clausal structure (i.e., the absence of a CP layer), which will form part of our analysis of clausal pied-piping below.

(11) a. Chuki$_i$ tyi a-tyech-e [a-mel t_i]?
what PFV A2-begin-TV A2-make
'What did you begin to make?'
b. *[Chuki$_i$ a-mel t_i]$_k$ tyi a-tyech-e t_k?
what A2-make PFV A2-begin-TV
intended: 'What did you begin to make?'

As noted in Vázquez Álvarez (2011), verbs which take finite complements in Ch'ol encode meanings typical of verbs which take clausal complements cross-linguistically (Cristofaro 2003; Noonan 2007), including utterance verbs like *al* 'say, believe', *su'b* 'say, accuse', *jak'* 'respond, admit' (12); propositional attitude verbs like *u'biñ* 'think, believe', *ch'ujbiñ* 'believe, accept, obey', *pi'tyañ* 'await, anticipate' (13); verbs of knowledge such as *ujil* 'know', *ña'tyañ* 'know, undertand' (14); desideratives like *om* 'want', *mulañ* 'like' *pi'tyañ* 'wait', (15), and verbs of perception like *ilañ* 'see', *k'el* 'see, watch', *chäñtyañ* 'watch, observe', *ñächtyañ* 'listen' (16). All of the bracketed forms below can serve as matrix clauses minus the complementizer *che'*.[3]

(12) Tyi a-su'b-u [che' tyi i-jats'-ä y-ijts'iñ li alob].
PFV A2-accuse-DTV that PFV A3-hit-TV A3-younger.sibling DET boy
'You accused the boy of hitting his younger sibling.'

(13) Tyi k-u'b-i [che' tyi i-tsäñ-s-ä chityam li wiñik].
PFV A1-understand-DTV that PFV A3-die-CAUS-DTV pig DET man
'I understood that the man killed a pig.'

(14) K-ujil [che' ta-x majl-i xk'aläl].
A1-know that PFV-already go-ITV girl
'I know that the girl already left.'

(15) Mi k-pi'ty-añ [che' mi i-k'än-'añ k-chol].
IPFV A1-wait-DTV that IPFV A3-ripe-INCH A1-field
'I'm waiting for my crops to ripen.'

(16) Tyi k-il-ä [che' tyi i-mäñ-ä wakax li wiñik].
PFV A1-see-DTV that PFV A1-buy-TV cow DET man
'I saw that the man bought a cow.'

[3] Throughout we have translated Ch'ol sentences into a natural-sounding English equivalent, though in many cases where English embeds non-finite forms, Ch'ol uses a full CP clause.

3 Clause type and long-distance extraction

Armed with a basic inventory of clausal embedding verbs, we now turn to patterns of long-distance wh-dependencies. As foreshadowed in section 1, we find two different strategies for long-distance wh-questions. Either the wh-word may extract to the edge of the matrix clause (see (2b)), or the wh-word moves to the left edge of the embedded clause, and the embedded clause moves to the left edge of the matrix clause, (2c).

In this section, we first investigate the distribution of the two types of extraction, showing initial evidence that irrealis complements only permit long-distance extraction, while realis complements permit either (§3.1). Next (§3.2), we show that despite no obvious difference in syntactic size, there are other differences between the two embedded complement clause types which allow us to further pin the distinction between extraction strategies to realis versus irrealis status. Section 4 then turns to our analysis of clausal pied-piping.

3.1 Clause type and extraction

We focus here on four of the verbs from section 2.2, shown in the table in (17). Two of these verbs, *al* 'say, tell' and *mulañ* 'like', allow for either realis or irrealis complements, while the other two, *ilañ* 'see' and *om* 'want', combine with only realis and irrealis complements, respectively.

(17)

verb		realis complement?	irrealis complement?
ilañ	'see'	✓	✘
al	'say, tell'	✓	✓
om	'want'	✘	✓
mulañ	'like'	✓	✓

We begin with the unambiguous forms, *ilañ* 'see' and *om* 'want'. In declarative contexts, both combine with complement clauses headed by the complementizer *che'*, and both appear with aspect and person marking.

(18) Tyi k-il-ä [**che' tyi** a-päk'-ä aw-ixim].
PFV A1-see-DTV that PFV A2-plant-TV A2-corn
'I saw that you planted your corn.' (realis complement)

(19) K-om [**che' mi** a-päk' aw-ixim].
A1-want that IPFV A2-plant A2-corn
'I want you to plant your corn.' (irrealis complement)

The examples in (20) illustrate that long-distance extraction of all core embedded argument types—intransitive subject, transitive object, and transitive subject—is possible from the complement of *om*. The complementizer *che'* is consistently judged as either ungrammatical

or dispreferred on the embedded clause any time it is crossed over by a wh-word; we illustrate this explicitly in (20), though it holds consistently for other long-distance extraction examples below.

(20) a. **Maxki**$_i$ aw-om [(*che') mi i-majl-el t_i]?
who A2-want that IPFV A3-go-DEP
'Who do you want to go?'

b. **Chuki**$_i$ aw-om [(*che') mi i-mäñ t_i ajMaria]?
what A2-want that IPFV A3-buy Maria
'What do you want Maria to buy?'

c. **Maxki**$_i$ aw-om [(*che') mi i-mäñ muty t_i]?
who A2-want that IPFV A3-buy chicken
'Who do you want to buy chicken?'

Clausal pied-piping is starkly ungrammatical with the complement of *om*, shown in (21).

(21) a. *[**Maxki**$_i$ mi i-majl-el t_i]$_k$ aw-om t_k?
who IPFV A3-go-DEP A2-want
intended: 'Who do you want to go.'

b. *[**Chuki**$_i$ mi i-mäñ ajMaria t_i]$_k$ aw-om t_k?
what IPFV A3-buy Maria A2-want
intended: 'What do you want Maria to buy?'

c. *[**Maxki**$_i$ mi i-mäñ muty t_i]$_k$ aw-om t_k?
who IPFV A3-buy chicken A2-want
intended: 'Who do you want to buy chicken?'

Turning to realis complements of *ilañ* 'see', we find a different pattern. Extraction of all core arguments triggers pied-piping with inversion of the wh-word, shown in (22). As with long-distance extraction, the complementizer *che'* may not appear anywhere in these examples.

(22) a. [**Maxki**$_i$ tyi majl-i t_i]$_k$ tyi aw-il-ä t_k?
who PFV go-ITV PFV A2-see-DTV
'Who did you see leave?'

b. [**Chuki**$_i$ tyi i-päk'-ä t_i ajJuan]$_k$ tyi aw-il-ä t_k?
what PFV A3-plant-TV Juan PFV A2-see-DTV
'What did you see that Juan planted?'

c. [**Maxki**$_i$ tyi i-päk'-ä ixim t_i]$_k$ tyi aw-il-ä t_k?
who PFV A3-plant-TV corn PFV A2-see-DTV
'Who did you see planted corn?'

All speakers consulted find both pied-piping and long-distance extraction of the wh-word—shown in (23)—grammatical. Two consultants expressed a preference for the pied-piping option in certain cases, but consistently agreed that both possibilities were good. We were unable to discern a pattern in the preference for one choice over another, and further work is needed to determine what factors may determine the choice between the two.

(23) a. **Maxki**$_i$ tyi aw-il-ä [tyi majl-i t_i]?
who PFV A2-see-DTV PFV go-ITV
'Who did you see leave?'

b. **Chuki**$_i$ tyi aw-il-ä [tyi i-päk'-ä t_i ajJuan]?
what PFV A2-see-DTV PFV A3-plant-TV Juan
'What did see you that Juan planted?'

c. **Maxki**$_i$ tyi aw-il-ä [tyi i-päk'-ä ixim t_i]?
who PFV A2-see-DTV PFV A3-plant-TV corn
'Who did you see planted corn?'

Despite the fact that the overt complementizer *che'* is impossible in both types of long-distance wh-dependency, these clauses nonetheless show no signs of being reduced. In addition to aspect marking, negation is also possible, in both pied-piped clauses like (24a), and clauses from which long-distance extraction has occurred, like (24b).

(24) a. [**Maxki**$_i$ ma'añ mi i-mäñ koya' t_i]$_k$ tyi y-äl-ä ajAna t_k?
who NEG IPFV A3-buy tomato PFV A3-say-TV Ana
'Who did Ana say did not buy tomatoes?' (realis complement)

b. **Maxki**$_i$ tyi y-äl-ä ajAna [ma'añ mi i-mäñ lembal t_i]?
who PFV A3-say-TV Ana NEG IPFV A3-buy liquor
'Who did Ana tell to not buy liquor?' (irrealis complement)

To capture the obligatory absence of *che'*, we assume that complementizers which drive wh-movement (both interrogative Cs, as well as embedded Cs responsible for "secondary wh-movement", discussed in §4) are morphologically null.[4]

3.2 Realis versus irrealis clauses

Having established the initial generalization, we now look deeper into the properties of realis and irrealis complement clauses, which will allow us to probe differences in the two verbs which take either realis or irrealis complements. First, however, we note that there is no obvious

[4] See also Vázquez Álvarez & Coon (2020), who note that interrogative wh-words in Ch'ol all end in *ki*, alternating with *ki*-less wh-indefinites (e.g., *chu* 'thing', *chuki* 'what'), and suggest an analysis in which *ki* is a clitic in interrogative C, which cliticizes to the fronted wh-word.

difference in *syntactic size* between the two embedded complement types. Vázquez Álvarez (2011) notes, for example, that fully finite embedded clause can be distinguished from the non-finite embedded clauses examined in section 2.2 by their ability to host fronted focussed constituents, negation, and second-position clitics. The sentences in (25) and (26) show that these properties hold not just for the realis clauses examined in Vázquez Álvarez (2011), but also for irrealis complements.

(25) K-om [che' **ajMariaj$_i$ = äch** mi a-koty-añ t_i].
A1-want that Maria = AFFR IPFV A2-help-TV
'I want you to indeed help Maria$_{[F]}$.'

(26) K-om [che' **ma'añ** mi a-koty-añ ajMaria].
A1-want that NEG IPFV A2-help-TV Maria
'I want you to not help Maria.'

Given the appearance of the complementizer *che'*, a focus position, and obligatory aspect marking, we take these complements to be CPs.

Despite apparent similarity of syntactic size, there are two differences which distinguish between realis and irrealis embedded clauses. First, Ch'ol has a number of second-position clitics, among them the irrealis clitic *=ik* and the realis clitic *=ta'* (Vázquez Álvarez 2011). The irrealis clitic may optionally appear on complements to desiderative verbs which take irrealis complements, like *om* 'want' (27), but is impossible in complements of realis verbs like *il* 'see' (28); *=ta'* has the opposite distribution in embedded clauses.

(27) K-om [che' muk'**(=ik/*ta')** a-päk' aw-ixim].
A1-want that IPFV = IRR/REA A2-plant A2-corn
'I want you to plant your corn.'

(28) Tyi k-il-Dä [che' ta'**(*=ik/ta')** a-päk'-ä aw-ixim].
PFV A1-see-TV that PFV = IRR/REA A2-plant-TV A2-corn
'I saw that you planted your corn.'

Second, while it is the case that non-control/finite complements to *om* 'want' must occur with an aspect marker (as with other finite clauses), only the imperfective aspect is possible, as in (29a). Any other choice of aspect is ungrammatical, as shown with the ungrammatical progressive and perfective forms in (29b) and (29c).

(29) a. K-om [che' *(**mi**) i-mäñ muty ajMaria].
A1-want that IPFV A3-buy chicken Maria
'I want Maria to buy chicken.'

b. *K-om [che' **choñkol** i-mäñ muty ajMaria].
A1-want that PROG A3-buy chicken Maria
intended: 'I want Maria to be buying chicken.'

c. *K-om [che' **tyi** i-mäñ-ä muty ajMaria].
A1-want that PFV A3-buy-TV chicken Maria
intended: 'I want Maria to have bought chicken.'

These differences between realis and irrealis clauses will help us disentangle the realis and irrealis uses of the other two embedding verbs we focus on here, *al* 'say, tell' and *mulañ* 'like'. As shown in (30), matrix *al* can embed realis or irrealis clauses, which may be morphologically identical. As expected, the embedded irrealis form in (30a) is necessarily in the imperfective aspect, while the aspect of the embedded realis form may vary with the context (i.e., whether he bought tomatoes yesterday, is currently buying tomatoes, etc.).

(30) a. **Irrealis context:** You know that Ana has sent her son Pedro shopping to buy tomatoes. You report:
Tyi y-äl-ä ajAna [che' mi i-mäñ koya' ajPedro].
PFV A3-say-TV Ana that IPFV A3-buy tomato Pedro
'Ana told Pedro to buy tomatoes.'

b. **Realis context:** Ana is proud that her son Pedro now regularly goes to the store by himself to buy tomatoes. Later she tells you about this, and you report:
Tyi y-äl-ä ajAna [che' mi i-mäñ koya' ajPedro].
PFV A3-say-TV Ana that IPFV A3-buy tomato Pedro
'Ana said that Pedro buys tomatoes.'

The strategy chosen for wh-extraction disambiguates. In irrealis contexts, long-distance extraction is the only possibility (31a); clausal pied-piping enforces a realis interpretation, (31b).[5]

[5] An anonymous reviewer asks about the possibility of focus and negation in embedded irrealis clauses from which wh-extraction has occurred—i.e., long-distance extraction versions of (25) and (26) above—in order to ensure that irrealis clauses are not reduced in extraction contexts. While it is impossible to long-distance extract from an embedded clause with a focussed element, shown in (ia), this is unsurprising given that foci and fronted wh-words are generally taken to compete for a single left-edge position (§2.1). Assuming the focussed element and the trace of long-distance movement compete for this position, the ungrammaticality of (ia) can be seen as providing support for successive cyclic movement. Extraction from an embedded clause containing negation is, however, possible, as shown in (ib), suggesting that these clauses are not reduced in extraction contexts.

(i) a. *Chuki$_i$ aw-om [**ajMaria** mi i-mäñ-e' t_i]?
what A2-want Maria IPFV A3-buy-DEP
intended: 'What do you want Maria$_{[F]}$ to buy?'

b. Maxki$_i$ aw-om [**ma'añ** mi k-pejk-añ t_i]?
what A2-want NEG IPFV A1-talk.to-DEP
'Who do you want me not to talk to?'

(31) **Irrealis context:** You know that Ana has sent one of her children shopping to buy tomatoes, but you don't know who. You ask:

a. **Maxki**$_i$ tyi y-äl-ä ajAna [mi i-mäñ koya' t_i]?
who PFV A3-say-TV Ana IPFV A3-buy tomato
'Who did Ana tell to buy tomatoes?'

b. #[**Maxki**$_i$ mi i-mäñ koya' t_i]$_k$ tyi y-äl-ä ajAna t_k?
who IPFV A3-buy tomato PFV A3-say-TV Ana
intended: 'Who did Ana tell to buy tomatoes?'

While the sentence in (31b) is infelicitous in the context in (31), it is the preferred response to the question in the realis context in (32). As above, for some speakers the pied-piping option in (32a) is preferred; for others, either pied-piping or long-distance extraction is possible, (32b).

(32) **Realis context:** Ana is proud that one of her sons goes to the store to buy tomatoes and is telling all her friends, but you're not sure which son she's talking about. You ask:

a. [**Maxki**$_i$ mi i-mäñ koya' t_i]$_k$ tyi y-äl-ä ajAna t_k?
who IPFV A3-buy tomato PFV A3-say-TV Ana
'Who did Ana say buys tomatoes?'

b. %**Maxki**$_i$ tyi y-äl-ä ajAna [mi i-mäñ koya' t_i]?
who PFV A3-see-TV Ana IPFV A3-buy tomato
'Who did Ana say buys tomatoes?'

The verb *mulañ* 'like' similarly may take either irrealis or realis complements, as shown in (33). The irrealis complement in (33a) is necessarily imperfective and may appear with the irrealis clitic. The realis form in (33b), in contrast, is not aspectually restricted and *=ik* is impossible.

(33) a. **Irrealis context:** It's getting late in the season and you still haven't planted your corn. I'm getting worried, so I tell you:
Mi k-mul-añ [che' muk'=ik a-päk' aw-ixim].
IPFV A1-like-TV that IPFV=IRR A2-plant A2-corn
'I'd like you to plant your corn.'

b. **Realis context:** I was concerned you weren't going to plant your corn on time, but you finally did and I'm happy about this. I tell you:
Tyi k-mul-ä [che' {tyi, *ta'=ik} a-päk'-ä aw-ixim].
PFV A1-like-TV that PFV PFV=IRR A2-plant-TV A2-corn
'I like that you planted your corn.'

Extraction patterns follow what we saw above. Long-distance extraction in (34a) either favours or forces an irrealis interpretation; clausal pied-piping in (34b) only receives a realis interpretation.[6]

(34) a. **Chuki**$_i$ mi a-mul-añ [mi i-päk' t_i li wiñik]?
who IPFV A2-like-DTV IPFV A3-plant DET man
'What would you like the man to plant?' (✓irrealis)
or: 'What do you like that the man plants?' (✓realis)

b. [**Chuki**$_i$ tyi i-päk'-ä t_i li wiñik]$_k$ tyi a-mul-ä t_k?
what PFV A3-plant-TV DET man PFV A2-like-DTV
'What do you like that the man planted?' (✓realis)

The above examples have all involved the extraction of argument wh-words, but the same patterns are found with adjunct wh-questions, like those in (35). Like its English equivalent, the question in (35a) is ambiguous: *baki* 'where' could be asking either about the time of you saying, or about the time of Ana going. The addition of the complementizer *che'* in (35b) forces a matrix association, compatible with the generalization that clauses from which extraction has taken place may not have overt *che'*. Finally, in (35c) *baki* in the embedded clause triggers clausal pied-piping, and as expected, *baki* can only be interpreted as modifying the embedded verb. The possibility of the realis clitic *=ta'* confirms that this pied-piped clause is realis.

(35) a. **Baki** tyi aw-äl-ä [mi kaje i-majl-el ajAna]?
where PFV A2-say-TV IPFV PROSP A3-go-DEP Ana
'Where did you say Ana will go?' (ambiguous association for *baki*)

b. **Baki** tyi aw-äl-ä [**che'** mi kaje i-majl-el ajAna]?
where PFV A2-say-TV that IPFV PROSP A3-go-DEP Ana
'Where did you say that Ana will go?' (*baki* only modifies matrix V)

c. [**Baki** = ta' mi kaje i-majl-el ajAna]$_k$ tyi aw-äl-ä t_k?
where = REA IPFV PROSP A3-go-DEP Ana PFV A2-say-TV
'Where did you say that Ana will go?' (*baki* only modifies embedded V)

4 Clausal pied-piping

This section offers a closer look at clausal pied-piping and an analysis for the syntactic mechanism underlying it. We begin in section 4.1 with evidence that clausal pied-piping in Ch'ol necessarily involves wh-movement, in contrast with "slifting" constructions in languages like English, and even with apparently more general "inversion" constructions in Popti' Mayan. Section 4.2 provides a brief overview of clausal pied-piping in other languages for which it has been described.

[6] The construction in (34b) is dispreferred to a long-distance extraction equivalent for the speakers consulted, in contrast with the clausal pied-piping constructions above. We do not have an explanation for this difference.

Here we look specifically at evidence that clausal pied-piping in Ch'ol is semantically equivalent to long-distance wh-extraction—and different from wh-scope marking constructions in languages like Hindi and German—following Arregi's (2003) analysis of clausal pied-piping in Basque. Here we propose following Cable (2007) that the semantic vacuity of clausal pied-piping, combined with evidence from patterns involving multiple levels of embedding, provides evidence in favour of a QP approach to pied-piping, and against an analysis involving feature percolation.

4.1 Clausal pied-piping requires wh-movement

Unlike certain inversion constructions in English and Popti', clausal pied-piping in Ch'ol necessarily involves wh-movement (in particular, wh-movement, seen throughout, or focus movement, discussed below). Ross (1973) dubs constructions like the English in (36a) "slifting" ("sentence lifting"), proposing that the initial clause originates as a finite embedded clause; a transformational rule deletes the complementizer and moves the embedded clause to adjoin to the left periphery of the matrix clause. These bear some obvious resemblance to the Ch'ol clausal pied-piping examples above. More recently, Haddican et al. (2014) and Stepanov & Stateva (2016) discuss examples of "interrogative slifting", as in (36b), noting the parallels to clausal pied-piping in Basque.

(36) a. There is something funny about Venus, the astronomers realized. (Ross 1973)
b. Where did John go, do you think? (Haddican et al. 2014)

In a potentially similar vein, Craig (1977) describes certain "inversion" constructions in the Mayan language Popti', which at first glance also look similar to the Ch'ol clausal pied-piping constructions introduced above. In particular, Craig notes that "inversion is preferred when an NP is questioned out of the complement sentence", as shown by the pair of examples in (37), mirroring the Ch'ol preference for clausal pied-piping in long-distance extraction.

(37) a. ?Mach xal naj [chubil xmakni xo']?
who said CLF.he that hit CLF.her
'Who did he say hit her?
b. [Mach xmakni xo'] yalni naj?
who hit CLF.her say CLF.he
'Who did he say hit her? (Popti'; Craig 1977: 263)

In Popti', however, as in English slifting, inversion is *not* restricted to wh-extraction environments, shown in (38). Furthermore, inversion only occurs with three embedding verbs, all involving the root *al* 'say' (an apparent cognate with the Ch'ol root *al* 'say, tell'): *hala* 'to say', *ay -ala* 'to desire', and *ham -alni* 'to think'. Additionally, Craig notes that the inverted clause must be in a reduced aspectless form.

(38) a. Xal naj jet an [tato x'apni ya' kumi'].
said CLF.he to.us 1PL that arrived CL.the lady
'He told us that the lady had arrived.'
b. [X'apni ya' kumi'] yalni naj jet an.
arrived CL.the lady says CLF.he to.us 1PL
'The lady arrived, he told us.' (Craig 1977: 259)

In contrast with English slifting and Popti' inversion, clausal pied-piping in Ch'ol *requires* wh-movement within the embedded clause. Compare the grammatical example in (39a) with the ungrammatical (39b). Here we observe that attempting to move the non-interrogative clause results in ungrammaticality, regardless of the presence or absence of a complementizer.[7]

(39) a. [Chuki tyi i-mäñ-ä ajAna]$_i$ [tyi aw-äl-ä t_i]?
what PFV A3-buy-TV Ana PFV A2-say-TV
'What did you say that Ana bought?'
b. *[(Che') tyi i-mäñ-ä koya' ajAna]$_i$ [tyi aw-äl-ä t_i].
that PFV A3-buy-TV tomato Ana PFV A2-say-TV
intended: 'You said that Ana bought tomatoes.'

Given that wh-movement is obligatory in Ch'ol wh-questions (see §2.1), it is perhaps unsurprising that wh-in situ is ungrammatical within the pied-piped clause. The examples in (40)—involving a pied-piped clause with an in situ wh-word—show this to be true for object and subject wh-questions within the pied-piped clauses, respectively.

(40) a. *[Tyi i-mäñ-ä **chuki** ajAna]$_i$ tyi aw-äl-ä t_i?
PFV A3-buy-TV what Ana PFV A2-say-TV
intended: 'What did you say Ana bought?'
b. *[Tyi i-mäñ-ä koya' **maxki**]$_i$ tyi aw-äl-ä t_i?
PFV A3-buy-TV tomato who PFV A2-say-TV
intended: 'Who did you say bought tomatoes?'

As an instance of wh-movement, we expect clausal pied-piping to be able to cross multiple clauses; this is indeed possible, and an important focus of section 4.3 below. Like other instances of wh-movement in Ch'ol, clausal pied-piping is also subject to island effects, shown for extraction out of an interrogative complement (i.e., a *whether*-island) in (41). The examples in (41b)–(41c) remain ungrammatical regardless of the placement of *mi* 'if'.

[7] The string in (39b) is grammatical (without the complementizer), with a pause appears between the two clauses. Speakers consulted comment that this sounds like two separate sentences: 'Ana bought tomatoes…you said (it)'; a reviewer asks whether this should be likened to English slifting, which seems plausible to us.

(41) a. Tyi j-k'ajty-i [mi tyi i-jap-ä i-kajpe' ajMaria].
PFV A1-ask-DTV if PFV A3-drink-TV A3-coffee Maria
'I asked if Maria drank her coffee.'

b. *$\textbf{Chuki}_i$ tyi a-k'ajty-i [mi tyi i-jap-ä t_i ajMaria]?
what PFV A2-ask-DTV if PFV A3-drink-TV Maria

c. *[$\textbf{Chuki}_i$ mi tyi i-jap-ä t_i ajMaria]$_k$ tyi a-k'ajty-i t_k?
what if PFV A3-drink-TV Maria PFV A2-ask-DTV
intended: 'What did you ask if Maria drank?'

While space prevents us from a detailed look at these effects, we note that in Basque, clausal pied-piping has been shown to repair some, but not all, island violations (Ortiz de Urbina 1989). We save a detailed comparison and understanding of this apparent difference for future work.

Finally, we note that while clausal pied-piping cannot cross an island, it can target the edge of an embedded interrogative CP, as in (42).

(42) Tyi j-k'ajty-i [[$\textbf{chuki}_i$ tyi i-mäñ-ä t_i ajCarol]$_k$ tyi aw-il-ä t_k]
PFV A1-ask-DTV what PFV A3-buy-TV Carol PFV A2-see-DTV
'I asked what you saw that Carol bought?'

4.2 Clausal pied-piping does not affect meaning

In a general survey of pied-piping, Heck (2008: 105) describes clausal pied-piping as a "rather rare phenomenon." Nonetheless, clausal pied-piping has been described in a range of unrelated and typologically diverse languages—including Wolof (Niger-Congo; Torrence 2013), Tlingit (Na-Dene; Cable 2007), Imbabura Quechua (Cole 1982; Hermon 1985), and Basque (Ortiz de Urbina 1989; 1993; Arregi 2003)–where it has been argued to provide evidence for the successive cyclic nature of movement in long-distance wh-dependencies. Common themes include (i) the dependence of clausal pied-piping on wh-movement *internal* to the pied-piped clause—what Heck terms "secondary wh-movement"— and (ii) the absence of a semantic distinction between pied-piped and long-distance extraction alternations (Arregi 2003; Cable 2007; Torrence 2013).

This latter property distinguishes clausal pied-piping from wh-scope marking constructions in languages like Hindi and German, a focus of Arregi (2003). Specifically, contra suggestions in Horvath (1997) and Lahiri (2002), Arregi (2003) argues that clausal pied-piping in Basque has no interpretive consequences, setting it apart from wh-scope marking constructions in languages like Hindi, Hungarian, and German. Arregi (2003) demonstrates specifically that Basque clausal pied-piping involves only a single presupposition, on par with typical long-distance wh-movement constructions, and *unlike* wh-scope marking constructions. Ch'ol behaves identically to Basque, as shown by the examples in (43), modelled after equivalent Basque examples in Arregi (2003). Specifically, either the clausal pied-piping question in (43a) or the long-distance extraction question in (43b) is a natural question given the context.

(43) **Context:** You know that Juan didn't kill anyone, but Maria said that he did. You ask:

a. Maxki tyi i-tsäñ-s-ä ajJuan mi y-äl ajMaria?
who PFV A3-die-CAUS-DTV Juan IPFV A3-think Maria

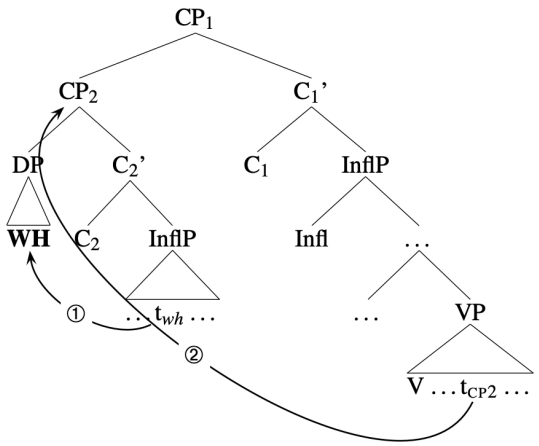
b. Maxki mi y-äl ajMaria tyi i-tsäñ-s-ä ajJuan?
who IPFV A3-think Maria PFV A3-die-CAUS-DTV Juan
'Who does Maria think Juan killed?'

If clausal pied-piping mirrored the behaviour of wh-scope marking, we would expect the pied-piped construction in (43a) to introduce two presuppositions: (i) that Juan killed someone, and (ii) that Maria thinks that Juan killed someone. Following Arregi (2003), given that (43a) is appropriate in a context in which it is clear that Juan hasn't killed anyone, we confirm that Ch'ol behaves like clausal pied-piping in Basque—and unlike wh-scope marking constructions in Hindi and German.

4.3 A QP analysis of clausal pied-piping in Ch'ol

In the sections above we have seen that clausal pied-piping in Ch'ol is: (i) possible only with realis CPs; (ii) that it is dependent on wh-movement within the pied-piped clause; and (iii) that it has apparently no semantic consequences, alternating freely with long-distance wh-extraction. A basic syntactic structure for clausal pied-piping in Ch'ol is provided in (44), following analyses of Basque (Ortiz de Urbina 1989; 1993; Arregi 2003) and Tlingit (Cable 2007) (setting aside QP for now to focus on the basic structure). Specifically, we find movement of the embedded wh-word to the edge of the embedded Spec,CP (CP_2). This may be followed either by continued movement of the wh-word alone to the specifier of the matrix clause, resulting in familiar long-distance extraction (not shown here), or by movement of the entire embedded CP to the specifier of the matrix clause, Spec,CP_1, as in (44).

(44) Clausal pied-piping in Ch'ol



The next question to address is how it is possible for CP_2 to be targeted for extraction, rather than the wh-word itself—the standard issue for pied-piping constructions. We assume that wh-movement is triggered by unvalued [uQ] features on interrogative C. One option for ensuring that CP_2 is targeted rather than the wh-word in its specifier is feature percolation, proposed for example by Ortiz de Urbina (1993) for Basque. Under a percolation account, the wh-word bears the valued [Q] features which ultimately drive wh-movement. To account for clausal pied-piping, the wh-word moves to the specifier of the embedded clause, and then its [Q] feature "percolates" to the phrasal projection which immediately dominates it, CP_2. Now that CP_2 bears the [Q] feature, it is targeted for movement to the specifier of the CP bearing [uQ]. This percolation operation would be optional in Basque and Ch'ol, accounting for the fact that either the wh-word or the embedded clause may be targeted for movement to Spec,CP_1.

An alternative account of pied-piping is proposed by Cable (2007), foreshadowed above, who argues that wh-movement is always mediated by a projection, QP, headed by a focus-sensitive operator, Q. Q is realized as an overt question particle in a language like Tlingit, but covert in other languages, like Ch'ol. QP ultimately raises to check the unvalued [uQ] features on the interrogative C. Optionality in *where* the Q head merges results in variation in pied-piping. In the long-distance extraction of a wh-word, the Q head merges directly with the wh-DP. To account for clausal pied-piping, Cable (2007) proposes that the Q head may instead merge with the embedded CP. Following Cable, secondary movement of the wh-word within the embedded clause is motivated by the need for the embedded wh-word to enter into a local relationship with Q; we assume that secondary wh-movement is driven by an [EPP] feature on the relevant embedded CP which triggers movement of the embedded wh-word to Spec,CP, ensuring this locality (see also Aissen & Polian 2024 on pied-piping with inversion in the DP domain). As argued in detail in Cable (2007), the choice of whether to merge Q directly with the wh-word, or with a higher projection containing the wh-word, does not have semantic consequences, so long as certain locality conditions are met.[8]

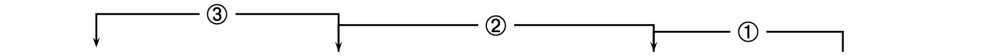
While a simple case of extraction does not provide an immediate way to distinguish between these two possible analyses, an example with multiple embedded clauses, like the sentence in (45), does; see Coon (2009) for analogous discussion in the case of DP-internal pied-piping in Ch'ol. In particular, we examine predictions for extraction of the embedded object *bu'ul* to the matrix clause, in order to form the question equivalent to 'What did you say that Carol liked that Nico planted?'.

[8] Specifically, Cable (2007: ch. 7) argues that wh-words themselves are semantically deficient and have only a focus-semantic value: a set of focus-alternatives (Beck 2006). Q particles are focus-sensitive operators (Rooth 1985), which take focus-semantic values as arguments, closing off the set of focus alternatives. Semantic composition requirements then ensure that one and only one Q must merge with a projection containing a wh-word, a restriction which will become relevant below.

(45) [$_{CP1}$ Tyi awälä [$_{CP2}$ che' tyi imulä ajCarol [$_{CP3}$ che' tyi ipäk'ä **bu'ul** ajNico]]].
PFV A2.say that PFV A3.like Carol that PFV A3.plant beans Nico
'You said that Carol liked that Nico planted his beans.'

Both percolation and QP analyses correctly predict the possibility of extraction of the embedded wh-word alone to the matrix clause, as shown in (46). In a feature percolation account, no feature percolation takes place and the wh-word is directly targeted for movement through the edges of each CP. In a QP account, on the other hand, the Q head (null in Ch'ol) merges directly with the wh-DP, here *chuki* 'what'. QP is then attracted to through the edge of each clause to form the matrix question. (As our aim here is to compare the two analyses, we do not represent either percolation or a QP projection in the examples below, but summarize the required mechanisms for each derivation in (50) below.)

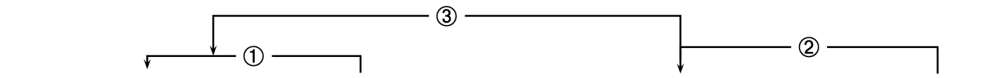
(46)



[$_{CP1}$ **Chuki**$_i$ tyi awälä [$_{CP2}$ t_i tyi imulä ajCarol [$_{CP3}$ t_i tyi ipäk'ä t_i ajNico]]]?
what PFV A2.say PFV A3.like Carol PFV A3.plant Nico
'What did you say that Carol liked that Nico planted?'

Under a QP account of pied-piping, variation in pied-piping can be connected to the selectional properties of Q. In Ch'ol (and Tlingit, and Basque), Q must be able to select CP. A QP analysis thus predicts two additional options for the original sentence in (46): Q could merge with the most embedded CP_3, or Q could merge with the intermediate CP_2. The first possibility is shown in (47). Here, the wh-word *chuki* first undergoes "secondary" movement to the edge of CP_3; as noted above, we assume that secondary movement is driven by edge features, which ultimately ensure that the wh-word enters into local relationship with Q (indicated with ①; see Cable 2007 for details). Next, CP_3, which is dominated by QP, is attracted to the edge of CP_2, and then further extracted to the edge of the matrix clause.

(47)



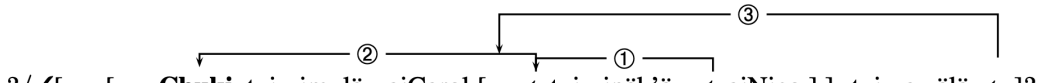
[$_{CP1}$ [$_{CP3}$ **Chuki**$_i$ tyi ipäk'ä t_i ajNico]$_k$ tyi awälä [$_{CP2}$ t_k tyi imulä ajCarol t_k]]?
what PFV A3.plant Nico PFV A2.say PFV A3.like Carol
'What did you say that Carol liked that Nico planted?'
(lit.: 'What did Nico plant did you say that Carol liked?')

A feature percolation account—in which the relevant features begin on the wh-word itself—requires an analysis in which the wh-word first moves to the edge of CP_3 and its [Q] features

percolate. Next, [Q]-bearing CP_3 moves to the edge of CP_2 and no percolation must occur; finally, CP_3 moves to the matrix clause.

For a QP account, the third predicted structure, not seen so far, is shown in (48). Here, the Q head selects the intermediate CP_2. The wh-word *chuki* undergoes secondary movement to the edge of the embedded CP_3, and then further to the edge of CP_2.[9] Next, CP_2, dominated by QP, is attracted to the specifier of of the matrix CP. Though some Ch'ol speakers found this construction more difficult than the options in (46) and (47), indicated here with "?/✓", it was judged to be acceptable (and in clear contrast with the next example discussed below).

(48)

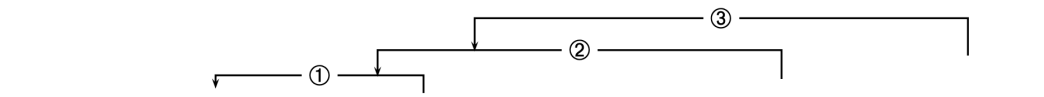


?/✓[$_{CP1}$ [$_{CP2}$ **Chuki**$_i$ tyi imulä ajCarol [$_{CP3}$ t_i tyi ipäk'ä t_i ajNico]]$_k$ tyi awälä t_k]?
what PFV A3.like Carol PFV A3.plant Nico PFV A2.say
'What did you say that Carol liked that Nico planted?'
(lit.: 'What did Carol like that Nico planted did you say?')

Under a feature percolation account of the clause in (48), the wh-word would hold its features after movement to Spec,CP_3, and then [Q] features would percolate after the next step of movement to Spec,CP_2, allowing CP_2 to be attracted to the matrix clause.

We consider next a final option, in which we find successive CP movement: CP_3 moves to the specifier of CP_2, followed by movement of CP_2 to the specifier of CP_1. This would derive the sentence in (49), judged by speakers to be ungrammatical (in contrast with the somewhat degraded (48)). Under a feature percolation account, this configuration would involve successive instances of percolation: first from the wh-word to CP_3, next from CP_3 to CP_2. A QP analysis, on the other hand, is unable to derive this structure. As discussed in detail in Cable (2007), Q is a focus-sensitive operator; more than one Q per question word would result in a semantic clash.

(49)



*[$_{CP1}$ [$_{CP2}$ [$_{CP3}$ **Chuki**$_i$ tyi ipäk'ä t_i ajNico]$_k$ tyi imulä ajCarol t_k]$_j$ tyi awälä t_j]?
what PFV A3.plant Nico PFV A3.like Carol PFV A2.say
intended: 'What did you say that Carol liked that Nico planted?'
(lit.: 'What did Nico plant did Carol like you said?')

[9] Note here that this first instance of movement of the wh-word to the specifier of CP_3 cannot be directly driven by Q, which only merges with the intermediate CP. We assume that wh-movement must proceed cyclically through the edge of CPs, and as noted above, that this secondary movement is driven by [EPP] features. The resulting derivation then only converges if Q is able to establish a relationship with the wh-word, as in Cable (2007).

The table in (50) presents a side-by-side comparison of the possible analyses of the four different derivations considered in (46)–(49). Crucially, under the assumption that Q may select for either DP or CP in Ch'ol, we correctly derive the three possible structures in (46)–(48). Given the impossibility of merging multiple Qs in a construction with a single wh-element, we immediately rule out the ungrammatical (49).[10] Under a percolation account, on the other hand, there is no principled way to rule in the grammatical structures while ruling out the ungrammatical structures. Specifically, a feature percolation account requires (i) that percolation be optional, and (ii) that features percolate from a specifier to the immediately dominating projection. There is then no principled way to rule out the successive percolation which would derive the ungrammatical (49), without the addition of stipulations. For example, one might stipulate that percolation could take place a maximum of one time. This would require the derivation to somehow "count" or keep a record of whether percolation has already taken place, adding unnecessary complexity to the already optional operation.[11]

(50) Side-by-side comparison of analyses

	ex. #	QP	feature percolation
wh-word extraction	(46)	Q merges with wh-DP	✘percolation > ✘percolation
CP_3 extraction	(47)	Q merges with CP_3	✓percolation > ✘percolation
CP_2 extraction	(48)	Q merges with CP_2	✘percolation > ✓percolation
*roll-up CP extraction	(49)	✘	✓percolation > ✓percolation

[10] A reviewer asks about configurations in which a wh-possessor pied-pipes its possessum, and in turn triggers clausal pied-piping. Such constructions are indeed possible in Ch'ol, as shown in (i). Here, the wh-possessor has moved to the left edge of the possessive phrase (recall from §2.1 that non wh-possessors follow the possessum) and pied-piped the possessum to the left edge of the embedded clause. The entire embedded clause is then pied-piped to the left edge of the matrix clause.

(i) [[Maxki$_j$ i-mama t_j]$_i$ tyi majl-i t_i]$_k$ tyi aw-äl-ä t_k?
who A3-mom PFV go-ITV PFV A2-say-TV
'Whose mom did you say left?'

On the face of it, such constructions might be assumed to require two Qs: one merged with the possessive phrase, driving movement to the left edge of the embedded clause, and a second merged with the embedded CP, triggering clausal pied-piping. Here we suggest instead that movement of *maxki* 'who' to the left edge of the possessive phrase, and movement of the possessive phrase to the left edge of the embedded clause are *both* instances of secondary wh-movement. This is in line with both Cable (2007) and Aissen & Polian (2024), who assume that the instances of apparent subextraction of wh-possessors discussed in Coon (2009)—which require optionality in merge of a QP internal to possessive phrases—in fact do not involve subextraction of the wh-possessor, but rather alternative constructions in which the wh-possessor is base-generated external to the possessive phase.

[11] See Moro et al. (2003) and subsequent work, discussed in Kayne (2022), on the absence of *counting* in syntactic operations. We thank a reviewer for pointing out the relevance of this work.

We thus argue that clausal pied-piping in Ch'ol provides additional evidence in favour of a QP approach to pied-piping.

We now return to the question of why clausal pied-piping is *ungrammatical* with irrealis complements. Recall from section 3.2 above that despite no obvious difference in syntactic size, irrealis CPs are *aspectually restricted*, appearing only with imperfective aspect. We propose that something about the defective nature of irrealis CPs makes them unselectable by Q. At the moment, this remains a stipulation, though we note it correctly derives the empirical patterns seen so far. We believe a fruitful avenue to pursue would be that there is something incompatible with the *semantic type* of the irrealis. Arregi (2003: 141), for example, argues that pied-piped CPs must reconstruct at LF, and that this requirement could be derived "from a strong condition on the interpretation of traces, which only allows them to be interpreted as variables over individuals." If irrealis clauses are not semantically *individuals*, then we would correctly rule out the possibility of pied-piping. We leave a full semantic analysis for future work. We note that the restriction that Q in Ch'ol selects for wh-DPs or (realis) CPs correctly rules out pied-piping of the non-finite embedded clauses examined in section 2.2 above.

Finally, we argue that a QP approach provides a better means to account for cross-linguistic variation in the availability of clausal pied-piping than a feature percolation account, and further, that Ch'ol provides some important data points regarding the typology of this construction. Specifically, we turn here to the question of why clausal pied-piping is possible in Ch'ol and Basque, but not in English and Spanish. Ortiz de Urbina (1993: 216) suggests two possible correlations, neither of which work for Ch'ol. At the time of Ortiz de Urbina's writing, the descriptions of clausal pied-piping involved SOV languages, Basque and Quechua, which he suggested as a possible correlation. Verb-initial Ch'ol obviously problematizes this possibility. A second, and more promising possibility, Ortiz de Urbina notes, is that while wh-words may not cooccur with overt complementizers in English and Spanish, they are obligatory in Basque embedded clauses. Specifically, Ortiz de Urbina (1993) suggests that the null complementizer in languages like English and Spanish is subject to constraints governing empty categories, following Stowell (1981), ruling out clausal pied-piping. Again, however, Ch'ol shows us that this story can't be right at a cross-linguistic level: the complementizer *che'* is ungrammatical in both long-distance wh-extraction and clausal pied-piping environments (see e.g., (22) above). The variation between Ch'ol and Basque in this respect suggests this should not relate to a deep property of clausal pied-piping.

Under a QP analysis, variation in the availability of clausal pied-piping is attributed to variation in the selectional properties of the Q head (in addition to restrictions noted in Cable 2007). Thus, while we do not arrive at an answer to *why* Q selects CP in Ch'ol but not in English, selectional variation is independently needed. Under a feature percolation account,

however, variation must be attributed to the availability of a core syntactic operation, a path we suggest—following extended discussion in Cable (2007) and subsequent work—adds unnecessary complication to the grammar.

Finally, though this paper has focused on clausal pied-piping driven by interrogative wh-words, recall that wh-questions, focus, and relativization all involve movement to the left edge of the clause, and are typically taken to be in competition for a single left-edge position. Focussed elements may also trigger clausal pied-piping with inversion, also described by Arregi (2003) for Basque. Specifically, the question in (51a) is felicitously answered by the sentence in (51b), in which the focussed subject has undergone secondary movement to the edge of the embedded clause, followed by pied-piping of the embedded clause to the matrix clause. Leaving the focussed element in situ within the fronted clause in (51b) results in ungrammaticality (unless a pause is present, in which case the configuration can be understood to involve two sentences; see fn. 7).

(51) a. [**Maxki**$_i$ mi i-mäñ koya' t_i]$_k$ tyi aw-äl-ä t_k?
who IPFV A3-buy tomato PFV A2-see-TV
'Who did you say buys tomatoes?'

b. [**Jiñ-äch alob**$_i$ mi i-mäñ koya' t_i]$_k$ tyi k-äl-ä t_k.
DET-AFFR boy IPFV A3-buy tomato PFV A1-see-TV
'I said the boy$_{[F]}$ buys tomatoes.'

Relativization may similarly trigger clausal pied-piping, at least for some speakers. An example of long-distance extraction for relativization is shown in (52a). Some speakers found the option with clausal pied-piping in (52b) good, while others found it degraded.

(52) a. Tyi majl-i jiñi [**alob**$_i$ ta'=bä aw-äl-ä [tyi i-mäñ-ä koya' t_i]].
PFV go-ITV DET boy PFV=REL A2-say-TV PFV A3-buy-TV tomato

b. %Tyi majl-i jiñi [**alob**$_i$ ta'=bä i-mäñ-ä koya' t_i]$_k$ tyi aw-äl-ä t_k.
PFV go-ITV DET boy PFV=REL A3-buy-TV tomato PFV A2-say-TV
'The boy who you said bought tomatoes left.'

We suggest that this variation could be related to a processing issue, having to do with complexity of the construction, but we save further investigation—as well as a more detailed look at focus and relativization in the domain of pied-piping—for future work.[12]

[12] Interestingly, all speakers consulted disliked pied-piping with inversion for *possessors*; this is in stark contrast with wh-possessors, which necessarily invert within the possessive phrase in Ch'ol, and also in contrast with Tsotsil, where Aissen (1996) documents pied-piping with inversion for focussed possessors.

5 Conclusion

This paper examined patterns of clausal embedding, long-distance extraction, and clausal pied-piping in Ch'ol Mayan. We began in section 2 with a general introduction to patterns of embedding in the language, focussing on extraction patterns from embedded CPs in section 3. Section 3 established the core empirical generalization: clausal pied-piping is optional (alongside long-distance extraction) when the wh-word is embedded in a *realis* CP, but prohibited for *irrealis* CPs. We examined the two clause types further, noting that irrealis CPs are aspectually deficient.

Section 4 turned to a more in-depth look at clausal pied-piping, situating Ch'ol within the (limited) cross-linguistic context in which these constructions have been discussed, and distinguishing clausal pied-piping in Ch'ol from other types of inversion constructions, as well as from wh-scope marking. Clausal pied-piping—in Ch'ol, and in the other languages for which it has been described—involves (i) obligatory secondary wh-movement of an embedded wh-word; and (ii) no detected semantic distinctions. Finally, through the examination of complex embedding configurations, we argued that Cable's (2007) QP analysis of pied-piping provides superior empirical coverage to a feature percolation account.

This paper thus contributes both to the empirical landscape of long-distance extraction patterns—an under-documented but fruitful area of investigation in Mayan languages—as well as to theoretical approaches to patterns of pied-piping. We leave open several areas for future investigation, including details of the semantics of embedded irrealis clauses, as well as questions about difference between focus movement and wh-movement. Finally, we note that Aissen & Polian (2024) identify information-structural differences between different extraction options within the DP domain in related Tsotsil and Tseltal, and it would be interesting to compare their discussion with our findings here.

Abbreviations

Glosses in follow Leipzig Glossing Conventions with the following Mayan-specific additions: A – "Set A" (ergative, possessive); AFFR – affirmative; B – "Set B" (absolutive); DEP – dependent suffix; DTV – derived transitive status suffix; ITV – intransitive status suffix; REA – realis; TV – transitive status suffix. In some cases, glosses from other sources have been modified for consistency.

Ethics and consent

Research with Ch'ol speakers presented here was approved by McGill University Research Ethics Board II as File #75-0716 *Mayan Grammar in Linguistic Theory: Documentation and Analysis*, Principal Investigator Jessica Coon.

Funding information

This work was funded in part by a SSHRC Insight Grant to the first author and PAPIIT funding (IA402023) to the second author.

Acknowledgements

Special thanks to Morelia Vázquez Martínez, Virginia Martínez Vázquez, José Ramírez López, and María Vázquez Martínez for Ch'ol judgments and discussion. We are grateful to three anonymous reviewers for extensive and helpful comments. Thanks also to Judith Aissen, Karlos Arregi, Chase Boles, Robert Henderson, Carol-Rose Little, Martina Martinović, Rodrigo Ranero, Justin Royer, Michael Wagner, McGill's syntax-semantics reading group, and participants at FAMLi VII. Any errors in data or interpretation are our own.

Competing interests

The authors have no competing interests to declare.

References

Aissen, Judith. 1992. Topic and focus in Mayan. *Language* 68(1). 43–80. DOI: https://doi.org/10.1353/lan.1992.0017

Aissen, Judith. 1996. Pied-piping, abstract agreement, and functional projections in Tzotzil. *Natural Language and Linguistic Theory* 14(3). 447–491. DOI: https://doi.org/10.1007/BF00133596

Aissen, Judith. 2017a. Complement clauses. In Aissen, Judith & England, Nora & Zavala Maldonado, Roberto (eds.), *The Mayan languages*, 259–292. New York: Routledge. DOI: https://doi.org/10.4324/9781315192345-10

Aissen, Judith. 2017b. Correlates of ergativity in Mayan. In Coon, Jessica & Massam, Diane & Travis, Lisa (eds.), *Oxford handbook of ergativity*. New York, NY: Oxford University Press. DOI: https://doi.org/10.1093/oxfordhb/9780198739371.013.30

Aissen, Judith & Polian, Gilles. 2024. Possessor extraction and categorical subject in Tseltalan. *Natural Language and Linguistic Theory*. DOI: https://doi.org/10.1007/s11049-024-09624-6

Arregi, Karlos. 2003. Clausal pied-piping. *Natural Language Semantics* 11. 115–143. DOI: https://doi.org/10.1023/A:1024499226731

Beck, Sigrid. 2006. Intervention effects follow from focus interpretation. *Natural Language Semantics* 14(1). 1–56. DOI: https://doi.org/10.1007/s11050-005-4532-y

Cable, Seth. 2007. *The grammar of Q: Q-particles and the nature of Wh-fronting, as revealed by the Wh-questions of Tlingit*. Cambridge, MA: Massachusetts Institute of Technology dissertation.

Can Pixabaj, Telma Angelina. 2020. *Las oraciones de complemento en k'iche'*. México: Centro de Investigaciones Multidisciplinarias sobre Chiapas y la Frontera Sur.

Clemens, Lauren Eby & Coon, Jessica. 2018. Deriving verb initial order in Mayan. *Language* 94(2). 237–280. DOI: https://doi.org/10.1353/lan.2018.0017

Cole, Peter. 1982. *Imbabura Quechua*. Amsterdam: North-Holland Publishing Company.

Coon, Jessica. 2009. Interrogative possessors and the problem with pied-piping in Chol. *Linguistic Inquiry* 40(1). 165–175. DOI: https://doi.org/10.1162/ling.2009.40.1.165

Craig, Colette Grinevald. 1977. *The structure of Jacaltec*. Austin, TX: University of Texas Press.

Cristofaro, Sonia. 2003. *Subordination*. Oxford: Oxford University Press.

Erlewine, Michael Yoshitaka. 2016. Anti-locality and optimality in Kaqchikel Agent Focus. *Natural Language and Linguistic Theory* 34. 429–479. DOI: https://doi.org/10.1007/s11049-015-9310-z

Ewing, Caleb. 2022. *Pied-piping and wh-movement: A syntactic analysis and comparison of Kaqchikel and K'iche'*. Gainesville: University of Florida dissertation.

Haddican, Bill & Holmberg, Anders & Tanaka, Hidekazu & Tsoulas, George. 2014. Interrogative slifting in English. *Lingua* 138. 86–106. DOI: https://doi.org/10.1016/j.lingua.2013.10.006

Heck, Fabian. 2008. *On pied-piping: Wh-movement and beyond*, vol. 98 (Studies in Generative Grammar). Berlin: Mouton de Gruyter. DOI: https://doi.org/10.1515/9783110211467

Hermon, Gabriela. 1985. *Syntactic modularity*. Dordrecht: Foris Publications. DOI: https://doi.org/10.1515/9783110849141

Horvath, Julia. 1997. The status of 'wh-expletives' and the partial wh-movement construction of hungarian. *Natural Language & Linguistic Theory* 15(3). 590–572.

Kayne, Richard S. 2022. Antisymmetry and externalization. *Studies in Chinese Linguistics* 43(1). DOI: https://doi.org/10.2478/scl-2022-0001

Lahiri, Utpal. 2002. On the proper treatment of "expletive wh" in Hindi. *Lingua* 112(7). 501–540. DOI: https://doi.org/10.1016/S0024-3841(01)00059-6

Little, Carol-Rose. 2020. Left branch extraction, object shift, and freezing effects in Tumbalá Ch'ol. *Glossa: A Journal of General Linguistics* 5(1). 26. DOI: https://doi.org/10.5334/gjgl.988

Mendes, Gesoel & Ranero, Rodrigo. 2021. Chain reduction via substitution: Evidence from Mayan. *Glossa: A Journal of General Linguistics* 6. 10. DOI: https://doi.org/10.5334/gjgl.1087

Moro, Andrea & Musso, Mariacristina & Glauche, Volkmar & Rijntjes, Michel & Reichenbach, Jürgen & Büchel, Christian & Weiller, Cornelius. 2003. Broca's area and the language instinct. *Nature and Neuroscience* 6. 774–781. DOI: https://doi.org/10.1038/nn1077

Noonan, Michael. 2007. Complementation. In Shopen, Timothy (ed.), *Language typology and syntactic description*, vol. 2, 55–150. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511619434.002

Norman, William. 1977. Topic and focus in Mayan. Paper presented at the Mayan Workshop II, San Cristóbal de las Casas, Chiapas, Mexico.

Ortiz de Urbina, Jon. 1989. *Parameters in the grammar of Basque*. Dordrecht: Foris. DOI: https://doi.org/10.1515/9783110876741

Ortiz de Urbina, Jon. 1993. Feature percolation and clausal pied-piping. In Hualde, José Ignacio & Ortiz de Urbina, Jon (eds.), *Generative studies in Basque linguistics*, 189–219. John Benjamins Publishing Company. DOI: https://doi.org/10.1075/cilt.105.08urb

Rooth, Mats. 1985. *Association with focus*. Cambridge, MA: Massachusetts Institute of Technology dissertation.

Ross, John R. 1973. Slifting. In Gross, Maurice & Halle, Morris & Schützenberger, Marcel-Paul (eds.), *The formal analysis of natural languages*, 131–169. The Hague: Mouton. DOI: https://doi.org/10.1515/9783110885248-009

Stepanov, Arthur & Stateva, Penka. 2016. Interrogative slifting: More syntactic, less parenthetical. *Glossa: A Journal of General Linguistics* 30. 1–33. DOI: https://doi.org/10.5334/gjgl.117

Stowell, Tim. 1981. *Origins of phrase structure*. Cambridge, MA: MIT Doctoral dissertation.

Torrence, Harold. 2013. A promotion analysis of Wolof clefts. *Syntax* 16(2). 176–215. DOI: https://doi.org/10.1111/synt.12000

Vázquez Álvarez, Juan J. 2011. *A grammar of Chol, a Mayan language*. Austin, TX: University of Texas Austin dissertation.

Vázquez Álvarez, Juan J. 2013. Dos tipos de cláusulas no finitas en chol. In Palancar, Enrique L. & Zavala, Roberto (eds.), *Estudios sintácticos en lenguas de Mesoamérica*. México: CIESAS.

Vázquez Álvarez, Juan J. & Coon, Jessica. 2020. Headless relative clauses in Ch'ol. In Caponigro, Ivano & Torrence, Harold & Maldonado, Roberto Zavala (eds.), *Headless relative clauses in languages of Mesoamerica*. Oxford University Press. DOI: https://doi.org/10.1093/oso/9780197518373.003.0011

Velleman, Leah Bridges. 2014. *Focus and movement in a variety of K'ichee'*. Austin, TX: University of Texas Austin dissertation.